



Studying topic engagement and synergy among candidates for 2020 US Elections

Manmeet Kaur Baxi¹ · Rajesh Sharma² · Vijay Mago¹

Received: 6 July 2022 / Revised: 13 August 2022 / Accepted: 21 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

This article provides a comprehensive summary of how candidates running in the 2020 US Presidential Elections used Twitter to communicate with the public. More specifically, it aims to uncover elements linked to public engagement and internal cooperation (in terms of content and stance similarity among the candidates from the same political front, and with respect to the official Twitter accounts of their political parties). Our main subjects are the Presidential and Vice-Presidential candidates who contested for the 2020 US Elections from the two major political fronts—Republicans and Democrats. Their tweets were evaluated for social reach, content similarity and stance similarity on 22 topics. According to the findings, Joe Biden had the highest engagement and impact (user impact: 177.08k, normalized to 0.99), followed by Donald Trump (user impact: 164.19k, normalized to 0.92). The Democrats depicted a clearer understanding of their audience, portraying an essential link between public participation, internal cooperation and the electoral campaign. The results also demonstrate that specific topics (like *US Elections*, and *Inauguration Ceremony*) were more engaging than others (*Trump Healthcare Plan*, and *The Supreme Court Appointments*). This study adds to the existing work on using social media platforms for electoral campaigns and can be effectively utilized by contesting candidates.

Keywords Social media · Electoral campaigns · Public engagement · Content similarity · 2020 US elections

1 Introduction

Social media platforms (SMPs) like Facebook, Twitter and Instagram have become the conventional modes of online campaigns for elections after being first used by Barack Obama while contesting for his 2008 candidacy (Effing et al 2011). Researchers from around the world have expressed a strong interest in analyzing and evaluating social media data in the context of elections across these different platforms, as discussed in the works of Borah (2016); Russmann

and Svensson (2017), and Vesnic-Alujevic and Van Bauwel (2014).

Various politicians have widely used SMPs to express their views on current topics, share the latest developments in their constituencies, and communicate with their potential voters strategically. From the statistics in the report by Pew Research Center (2020), out of the top 10% adult Twitter users in the US, 92% of them are politicians, with Democrats or Democratic-leaning independents being hyperactive and capturing 69% positions in the top Twitter users; while Republicans or Republic-leaning independents occupying the remaining 26%. Additionally, recent studies have shown that maintaining an active presence on SMPs has helped politicians address social concerns and build a stronger relationship with the audience (Bonsón et al. 2019; Gruzd et al. 2018; Sahly, Shao and Kwon 2019). Furthermore, the cooperation among the candidates from the same political front has helped them communicate their policy initiatives clearly and organize support from the related interest groups (Grossmann 2014). The authors of (Wonka and Haunss 2020) emphasize the benefits of cooperation among the political parties and interest groups in European Union

✉ Manmeet Kaur Baxi
mbaxi@lakeheadu.ca

Rajesh Sharma
rajesh.sharma@ut.ee

Vijay Mago
vmago@lakeheadu.ca

¹ Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada

² Institute of Computer Science, University of Tartu, Tartu, Estonia

policy-making by examining their information networks. Hence, there is a requirement to investigate the influence of citizens' engagement and the internal cooperation among the politicians on the election results. The authors of Bonsón et al (2012); Norris and Reddick (2013) have stressed the importance of a future qualitative study to quantify the genuine impact of social media on Government to Citizen (G2C) interactions. As a result, the goal of this research is to measure the utility of Twitter for politicians during the 2020 Presidential elections in the United States. In particular, we investigate the following two research questions:

1.1 Research Question 1 —What topics did the candidates discuss through online (Twitter) and offline (Presidential debates) mediums? How engaging were these topics and to what extent during the different phases of the electoral campaign?

Understanding what information and topics appeal to the audience the most, may be an effective method for gaining attention and increasing involvement (Bonsón et al 2015). Bonsón and Bednárová (2018) found that certain types of contents are more engaging than others. Therefore, identifying such materials and developing thorough plans for maintaining a continuous dialog with the citizens, responding to their grievances, recommendations and wants, would help improve governance quality. Following the identification of objectives and goals, norms could be produced to facilitate the contesting candidates with an effective tool for communication on SMPs. If implemented properly, it is indeed a win–win situation for both the candidates and the citizens. On that account, we calculate the impact of candidates and the engagement received by them on various topics, followed by classifying them according to the topic stickiness in different phases of the election campaign.

1.1.1 Findings

Joe Biden was the most impactful candidate among all, and Democrats tweeted more about topics of public interest during the electoral campaign as compared to Republicans. The detailed observations can be found in Sect. 4.

1.2 Research Question 2—Did the candidates from the same political front have similarities in their tweets and the stance for the topics with respect to their political front ?

Politicians collaborate to share resources and coordinate political support. The various types of cooperation networks formed inside a political front during the European Union policy-making have been highlighted by Wonka and Haunss

(2020). Furthermore, the smaller networks inside a political party or interest group may reconfigure themselves based on the reputation (impact of the candidates) and the internal reciprocity (similarity in thoughts/actions) (Balliet et al. 2014; Gallo and Yan 2015), and Gross and De Dreu (2019). Thus, to analyze the synergy among the candidates during the electoral campaign of 2020 US Elections, we employ two methods, i.e., content similarity-based on the tweets, and stance similarity-the standpoint of candidates with respect to different topics. Understanding these aspects would help us identify which political front was more cooperative among themselves and echoed similar thoughts on Twitter.

1.2.1 Findings

Kamala Harris depicted a higher amount of cooperation with both Joe Biden and the official account of Democrats in both — content and stance. On the other hand, Republicans portrayed comparatively lower synergy in their stance with respect to different topics. Refer to Sect. 5 for more details.

This research provides a comprehensive analysis of the contesting candidates, a combination of both qualitative and quantitative insights into their online behavior. Previously, this form of hybrid research has proven to be beneficial (Sampieri 2018). Our study uses statistical methods to examine the public metrics of candidates' tweets (the number of likes, replies, retweets, and quotes) and determine their social reach. It is qualitative as we infer the topics the candidates tweeted about and the similarity in the content and stance of the candidates from the same political front. Therefore, the objectives of this study are to discover the topics discussed by the candidates through online (Twitter) and offline (Presidential debates) channels, as well as the civic engagement on these topics during the different phases of the electoral campaign, and throughout the whole election campaign. Additionally, we also investigate the similarities in the tweets of the candidates from the same political front with respect to the tweet content and their stance on the topics. Furthermore, we try to uncover any relationships between public engagement and internal cooperation (content and stance similarities) that might have aided the candidates in contesting the 2020 US Presidential elections. Figure 1 presents an overview of the research framework.

The remaining part of the article is arranged as follows: Section 2 gives a summary of prior work linked to the use of SMPs in the electoral campaigns. Section 3 explains the data utilized in this study. The techniques used and insights of public engagement during the electoral campaign are presented in Sect. 4. Section 5 discusses the approach followed and observations for the identifying the collaboration among the candidates from the same political front. Finally, the key inferences and further research directions are presented in Sect. 6.

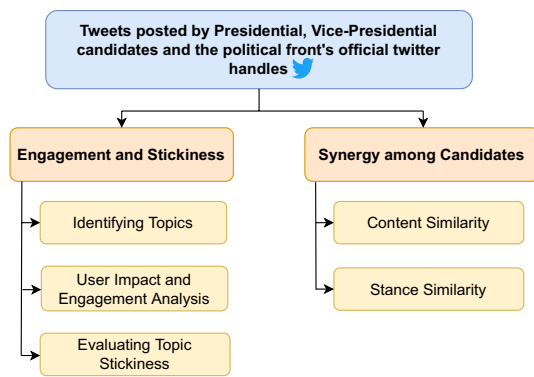


Fig. 1 Overall research framework

2 Related work

The analysis on elections has been widespread across different SMPs, like Bruns and Moon (2018) compared the candidate and audience activity on Twitter for the 2013 and 2016 federal elections in Australia. Additionally, the role of SMPs in enhancing democratic participation during the 2012 and 2016 Ghana elections was studied by Dzisah (2018). Praznik et al (2021) took a different approach by analyzing the strength of networks based on the usage of hashtags on Twitter during electoral campaigns (Praznik et al 2021). Gaisbauer et al. (2021) tried to contrast different opinion groups using network representations for replies and retweets in the context of Saxon state elections and violent riots in the city of Leipzig, Germany, in 2019, and Bilal et al (2019)'s work surveyed the current state-of-the-art approaches to analyze the election prediction mechanisms in use (Bilal et al 2019). As far as SMPs are concerned, Borah (2016) investigated the use of Facebook for campaign strategies used in 2008 and 2012 US Presidential elections, Russmann and Svensson (2017) analyzed Instagram for Swedish elections, and Vesnic-Alujevic and Van Bauwel (2014) studied the use of YouTube as an advertising tool during the campaign of European Parliament elections.

According to previous studies (Bertot et al. 2010; Bonsón et al 2012; Chun et al 2010), SMPs can help enhance the transparency, involvement, and correspondence in governance. Also, researchers have revealed the influence on different features of public interaction through various instances (Gruzd and Roy 2016; Hollebeek et al. 2014; Ríos et al. 2017). Several authors (Bonsón et al. 2019; Bonsón and Ratkai 2013; Bonsón et al. 2017; Gruzd et al. 2018; Sahly, Shao and Kwon 2019; Siebers et al. 2019) have highlighted the relevance of SMPs as a vital instrument for amplifying social reach, and to help understand the audience better. However, earlier studies have also found that the sentiment, emotion, stance of the tweets, the promotion of tweets by bots; and collaboration between interest groups (polarization) may

mitigate the impact of different aspects of public involvement and steer the change in public opinions (Bhat et al 2021; Galgoczy et al. 2022; Grover et al. 2019; Sandhu et al. 2019; Sandoval-Almazan and Valle-Cruz 2018).

Internal cooperation and collaboration play an essential role for a political party to convey their policy initiatives during the electoral campaigns and gather support. The survey by Khanam et al. explores various methodologies proposed till date on the usage of the *Homophily* principle (likelihood of similar-minded people to engage with one another in communities) across different domains (Khanam et al 2022). Another survey by Chandrasekaran and Mago (2021) lists the different methods available to evaluate the semantic similarity between texts, ranging from traditional Natural Language Processing (NLP) techniques to deep neural-network-based hybrid methods (Chandrasekaran and Mago 2021), out of which, we alter one method as per the objectives of this study, to measure the internal cooperation among candidates based on their tweets. However, several factors affect the conflicts and synergy within a group as highlighted in the studies (Gross and De Dreu 2019; Larson 2021; Madeo and Mocenni 2020; Perkoski 2019). It is important to identify them for effective operations and governance. Our study evaluates the cooperation among the candidates by comparing the content of their tweets and collating the similarities in their stance on different topics. We also uncover the relationship between internal synergy and electoral campaigns.

Formerly, Bansal and Srivastava (2018); Joseph (2019); Nugroho (2021), and Tsai et al (2019) have performed sentiment analysis on tweets using machine learning techniques, like lexicon-based models-VADER, and decision trees to predict the election results by focusing on a single aspect. Additionally, Chen et al (2021) have released a dataset for analyzing the 2020 US Elections. However, there has not been an empirical analysis to understand and examine the utility of Twitter (with emphasis on public participation and internal cooperation) as a communication tool during the 2020 US electoral campaign, considering different factors like social reach and internal cooperation (stance and content similarity), which is hence the focus of this work.

3 Dataset

We collected a total of 117,217 tweets authored from the accounts of Presidential and Vice-Presidential candidates of two major political fronts—Republicans and Democrats using the Twitter API v2¹, during the time frame of January 21, 2019, to January 27, 2021. Additionally, the tweets created by the official Twitter handles of both the

¹ <https://developer.twitter.com/en/docs/twitter-api>

Table 1 Tweet distribution of the candidates selected from both the political fronts

Political party	Candidates (Twitter handle)	Number of tweets
Democrats	<i>Presidential Candidate</i> : Joe Biden (@JoeBiden)	5,486
	<i>Vice-Presidential Candidate</i> : Kamala Harris (@KamalaHarris)	5,835
	<i>OfficialDemocrats</i> (@TheDemocrats, @HouseDemocrats, @SenateDems)	41,728
	<i>Total</i>	53,049
Republicans	<i>Presidential Candidate</i> : Donald J. Trump (@realDonaldTrump, @POTUS)	21,007
	<i>Vice-Presidential Candidate</i> : Mike Pence (@Mike_Pence, @VP, @VP45)	12,003
	<i>OfficialRepublicans</i> (@GOP, @HouseGOP, @SenateGOP)	31,158
	<i>Total</i>	64,168

The italicized text signifies the rank of contestants for 2020 US Presidential Elections along with the total number of tweets for each political front

political fronts were also collected. The candidates selected for our analysis and the number of tweets scraped from their accounts are discussed in Table 1. For the scope of our research, the political fronts (Democrats/Republicans) confine to the Presidential candidate, Vice-Presidential candidate, and the official Twitter handles of the political fronts. The official Twitter handles of the political fronts are referred to as ‘*OfficialDemocrats*’ and ‘*OfficialRepublicans*’ for Democrats and Republicans, respectively.

4 Engagement and stickiness of topics

4.1 Identifying topics

We analyze the most and least discussed topics by the candidates from the political fronts through two sources—*offline* and *online* as defined below:

1. The *offline* source is the topics that were discussed in the Presidential debates by both the candidates as given by ‘The Commission on Presidential Debates’² and the events synchronous with the US Elections (Current/Snapshot events). The Presidential debates hosted by The Commission on Presidential Debates offer equal opportunities for candidates to express their opinions on the pressing issues (or political agendas), whereas the TV/YouTube commercials and Zoom rallies could vary according to the availability of resources a candidate has access to, and
2. The *online* source of topics is topic modeling on the tweets authored by the candidates. We first preproc-

ess and then cluster the tweets using various clustering algorithms and leverage the topics yielded by the best-performing topic model.

Preprocessing: Firstly, all the non-alphabets (numbers, punctuation, new-line characters and extra spaces) were removed from the text using the regular expression module (*re 2.2.1*). Then, the text was tokenized using *nlTK 3.2.5*, followed by the removal of stopwords. Also, tiny words (i.e., words with a length of fewer than three characters) were removed from the text. This was followed by stemming the text using *PorterStemmer* and lemmatizing it using the *WordNetLemmatizer* from *nlTK*.

Topic Modeling: Researchers have relied on Term Frequency-Inverse Document Frequency (TF-IDF) for generating document embeddings for short-text (Lilleberg et al 2015; Sari and 2016). Tweets are categorized as short texts³, and after preprocessing them, we generate document embeddings using TF-IDF and then pass them to five different clustering algorithms, namely Latent Dirichlet Allocation (LDA), Parallel LDA, Non-negative matrix factorization (NMF), Latent Semantic Indexing (LSI), and Hierarchical Dirichlet Process (HDP), to generate topic clusters. Due to the short and noisy nature of the data, we ran these models five times over the data with varying random seeds. We check the coherence scores of topic models based on words, the ‘*c_umass*’ (Newman et al 2010) and ‘*c_v*’ (Röder et al 2015) measure, to confirm the performance consistency over multiple runs and finally use the best model to extract the top five topics.

We used Latent Dirichlet Allocation (LDA)⁴ and LDA multi-core⁵(Parallel LDA) provided by Gensim.

² <https://www.debates.org/2020/09/22/moderator-announces-topics-for-first-presidential-debate-2/>, <https://www.debates.org/2020/10/16/moderator-announces-topics-for-oct-22-presidential-debate/>

³ <https://developer.twitter.com/en/docs/counting-characters>

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

⁵ <https://radimrehurek.com/gensim/models/ldamulticore.html>

Table 2 Model parameters for topic clustering with TF-IDF document embeddings

Clustering Algorithm	Epochs	Chunk size	Workers (number of CPU cores)	Evaluation Period (seconds)	Alpha (A-priori belief on document-topic distribution)	Eta (A-priori belief on topic-word distribution, also known as beta)	Kappa (gradient descent step-size)	Minimum normalizing probability
LDA	205	1000	NA	10	0.01	0.9	NA	NA
Parallel LDA	205	1000	7	10	0.01	0.9	NA	NA
LSI	NA	1000	NA	NA	NA	NA	NA	NA
NMF	205	1000	NA	10	NA	NA	1	0
HDP	NA	1000	NA	NA	0.01	NA	1	NA

Non-negative matrix factorization (NMF) model⁶ uses the online NMF proposed in Zhao and Tan (2016) for large corpora. Latent Semantic Indexing (LSI) model⁷ implements fast truncated SVD (Singular Value Decomposition). And, for HDP⁸, we use the improved online variational inference model proposed in Wang et al (2011). The details of parameters used for each of the models are listed in Table 2, and performance for each clustering algorithm in terms of their coherence scores ('c_v' and 'c_umass') and the amount of CPU time taken are mentioned in Table 4.

From Table 4, we can see that NMF had the highest coherence scores ('c_v' and 'c_umass'), followed by LDA and HDP. Hence, we selected the top five topics yielded by NMF to search across the first page of Google search results. The content from the first page of Google search results was then retrieved to make sense of the extracted topic keywords to suggest a good topic name. For example, for the set of keywords yielded by the topic model: ['Paris', 'climate', 'green', 'change', 'science', 'reforms', 'environment', 'sustainable', 'urgency'], we did a Google search with these keywords and looked up for content and connections between them to deduce a suitable topic-phrase, i.e., (*Paris Climate Agreement*).

Hence, for each of the 22 selected topics (combined from both the sources, i.e., *online and offline*), we assign them an abstract category out of 'Social Issues,' 'Healthcare,' 'Elections' and 'National Security.' The abstract categories chosen serve as the foundation for political campaigns (Liu

and Lei 2018), and they are the most significant categories to consider when trying to persuade the public to vote for a particular political front. We utilize all these topics to analyze the engagement, stickiness, and to predict the candidate's stance. Table 3 presents the details of the selected topics as per their source and the abstract category they fall into. Each topic category consists of topics from at least two abstract categories. There are nine topics in 'Social Issues,' five in 'Healthcare,' and four in both 'Elections' and 'National Security.' The distribution of topics as per their abstract categories can be seen in Fig. 2. Furthermore, we analyze the most and least talked about topics and abstract categories for each candidate and political party based on engagement and stickiness.

4.2 Engagement on topics

The amount of engagement received on a particular topic helps us quantify how popular the topic was among the general public. To quantify a topic's engagement on Twitter, we first define each of the selected candidate's (user's) engagement on Twitter and then aggregate the tweets published by them, as well as their engagement as per the topic categories defined. The engagement for each user is defined as the product of average engagement per day and their impact.

⁶ <https://radimrehurek.com/gensim/models/nmf.html>

⁷ <https://radimrehurek.com/gensim/models/lmodel.html>

⁸ <https://radimrehurek.com/gensim/models/hdpmodel.html>

Table 3 Topics selected for analysis

Topic source	Topic category	Topic	Abstract category
Online	Modeled Topics (Topics generated from NMF)	Legalization of Medical Marijuana	Social Issues
		Equality rights for LGBTQ	Social Issues
		Weapon Ban	Social Issues
		Build Back Express Tour	Social Issues
		Affordable Health Care Act	Healthcare
Offline	Presidential Debate (1st)	The Economy	Social Issues
		The Supreme Court Appointments	National Security
		COVID-19	Healthcare
		Race & Violence in our cities	Social Issues
		The Integrity of Elections	Elections
		The Trump Biden Records	Elections
		Trump Healthcare Plan	Healthcare
	Presidential Debate (2nd)	Fighting COVID-19	Healthcare
		American Families & The Economy	Healthcare
		Race in America	Social Issues
		Climate Change	Social Issues
		National Security	National Security
		Leadership	National Security
	Snapshot Events	Black Lives Matter	Social Issues
		Capitol Hill Incident	National Security
		US Elections	Elections
		Inauguration Ceremony	Elections

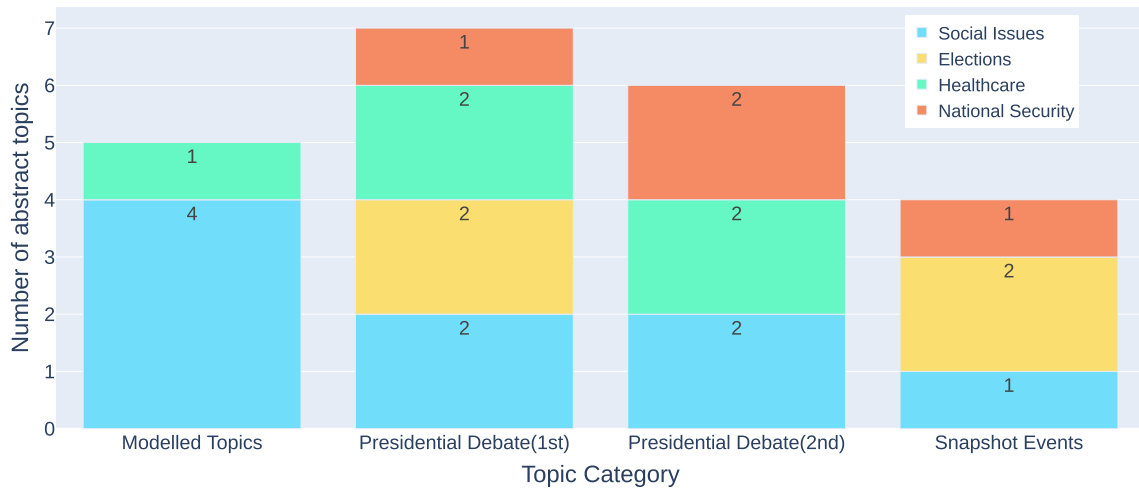


Fig. 2 Distribution of topics as per their abstract categories

Algorithm 1: Pseudocode for engagement of a user per day

```

Input : likesCount, repliesCount, retweetsCount, quotesCount, tweetCreationDate, tweet, user_df,
         followers, following, listedCount, totalTweets, profileCreationDate
Output: engagementPerDayuser (Engagement of a user per day)
1 Function avgEngagementPerDay (likesCount, repliesCount, retweetsCount, quotesCount,
  tweetCreationDate, tweet, user_df):
  // Count the number of tweets in a day
2 tweetsPerDay ← user_df.groupby(['tweetCreationDate'])['tweet'].count()
  // Calculate total engagement for the day
3 user_df['engagement_rate'] ← likesCount + repliesCount + retweetsCount + quotesCount
4 engagement_day ← user_df.groupby(['tweetCreationDate'])
  // Calculate average engagement per day
5 avgEngagement_day ← engagement_day / (4 * tweetsPerDay)
  // Weighted exponential moving average of avgEngagement_day
6 avgEngagement_day ← avgEngagement_day.ewm(span = 20, adjust = FALSE).mean()
7 z_score ← stats.zscore(avgEngagement_day)
  // Remove outliers using z-score
8 avgEngagement_day ← avgEngagement_day[-3 ≤ z_score ≤ 3]
  // Smoothen the average engagement per day to 8th degree using Savitzky-Golay
  filter
9 avgEngagement_day ← savgol_filter(avgEngagement_day, polyorder = 8)
10 return avgEngagement_day
11 Function userImpact (followers, following, listedCount, totalTweets, profileCreationDate):
12 profileAge ← days(January 27, 2021 – profileCreationDate)
  // To quantify whether a user is active producer/consumer of content
13 FtFRatio ← log10 ( (followers / following) + 1 )
14 userImpact ← (followers * listedCount * FtFRatio) / (totalTweets * profileAge)
15 return userImpact
16 engagementPerDayuser ← avgEngagement_day * userImpact
    
```

User engagement was formerly quantified in terms of community features (the number of communities a user is a member of), author features (number of followers/ following, author influence) and content features (the number of retweets, mentions, URLs, hashtags, keywords, comments, and sentiment subjectivity) (He et al. 2020; Purohit et al 2011). Similarly, we aim to include all the accessible features through the Twitter Academic API in this work, and the average engagement per day for a user (*engagementPerDay_{user}*) is computed as the product of average engagement for a

tweet each day (*avgEngagement_day*) and the user impact (*userImpact*). The average engagement for a tweet is aggregated by measuring the reactions received on tweets (such as the number of replies, retweets, likes, and quotes) over the course of the day. The data obtained from the Twitter API for the reactions to each tweet have been aggregated from January 21, 2019, to January 27, 2021. We propose average engagement per day for a tweet by taking inspiration from the *Engagement rate* defined by Twitter⁹. For a given user, Twitter defines *Engagement Rate* as:

$$\text{Engagement rate} = \frac{\text{Engagement}}{\text{Impressions}} * 100 \tag{1}$$

where *Engagement* is the summation over the number of likes, replies, retweets, media views, tweet expansion, profile/ hashtag/ URL clicks, and new followers gained for every tweet, and *Impressions* is the total number of times a tweet has been seen on Twitter, such as through a follower’s timeline, Twitter search, or as a result of someone liking your tweet⁹.

Table 4 Mean coherence scores and CPU time for different clustering algorithms with TF-IDF embeddings over five runs with varying random states

Clustering Algorithm	c_v	c_umass	CPU time (min:sec)
LDA	0.70	-2.26168	52:52
Parallel LDA	0.5921	-2.41955	12:12
NMF	0.773022	-1.61094	07:37
LSI	0.585223	-2.59355	00:27
HDP	0.640714	-17.3223	01:38

Bold signifies the best performing Clustering Algorithm

⁹ <https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard>

Due to limitations with the API, we only have access to the public metrics, i.e., number of likes, retweets, replies, and quotes. Therefore, to calculate the average engagement rate for a user per day, we use the function *avgEngagementPerDay* as proposed in Algorithm 1 (line number: 1). To normalize the fluctuating values of Average Engagement, we calculate its exponential moving average (EMA) with a window span of 20 days for every candidate¹⁰ and remove the outliers using z-score, followed by smoothing the average engagement per day to the 8th degree using Savitzky–Golay filter¹⁰

Every user has a different number of followers, following and they receive varied responses from the users on Twitter (which may or may not be their followers); hence, it is essential to consider their impact (popularity) on Twitter to calculate the number of users they reach through their tweets. Researchers have tried to analyze the impact of users by proposing heuristic and neural-network-based models (Daniluk et al 2021; Razis and Anagnostopoulos 2014; Son et al. 2020). We define the impact of a user (*impact_u*) inspired from the previous work done in Razis and Anagnostopoulos (2014) and define it as a function of followers, following, the total number of tweets, and the profile age, as in Algorithm 1 (line number: 1), where *followers* is the total number of followers a user has, *listedCount* is the number of public lists a user is a part of, *following* is the number of people that the user follows, $\log_{10} \left(\frac{\text{followers}}{\text{following}} + 1 \right)$ is the ratio of followers to following (*FtF* ratio) to check whether a user is an active user (with more followers, producing content) or a passive user (with more following, consuming content). To avoid outliers, we take log base 10 and add one to prevent the metric from being zero when the value of followers equals to following. *tweetCount* is the total number of tweets produced by the user for the scope of our analysis, and *profileAge* is the difference between the profile creation date reported by Twitter and January 27, 2021, i.e., the last day for our data collection, quantified as the number of days. Our algorithm overcomes the shortcomings of Razis and Anagnostopoulos (2014) by incorporating the *listedCount* factor and changing the placement of *tweetCount*. The *tweetCount* has been deemed inversely related to the user impact, because a user tweeting sporadically but obtaining high interaction is more significant than a person tweeting recurrently but receiving low engagement.

The engagement for a user is the product of average engagement per day and the user's impact. The engagement value helps us in quantifying the user's social reach.

4.2.1 Findings

Joe Biden had the highest impact, followed by Donald Trump, Mike Pence, Kamala Harris, *OfficialDemocrats* and *OfficialRepublicans*. We normalize the user impact between the range 0 and 1 to calculate the engagement on tweets for each topic, where 0 is the lowest user-impact, and 1 is the highest.

For Joe Biden, the top three topics receiving the maximum engagement during the scope of our analysis were – *The Integrity of Elections*, *Weapon Ban*, and *US Elections*. As for Kamala Harris, they were *US Elections*, *Fighting COVID-19* and *The Integrity of Elections*, and however, for the *OfficialDemocrats*, the most engaging topics differed from these, and they were *The Supreme Court Appointments*, *Equality Rights for LGBTQ*, and *Fighting COVID-19*. Overall, for Democrats, the top three most engaging topics were *The Integrity of Elections*, *US Elections*, and *Weapon Ban*.

In the case of Donald Trump, the top three topics receiving the maximum engagement were *US Elections*, *The Integrity of Elections*, and *Affordable Healthcare Act*. For Mike Pence, they were *Inauguration Ceremony*, *US Elections* and *Fighting COVID-19*. However, for the *OfficialRepublicans*, as they had the lowest impact, the engagement received on the tweets was too low to be quantified. Overall, for Republicans, the top three most engaging topics were *US Elections*, *Inauguration Ceremony* and *Affordable Healthcare Act*.

4.3 Stickiness of topics

Stickiness helps us to identify the favorite topics for each candidate within the scope of our analysis. We quantify stickiness based on the repetitiveness of the topics spanning across different election phases. The Presidential candidate's timeline is divided into three phases, and the Vice-Presidential candidate's timeline is divided into four phases. See Fig. 3 for more detailed information about the timelines. The topic stickiness is checked for three candidates only, because of the unavailability of exact campaigning dates for Mike Pence. We segregate the topics into three classes based on stickiness:

1. *Very Sticky*: Topics are *Very Sticky* when they have been tweeted about in every election phase,
2. *Sticky*: If the topics were tweeted in $(n - 1)$ election phases, then they were classified as *Sticky*. Here, n is the total number of election phases (i.e., $n = 3$ for Presidential, and $n = 4$ for Vice-Presidential candidates).
3. *Loose*: If the topics were tweeted only once for Presidential candidates, and twice for Vice-Presidential candidates across all the election phases, then they were classified as *Loose*.

¹⁰ A grid-search analysis was performed to find the best value.

Election Phase Timelines for Presidential and Vice Presidential Candidates

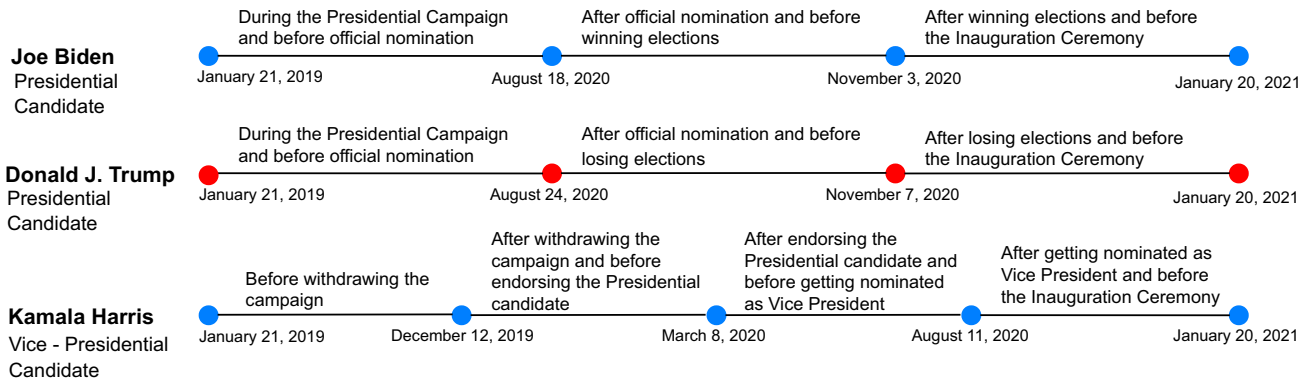


Fig. 3 Electoral campaign timelines for Presidential and Vice-Presidential candidates. The timeline is divided as per the general election phases and the ranks each candidate was contesting for. The details

of campaigning for each candidate have been taken from the news reports of the campaigns on CNBC and Politico

4.3.1 Findings

Joe Biden and Kamala Harris tweeted mostly about *Social Issues and Healthcare* as we can see these categories dominate both the *Very Sticky* and *Sticky* levels from Fig. 4a and b. However, *National Security* topics were loose in nature. Donald Trump tweeted differently from the Democrats and tweeted the most about *Elections and Healthcare*, followed by *Social Issues*.

Doing a micro-analysis, we found that Joe Biden tweeted about twenty topics out of twenty-two, with fourteen of them being *Very Sticky*, five being *Sticky*, and one being *Loose*.

Legalization of Medical Marijuana and *Trump Healthcare Plan* were the topics that Joe Biden had not tweeted about, even once. However, Kamala Harris and Donald Trump tweeted about 21 topics. Kamala Harris had eleven topics in the *Very Sticky* category, seven in *Sticky* and three in *Loose*. For Donald Trump, the distribution of topics per their stickiness levels was slightly different, with thirteen being *Very Sticky*, five being *Sticky*, and three being *Loose*. Kamala Harris did not tweet about *Trump Healthcare Plan*, and Donald Trump did not tweet about the *Legalization of Medical Marijuana*.

Algorithm 2: Content Similarity of two users

```

Input : tweetsuser1, tweetsuser2
Output: Average cosine similarity of users' content (avgCosineSim)
1 embeddingsuser1 ← model.encode(tweetsuser1)
2 embeddingsuser2 ← model.encode(tweetsuser2)
3 cosineSimilarity = [ ]
  /* Compare each tweet embedding of user 1 with all tweet embeddings of user 2 and
   store the max similarity in list */
4 for embeddinguser1 ← 0 to len(embeddingsuser1) do
5   cosineScoresList ← cosineSimilarity(embeddinguser1, embeddingsuser2[:])
6   maxScore ← max(cosineScoresList)
7   cosineSimilarity.append(maxScore)
8 end for
9 avgCosineSim = mean(cosineSimilarity)
    
```

Table 5 Appearance of *Loose* topics in different election phases

Candidate	Election Phase 1	Election Phase 2	Election Phase 3	Election Phase 4
Joe Biden	No loose topics	The Supreme Court Appointments	No loose topics	Not applicable (NA)
Kamala Harris	The Economy, The Trump & Biden Records	National Security	No loose topics	No loose topics
Donald Trump	The Economy, Trump Healthcare Plan, Build Back Express Tour	No loose topics	No loose topics	Not applicable (NA)

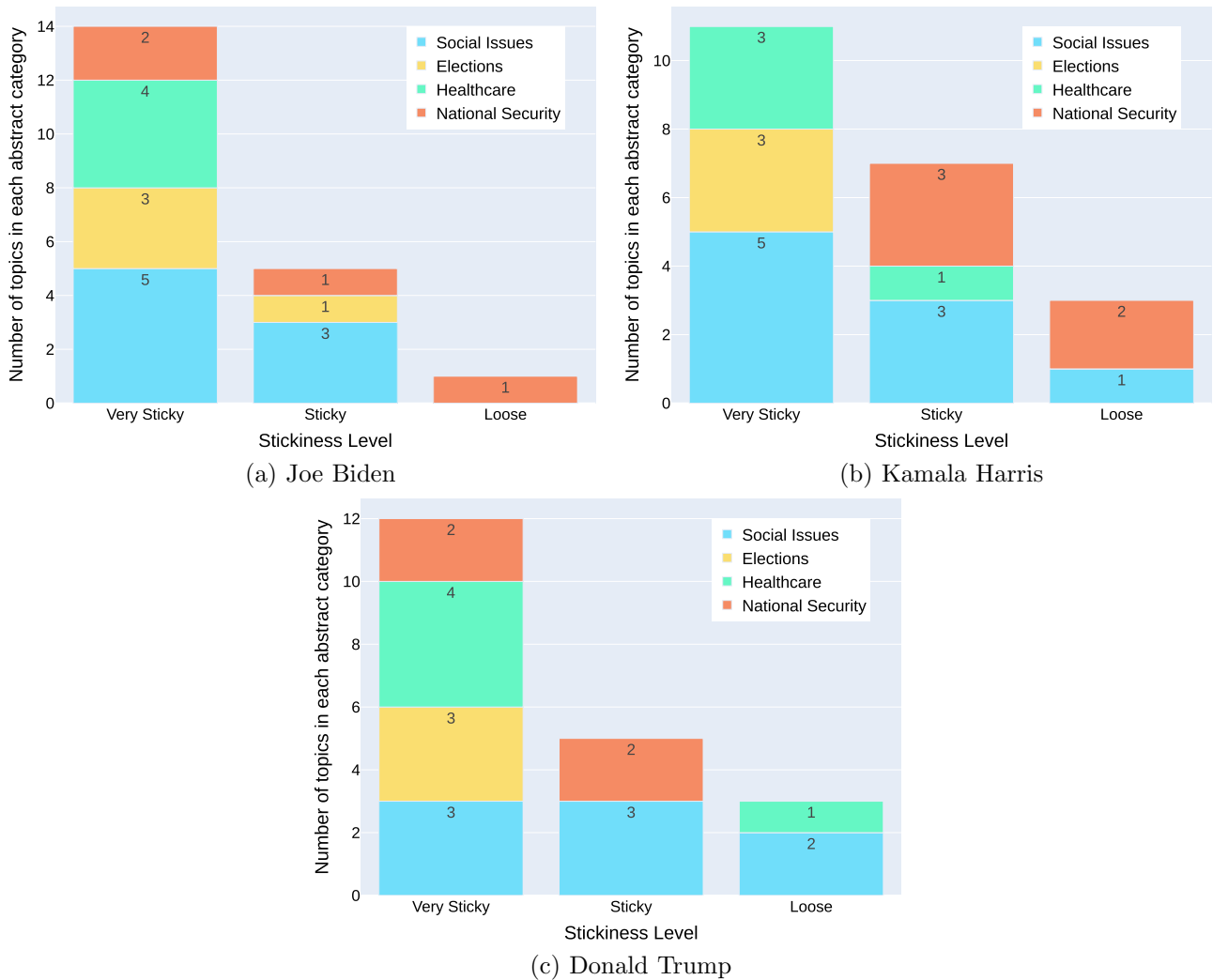


Fig. 4 Abstract categories of topics segregated according to stickiness levels for all three candidates (a) Joe Biden, (b) Kamala Harris, and (c) Donald Trump

Regardless of their political fronts, Joe Biden, Kamala Harris, and Donald Trump had nine *Very Sticky* topics in common (i.e., *The Integrity of Elections*, *Affordable Healthcare Act*, *Equality Rights for LGBTQ*, *Weapon Ban*, *Inauguration Ceremony*, *US Elections*, *American Families & The Economy*, *COVID-19*, *Race & Violence in our cities*). Joe

Biden and Donald Trump stuck to *The Integrity of Elections*; however, Kamala Harris stuck to *Affordable Healthcare Act*. Comparing the candidates from the same political front, Joe Biden and Kamala Harris had eleven common topics in the *Very Sticky* category. For the Presidential candidates, in addition to the nine common topics, they also had *Fighting*

Candidate	Political Party	all-MiniLM-L6-v2		paraphrase-xlm-r-multilingual-v1		paraphrase-mpnet-base-v2		bert-base-nli-mean-tokens		paraphrase-multilingual-MiniLM-L12-v2	
		wrt Presidential Candidate	wrt Political Party (Official Party username)	wrt Presidential Candidate	wrt Political Party (Official Party username)	wrt Presidential Candidate	wrt Political Party (Official Party username)	wrt Presidential Candidate	wrt Political Party (Official Party username)	wrt Presidential Candidate	wrt Political Party (Official Party username)
Kamala Harris	Democrats	0.6537	0.6474	0.6466	0.6256	0.6960	0.7016	0.8334	0.8387	0.6915	0.6795
Joe Biden		NA	0.6549	NA	0.6457	NA	0.7144	NA	0.8497	NA	0.7192
Mike Pence	Republicans	0.6829	0.6974	0.6452	0.6681	0.6670	0.7263	0.8439	0.8562	0.7177	0.7319
Donald Trump		NA	0.6453	NA	0.6262	NA	0.6986	NA	0.8485	NA	0.7187

Fig. 5 Content similarity between candidates using different BERT-based embeddings

COVID-19 and Capitol Hill Incident repeating in the *Very Sticky* category.

Furthermore, the topics classified as *Loose* from all three candidates appeared in only one of the election phases. For example, in Joe Biden’s case, topic *Supreme Court Appointments* appeared only in the second election phase as per his timeline, and similar behavior can be seen for the other two candidates (refer Table 5 for details). Also, the *Loose* topics, i.e., *Trump Healthcare Plan*, *Supreme Court Appointments*, *Build Back Express Tour*, *The Economy*, and *The Trump & Biden Records*, are among the rarely tweeted topics, and corresponding behavior can be seen for the *Very Sticky* topics, i.e., the top three *Very Sticky* topics for each candidate are the most frequently tweeted and highly engaging topics.

5 Synergy among candidates

To quantify the cooperation among the candidates from the same political front, we highlight the content similarity and the congruities and contrasts in the stance of various topics they tweeted about, as discussed below.

5.1 Content similarity

We check the alignment of the Presidential and Vice-Presidential candidates with the political party by comparing the similarity of their tweets as per our proposed Algorithm 2. When comparing the similarity of two users, there is a high probability that the topic of tweets from one user may be repeated by the second user a couple of days before or after the first user’s tweet during the election campaign. So, to address this, we compare each tweet of $user_1$ with all the tweets of $user_2$ and store the maximum similarity between the tweet text. We repeat this process for all the tweets, then average the results to determine how similar two candidates’ content is. The tweets of Vice-Presidential candidates are compared with both the Presidential candidate and the political party, and however, the tweets of the

Presidential candidate are compared with the political party only. We compute the content similarity (cosine similarity) between the candidates by using the top-5 models from HuggingFace¹¹ (grouped by the sentence-similarity task, sorted by the number of downloads) to generate text embeddings. Figure 5 elaborates the performance details for each of them.

5.1.1 Findings

From all the models tested, ‘bert-base-mean-nli-tokens’ performs the best in computing the content similarity between the candidates, followed by ‘paraphrase-multilingual-MiniLM-L12-v2.’ The common trend noticed while computing the content similarity is that the tweets by Kamala Harris are more aligned with the Presidential candidate than the political party for three out of five models; however, for Mike Pence, it’s the opposite, i.e., Mike Pence aligns more with the political party instead of the Presidential candidate. Also, the Presidential candidates have a high similarity rate with the political party. Therefore, the results portray coordination in the candidates’ tweets and the tweets from their political parties.

5.2 Stance similarity

Stance Similarity is an important technique to analyze textual data and is frequently used in NLP to analyze the standpoint of a person toward a topic or an event. We test different models that classify the candidates’ stance for the selected topics in three categories: favor, against, and neutral. Although there is no universal number for how many tweets should be sampled, for testing a model, we observe a range across studies from under 2,000 labeled tweets (Alomari et al 2017; Peisenieks and Skadiņš 2014; Şaşmaz and Tek 2021) to several thousand (Golubev and Loukachevitch 2020; Nabil et al 2015; Rustam et al. 2021;

¹¹ <https://huggingface.co/models>

Table 6 Stance classification performance on the testing set (i.e., 20% of the sampled dataset) using different algorithms

Algorithm used	Oversampled (Yes/No)	Classification performance
Hashing Vectorizer and Linear SVM	No	0.70
	Yes	0.54
<i>Hashing Vectorizer and SVM</i>	No	0.73
	Yes	0.66
<i>Hashing Vectorizer and Logistic Regression</i>	No	0.73
	Yes	0.57
TF-IDF and Linear SVM	No	0.72
	Yes	0.54
TF-IDF and SVM	No	0.70
	Yes	0.66
<i>TF-IDF and Logistic Regression</i>	No	0.73
	Yes	0.56
Spark NLP (Universal Sentence Encoder and Deep Learning Classifier)	No	0.72
BERT-base (uncased)	No	0.69
XLNet (base, epochs=10)	No	0.71
XLNet (large, epochs=10)	No	0.71
facebook/bart-large-mnli (fine tuned)	No	0.75

Zhang et al 2020). For this study, we sample 3,015 tweets evenly distributed across the timeframe of our data collection and from all the candidates. Three different annotators then labeled these tweets, and the stance category having the majority among the three annotators was chosen as the overall response. The annotators had no known prior political biases, and they annotated the tweets solely on the basis of the tweet content. The Fleiss' Kappa statistical test was performed to determine the inter-annotator agreement in labeling, and the *kappa* score is '0.7'. We divide the data into an 80:20 ratio for training and testing multiple classification methods, and we annotate the dataset using standard procedures, as defined above.

We try various traditional and modern algorithms to estimate the performance for stance classification on the labeled tweets. We use TF-IDF and Hashing Vectorizer for the conventional algorithms to generate embeddings as inputs to support vector machine (SVM), linear SVM, and logistic regression to compute the performance. Synthetic Minority Oversampling Technique (SMOTE) is used to oversample the tweets' vectorized features. However, we do not notice a rise in classification performance after oversampling. We also report the classification performances on modern algorithms, like the Deep Neural Network (DNN)-based classifier from the Spark NLP pipeline, which takes Universal Sentence Encodings of tweets as inputs¹²,

BERT-base-uncased¹³, XLNet (base-cased¹⁴, large-cased¹⁵) and fine-tuned Facebook's Zero-shot learning-based, bart-large-mnli¹⁶. From the traditional algorithms, TF-IDF combined with Logistic Regression, Hashing Vectorizer with SVM, and Logistic Regression perform equally well with their stance classification accuracies on the testing set of the sampled data as 73%. From the modern ones, Facebook's 'bart-large-mnli' performs at par with 75% classification accuracy on the test set (Table 6). We then use Facebook's bart-large-mnli to predict the stance on the remaining tweets.

5.2.1 Findings

From the predicted stance, we notice that Democrats favored most of the topics, apart from *Capitol Hill Incident* and *Build Back Express Tour*, where they had a neutral stance. However, Republicans had a favorable outlook for 14 topics, an unbiased view for seven (*The Trump & Biden Records*, *National Security*, *Leadership*, *Black Lives Matter*, *Capitol Hill Incident*, *Inauguration Ceremony*, and *Build Back Express Tour*), and they were against one topic (*Trump Healthcare Plan*). Additionally, the predicted stance illustrates symmetry between the candidates' tweets and the tweets from their political parties.

¹² <https://nlp.johnsnowlabs.com/api/com/johnsnowlabs/nlp/annotators/classifier/dl>

¹³ <https://huggingface.co/bert-base-uncased>

¹⁴ <https://huggingface.co/xlnet-base-cased>

¹⁵ <https://huggingface.co/xlnet-large-cased>

¹⁶ <https://huggingface.co/facebook/bart-large-mnli>

6 Conclusion & future work

Social media platforms (SMPs) have evolved into strategic spaces critical to modern political campaigns. Because of the interactivity of social media, candidates can establish direct relations with their audience without an intermediary (e.g., newspapers, news channels). Not surprisingly, campaign strategists prioritize social media as the primary channel for delivering and persuading messages. Therefore, understanding how the internal co-operation of a political front helps in contesting for elections is just as important as understanding the external factors, like the candidate's impact and the audience's engagement.

In terms of the candidate's impact and engagement on topics, it was found that the candidate with the highest impact did not have a higher number of tweets as compared to the candidate with a lower impact. User impact ultimately depends on the topic referred to in the tweets. Also, Joe Biden and Kamala Harris understood their audience well. They tweeted about the topics receiving higher engagements than Donald Trump and Mike Pence, who were unable to identify the topics of social interest. Additionally, the Democratic candidates displayed higher internal cooperation through their tweets and stance on different topics than the Republican candidates.

This study extends political campaign research by investigating two broad aspects — engagement and stickiness of topics, the candidate synergy (i.e., the content and stance similarity) and their social reach on Twitter during the 2020 US Presidential elections. The results clearly indicate that the tweet's topic, and the candidate's influence, both have an impact on the amount of public engagement it receives. Internal cooperation (i.e., similarities in the content and stance for specific topics) also helps the candidates in creating a stronger hold on the thoughts of the public during the election campaign. Furthermore, this study employed an empirical approach by examining how internal cooperation between candidates and their engagement influenced the election results and it may be utilized to create a Twitter communication model for the candidates and government to assist in effective campaigning and governance.

Several restrictions, as well as recommendations, must be noted for further research. Additionally, as the data for this study are focused on the tweets from Presidential and Vice-Presidential candidates for the 2020 US Elections, it was not feasible to distinguish between organic and synthetically generated (i.e., through bots, masked profiles) engagements. The impact of moving window for the content similarity between two candidates needs to be investigated further. Also, we do not intend to generalize our results for every political campaign. Moreover, our research is based on text features; we could not account for the influence of

image features and knowledge graphs related to a particular tweet. Investigating the separation and impact of organic and sponsored audience involvement may benefit future studies. Additionally, factors that might influence public engagement and internal cooperation, such as emotions, timing impacts, and tweet formats, should be examined precisely. Furthermore, investigating the extent of interactions between users and political candidates through their social media managers will contribute significantly to a better understanding of the democratic ability of SMPs.

Acknowledgments The authors are grateful to the research team at Lakehead University's DaTALab for their help - Andrew Fisher and Aditya Singhal for proof-reading this paper, along with Aditya Singhal, Amir Shahbandegan and Subhodip Panda for annotating the sampled tweets, to evaluate the stance similarity among candidates from the same political front. We would also like to thank the CASES building, Lakehead University for providing the infrastructure; the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant and Vector Institute since this study would not have been feasible without their support. Rajesh Sharma is supported by EU H2020 program under the SoBigData++ project (grant agreement No. 871042).

Author contributions Manmeet Kaur Baxi contributed to formal analysis, investigation, validation, visualization, and writing — original draft. Rajesh Sharma and Vijay Mago were involved in supervision, and writing — review & editing.

Code availability The additional information for this work — data, code, and findings, is accessible on the GitHub repository (<https://github.com/manmeetkaurbaxi/2020-US-Elections>).

Declarations

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

References

- Alomari KM, ElSherif HM, Shaalan K (2017) Arabic tweets sentiment analysis using machine learning. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, pp 602–610
- Balliet D, Wu J, De Dreu CK (2014) Ingroup favoritism in cooperation: a meta-analysis. *Psychol Bull* 140(6):1556
- Bansal B, Srivastava S (2018) On predicting elections with hybrid topic based sentiment analysis of tweets. *Procedia Comput Sci* 135:346–353
- Bertot JC, Jaeger PT, Munson S, Glaisyer T (2010) Social media technology and government transparency. *Computer* 43(11):53–59
- Bhat V, Yadav A, Yadav S, Chandrasekran D, Mago V (2021) Adcofe: advanced contextual feature extraction in conversations for emotion classification. Preprint [arXiv:2104.04517](https://arxiv.org/abs/2104.04517)
- Bilal M, Gani A, Marjani M, Malik N (2019) Predicting elections: Social media data and techniques. In: 2019 international conference on engineering and emerging technologies (ICEET), IEEE, pp 1–6.
- Bonsón E, Bednářová M (2018) The use of youtube in western European municipalities. *Gov inf Q* 35(2):223–232

- Bonsón E, Perea D, Bednárová M (2019) Twitter as a tool for citizen engagement: an empirical study of the Andalusian municipalities. *Gov Inf Q* 36(3):480–489
- Bonsón E, Ratkai M (2013) A set of metrics to assess stakeholder engagement and social legitimacy on a corporate facebook page. *Online Information Review*
- Bonsón E, Royo S, Ratkai M (2015) Citizens' engagement on local governments' Facebook sites. An empirical analysis: the impact of different media and content types in western Europe. *Gov Inf Q* 32(1):52–62
- Bonsón E, Royo S, Ratkai M (2017) Facebook practices in western European municipalities: an empirical analysis of activity and citizens' engagement. *Adm Soc* 49(3):320–347
- Bonsón E, Torres L, Royo S, Flores F (2012) Local e-government 2.0: Social media and corporate transparency in municipalities. *Gov Inf Q* 29(2):123–132
- Borah P (2016) Political Facebook use: campaign strategies used in 2008 and 2012 presidential elections. *J Inf Technol Polit* 13(4):326–338
- Bruns A, Moon B (2018) Social media in Australian federal elections: comparing the 2013 and 2016 campaigns. *Journal Mass Commun Q* 95(2):425–448
- Centre PR (2020) Differences in how democrats and republicans behave on twitter. <https://www.pewresearch.org/politics/2020/10/15/differences-in-how-democrats-and-republicans-behave-on-twitter/>
- Chandrasekaran D, Mago V (2021) Evolution of semantic similarity-a survey. *ACM Comput Surv (CSUR)* 54(2):1–37
- Chen E, Deb A, Ferrara E (2021) #Election2020: the first public twitter dataset on the 2020 us presidential election. *J Comput Soc Sci* 5:1–18
- Chun S, Shulman S, Sandoval R, Hovy E (2010) Government 2.0: making connections between citizens, data and government. *Inf Polity* 15(1, 2):1–9
- Daniluk M, Dabrowski J, Rychalska B, Goluchowski K (2021) Synerise at recsys 2021: Twitter user engagement prediction with a fast neural model. In: *RecSysChallenge'21: proceedings of the recommender systems challenge 2021*, pp 15–21
- Dzisah WS (2018) Social media and elections in Ghana: enhancing democratic participation. *Afr Journal Stud* 39(1):27–47
- Effing R, Van Hillegersberg J, Huibers T (2011) Social media and political participation: Are Facebook, Twitter and YouTube democratizing our political systems? In: *International conference on electronic participation*, Springer, pp 25–35
- Gaisbauer F, Pournaki A, Banisch S, Olbrich E (2021) Ideological differences in engagement in public debate on twitter. *Plos One* 16(3):e0249241
- Galgoczy MC, Phatak A, Vinson D, Mago VK, Giabbanelli PJ (2022) (re) shaping online narratives: when bots promote the message of president trump during his first impeachment. *PeerJ Comput Sci* 8:e947
- Gallo E, Yan C (2015) The effects of reputational and social knowledge on cooperation. *Proc Natl Acad Sci* 112(12):3647–3652
- Golubev A, Loukachevitch N (2020) Improving results on Russian sentiment datasets. In: *Conference on artificial intelligence and natural language*, Springer, pp 109–121
- Gross J, De Dreu CK (2019) The rise and fall of cooperation through reputation and group polarization. *Nat Commun* 10(1):1–10
- Grossmann M (2014) *Artists of the possible: Governing networks and American policy change since 1945*. Oxford University Press, Oxford
- Grover P, Kar AK, Dwivedi YK, Janssen M (2019) Polarization and acculturation in us election 2016 outcomes-can twitter analytics predict changes in voting preferences. *Technol Forecast Soc Change* 145:438–460
- Gruzd A, Lannigan J, Quigley K (2018) Examining government cross-platform engagement in social media: Instagram versus twitter and the big lift project. *Gov Inf Q* 35(4):579–587
- Gruzd A, Roy J (2016) Social media and local government in Canada: an examination of presence and purpose. In: *Social Media and Local Governments*, Springer, pp 79–94
- He L, Han C, Mukherjee A, Obradovic Z, Dragut E (2020) On the dynamics of user engagement in news comment media. *Wiley Interdiscip Rev: Data Min Knowl Discov* 10(1):e1342
- Hollebeek LD, Glynn MS, Brodie RJ (2014) Consumer brand engagement in social media: Conceptualization, scale development and validation. *J Interact Mark* 28(2):149–165
- Joseph FJJ (2019) Twitter based outcome predictions of 2019 Indian general elections using decision tree. In: *2019 4th international conference on information technology (InCIT)*, IEEE, pp 50–53
- Khanam KZ, Srivastava G, Mago V (2022) The homophily principle in social network analysis: a survey. *Multimed Tools Appl* 1–44
- Larson JM (2021) Networks of conflict and cooperation. *Annu Rev Polit Sci* 24:89–107
- Lilleberg J, Zhu Y, Zhang Y (2015) Support vector machines and word2vec for text classification with semantic features. In: *2015 IEEE 14th international conference on cognitive informatics & cognitive computing (ICCI* CC)*, IEEE, pp 136–140
- Liu D, Lei L (2018) The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 us presidential election. *Discourse Context Media* 25:143–152
- Madeo D, Mocenni C (2020) Self-regulation versus social influence for promoting cooperation on networks. *Sci Rep* 10(1):1–10
- Nabil M, Aly M, Atiya A (2015) Astd: Arabic sentiment tweets dataset. In: *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp 2515–2519
- Newman D, Lau J.H, Griesser K, Baldwin T (2010) Automatic evaluation of topic coherence. In: *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, pp 100–108
- Norris DF, Reddick CG (2013) Local e-government in the united states: Transformation or incremental change? *Public Adm Rev* 73(1):165–175
- Nugroho D.K (2021) Us presidential election 2020 prediction based on twitter data using lexicon-based sentiment analysis. In: *2021 11th international conference on cloud computing, data science & engineering (confluence)*, IEEE, pp 136–141
- Peisenieks J, Skadiņš R (2014) Uses of machine translation in the sentiment analysis of tweets. In: *Human language technologies—the Baltic perspective*. IOS Press, pp 126–131
- Perkoski E (2019) Internal politics and the fragmentation of armed groups. *Int Stud Q* 63(4):876–889
- Praznik L, Qudar MMA, Mendhe C, Srivastava G, Mago V (2021) Analysis of link prediction algorithms in hashtag graphs. In: *Big data and social media analytics*, Springer, pp 221–245
- Purohit H, Ruan Y, Joshi A, Parthasarathy S, Sheth A (2011) Understanding user-community engagement by multi-faceted features: a case study on twitter. In: *WWW 2011 Workshop on Social Media Engagement (SoME)*
- Razis G, Anagnostopoulos I (2014) Influcetracker: rating the impact of a twitter account. In: *IFIP international conference on artificial intelligence applications and innovations*, Springer, pp 184–195
- Ríos AM, Benito B, Bastida F (2017) Factors explaining public participation in the central government budget process. *Aust J Public Adm* 76(1):48–64
- Röder M, Both A, Hinneburg A (2015) Exploring the space of topic coherence measures. In: *Proceedings of the eighth ACM international conference on Web search and data mining*, pp 399–408
- Russmann U, Svensson J (2017) Interaction on Instagram?: Glimpses from the 2014 Swedish elections. *Int J E-Polit (IJEP)* 8(1):50–66

- Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS (2021) A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos One* 16(2):e0245909
- Sahly A, Shao C, Kwon KH (2019) Social media for political campaigns: An examination of Trump's and Clinton's frame building and its effect on audience engagement. *Soc Media+ Soc* 5(2):2056305119855141
- Sampieri RH (2018) *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw Hill, México
- Sandhu M, Vinson CD, Mago VK, Giabbanelli PJ (2019) From associations to sarcasm: mining the shift of opinions regarding the supreme court on twitter. *Online Soc Netw Media* 14:100054
- Sandoval-Almazan R, Valle-Cruz D (2018) Facebook impact and sentiment analysis on political campaigns. In: *Proceedings of the 19th annual international conference on digital government research: governance in the data age*, pp 1–7
- Sari Y, Stevenson M (2016) Exploring word embeddings and character n-grams for author clustering. In: *CLEF (Working Notes)*, Citeseer, pp 984–991
- Şaşmaz E, Tek FB (2021) Tweet sentiment analysis for cryptocurrencies. In: *2021 6th international conference on computer science and engineering (UBMK)*, IEEE, pp 613–618
- Siebers V, Gradus R, Grotens R (2019) Citizen engagement and trust: A study among citizen panel members in three Dutch municipalities. *Soc Sci J* 56(4):545–554
- Son J, Lee J, Oh O, Lee HK, Woo J (2020) Using a heuristic-systematic model to assess the twitter user profile's impact on disaster tweet credibility. *Int J Inf Manag* 54:102176
- Tsai MH, Wang Y, Kwak M, Rigole N (2019) A machine learning based strategy for election result prediction. In: *2019 international conference on computational science and computational intelligence (CSCI)*, IEEE, pp 1408–1410
- Vesnic-Alujevic L, Van Bauwel S (2014) YouTube: a political advertising tool? A case study of the use of YouTube in the campaign for the European parliament elections. *J Polit Mark* 13(3):195–212
- Wang C, Paisley J, Blei D (2011) Online variational inference for the hierarchical dirichlet process. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*, pp.752–760
- Wonka A, Haunss S (2020) Cooperation in networks: political parties and interest groups in EU policy-making in Germany. *Eur Union Polit* 21(1):130–151
- Zhang L, Fan H, Peng C, Rao G, Cong Q (2020) Sentiment analysis methods for hpv vaccines related tweets based on transfer learning. In: *Healthcare*, vol 8, Multidisciplinary Digital Publishing Institute, p 307
- Zhao R, Tan VY (2016) Online nonnegative matrix factorization with outliers. *IEEE Trans Signal Process* 65(3):555–570

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