


ORIGINAL ARTICLE

Violent political rhetoric on Twitter

Taeyoon Kim 

Political Science and Social Data Analytics, Pennsylvania State University, University Park, Pennsylvania, USA
Corresponding author. Email: taeyoon@psu.edu

(Received 12 March 2021; revised 21 November 2021; accepted 26 January 2022; first published online 31 May 2022)

Abstract

Violent hostility between ordinary partisans is undermining American democracy. Social media is blamed for rhetoric threatening violence against political opponents and implicated in offline political violence. Focusing on Twitter, I propose a method to identify such rhetoric and investigate substantive patterns associated with it. Using a data set surrounding the 2020 Presidential Election, I demonstrate that violent tweets closely track contentious politics offline, peaking in the days preceding the Capitol Riot. Women and Republican politicians are targeted with such tweets more frequently than men and non-Republican politicians. Violent tweets, while rare, spread widely through communication networks, reaching those without direct ties to violent users on the fringe of the networks. This paper is the first to make sense of violent partisan hostility expressed online, contributing to the fields of partisanship, contentious politics, and political communication.

Keywords: American politics; civil/domestic conflict; computational models; mass media and political communication; political parties and interest groups

The emergence of social media platforms was widely touted as a technological revolution that would bring about many beneficial outcomes such as political learning and participation (Dimitrova et al., 2014; Tucker et al., 2017). However, such early hopes are being overshadowed by mounting concerns about aggressive political communication. In recent days, one can easily encounter uncivil political discussion both from political elites as well as ordinary users. Also, various types of hate speech—targeted at women, ethnic minorities, and partisan opponents—are common and viral on social media (Mathew et al., 2019). Accordingly, much scholarly attention has been paid to detect such speech and curb its spread (Siegel, 2020). However, we know very little about another, perhaps most deleterious, type of aggressive political speech: violent political rhetoric. Violent political rhetoric, expressing the intention of physical harm against political opponents, has drawn significant media attention. Numerous media reports show that malevolent users on social media write posts that threaten violence against political opponents on the basis of partisanship, ideology, and gender and that such posts are even associated with the actual incidences of offline violence (Brice-Saddler, 2019; Daugherty, 2019; Vigdor, 2019). In particular, many social media platforms are implicated in the extremist effort to motivate and organize the Capitol Riot that left a vivid and deep scar on American democracy. Plenty of evidence shows that not only niche extremist online forums but also mainstream social media platforms, including Twitter, were exploited by users who called for violence in the days preceding the riot on January 6, 2021 (Guynn, 2021; Lytvynenko and Hensley-Clancy, 2021; Romm, 2021).

Violent political rhetoric is worrisome not only because it serves as a harbinger of extremist offline violence but also because exposure to such rhetoric has harmful consequences such as

increased tolerance for offline violence against political opponents (Kalmoe, 2014) and ideological polarization (Kalmoe et al., 2018). It is particularly concerning because violent political rhetoric can widely spread through the communication network on social media, amplifying its negative effects. Besides, such rhetoric is in itself a behavioral manifestation of violent partisanship where individuals not just hate out-partisans (Abramowitz and Webster, 2018) but also support and even enjoy the use of violence against them (Kalmoe and Mason, 2018). The rhetoric is an online mirror image of the recent instances of inter-partisan offline violence surrounding contentious political issues (e.g., Black Lives Matter movements, the controversies about the 2020 Presidential Election) and is no less concerning than its offline counterpart (Pilkington and Levine, 2020).

How prevalent is violent political rhetoric on social media? How do posts containing such rhetoric relate to offline-world politics? What types of politicians are targeted? What users use violent rhetoric against political opponents? How diffusive is violent political rhetoric and what predicts its spread? Given the significance of violent political rhetoric, it is urgent to investigate these questions. Due to the massive size of the content generated in real time, however, it is prohibitively expensive to manually identify violent content on a large scale, leaving only anecdotal and incomprehensive evidence (Lytvynenko and Hensley-Clancy, 2021; Romm, 2021). Therefore, I propose an automated method for detecting violent political rhetoric from a continuous stream of social media data, focused on Twitter. I then apply the method to build a data set of tweets containing violent political rhetoric over a 16-week period surrounding the 2020 Presidential Election. Finally, I provide comprehensive data analyses on the characteristics and spread of violent political rhetoric.

By doing so, I contribute to three areas of research in political science. First, I shed light on the literature on political violence by extending the study of individuals' engagement in political violence to online domains. While a body of research in offline political violence has taken a bottom-up approach to study individuals who take part in collective violence in the offline world (Horowitz, 1985; Scacco, 2010; Fujii, 2011; Tausch et al., 2011; Claassen, 2016), few studies have taken a similar approach to investigate individuals who threaten violence against political opponents in online space. I fill part of the gap by showing that individuals who threaten violence against political opponents on social media are ideologically extreme and located on the fringe of the online communication network. I also show that they threaten opposition politicians, in the context of heightened contentious politics offline. The online–offline links identified in my study open up a future research agenda on what causal mechanisms connect threats of political violence online and contentious offline politics, including offline political violence.

By identifying and characterizing violent political rhetoric on Twitter, I also extend the study of aggressive online political communication where incivility and hate speech have been the key areas of inquiry (Berry and Sobieraj, 2013; Gervais, 2015; Munger, 2017; Suhay et al., 2018; Gervais, 2019; Popan et al., 2019; Sydnor, 2019; Siegel, 2020; Munger, 2021; Siegel et al., 2021). Building on a new data set spanning the crucial period surrounding the 2020 Presidential Election, I show that, although tweets containing violent political rhetoric are rare (0.07 percent of political tweets, on average), they spread beyond those without direct ties to violent users. I find that almost 40 percent of the retweets of such content spread through indirect ties (i.e., my friend's friend, a friend of my friend's friend, etc.), thereby creating huge potential for incidental exposure to such abhorrent language. I also demonstrate that, although threatening tweets are shared primarily among ideologically similar users, there is a considerable amount of cross-ideological exposure as well, calling for further investigation into the effects of exposure to violent political rhetoric both from an in-party member and from an out-party member.

Finally, I shed light on the literature on mass partisan polarization and negative partisanship by demonstrating that violent partisanship is manifested online in the form of threats against partisan opponents. Recent studies on mass partisan polarization highlight that partisans are not just ideologically far apart (Abramowitz and Saunders, 2008; Fiorina and Abrams, 2008) but also

dislike or even endorse violence against our-party members (Iyengar et al., 2012; Abramowitz and Webster, 2018; Kalmoe and Mason, 2018; Iyengar et al., 2019). However, there was little effort to explore how violent partisanship is expressed online. My work contributes to the literature by providing an easy-to-access indicator for tracking the level of violent partisanship. Considering the evidence that there are significant discrepancies between survey self-reports and actual online behavior (Guess et al., 2019), my study provides an excellent complement to survey-based measurement as it enables researchers to directly observe the over-time trend of violent partisan behavior expressed online.¹ For instance, I illustrate that the level of violent political rhetoric on Twitter corresponds to the violent partisan tension offline, reaching its peak in the days preceding the Capitol Riot.

1 Related work

In this paper, I build on and contribute to three streams of literature. First, a large body of works takes a micro-level approach to study participation in offline political violence, helping shed light on those who threaten political opponents online. Second, an extensive body of research in political communication investigates violent political metaphors offline and aggressive speech in online political discussion, providing a rich context for an inquiry into violent rhetoric in online political communication. Third, research on political polarization and negative partisanship helps understand why social media users express a violent intention against out-partisans (a form of behavioral manifestation of extreme negative partisanship) and what consequences such behavior has.

1.1 Offline political violence

Although few studies exist to explain political violence online, there is an extensive body of literature explaining why individuals engage in offline political violence in various settings. Focused on conflict-ridden contexts, studies seek to explain why individuals participate in inter-group violence (ethnic, religious, partisan). Major explanations include selective incentives that alleviates the problem of free-riding (DiPasquale and Glaeser, 1998; Humphreys and Weinstein, 2008), social pressure (Scacco, 2010; Fujii, 2011), and perceived distributive inequality (Claassen, 2016). Also, an interdisciplinary stream of studies on violent extremism seeks to identify a host of risk factors associated with individuals' tendency to join violent extremist activities (LaFree and Ackerman, 2009; Borum, 2011a, b; Gill et al., 2014; McGilloway et al., 2015). Lack of stable employment, history of mental illness, low self-control, perceived injustice, and exposure to violent extremism are among the factors highlighted in the literature (Schils and Pauwels, 2016; Pauwels and Heylen, 2017; LaFree et al., 2018).

1.2 Aggressive political communication

Raising concerns about political elites' violent rhetoric in the USA, a recent strand of studies investigates its political consequences (Kalmoe, 2014; Matsumoto et al., 2015; Kalmoe et al., 2018; Kalmoe, 2019). Kalmoe (2019) shows that violent political metaphors (metaphors that describe politics as violent events such as a battle or a war) increase willingness to vote among individuals with highly aggressive personalities but the opposite effect is found among individuals low in aggressive personalities. Focusing on issue polarization, Kalmoe et al. (2018) find that violent political metaphors prime aggression in aggressive partisans and thus lead to intransigence on issue positions.

¹Although this approach shares with survey self-reports a concern that they both can be susceptible to intentional exaggeration or suppression resulting from social norms, the former nonetheless is far less reactive than the latter.

While violent political rhetoric is studied mainly in the context of political elites' offline speech, many works in online political communication focus on incivility and hate speech. They point out that the reduced gate-keeping power of traditional media outlets and online anonymity gave rise to uncivil and hateful content targeted at people of a different race, gender, and partisan affiliation (Kennedy and Taylor, 2010; Berry and Sobieraj, 2013; Munger, 2017; Shandwick, 2019; Munger, 2021). Aggressive online speech is reported to have crucial consequences for many political outcomes, including participation (Henson et al., 2013; Sydnor, 2019), information seeking (Sydnor, 2019), inter-group evaluations, and deliberative attitudes (Gervais, 2019).² Accordingly, a large body of works is devoted to detecting (Waseem and Hovy, 2016; Davidson et al., 2017; Zimmerman et al., 2018; Siegel, 2020) and discouraging uncivil and hateful speech (Munger, 2017, 2021).

1.3 Affective polarization and negative partisanship

Recent scholarship on political polarization highlights affective polarization, the degree to which citizens dislike and distrust out-partisans (Iyengar et al., 2012, 2019). Documenting an increase in affective polarization over the last several decades (Iyengar et al., 2019), the scholarship seeks to uncover its negative consequences, including anti-deliberative attitudes, social avoidance, and outright social discrimination (MacKuen et al., 2010; Iyengar et al., 2012; Abramowitz and Webster, 2016; Huber and Malhotra, 2017; Hutchens et al., 2019; Broockman et al., 2020; Druckman et al., 2020). Extending the study of negative partisanship, some works take one step further, evaluating the extent to which partisans rationalize harm and even endorse violence against partisan opponents (Kalmoe and Mason, 2018; Westwood et al., 2021). Such negative partisanship has mainly been measured using survey self-reports. While there exist a handful of other approaches, such as implicit association test (Iyengar et al., 2019), survey self-reports have been the only strategy to measuring violent partisanship (Kalmoe and Mason, 2018; Westwood et al., 2021).

2 Targeted violent political rhetoric

Building on the psychology literature on aggression (Anderson and Bushman, 2002), I define violent political rhetoric as rhetoric expressing the intention of severe physical harm against political opponents.³ This involves a threat and support of physical harm against political opponents and hopes of extreme physical harm inflicted on them (*schadenfreude*).⁴

Existing studies on violent political rhetoric have employed various conceptualizations (Kalmoe, 2014; Kalmoe et al., 2018; Kalmoe, 2019; Zeitzoff, 2020). Zeitzoff (2020) employs an expansive definition of violent political rhetoric: "any type of language that defames, dehumanizes, is derogatory, or threatens opponents." Thus, violent political rhetoric is conceptualized as a spectrum that encompasses "name-calling and incivility at the lower end and threats or calls for violence at the upper end." Closely related to my study is the type of violent political rhetoric at the upper end of the spectrum.

Kalmoe and his coauthors' works focus specifically on violent political metaphors (Kalmoe, 2014; Kalmoe et al., 2018; Kalmoe, 2019). In their work, violent political metaphors are defined

²For a comprehensive review of behavioral consequences of political incivility and a discussion of related psychological processes, see Sydnor (2019).

³Anderson and Bushman (2002) define aggression as "any behavior directed toward another individual that is carried out with the proximate (immediate) intent to cause harm" and violence as a form of "aggression that has extreme harm as its goal."

⁴There is an active discussion on how to conceptualize and measure political violence in survey research (Kalmoe and Mason, 2018; Westwood et al., 2021). Here, violent political rhetoric is targeted at a specific political entity. The target is typically a partisan opponent, either a group (e.g., Republican representatives, Democratic senators) or an individual politician (e.g., Donald Trump, Joe Biden).

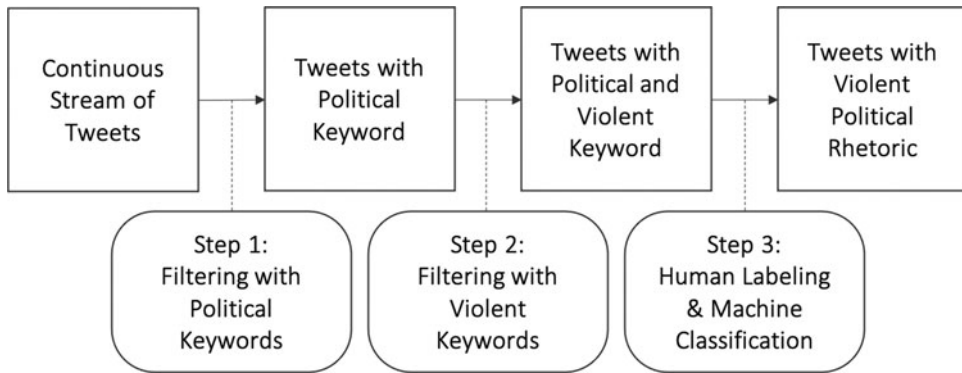


Figure 1. Data collection pipeline.

as “figures of speech that cast nonviolent politics of campaigning and governing in violent terms, that portray leaders or groups as combatants, that depict political objects as weapons, or that describe political environments as sites of non-literal violence.” In contrast to the definition employed in my study, this type of violent political rhetoric does not threaten (or support, incite) any physical violence against political opponents.

3 Detecting violent political rhetoric on Twitter

Many approaches have been proposed to detect hostile speech on social media, including incivility (Davidson et al., 2020; Theocharis et al., 2020) and hate speech (Siegel, 2020), employing various approaches from dictionary (Dadvar et al., 2012; Magu et al., 2017; Isbister et al., 2018) to machine learning methods (Nikolov and Radivchev, 2019; Williams et al., 2020).⁵ However, there has been little effort to identify violent political rhetoric, a distinct form of hostile speech. While a small body of research on YouTube proposes several methods to identify threatening comments from YouTube videos (Wester, 2016; Wester et al., 2016; Hammer et al., 2019), they are narrowly focused on a small sample of videos in a highly specific context.⁶ In this section, I introduce a new method that combines keyword filtering and machine learning to detect violent political rhetoric from a massive stream of content on Twitter (Figure 1).

3.1 Step 1: filtering through political keywords

I start with compiling a list of political keywords to download tweets from a Twitter API (Application Programming Interface).⁷ Since a massive number of heterogeneous tweets are generated in real time, I first filter the tweet stream through a set of political keywords. The keywords involve a broad sample of politicians’ accounts (members of Congress, governors, and the four candidates of the 2020 Presidential Election) as well as those belonging to major parties. The tweets filtered and downloaded through the list of accounts “mention” (Twitter, 2021a) at least one of the political accounts in the list.⁸ Naturally, the keywords of my choice make the

⁵Machine learning methods typically outperform dictionary methods. For a comprehensive review of works focused on detecting incivility and hate speech, see Davidson et al. (2020) and Siegel (2020), respectively.

⁶See Online Appendix D for more information.

⁷A Twitter API is used to retrieve data and engage with the communication on Twitter.

⁸Mentions appear in tweets (a) when users reply to other users’ tweets (then, the account of the original tweeter automatically appears in the reply) and (b) when users simply include the account in their tweet text. The function is the key communicative component on Twitter with which users initiate and keep engaging with each other (Twitter, 2021a).

downloaded tweets political in nature. In addition, focusing on tweets mentioning these accounts is an effective approach to gather political tweets that engage (and threaten) politicians in conversation.⁹

I then run a computer program that scrapes live tweets that contain any of the keywords in the list. The program is designed to scrape live tweets continuously via the Streaming API (Twitter, 2021c). This API allows researchers to scrape live tweets as they are published while another major API, the Search API, provides access to historical tweets up to a certain number of days in the past (Twitter, 2021f). The decision to opt out of the Search API is due to the potential for the platform to engage in censorship. That is, a set of tweets retrieved via the Search API will leave out violent tweets that have been deleted by Twitter for violating its terms of service.¹⁰

3.2 Step 2: filtering through violent keywords

Once I have collected a corpus of tweets with at least one political keyword, I move on to the task of splitting it into violent and non-violent tweets. Here, my approach is very similar to the one taken in the previous step. I first compile a list of violent keywords and filter the existing tweets through those keywords. A challenge here is that any human-generated list of keywords might leave out potentially relevant tweets. As King et al. (2017) demonstrate, humans are not particularly capable of coming up with a representative list of keywords for a certain topic or concept. In other words, it is hard for any single researcher to compile a comprehensive set of keywords used to express a violent intention against partisan opponents (e.g., kill, shoot, choke, etc.).

To deal with this, I combine model-based extraction of keywords with human judgment. First, I start with fitting a model to score terms in an external corpus that was already human-labeled in terms of whether a text is threatening or not. Here, I intend to extract violent keywords from a corpus that already contains information about what multiple people deem to be threatening. Specifically, I use a data set built by Jigsaw, a unit within Google (Jigsaw, 2020). The data set contains around two-million online comments labeled by human coders for various toxic conversational attributes, including “threat.” I fit a logistic regression model and extract terms (uni- and bi-gram features) that are most predictive of perceived threat (in terms of the size of the weights assigned to them). Second, given the weighted terms, I then use human judgment to set a threshold above which terms are included in the list of violent keywords. I set the threshold at the top-200 because over the top-200 terms, the terms were too generic to indicate any intention of violence. Using the list of terms, I divided the political tweets from step 1 into ones with and without at least one violent keyword. For more detailed information about keyword filtering in general and my violent keywords, see Online Appendix B.

3.3 Step 3: manual labeling and machine classification

Although the previous round of filtering relies on a list of violent keywords that people frequently use online and consider violent, only a small fraction of the violent-keyword tweets contain the

⁹Though I do not intend to build a sample of “all political tweets,” focusing on mention tweets might not represent all political tweets engaging politicians in conversation. This is primarily because users still can and do reference politicians using their name (“Donald Trump” as opposed to “@realDonaldTrump”). To evaluate the extent to which focusing on mention tweets bias any downstream analysis, I calculated the proportion of the number of tweets including a given politician’s full names to the number of tweets including their accounts and compared the proportion across major politician-level attributes highlighted in the analysis. I report the results in Online Appendix A. I find no evidence for any tendency that politicians are referenced differently in terms of the choice of the full name and the account, across gender, political party, and position.

¹⁰Twitter has detailed policies on violent threats (Twitter, 2021h). Essentially, its approach is post hoc in that it reviews what is already publicly published and decides whether to moderate content or sanction users. The data set I gather through the Streaming API (Twitter, 2021c) avoids such post hoc moderation. In addition, to the best of my knowledge, it is unclear whether Twitter has a mechanism that prevents users from writing violent content in the first place.

intention of violence. This is because many tweets contain a violent keyword without expressing any intention of physical harm against political opponents. The major sources of false positives involve (a) when violent keywords are used as a metaphor that describes non-violent political events (Kalmoe, 2013, 2014; Kalmoe et al., 2018; Kalmoe, 2019), (b) a religious curse that does not threaten physical harm (e.g., “burn in hell!”), (c) quoting (or even criticizing) violent political rhetoric from someone else, and (d) irony (e.g., “why don’t you just shoot them all if you believe violence solves the problem?”). To more accurately identify tweets containing violent political rhetoric, three human coders, including myself and two undergraduate assistants, classified tweets in terms of whether the author expresses the intention of severe physical harm against a political opponent (see Supplementary materials for detailed coding rules). The coders manually labeled a set of 2500 tweets together and then individually labeled over 7500 tweets. The inter-coder agreement score in terms of Krippendorff’s alpha is around 0.6, higher than the standard in the relevant literature (Krippendorff, 2018). For more information on the manual labeling, see Online Appendix C.

In addition, I used active learning (Settles, 2009; Linder, 2017; Miller et al., 2020) to more efficiently identify tweets with violent political rhetoric. Since the corpus compiled through steps 1 and 2 is highly imbalanced with only a small fraction containing violent political rhetoric, randomly sampling a training set for regular supervised learning will lead to inefficiency. That is, the training set will contain too few relevant tweets for any classifier to learn about what features predict violent political rhetoric. Using active learning, I go through an iterative process where I start with manually labeling *randomly sampled* texts to train a classifier, *select (not randomly)* texts whose predicted probabilities are around the decision threshold (ones whose class the classifier is most uncertain about), manually label the around-the-threshold texts, and finally accumulate those texts to re-train the classifier.

Through the iterative process, I compiled a training set of violent-keyword tweets labeled for violent rhetoric. I then trained various machine learning classifiers and the performance of the classifiers was evaluated on unseen (or held-out) data using fivefold cross validation in terms of precision, recall, and F-1 (Han et al., 2011). To label the rest of the tweets, I selected the best performing classifier (precision: 71.8, recall: 65.6, F-1: 68.4), one built on BERT (bidirectional encoder representations from transformers) (Devlin et al., 2018). For more information on the active learning and machine classification process, see Online Appendix D.

4 Characteristics and spread of tweets containing violent political rhetoric

How prevalent is violent political rhetoric on social media? How do posts containing such rhetoric relate to offline-world politics? What types of politicians are targeted? What users use violent rhetoric against political opponents? How diffusive is violent political rhetoric and what predicts its spread? In this section, I provide comprehensive data analyses concerning the characteristics and spread of tweets containing violent political rhetoric. The following analysis is based on a data set of tweets collected between September 23, 2020 and January 8, 2021.¹¹ This 16-week period covers major political events concerning the 2020 Presidential Election, including the Capitol Riot and the suspension of Trump’s Twitter account.

The key findings include the following. Violent political rhetoric on Twitter is closely related to offline contentious politics, spiking to its highest level in the days preceding the Capitol Riot. In terms of targeting, women and Republican politicians are more frequently targeted than men and non-Republican politicians. Violent users are ideologically extreme, located on the fringe of the communication network, and their ideological makeup varies over time depending on what issues violent political rhetoric arises from. The spread of violent tweets takes place primarily among ideologically similar users but there is also a substantial amount of cross-ideological

¹¹The data set includes 343,432,844 political tweets (235,019 are classified as violent). For computational efficiency, I randomly sampled 1/2000 of the non-violent political tweets from each day and used the sampled tweets in the analysis.

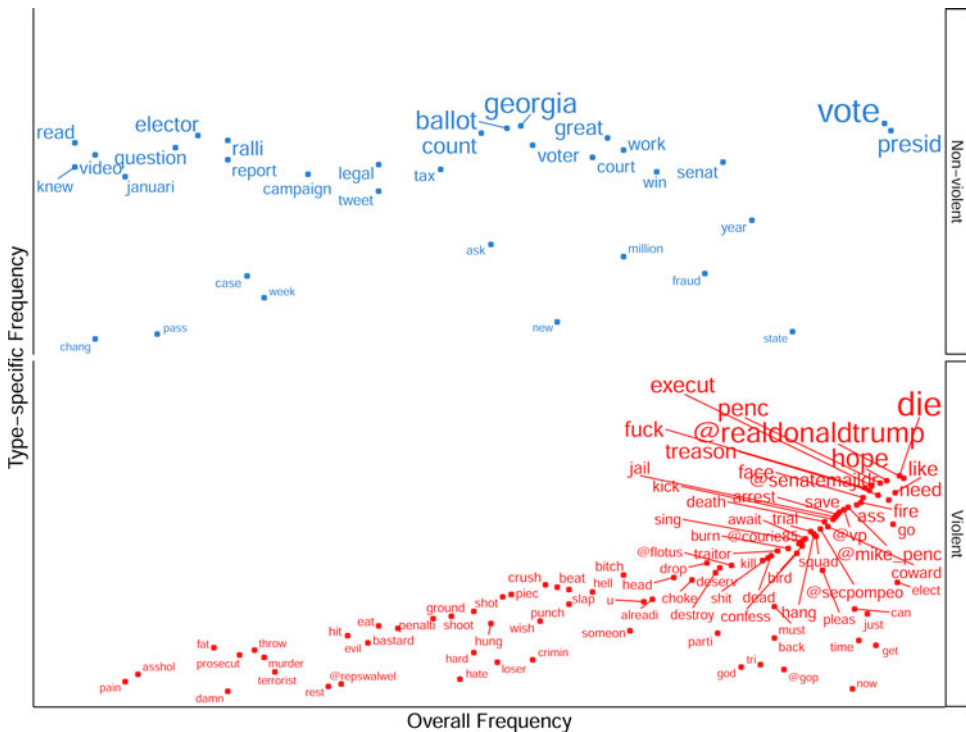


Figure 2. Comparison of terms by type of tweets. *Note:* For the analysis, I took a sample of 10,000 tweets, with 5000 from each type. I used an R package *quanteda* for text preprocessing. Punctuation, symbols, numbers, stopwords, and URLs were removed from the text. The text was lower-cased and stemmed.

spread, raising concerns about co-radicalization. While violent political rhetoric is rare (0.07 percent of political tweets) but almost 40 percent of retweets of violent tweets take place between users without a direct following tie, incidentally exposing a potentially huge audience to such appalling content.

4.1 Content and timeline of violent tweets

To shed light on how tweets containing violent political rhetoric differ from non-violent political tweets in terms of content, [Figure 2](#) shows the terms that divide non-violent political tweets from violent political tweets. I rely on a feature selection/weighting method for comparing word usage across different groups called *Fightin' Words* (Monroe et al., 2008).¹² In the figure, the x-axis indicates the relative frequency with which the keyword occurs in each type. The y-axis in each panel depicts the extent to which the keyword is associated with each type (see [Online Appendix E](#) for the top-30 keywords). Note that some of the words included as indicating violent political tweets have already been baked in as part of the violent-keyword filtering.

What is most noteworthy is that words that indicate certain political entities are much more frequent for violent tweets than for non-violent ones. We can see that, while no entity-specific words were included in the keywords for non-violent tweets, the violent keywords include many accounts that belong to high-profile political figures such as @realdonaldtrump (Donald

¹²The method models word usage differences across different groups in a way that reduces the prominence of words used too frequently or too infrequently. The method produces a z-score that quantifies the significance with which the use of a word differs between two groups of documents.

Table 1. Most frequent hashtags in violent political rhetoric (entire period)

Rank	Hashtag	Count	Rank	Hashtag	Count
1	#wethepeople	1511	16	#pardonsnowden	365
2	#1	1398	17	#traitortrump	358
3	#pencecard	1341	18	#freeassange	356
4	#maga	881	19	#punkaf	354
5	#fightback	702	20	#godwins	244
6	#1776again	672	21	#execute	241
7	#antifaarefascists	607	22	#covidiot	231
8	#blmareracists	607	23	#arrest	228
9	#covid19	606	24	#trampicntraitors	225
10	#treason	555	25	#brandonbernard	223
11	#vote	498	26	#mcenemy	218
12	#trump	452	27	#moscowmitch	215
13	#trump2020	434	28	#againstrump	199
14	#walterreed	428	29	#makeassholegoaway	199
15	#savebrandonbernard	421	30	#jesuschrist	187

Trump), @senatemajldr (Mitch McConnell), @mike_pence (Mike Pence), and @secpompeo (Mike Pompeo). In particular, the account for Trump, “@realdonaldtrump,” demonstrates that he was at the center of violent and divisive communication on Twitter. The prevalence of entity-specific words is also consistent with our focus on targeted violent political rhetoric. For the words indicating non-violent tweets, many general political terms are included (e.g., presid, vote, tax, elector, campaign) along with words that represent particular political events such as “georgia” (the Senate election in Georgia) or “fraud” (misinformation about election fraud).

Now that we understand the stylistic characteristics of violent political rhetoric, what is talked about in violent tweets? To provide a general sense of the content in violent tweets, Table 1 reports the top-30 hashtags that are most frequently used in violent tweets.¹³ Note that I had lower-cased the text of the tweets before extracting hashtags to match ones that only differ in capitalization. In general, the hashtags together show that the content of violent political rhetoric is highly variegated, revolving around diverse political/social issues: general partisan hostility (#wethepeople, #1), racial conflict (#antifaarefascists, #blmareracists), moral issues (#brandonbernard, #pardonsnowden, #freeassange), election campaigning (#vote, #trump2020), disputes over the election result (#pencecard, #fightback, #1776again), and the COVID-19 pandemic (#covid19, #walterreed, #covidiot). For the hashtags reflecting general partisan hostility (“#wethepeople” and “#1”), close manual reading reveals that they are used when users emphasize their in-partisans as representing the whole country (the former) and their out-partisans as the foremost enemy of the country (the latter). Although it is beyond the scope of this study to review every hashtag in the list, they together make it clear that violent political rhetoric is closely related to various political/social issues in offline politics.

Then, how frequent are violent tweets over time? Figure 3 illustrates the timeline of tweets containing violent political rhetoric. The trend is expressed in their count and proportion to the total number of political-keyword tweets. Regardless of the metric, the figure shows very similar trends. First, we can see that the proportion of violent political rhetoric is quite rare: an average of 0.07 percent of the tweets that include the political keyword(s) contain violent political rhetoric. Such rarity is consistent with findings from recent research on aggressive political communication on social media. For instance, Siegel et al. (2021) report that around 0.2 percent of political tweets contain hate speech during the period from June 2015 to June 2017. Although violent tweets comprise only a small fraction of political discussion, it is important to note that it amounts

¹³Not all violent tweets contain a hashtag so the partisan source of the hashtags in Table 1 (or Table 2) does not necessarily correspond to the distribution of violent users’ ideology or their partisanship in the entire data set.

Table 2. Most frequent hashtags in violent political rhetoric (weekly)

	(2020) 9/23–9/29	9/30–10/6	10/7–10/13	10/14–10/20
1	#trump2020	#covid19	#executed	#treason
2	#maga	#vote	#amendments	#biden
3	#treason	#walterreed	#bancapitalisim	#sealteam6
4	#debates2020	#trump	#constitution	#hillaryclinton
5	#whenthesecondwavehits #10/21–10/27	#covidiot 10/29–11/3	#government 11/4–11/10	#obama 11/11–11/17
1	#crimesagainstchildren	#endnigeria	#jesuschrist	#antifaarefascists
2	#crimesagainsthumanity	#endsars	#trump2020	#blmareracists
3	#laptopfromhell	#vote	#trump	#marchfortrump
4	#tonybobulinski	#trump2020	#maga	#trump2020
5	#moscowmitch	#electionday	#trumpcrimefamily	#treason
	#11/18–11/24	11/25–12/1	12/2–12/8	12/9–12/15
1	#treason	#maga	#treason	#savebrandonbernard
2	#maga	#diaperdon	#magabusmusters	#brandonbernard
3	#scif	#fightbacknow	#magaqueentrains	#gopisover
4	#trump	#richardmoore	#bidencheated2020	#abolishthedeathpenalty
5	#democracydemandsit #12/16–12/22	#headsmustroll 12/23–12/29	#kag2020 12/30–1/5 (2021)	#treason 1/6–1/8
1	#pardonsnowden	#wethepeople	#fightback	#maga
2	#freeassange	#1	#1776again	#traitortrump
3	#punkaf	#pencecard	#godwins	#execute
4	#wethepeople	#pardonsnowden	#divinetiming	#arrest
5	#stopthesteal	#freeassange	#trustgod	#trampicantraitors

to hundreds of thousands of tweets containing violent political rhetoric, per day, and it is seen by the number of users that is far greater than that of such tweets themselves.¹⁴

As illustrated in Figure 3, there is a considerable over-time variation in the trend of violent political rhetoric. In particular, two big spikes are prominent in early October 2020 and early January 2021 along with a steady increase toward the election and the period of power transition. To provide a detailed look into issues driving the trend, Table 2 reports the weekly top-5 hashtags included in violent tweets. While the steady uptrend toward the election and the period of power transition appears associated with the partisan competition/tension over the election and its results (#vote, #trump2020, #electionday, #laptopfromhell, #tonybobulinski), the two big spikes require further explanation. First, the hashtags for the week from September 30 to October 6 (e.g., #walterreed, #trump, #covidiot, #covid19) show that the earlier spike reflects political animosity surrounding Trump's infection of COVID-19 and his much-criticized behavior during his three-day hospitalization at Walter Reed military (O'Donnell, 2020). In addition, manual reading of the tweets on October 2 and the following several days verifies that numerous tweets express a violent intention against Trump.

As for the later spike, the hashtags for the last couple of weeks, such as #fightback, #1776again, and #pencecard, are the ones that grew substantially among far-right extremists and conspiracy theorists who attempt to delegitimize the election results. We can also see that anti-Trump users, in turn, responded to the far-right discourse using hashtags such as "#arrest and #execute #traitortrump," leading to the massive upsurge in the amount of violent political rhetoric during the last phase of the period under study.¹⁵ It is also important to note that, while the general prevalence of violent political rhetoric in November and December reflects the partisan tension over the election results (#treason, #diaperdon, #fightbacknow, #stopthesteal) along with other

¹⁴Note that the Streaming API returns 1 percent of all tweets in real time. Therefore, the estimated number of violent tweets will be roughly 100 times greater than what we see in the data.

¹⁵For more detailed information about the context in which these hashtags were used, see Blumenthal (2021), Itkowitz and Dawsey (2020), and Lang et al. (2021).

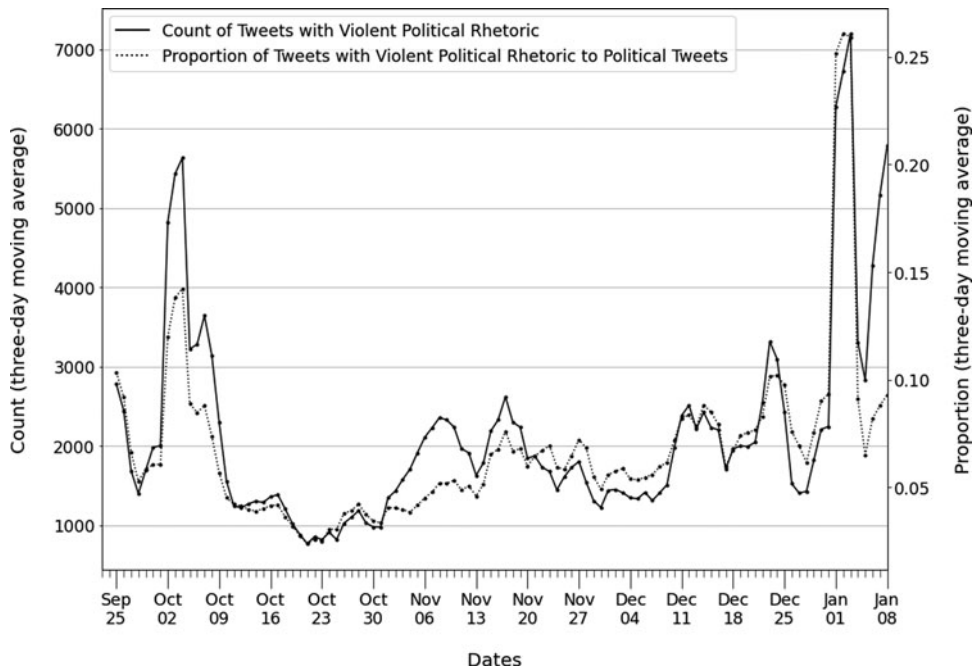


Figure 3. Timeline of violent political rhetoric (September 23, 2020–January 8, 2021). *Note:* The y-axis on the left side indicates the number of tweets containing violent political rhetoric while the other y-axis on the right side depicts the proportion of such tweets relative to tweets containing a political keyword. Each point in the lines indicates the three-day moving average.

politically salient issues, the drastic uptrend starting in the last week of 2020 appears to be predominantly driven by the extremist discourse agitated by Trump’s continuous mobilization effort, on and off Twitter. Considering Trump’s tweet instigating his radical supporters to gather in D.C. on January 6 and the riot on that day,¹⁶ it is abundantly clear that offline political conflict is intertwined with violent political rhetoric on Twitter.¹⁷

4.2 Politicians in violent tweets

Then, what politicians are mentioned in violent tweets? Tweets can “mention” an account either by directly including it in its text or by replying to tweets written by the account (Twitter, 2021a). Table 3 reports what politicians’ accounts are mentioned in violent tweets and presents them by the type of position, political party, and gender. Each cell records the average number of violent tweets that mention politicians’ accounts in a given category. First, the table shows that Trump is at the center of violent partisan expressions on Twitter. As a single political figure, he appears in far more violent tweets than all the other political accounts combined. Pence, the former vice president, attracts the second largest number of violent tweets followed by the contender for

¹⁶On December 26, 2020, Trump tweeted that “*The ‘Justice’ Department and the FBI have done nothing about the 2020 Presidential Election Voter Fraud, the biggest SCAM in our nation’s history, despite overwhelming evidence. They should be ashamed. History will remember. Never give up. See everyone in D.C. on January 6th.*” On January 6, 2021, a joint session of Congress was scheduled to be held to count the Electoral College and to formalize Biden’s victory.

¹⁷It is important to note that I am not making causal claims between violent political rhetoric online and offline political conflict. This is an important direction for future research. For existing works on the relationships between the two, see Chan et al. (2016), Mooijman (2018), Olteanu et al. (2018), Klein (2019), van der Vegt et al. (2019), Wei (2019), Siegel (2020), Gallacher (2021), Gallacher et al. (2021), and Gallacher and Heerdink (2021).

Table 3. Mean mention count

		Mention count
Position	Trump (incumbent president)	137,475
	Pence (incumbent vice president)	18,506
	Biden (candidate for presidency)	8759
	Harris (candidate for vice presidency)	467
	Governors	165
	Senators	479
Party	Representatives	56
	Republican	257
	Non-Republican	117
Gender	Women	103
	Men	207

the presidency, Biden, and by the vice-presidential candidate from the Democratic Party, Harris. Also, representatives, compared to governors and senators, receive a small amount of attention in violent political tweets. Presumably, it might be due to the large number of representatives that makes them less likely to get sufficient individualized attention to stimulate violent partisan expressions.

Given that Trump (Republican and man) can obscure the comparison based on political party and gender, statistics for political party and gender are reported without violent tweets that mention his account. The second part of the table shows that Republicans appear more frequently than non-Republicans (Democrats and a handful of independent/minor party politicians). Also, we can see that, on average, men politicians appear more frequently in violent tweets than women politicians.

To further explore how political party, gender, and the type of position correlate with the mentioning of politicians in violent tweets,¹⁸ Table 4 reports the results from a negative binomial regression where the count of mentions in violent tweets, the outcome variable, is regressed against the type of position, political party, and gender. In line with the literature (Southern and Harmer, 2019), I include the number of followers to consider the amount of attention given to each politician. To prevent a tiny subset of the observations from being overly influential, I exclude the candidates for the presidential election (Biden, Trump, Harris, Pence) who attracted so much attention during the period around the election. For the details of modeling and robustness analysis, see Online Appendix F.

First, the results reveal that being Republican correlates positively with mentioning in violent tweets (model 5). Why do Republican politicians appear more frequently in violent tweets than Democratic ones? One possibility is that politicians who belong to the party holding presidency are more frequently targeted as they might draw more attention and criticism, particularly given the amount of violent intention directed at Trump. Also, as often pointed out in the literature, Twitter users are younger and more likely to be Democrats than the general population (Wojcik and Hughs, 2019). Therefore, liberal users who outnumber conservative ones might write more violent tweets that target Republican politicians than their conservative counterparts do against Democratic politicians. Second, the results show that being a woman is positively associated with mentioning in violent tweets (model 5). This is consistent with both academic and

¹⁸blackWhile mentioning does not necessarily indicate targeting, mentioning is a good proxy for targeting and is often used to measure targeting in the literature (Munger, 2021; Siegel et al., 2021). To evaluate the extent to which mentioning indicates targeting in my data, I manually labeled a random sample of 500 tweets taken from the entire data of violent tweets, in terms of whether a violent tweet is targeting the mentioned politician. Specifically, I labeled a tweet as relevant when (a) the intention of violence in the tweet is targeted at a specific politician and (b) the tweet targets the politician using their account. The result shows that about 40 percent of violent tweets target politicians using their accounts. This proportion is higher than what is found in a similar study on hate speech in online political communication (about 25 percent in Siegel et al., 2021).

Table 4. Mentioning of political accounts: negative binomial regression

	Model 1	Model 2	Model 3	Model 4	Model 5
Position:governor	1.08*** (0.30)				0.51* (0.22)
Position:senator	2.15*** (0.23)				0.18 (0.18)
Woman		- 0.38 (0.21)			0.97*** (0.15)
Republican			0.78*** (0.18)		0.99*** (0.13)
Follower count (log)				2.57*** (0.11)	2.52*** (0.13)
(Intercept)	4.02*** (0.10)	5.00*** (0.10)	4.47*** (0.12)	- 8.79*** (0.52)	- 9.44*** (0.59)
AIC	5255.01	5364.26	5348.76	4689.54	4636.13
BIC	5272.50	5377.38	5361.88	4702.53	4666.46
Log likelihood	- 2623.51	- 2679.13	- 2671.38	- 2341.77	- 2311.07
Deviance	734.63	747.13	745.37	664.94	658.53
Number of accounts	585	585	585	562	562

* Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

* For models 4 and 5, the follower count was not retrieved for some accounts due to screen name change, suspension, etc.

journalistic evidence for online abuse against women politicians (Fuchs and SchÄfer, 2019; Rheault et al., 2019; Southern and Harmer, 2019; Felmlee et al., 2020; Cohen, 2021; Di Meco and Brechenmacher, 2021).

4.3 Engagement in political communication network by tweeter type

How central and active are violent and non-violent users in the political communication network on Twitter? This question is important because the more central to the network and active violent users are, the more likely ordinary users are exposed to violent political rhetoric. Figure 4 depicts the logged distribution of four user-level indicators in the political communication network (see Online Appendix G for the median values). Here, violent users follow (and are followed by) other users, “like” others’ tweets, and write tweets to a lesser degree than non-violent users, implying that violent users are on the fringe of the communication network (the number of friends and followers, and likes) and less active (the number of tweets).¹⁹

4.4 Distribution of ideology by tweeter type

While there is plenty of evidence that far-right extremism is more responsible for offline political violence in the USA than their left-wing counterpart (e.g., Jones, 2020), it is unclear whether such asymmetry holds in online political communication. How are violent users distributed on the ideological continuum? In panel (a) in Figure 5, I report the distribution of an ideology score for violent and non-violent tweeters, measured using an ideal point estimation approach introduced by Barberá (2015).²⁰ Here, higher scores indicate greater conservatism. First, the

¹⁹blackThe term “fringe” is used to indicate that violent users are less central in key communication ties on Twitter (i.e., friending, following, and liking). It is most closely related to “degree centrality” in network analysis (i.e., the number of ties that a node has) (Newman, 2018).

²⁰This method is well established and has been used in many other studies in both political science and other social science disciplines (Vaccari et al., 2015; Imai et al., 2016; Brady et al., 2017; Jost et al., 2018; Gallego et al., 2019; Hjorth and Adler-Nissen, 2019; Freelon and Lokot, 2020; Sterling et al., 2020; Kates et al., 2021; Munger, 2021). It is based on the assumption that Twitter users follow political actors (e.g., politicians, think tanks, news outlets) whose positions on the latent ideological dimension are similar to theirs (i.e., homophily in social networks). Considering following decisions a costly signal

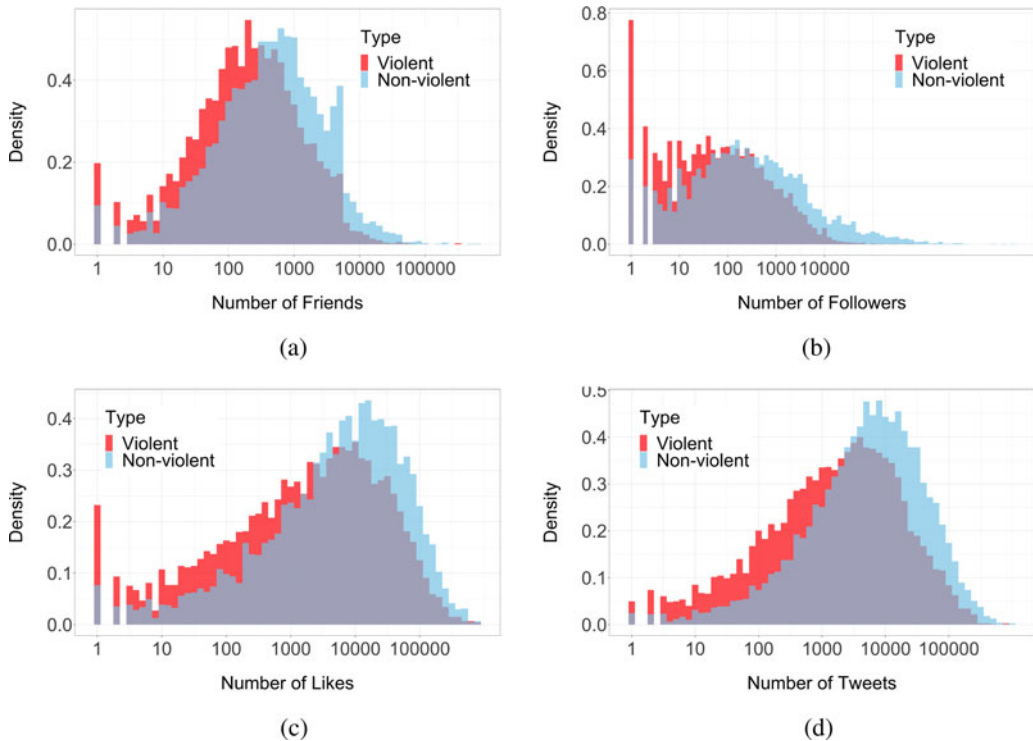


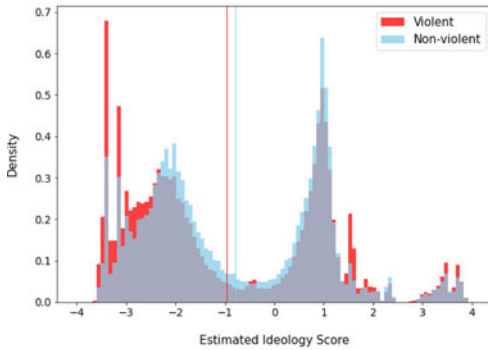
Figure 4. Distribution for network engagement indicators. *Note:* The unit of observation is an account. Each of the four network engagement indicators is depicted on the x-axis. The original linear distribution for each indicator was log-transformed (base 10) after adding 1 in order to clearly visualize outliers. The y-axis depicts the probability density. “Friends” are whom a given user follows and “followers” are those who follow a given user.

distribution for non-violent tweeters shows that they are slightly more liberal (since the vast majority of political tweeters are non-violent ones, the distribution for non-violent tweeters is nearly identical to that of political tweeters). This is consistent with the fact that Twitter users tend to be liberal, younger, and Democrats (Wojcik and Hughs, 2019). Second, it is noteworthy that violent tweeters are more liberal than non-violent tweeters. We can see that the mean ideology score of violent tweeters leans toward the liberal direction. The results of Welch two-sample *t*-test also show that the difference is 0.18 and statistically significant (95 percent confidence interval (C.I.): 0.15, 0.20). This analysis reveals that liberals are no less violent than conservatives in online political communication, in contrast to the asymmetry in the offline world.

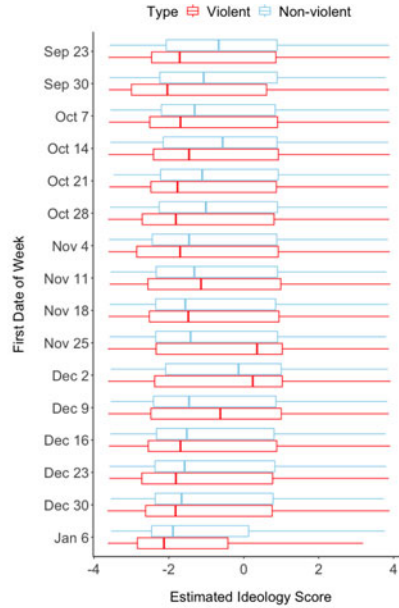
Certainly, the liberal slant might be affected by the fact that the data covers a period that only includes a Republican president. Indeed, a huge number of threatening tweets were targeted at Trump (see Table 3). Considering the level of hostility an incumbent president can provoke from the partisan opposition, liberals might be over-represented in violent tweets in the data. However, the liberal slant still exists after removing all the tweets that mention Trump’s account (see Online Appendix H).

Here, it is important to note that there is over-time heterogeneity. Panel (b) in Figure 5 shows that, while violent users tend to be more liberal than non-violent ones for the first seven weeks, the trend flips for the next five weeks, and again flips back for the last four weeks. These findings

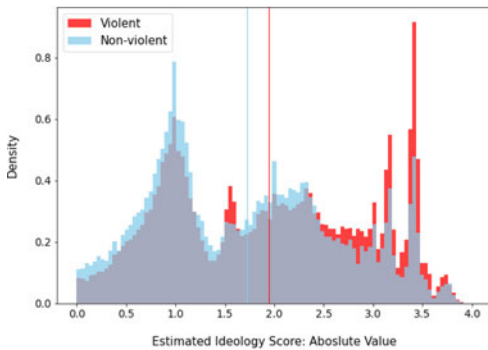
about users’ perceptions of both their latent ideology and that of political actors, the method estimates ideal points of Twitter users based on the structure of following ties. It produces a uni-dimensional score in which negative values indicate liberal ideology and positive values indicate conservative ideology.



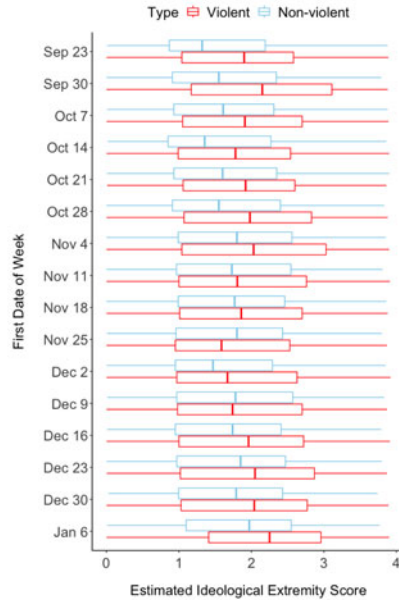
(a) Ideology



(b) Ideology (weekly)



(c) Ideological extremity



(d) Ideological extremity (weekly)

Figure 5. Ideology and ideological extremity by type of political tweeters. *Note:* The unit of observation is an account. For panels (a) and (b), larger values indicate greater conservatism. For panels (c) and (d), larger values indicate greater extremity. The vertical lines in panels (a) and (c) indicate the mean value for each group.

imply that the use of violent language in online political communication is likely to reflect particular phrases of politics that stimulate violent partisan hostility—as seen in the hashtags in Table 2—rather than the use of violent political rhetoric bears an inherent relationship with ideology.

Finally, to get a sense of how ideologically extreme violent users are compared to non-violent users, I computed an ideological extremity score by taking the absolute value of the ideology score. Panel (c) in [Figure 5](#) demonstrates that violent tweeters are more ideologically extreme than non-violent tweeters. The same pattern is also found for almost all the weekly distributions shown in panel (d). These results make intuitive sense in that those who display such radical online behavior are unlikely to be ideologically moderate just like offline political violence is committed by extremists on the far ends of the ideological spectrum.

4.5 Spread of violent political rhetoric

How do tweets containing violent political rhetoric spread? Existing research on online political communication suggests that, while political information is exchanged primarily among individuals who are ideologically similar (Barberá et al., 2015), there is also a significant amount of cross-ideological communication (Barberá, 2014; Bakshy et al., 2015). Then, in terms of retweeting, do violent tweets spread primarily among ideologically homogeneous users?²¹ The first two panels in [Figure 6](#) present two scatter plots for violent and non-violent tweets where tweeter's ideology score is on the x-axis and retweeters' is on the y-axis. We can see the retweets are highly concentrated in the areas of similar ideology scores. The Pearson's *R* scores are around 0.7 (0.696 for the violent, 0.713 for the non-violent).

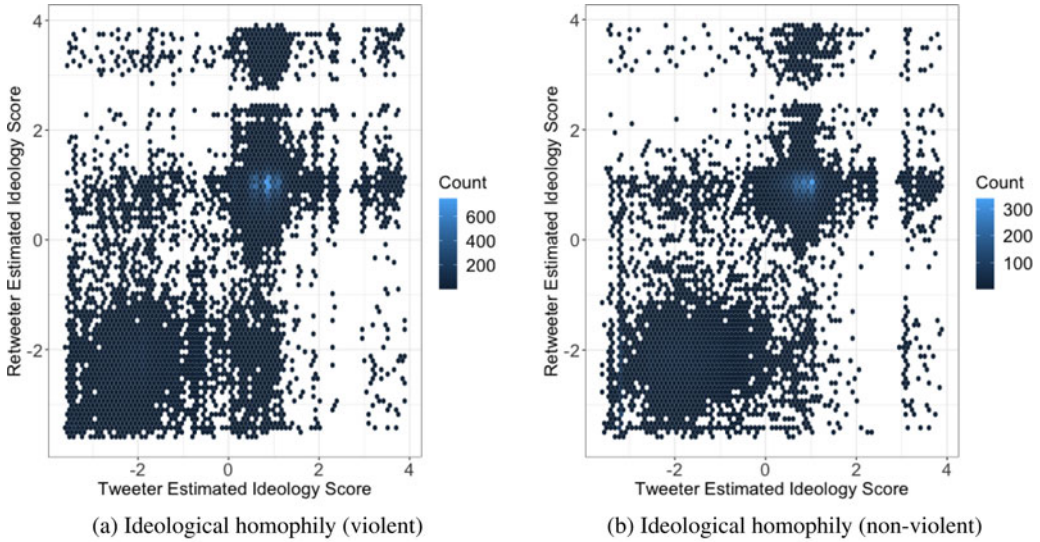
While the findings confirm that retweeting, both violent and non-violent, is affected by ideological homophily, there is a substantial amount of cross-ideological spread in both types of political communication (expressed on the top-left and bottom-right sides of the plots). Although the spread of violent political rhetoric takes place primarily among ideologically similar users, the findings imply that users encounter and spread partisan opponents' violent behavior, potentially co-radicalizing each other by feeding off political opponents' violent behavior (Ebner, 2017; Pratt, 2017; Knott et al., 2018; Moghaddam, 2018).

Then, how far do violent tweets travel on the Twitter communication network?²² As previously discussed, violent tweeters tend to lie on the fringe of the communication network. However, their content still can travel to a large audience through indirect ties. Panel (c) in [Figure 6](#) describes the distribution of the shortest path distance on the following network for all the retweets of violent and non-violent tweets in the data set. Here, the shortest path distance is the minimum number of following ties necessary to connect two users. The distance is estimated as one if the retweeter is in the tweeter's followers list (or the tweeter is in the retweeter's friends list). Similarly, the distance is estimated as two if the intersection between the retweeter's friends list and the tweeter's follower list is not an empty set (and if there is no direct follower/following relationship). If neither condition is met, the shortest distance is estimated as three or more.

As shown in panel (c), for both violent and non-violent tweets, around two-thirds of the retweets take place between pairs of users with a direct tie (62 and 67 percent, respectively). However, there is a substantial minority of retweets that travel beyond the tweeter's followers. Around one-third of the retweets take place between users whose estimated shortest path distance is two (31 and 27 percent). For the rest, tweets were retweeted over three or more ties (7 and 6

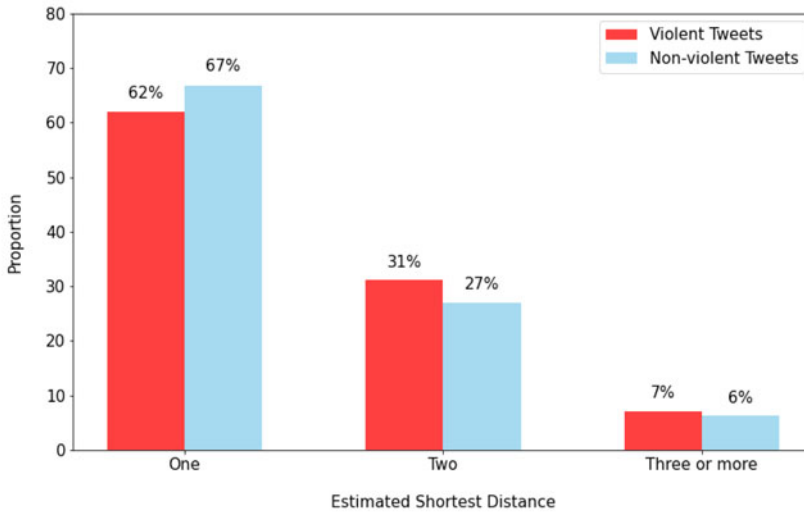
²¹In the data set, approximately 53 percent of (non-violent) political tweets are original tweets while the proportion is 75 percent for violent political tweets. It implies that violent tweets are retweeted less than non-violent tweets, which is consistent with the finding that violent users have fewer followers and their tweets get fewer likes (see [Figure 4](#)). However, even though such rhetoric is not written by those who share it, retweets of violent tweets still contain violent political rhetoric and ordinary users are exposed to and influenced by them.

²²Note that not all retweets take place through the following network although most retweets take place between connected users on the network (Fábrega and Paredes, 2012, 2013). For instance, Twitter has various affordances that enable users to connect with each other. For instance, users can simply search content and other users (Twitter, 2021e). Similarly, Twitter's algorithms provide users with popular topics or news, tailored based on who they follow, their interests, and their location (Twitter, 2021g).



(a) Ideological homophily (violent)

(b) Ideological homophily (non-violent)



(c) Reach of political tweets on the following network

Figure 6. Spread of tweets containing violent political rhetoric.

Note: For panels (a) and (b), each point in the plots expresses the number of retweets where the ideology scores of the tweeter and the retweeter correspond to the x-y coordinates. Higher values indicate greater conservatism. For panel (c), the height of the bars depicts the proportion of tweets containing violent political rhetoric whose shortest distance on the following network belongs to each category. For non-violent tweets, I use a random sample of 238 tweets due to a heavy limit on retrieving follower IDs in the Twitter API (Twitter, 2021d).

percent). The figure shows that political tweets in general spread widely and that violent tweets appear to spread just as far as non-violent tweets despite the offensive nature of the content.²³

²³Research on information diffusion in online platforms demonstrates that diffusion between users who are not directly connected is rare and becomes even rarer as the social distance between them increases. Working on seven different online diffusion networks (including the diffusion of US news articles and YouTube videos on Twitter), Goel et al. (2012) find that approximately 90 percent of the diffusion takes place between directly connected users. Similarly, focusing on randomly

Importantly, the findings imply that even if users do not follow a violent tweeter (even a violent tweeter's followers), it is still possible that they get exposed to such discomfiting content against one's intent. Also, the impact of violent tweets can be dramatically amplified beyond the personal follower networks of violent tweeters, if highly popular users—themselves not violent—retweet violent tweets thereby exposing a large number of users to them.

5 Conclusion

The recent violent hostility among ordinary American partisans, as dramatically expressed in the Capital Riot, has drawn immense attention both from the media and academia. While the previous literature tends to view partisanship positively as guidance for policy stance and vote choice (Campbell et al., 1980), such view is increasingly replaced by concerns about its destructive potential. At the same time, despite the clear benefits of social media for democracy such as political learning and participation (Dimitrova et al., 2014; Tucker et al., 2017), social media platforms are criticized and scrutinized for hateful and violent political communication and their role in stimulating and exacerbating offline violence between confronting partisans.

This paper is among the first to make sense of violent partisan hostility expressed online and thus contribute to the fields of grassroots political violence, online political communication, and violent partisanship. Methodologically, I introduce a new automated method that identifies violent political rhetoric from a massive stream of social media data, adding to the toolkit for measuring violent partisanship. Substantively, I demonstrate that violent political rhetoric on Twitter peaks in the days preceding the Capitol Riot, revealing its close relationship with contentious offline politics. Also, users who threaten violence are ideologically extreme and located on the fringe of the communication network. In terms of targeting, violent tweets are more frequently targeted at women and Republican politicians. While the number of violent tweets is small, such tweets often transcend direct inter-personal connections on the following network, amplifying their negative effects. Finally, such tweets are shared not only among like-minded users but also across the ideological divide, creating the potential for co-radicalization where ideologically extreme users further radicalize each other (Ebner, 2017; Pratt, 2017; Knott et al., 2018; Moghaddam, 2018).

In addition, the findings in this paper call for further research on the causes and consequences of violent political rhetoric. First, what are the causal relationships between violent political rhetoric online and offline political violence? While this paper presents abundant evidence for close relationships between the two, it is pressing for future research to scrutinize whether/how online and offline violent acts stimulate each other. Second, while recent research in political communication investigates the consequences of exposure to mildly violent political metaphors (Kalmoe, 2013, 2014, 2019; Kalmoe et al., 2018), little attention has been paid to an extreme form of violent language such as threats of violence. Therefore, it is crucial to investigate the effects of exposure to speech threatening violence against out-partisans. Does exposure to threatening messages have a contagion effect where exposed individuals come to endorse political violence? Alternatively, does it stimulate any corrective effort where individuals who encounter such norm-violating behavior oppose political violence?

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/psrm.2022.12>. To obtain replication material for this article, please visit <https://doi.org/10.7910/DVN/NEC17Z>.

sampld retweets (worldwide), Fábrega and Paredes (2012) report that over 80 percent of retweets are between directly connected users and approximately 7 percent of retweets are between pairs of users whose following distance is two. Retweeting beyond two hops on the following network was less than 2 percent. Given this, relative to general sharing behavior, the results in panel (c) provide evidence that the retweeting of violent tweets is generally far-reaching.

References

- Abramowitz A and Saunders K (2008) Is polarization a myth?. *The Journal of Politics* **70**, 542–555.
- Abramowitz A and Webster S (2016) The rise of negative partisanship and the nationalization of US elections in the 21st century. *Electoral Studies* **41**, 12–22.
- Abramowitz A and Webster S (2018) Negative partisanship: why Americans dislike parties but behave like rabid partisans. *Political Psychology* **39**, 119–135.
- Anderson CA and Bushman BJ (2002) Human aggression. *Annual Review of Psychology* **53**, 27–51.
- Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132.
- Barberá P (2014) How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University* **46**.
- Barberá P (2015) Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* **23**, 76–91.
- Barberá P, Jost JT, Nagler J, Tucker JA and Bonneau R (2015) Tweeting from left to right: is online political communication more than an echo chamber?. *Psychological Science* **26**, 1531–1542.
- Barrie C and Ho JC ting (2021) academictwitter: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software* **6**, 3272. <https://doi.org/10.21105/joss.03272>.
- Berry JM and Sobieraj S (2013) *The outrage industry: Political opinion media and the new incivility*. Oxford, United Kingdom: University Press.
- Blumenthal S (2021) The martyrdom of Mike Pence. *The Guardian*. <https://www.theguardian.com/commentisfree/2021/feb/07/mike-pence-donald-trump-republicans-religion-evangelical>.
- Borum R (2011a) Radicalization into violent extremism I: a review of social science theories. *Journal of Strategic Security* **4**, 7–36.
- Borum R (2011b) Radicalization into violent extremism II: a review of conceptual models and empirical research. *Journal of Strategic Security* **4**, 37–62.
- Brady WJ, Wills JA, Jost JT, Tucker JA and Bavel JJVan (2017) Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* **114**, 7313–7318.
- Brice-Saddler M. (2019) A man wrote on Facebook that AOC “should be shot,” police say. Now he’s in jail. *The Washington Post*. <https://www.washingtonpost.com/politics/2019/08/09/man-said-aoc-should-be-shot-then-he-said-he-was-proud-it-now-hes-jail-it/>.
- Broockman D, Kalla J and Westwood S (2020) Does affective polarization undermine democratic norms or accountability? Maybe not.”.
- Campbell A, Converse PE, Miller WE and Stokes DE (1980) *The American voter*. Chicago, Illinois, United States: University of Chicago Press.
- Chan J, Ghose A and Seamans R (2016) The internet and racial hate crime: offline spillovers from online access. *MIS Quarterly* **40**, 381–403.
- Cheng J, Danescu-Niculescu-Mizil C and Leskovec J (2015) Antisocial behavior in online discussion communities. In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 9.
- Claassen C (2016) Group entitlement, anger and participation in intergroup violence. *British Journal of Political Science* **46**, 127–148.
- Cohen M (2021) “Capitol rioter charged with threatening to ‘assassinate’ Rep. Ocasio-Cortez.”.
- Dadvar M, de Jong FMG, Ordelman R and Trieschnigg D (2012) Improved cyberbullying detection using gender information. In *Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012)*. University of Ghent.
- Daugherty N (2019) Former MLB player Aubrey Huff says he’s teaching his children about guns in case Sanders beats Trump. *The Hill*. <https://thehill.com/blogs/blog-briefing-room/news/472266-former-mlb-player-aubrey-huff-teaching-his-children-how-to-use>.
- Davidson T, Warmesley D, Macy M and Weber I (2017) Automated hate speech detection and the problem of offensive language. In *Eleventh international AAAI conference on web and social media*.
- Davidson S, Sun Q and Wojcieszak M (2020) Developing a new classifier for automated identification of incivility in social media. In *Proceedings of the fourth workshop on online abuse and harms*. pp. 95–101.
- Devlin J, Chang M-W, Lee K and Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. Preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805).
- Di Meco L and Brechenmacher S (2021) Tackling online abuse and disinformation targeting women in politics.
- Dimitrova DV, Shehata A, Strömbäck J and Nord LW (2014) The effects of digital media on political knowledge and participation in election campaigns: evidence from panel data. *Communication Research* **41**, 95–118.
- DiPasquale D and Glaeser EL (1998) The Los Angeles riot and the economics of urban unrest. *Journal of Urban Economics* **43**, 52–78.
- Druckman J, Klar S, Kkrupnikov Y, Levendusky M and Ryan JB (2020) The political impact of affective polarization: How partisan animus shapes COVID-19 attitudes.

- Ebner J** (2017) *The rage: The vicious circle of Islamist and far-right extremism*. London, United Kingdom: Bloomsbury Publishing.
- Fábrega J and Paredes P** (2012) Three degrees of distance on Twitter. preprint [arXiv:1207.6839](https://arxiv.org/abs/1207.6839).
- Fábrega J and Paredes P** (2013) Social contagion and cascade behaviors on Twitter. *Information* 4, 171–181.
- Felmlee D, Rodis PI and Zhang A** (2020) Sexist slurs: reinforcing feminine stereotypes online. *Sex Roles* 83, 16–28.
- Fiorina MP and Abrams SJ** (2008) Political polarization in the American public. *Annual Review of Political Science* 11, 563–588.
- Freelon D and Lokot T** (2020) Russian disinformation campaigns on Twitter target political communities across the spectrum. Collaboration between opposed political groups might be the most effective way to counter it. *Misinformation Review*.
- Fuchs Tamara and Schäfer Fabian** (2019) Normalizing misogyny: Hate speech and verbal abuse of female politicians on Japanese Twitter. In *Japan Forum*. Taylor & Francis, pp. 1–27.
- Fujii LA** (2011) *Killing neighbors: Webs of violence in Rwanda*. Ithaca, New York, United States: Cornell University Press.
- Gallacher JD** (2021) Online intergroup conflict: How the dynamics of online communication drive extremism and violence between groups. PhD thesis, University of Oxford.
- Gallacher J and Heerdink M** (2021) Mutual radicalisation of opposing extremist groups via the Internet.
- Gallacher JD, Heerdink MW and Hewstone M** (2021) Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media+ Society* 7, 2056305120984445.
- Gallego J, Martinez JD, Munger K and Vásquez-Cortés M** (2019) Tweeting for peace: experimental evidence from the 2016 Colombian Plebiscite. *Electoral Studies* 62, 102072.
- Gervais BT** (2015) Incivility online: affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics* 12, 167–185.
- Gervais BT** (2019) Rousing the partisan combatant: elite incivility, anger, and antideliberative attitudes. *Political Psychology* 40, 637–655.
- Gill P, Horgan J and Deckert P** (2014) Bombing alone: tracing the motivations and antecedent behaviors of lone-actor terrorists. *Journal of Forensic Sciences* 59, 425–435.
- Goel S, Watts DJ and Goldstein DG** (2012) The structure of online diffusion networks. In *Proceedings of the 13th ACM conference on electronic commerce*. pp. 623–638.
- Grimmer J and Stewart BM** (2013) Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21, 267–297.
- Guess A, Munger K, Nagler J and Tucker J** (2019) How accurate are survey responses on social media and politics. *Political Communication* 36, 241–258.
- Guynn J** (2021) “Burn down DC”: Violence that erupted at Capitol was incited by pro-Trump mob on social media. *USA Today*. <https://www.usatoday.com/story/tech/2021/01/06/trump-riot-twitter-parler-proud-boys-boogaloos-antifa-qanon/6570794002/>.
- Hammer HL, Riegler MA, Øvrelid L and Veldal E** (2019) Threat: A Large Annotated Corpus for Detection of Violent Threats. In *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, pp. 1–5.
- Han J, Pei J and Kamber M** (2011) *Data mining: Concepts and techniques*. Amsterdam, Netherlands: Elsevier.
- Hayes PJ and Weinstein SP** (1990) CONSTRUE/TIS: A system for content-based indexing of a database of news stories. In *IAAI*. Vol. 90, pp. 49–64.
- Henson B, Reynolds BW and Fisher BS** (2013) Fear of crime online? Examining the effect of risk, previous victimization, and exposure on fear of online interpersonal victimization. *Journal of Contemporary Criminal Justice* 29, 475–497.
- Hjorth F and Adler-Nissen R** (2019) Ideological asymmetry in the reach of pro-Russian digital disinformation to United States audiences. *Journal of Communication* 69, 168–192.
- Horowitz DL** (1985) *Ethnic groups in conflict*. Berkeley, CA: Univ. California Press.
- Huber GA and Malhotra N** (2017) Political homophily in social relationships: evidence from online dating behavior. *The Journal of Politics* 79, 269–283.
- Humphreys M and Weinstein JM** (2008) Who fights? The determinants of participation in civil war. *American Journal of Political Science* 52, 436–455.
- Hutchens MJ, Hmielowski JD and Beam MA** (2019) Reinforcing spirals of political discussion and affective polarization. *Communication Monographs* 86, 357–376.
- Imai K, Lo J and Olmsted J** (2016) Fast estimation of ideal points with massive data. *American Political Science Review* 110, 631–656.
- Isbister T, Sahlgren M, Kaati L, Obaidi M and Akrami N** (2018) Monitoring targeted hate in online environments. Preprint [arXiv:1803.04757](https://arxiv.org/abs/1803.04757).
- Itkowitz C and Dawsey J** (2020) Pence under pressure as the final step nears in formalizing Biden’s win. *The Washington Post*. <https://www.washingtonpost.com/politics/pence-biden-congress-electoral/2020/>.
- Iyengar S, Sood G and Lelkes Y** (2012) Affect, not ideology: a social identity perspective on polarization. *Public Opinion Quarterly* 76, 405–431.

- Iyengar S, Leikes Y, Levendusky M, Malhotra N and Westwood SJ (2019) The origins and consequences of affective polarization in the United States. *Annual Review of Political Science* 22, 129–146.
- Jigsaw (2020). <https://jigsaw.google.com/>.
- Jones SG (2020) War comes home: The evolution of domestic terrorism in the United States.
- Just JT, Barberá P, Bonneau R, Langer M, Metzger M, Nagler J, Sterling J and Tucker JA (2018) How social media facilitates political protest: information, motivation, and social networks. *Political Psychology* 39, 85–118.
- Kalmoe NP (2013) From fistfights to firefights: trait aggression and support for state violence. *Political Behavior* 35, 311–330.
- Kalmoe NP (2014) Fueling the fire: violent metaphors, trait aggression, and support for political violence. *Political Communication* 31, 545–563.
- Kalmoe NP (2019) Mobilizing voters with aggressive metaphors. *Political Science Research and Methods* 7, 411–429.
- Kalmoe NP and Mason L (2018) Lethal mass partisanship: Prevalence, correlates, and electoral contingencies. In *American Political Science Association Conference*.
- Kalmoe NP, Gubler JR and Wood DA (2018) Toward conflict or compromise? How violent metaphors polarize partisan issue attitudes. *Political Communication* 35, 333–352.
- Kates S, Tucker J, Nagler J and Bonneau R (2021) The times they are rarely A-Changin': circadian regularities in social media use. *Journal of Quantitative Description: Digital Media* 1.
- Kennedy MA and Taylor MA (2010) Online harassment and victimization of college students. *Justice Policy Journal* 7, 1–21.
- King G, Lam P and Roberts ME (2017) Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science* 61, 971–988.
- Klein A (2019) From Twitter to Charlottesville: analyzing the fighting words between the Alt-Right and Antifa. *International Journal of Communication* 13, 22.
- Knott K, Lee B and Copeland S (2018) Briefings: Reciprocal radicalisation. CREST. Online document <https://crestresearch.ac.uk/resources/reciprocal-radicalisation>.
- Krippendorff K (2018) *Content analysis: An introduction to its methodology*. Thousand Oaks, California, United States: Sage Publications.
- LaFree G and Ackerman G (2009) The empirical study of terrorism: social and legal research. *Annual Review of Law and Social Science* 5, 347–374.
- LaFree G, Jensen MA, James PA and Safer-Lichtenstein A (2018) Correlates of violent political extremism in the United States. *Criminology* 56, 233–268.
- Lang M, Nakhlawi R, Peter F, Moody F, Chen Y, Taylor D, Usero A, DeMarco N and Vitkovskaya J (2021) Identifying far-right symbols that appeared at the U.S. Capitol riot. *The Washington Post*. <https://www.washingtonpost.com/nation/interactive/2021/far-right-symbols-capitol-riot/>.
- Linder F (2017) Improved data collection from online sources using query expansion and active learning. Available at SSRN 3026393.
- Lytvynenko J and Hensley-Clancy M (2021) The rioters who took over the Capitol have been planning online in the open for weeks. *BuzzFeed*. <https://www.buzzfeednews.com/article/janelytyvnenko/trump-rioters-planned-online?scrolla=5eb6d68b7fedc32c19ef33b4>.
- MacKuen M, Wolak J, Keele L and Marcus GE (2010) Civic engagements: resolute partisanship or reflective deliberation. *American Journal of Political Science* 54, 440–458.
- Magu R, Joshi K and Luo J (2017) Detecting the hate code on social media. In *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 11.
- Mathew B, Dutt R, Goyal P and Mukherjee A (2019) Spread of hate speech in online social media. In *Proceedings of the 10th ACM Conference on Web Science*. pp. 173–182.
- Matsumoto D, Frank MG and Hwang HC (2015) The role of intergroup emotions in political violence. *Current Directions in Psychological Science* 24, 369–373.
- McGilloway A, Ghosh P and Bhui K (2015) A systematic review of pathways to and processes associated with radicalization and extremism amongst Muslims in Western societies. *International Review of Psychiatry* 27, 39–50.
- Miller B, Linder F and Mebane WR (2020) Active learning approaches for labeling text: review and assessment of the performance of active learning approaches. *Political Analysis* 28, 532–551.
- Moghaddam FM (2018) *Mutual radicalization: How groups and nations drive each other to extremes*. Washington, D.C., United States: American Psychological Association.
- Monroe BL, Colaresi MP and Quinn KM (2008) Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16, 372–403.
- Mooijman M, Hoover J, Lin Y, Ji H and Dehghani M (2018) Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour* 2, 389–396.
- Munger K (2017) Tweetment effects on the tweeted: experimentally reducing racist harassment. *Political Behavior* 39, 629–649.
- Munger K (2021) Don't@ me: experimentally reducing partisan incivility on Twitter. *Journal of Experimental Political Science* 8, 102–116.
- Newman M (2018) *Networks*. Oxford, United Kingdom: Oxford University Press.

- Nikolov A and Radivchev V** (2019) Nikolov–Radivchev at SemEval-2019 task 6: Offensive tweet classification with BERT and ensembles. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. pp. 691–695.
- O'Donnell C** (2020) Timeline: History of Trump's COVID-19 illness. *Reuters*. <https://www.reuters.com/article/us-health-coronavirus-trump>.
- Olteanu A, Castillo C, Boy J and Varshney KR** (2018) The effect of extremist violence on hateful speech online. In *Twelfth International AAAI Conference on Web and Social Media*.
- Pauwels LJR and Heylen B** (2017) Perceived group threat, perceived injustice, and self-reported right-wing violence: an integrative approach to the explanation right-wing violence. *Journal of Interpersonal Violence* **35**, 4276–4302.
- Pilkington Ed. and Levine S** (2020) "It's surreal": The US officials facing violent threats as Trump claims voter fraud. *The Guardian*. <https://www.theguardian.com/us-news/2020/dec/09/trump-voter-fraud-threats-violence-militia>.
- Popan JR, Coursey L, Acosta J and Kenworthy J** (2019) Testing the effects of incivility during internet political discussion on perceptions of rational argument and evaluations of a political outgroup. *Computers in Human Behavior* **96**, 123–132.
- Pratt D** (2017) Islamophobia as reactive co-radicalization. In *Religious Citizenships and Islamophobia*. Routledge, pp. 85–98.
- Rheault L, Rayment E and Musulan A** (2019) Politicians in the line of fire: incivility and the treatment of women on social media. *Research & Politics* **6**, 2053168018816228.
- Romm T** (2021) Facebook, Twitter could face punishing regulation for their role in U.S. Capitol riot, Democrats say. *The Washington Post*. <https://www.washingtonpost.com/technology/2021/01/08/facebook-twitter-congress-trump-riot/>.
- Scacco A** (2010) *Who riots? Explaining individual participation in ethnic violence* (Doctoral Dissertation), Columbia University, New York, New York, United States.
- Schils N and Pauwels LJR** (2016) Political violence and the mediating role of violent extremist propensities. *Journal of Strategic Security* **9**, 70–91.
- Settles B** (2009) Active learning literature survey.
- Shandwick W.** (2019) Civility in America 2019: Solutions for tomorrow.
- Siegel AA** (2020) Online hate speech. *Social Media and Democracy: The State of the Field, Prospects for Reform* pp. 56–88.
- Siegel AA, Nikitin E, Barberá P, Sterling J, Pullen B, Bonneau R, Nagler J, Tucker JA, et al.** (2021) Trumping hate on Twitter? Online hate speech in the 2016 US election campaign and its aftermath. *Quarterly Journal of Political Science* **16**, 71–104.
- Southern R and Harmer E** (2019) Twitter, incivility and “everyday” gendered othering: an analysis of tweets sent to UK members of Parliament. *Social Science Computer Review* **39**, 259–275.
- Stenberg CE** (2017) Threat detection in online discussion using convolutional neural networks. Master's thesis.
- Sterling J, Jost JT and Bonneau R** (2020) Political psycholinguistics: a comprehensive analysis of the language habits of liberal and conservative social media users. *Journal of Personality and Social Psychology* **118**, 805.
- Suhay E, Bello-Pardo E and Maurer B** (2018) The polarizing effects of online partisan criticism: evidence from two experiments. *The International Journal of Press/Politics* **23**, 95–115.
- Sydnor E.** (2019) *Disrespectful democracy: The psychology of political incivility*. New York, New York, United States: Columbia University Press.
- Tausch N, Becker JC, Spears R, Christ O, Saab R, Singh P and Siddiqui RN** (2011) Explaining radical group behavior: developing emotion and efficacy routes to normative and nonnormative collective action. *Journal of Personality and Social Psychology* **101**, 129.
- Theocharis Y, Barberá P, Fazekas Z, Popa S. and Parnet O** (2016) A bad workman blames his tweets: the consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of Communication* **66**, 1007–1031.
- Theocharis Y, Barberá P, Fazekas Z and Popa SA** (2020) The dynamics of political incivility on Twitter. *Sage Open* **10**, 2158244020919447.
- Tucker JA, Theocharis Y, Roberts ME and Barberá P** (2017) From liberation to turmoil: social media and democracy. *Journal of Democracy* **28**, 46–59.
- Twitter** (2021a) About replies and mentions. Accessed: 25 October 2021. <https://help.twitter.com/en/using-twitter/mentions-and-replies>.
- Twitter** (2021b) Academic research product track. Accessed: 25 October 2021. <https://developer.twitter.com/en/products/twitter-api/academic-research>.
- Twitter** (2021c) Filter realtime tweets. Accessed: 14 October 2021. <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/overview>.
- Twitter** (2021d) GET followers/ids. Accessed: 19 November 2021. <https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-followers-ids>.
- Twitter** (2021e) How to use Twitter search. Accessed: 3 November 2021. <https://help.twitter.com/en/using-twitter/twitter-search>.
- Twitter** (2021f) Search tweets: standard v1.1. Accessed: 14 October 2021. <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets>.
- Twitter** (2021g) Twitter trends FAQ. Accessed: 3 November 2021. <https://help.twitter.com/en/using-twitter/twitter-trending-faqs>.

- Twitter** (2021h) Violent threats policy. Accessed: 14 October 2021. <https://help.twitter.com/en/rules-and-policies/violent-threats-glorification>.
- Vaccari C, Valeriani A, Barberá P, Bonneau R, Jost JT, Nagler J and Tucker JA** (2015) Political expression and action on social media: exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy. *Journal of Computer-Mediated Communication* **20**, 221–239.
- van der Vegt I, Mozes M, Gill P and Kleinberg B** (2019) Online influence, offline violence: Linguistic responses to the “Unite the Right” rally. Preprint [arXiv:1908.11599](https://arxiv.org/abs/1908.11599).
- Vigdor N** (2019) Police officer suggests AOC should be shot: “She needs a round”. *Independent*. https://www.independent.co.uk/news/world/americas/us-politics/aoc-trump-twitter-democrats-louisiana-police-charlie-rispoli-a9015301.html&utm_source=shareutm_medium=ios_app.
- Wang S, Chen Z, Liu B and Emery S** (2016) Identifying search keywords for finding relevant social media posts. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 30.
- Waseem Z and Hovy D** (2016) Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*. pp. 88–93.
- Weerkamp W, Balog K and Rijke M de** (2012) Exploiting external collections for query expansion. *ACM Transactions on the Web (TWEB)* **6**, 1–29.
- Wei K** (2019) Collective action and social change: How do protests influence social media conversations about immigrants? PhD thesis University of Pittsburgh.
- Wester AL** (2016) Detecting threats of violence in online discussions. Master’s thesis.
- Wester A, Øvrelid L, Veldal E and Hammer HL** (2016) Threat detection in online discussions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. pp. 66–71.
- Westwood S, Grimmer J, Tyler M and Nall C** (2021) American support for political violence is low.
- Williams ML, Burnap P, Javed A, Liu H and Ozalp S** (2020) Hate in the machine: anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology* **60**, 93–117.
- Wojcik S and Hughs A** (2019) Sizing up Twitter users. <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>.
- Wulczyn E, Thain N and Dixon L** (2017) Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th international conference on world wide web*. pp. 1391–1399.
- Zeitsoff T** (2020) The nasty style: Why politicians use violent rhetoric. *Unpublished working paper*.
- Zimmerman S, Kruschwitz U and Fox C** (2018) Improving hate speech detection with deep learning ensembles. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.