RESEARCH ARTICLE

Online influence, offline violence: language use on YouTube surrounding the 'Unite the Right' rally

Isabelle van der Vegt[1] · Maximilian Mozes[1,2,3] · Paul Gill[1] ·
Bennett Kleinberg[1,2]

Abstract

The media frequently describes the 2017 Charlottesville 'Unite the Right' rally as a turning point for the alt-right and white supremacist movements. Social movement theory suggests that the media attention and public discourse concerning the rally may have engendered changes in social identity performance and visibility of the alt-right, but this has yet to be empirically tested. The presence of the movement on YouTube is of particular interest, as this platform has been referred to as a breeding ground for the alt-right. The current study investigates whether there are differences in language use between 7142 alt-right and progressive YouTube channels, in addition to measuring possible changes as a result of the rally. To do so, we create structural topic models and measure bigram proportions in video transcripts, spanning approximately 2 months before and after the rally. We observe differences in topics between the two groups, with the 'alternative influencers', for example, discussing topics related to race and free speech to a larger extent than progressive channels. We also observe structural breakpoints in the use of bigrams at the time of the rally, suggesting there are changes in language use within the two groups as a result of the rally. While most changes relate to mentions of the rally itself, the alternative group also shows an increase in promotion of their YouTube channels. In light of social movement theory, we argue that language use on YouTube shows that the Charlottesville rally indeed triggered changes in social identity performance and visibility of the alt-right.

Keywords Alt-right · Charlottesville rally · Structural topic modelling · YouTube · Quantitative text analysis

✉ Isabelle van der Vegt
   isabelle.vandervegt@ucl.ac.uk

[1]   Department of Security and Crime Science, University College London, London, UK

[2]   Dawes Centre for Future Crime, University College London, London, UK

[3]   Department of Computer Science, University College London, London, UK

Introduction

On 11 and 12 August 2017, dozens of alt-right, white supremacist and neo-Nazi individuals descended on Charlottesville, Virginia. The event, known as the 'Unite the Right' rally, turned fatal on the second day when a white supremacist deliberately drove into a crowd of counter-protestors, resulting in the death of one person and leaving several others injured [1, 2]. In recent years, the rise of the alt-right has been accompanied by several other acts of violence and terror attacks motivated by white supremacist ideologies, with 18 out of 34 extremist-related deaths in 2017 attributed to this group [3]. In 2019, 90% of all 42 extremist murders in the United States were linked to right-wing extremism [4]. At the same time, alt-right ideologies have become widespread online. Their content is easily accessible through social media platforms, and ideas are amplified on websites such as 4chan [5] and Gab [6]. YouTube, in particular, has been described as a breeding ground for the alt-right [7, 8].

This paper sheds light on the alt-right as a social movement by studying its language use in a unique dataset of YouTube video transcripts. We examine whether the Charlottesville rally functioned as a critical juncture in the online behaviour of the alt-right, and additionally contrast this with language use in a progressive sample of YouTube channels. In the next section, we discuss the alt-right and the Charlottesville rally. Thereafter, we outline the wider social movement literature as well as previous work on the effect of offline trigger events for online behaviour. Following this, we introduce our empirical examination of differences in language use on YouTube within and between alt-right and progressive channels, shortly before and after the Charlottesville rally.

The alt-right and Charlottesville

The alt-right is not defined by a central organisation [9], nor does it 'offer a coherent or well-developed set of policy proposals' [10]. Instead, it has been referred to as a 'mix of rightist online phenomena' [11] with white identity at its core [12]. The alt-right is variously characterised as anti-political correctness, anti-immigration, anti-Semitist, and anti-feminist [12], ideologies which are commonly spread online through irony and dark humour. Scholars have begun to study the alt-right as a social movement, following the definition of 'a cluster of performances organised around a set of grievances or claims' [9, 13]. It has been argued that the alt-right, mainly through online activity, engages in promoting a shared identity, fostering a commitment to a common cause, and proclaiming the 'worthiness, unity, and size' of its movement [9].

The presence of the alt-right on social media has been particularly salient on YouTube [7, 8]. A 2018 report described an 'Alternative Influencer Network' on YouTube consisting of content creators 'who range in ideology from mainstream libertarian to openly white nationalist' [8]. It was found that alternative political influencers on YouTube adopt strategies of mainstream popular Youtubers to gain popularity, engaging in tactics for search engine optimisation and cultivating a

relatable 'underdog' image [8]. Further research on this network argued that a "supply-and-demand framework" is needed to understand the popularity of alternative influencers, where the ease of uploading and monetizing fringe political videos on YouTube enables a supply that is in demand for viewers who feel alienated from mainstream media [14]. It has also been noted that the audience of alt-right YouTube videos is highly engaged with the content, displaying more likes and comments per view than other less extreme or mainstream media videos [14]. While some videos or channels of extreme influencers may have been demonetized because advertisers do not want to be associated with the content, many have adopted alternative strategies to raise revenue. This includes the use of crowdfunding platforms such as Patreon or so-called "super-chats" where viewers make a donation for their message to be read out on a livestream [14].

After the Charlottesville rally, various media outlets declared that 'white nationalists are winning' [15] and 'the genie is out of the bottle' [1]. In addition, President Trump stated that there was 'blame on both sides' [16], which prompted the suggestion that his claims 'reinvigorated' the alt-right movement [16]. In the aftermath of the rally, various reports also noted that white nationalists have entered mainstream conversation [1, 17] and some say they were aided in doing so by the Trump administration [17]. Drawing from the study of protests by extreme right-wing groups and other social movements, one might argue that that the rally was not only important for the effects it had outside of the movement (e.g., in the media and politics), but also within the movement itself [18, 19], which could be assessed by studying its YouTube videos. In the next section, we outline some of the social movement literature to better understand the alt-right and possible effects of the Charlottesville rally.

Social movement theory

Social movements have been studied for decades [20–25]. One definition states that a social movement is a group containing 'a plurality of individuals, groups and/or organizations, engaged in political or cultural conflicts, on the basis of shared collective identities' [20]. Within these movements, the collective identity of the group can be actively emphasized through distinguishing between "us" and "them" [22, 24]. These identities crucially need to be "framed" to mobilize supporters, where the frame generally serves to identify an injustice which can be addressed through a collective agency [25]. Importantly, it is consistently shown that social movements make extensive use of the internet for communication and organisation [21, 23]. Indeed, further definitions of social movements state that resources are generally shared through informal networks [21, 26]. This phenomenon has also been studied within the context of white supremacist groups, for which the internet serves to reinforce their sense of collective identity, where white supremacy and difficulties faced by white people are emphasized [27].

Elements of social movement theory propose that people engage in social identity performance, which refers to behaviour that serves to express the norms of the social group one aims to belong to [28, 29]. Such behaviour includes affirming ones social identity, conforming to a social movement, strengthening one's identity, or

mobilising others [28]. Within the context of the alt-right, social identity performance may, for example, include using community-specific language [5] or memes online (e.g., Pepe the Frog, a popular internet meme appropriated by the alt-right [5, 10]), or to publicly adopt symbols related to white supremacism.

Research on the effect of media coverage and public discourse on social movements might explain the potential effect of the Charlottesville rally on the alt-right [30, 31]. For example, research on right-wing violence in Germany suggests that both positive and negative reactions from public figures to violent events may help to lend prominence to the movement [31]. That is, even if one aims to condemn a violent movement's message, the message is (at least partially) reproduced [31]. By studying newspaper sources, this line of research suggested that discursive opportunities, summarised as public visibility, resonance, and legitimacy affected the behaviour of right-wing movements, measured in terms of violent events against different target groups [31]. Public visibility refers to the number of outlets reporting on the movement and the prominence of the movement's message within those outlets [31]. Resonance is defined as the (positive or negative) reaction from public figures to the movement's message as well as the associated ripple effect in the media [31]. Legitimacy involves the general public's support of a message [31]. Similar discursive opportunities were also studied in relation to the rise in popularity of right-wing populist Pim Fortuyn in the Netherlands [30]. In a similar vein, visibility (e.g., the extensive media coverage), resonance (e.g., responses to the rally from President Trump and other politicians), and legitimacy (e.g., subsequent protests and vigils denouncing the rally [32]) can be observed in the context of the alt-right and Charlottesville rally.

The effect of discursive opportunities has yet to be examined for the specific case of the alt-right and the Charlottesville rally. If indeed the visibility of the alt-right increased following the rally, the message of the movement resonated in the media and public discourse, and the alt-right gained legitimacy through acknowledgement from opponents and the general public, we may expect to see changes in behaviour within the movement. Within the context of social identity performance, one may expect to see strengthened social identity consolidation within the alt-right movement as a result of the rally, President Trump's comments, and the media coverage of the rally. After the rally, we might expect increased expression of norms from the alt-right movement, for example in the form of stronger endorsement or more extreme expressions of in-group ideology. As has been raised previously, such behaviour may serve to further strengthen the movement or mobilise others to join.

In summary, the Charlottesville rally may have had an effect both within and beyond the alt-right, namely on their social identity performance and visibility, respectively. Similar theories have been proposed within the study of protests by extreme right-wing groups and other social movements. Large (sudden) protests are sometimes said to not only have important effects outside a social movement but also within the movement itself by further radicalising or mobilising (non) members [18, 19]. Within this context, it is said that protests sometimes can trigger critical junctures that bring about abrupt and lasting changes both within and beyond a social movement [19]. In the next section, we further examine the effect of offline trigger events on online behaviour, as well as the interaction between the two domains.

Reactions to 'trigger' events

A large body of research has examined the interplay between online activity and offline events, particularly how both domains may influence each other. Early work in this area already suggested that the internet was transforming collective action by having a mobilizing influence on its users [33]. For instance, it has been argued that the online discussion within social movements influences the politicized identity of individuals (e.g., identification with a movement), which in turn influences their intentions to engage in collective action (e.g., attending a rally) [34]. Similar claims have been made in light of the Arab spring, where online activity has been said to enable the formation of a new social identity (i.e., opposing the government) and mobilized people to engage in mass protests [35]. Besides political contexts, it has for example also been shown that online interactions in addiction recovery support groups (e.g., affirmation through likes, identification with the recovery community expressed in language) predicted offline retention in the program [36, 37]. Besides offline (collective) action, the online activity also seems to have an effect on offline media. For example, a 'symbiotic' relationship has been identified between Twitter feeds and top newspapers. Examining tweets from 2016 US presidential candidates and issue agendas in five US newspapers, it was found that tweets (e.g., on employment, immigration, national security) frequently predicted news agendas, and vice versa [38]. In another study, it was suggested Tweets can be used to infer voter preferences [39]. Political party mentions and tweet sentiment were said to reflect actual election results in Germany [39].

Of particular interest is the measurement of (hate) crimes in response to specific 'trigger' events, such as terrorist attacks [40]. Several studies have reported spikes in hate crimes following 9/11 or the 7/7 London attacks [41, 42]. In the online sphere, similar patterns can be observed. A survey conducted between 2013 and 2015 also showed that young people in Finland witnessed increased hate online shortly after the 2015 Paris attacks [43]. In the aftermath of the 2013 Woolwich terrorist attack, researchers observed hate directed at black and minority ethnic groups in Tweets directly related to the attack [40]. In another study on information flows on Twitter in the aftermath of the Woolwich attack, it was found that tweet sentiment was predictive of the number retweets and their timespan (i.e., time between first and last retweet). Offline news reports also (positively) predicted the number of retweets on the same topic [44]. Another study examined the effect of jihadist terrorism and Islamophobic attacks on hate speech on Twitter and Reddit, measured over a period of 19 months shortly after 13 extremist attacks [45]. It was found that, following jihadist terrorist attacks, hate speech targeting Muslims, particularly those advocating violence, increased more after terror attacks compared to a counterfactual simulation [45]. An increase in hate speech targeting Muslims was not found following Islamophobic attacks, with the exception of messages posted after the 2017 Finsbury Park Mosque attack [45]. Hate speech and white nationalist rhetoric have also been measured during the 2016 US elections on Twitter, using a dictionary approach with Hatebase, the Racial Slur Database, and the Anti-Defamation League's database of

white-nationalist language[1] [46]. Tweets were examined by means of an interrupted time series analysis, showing a spike in hate speech in the Trump dataset following the imposed travel ban in early 2017 [46].

A small number of studies have looked at the specific effect of the Charlottesville rally on online behaviour. In a qualitative study of Twitter accounts of two alt-right and one far-left organisation in the 6 weeks leading up to the Charlottesville rally, it was observed that the two sides frequently targeted each other, framing the opposing group as the enemy [47]. Manual examination of the tweets showed that the alt-right accounts frequently referred to 'the left' and 'liberals' as unpatriotic and communist. At the same time, the far-left accounts dubbed the alt-right 'suit and tie Nazis'. Furthermore, both the alt-right and far-left groups incited violence in the weeks leading up to the Charlottesville rally and called for action among their supporters. A tweet from one of the alt-right groups read 'The left is preparing to lynch mobs to descend on the Unite The Right rally in Charlottesville, VA… This is going to be fun' [47]. Other research has shown that anti-Semitic memes and rhetoric increased after the 2016 US elections and the Charlottesville rally [48]. Several million posts and images from 4chan and Gab were studied for racial slurs and anti-Semitic terms, with a case study of a specific anti-Semitic meme showing that such content also spreads to mainstream platforms such as Twitter and Reddit [48].

Aims of this paper

Taken together, the theoretical lines discussed would suggest that the Charlottesville rally functioned as a critical juncture for the alt-right, engendering changes in online social identity performance and visibility of the social movement. To empirically examine this claim, the present study takes a closer look at a network of alternative political influencers [7] (hereafter, 'alternative group'), by examining YouTube video transcripts extracted from channels by these individuals. These video transcripts are compared to those from YouTube channels whose political orientation can be considered more progressive (hereafter, 'progressive group'), to assess whether the Charlottesville rally also had an effect outside of the alt-right movement itself.

This paper has two aims. First, we compare language use *between* the alternative and progressive group in a sixteen-week timeframe surrounding the Charlottesville rally. Second, we assess whether the rally had an effect on language use *within* the two groups. For the alternative group, we do not postulate any directional hypotheses about changes in language use. Nevertheless, in light of the social movement and social identity performance literature, we expect to see changes in social identity performance after the rally reflected in language use on YouTube. For the progressive group, we do not claim that the channels studied act as a social movement, and thus we have no expectations of social identity performance. However, we are

---

interested in seeing whether the channels lend any discursive opportunities to the alt-right through language use in their videos, thereby potentially contributing to the increased visibility of the alternative group.

The first aim is addressed through structural topic modelling, to compare the prevalence and content of topics between the two groups. The second aim is addressed using a word frequency approach, in which we examine the frequency of common phrases before and after the Charlottesville rally, searching for sudden increases or decreases as a result of the rally.

Method

Data availability statement

Supplemental materials, data and code to reproduce the analysis are available on the Open Science Framework: https://osf.io/yedt7/.

Data

Channel selection

YouTube channels were selected for analysis from two main sources. First, we drew from the list of 65 YouTube users referred to as the 'Alternative Influence Network' in the 2018 Data & Society report on political influencers [8]. Based on this list, we searched for a designated YouTube channel for each individual. If an individual did not have a designated YouTube channel or their channel was no longer available, we searched for the individual's name through the YouTube search function. For example, videos featuring Alex Jones (who was banned from YouTube so no longer has a designated channel) were obtained through the search query 'alex jones full show'. The group of alternative YouTube channels consisted of 56 channels and search queries used for transcript retrieval. Because data collection was done retrospectively, some channels appearing in the Data & Society report may not have been available (also when searched) because they were banned or deleted (total of 9 channels, 13.85%). Second, for the comparison group of progressives, we drew from two online lists of progressive YouTube channels.[2] Since the lists referred to specific existing channels, search queries for specific persons were not necessary. In total, 13 progressive channels and 56 alternative channels were used for transcript retrieval. For all channels and search queries, we retrieved the URLs for all available videos on 1 October 2018.

---

[2] http://the2020progressive.com/top-13-progressive-news-shows-youtube/ and https://medium.com/@tejazz89/top-5-youtube-channels-to-follow-if-you-are-a-true-progressive-ee2abc78d58f.

Table 1 Descriptive statistics video sample

	Alternative	Progressive
Total videos	2684	4458
Total word count	3,868,744	2,804,703
Word count	Mean: 1448 (SD: 1612) Min: 34, Max: 12,085	Mean: 632 (SD: 860) Min: 33, Max: 11,645
View count	Mean: 120,639 (SD: 234,513) Min: 2, Max: 17,340,303	Mean: 24,787 (SD: 47,468) Min: 12, Max: 2,475,766

Transcript retrieval

The method for retrieving YouTube video transcripts follows the procedure of related research [49, 50]. To retrieve the transcripts, a Python script was written using www.downsub.com to obtain XML-encoded transcripts. The transcripts were either automatically generated by YouTube or manually added by the YouTube user. In some cases, no transcript was available, because users disabled the transcript availability. XML-tags and time-stamps were removed, resulting in a single, non-punctuated string for each video transcript.

Data cleaning

Videos that contained fewer than 100 words were not considered for analysis, following previous work on YouTube transcripts [49, 50]. Using R software, each video was checked for English language and was excluded if it contained fewer than 50% English words. Videos were also excluded if they contained fewer than 90% ASCII characters. The video transcript strings were lower-cased and stopwords, unnecessary whitespace or punctuation were removed using the R packages *tidytext* [51], *tm* [52] and *qdap* [53].

Sample

To capture the immediate and continuing effects of the rally we sampled video transcripts up to approximately 2 months after the rally, as well as an equal timeframe preceding the rally. Previous works assessing the online effects of offline events have examined timeframes ranging from 2 weeks [44], a month [39, 40, 43], to one or several years [42, 45, 54]. Because no consensus seems to exist in the literature, we opted for a middle ground of 2 months pre- and post-event (data from a longer timeframe is available on request). This resulted in a total sample of videos spanning 16 weeks (i.e., 8 weeks pre- and post-rally). Descriptive statistics for this sample are given in Table 1.

Analysis plan

Structural topic model

To assess the differences in language use between the alternative and progressive groups, we construct a structural topic model. This method can be used to automatically extract underlying latent topics in a corpus [55, 56]. Common approaches include Latent Dirichlet Allocation [55] and Correlated Topic Models [57], probabilistic models which are based on the assumption that a piece of text consists of a mix of topics, which in turn are a mix of words with probabilities of belonging to a topic [55, 56]. A structural topic model is a type of Correlated Topic Model, with the added benefit that one can incorporate document-level covariates (e.g., document author, political orientation, date) and assess whether these covary with topic prevalence (i.e., the degree to which documents in a corpus are assigned a specific topic) and content (i.e., the terms in a topic) [56].

We first define a document-frequency-matrix with both unigrams (e.g., 'president') and bigrams (e.g., 'donald trump') in the corpus, which is then used to construct the structural topic model. We include group (alternative vs. progressive) as a covariate for topic prevalence and content. Topic models are fit with a varying number of topics,[3] after which we select the best fitting model based on the trade-off between semantic coherence and exclusivity [56, 58], two metrics frequently used to assess whether a topic is semantically useful [56, 59]. Semantic coherence is a measure of the co-occurrence of highly probable words in a topic and has been shown to correlate with expert judgments of topic quality [58]. It has been proposed that a measure of exclusivity of words to topics is needed to further determine topic quality, otherwise, several topics may be represented by the same highly probable words if one relies on semantic coherence alone. Exclusive topics are made up of words that have a high probability under one topic, but a low probability under other topics [59].

After selecting a model, we present topics for which a significant effect of the covariate group was found for topic prevalence (i.e., a significant difference between alternative and progressive channels), in order of total expected topic proportion for the corpus. Based on manual inspection of frequent and exclusive topic words [56, 60], we assign labels to topics. We also present a selection of three topics along with the words which differed between the alternative and progressive groups, to illustrate how the alternative and progressive groups talk about the same topic in different ways (i.e., content differences).

Word frequency

To examine possible changes in word frequency within the alternative and progressive group as a result of the rally, we compute the frequency of all bigrams for each

---

[3] 20, 40, 50, 60, 70, 80, and 90 topics.
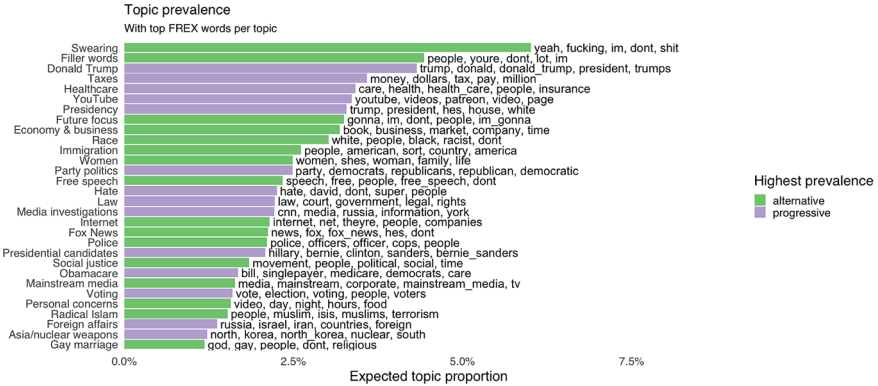
Fig. 1 Topic prevalence per group

week in both groups separately. By dividing these values by the total number of bigrams for each day, we obtain the daily proportion for each bigram. Thereafter, we can assess whether there is a structural breakpoint in the proportion of each bigram as a result of the rally. This is done by means of the Chow test [61, 62], with which we determine whether a breakpoint in the intercept and slope occurred at the time of the rally. To do so, we test for the equality between a model of bigram proportions before the Charlottesville rally and a model of bigram proportions after the rally. In both models, the proportion of each bigram is represented as a function of Date (day on which the proportion was measured, between 15 June and 7 Oct 2017). We compute an F-value for the equality between the two models for each bigram and report those which are found to differ significantly pre- and post-rally. We also report the associated Cohen's *d* effect size, where a value of of 0.2, 0.5, and 0.8 constitute a small, medium, and large effect, respectively [70]. In addition, we present associated intercept and slope changes.

Results

Structural topic model

We decided on a structural topic model with 40 topics based on examination of semantic coherence and exclusivity (see the OSF page for results with different numbers of topics). Thereafter, we found that the covariate Group was significant for the prevalence of 30 topics. Figure 1 shows the topics for which Group significantly covaried with topic prevalence. We assigned labels (e.g., 'Obamacare') based on the examination of highly probable as well as frequent and exclusive words. The alternative group discussed more of the topics which were labelled as swearing, filler words, future focus, economy & business, race, immigration, women, free speech, internet, Fox News, police, social justice, mainstream media, personal concerns,

Table 2 Topic content per group

Topic label	Group	Common terms
Social justice	Alternative	diversity, social justice, google, differences, identity politics
	Progressive	movement, conference, resistance, africa, organization
Women	Alternative	women, female, sexual, male, rape
	Progressive	women, sex, father, child, family
Race	Alternative	white people, black people, racist, blacks, racism
	Progressive	charlottesville, white supremacist, racism, racist, nazis

Table 3 Ten bigrams with largest Chow test (F) statistic in an alternative group

Bigram	Chow test (F)[a]	Effect size d	Intercept change	Slope change[b]
White nationalist	42.84	0.89	1.12	−0.64
Happen charlottesville	38.09	0.84	0.20	−0.11
Hit bell	35.47	0.81	0.26	−0.15
Video bitcoin	31.85	0.76	0.20	−0.11
Subscribe hit	28.92	0.73	0.24	−0.14
Charlottesville virginia	28.10	0.72	0.11	−0.06
nazi flag	26.86	0.70	0.19	−0.11
Descript patreon	26.70	0.70	0.19	−0.11
Neonazi white	26.00	0.69	0.11	−0.06
Begin video	25.90	0.69	0.33	−0.19

[a]For all coefficients (F, intercept and slope changes): *p* < 0.001

[b]Slope change estimates have been multiplied by 10,000 for interpretability

radical Islam, and gay marriage. The progressive channels focused more on Donald Trump, taxes, healthcare, YouTube, the presidency, party politics, hate, law, media investigations, presidential candidates, Obamacare, voting, foreign affairs and Asia/nuclear weapons.
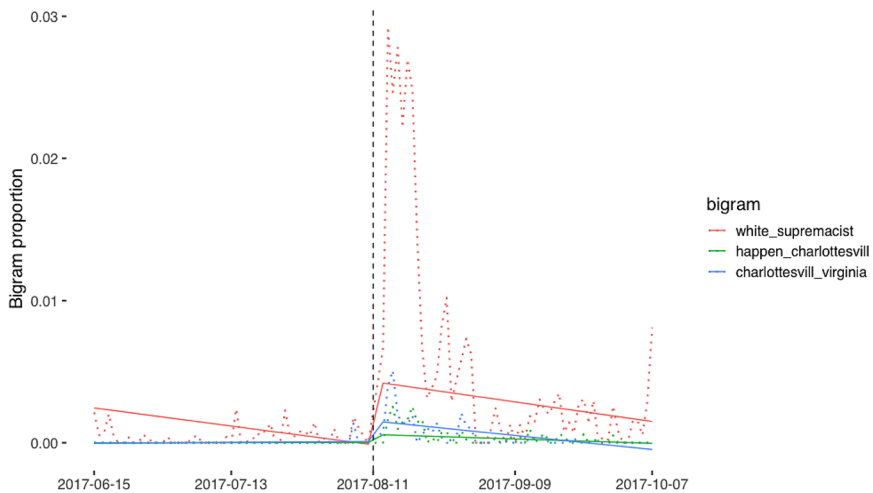
Table 2 shows how three topics are discussed differently by alternative and progressive channels (full list of topics available on the OSF). By including topic content as a covariate, we are able to see which words are more associated with each group per topic. For example, the 'social justice' topic is discussed as a 'movement' and 'resistance' by progressive channels, whereas the alternative group uses the term 'identity politics'. The topic of 'women' is discussed with terms referring to sexuality by both groups, but the progressive group also includes terms referring to

Table 4 Ten bigrams with largest Chow test (F) statistic for progressive group

Bigram	Chow test $(F)^a$	Effect size d	Intercept change	Slope change[b]
White nationalist	42.84	0.89	1.12	−0.64
Happen charlottesville	38.09	0.84	0.20	−0.11
Hit bell	35.47	0.81	0.26	−0.15
Video bitcoin	31.85	0.76	0.20	−0.11
Subscribe hit	28.92	0.73	0.24	−0.14
charlottesville virginia	28.10	0.72	0.11	−0.06
Nazi flag	26.86	0.70	0.19	−0.11
Descript patreon	26.70	0.70	0.19	−0.11
Neonazi white	26.00	0.69	0.11	−0.06
Begin video	25.90	0.69	0.33	−0.19

[a]For all coefficients (F, intercept and slope changes): $p < 0.001$

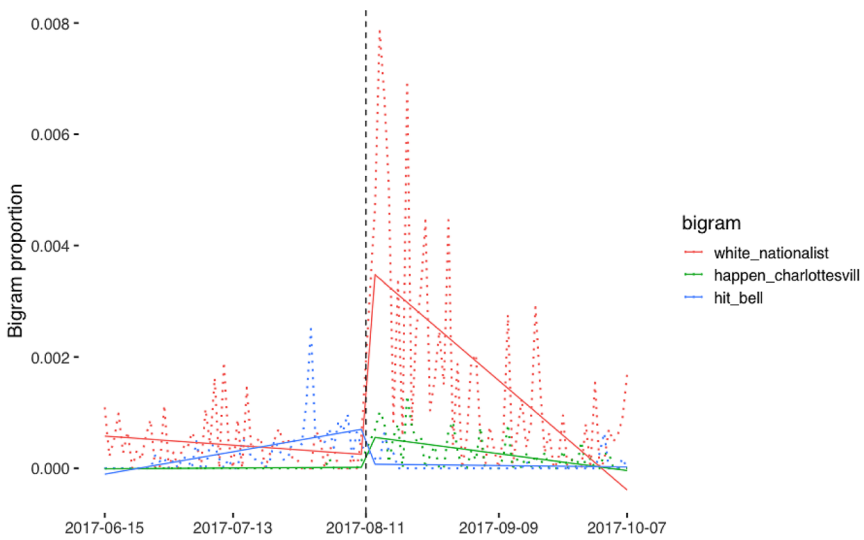[b]Slope change estimates have been multiplied by 10,000 for interpretability

Fig. 2 Observed proportion of three bigrams with largest F-value in an alternative group

family. Although both topics discuss 'race' with terms relating to racism, the progressive group uses terms such as 'white supremacist' and 'nazis'.

Word frequency approach

We show the ten bigrams for which the Chow test F statistic (and associated effect size), indicative of a joint breakpoint in intercept and slope, was largest,[4] in the

---

[4] Further bigrams that exhibited breakpoints are available in the supplemental materials.

Fig. 3 Observed proportions of three bigrams with largest F-value in the progressive group

alternative group (Table 3) and progressive group (Table 4). We also show the direction and magnitude of intercept and slope changes after the rally; please note that the values of slope changes were small (albeit statistically significant and large in terms of effect size *d*) and, therefore have been multiplied by 10,000 for interpretability. Among the top ten bigrams with breakpoints for both groups, the majority relate to the rally itself, such as 'white nationalist' and 'happen charlottesville'. Note that some bigrams showed a breakpoint in both groups, namely, 'white nationalist', 'happen charlottesville', 'charlottesville virginia', and 'neonazi white'. In the alternative group, several bigrams unrelated to the rally (e.g., 'hit bell', 'video bitcoin') also exhibit strong breakpoints. In the progressive group, only one bigram with a strong breakpoint in the top ten seems to be unrelated to the rally, namely 'hurricane maria'. To further illustrate the bigram proportion breakpoints, we show the progression of the first three (based on the magnitude of the Chow test *F*) bigrams for the alternative group (Fig. 2) and the progressive group (Fig. 3). In both groups, the proportion of the bigrams depicted significantly increases in terms of intercept, with slight (negative) changes in slopes.

Discussion

The current study examined language use for alternative and progressive YouTube channels around the time of the Charlottesville 'Unite the Right' rally. The aims of this paper were to compare language use between the groups surrounding the rally and to assess whether the rally had an effect on language use within the two groups.

We examined language use in both groups in terms of structural topic models and searched for structural breakpoints in a change of content as a result of the rally. We consider the outcome of both approaches, in turn, followed by an interpretation of the results in light of social movement theory.

Differences between alternative and progressive channels

The first line of inquiry examined whether there were structural differences in the prevalence and content of topics between groups. This analysis illustrates the matters discussed in videos throughout the studied period in the two groups. Perhaps unsurprisingly, several topics in both groups related to politics and current events (e.g., taxes, healthcare and the economy). We found that the prevalence of the majority of topics covaried with the political orientation of channels (alternative or progressive). For instance, topics that may be loosely associated with the 'ideology' of the alt-right were found to be used more by the alternative group, such as race, immigration, radical Islam, gay marriage, and free speech [11, 12]. Indeed, the concept of free speech has frequently been linked to the alt-right and white nationalism, where the right to free speech is used to "advance racist and sexist ideas" [63]. In a similar vein, discussions relating to women's and LGBT rights as well as social justice which appeared in our corpus have also been linked to the far right [64], a further potential indicator of expressing social norms within this group. The topic of so-called mainstream media was also discussed more by alternative channels, as well as Fox News in particular. In contrast, the progressive channels discussed Donald Trump to a larger extent, as well as other more general current affairs, such as the Democratic and Republican parties, legal matters, Obamacare, and foreign politics. Interestingly, we also observed a difference in the prevalence of swearing, which was significantly higher for alternative influencers. Swearing may be a way of conforming to a social group, and our results suggest that this kind of language is more common among alternative than progressive YouTube channels. The content of topics further elucidated differences between groups, for example, the way in which the alternative and progressive channels discussed the topic of race with differential terms, with the latter using terms that seem to condemn racism (e.g., 'white supremacist', 'nazis'). In short, the structural topic models indeed show that there are differences in topics between alternative and progressive YouTube channels. Some of these patterns in topics may support previous claims that the alt-right behaves as a social movement [9].

Effects of the rally within alternative and progressive channels

The word frequency approach showed the rally had an effect on language use within the two groups, illustrated by several breakpoints in bigram proportions that coincided with the Charlottesville rally. Unsurprisingly, the use of words relating to the rally (e.g., confederate monument, white nationalist, white supremacist) increased at this point. While the proportions of these bigrams all exhibited sudden increases, the mentions did decrease over time in the post-rally timeframe. This possibly reflects

a 'natural' descending trend for discussions of an event as time progresses, which potentially adds to the justification of measuring bigram proportions over time to assess reactions to events in language.

Although there was some overlap between groups in bigram use, it also appears that both groups discussed the events in a different light. The progressive group increasingly mentions 'white supremacists' after the rally, whereas the alternative group increasingly mentions 'white nationalists'. These differences in terminology seem to reflect a more general divide between groups. Indeed, 'white supremacists' is a term preferred by people who study or condemn the movement, but the term is not preferred among the extreme right itself [12]. Among the alt-right, the preferred term is 'white nationalist', which indeed emerges from our data [12]. This preference relates to the wish to establish separate white *nations*, in contrast to multiracial nations where whites are the dominant ('supreme') group [12]. One could argue that this difference in terminology may reflect increased expressions of in-group (alt-right) norms, an aspect of social identity performance.

Further breakpoints observed in the progressive group refer to several details related to the rally, such as the confederate statue of Robert Lee, the removal of which gave rise to the Charlottesville rally [65]. A strong increase within progressive post-rally videos was observed for the mention of counter-protestors, highlighting potential condemnation of the rally and the violence that ensued against counter-protestors [66]. Interestingly, none of these details appear in the top ten of breakpoints for the alternative group. We do not propose that these patterns in language use provide evidence for social identity performance on part of the progressive group, as we studied a user-generated and highly heterogenous list of channels, for which, in contrast to the alternative group, no claims have been made that they form a specific social movement. However, mentions of the rally on part of the progressive group may have lent further discursive opportunities and resulting visibility to the alternative group [30, 31].

Interestingly, a large number of the top ten bigrams in the alternative group for which a breakpoint was observed did not relate to the rally, but the promotion of YouTube channels, for instance urging viewers to subscribe to a channel, enable notifications, or donate to Patreon, a platform where content creators can crowdsource donations [67]. This behaviour lends further support to previous findings that the alternative YouTubers promote their channels such as mainstream influencers, and monetize their videos through donations to create a devoted fanbase [8, 14]. In short, the examination of bigram proportion breakpoints showed that the Charlottesville rally did seem to have an effect on language use in both groups separately.

The alt-right as a social movement

Both the language differences between and within the progressive and alternative video transcripts can be interpreted in light of social movement theory, and also add to our understanding of the effect of offline events on online behaviour. First, we observed several topics prevalent among alternative channels that could be seen as in line with the social identity of the alt-right. Swearing, distrust in mainstream

media, white nationalism, and an emphasis on free speech distinguished the alternative group from the progressive group. Second, we saw marked changes in language after the Charlottesville rally. The alternative YouTubers not only discussed the rally but seemingly also promoted their channels more. While the further examination of the contexts in which these calls are made will be needed, the fact that (positive) breakpoints (in intercept) appear at the time of the rally may be a sign of mobilising others, urging viewers to show their support for the alternative channels and related movements. Indeed, if these calls are a direct result of the rally, the event may be viewed as a critical juncture for the alt-right movement, where the rally served as a triggering event for increased social identity performance and mobilisation, aimed at strengthening the movement. Furthermore, the progressive group was also shown to lend resonance and visibility to the alt-right by discussing the rally, even if condemning language (e.g., 'white supremacist' over 'white nationalist') was used. These discursive opportunities may, in turn, have fuelled social identity performance on part of the alt-right [30, 31]. That is, by discussing and even condemning the alt-right rally, the progressive group lends further resonance and visibility to the movement [31]. All in all, results of this study may support the notion that the alt-right behaves as a social movement and that the (offline) Charlottesville rally had an effect on online social identity performance within the alt-right on YouTube, and possibly also outside of the movement as demonstrated by analyses of progressive YouTube video transcripts.

Limitations and future work

The current study is not without limitations. First, data selection and subsequent operations may have impacted the results of our analysis. For example, the sources that we have drawn on for the YouTube videos were unbalanced in nature, with the progressive sample consisting of more videos than the alternative sample. Furthermore, the two groups also differed in terms of view counts and video length, both factors which may have impacted on language use. In addition, while the alternative channels were drawn from a research report, the list of progressive channels was drawn from user-generated online lists. Future research may be aimed at curating an expert-verified or crowd-sourced dataset of channels with different political biases.[5] Other search strategies to identify alt-right channels that do not rely on keyword searches, for example using hyperlinks posted on alt-right forums [68], should also be considered in future work. Furthermore, when we selected videos for analysis only transcripts with more than 100 words and a pre-specified percentage of English words were retained. These decisions were guided by previous research [49, 50] and our aim to retain only high-quality transcripts suitable for topic modelling. A full dataset without these filters applied is available for other researchers to experiment with other constraints. In a similar vein, researchers may be interested

---

[5] Similar to https://mediabiasfactcheck.com/ and https://www.allsides.com/media-bias/media-bias-ratings but for YouTube channels.

in examining longer or shorter timeframes surrounding the Charlottesville rally or even other events, and further data from our transcript retrieval (all videos available until 1 October 2018) is available on request. Lastly, transcript quality may have varied based on whether they were generated through automatic speech recognition or manually reviewed and/or added to a video. YouTube notes that automatic captions may be inaccurate due to mispronunciations, accents, or dialect.[6] Nevertheless, relying on the provided captions was the most straightforward way to obtain transcripts, and future work may examine what the effect is of different automatic speech recognition technologies on linguistic analyses.

Topic modelling involves several decisions on part of the researcher. For instance, various approaches exist for selecting the number of topics for a model, with no consensus in the research community [56]. Furthermore, assigning labels to topics is based on the interpretation of the researcher, with decisions highly sensitive to human bias. Nevertheless, we provide alternative models (with different numbers of topics) and further terms associated with topics on our OSF project page, for the reader to examine the outcome of our analyses, giving way to alternative explanations. Furthermore, some topics were difficult to interpret (e.g., 'Filler words' and 'Future focus'), mostly because they were composed of parts-of-speech with little meaning, or because the words did not form a coherent topic, and merely consisted of words that were used in the same way.

A bag-of-words approach utilised in both the topic modelling and the word frequency approach also holds its limitations. Specifically, bag-of-words models disregard word order and context. Furthermore, when measuring the prevalence of bigrams, polarity words or adjectives (e.g., not, very, super) that preceded each bigram may not have been captured. This issue may be solved in future using trigrams, although relevant *n*-grams that occur even further away from the keyword will still not be captured and further context will still be disregarded. As has been raised in the discussion, the breakpoints we observed only show that there was a change in frequency (proportion) of a bigram, and say nothing about the context in which bigrams occurred. For example, mentions of 'white nationalist' may have appeared in a negative context in the progressive group, and a positive context in the alternative group, but further analyses will be needed to make such claims. A further noteworthy solution to this problem is the use of word embeddings, an approach used to learn vector representations for individual words that aim to capture semantic relationships between words based on the contexts in which they appear. This approach has already been used within the context of the Charlottesville rally, showing that US media associated African-Americans (e.g., the term 'black') less with negative character traits (e.g., 'silly', 'extreme') after the rally [69].

It can be argued that understanding changes in the language use of potentially violent groups on social media may be of particular interest to policymakers and security officials aiming to prevent or de-escalate violence. Future research may focus on extending the present approach to measuring changes in language over time on other social media platforms where alt-right supporters are active, such as 8Kun

---

[6] https://support.google.com/youtube/answer/6373554?hl=en.

and Gab. It may also be of interest to measure concepts other than topics and *n*-gram frequencies, such as hate speech and abusive language, in response to the Charlottesville rally and perhaps other events of interest. Although it is beyond the scope of the current paper, a follow-up study of the specific contexts in which certain topics and *n*-grams occur may be interesting. For example, is the sentiment regarding 'white people' or 'feminism' negative or positive in polarity?

Conclusion

Following the violent rally in Charlottesville, the alt-right received significant attention in the media and public discourse. As a result, we expected to see differences in social identity performance and visibility of the alt-right movement, which was measured through examining language use. Contrasting a unique dataset of YouTube video transcripts from alternative, right-leaning channels to progressive, left-leaning channels, the present investigation indeed observed differences in language within and between the alternative and progressive groups. Results potentially reflect changes in social identity performance and visibility after the rally, as well as differences between the two groups more generally.

Compliance with ethical standards

References

1. Hughes, T. (2018). A year after Charlottesville rally, white nationalists enter mainstream conversation. *Usatoday*. https://www.theatlantic.com/ideas/archive/2018/08/the-battle-that-erupted-in-charlottesville-is-far-from-over/567167/.
2. Yan, H., & Sayers. (2017). Virginia governor on white nationalists: They should leave America. *CNN*. https://www.cnn.com/2017/08/13/us/charlottesville-white-nationalist-rally-car-crash/index.html.

3. Anti-Defamation League. (2017). Murder and extremism in the United States in 2017. *Anti-Defamation League*. https://www.adl.org/resources/reports/murder-and-extremism-in-the-united-states-in-2017.

4. Anti-Defamation League. (2018). Murder and extremism in the United States in 2018. *Anti-Defamation League*. https://www.adl.org/murder-and-extremism-2018.

5. Hine, G., Onaolapo, J., Cristofaro, E. D., Kourtellis, N., Leontiadis, I., Samaras, R., Blackburn, J. (2017). Kek, Cucks, and God Emperor Trump: a measurement study of 4chan's politically incorrect forum and its effects on the web. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (p. 10).

6. Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018). What is gab: a Bastion of free speech or an alt-right echo chamber. In *Companion Proceedings of the The Web Conference 2018* (pp. 1007–1014). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3184558.3191531.

7. Ellis, E. G. (2018). The alt-right are savvy internet users. Stop letting them surprise you. *Wired*. https://www.wired.com/story/alt-right-youtube-savvy-data-and-society/.

8. Lewis, R. (2018). Broadcasting the reactionary right on YouTube. *Data & Society*. Retrieved from: https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf.

9. Hodge, E., & Hallgrimsdottir, H. (2019). Networks of hate: the alt-right, "troll culture", and the cultural geography of social movement spaces online. *Journal of Borderlands Studies*. https://doi.org/10.1080/08865655.2019.1571935.

10. Hawley, G. (2018). *The alt-right: what everyone needs to know*®. Oxford: Oxford University Press.

11. Nagle, A. (2017). *Kill all normies: the online culture wars from Tumblr and 4chan to the alt-right and Trump*. Winchester, Washington: Zero Books.

12. Hawley, G. (2017). *Making sense of the alt-right*. Columbia: Columbia University Press.

13. Tilly, C. (1993). Contentious Repertoires in Great Britain, 1758–1834. *Social Science History, 17*(2), 253–280. https://doi.org/10.1017/S0145553200016849.

14. Munger, K., & Phillips, J. (2019). A supply and demand framework for YouTube politics. *Pre-print*. Retrieved from: https://osf.io/73jys/download.

15. Serwer, A. (2018). The white nationalists are winning. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2018/08/the-battle-that-erupted-in-charlottesville-is-far-from-over/567167/.

16. Shear, M. D., & Haberman, M. (2018, January 20). Trump defends initial remarks on charlottesville; again blames 'Both Sides.' *The New York Times*. https://www.nytimes.com/2017/08/15/us/politics/trump-press-conference-charlottesville.html.

17. Atkinson, D. C. (2018). Charlottesville and the alt-right: a turning point? *Politics, Groups, and Identities, 6*(2), 309–315. https://doi.org/10.1080/21565503.2018.1454330.

18. Caiani, M., della Porta, D., & Wagemann, C. (2012). *The Action Repertoires of the Radical Right: Violence and Beyond. Mobilizing on the Extreme Right*. Oxford: Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199641260.001.0001/acprof-9780199641260-chapter-5.

19. Porta, D. D. (2018). Protests as critical junctures: some reflections towards a momentous approach to social movements. *Social Movement Studies*. https://doi.org/10.1080/14742837.2018.1555458.

20. Diani, M. (1992). The concept of social movement. *The Sociological Review, 40*(1), 1–25. https://doi.org/10.1111/j.1467-954X.1992.tb02943.x.

21. Ackland, R., & O'Neil, M. (2011). Online collective identity: the case of the environmental movement. *Social Networks, 33*(3), 177–190. https://doi.org/10.1016/j.socnet.2011.03.001.

22. Hunt, S. A., & Benford, R. D. (1994). Identity talk in the peace and justice movement. *Journal of Contemporary Ethnography, 22*(4), 488–517. https://doi.org/10.1177/089124194022004004.

23. Langman, L. (2005). From virtual public spheres to global justice: a critical theory of internetworked social movements*. *Sociological Theory, 23*(1), 42–74. https://doi.org/10.1111/j.0735-2751.2005.00242.x.

24. Snow, D. (2001). Collective identity and expressive forms. https://escholarship.org/uc/item/2zn1t7bj;jsessionid&.

25. Polletta, F., & Jasper, J. M. (2001). Collective identity and social movements. *Annual Review of Sociology, 27*(1), 283–305. https://doi.org/10.1146/annurev.soc.27.1.283.

26. Diani, M. (2003). *Networks and social movements: a research programme*., Social Movements and networks: relational approaches to collective action Oxford: Oxford University Press.

27. Adams, J., & Roscigno, V. J. (2005). White supremacists, oppositional culture and the World Wide Web. *Social Forces, 84*(2), 759–778. https://doi.org/10.1353/sof.2006.0001.

28. Klein, O., Spears, R., & Reicher, S. (2007). Social identity performance: extending the strategic side of SIDE. *Personality and Social Psychology Review, 11*(1), 28–45. https://doi.org/10.1177/1088868306294588.

29. Simon, B., Trötschel, R., & Dähne, D. (2008). Identity affirmation and social movement support. *European Journal of Social Psychology, 38*(6), 935–946. https://doi.org/10.1002/ejsp.473.

30. Koopmans, R., & Muis, J. (2009). The rise of right-wing populist Pim Fortuyn in the Netherlands: a discursive opportunity approach. *European Journal of Political Research, 48*(5), 642–664. https://doi.org/10.1111/j.1475-6765.2009.00846.x.

31. Koopmans, R., & Olzak, S. (2004). Discursive opportunities and the evolution of right-wing violence in Germany. *American Journal of Sociology, 110*(1), 198–230. https://doi.org/10.1086/386271.

32. Peltz, J. (2017). Protests, vigils around US decry white supremacist rally-NY Daily News. https://web.archive.org/web/20170814021328/http://www.nydailynews.com/newswires/new-york/protests-vigils-decry-white-supremacist-rally-article-1.3408420.

33. Postmes, T., & Brunsting, S. (2002). Collective action in the age of the internet: mass communication and online mobilization. *Social Science Computer Review, 20*(3), 290–301. https://doi.org/10.1177/089443930202000306.

34. Alberici, A. I., & Milesi, P. (2016). Online discussion, politicized identity, and collective action. *Group Processes & Intergroup Relations, 19*(1), 43–59. https://doi.org/10.1177/1368430215581430.

35. McGarty, C., Thomas, E. F., Lala, G., Smith, L. G. E., & Bliuc, A.-M. (2014). New technologies, new identities, and the growth of mass opposition in the Arab Spring. *Political Psychology, 35*(6), 725–740. https://doi.org/10.1111/pops.12060.

36. Bliuc, A.-M., Best, D., Iqbal, M., & Upton, K. (2017). Building addiction recovery capital through online participation in a recovery community. *Social Science and Medicine, 193,* 110–117. https://doi.org/10.1016/j.socscimed.2017.09.050.

37. Best, D., Bliuc, A.-M., Iqbal, M., Upton, K., & Hodgkins, S. (2018). Mapping social identity change in online networks of addiction recovery. *Addiction Research & Theory, 26*(3), 163–173. https://doi.org/10.1080/16066359.2017.1347258.

38. Conway-Silva, B. A., Filer, C. R., Kenski, K., & Tsetsi, E. (2018). Reassessing Twitter's agenda-building power: an analysis of intermedia agenda-setting effects during the 2016 presidential primary season. *Social Science Computer Review, 36*(4), 469–483. https://doi.org/10.1177/0894439317715430.

39. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter: how 140 characters reflect the political landscape. *Social Science Computer Review, 29*(4), 402–418. https://doi.org/10.1177/0894439310386557.

40. Williams, M. L., & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: a case study in computational criminology and big data. *British Journal of Criminology, 56*(2), 211–238. https://doi.org/10.1093/bjc/azv059.

41. King, R. D., & Sutton, G. M. (2013). High times for hate crimes: explaining the temporal clustering of hate-motivated offending. *Criminology, 51*(4), 871–894. https://doi.org/10.1111/1745-9125.12022.

42. Hanes, E., & Machin, S. (2014). Hate crime in the wake of terror attacks: evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice, 30*(3), 247–267. https://doi.org/10.1177/1043986214536665.

43. Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior, 78,* 90–97. https://doi.org/10.1016/j.chb.2017.09.022.

44. Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., et al. (2014). Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining, 4*(1), 206. https://doi.org/10.1007/s13278-014-0206-4.

45. Olteanu, A., Castillo, C., Boy, J., & Varshney, K. R. (2018). The Effect of Extremist Violence on Hateful Speech Online. In *Twelfth International AAAI Conference on Web and Social Media*. Presented at the Twelfth International AAAI Conference on Web and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17908.

46. Siegel, A. A., Nikitin, E., Barbera, P., Sterling, J., Pullen, B., Bonneau, R.,Tucker, J. A. (2018). Measuring the prevalence of online hate speech, with an application to the 2016 U.S. election, 22.
47. Klein, A. (2019). From Twitter to Charlottesville: analyzing the fighting words between the alt-right and antifa. *International Journal of Communication, 13,* 22.
48. Zannettou, S., Finkelstein, J., Bradlyn, B., & Blackburn, J. (2019). A quantitative approach to understanding online Antisemitism. http://arxiv.org/abs/1809.01644.
49. Kleinberg, B., Mozes, M., & van der Vegt, I. (2018). Identifying the sentiment styles of You-Tube's vloggers. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 3581-3590).
50. Soldner, F., Ho, J. C., Makhortykh, M., van der Vegt, I. W. J., Mozes, M., & Kleinberg, B. (2019). Uphill from here: Sentiment patterns in videos from left- and right-wing YouTube news channels. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science* (pp. 84–93). Minneapolis, Minnesota: Association for Computational Linguistics. https://www.aclweb.org/anthology/W19-2110.
51. Silge, J., & Robinson, D. (2019). *Text Mining with R*. https://www.tidytextmining.com/.
52. Feinerer, I., & Hornik, K. (2018). *tm: Text Mining Package*. https://CRAN.R-project.org/package=tm.
53. Rinker, T. W. (2019). Quantitative discourse analysis package. https://cran.r-project.org/web/packages/qdap/citation.html.
54. Bliuc, A.-M., Betts, J., Vergani, M., Iqbal, M., & Dunn, K. (2019). Collective identity changes in far-right online communities: the role of offline intergroup conflict. *New Media & Society, 21*(8), 1770–1786. https://doi.org/10.1177/1461444819831779.
55. Blei, D. M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3,* 993.
56. Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). stm: R package for structural topic models. *Journal of Statistical Software, 10,* 1.
57. Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *Annals of Applied Statistics*. https://projecteuclid.org/euclid.aoas/1183143727.
58. Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.*
59. Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., et al. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103.
60. Bischof, J. M., & Airoldi, E. M. (n.d.). Summarizing topical content with word frequency and exclusivity, 8.
61. Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28*(3), 591–605. https://doi.org/10.2307/1910133.
62. Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange : an *R* Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v007.i02.
63. Mayer, H. (2018). The alt-right manipulates free-speech rights. We should defend those rights anyway. The Washington Post. https://www.washingtonpost.com/news/made-by-history/wp/2018/08/21/the-alt-right-manipulates-free-speech-rights-we-should-defend-those-rights-anyway/.
64. Lewis, H. (2019). How Anti-feminism Is the Gateway to the Far Right. The Atlantic. https://www.theatlantic.com/international/archive/2019/08/anti-feminism-gateway-far-right/595642/.
65. White nationalist leads torch-bearing protesters against removal of confederate statue. (n.d.). https://www.nbcnews.com/news/us-news/white-nationalist-leads-torch-bearing-protesters-against-removal-confederate-statue-n759266.
66. Reuters. (2019, June 28). Charlottesville: white supremacist gets life sentence for fatal car attack. *The Guardian*. https://www.theguardian.com/us-news/2019/jun/28/charlottesville-james-fields-life-sentence-heather-heyer-car-attack.
67. Regner, T. (2020). Crowdfunding a monthly income: an analysis of the membership platform Patreon. *Journal of Cultural Economics*. https://doi.org/10.1007/s10824-020-09381-5.
68. Mariconti, E., Suarez-Tangil, G., Blackburn, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I.,Stringhini, G. (2018). *"You Know What to Do": proactive detection of youtube videos targeted by coordinated hate attacks*. arXiv:1805.08168. http://arxiv.org/abs/1805.08168.

69. Leschke, J., & Schwemmer, C. (2019). Media bias towards African-americans before and after the Charlottesville rally. *Presented at the Proceedings Of The Weizenbaum Conference 2019 Challenges Of Digital Inequality*. https://www.ssoar.info/ssoar/handle/document/62623.
70. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Academic Press.

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