Review

SM+S
social media + society

Social Media + Society
July-September 2021: 1–20

DOI: 10.1177/20563051211027202
journals.sagepub.com/home/sms
SAGE

The Role of Suspended Accounts in Political Discussion on Social Media: Analysis of the 2017 French, UK and German Elections

Silvia Majó-Vázquez[1], Mariluz Congosto[2], Tom Nicholls[3], and Rasmus Kleis Nielsen[1]

Abstract

Content moderation on social media is at the center of public and academic debate. In this study, we advance our understanding on which type of election-related content gets suspended by social media platforms. For this, we assess the behavior and content shared by suspended accounts during the most important elections in Europe in 2017 (in France, the United Kingdom, and Germany). We identify significant differences when we compare the behavior and content shared by Twitter suspended accounts with all other active accounts, including a focus on amplifying divisive issues like immigration and religion and systematic activities increasing the visibility of specific political figures (often but not always on the right). Our analysis suggests that suspended accounts were overwhelmingly human operated and no more likely than other accounts to share "fake news." This study sheds light on the moderation policies of social media platforms, which have increasingly raised contentious debates, and equally importantly on the integrity and dynamics of political discussion on social media during major political events.

Keywords

social media, content moderation, information disorders, platform governance, elections, democracy, Twitter, comparative research, bots, news consumption

Introduction

Democracies hinge in part on rich information environments (Dahl, 1973), and social media platforms are increasingly, and sometimes problematically, important parts of them. They provide users with access to diverse information and opportunities to discuss with others, and they are central to how people access and engage with news. That is why the growing problems associated with social media platforms—including the spread of disinformation, harassment, hate speech, and incitement to violence—are important for our democracies. Since the suspension of the accounts of the former president of the United States, Donald Trump, content moderation has become an issue not only for social media companies but also for governments around the globe, and there is increasing interest in the content moderation practices of social media, and the impact these practices have, for example, on different political voices' ability to freely express themselves and take part in public debate. Especially right-wing politicians have, without offering any systematic evidence, asserted that they are being "censored" by social media. In Germany, Alternative für Deutschland has, without proof, claimed that "Internet giants like Google, Facebook, Twitter, Amazon, are abusing their dominant position in the market to abolish freedom of expression" (Deutsche Welle, 2021), and some individual politicians, often from the right, have made similar unsubstantiated allegations in France and the United Kingdom too. Beyond investigations of individual cases including some of the most popular social media accounts (like Trump's), or analysis of far-right groups and foreign interference in US elections, there are, however, few

[1]University of Oxford, UK
[2]Universidad Juan Carlos III, Spain
[3]University of Liverpool, UK

Corresponding Author:
Silvia Majó-Vázquez, Reuters Institute for the Study of Journalism, University of Oxford, Oxford OX1 2JD, UK.
Email: silvia.majo-vazquez@politics.ox.ac.uk

large-scale analyses of how social media platforms moderate election-related content online (Partnership, 2021).

In this article, we analyze the activity of 8,058,085 million Twitter accounts active around three of the most important elections in Europe in 2017 (in France, the United Kingdom, and Germany), and present the first cross-national analysis comparing the behavior of and content shared by accounts suspended by Twitter and all other accounts active around these key political events. The main question we address is the following: What content and behavioral features differentiate suspended accounts from other active accounts that participated in social media debates during these election campaigns in Europe? Addressing this question, we aim to provide evidence to evaluate narratives of "censorship" targeting social media platforms, made by some politicians but also a few mainstream media amid the suspension of popular figures (Dorsey, 2020, 2021; Ingram, 2018; Twitter, 2021). Our goal is not a normative or legal investigation of the debate on freedom of expression on social media, but instead an empirical contribution to advance our understanding of the political consequences or current content moderation practices. An issue we believe is intrinsically important to understand, and that can in turn, inform normative and legal discussions. Our research offers a large-scale study on how moderation efforts play out at scale in Western democracies during major political events.

Moreover, this study contributes to our understanding of how misinformation problems unfold in different democracies, an issue dominated by economists, political scientists, and psychologists in public and policy debate, but to which we believe (and can hopefully show) communication research can make major contribution focusing on the nature of the content shared.

We first consider the distinctive traits of the accounts that were both reported to be suspended by Twitter and active in political conversations. We assess whether the patterns of their activity suggest that they operated, as some bots do, to alter political conversations. Then, we provide novel cross-country evidence that suspended accounts actively spread political messages, sometimes at the extreme of the ideological spectrum, and systematically worked to increase the prominence of the content produced by news media outlets covering divisive issues (such as immigration, terrorism, and religion), demonstrating how problems of "information disorder" (Wardle & Derakhshan, 2017) involve not only bots and social media platforms but also users strategically amplifying specific content from domestic politicians and news media.

In the rest of the article, we first review the literature on information operations, disinformation, and platform moderation. Then, we describe our data and methods, and finally, we present the main results of our analyses. We conclude by assessing what can be generalized beyond Twitter and discuss the implications of our results for both the ongoing debate on how social media platforms mediate the flow of information online and the consequences for public debate, and thus our democracies.

Literature Review

While social media use may broadly be associated with more diverse news diets (Fletcher & Nielsen, 2018; Yang et al., 2020) and higher levels of political participation, engagement, and expression (Boulianne, 2019), they are also associated with a range of problems, including attempts to artificially shape the public agenda (Vargo et al., 2018), spread disinformation (Grinberg et al., 2019), or inflate the popularity of specific candidates (Allcott & Gentzkow, 2017). Social media activities may also contribute to the polarization of political debates and stoke incivility in the public sphere (Stella et al., 2018; Theocharis et al., 2016). They have demonstrably enabled hate speech, harassment, and trolling, especially oriented toward women and minorities too (Matamoros-Fernández & Farkas, 2021; Sobieraj, 2020; Stecklow, 2018).

In parallel, social media platforms have been investing to both assess and tackle disinformation problems, and in several cases have expanded their content moderation efforts and security more broadly (Conger, 2020; Gadde, 2018; Kelly, 2020; Roth & Harvey, 2018; Weedon et al., 2017). However, because the platform companies generally disclose limited details on those efforts and data access for independent researchers has been scarce or in some cases virtually non-existent (aside from special deals for specific teams at a few institutions), we still have a very limited understanding of the scale, scope, and nature of the problems affecting mediated political debates (for exception, see Twitter, 2020a).

According to Facebook's policies, accounts are suspended when they violate its terms and conditions and also if they have been *mis-classified*. The company says that around 3% of Facebook accounts might be false. Yet, the platform acknowledges this figure has been potentially much higher during episodic spikes, especially in countries such as Indonesia, Turkey, and Vietnam (for more, see Facebook, 2018a). Twitter says they may suspend accounts when they are created from the same IP or linked to the same email; when they result from an automated sign-up process; or when they show an exceptionally high volume of tweeting with the same hashtag or the same username without a reply from that account (Roth, 2019; Roth & Harvey, 2018).[1] Previous research on spam and unsolicited market offers on Twitter has shown that most of the suspensions take place within 3 days of the fraudulent activity—up to 92% of the accounts analyzed (Thomas et al., 2011).

This information though, tells us little about the narratives pushed by suspended accounts. The scarce systematic analyses we have on the role of suspended accounts pertain mainly to the US political scenario (Marcellino et al., 2020; O'Sullivan, 2018). From those studies, we know that social media platforms

have acted—frequently, after the fact—against foreign interferences to inflamate partisan divides or discorage citizens from voting. Something is anecdotally known about accounts suspended elsewhere too. In France for instance, Facebook suspended 30,000 accounts during the Presidential elections for repeatingly posting the same content or an increase in messages sent (Shaik, 2017). Yet, the nature of that content and the ultimate goal of those behind the suspended accounts was not clarified. During the Brexit referendum, about 30,000 bot accounts were either deactivated or removed by Twitter after spreading messages in support of the Leave campaign (Bastos & Mercea, 2019). In Venezuela (Twitter, 2019); Egypt, UAE, and Saudi Arabia (Grossman et al., 2020); China (Twitter, 2020a); and Myanmar (Facebook, 2018b), suspended accounts were behind activities directed at pushing narratives either against or in favor of current regimes and country allies. Computational approaches to the study of suspended accounts can help us go beyond company press releases and transparency reports, and have shown, for example, that the pursue of economic gains drives much of suspended accounts' activity (Thomas et al., 2011). Perhaps more relevant, those approaches demonstrate that while bots may be just noise exerting little influence over the narratives being part of online debates, suspended users that are not bots may actually be exerting true social influence in this regard (Wei et al., 2016). We already know that there are unintended negative consequences of the operation of these types of accounts, like the public disengagement from democracy or more generally increasing distrust in public institutions.

As pointed above, a lot of research in this area has focused on the identification of automated accounts (i.e., bots) and more recently the detection of real-time misinformation campaigns (Alizadeh et al., 2020). Techniques have been improved to identify bots, while at the same time, the actors behind the bots' activities have made their strategies more sophisticated. But, there is an increasing reliance on human-operated accounts to bypass social media platforms' defensive systems and keep altering political discussion (Grimme et al., 2017). Some of these accounts, although more sophisticated, can also be identified and, if so, suspended for violating the terms and conditions of the platform where they operate. However, we have only limited information—cited above—on *why* or *when* social media accounts are suspended, and what, if any, patterns there are to the effect on particular political voices. We do know, though, that they are commonly suspended around or during major political events, and in some single-case studies, we have seen they pursued a specific political agenda. Yet, the extent that this is different from the active account's remains to be investigated. Taken together, the literature reviewed above, even if relevant for our purposes, cannot be generalized to understand the specific nature of that agenda elsewhere. Also, whether there are communalities across countries remains to be investigated because most of the cited studies are a single-case approach. This leads us to pose the following research questions:

RQ1. What *behavioral features* differentiated suspended accounts from active accounts during electoral campaigns in France, the United Kingdom, and Germany?

RQ2. What type of *content* was spread by suspended accounts during the elections in France, the United Kingdom, and Germany?

Data and Methods

For this study, we analyze Twitter data during the presidential elections in France and the general elections in the United Kingdom and Germany in 2017. Our data time windows vary, to cover each election campaign until the day after the polling days. For France, we collected data from 2 April to 8 May; for the United Kingdom, we collected tweets between 5 May and 9 June; and finally, for the elections in Germany, we started the data gathering process on 21 August and finished it on 25 September. On the basis of these election-related tweets, we proceed with the following analyses to address our research questions: First, we identify the accounts that were suspended by Twitter and assess their main behavioral features. Then, we follow by comparing those features to those of the active accounts and bring evidence to the extent their activity impacted the election conversations; we end by measuring the topics discussed and shared by suspended account and the significant differences between those spread by the active accounts. To do this, we apply a standard scoring technique (more below) that allows us to understand the type of content and debates that focused the attention of the suspended accounts.

Notably, we focus our study on Twitter for several reasons. First, although Facebook, YouTube, and other platforms are more widely used, Twitter is a significant platform in itself with 166 million active users in 2020 (Twitter, 2020b). It is used by 9% of the population in France, 14% in the United Kingdom, and 6% in Germany for accessing news content. These percentages, which have been broadly steady through recent years, might look negligible when compared with those of Facebook—43%, 24%, and 22%, respectively (N. Newman et al. 2018). Yet, among those who use Twitter the most, one finds journalists and elites who in turn, have a greater influence over the offline media agenda. There is evidence that information circulating around Twitter might be deemed more newsworthy by journalists, especially by the youngest ones, than information from standard news sources like news wire agencies (McGregor & Molyneux, 2020). This evidence pairs with the fact that, ultimately, those behind information disorders online are interested in their stories being picked up by professional journalists. Unfortunately, they sometimes also serve as amplifiers of false stories or non-existent social concerns, just by covering them (Wardle, 2018).

Second, we focus on Twitter here because the company—unlike, for example, Facebook—has so far openly enabled researchers to conduct large-scale data gathering, thus

Table 1. Comparison Between Active and Suspended Users Across the Three Countries.

Dataset actives	DE	FR	UK	DE	FR	UK
Tweets				3,975,288	40,473,860	25,956,620
Users				319,032	2,817,207	4,328,323
Tweets/user				12.46	14.37	6.00
n_tweets_url	2,284,345	18,522,578	10,648,984	51.95%	41.93%	38.20%
n_tweets_media	1,073,842	14,262,659	11,200,421	24.42%	32.29%	40.17%
Original	1,006,249	5,360,092	2,560,847	22.89%	12.13%	9.19%
Replies out	480,181	2,739,351	2,002,547	10.92%	6.20%	7.18%
RTs out	2,025,439	25,305,604	17,649,398	46.06%	57.28%	63.30%
Quotes out	463,419	7,068,813	3,743,828	10.54%	16.00%	13.43%
HT	2,375,369	17,685,620	16,732,582	54.02%	40.03%	60.02%
Users suspended	DE	FR	UK	DE	FR	UK
Tweets				266,995	1,877,603	1,120,641
Users				20,674	158,495	214,965
Tweets/user				12.91	11.85	5.21
n_tweets_url	187,273	950,053	511,312	70.14%	50.60%	45.63%
n_tweets_media	66,544	691,075	441,316	24.92%	36.81%	39.38%
Original	111,932	351,377	88,196	41.92%	18.71%	7.87%
Replies out	32,034	171,871	143,915	12.00%	9.15%	12.84%
RTs out	93,853	1,025,952	746,865	35.15%	54.64%	66.65%
Quotes out	29,176	328,403	141,666	10.93%	17.49%	12.64%
HT	178,306	866,105	732,242	66.78%	46.13%	65.34%

n_tweets_urls, n_tweets_media, RT and HT stand for tweets including URLs, pictures or videos, retweets and hashtags, respectively.

allowing detailed empirical analysis. Yet, because activities on a social media platform will reflect both the particular user base and the particular affordances and policies of the platform in question, one cannot always generalize from studies of one platform to others. Nonetheless, as we discuss toward the end of the article, we believe there are some generalizable takeaways from this study.

The Sample

Our data were obtained first by calling the Twitter Streaming API and using semi-supervised third-party software, called Kalium (Napalkova et al., 2018). As suggested elsewhere, electoral campaigns are a multithreaded event (Jungherr, 2016). This poses a challenge to identify all the search parameters (i.e., keywords) that will allow the most comprehensive data collection process. To minimize the risk of missing relevant tweets during the elections, we checked twice per day for trending hashtags and keywords using tools external to Twitter during the whole data collection period. The most important hashtags in terms of the overall conversation were daily manually fed into the crawler (for detailed information, see Majó-Vázquez, Zhao, Nielsen, 2017).

For this study, we collected 76,452,886 tweets. They correspond to 8,058,085 unique accounts that actively participated on Twitter conversations during the elections in France, the United Kingdom, and Germany.[2] As Table 1 shows, our biggest dataset corresponds to France, measured by the overall number of tweets posted (N=42,351,463).[3]

We ran a second crawling process on 22 May 2018 to obtain basic information about the unique users who participated in the three elections. Notably, although Twitter has been reactively suspending accounts since it was created—mainly after receiving a report from another user—it was not until the beginning of 2018 that the company decided to proactively and massively suspend accounts when their behavior conflicted with the platform's policies. Worldwide and only in May 2018, 9.9 million users were flagged as *malicious* accounts. There is no public information on how many of those flagged accounts were eventually suspended (Roth & Harvey, 2018). Yet that was a major step in the content moderation strategy of the platform. We ran our second crawling process at that point to identify how many of the accounts that participated in the previous elections in Europe had been suspended by then. Undoubtedly, the time windows for the data collection limit us in establishing any causal link between the participation of the accounts in the election debates and their later suspension. Regardless, we contribute to the current research by studying the behavior and content shared by suspended accounts and bringing evidence of their active attempts to undermine the integrity of public debate around the elections in Europe. Equally important, we bring novel evidence to inform policy making at the time when many Western democracies tackle contentious debates about platform moderation.

Table 2. Statistics: URLs Fetched.

N URLs/(*N* of links)	*N*	Fetched	Excluded: not journalistic	Unable to crawl
UK active	9,999	5,507	4,343	149
	(5,315,325)	(2,518,232)	(2,679,791)	(117,302)
UK suspended	9,999	4,742	5,089	168
	(297,893)	(100,526)	(190,474)	(6,893)
France active	9,999	4,580	4,980	439
	(6,815,145)	(2,683,156)	(3,912,486)	(219,503)
France suspended	9,999	3,665	5,813	521
	(391,659)	(134,408)	(237,017)	(20,234)
Germany active	9,999	5,765	4,015	219
	(896,448)	(503,266)	(377,352)	(15,830)
Germany suspended	9,999	4,300	5,568	131
	(104,303)	(29,114)	(74,496)	(693)
Total	59,994	28,561	29,808	1,685
	(13,820,773)	(5,968,702)	(7,471,616)	(380,455)

Note. Statistics for the "not found" accounts are included in Table A2 in the Appendix.

To classify all users as active or suspended, we called the Twitter REST API. We individually checked each user for which we did not obtain specific profile information and classified them as suspended or not found based on the information returned by the API. We stored information on whether users were still active; when they joined the social platform; and on how many days, out of the total election campaign period, they participated. We detected 349,134 suspended users, which represents 4.89% of the total population studied. Of those, 2,694 accounts, equivalent to 0.77% of the total number of suspended accounts, participated in the three elections pointing to the interest of these small populations to play a role in the three major political events in Europe.

In total, 199,389 users (2.47% of the total) were not found. There is not clear information from Twitter on when and how a user is completely removed from the platform. Hence, one cannot determine whether those users deleted their profile themselves or Twitter removed them from the platform. We treat this population aside from the suspended accounts and report the results of this separate analysis in the Appendix.

Content Analysis

Since our main research goal is to identify the type of *content the* suspended accounts amplified during the elections, we tackle this goal with a twofold strategy. First, we built an additional dataset containing all the URLs shared by the users. We detected almost all the URLs included in the tweets of actives and suspended accounts.[4] In total, this dataset contains 34,236,321 tweets with a URL. That is 44.78% of the tweets posted during the elections in France, the United Kingdom, and Germany included a URL linking to external content. Table 1 shows that suspended users, especially in Germany, more frequently added links to external content in their messages.

To identify which domains those URLs pointed to, we had to expand 1.8 million URLs.[5] As a previous step, we manually compiled a list of the most common link shorteners—alongside the commercial ones, several news media use their own shortener. Building on the previous research, we enhanced our list of URL shorteners, including the most common shorteners among bots (e.g., dlvr.it, dld.bz, viid.me, or ln.is). In the study by Chen et al. (2017), which compiled a comprehensive list of this type of shortener, more than 10% of tweets including them would be generated by bots.

For each group of users (i.e., active and suspended), we identified the 9,999 distinct URLs most frequently linked to. We manually filtered these lists, examining each distinct domain and retaining journalistic targets while filtering out the few sites that were spam or non-journalism (see Table 2). Notably, the ranking of top domains shared by suspended and active users was dominated by news sources (see ranking of top domains in Figure 3 below). We fetched each of the URLs using a web crawler during August 2018 (Nicholls, 2018). The main text was extracted, cleaned, and stemmed—running, runner, and run were all normalized to run, for example. We did this in a language-aware way to reflect that our dataset is multilingual. We then ran an automated content analysis to identify the terms that were significantly more used in the corpus of content shared by suspended users than by active users (more details in the next section).

Not all URLs were successfully fetched, though. There were several sites for which we could not fetch their content due to end point disruptions, mainly, because at the time of running the content data collection, those sites were not accessible, but also due to paywalls or restrictions on crawling. For instance, the content of Citizen Slant, allegedly a website based in Los Angeles, could not be fetched because it was not active at the time the data collection process took place. Neither were its Twitter account—created in May

2016 and with almost 22,000 followers—and Facebook account. While active, its Twitter account posted links to take audience to its main site, the almost sole purpose of which was spreading information about the Trump administration and criticizing some of the decisions of the US president and his aides.

Table 2 reports the statistics on successful fetches. We were able to fetch 47.6% of all links included in the tweets of active and suspended accounts. Worthy of note, we fetched 94.1% of all *news* URLs included in those tweets. As mentioned before, news sites dominated the ranking of most shared domains among active and suspended accounts (see Figure 3).

Then, we crawled all the content shared by accessing the URLs included in the tweets. We built a corpus of news articles for each of the three countries and extracted the top 250 key terms.[6] Our goal is to calculate the relative prominence of the words in the corpora of the suspended accounts and compare this against that of the active accounts. For this, we first identify important *terms* in each news article as a whole and then look at the relative frequencies of these words among the two groups.

To identify the key terms, we use a tf–idf model, a standard scoring technique for keywording (Manning, Raghavan, and Schütze, 2008, pp. 116–121), where the score for each term is its frequency downweighted by the proportion of documents in the corpus in which it occurs.[7] This particularly significantly reduces the scores of terms that appear in nearly all documents and have little semantic meaning. By aggregating these scores across the corpus, we select for terms that are key to multiple documents but are still distinctive.

Then, we rank the relative prominence of each term based on a different form of term frequency scoring. We built separate sub-corpora for each of the groups of users and, for each keyword, calculated modified term frequencies for each of the sub-corpora. The term frequency in this case is the frequency of the term in each document, multiplied by the number of times that document was shared, summed across the sub-corpus, and then normalized by the total such frequency of all terms in the group. This measures the prominence of each keyword in each group. At the same time, it takes account of the frequency of sharing of each link within the dataset. In other words, the most prominent articles, shared thousands of times, are weighted higher than those shared a dozen times.

More formally, the final scoring for each term is the ratio of these term frequency scores in the comparator groups

$$wt_{term,group} = \sum_{doc \in group} f_{term,doc} \cdot linkct_{doc}$$

$$prominence_{term,group_1,group_2} = \frac{wt_{term,group_1}}{N_{group_1}} \div \frac{wt_{term,group_2}}{N_{group_2}}$$

with :

$$doc \in group; term \in doc$$

The formula above is used to calculate the relative prominence of terms in the articles linked to by suspended accounts, compared with those linked to by active accounts.

However, Table 1 shows that 48.05% of tweets from Germany, 58.08% from France, and 61.81% from the United Kingdom did not include any URL. Consistently, we finally built a second corpus of text including all messages posted by active and suspended accounts and apply the same methodological approach to identify the significant differences among the topics on those messages, which did not include a URL.

Bot Identification

Finally, it is pertinent to ask whether the suspended accounts under analysis reproduced the behavior of bots. Notably, there is no complete consensus on the defining traits of bot activity. The literature varies regarding the activity features and metadata assessed to detect automated accounts (see for instance Bastos & Mercea, 2019; Varol et al., 2017). Some later studies, highlighting the challenge faced, call them *suspected or potential* bots (Wojcik et al., 2018). No matter how different those approaches are, they all converge on identifying the frequency of posting in a given time window as the most important indicator to identify bots; the total number of posts since joining the platform is argued as a revealing feature too; and the ratio between followers and followees is an effective measure to identify automated accounts or approximate their behavior.

Building on the previous criteria, we filtered out the subset of users that were not media outlets and participated on at least 70% of the days of the electoral campaign; also, those that published on average 20 posts per day; and finally, those whose ratio of followees over followers was greater than 1.5. That is, they followed one user and a half for each follower they had. Once we excluded all media outlets, we found 2,349 users that matched those criteria out of the over eight million users studied (Germany = 168; France = 1686; United Kingdom = 495). Hence, our sample of suspended accounts (N = 349,134), as identified by calling the Twitter API, is much larger than that of potential bots and therefore so is their potential impact on the overall conversation online.

Results

The biggest population of suspended accounts, as shown in Table 1, was found in the United Kingdom, but the most active one was in Germany. To respond to our first research question, Figure 1 summarizes the main differences between active and suspended accounts. Here, we calculate the average of the z scores of different indicators[8] for actives and suspended accounts and compare it against the overall sample. As the comparison shows, across the three countries, the suspended accounts are younger (in terms of how long ago they were created), have fewer followers, and the length of

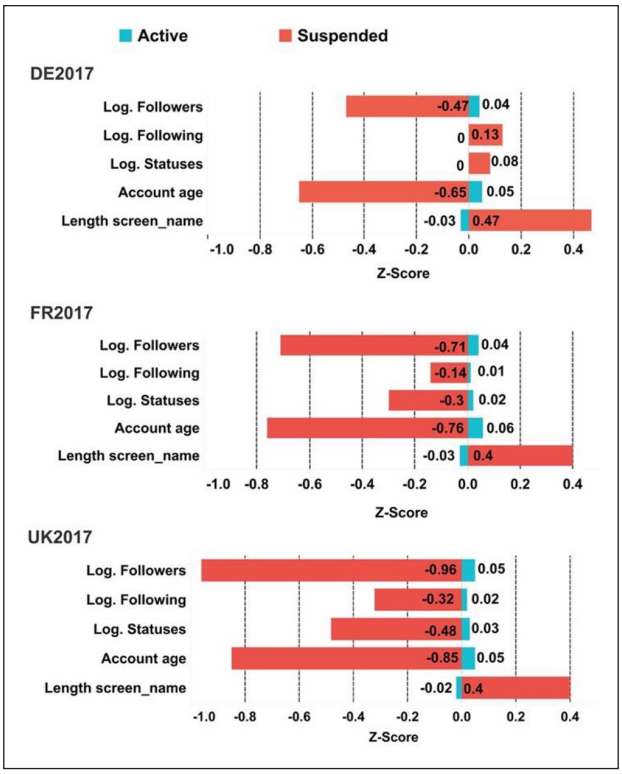

Figure 1. Z-score main traits of suspended and active accounts.

their screen names is larger than that of the overall sample. Thus, suspended accounts clearly exhibit exceptional values when measuring their age, number of followers, and the length of the screen name. There are some differences across countries though. In France and the United Kingdom, overall suspended accounts have posted fewer messages in their lifetime in comparison with the mean of the overall sample. This would be in consonance with their younger ages. In contrast though, in Germany, this pattern is not consistent. There, suspended accounts have been more active in their shorter life, in comparison with the mean of the overall sample. Hence, they exhibit a more exceptional active posting activity.

In addition, in Figure 2, we provide more details of when the suspended accounts were created. We show the distribution of the join date of the active users and that of the suspended accounts. Again, a common pattern arises across the three countries. The bulk of suspended users were created immediately before the elections, whereas the active users joined the platform on various dates as shown by the higher variance in the distribution. A similar strikingly different pattern arises when looking at the total activity of active versus suspended accounts. As shown in Figure 3, across the three countries, there is a high concentration of suspended users, between 30% and 40%, who tweeted between 550 and 1,700 tweets during the elections. In contrast, the posting activity of active accounts (i.e., non-suspended) has a much wider range of posting rate.

Our analyses reveal that the majority of the suspended accounts in the three countries have between10 and 100 followers, whereas the most common number of followers of the active accounts overall is between 100 and 1,000—see Figure A3 in the Appendix. We also find that across the three countries, the standard behavior of active users is that the more they publish, the higher is their number of original tweets. Yet, this relation does not arise when looking at the activity of suspended accounts—see Figure A4 in the Appendix. For suspended users, we find that they published fewer original tweets than the active users. Besides, in Germany and the United Kingdom, the number of original tweets is not even proportional to the frequency of posting. This results might suggest that some of those accounts functioned as amplifiers of, as we will see, divisive content posted by others.

Weather the activity of the suspended accounts had any impact in the electoral discussions on the platform can be measured by the number of retweets or quotes they received. Figure 4 shows that there were several hundreds of accounts that received more than a thousand Retweets (RTs) in France and the United Kingdom. Although in much fewer cases, some suspended accounts even received 10,000 RTs. We found a similar pattern when analyzing the quote activity around tweets posted by suspended accounts. These results point to the existence of a kind of highly connected accounts, identified in previous elections too (Marcellino et al., 2020), which spread messages effectively and quickly. The reach of their activity confirms the impact of the content posted by those users beyond the boundaries of their own feed.

To answer RQ2, we look at the content spread by these accounts. Starting by assessing the ranking of the most shared domains by these accounts, in Figure 3, we show that during the elections, suspended users in France, the United Kingdom, and Germany were mostly interested in spreading news content. Leaving aside the sites in France and Germany, that were straightforward commercial spam (e.g., *1001portails.com* or *benzinepreis-aktuell.de*), the vast majority of the top 10 most popular domains among suspended users are news media sites. Figure 5 also shows the difference in the percentage of URLs containing those domains in each of the groups (i.e., suspended versus active). Interestingly, in both France and the United Kingdom, the right-leaning digital-born outlet, Breitbart, was not only among the top 10 more shared URLs but also shared more by suspended than by active accounts.

The prominent role of YouTube in each of the three countries points to the interest of the suspended accounts in sharing video content. Unfortunately, the text-based focus of this study limits our ability to provide further information on the actual content of the videos. Nonetheless, these results reinforce the importance of integrating video and images in the study of information operations online (Wardle & Derakhshan, 2017).

These results at the domain level tell us little about the actual content suspended accounts were posting and sharing. Consistently, Figure 6 shows the highest and lowest results for each country for the relative prominence of terms in news content accessed through links. These are the words used disproportionately frequently in content shared by suspended

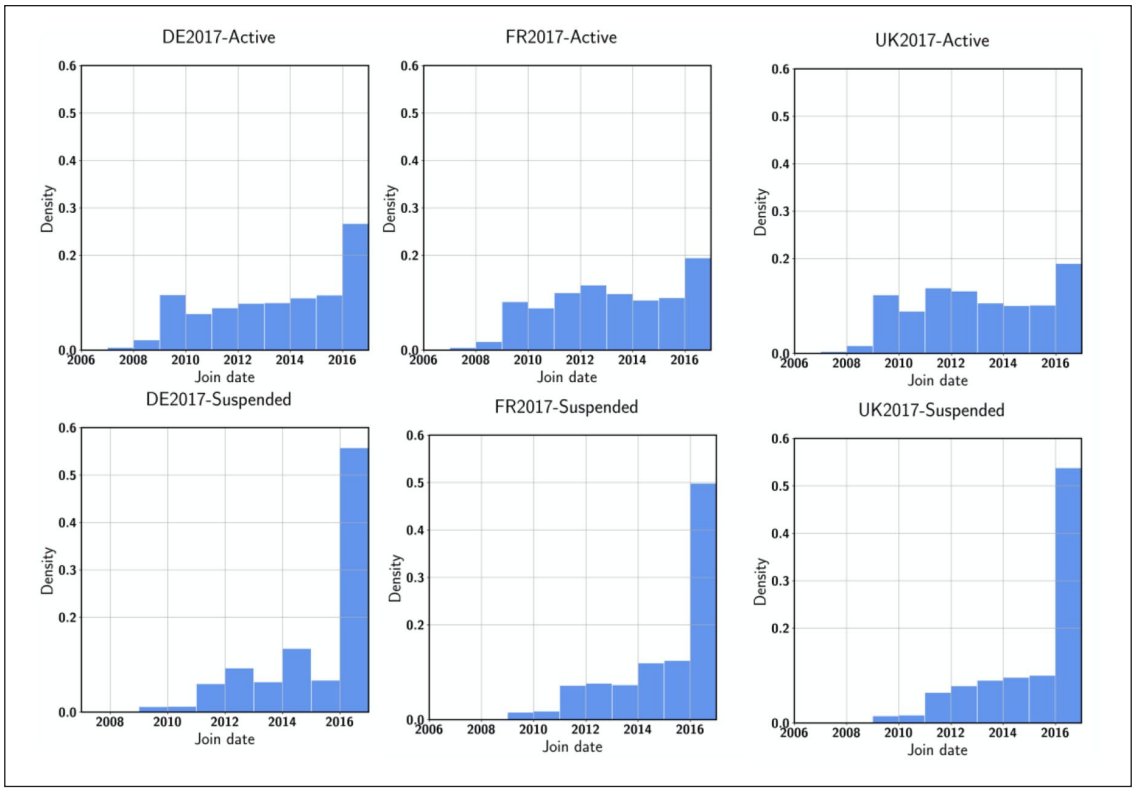

Figure 2. Distribution of the join date of active and suspended users across countries.
Note. Figure A1 in the Appendix shows the join date of the "not found" accounts, which mirrors the distribution of that of the suspended accounts.

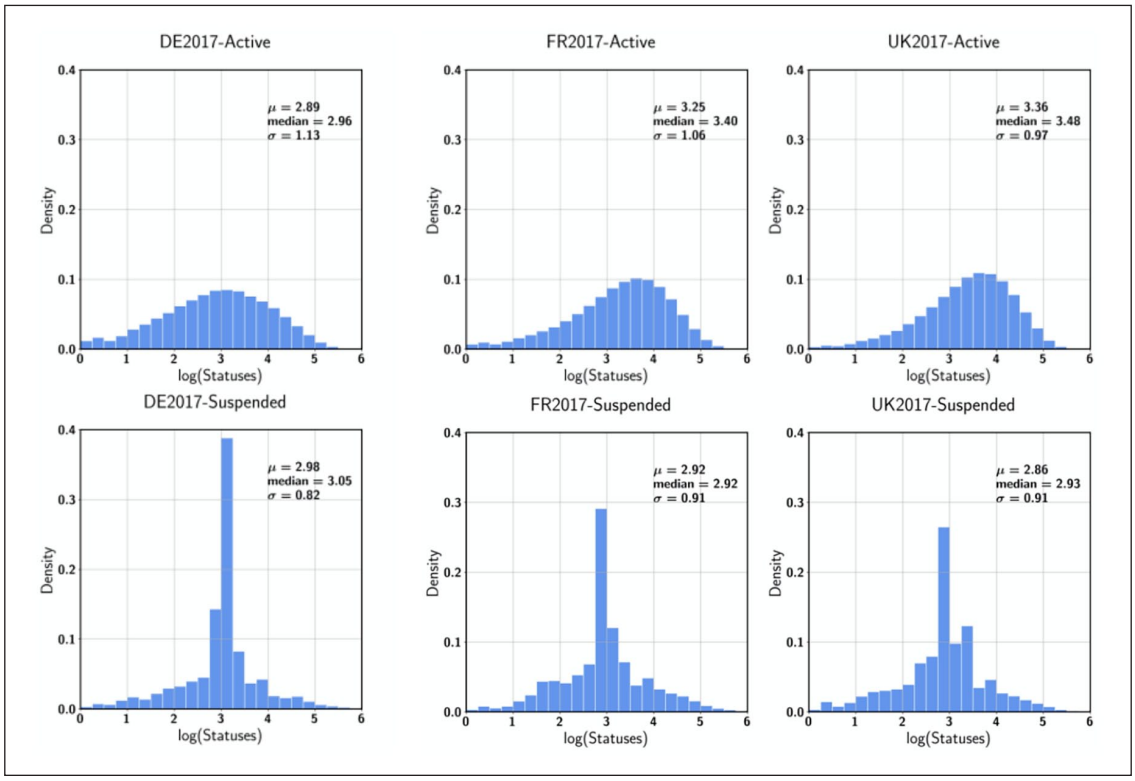

Figure 3. Distribution of the tweeting activity of active and suspended users across countries.
Note. Figure A2 in the Appendix shows the tweeting activity of the "not found" accounts.

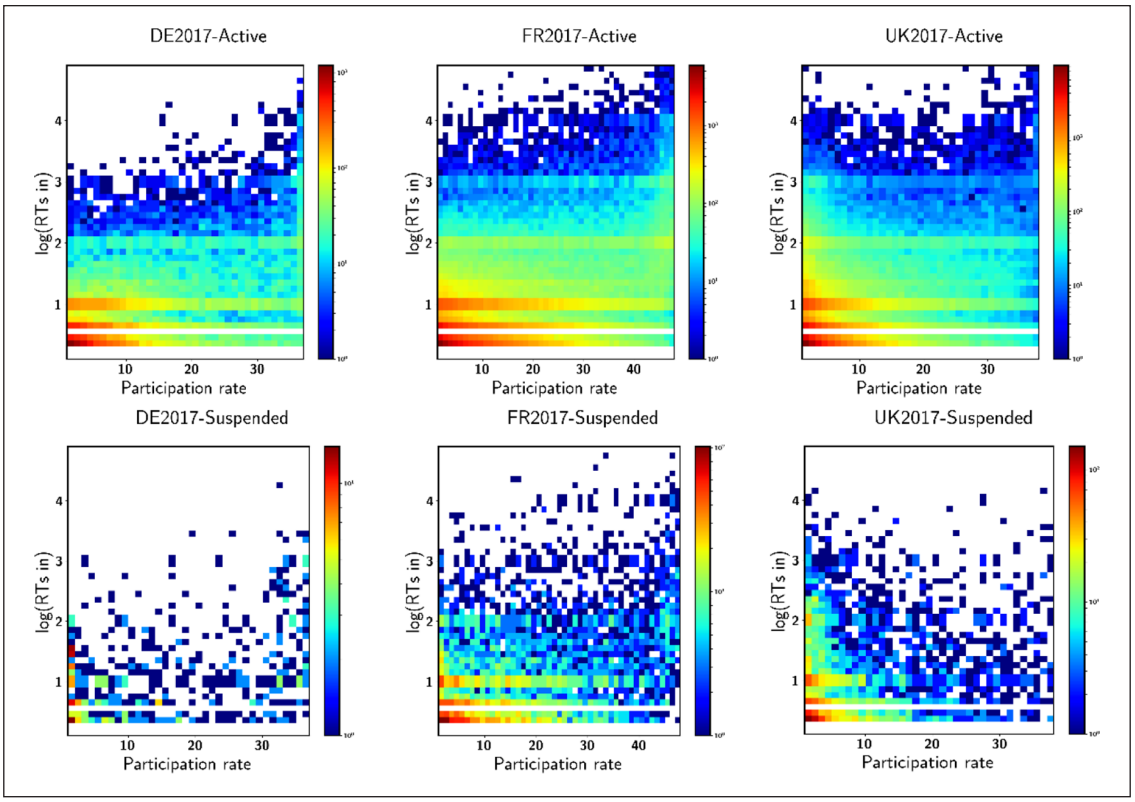

Figure 4. Distribution of RTs by participation of suspended versus active accounts.
Note. Figure A5 in the Appendix shows the relation between RTs and posting activity of the "not found" accounts.

accounts (high prominence ratio) and by active accounts (low prominence ratio).

In the UK and Germany datasets, the results clearly show that the most prominent terms included in content shared by suspended accounts are all associated with divisive issues or the online right (i.e., Muslims, terrorism, or Islam). In Germany, for example, the term "Turkei" is 2.3 times more frequent as a proportion of content linked from suspended accounts than from active accounts.

By contrast, those with the lowest ratios and therefore disproportionately heavily used by active accounts in the United Kingdom are predominantly scene-setting and descriptive, such as Britain goes (to the polls), Scottish, party leader, and (Prime) Minister Theresa (May). In Germany, a mixture of electoral terms is more common among the content shared by active accounts than by the suspended ones. This includes terms like wahlkreis, mitgleid or burg, and, interestingly although not surprisingly, (Frauke) Petry, the candidate for the right-wing AfD Party.

For the French dataset, a different cleavage emerges. The key distinction is not simply between right-wing and mainstream views of the world, but between an English-language dialogue and an institutional center. The most prominent terms among suspended accounts are English (e.g., first round, per cent, Sunday, poll). By contrast, the prominent terms among articles shared by active accounts are entreprise, projet, débat, François Fillon, and other French terms of political institutions. That the French results are showing an unexpected cleavage (language, rather than politics) is consistent with two pieces of evidence: (a) our previous results indicating the prominence of English sites (like, for example, Breitbart) among those linked by the suspended accounts in France; and (b) the results of the overlapping audiences analysis. This analysis, which identifies the accounts that posted tweets including any of the hashtags or keywords used in each of the three elections, shows that in total, 2,694 suspended accounts were active in the three elections. Even if this number only represents less than 1% of the total accounts suspended, this result is in line with previous research showing that part of election interference includes accounts that are repurposed in different political events at different times (Marcellino et al., 2020).

Using the same relative prominence method explained above, we also analyze the content of the tweets posted by suspended and active accounts. The reason for this is that a large portion of these tweets do not link to the external content—the largest being in the United Kingdom, as shown in Table 1. Hence, the question is what issues were discussed in those tweets? And were they significantly different from those discussed by active accounts? Our final corpus contains the text of all the tweets that did not link to content external to Twitter. We cleaned the text of these tweets by removing URLs pointing to embedded images.

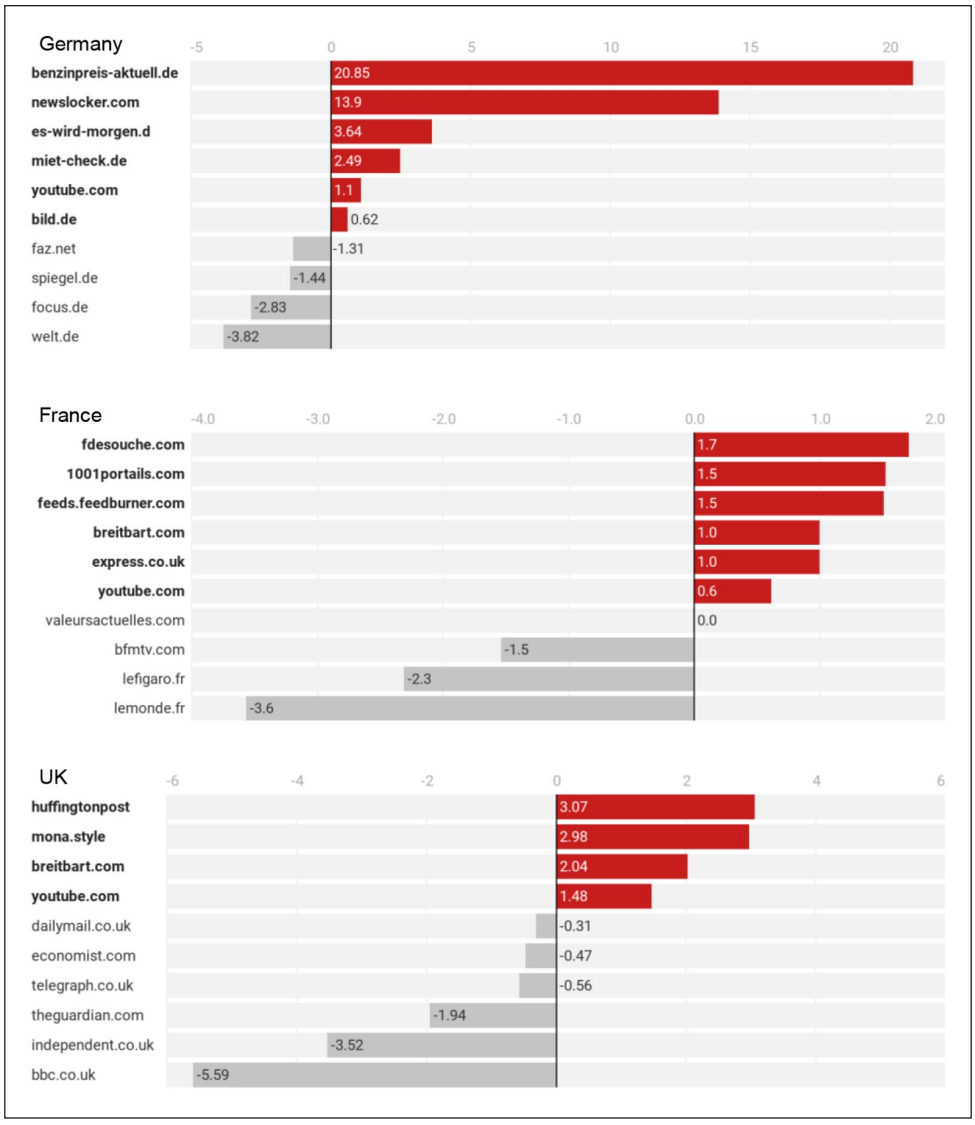

Figure 5. Top domains shared by suspended accounts and difference in percentage of sharing between suspended and active users.

Figure 7 shows that in Germany and France, suspended accounts were more frequently posting messages about the far-right parties and candidates—if not in favor, see for instance #jevotemarine—and at the same time against the main opposition candidates (i.e., #nichtmeinekanzlerin and #jamaismacron). Whereas in the United Kingdom, the terrorist attacks in Manchester and London were more closely discussed by suspended accounts than by the active ones. The latter were significantly more concerned with the support for the Labor Party campaign, as revealed by the hashtag #forthemany, and the Scottish National Party—see #votesnp.

Finally, to better understand the behavioral traits of the suspended accounts, and further address RQ1, we mapped the structure of their activity by building three retweet networks. We first tracked all the retweets by suspended accounts during the election campaigns in France, the United Kingdom, and Germany. Then, we built one network map for each country where *nodes* are the users—both suspended and active—and *ties* (i.e., the links between the nodes) represent retweets only by suspended accounts. Therefore, in these networks, two nodes are connected if the same suspended accounts retweet their content and are disconnected otherwise. These network structures allow us to identify where the suspended accounts' attention was focused the most across the campaign.

We analyze the networks by applying a community detection method to identify the different groups arising in each network. Community detection is a technique for the reduction of networks that classifies nodes into modules according to the density of connections: nodes in the same module have more connections to each other than to nodes in other modules (Girvan & Newman, 2002; M. E. J. Newman, 2012).

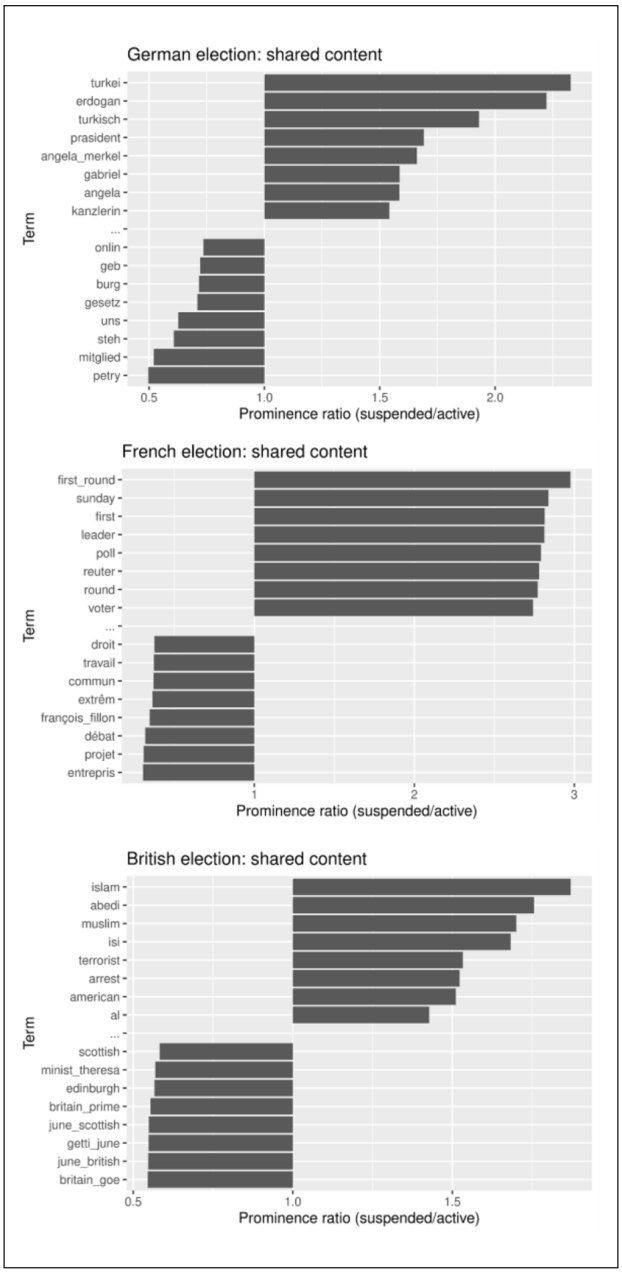

Figure 6. Terms significantly more used in the corpus of content shared by suspended accounts with respect to active users.[9]

Note. The scores are comparable within groups, but the overall magnitude of the scores varies with the relative size of each group.

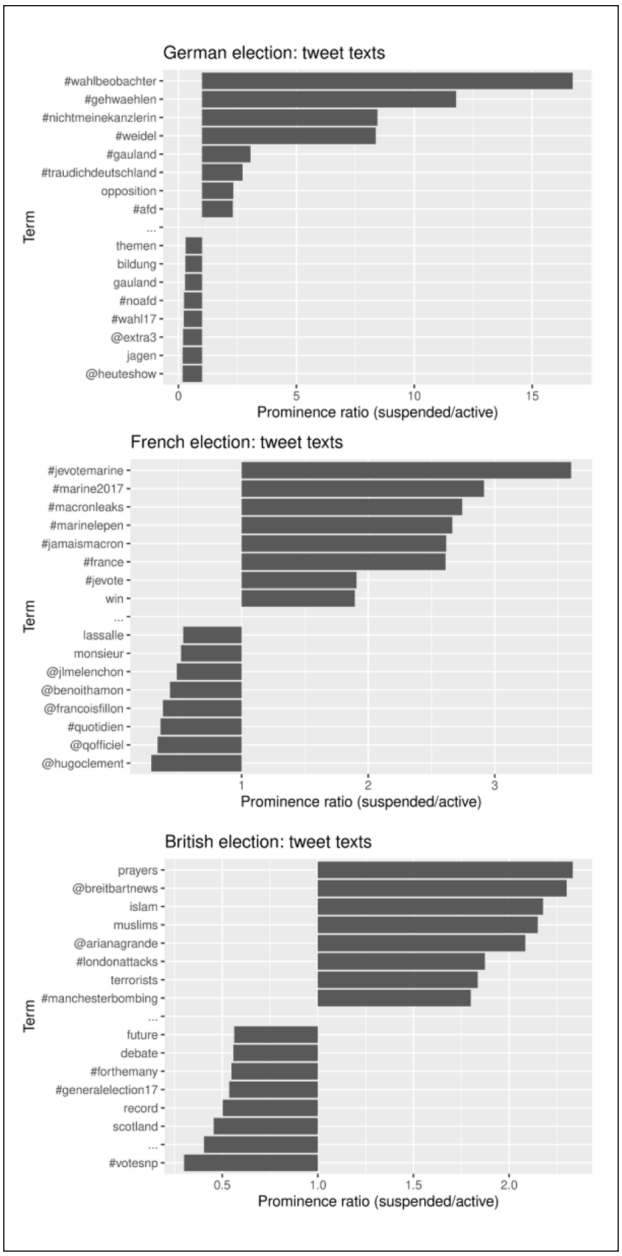

Figure 7. Terms significantly more used in tweets posted by suspended accounts in comparison with the active accounts.[10]

This approach helps characterize the organizational logic of a network by delineating areas where the network is denser and, therefore, more likely to channel information. In our case, it signals groups of Twitter users that are more likely to attract the attention of the suspended accounts.

There are different community detection methods. Here, we apply the Louvain method (Blondel et al., 2008), which is specifically designed for large network structures matching the nature of our data.[11] The visualization of the results is shown in Figure 8 and can be interpreted as the maps of suspended accounts' attention patterns.

Although there is variation across the three countries, these figures show that suspended accounts' retweets mainly aimed to promote the prominence of certain public and political figures or news media outlets. In France, Marie le Pen, the candidate of the far-right Front National, got most of the retweets by suspended accounts—a total of 37,709—followed by the profile of the public figure Paul Joseph Watson—a total of 16,970 retweets. Watson is a heavily

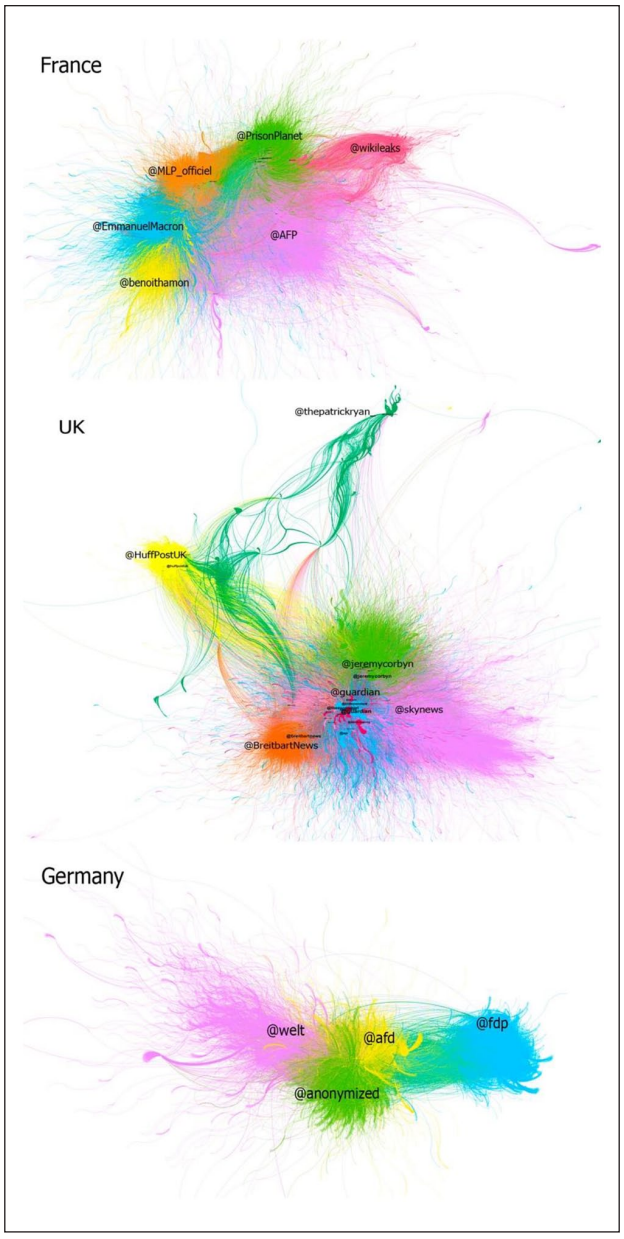

Figure 8. Communities in the networks of retweets by suspended accounts.
Note. For each network, only the giant connected component is graphed, which is the largest component that encompasses a significant fraction of the graph, and hence, it occupies most of the network. Hubs belonging to individual accounts have been anonymized.

followed YouTuber who built this popularity by collaborating with the US extreme right site InfoWars and joined Britain's UKIP in 2018. Finally, the third most retweeted user by the suspended accounts is a site called Voice of Europe, which shows a clear right-wing bias focusing on news related to immigration, anti-Europe messages, and the refugee crises.

In the case of the United Kingdom, the most retweeted users by suspended accounts were the following: The Guardian, the left-leaning newspaper—18,488 retweets; Jeremy Corbyn, the candidate of the Labour Party—16,329 retweets; and Breitbart, the US-based right-wing news site—14,265 retweets. And in Germany, the pattern found in France persists. Again, a right-wing political party, AfD, is the most retweeted user by suspended accounts—3,069 retweets, followed by an individual user, with more than 11,000 followers, whose activity is focused on spreading anti-EU and immigration and pro-Trump messages. The Free Democratic Party (FDP), the German liberal party, closely followed by the former Christian Democratic Union (CDU) politician Erika Steinbach, who openly supports the AfD, were the next most retweeted users by suspended accounts in this country.

Finally, the main *hub* of each community is named in each network in Figure 8. In our context, a hub is the account within a specific community that received most of the retweets of the suspended accounts. In other words, those accounts were promoted the most by suspended users. As we can see in Figure 8, they are yet again mostly politicians, political parties, or public figures specializing in divisive issues. On some occasions though, we also find that suspended accounts were boosting the visibility of legacy news media. Our analyses show that this is the case when these news outlets touched on issues like immigration, religion, or terrorism, or unveiled a political scandal that affected one of the candidates.

Conclusion and Discussion

On average, 5.6% of the users that actively participated in the elections in France, the United Kingdom, and Germany were suspended by Twitter. Our results show significant differences between the news content shared by and the patterns that shaped the behavior of the suspended accounts compared with that of the rest of the active users. We provide evidence that suspended accounts that participated in the elections in France, the United Kingdom, and Germany in 2017 aimed to increase the salience of *specific political figures* and *divisive content* related, among other issues, to religion or immigration (RQ2). We find that news content from legacy media as well as from right-wing digital-born outlets was shared significantly more by suspended accounts as opposed to content from fake news sources or any other type of text-based content. However, whether those accounts were suspended in relation to their behavior remains unknown because, unfortunately, the Twitter API does not return information on the exact point in time when the accounts were suspended.

In line with previous reviewed research, focused on a single country study, we have seen that the accounts suspended pursued a specific political agenda. The content spread by the suspended accounts in the United Kingdom, France and Germany feeds into a context of polarization and increasing divides across lines of the ideological spectrum. In these three European countries, traditional parties were losing ground in

favor of populist challenger parties both from the left and the right of the political spectrum (De Vries & Hobolt, 2020). The vast majority of the divisive topics identified in our study tap into the agendas of these outsiders and away from those of the mainstream parties. The question though is what effect those contents have on those exposed to them. Here, we cannot measure the exact number of people exposed to the content spread by suspended accounts or the magnitude of their effect on people's opinion over the course of the political debate. Yet, we can safely say that, considering their level of activity—they contribute more than three million tweets, 4.2% of the total, to the election debates—and the number of RTs and quotes they received, but also the differences between the content they spread and that spread by the other active accounts, they increased the visibility on Twitter of divisive issues in three countries where those same topics already played an important role in polarizing the electorate. The magnitude of the effect of that content, if any, is a matter for a future empirical question though.

Interestingly, only a small minority of the accounts suspended around the three elections are suspected or potential bots (0.67% of all the suspended accounts). This result points to the human labor behind the operation of the suspended accounts that participated in the elections in Europe. While important, "computational propaganda" (Woolley & Howard, 2018) reliant on automation is only a small part of wider problems of disinformation, many of which are driven by ordinary human user behavior on social media platforms, especially short-lived, human-operated, very active accounts with few followers. This creates a challenging scenario for democracies to the extent that this type of operation may be more difficult to detect, and the line between this and legitimate albeit sometimes controversial political expression is hard to draw, but is still potentially effective in swaying the public opinion (Boichak et al., 2021).

Another relevant finding is that video-based content, which we were limited to study due to the analytical tools available to us, represented an important fraction of the content spread by suspended accounts. This result reinforces the calls for future research in the specific field of automated image recognition.

In times when moderation policies enforced by social media platforms are under public and political scrutiny, our results suggest that accounts suspended by Twitter were more often active at the outer ends of the political spectrum, and hostile to established centrist political actors (or engaged in commercial spam). To those who are suspended, or who may see their supporters suspended, this may come across as bias at best and censorship at worst. But of course, it can also reflect the fact that some political groups are more likely to engage in behavior and speech that is deemed to have broken a particular social media platform's terms and conditions. The tension between a principled commitment to procedural fairness (same rules consistently applied to all, no matter what the consequences) and a more political calculation (seeking to develop rules and enforcement practices that impact different political actors equally, even if they act very differently) has been forcefully illustrated by reporting and suggesting that, for example, Facebook changed its content moderation in part to avoid charges of anti-conservative bias from President Trump and some of his supporters (Hao, 2021). Currently, these decisions are taken almost unilaterally by private, for-profit companies with little accountability, oversight, or transparency. Perhaps that will change in the years to come—in part in response to charges of censorship by social media, German courts are currently considering whether citizens have a positive right to free expression, also on private platforms, a horizontal application of a fundamental right that would limit companies' ability to engage in content moderation on the basis of their own terms and conditions, at least on political matters.

Our analysis is not a contribution to the legal and normative debate over what free expression looks like on social media (see instead Gillespie, 2018; Kaye, 2019; York & Zuckerman, 2019) but an empirical contribution to what the consequences are of current content moderation practices during important political events. Of course, it also remains as future empirical question whether the patterns we identify in this cross-national comparative study are reproduced in other countries with different political systems, especially developing countries with varying levels of internet penetration and where different levels of divisive issue arise, including terrorist attacks or geopolitical tensions (Neyazi, 2020). Similarly, it is left for future research to confirm that the content spread by and role of suspended accounts is recurrent across other structurally different and far more used platforms such as Facebook. Moderation policies are inconsistent across platforms and even some lack any kind of moderation of the most extreme content, for example, QAnon or Deep State (Partnership, 2021).

While research so far has mostly focused on the corrosive effects of misinformation and content spread by automated accounts on democratic quality, mainly through impacts on political engagement (Tucker et al., 2018), it remains an open question to assess the impact of removal of divisive topics. Only answering this question one can truly inform the current debates about the design features that can promote a more desirable political discussion for democratic ends, which should include lowering levels of political polarization and toxic conversations.

The results here point to the need to publish more detailed information on the moderation policies of social media platforms, not only when it comes to highly relevant public figures (Byers, 2021), but on standard users with fewer followers but who are equally important regarding their right to free speech and the need for a democratic public debate protected from information operations. To understand the wider implications of private for-profit companies moderating political speech at scale, we need an understanding both of individual high-profile cases of accounts sanctioned and

suspended (like the case of President Trump) and of what characterize the thousands and thousands of other accounts who are suspended. That is what we have provided here.

Acknowledgements

The authors would like to thank Felix Simon for his collaboration as a research assistant and for his thoughtful comments. They are also grateful for the helpful comments by Ralph Schroeder, Ben Toff, and Iván Lacasa as well as the research team of the Reuters Institute for the Study of Journalism and the researchers of the Oxford Internet Institute, where an early version of this article was presented in 2019. Finally, they are also thankful for the insightful comments of several anonymous reviewers. The authors thank the valuable contribution of Leonie Riviere as research assistant, and Andreas Kalterbrunner, Pablo Aragón, and Matteo Manca for helping us with the data collection process.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Work on this article has been partially supported by Google UK as part of the Google News Initiative.

ORCID iD

Silvia Majó-Vázquez https://orcid.org/0000-0002-2312-7907

Notes

1. In the case of Google and its products (i.e., Google News, YouTube), we know that they have defensive measures in place to detect and remove from the search engine: sites or accounts that impersonate any person or organization, including news organizations; those that conceal or misrepresent their ownership or primary purpose; or those that engage in coordinated activity to mislead users (Google, 2019; Walker, 2018).
2. We classify accounts by countries based on their tweeting activity. Hence, for instance, tweets posted during the French elections and using any of the relevant keywords or hashtags that emerged during that event are classified into the French dataset.
3. Differences between the sum of active and suspended users and the total number of users reported (8,058,085 users) correspond to "not found" accounts, which amount to 199,389 users across the three countries (2.47% of the total population studied). See the statistics in Table A1 in the Appendix.
4. We successfully detected 99.98% of the URLs in the French dataset; 99.99% of the URLs in the British dataset; and 100% of the URLs in the German dataset. In total, 4,901 URLs, out of 34,241,222, were unsuccessfully detected.
5. Those URLs that were shortened using trib.al had to be decompressed twice as they were shortened twice.
6. These are mostly unigrams (i.e., single words), but bigrams (two-word phrases) were also included as terms where they appeared at least 20 times in the corpus.
7. In the implementation we use here, the weight of a term, i, in document j, part of a corpus of size D, equals the frequency of term i in document j multiplied by the logarithm of the corpus size D divided by the number of documents in the corpus in which i appears, $weight_{i,j} = termfreq_{i,j} * log_2\left(\frac{D}{docfreq_i}\right)$ the keywords are those terms for which the sum of these weights over each document j is highest.
8. The z score is measured in terms of standard deviations from the mean. Hence, they are only computed for indicators whose values exhibit a normal distribution and not for those that follow a power-law distribution such as RT received and RT sent.
9. See results for "not found" accounts in Figure A6.
10. See results for "not found" accounts in Figure A7.
11. The network modularity scores for each country are as follows: Germany = 0.505; France = 0.664; United Kingdom = 0.715. These results highlight that the three networks are highly partitioned (Gonzalez-Bailon & Wang, 2016), which signals the divergent interests of the suspended accounts in each country.

References

Alizadeh, M., Shapiro, J. N., Buntain, C., & Tucker, J. A. (2020). Content-based features predict social media influence operations. *Science Advances*, *6*(30), eabb5824.

Allcott, H., & Gentzkow, M. (2017). *Social media and fake news in the 2016 election* (No. w23089). National Bureau of Economic Research.

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and user-generated hyperpartisan news. *Social Science Computer Review*, *37*(1), 38–54.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Boichak, O., Hemsley, J., Jackson, S., Tromble, R., & Tanupabrungsun, S. (2021). Not the bots you are looking for: Patterns and effects of orchestrated interventions in the US and German elections. *International Journal of Communication*, *15*, 26.

Boulianne, S. (2019). Revolution in the making? Social media effects across the globe. *Information, Communication & Society*, *22*(1), 39–54.

Byers, D. (2021, January 28). Facebook's "oversight boar," overturns 4 cases in first rulings. *NBC News*. https://www.nbcnews.com/tech/tech-news/facebook-s-first-oversight-board-rulings-overturn-four-five-cases-n1255960

Chen, Z., Tanash, R. S., Stoll, R., & Subramanian, D. (2017). Hunting malicious bots on Twitter: An unsupervised approach. In G. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), International Conference on Social Informatics (pp. 501–510). Springer.

Conger, K. (2020, June 11). Twitter removes Chinese disinformation campaign. *The New York Times*. https://www.nytimes.com/2020/06/11/technology/twitter-chinese-misinformation.html

Dahl, R. A. (1973). *Polyarchy: Participation and opposition*. Yale University Press.

Deutsche Welle. (2021, January 12). *Germany at odds over Twitter ban for far-right AfD party*. https://www.dw.com/en/germany-at-odds-over-twitter-ban-for-far-right-afd-party/a-56204646

De Vries, C., & Hobolt, S. (2020). The rise of challenger parties. *Political Insight*, *11*(3), 16–19.

Dorsey, J. (2020). *Fack check*. Twitter. https://twitter.com/jack/status/1265837138114830336

Dorsey, J. (2021). Twitter. https://twitter.com/jack/status/1349510769268850690?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1349510769268850690%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.theguardian.com%2Ftechnology%2F2021%2Fjan%2F13%2Ftrump-twitter-ban-jack-dorsey-c

Facebook. (2018a). *Facebook annual report 2017*. https://s21.q4cdn.com/399680738/files/doc_financials/annual_reports/FB_AR_2017_FINAL.pdf

Facebook. (2018b). *Removing Myanmar Military Officials from Facebook*. https://newsroom.fb.com/news/2018/08/removing-myanmar-officials/

Fletcher, R., & Nielsen, R. K. (2018). Automated serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism*, *6*(8), 976–989.

Gadde, V. (2018). *Confidence in follower counts*. Twitter. https://blog.twitter.com/official/en_us/topics/company/2018/Confidence-in-Follower-Counts.html

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Gonzalez-Bailon, S., & Wang, N. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks*, *44*, 95–104.

Google. (2019). *How Google fights disinformation*. https://kstatic.googleusercontent.com/files/388aa7d18189665e5f5579aef18e181c2d4283fb7b0d4691689dfd1bf92f7ac2ea6816e09c02eb98d5501b8e5705ead65af653cdf94071c47361821e362da55b

Grimme, C., Preuss, M., Adam, L., & Trautmann, H. (2017). Social bots: Human-like by means of human control? *Big Data*, *5*(4), 279–293.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, *363*(6425), 374–378.

Grossman, S., Khadija, H., DiResta, R., Kheradpir, T., & Miller, C. (2020). *Blame it on Iran, Qatar, and Turkey: An analysis of a Twitter and Facebook operation linked to Egypt, the UAE, and Saudi Arabia*. https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf

Hao, K. (2021, March 11). How Facebook got addicted to spreading misinformation. *MIT Technology Review*. https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/

Ingram, M. (2018, July 18). Republicans still convinced Facebook and Twitter are biased against them. *Columbia Journalism Review*. https://www.cjr.org/the_media_today/tech-biased-against-conservatives.php

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, *13*(1), 72–91.

Kaye, D. (2019). *Speech police: The global struggle to govern the Internet*. Columbia Global Reports.

Kelly, H. (2020, August 6). Facebook, Twitter penalize Trump for posts containing coronavirus misinformation. *The Washington Post*. https://www.washingtonpost.com/technology/2020/08/05/trump-post-removedfacebook/?hpid=hp_hp-top-table-high_fb-trump-656pm%3Ahomepage%2Fstoryans&itid=hp_hp-top-table-high_fb-trump-656pm%3Ahomepage%2Fstory-ans

Majó-Vázquez, S., Zhao, J., & Nielsen, R. K. (2017). *The digital-born and legacy media news media on Twitter during the French presidential elections*.

Marcellino, W., Johnson, C., Posard, M. N., & Helmus, T. C. (2020). *Foreign interference in the 2020 election*. www.rand.org/t/RRA704-2

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Xml retrieval. In *Introduction to Information Retrieval*. Cambridze University Press.

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, *22*(2), 205–224.

McGregor, S. C., & Molyneux, L. (2020). Twitter's influence on news judgment: An experiment among journalists. *Journalism*, *21*(5), 1–17.

Napalkova, L., Aragón, P., & Castro Robles, J. (2018, October 1–3). *Big data-driven platform for cross-media monitoring* [Conference session]. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy.

Newman, M. E. J. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, *8*(1), 25–31.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2018). *Reuters Institute Digital News Report 2018*. http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475

Neyazi, T. A. (2020). Digital propaganda, political bots and polarized politics in India. *Asian Journal of Communication*, *30*(1), 39–57.

Nicholls, T. (2018). RISJbot: A scrapy project to extract the text and metadata of articles from news websites (v1.1.0) [Computer software].

O'Sullivan, D. (2018, November 3). Twitter took down thousands of accounts that discouraged voting in midterms. *CNN Business*. https://edition.cnn.com/2018/11/02/tech/twitter-accounts-discourage-voting/index.html

Partnership, E. I. (2021). *The long fuse: Misinformation and the 2020 election*. https://www.atlanticcouncil.org/in-depth-research-reports/the-long-fuse-eip-report-read/

Roth, Y. (2019). *Information operations on Twitter: Principles, process, and disclosure*. Twitter. https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html

Roth, Y., & Harvey, D. (2018). *How Twitter is fighting spam and malicious automation*. Twitter. https://blog.twitter.com/official/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html

Shaik, S. (2017). *Improvements in protecting the integrity of activity on Facebook*. Facebook. https://www.facebook.com/notes/facebook-security/improvements-in-protecting-the-integrity-of-activity-on-facebook/10154323366590766/

Sobieraj, S. (2020). *Credible threat: Attacks against women online and the future of democracy*. Oxford University Press.

Stecklow, S. (2018, August 15). Why Facebook is losing the war on hate speech in Myanmar. *Reuters*. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/

Stella, M., Ferrara, E., & De Domenico, M. (2018). *Bots increase exposure to negative and inflammatory content in online social systems*. http://arxiv.org/abs/1802.07292

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & 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*(6), 1007–1031.

Thomas, K., Grier, C., Song, D., & Paxson, V. (2011). Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (pp. 243–258). Association for Computing Machinery. https://dl.acm.org/doi/abs/10.1145/2068816.2068840

Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. https://doi.org/10.2139/ssrn.3144139

Twitter. (2019). *Elections integrity* [Twitter's Focus is on a Healthy Public Conversation]. https://about.twitter.com/en_us/values/elections-integrity.html#data

Twitter. (2020a). *Disclosing networks of state-linked information operations we've removed*. https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html

Twitter. (2020b). *Q1 2020 letter to shareholders*. https://s22.q4cdn.com/826641620/files/doc_financials/2020/q1/Q1-2020-Shareholder-Letter.pdf

Twitter. (2021). *Permanent suspension of @realDonaldTrump*. https://blog.twitter.com/en_us/topics/company/2020/suspension.html

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, *20*(5), 2028–2049.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). *Online human-bot interactions: Detection, estimation, and characterization*. https://arxiv.org/pdf/1703.03107.pdf

Walker, K. (2018, August 23). An update on state-sponsored activity. *Google Official Blog*. https://www.blog.google/technology/safety-security/update-state-sponsored-activity/

Wardle, C. (2018, December 28). 5 lessons for reporting in an age of disinformation. *FirstDraft*. https://medium.com/1st-draft/5-lessons-for-reporting-in-an-age-of-disinformation-9d98f0441722

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Report DGI[2017]09). Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Weedon, J., Nuland, W., & Stamos, A. (2017, April 27). *Information operations and Facebook* (Version 1). https://i2.res.24o.it/pdf2010/Editrice/ILSOLE24ORE/ILSOLE24ORE/Online/_Oggetti_Embedded/Documenti/2017/04/28/facebook-and-infor

Wei, W., Joseph, K., Liu, H., & Carley, K. M. (2016). Exploring characteristics of suspended users and network stability on Twitter. *Social Network Analysis and Mining*, *6*(1), 1–18.

Wojcik, S., Messing, S., Smith, A., Rainie, L., & Hitlin, P. (2018). *Bots in the Twittersphere*. http://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere/

Woolley, S. C., & Howard, P. N. (2018). *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press. https://doi.org/10.1093/oso/9780190931407.001.0001

Yang, T., Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2020). Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(46), 28678–28683. https://doi.org/10.1073/pnas.2006089117

York, J. C., & Zuckerman, E. (2019). Moderating the public sphere. In R. F. Jørgensen (Ed.), *Human rights in the age of platforms* (pp. 137–162). MIT Press.

Author Biographies

Silvia Majó-Vázquez, PhD, is a research fellow at the Reuters Institute for the Study of Journalism, University of Oxford, UK. Previously, she has been a visiting researcher at the Annenberg School for Communication (UPenn). She studies how digital technologies are reconfiguring the news media ecology, borrowing tools from network science, and its effects on public opinion.

Mariluz Congosto, PhD, is a researcher at the Department of Telematics at Universidad Carlos III of Madrid in Spain. Her research interests include computational methods, political communication, digital data, and social media content.

Tom Nicholls, PhD, is a lecturer in Data Science, School of Communication and Media, University of Liverpool. Previously, he was a research fellow at the Reuters Institute for the Study of Journalism, University of Oxford, UK. His research interests include computational methods, political communication, news content at scale, and internet politics.

Rasmus Kleis Nielsen, PhD, is the Director of the Reuters Institute for the Study of Journalism, University of Oxford. He served as Editor-in-Chief of the International Journal of Press/Politics and won the Doris Graber Award for best book published in political communication. His work focuses on changes in the news media, political communication, and the role of digital technologies in both.

Appendix

Table A1. Statistics of "Not Found" Accounts Across the Three Countries.

Users not found	DE	FR	UK	DE	FR	UK
Tweets				154,682	1,824,476	802,537
Users				6,725	82,177	110,487
Tweets/user				23.00	22.20	7.26
n_tweets_url	91,619	748,290	296,691	59.23%	41.01%	36.97%
n_tweets_media	28,958	622,933	322,753	18.72%	34.14%	40.22%
Original	54,022	213,584	71,144	34.92%	11.71%	8.86%
Replies out	29,971	178,482	111,787	19.38%	9.78%	13.93%
RTs out	55,131	1,107,415	502,944	35.64%	60.70%	62.67%
Quotes out	15,558	324,995	116,662	10.06%	17.81%	14.54%
HT	98,060	815,509	488,893	63.39%	44.70%	60.92%

Table A2. Statistics of URLs Fetched Including "Not Found" Accounts.

N URLs/(*N* of links)	*N*	Fetched	Excluded: not journalistic	Unable to crawl
UK active	9,999	5,507	4,343	149
	(5,315,325)	(2,518,232)	(2,679,791)	(117,302)
UK suspended	9,999	4,742	5,089	168
	(297,893)	(100,526)	(190,474)	(6,893)
UK not found	9,999	5,329	4,520	150
	(163,393)	(75,850)	(83,998)	(3,545)
France active	9,999	4,580	4,980	439
	(6,815,145)	(2,682,897)	(3,912,486)	(219,762)
France suspended	9,999	3,665	5,813	521
	(391,659)	(134,376)	(237,017)	(20,266)
France not found	9.999	4,229	5,345	425
	(316,100)	(115,679)	(190,403)	(10,018)
Germany active	9,999	5,765	4,015	219
	(896,448)	(503,266)	(377,352)	(15,830)
Germany suspended	9,999	4,300	5,568	131
	(104,303)	(29,114)	(74,496)	(693)
Germany not found	9,999	5,992	3,790	217
	(35,770)	(21,381)	(13,744)	(645)
Total	89,991	44,109	43,463	2,419
	(14,336,036)	(6,181,321)	(7,759,761)	(394,954)

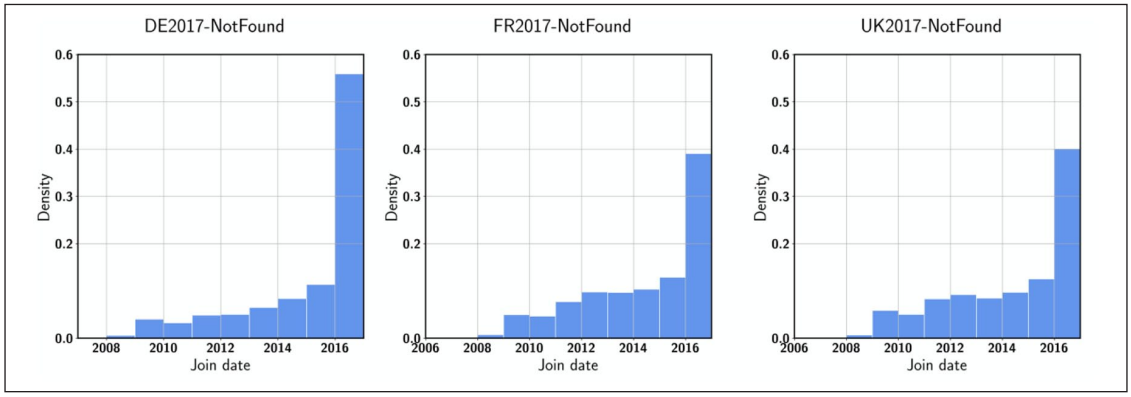

Figure A1. Distribution of the "join date" of "not found" accounts across countries.

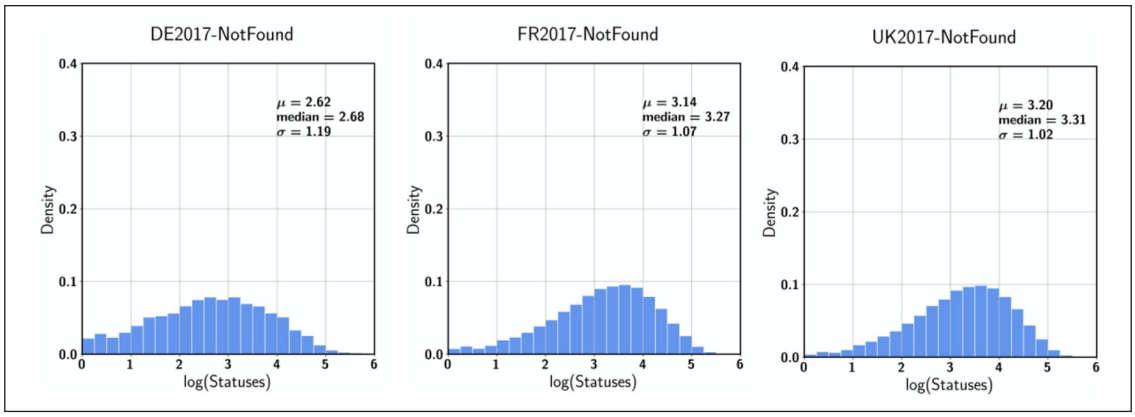

Figure A2. Distribution of the tweeting activity of "not found" accounts across countries.

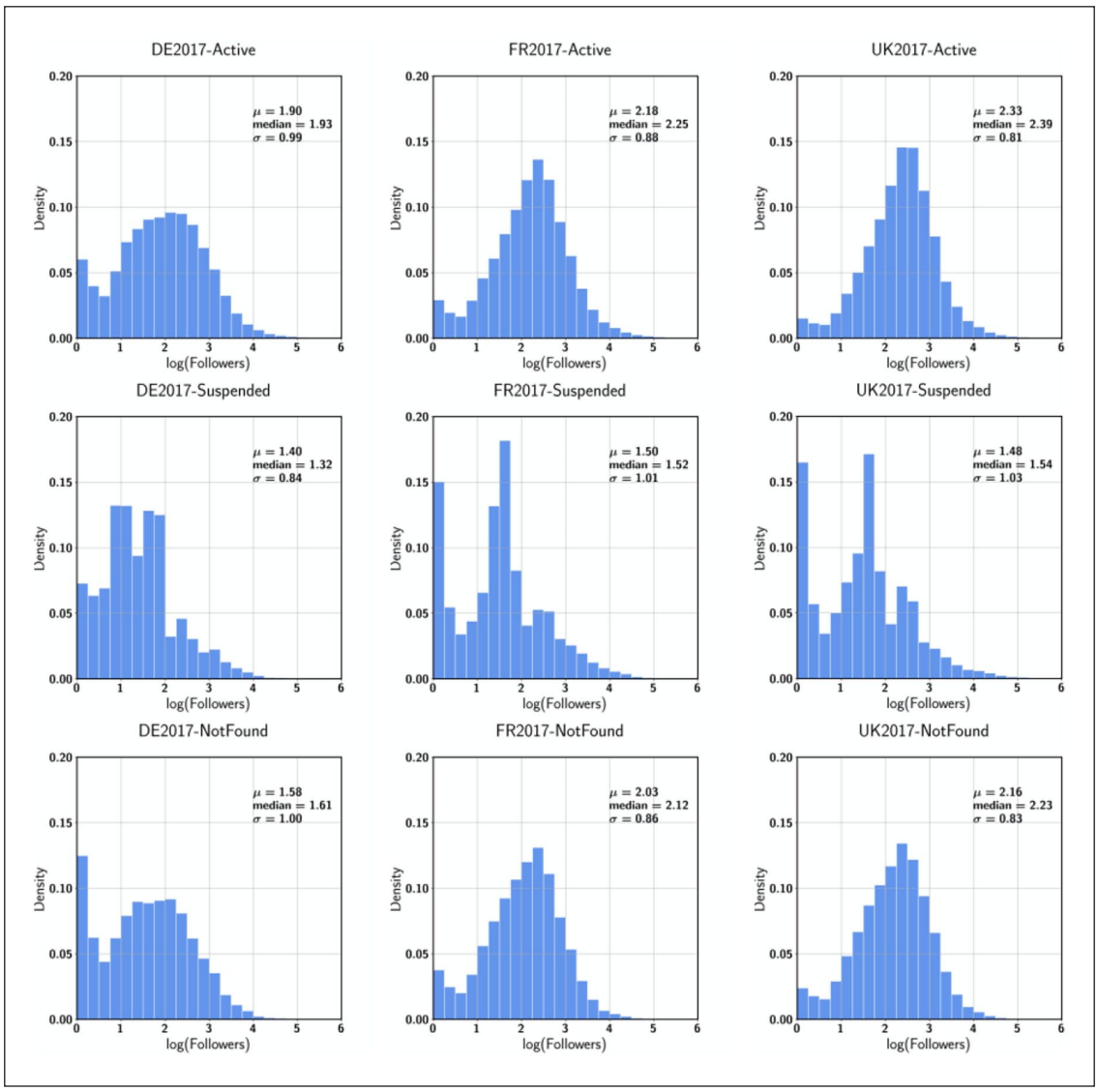

Figure A3. Distribution of followers of active and suspended users across countries.

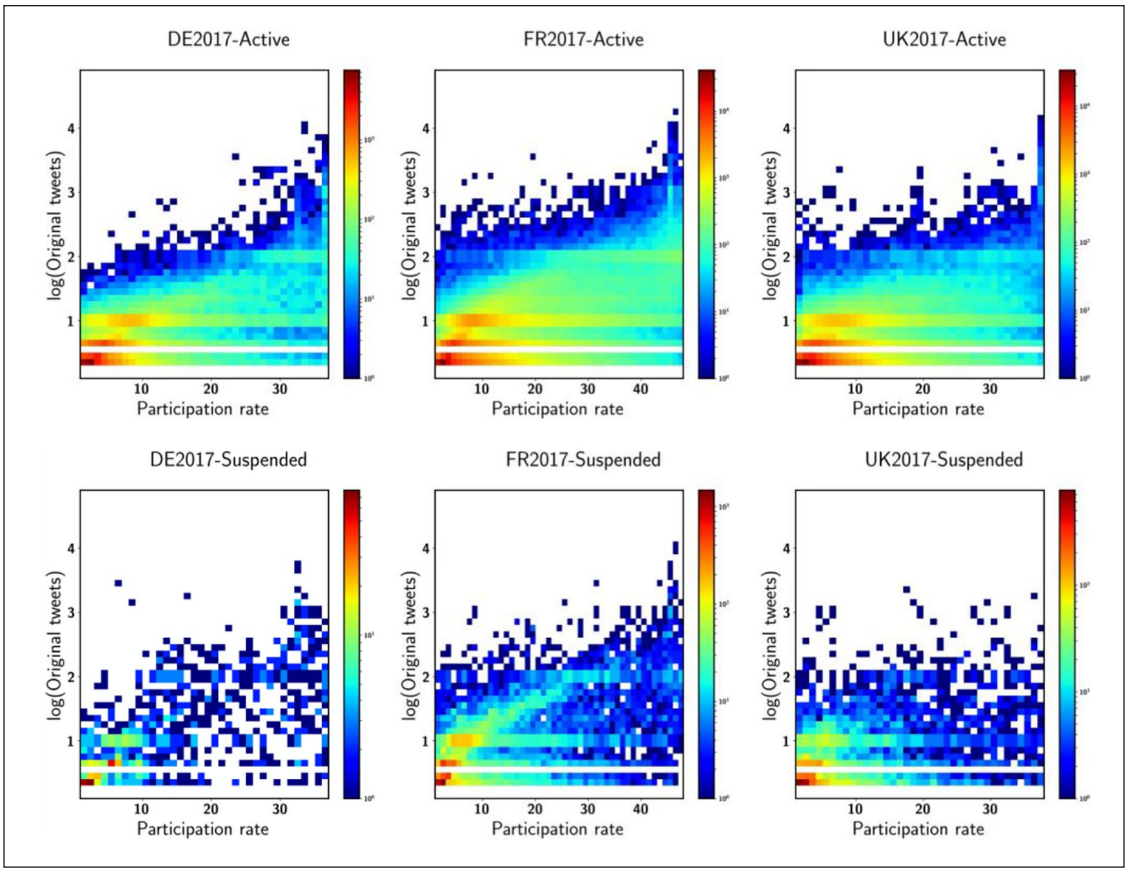

Figure A4. Distribution of original tweets versus frequency of active and suspended users across countries.

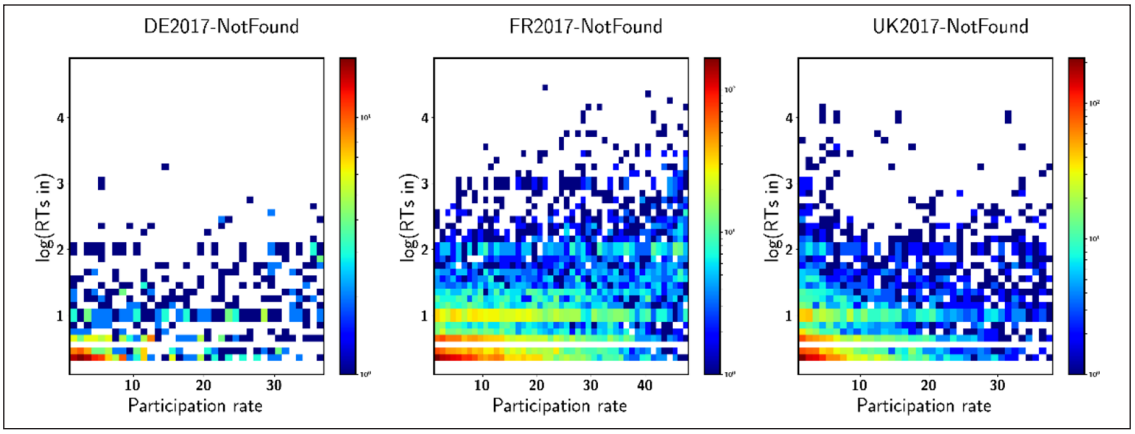

Figure A5. Distribution of RTs by participation of "not found" accounts.

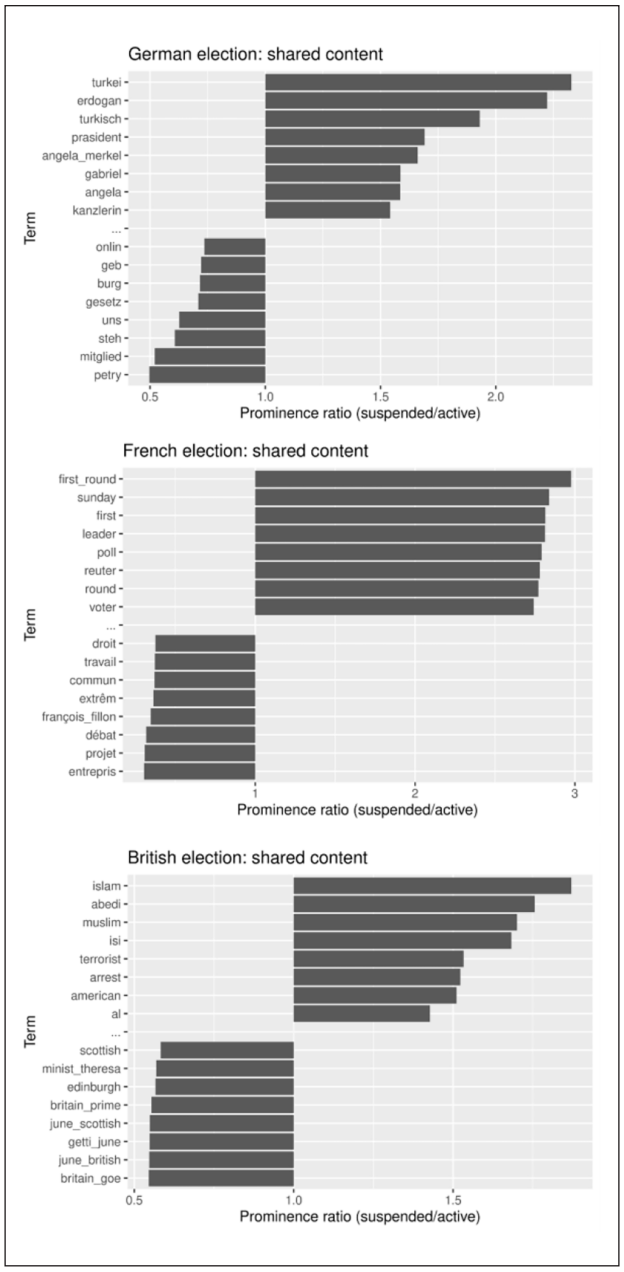

Figure A6. Terms significantly more used in the corpus of content shared by "not found" accounts compared with that of the active accounts.

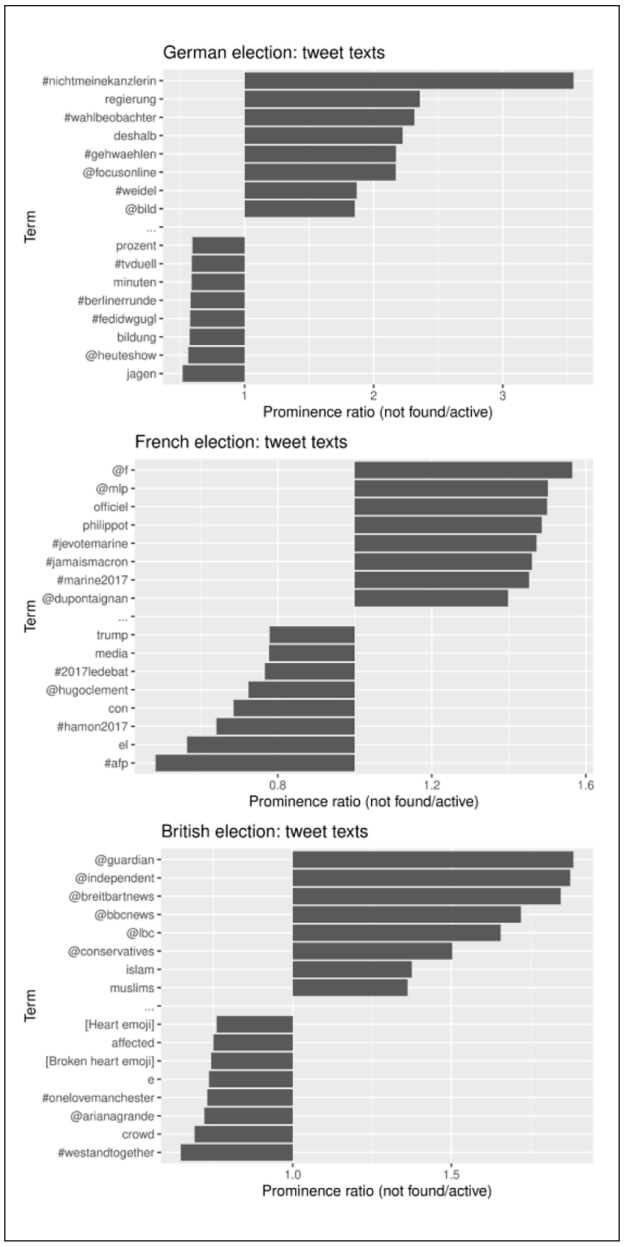

Figure A7. Terms significantly more used in tweets posted by "not found" accounts compared with the active accounts.