Bots are less central than verified accounts during contentious political events

Sandra González-Bailón[a,1] and Manlio De Domenico[b,1]

[a]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104; and [b]Center for Information and Communication Technology, Fondazione Bruno Kessler, 38123 Trento, Italy

Edited by Nicole Ellison, University of Michigan, Ann Arbor, MI, and accepted by Editorial Board Member Margaret Levi January 30, 2021 (received for review June 27, 2020)

Information manipulation is widespread in today's media environment. Online networks have disrupted the gatekeeping role of traditional media by allowing various actors to influence the public agenda; they have also allowed automated accounts (or bots) to blend with human activity in the flow of information. Here, we assess the impact that bots had on the dissemination of content during two contentious political events that evolved in real time on social media. We focus on events of heightened political tension because they are particularly susceptible to information campaigns designed to mislead or exacerbate conflict. We compare the visibility of bots with human accounts, verified accounts, and mainstream news outlets. Our analyses combine millions of posts from a popular microblogging platform with web-tracking data collected from two different countries and timeframes. We employ tools from network science, natural language processing, and machine learning to analyze the diffusion structure, the content of the messages diffused, and the actors behind those messages as the political events unfolded. We show that verified accounts are significantly more visible than unverified bots in the coverage of the events but also that bots attract more attention than human accounts. Our findings highlight that social media and the web are very different news ecosystems in terms of prevalent news sources and that both humans and bots contribute to generate discrepancy in news visibility with their activity.

social media | computational social science | online networks | information diffusion | political mobilization

Online networks have become an important channel for the distribution of news. Platforms like Twitter have created a public domain in which longstanding gatekeeping roles lose prominence and nontraditional media actors can also shape the agenda, increasing the public representation of voices that would otherwise be ignored (1, 2). The role of online networks is particularly crucial to launch mobilizations, gain traction in collective action efforts, and increase the visibility of political issues (3–8). However, online networks have also created an information ecosystem in which automated accounts (human and software-assisted accounts, such as bots) can hijack communication streams for opportunistic reasons [e.g., to trigger collective attention (9, 10), gain status (11, 12), and monetize public attention (13)] or with malicious intent [e.g., to diffuse disinformation (14, 15) and seed discord (16)]. During contentious political events, such as demonstrations, strikes, or acts of civil disobedience, online networks carry benefits and risks: They can be tools for organization and awareness or tools for disinformation and conflict. However, it is unclear how human, bot, and media accounts interact in the coverage of those events, or whether social media activity increases the visibility of certain sources of information that are not so prominent elsewhere online. Here, we cast light on these information dynamics, and we measure the relevance of unverified bots in the coverage of protest activity, especially as they compare to public interest accounts.

Prior research has documented that a high fraction of active Twitter accounts are bots (17); that bots are responsible for much disinformation during election periods (18, 19); and that bots exacerbate political conflict by targeting social media users with inflammatory content (20). Past research has also looked at the role bots play in the diffusion of false information, showing that they amplify low-credibility content in the early stages of diffusion (21) but also that they do not discriminate between true and false information (i.e., bots accelerate the spread of both) and that, instead, human accounts are more likely to spread false news (22). Following this past work, this paper aims to determine whether bots distort the visibility of legitimate news accounts as defined by their audience reach off-platform. Unlike prior work, we connect Twitter activity with audience data collected from the web to determine whether the social media platform changes the salience that legitimate news sources have elsewhere online and, if so, determine whether bots are responsible for that distortion. We combine web-tracking data with social media data to characterize the visibility of legitimate news sources and analyze how bots create differences in the news environment. This is a particularly relevant question in the context of noninstitutional forms of political participation (like the protests we analyze here) because of their unpredictable and volatile nature.

The use of the label "bot" often blurs the diversity that the category still contains. This label serves as a shorthand for accounts that can be fully or partially automated, but accounts that exhibit bot-like behavior can have very different goals and levels of human involvement in their operation. Traditional news

Significance

Online networks carry benefits and risks with high-stakes consequences during contentious political events: They can be tools for organization and awareness, or tools for disinformation and conflict. We combine social media and web-tracking data to measure differences on the visibility of news sources during two events that involved massive political mobilizations in two different countries and time periods. We contextualize the role of social media as an entry point to news, and we cast doubts on the impact that bot activity had on the coverage of those mobilizations. We show that verified, blue-badge accounts were significantly more visible and central. Our findings provide evidence to evaluate the role of social media in facilitating information campaigns and eroding traditional gatekeeping roles.

Author contributions: S.G.B. and M.D.D. designed research, performed research, contributed analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. N.E. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: sgonzalezbailon@asc.upenn.edu or mdedomenico@fbk.eu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2013443118/-/DCSupplemental.

Published March 8, 2021.

organizations, for instance, manage many of the accounts usually classified as bots—but these accounts are actually designed to push legitimate news content. Other accounts with bot-like behavior belong to journalists and public figures—many of which are actually verified by the platform to let users know that their accounts are authentic and of public interest. Past research does not shed much light on how different types of bots enable exposure to news from legitimate sources, or how the attention they attract compares to the reach of mainstream news—which also generate a large share of social media activity (23–25). More generally, previous work does not address the question of how news visibility on social media relates to other online sources (most prominently, the web), especially in a comparative context where different political settings other than the United States are considered. Are bots effective in shifting the focus of attention as it emerges elsewhere online?

This paper addresses these questions by analyzing Twitter and web-tracking data in the context of two contentious political events. The first is the Gilets Jaunes (GJ) or Yellow Vests movement, which erupted in France at the end of 2018 to demand economic justice. The second is the Catalan referendum for independence from Spain, which took place on October 1 (1-O) of 2017 as an act of civil disobedience. These two events were widely covered by mainstream media (nationally and internationally), but they also generated high volumes of social media activity, with Twitter first channeling the news that was coming out of street actions and confrontations with the police. According to journalistic accounts, Twitter helped fuel political feuds by enabling bots to exacerbate conflict (26, 27). Our analyses aim to compare the attention that unverified bot accounts received during these events of intense political mobilization with the visibility of verified and mainstream media accounts, contextualizing that activity within the larger online information environment. In particular, we want to determine whether there is a discrepancy in the reach of news sources across channels (i.e., social media and the web) and, if so, determine whether bot activity helps explain that discrepancy (e.g., for instance, by having bots retweet sources that are less visible on the web). Ultimately, our analyses aim to identify changes in the information environment to which people are exposed depending on the channel they use to access news—a process of particular relevance during fast-evolving political events of heightened social tension.

Data

We collected social media data through Twitter's publicly available application programming interface (API) by retrieving all messages that contained at least one relevant hashtag for each of the two mobilization events (see *SI Appendix* for a list of keywords). Our data collection missed less than 1% of all messages with relevant hashtags during the two periods we consider (mid-November to late December of 2018 for the GJ dataset, mid-September to early October of 2017 for the 1-O dataset). In total, we collected tweets sent by hundreds of thousands of unique users (~880,000 for GJ, ~630,000 for 1-O). We used bot detection techniques to identify automated accounts (see *SI Appendix* for the technical details). We then used the "verified" feature to identify the accounts that the platform recognizes as being of public interest. Accounts that exhibit bot-like behavior but are also verified by the platform include news organizations, journalists, and public figures, a category we label with the shorthand "media" even though they have different levels of organizational and software support. Fig. 1 gives a first description of the data: In both datasets, only a small fraction (<1%) of all accounts fall in this media category; about 4 in 10 are classified as unverified bots; and about 6 in 10 are classified as human.

In addition to the Twitter data, we also analyze web-tracking data derived from two representative samples of the online population in France and Spain. In particular, we obtain the audience reach of news sites during the same months for which we have the Twitter data. These measures are based on multiplatform news consumption, that is, access to web domains through desktop, tablet, and mobile (see *SI Appendix* for more details, including the full list of news sites considered). Audience reach is a measure of market share that estimates the fraction of the online population in each country that accessed a given news site during the time in which the mobilizations took place. Fig. 2 summarizes the reach distribution for the two countries. News consumption is more concentrated in France, where it is also a less frequent activity among the online population: Only 5% of all news sites have an audience reach of 13% or higher, with a maximum of 41% (which means than less than half of the online

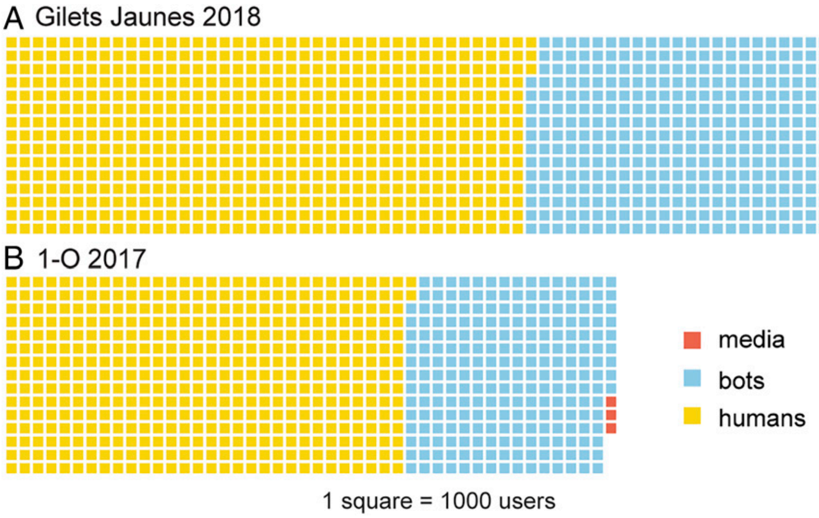

Fig. 1. Classification of Twitter accounts. The "media" category refers to accounts with bot-like behavior verified by the social media platform. These accounts include news organizations, journalists, and public figures. The "bots" category refers to unverified accounts. We classify the rest of the accounts as "human." Media accounts amount to less than 1% of all users engaged in communication around the two contentious mobilizations ($n = 4,117$ for GJ in panel *A*; n = 2,958 for 1-O in panel *B*).

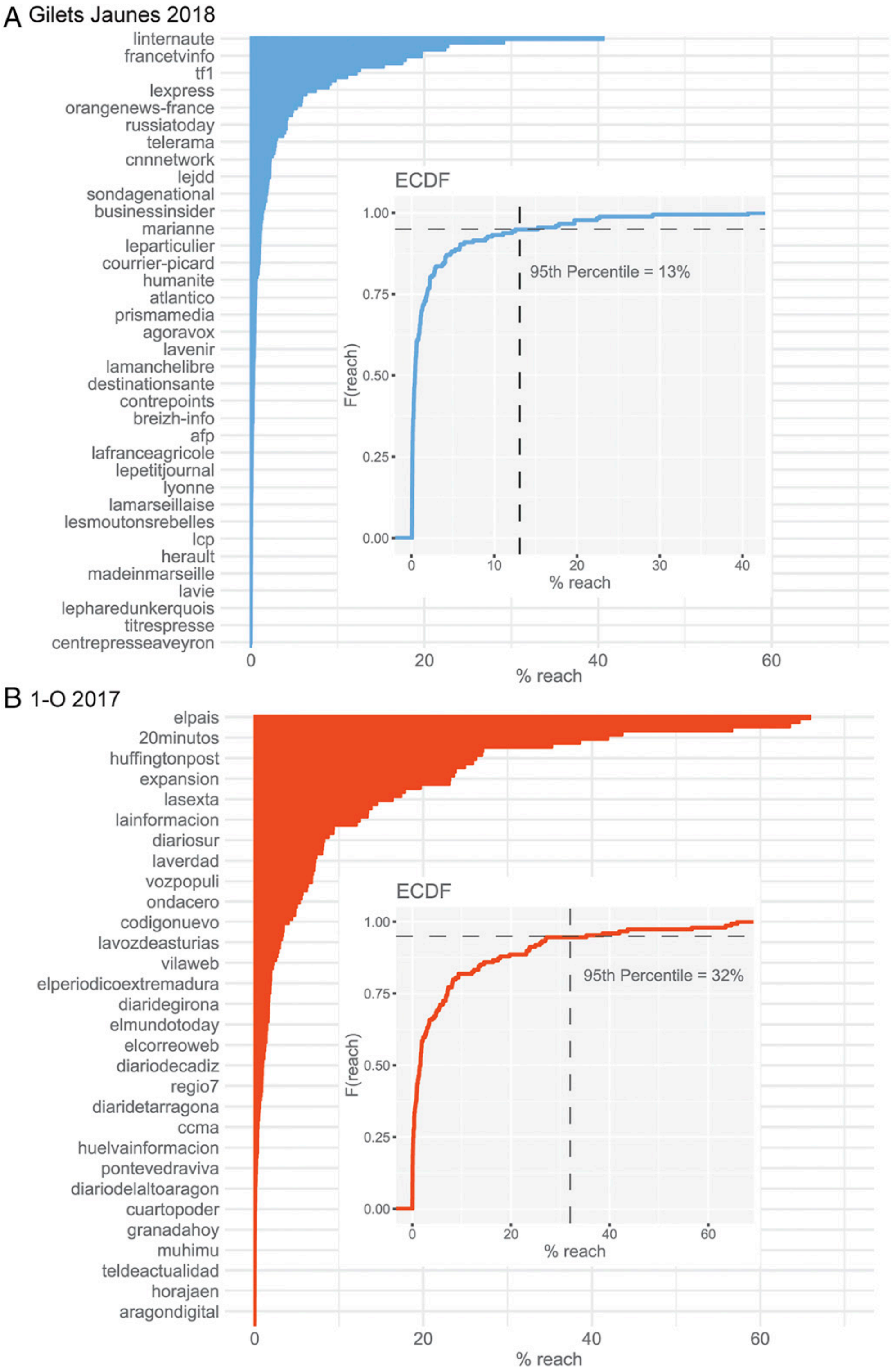

Fig. 2. Audience reach of news sites on the web. Percentage reach refers to the fraction of the online population in France (*A*) and Spain (*B*) that accessed a given news site during the period in which the mobilizations took place (for the full list of sites considered in our analyses, see *SI Appendix*, Tables S1 and S2). The *Insets* show the empirical cumulative distribution function: In France, only 5% of all news sites have an audience reach of 13% or higher; in Spain, only 5% of all news sites have an audience reach of 32% or higher.

population in France were consuming news on the web in this period). In Spain, only 5% of all news sites have an audience reach of 32% or higher, with a maximum of 66% for the most read newspaper. These numbers are consistent with prior work showing that news consumption amounts to a small fraction of all online activity (28, 29), but they also suggest that the overall level of user engagement with digital news and the degree of attention concentration can change substantially across countries.

Of more relevance for our purposes, the data summarized in Fig. 2 allow us to identify the news sites that acted as the main sources of information online during the mobilizations. During the time in which these mobilizations took place, web traffic data suggests that only 26% of the online population in France, and 31% of the online population in Spain, accessed Twitter. According to survey data, the percentage that declared using social media for news was 36% in France and 59% in Spain [data for 2018 and 2017 (30, 31)]. In other words, news consumption on the web and on social media attracts a different user base. Both the design of social media platforms and the fact that Twitter attracts a very specific segment of the online population suggest that, by necessity, these two information environments will differ. The question is how and whether there is evidence that unverified bots contribute, with their activity, to generate those discrepancies.

Results

We manually identified the Twitter handles for the news outlets included in the web-tracking data. Of the 177 news outlets available in France, we could match 126. In Spain, we matched 73 out of the total 149. This means that many of the news sources with audience on the web are not present in the social media coverage of the mobilizations. Fig. 3 summarizes the correlation between Twitter and web visibility for the news outlets that are present in both platforms. The left column shows the association between the rank position according to percentage reach and the number of followers in Twitter (a measure of global visibility, beyond the events surrounding the political mobilizations). The right column shows the association between rank position in percentage reach and centrality in the retweet (RT) network (e.g., number of RTs received). We focus on this measure of centrality because RTs are the main mechanism to diffuse information on the platform and reach a wider audience, and the most straightforward measure of the broadcasting potential of the different accounts covering the events. (Additional statistics describing the RT network as well as the network of mentions can be found in *SI Appendix*, Table S3.) The correlation between centrality in the RT network and percentage reach on the web is very low in both cases. This means that news sources with high visibility on the web did not attract the same amount of attention on the Twitter stream related to these political mobilizations. (Note that the correlation with number of followers is moderate to high, suggesting that the bigger outlets on the web, in terms of audience size, are also more prominent on social media, in terms of followers.) This discrepancy in visibility begs two questions: Is unverified bot activity responsible for this discrepancy? Also, did unverified bots attract more attention than verified users and

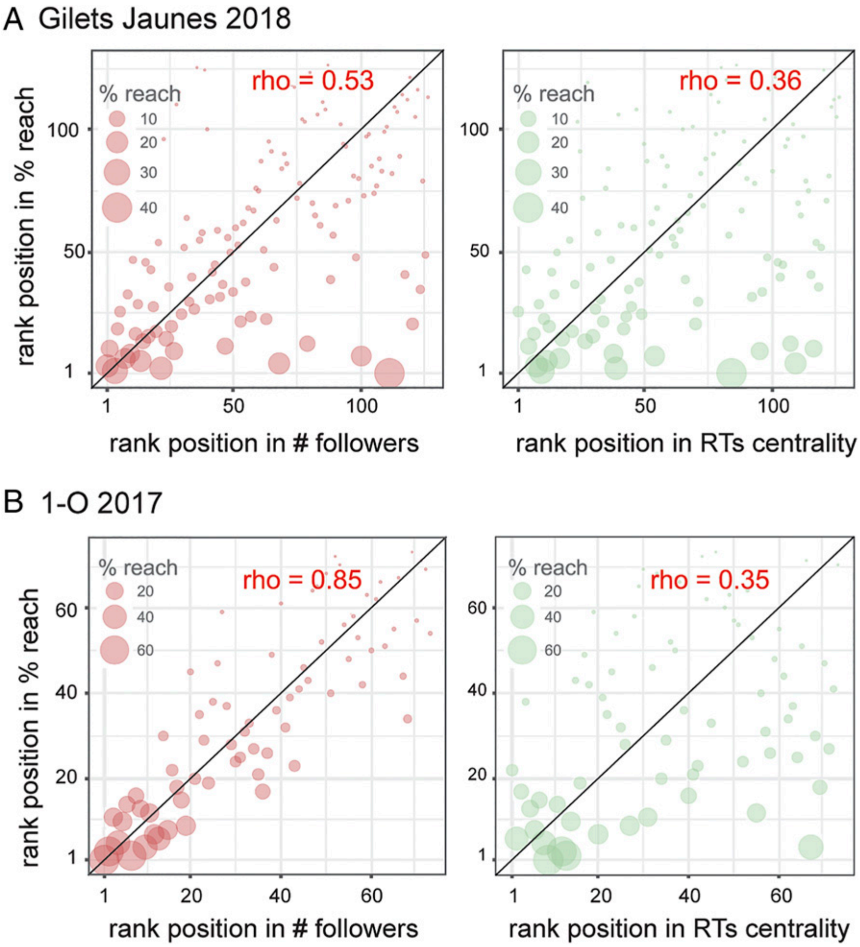

Fig. 3. Correlation of rank position in percentage reach (web) and RT centrality. Panel *A* shows correlations for the Yellow Vests data; panel *B* shows correlations for the Catalan Referendum data. The scatterplots on the *Left* column measure the association between audience reach on the web and number of followers on Twitter; the scatterplots on the *Right* measure the association of reach with number of RTs received in the stream of information related to the protests. For the latter, the correlation coefficients show that the association is weak: The most accessed news sources on the web are not the most of salient in the Twitter stream covering the mobilizations (even though the most accessed web sources tend to also have a higher number of followers).

news outlets, therefore having more potential to manipulate the coverage of the mobilizations?

Fig. 4 shows the RT networks contracted to the three account categories. On the aggregate, most RT activity happens between human accounts—there are many more of these in both datasets so, overall, they accumulate most of the activity. However, proportionally, media accounts receive many more RTs than expected by chance. The boxplots on the *Left* summarize the values obtained from permuting the data by randomly relabeling the category of individual accounts. What these randomizations reveal is that bots retweet significantly more (human accounts retweet less) than expected, but this level of activity does not translate into a significantly higher centrality in the number of RTs received: In both mobilizations, verified media accounts are the most often retweeted accounts. These differences in centrality remain significant after controlling for number of followers, friends, and number of RTs made: Verified media accounts are significantly and substantially more central in the RT network, with bots being only slightly more central than human accounts (*SI Appendix*, Figs. S6 and S7). Reciprocity is very low in both the RT and the mention networks (*SI Appendix*, Table S3), and the few reciprocated connections that exist are concentrated among human accounts (*SI Appendix*, Fig. S11); this suggests that if bots are trying to engage other users in reciprocated exchange (with the goal of boosting their own visibility), the strategy is not successful. However, bot and human accounts form clear communities of information exchange (*SI Appendix*, Fig. S5), which means that, even if verified media accounts are significantly more central, human accounts still have large exposure to bot-generated content.

Fig. 5 zooms into the subset of media accounts for which we have web-tracking data to examine the composition of their neighborhood in the RT network. In particular, we count the number of unverified bots retweeting these news outlets. If bot activity is responsible for the discrepancy in rankings identified in Fig. 3 (for instance, if bots retweet more frequently sources that are less visible on the web, boosting their social media visibility), we would expect a positive association between being more visible on Twitter and having more bots retweeting their content. The left column confirms this intuition, showing that there is a linear association and moderate correlation between bot retweeting activity and rank differences. However, this correlation disappears once we factor in the total number of retweeting accounts and normalize the number of bots as a fraction of the neighborhood size (right column). This means that the differences in visibility identified in Fig. 3 are a function of the

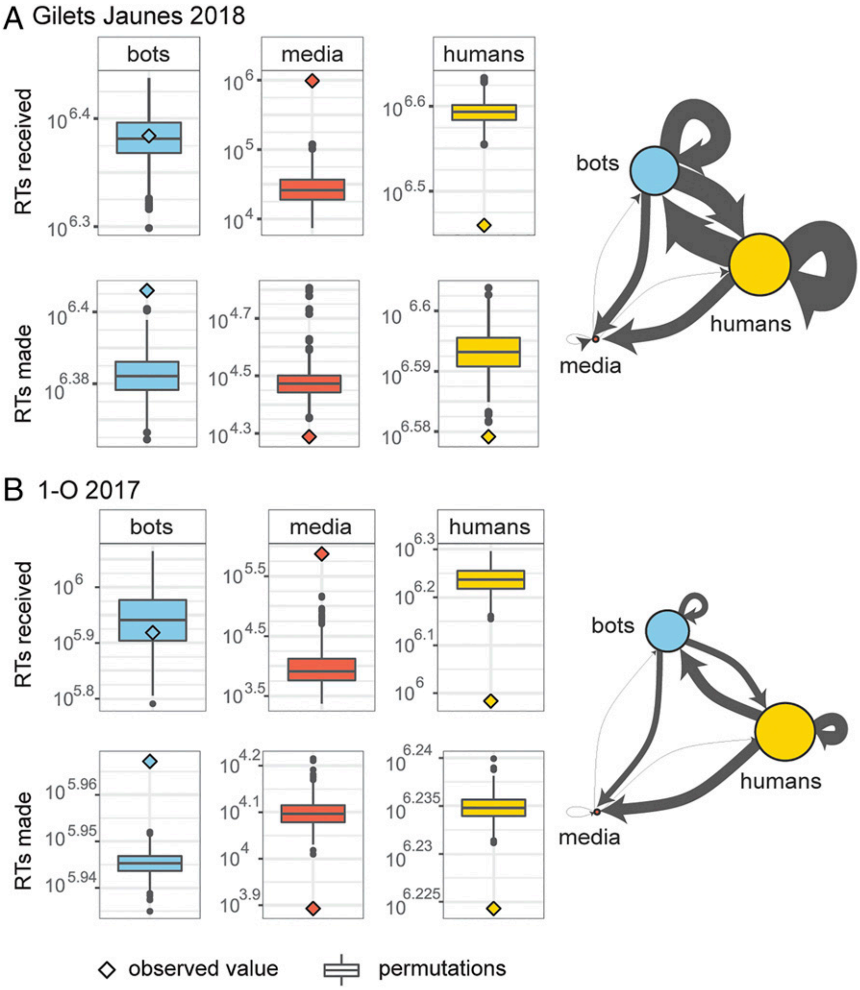

Fig. 4. Centrality in the RT network. Panel *A* shows the network built with the Yellow Vests data; panel *B* shows the network built with the Catalan Referendum data. The boxplots summarize the values obtained from permutations of the data where the category labels were randomly reshuffled across accounts. The observed centrality of media accounts is significantly higher than expected by chance in both mobilizations. Human accounts, on the other hand, receive significantly fewer RTs. The axes preserve different scales to allow visual identification of distance between permutations and observed values. Regression models predicting centrality can be found in *SI Appendix*, Figs. S6 and S7.

centrality of news outlets in the RT network, for which humans are as responsible as bots (in fact, for most outlets, there are more humans retweeting their messages than unverified bots).

In order to illuminate the factors predicting the number of RTs that individual messages receive, we fitted mixed-effects models at the message level (since messages are nested within unique users, we treat account-level variability as the random effect; see *SI Appendix* for more details on specification). Fig. 6 summarizes the results of these models. We used as controls structural features (e.g., the number of followers and friends of the accounts publishing the messages) and the content of the tweets. For this, we used a lexicon and rule-based sentiment analysis technique to extract message scores that range from −1 (extremely negative sentiment) to +1 (extremely positive sentiment), with 0 values representing neutral messages (32) (see *SI Appendix* for more details on the technique and for the distribution of sentiment scores). More importantly, the models also include binary variables identifying messages sent by media and human accounts (bots are the base category). As the figure shows (top panels), controlling for structure and content, and for random variability at the user level, messages published by verified media accounts receive significantly more RTs than messages published by bots. Human messages, on the other hand, receive less diffusion. The models summarized in the lower panels add another binary variable identifying the messages published by the subset of news outlets for which we could measure visibility on the web. Overall, the estimates for the other parameters do not change, but the models suggest that, controlling for content, the messages published by news outlets with wider web audience actually receive fewer RTs than the messages of other verified media accounts. In addition, negative content seems to have resonated more (the effect, however, is small).

Discussion

Social media create a very different landscape in which to obtain news: Compared to the web, Twitter has many more verified sources that get amplified with a combination of bot and human effort. We measured discrepancies in the salience of news outlets on Twitter and on the web, and we presented evidence suggesting that bots did not play a fundamental role in creating those discrepancies or taking centrality positions. We show that media accounts (e.g., accounts with bot-like behavior that are also verified by Twitter) are more likely to receive RTs and, therefore, that they were reference points in the coverage of the mobilizations. Their messages actually received more attention than news outlets with sustained traffic on the web. We also show

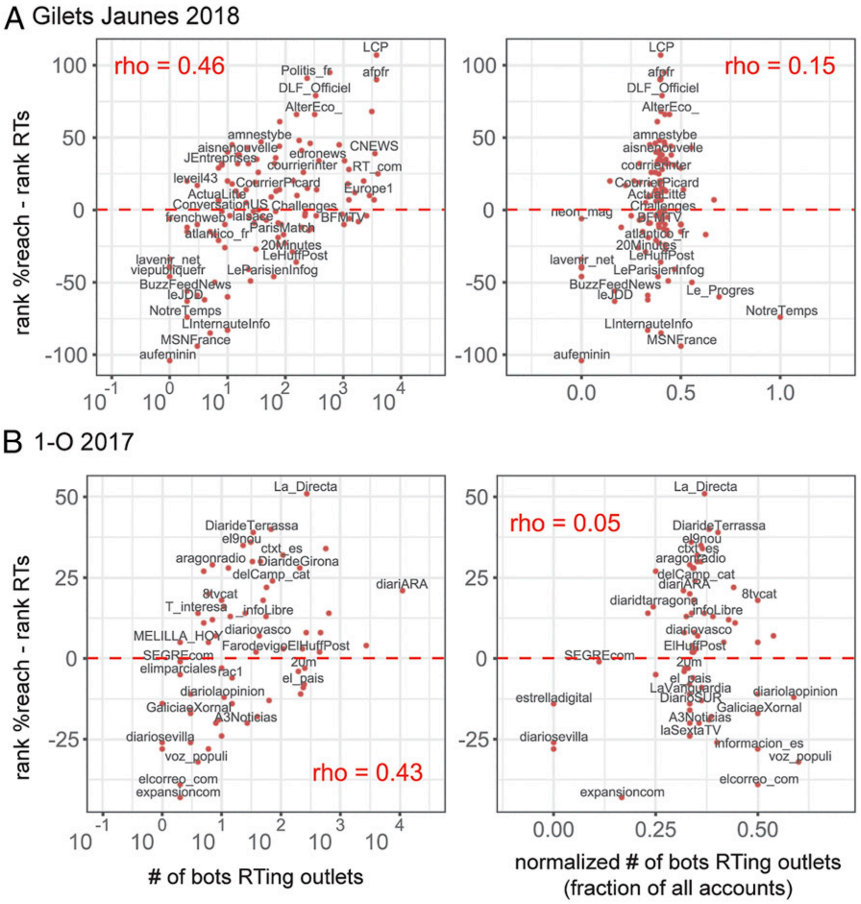

Fig. 5. Number of bots retweeting news outlets. Panel *A* shows correlations in the Yellow Vests data; panel *B* shows correlations in the Catalan Referendum data. The scatterplots measure the association between the number of unverified bots retweeting news outlets and the differences in visibility rankings (identified in Fig. 3, operationalized here as rank $_{\%\ \text{reach}}$ – rank $_{\text{RTs}}$). This allows us to relate the difference between Twitter and web-tracking ranks with bot retweeting activity. Outlets above the 0 line have better ranking on Twitter; outlets below the 0 line have better ranking on the web. The horizontal axes measure the number of unverified bots retweeting news outlets (left column) and the number of bots normalized by neighborhood size (right column). The scatterplots on the *Left* suggest that news outlets that have a higher number of retweeting bots have more salience on Twitter than on the web. However, the association disappears once we normalize the number of bots as a fraction of the neighborhood. News outlets that are more visible on Twitter simply have more accounts retweeting their content, including human accounts.

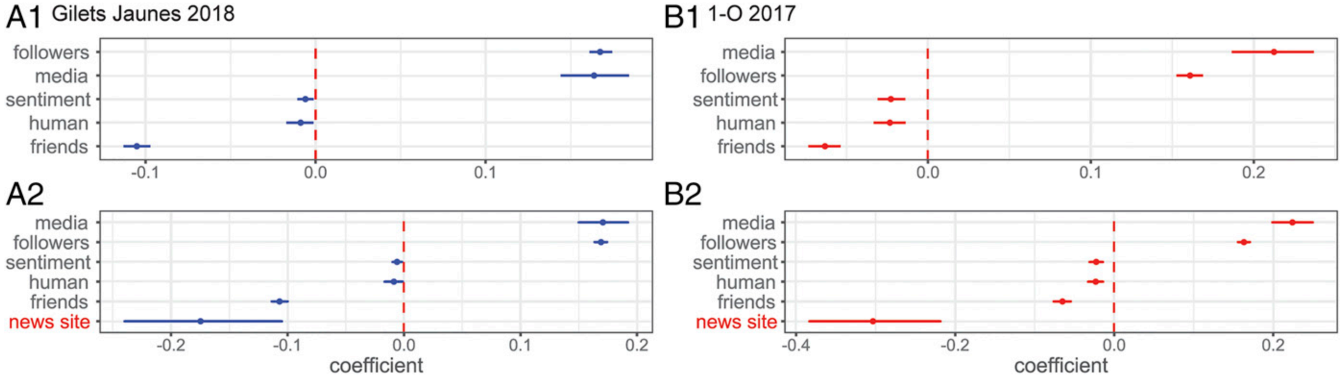

Fig. 6. Predictors of the number of RTs received by messages. Panels *A1* and *A2* show estimates for the Yellow Vests data; panels *B1* and *B2* show estimates for the Catalan Referendum data. The estimates result from message-level mixed-effects models (with 95% bootstrapped CI). Upper row: controlling by structure and content (and random variability at the account level), messages published by media accounts receive more RTs than messages published by unverified bots (the base category). Messages posted by human accounts, however, receive less diffusion. Lower row: Tweets published by the subset of news organizations for which we have web reach data also receive less attention than verified media accounts, an effect that is particularly strong and significant during the Catalan referendum.

that bot activity is not responsible for creating discrepancies in the visibility of news outlets: If anything, human retweeting is more important in amplifying differences of visibility on social media compared to the web.

A more likely cause for these discrepancies is that social media and web news content attract different populations. Users consuming news on the web or actively engaging with political content on Twitter are not representative of the population at large (e.g., the vast majority of online users do not proactively consume news, hence the low audience reach of news sites displayed in Fig. 2). However, it is unclear whether the Twitter and web populations are biased in different ways, which would explain the discrepancies documented here; or whether, instead, those discrepancies arise from the specific affordances of the social media platform (e.g., ref. 33)—exposure to news, in the end, responds to different mechanisms on the web and on social media. Our analyses suggest, in any case, that unverified bots are not the cause of those discrepancies.

Our analyses also reveal that the number of media accounts amount to a very small minority. In aggregate terms, these verified accounts generated only a small fraction of all messages covering the mobilizations and attracted a small fraction of all RTs. Unverified bots, by virtue of being more numerous, generated more content and interacted with more humans. Human accounts, on the other hand, received significantly less attention than automated accounts. This means that in certain pockets of the user base following the mobilizations through Twitter, exposure to unverified bots might have counteracted exposure to the content published by media accounts. Based on our observational data, it is not possible to determine whether these interactions and types of exposure had any effects on opinions and attitudes; for that, longitudinal survey measures need to be combined with Twitter data. Because of the difficulties of obtaining such data, there is not much research evaluating the effects side of exposure to content on social media, but the research that exists (e.g., ref. 14) presents evidence that attempts to manipulate information in social media led to no identifiable changes in the attitudes and behaviors of the users exposed. This finding casts doubts on the actual behavioral impact of information campaigns on social media, but more research is necessary to provide additional evidence. Likewise, it is also difficult to determine whether exposure to other media content, on the web or, more generally, in the wider news ecosystem, can counteract these effects [TV news exposure still is, after all, the most prevalent form of news consumption (28)].

Compared to other media accounts, the tweets published by established news outlets (e.g., those with enough traffic to appear in web-tracking data) do not gain much traction during the coverage of the political events—their messages are actually less visible (i.e., they receive less RTs) than those of other verified accounts, especially in the case of the Catalan referendum. The Spanish online population consume more news than the French online population (Fig. 2), and survey data suggest that they also trust news more: 51% of the Spanish population think that "you can trust most news most of the time" versus 30% of the French population [where trust levels in the news are among the lowest in Europe (30, 31)]. However, despite this, Spanish Twitter users chose to amplify the messages of other verified accounts over the accounts of established news organizations to a more significant extent. Many of these verified accounts actually belong to journalists working for established news outlets, but also to public figures and civil society representatives with no affiliation to news organizations. This offers additional evidence that social media is eroding traditional gatekeeping roles.

Our findings also highlight the importance of clearly defining the policies that underlie the verification process of social media accounts. Our analyses do not directly address the question of whether verified, blue-badge accounts are more reliable sources of information, but there is abundant evidence that misleading information is often spread from verified accounts, including those of government officials and recognized public figures. Verified accounts have also amplified the messages of dubious sources that would be fringe and less visible otherwise. Future research should shift attention from bots to verified accounts and the complex relationships that emerge in their interactions with other user accounts—effective disinformation campaigns certainly involve a range of different participants (34), and the actors behind those campaigns work around platform policies, including public verification programs.

Given the impact that misinformation can have on democracies, it is important to understand the role that bots play in the dissemination of news, especially in the context of contentious political events. However, it is also important to contextualize the role of bots in the broader media landscape. Here, we confirm that Twitter and the web are very different news ecosystems but that unverified bots are not responsible for the discrepancy in source salience. We show that verified media accounts are still more central in the diffusion of information, but also that sources that are salient on the web are less salient in the social media platform. We also show that human accounts are still significantly less visible than unverified bots so more research is necessary to determine whether, for certain populations, exposure to bot content has any effects on opinions and behavior. More generally, this research does not speak directly to broader questions of social media manipulation, like hijacking hashtags or gaming the algorithmic ranking of content—on these fronts, bot activity may turn out to be more successful. In the context of the two contentious events we analyze, however, we find little evidence that bots managed to gain comparable prominence to verified media accounts.

Materials and Methods

Data. We collected social media data through Twitter's publicly available API by retrieving all messages that contained at least one relevant hashtag (see *SI Appendix* for the list of keywords). Based on Twitter's rate limits, we estimate that our data collection missed less than 1% of all messages during the period we consider (mid-November to late December 2018 for the GJ dataset, and mid-September to early October for the 1-O dataset). The web-tracking data are based on Comscore's MMX Multi-Platform panel (i.e., key measures reports). We averaged the estimates for November to December 2018 (GJ data) and for September to October 2017 (1-O data). See *SI Appendix* for a full list of news sites included.

Methods. We build the RT and mention networks (weighted) and calculate centrality scores on the largest connected components (see *SI Appendix* for descriptive statistics). We identify automated accounts using a bot classification technique trained and validated on publicly available datasets. We follow state-of-the-art techniques in building this classifier, and cross-validation of model performance on an independent dataset suggests that the classifier generalizes in a satisfactory manner (see *SI Appendix* for more details on our model and cross-validation checks). However, out-of-domain performance is still an open problem for many ML systems, including ours, and this should be borne in mind when assessing our results. We quantify the sentiment of each Tweet using a well-established natural language processing technique named VADER (Valence Aware Dictionary and Sentiment Reasoner) (32). VADER is a lexicon and rule-based sentiment analysis tool that is specifically designed to analyze sentiments expressed in social media. The scores range from −1 (extremely negative sentiment) to +1 (extremely positive sentiment), with 0 values representing neutral messages. We modify VADER to natively support sentiment analysis of texts in French, Spanish, and Catalan (in addition to English; see *SI Appendix* for more details).

Models. To identify the factors that predict the number of RTs messages receive, we used mixed-effects models at the message level (35, 36). We use user account ID as the random effect. Fixed effects include three control variables (number of followers, friends, and the sentiment score of messages), and two explanatory variables (i.e., whether the account posting the messages is classified as media or as human, with "bots" as the base category). See *SI Appendix* for additional details and robustness checks.

Data Availability. Data and code to reproduce these results have been deposited in the Open Science Framework (https://osf.io/j65de/) (37).

ACKNOWLEDGMENTS. Work on this paper was partly funded by NSF Grants 1729412 and 2017655, and it was developed while S.G.B. was on sabbatical leave at the Center for Advanced Study in the Behavioral Sciences at Stanford University.

1. S. J. Jackson *et al.*, *HashtagActivism: Networks of Race and Gender Justice* (MIT Press, Cambridge, MA, 2020).
2. Z. Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (Yale University Press, New Haven, CT, 2017).
3. D. Freelon, C. McIlwain, M. Clark, Quantifying the power and consequences of social media protest. *New Media Soc.* **20**, 990–1011 (2016).
4. S. González-Bailón, J. Borge-Holthoefer, A. Rivero, Y. Moreno, The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011).
5. R. M. Bond *et al.*, A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
6. P. Barberá *et al.*, The critical periphery in the growth of social protests. *PLoS One* **10**, e0143611 (2015).
7. J. M. Larson *et al.*, Social networks and protest participation: Evidence from 130 million twitter users. *Am. J. Pol. Sci.* **63**, 690–705 (2019).
8. N. F. Johnson *et al.*, The online competition between pro- and anti-vaccination views. *Nature* **582**, 230–233 (2020).
9. J. Lehmann *et al.*, "Dynamical classes of collective attention in Twitter" in *Proceedings of the 21st International Conference on World Wide Web* (ACM, 2012), pp. 251–260.
10. M. De Domenico, E. G. Altmann, Unraveling the origin of social bursts in collective attention. *Sci. Rep.* **10**, 4629 (2020).
11. M. Cha *et al.*, "Measuring user influence in Twitter: The million follower fallacy" in *International AAAI Conference on Weblogs and Social Media (ICSWM)* (Association for the Advancement of Artificial Intelligence, 2010).
12. M. Stella, M. Cristoforetti, M. De Domenico, Influence of augmented humans in online interactions during voting events. *PLoS One* **14**, e0214210 (2019).
13. D. Carter, Hustle and brand: The sociotechnical shaping of influence. *Soc. Media Soc.* **2**, 2056305116666305 (2016).
14. C. A. Bail *et al.*, Assessing the Russian internet research agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 201906420 (2019).
15. D. Freelon *et al.*, Black trolls matter: Racial and ideological asymmetries in social media disinformation. *Soc. Sci. Comput. Rev.*, 10.1177/0894439320914853 (2020).
16. S. C. Woolley, P. N. Howard, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media* (Oxford University Press, 2018).
17. O. Varol *et al.*, "Online human-bot interactions: Detection, estimation, and characterization" in *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)* (Association for the Advancement of Artificial Intelligence, 2017), pp. 280–289.
18. E. Ferrara, Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday* **22**, https://doi.org/10.5210/fm.v22i8.8005 (2017).
19. A. Bessi, E. Ferrara, Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday* **21**, https://doi.org/10.5210/fm.v21i11.7090 (2016).
20. M. Stella, E. Ferrara, M. De Domenico, Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 201803470 (2018).
21. C. Shao *et al.*, The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 4787 (2018).
22. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
23. S. Wu *et al.*, "Who says what to whom on Twitter" in *World Wide Web 2011 Conference* (ACM, 2011), 978-1-4503-0637-9/11/03.
24. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374–378 (2019).
25. O. Varol, I. Uluturk, Journalists on Twitter: Self-branding, audiences, and involvement of bots. *J. Comput. Soc. Sci.* **3**, 83–101 (2020).
26. C. Baraniuk, How Twitter bots help fuel political feuds. *Scientific American*, **27** March 2018. https://www.scientificamerican.com/article/how-twitter-bots-help-fuel-political-feuds/. Accessed 27 June 2020.
27. C. Matlack, R. Williams, France to probe possible Russian influence on yellow vest riots. Bloomberg, 7 December 2018. https://www.bloomberg.com/news/articles/2018-12-08/pro-russia-social-media-takes-aim-at-macron-as-yellow-vests-rage. Accessed 27 June 2020.
28. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
29. T. Yang *et al.*, Exposure to news grows less fragmented with an increase in mobile access. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28678–28683 (2020).
30. Reuters Institute, Digital News Report *2018* (Reuters Institute for the Study of Journalism, 2018).
31. Reuters Institute, Digital News Report *2017* (Reuters Institute for the Study of Journalism, 2017).
32. C. Hutto, E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text" in *Eighth International Conference on Weblogs and Social Media (ICWSM-14)* (Association for the Advancement of Artificial Intelligence, 2014).
33. K. Jaidka, A. Zhou, Y. Lelkes, Brevity is the soul of Twitter: The constraint affordance and political discussion. *J. Commun.* **69**, 345–372 (2019).
34. K. Starbird, Disinformation's spread: Bots, trolls and all of us. *Nature* **571**, 449 (2019).
35. A. Galecki, T. Burzykowski, *Linear Mixed-Effects Models Using R* (Springer, New York, 2013).
36. J. J. Faraway, *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models* (Chapman and Hall, New York, 2005).
37. S. Gonzalez-Bailon, M. De Domenico, BotsLessCentral. Open Science Framework. https://osf.io/j65de/. Deposited 23 February 2021.