Bots increase exposure to negative and inflammatory content in online social systems

Massimo Stella[a], Emilio Ferrara[b,1], and Manlio De Domenico[a,1]

[a]Center for Information and Communication Technology, Fondazione Bruno Kessler, 38123 Trento, Italy; and [b]USC Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292

Edited by Jon Kleinberg, Cornell University, Ithaca, NY, and approved October 19, 2018 (received for review February 27, 2018)

Societies are complex systems, which tend to polarize into subgroups of individuals with dramatically opposite perspectives. This phenomenon is reflected—and often amplified—in online social networks, where, however, humans are no longer the only players and coexist alongside with social bots—that is, software-controlled accounts. Analyzing large-scale social data collected during the Catalan referendum for independence on October 1, 2017, consisting of nearly 4 millions Twitter posts generated by almost 1 million users, we identify the two polarized groups of Independentists and Constitutionalists and quantify the structural and emotional roles played by social bots. We show that bots act from peripheral areas of the social system to target influential humans of both groups, bombarding Independentists with violent contents, increasing their exposure to negative and inflammatory narratives, and exacerbating social conflict online. Our findings stress the importance of developing countermeasures to unmask these forms of automated social manipulation.

computational social science | complex networks | machine learning | sociotechnical systems | human behavior

Societies consist of agents engaging in multimodal social actions with one another in a complex system (1). This "society-as-system" metaphor inspired many computational studies aimed at identifying, at a microscopic level, how social interactions might lead to emergent global phenomena such as social segregation (2), spreading of information (3), and behavior (4, 5). The recent advent of digital communication systems has dramatically shifted the investigation from empirical social interactions in the physical world to online social platforms and technology-mediated interactions (6). Online platforms revolutionized the society-as-system metaphor (7) by providing detailed datasets suitable for large-scale investigation of patterns reflecting real-world social phenomena such as the presence and role of influencers in information diffusion (8–11), the effect of emotions on social ties (12), or the polarization of agents according to stances (13–15). Social media yields an invaluable source of information for learning the mechanisms behind social influence and social dynamics (16–18). However, digital systems are not populated only by humans, but also by software-controlled agents, better known as bots, programmed to pursue specific tasks, from sending automated messages to assuming specific social or antisocial behaviors (19, 20). Similarly to human interactions, bots might be able to affect structure and function of a social system (18). Understanding how human–bot dynamics drive social behavior is of utmost importance: As postulated by the theory of embodied cognition (21), the presence of robots in a social system affects the way humans perceive social norms and how they interact with one another and with the robots.

Here, we show how social bots play a central role in the collective dynamics taking place on online social systems during a voting event, namely, the Catalan Referendum of October 1, 2017. To this end, we monitored the discussion on a popular microblogging platform (Twitter) from September 22, 2017, to October 3, 2017. We discovered that bots generated specific content with negative connotation that targeted the most influential individuals among the group of Independentists (i.e., Catalan independence supporters). For our analysis, we first detect bots by using a cutting-edge scalable approach and find that nearly one in three users in this conversation is a bot.

Results

By disentangling the observed social interactions in retweets (who reshares the content posted by whom), replies (who responds to whom), and mentions (who attracts the attention of whom), we find that humans and bots share similar temporal behavioral patterns in the volume of messages. Both groups display daily excursions resembling a circadian rhythm, with a dramatic increase in the activity rate on October 1. Fig. 1 *B*, *Lower*, shows that bots produced 23.6% of the total number of posts during the event (retweets and mentions show comparable values). Notably, the percentage of Replies generated by bots increases to 38.8%, suggesting that during this event, bots preferred this form of targeted responses.

To better characterize the nature of the observed interactions, we investigate the targets of such intensive social activities. Fig. 1*A* and *SI Appendix*, Fig. S1*A* summarize the structure of human–bot interactions. While humans interact mostly with other humans, 19% of overall interactions are directed from bots to humans, mainly through retweets (74%) and mentions (25%), *SI Appendix*, Fig. S1 *B–D*.

To shed light on the nature of these human–bot interactions, we focus on the semantic content of posted messages. Sentiment

Significance

Social media can deeply influence reality perception, affecting millions of people's voting behavior. Hence, maneuvering opinion dynamics by disseminating forged content over online ecosystems is an effective pathway for social hacking. We propose a framework for discovering such a potentially dangerous behavior promoted by automatic users, also called "bots," in online social networks. We provide evidence that social bots target mainly human influencers but generate semantic content depending on the polarized stance of their targets. During the 2017 Catalan referendum, used as a case study, social bots generated and promoted violent content aimed at Independentists, ultimately exacerbating social conflict online. Our results open challenges for detecting and controlling the influence of such content on society.

Author contributions: M.S., E.F., and M.D.D. designed research; M.S., E.F., and M.D.D. performed research; M.S., E.F., and M.D.D. contributed new reagents/analytic tools; M.S. and M.D.D. analyzed data; and M.S., E.F., and M.D.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: emiliofe@usc.edu or mdedomenico@fbk.eu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1803470115/-/DCSupplemental.

Published online November 20, 2018.

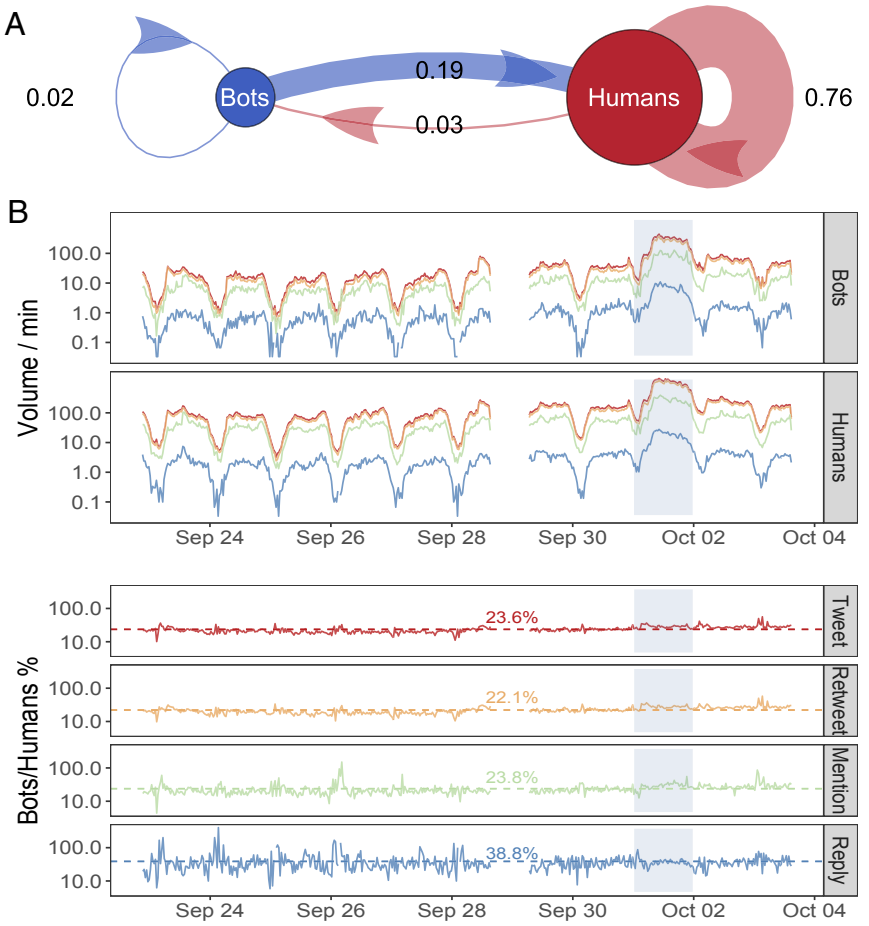

Fig. 1. Social activity of humans and bots over time. (*A*) Flowchart of human–bot Twitter interactions across the whole time window. A total of 19% of the considered interactions are from bots to humans. (*B*, *Upper*) The volume per minute for different social actions (tweet, retweet, mention, and reply). (*B*, *Lower*) The fraction of volume generated by bots. Shaded areas highlight October 1, 2017, the day of the Catalan referendum.

analysis (Materials and Methods) reveals interesting differences in emotional trends between humans and bots (Fig. 2). Retweets directed to bots do not display any evident deviation from neutrality (0 sentiment score), while interactions directed toward humans display marked positive and negative trends of sentiment intensity. An analogous behavior happens also for mentions (light colors). These differences indicate that bot-targeted interactions are not significantly influenced by the underlying social dynamics, and hence the analysis should focus more on human-targeted interactions (i.e., human-to-human and bot-to-human). The sentiment of human-to-human interactions displays marked trends in different phases: (*i*) a trending positive average sentiment score in the days before September 30 (Fig. 2, *Upper*, HH); (*ii*) a sudden drop in sentiment starting from the midnight of October 1 (Fig. 2, *Lower*, HH) after negative contents start getting reshared; (*iii*) a peak of negative sentiment in the midday of October 1; and (*iv*) a later increase in sentiment toward neutrality. These sentiment scores and their related content both indicate that human-to-human interactions are a powerful proxy of the dynamics of underlying real-world social systems. The drastic drop of sentiment score from positive to negative among >300,000 human users signals the presence of polarization in the social system, due either to opposing factions exchanging positive/negative messages or to the influence of nonhumans. Fig. 2 also highlights important differences between human and bot interactions: The drop in average sentiment evident in bot-to-human interactions is not present in bot-to-bot interactions. This difference indicates that automated content generated and endorsed by bots is not influenced by the social dynamics relative to the referendum: On average, bot-to-bot interactions are not influenced by the human polarization relative to the referendum. Such human polarization is captured by bot-to-human interactions instead: This distinctiveness indicates that bot-to-human interactions promote human-generated content, which is subject to polarization.

Identifying user polarization (i.e., users being in favor of or against a given event or topic) cannot be performed with sentiment only (22). We overcome this limitation by exploiting a synergy between the network structure of social actions and their emotional intensities, with the aim of identifying stances focused on the voting event in our dataset: Constitutionalists and Independentists to the Catalan referendum. Notice that our network-enhanced stance detection analysis has two major elements compared with previous approaches (22), as it not only considers semantic features of messages but also the structure of their exchanges and the nature of their recipients.

To capture pivotal trends in the structure of social interactions, we focus on the core of the network of social interactions (Materials and Methods). It is well documented that people tend to retweet each other as a form of social endorsement (23). To filter out spurious or infrequent interactions, we consider the available multimodal information and focus on strong social interactions (i.e., those actions where users perform at least a retweet and either a reply or a mention) during the considered time window. We use strong ties to identify the network core, shown in Fig. 3. To determine the two underlying polarized groups, we

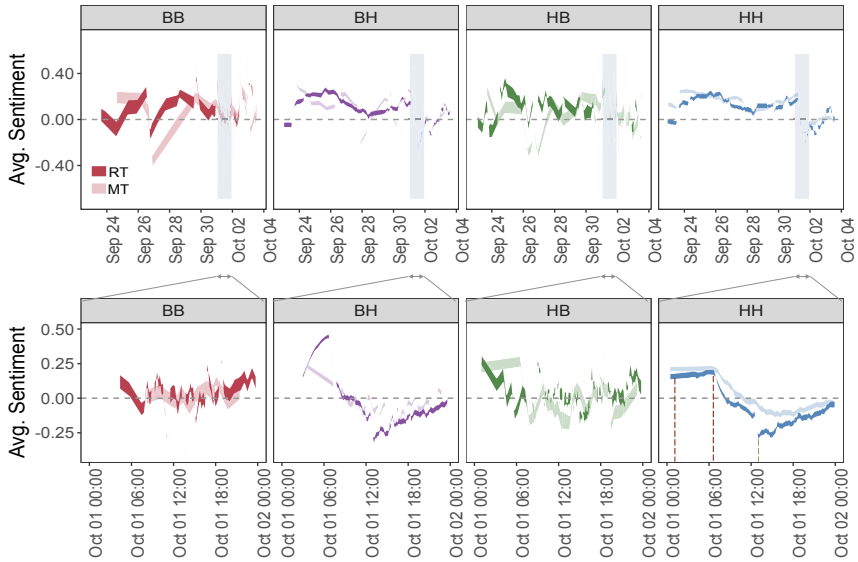

Fig. 2. Sentiment evolution before, during, and after the Catalan referendum. Average (Avg.) sentiment scores for retweets (darker colors) and mentions (lighter colors) over time for human-to-human (HH), human-to-bot (HB), bot-to-human (BH), and bot-to-bot (BB). The gray box highlights the day before the Catalonia ballot. While bot-to-bot and human-to-bot display no clear trend over time, human interactions display a positive pattern of sentiments until September 30, after which a drop in sentiment up to negative values appears in human-to-human and bot-to-human interactions. In the lower right, negative tweets are generated around 1:00 AM October 1, but they start spreading only in the morning, after 7:00 AM. Positive tweets start spreading after noon.

look for a partition that minimizes intergroup interactions and use the Fiedler vector approach (24) for an efficient estimation (compare Materials and Methods). The results are shown in Fig. 3*A*. Each group includes ∼6,300 users, with 18% (12%) of them being bots in group 1 (group 2). Within both groups, human-to-human interactions are the most frequent ones, followed by bot-to-human (Fig. 3*B*). Humans in group 1 direct toward bots almost 100 times more social interactions than in group 2, suggesting a larger influence of bots on the social dynamics in group 1 rather than in group 2. Bot–bot interactions across the two groups are absent since bots mostly interact with humans.

To understand the importance of humans and bots in this network, we calculate the PageRank, a widely used measure of users' importance in online networks (25). On average, we find

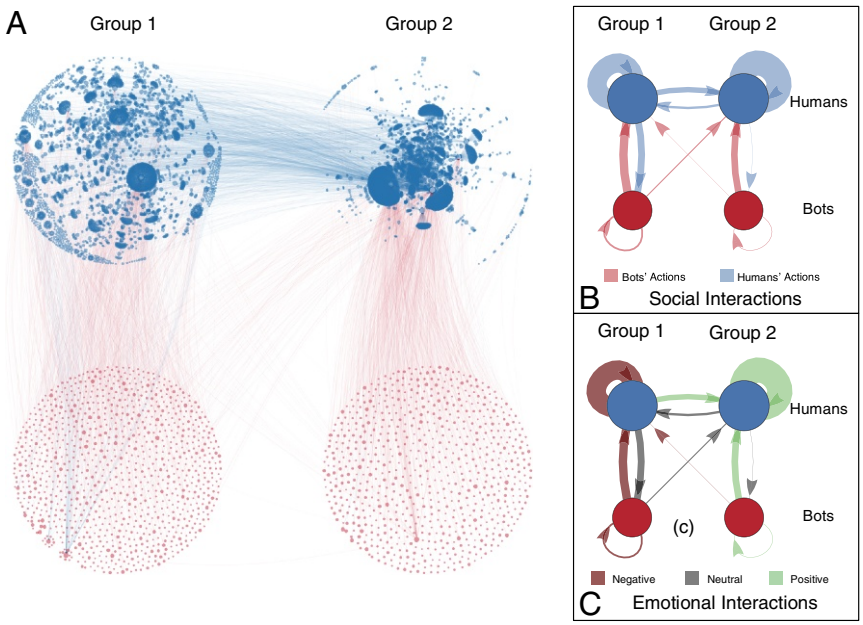

Fig. 3. Network of Twitter interactions. (*A*) Visualization of the network among users classified with respect to faction and bot/human class. Nodes indicate users, and links encode their social interactions (retweet and reply or mention). (*A*, *Upper*) Subnetworks corresponding to the factions consisting of humans. (*A*, *Lower*) Subnetworks of bot factions. Colors encode interactions started by humans (blue) or bots (red). (*B*) Total traffic of Twitter interactions among humans and bots. Thicker edges indicate higher traffic volume. (*C*) Median sentiments of Twitter interactions among factions. Interactions with average negative (positive) sentiment are in dark red (green). Black corresponds to interactions on average compatible with neutrality. Distributions of sentiments are tested against neutrality (i.e., 0 sentiment score) with a sign test at a 95% confidence level.

that humans are 1.8 times more central than bots, highlighting that the latter tend to act from the periphery of the social system. Interestingly, despite their peripheral position, bots target their interactions strategically, mostly directing their activity toward human hubs, playing an influential role in the system. If we define the in-degree of a user as the number of its incoming interactions, then the in-degree of humans with respect to interactions incoming only from bots correlates positively with the in-degree with respect to interactions incoming only from humans (Kendall tau $\kappa \approx 0.62$, $P < 10^{-4}$), indicating that bots tend to target their interactions mainly with the most connected humans. Analogously, humans also tend to interact mainly with the most connected bots (Kendall tau $\kappa \approx 0.75$, $P < 10^{-4}$). To verify if these effects are genuine, we performed the same analysis on randomized realizations of the network while preserving the empirical degree distribution. In this test, the observed correlations are no longer present, supporting the hypothesis of a strategic targeting of social interactions. Since hubs in online social networks like Twitter characterize broadcasters and influencers (10), the above results suggest that bots interacting with human hubs can influence the social dynamics of both groups, while remaining in the periphery of the microblogging social system. The volume of bot/human endorsements in the two groups and the fact that bots mainly target human hubs indicate that social bots can be influential: They promote human-generated content from hubs, rather than automated tweets, and target significant fractions of human users—as evidenced by the fraction of endorsements shown in Fig. 3*B* and reported in *SI Appendix*.

To harness the emotional structure of the links in the network core, we perform a sentiment analysis of the interactions among humans and bots in the two groups (Fig. 3*C* and Materials and Methods). The resulting atlas of emotional interactions indicates that the average sentiments of human-to-human and bot-to-human interactions are negative within group 1 and positive within group 2. This substantial difference in sentiment suggests that the two identified groups endorse their exchanged messages in a different way. In fact, group 1 preferentially endorses negative content. The volumes and sentiment polarities reported in Fig. 3*C* highlight an important mechanism of social contagion played by bots. First, bots direct significant fractions of endorsements to human users, thus actively exposing humans to some type of automatically generated content. However, this content crucially depends on the targets of the interaction: The polarity of endorsements from bots to humans coincide in both groups with the average sentiment of human–human interactions. In turn, this indicates that bots exploit and promote human-generated content, with the same polarity of the endorsements in a given group of human users. In this way, social bots accentuate the exposure of opposing parties to negative content, with the potential to exacerbate social conflict online.

To characterize the semantic nature of group-specific endorsements (e.g., aggressive, pessimistic, etc.), we build and analyze networks of hashtag co-occurrences (Materials and Methods), providing a proxy of users' mindset—that is, the way users perceive and associate concepts (26–28). A consistency analysis indicates that the two groups post messages about a common set of 4,132 hashtags but associate the corresponding concepts in different ways. Fig. 4 shows how the same hashtags co-occur differently in group 1 and 2. Capitalizing on this finding, we focus on those specific concepts that are most important for one group but most peripheral in the other one. We quantify the importance of concepts by identifying the hashtags with the highest degree, strength, and closeness centrality—characterizing number of different associations, total frequency of co-occurrences, and how closely hashtags are associated, respectively (26, 29, 30).

In group 1, concepts of "freedom" and "independence" are dramatically associated with "fight," "shame" against the Spanish government, "dictatorship," and blame against "police violence."

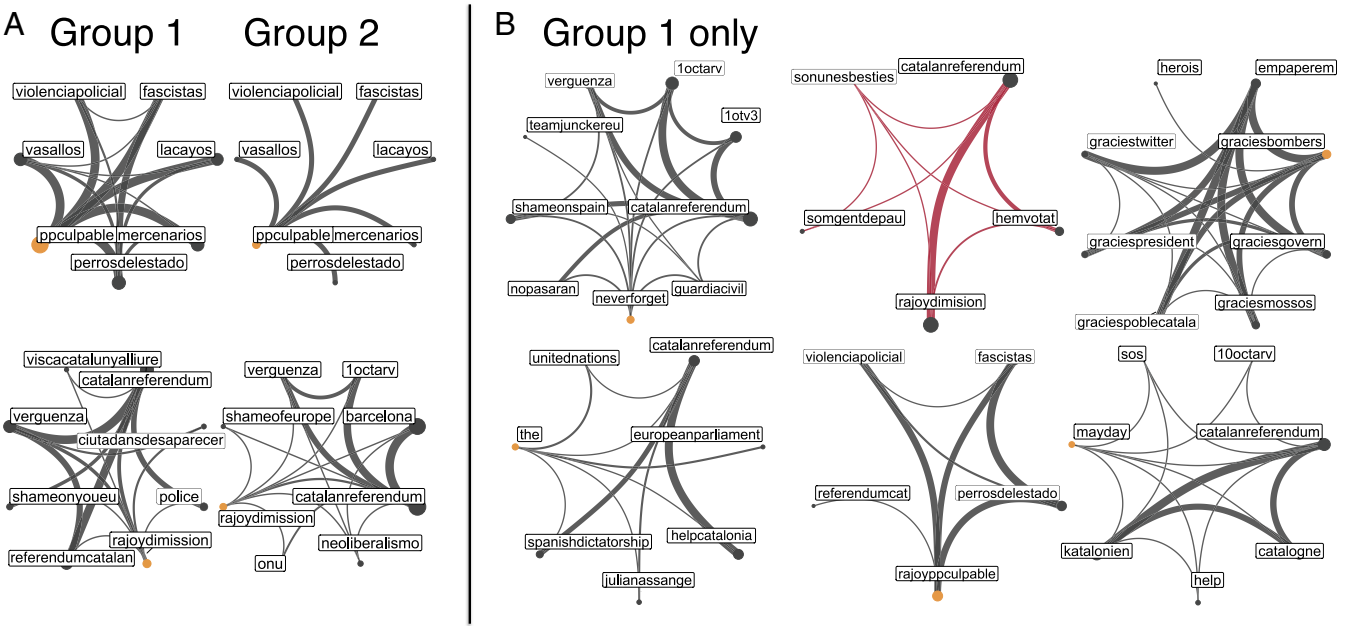

Fig. 4. Hashtag ecosystem reveals group identity. Hashtags are coupled together if they appear simultaneously in a message, building a network of concepts. Analyzing the hashtag networks obtained from each group, we identify the hashtags which are ranked similarly (*A*) and very differently (*B*) in the two groups to visualize the corresponding neighboring concepts. In *A*, low-ranked hashtags coexist in both groups and do not allow us to identify the underlying ideology of each group. In *B*, top-ranked hashtag that exist only in group 1 are strongly related to concepts of freedom, independence, fight, shame against the Spanish government, dictatorship, and blame against police violence, providing evidence that group 1 consists of Catalan Independentists. Remarkably, concepts related to "sonunesbesties" (translated as "they are beasts")—highlighted in *B*—are posted by bots only, whereas the other hashtag networks have contributions by both humans and bots. Note that, for clarity, in *B* we show only hashtags fully characterizing accounts associated with Independentists.

In group 2, these associations are completely missing, providing strong quantitative evidence that group 1 consists of Independentists. By combining this finding with the analysis shown in Fig. 3, which highlights the existence of only two groups, we deduce that group 2 consists of Constitutionalists and non-Independentists. We further distinguish between associations coming from bots and humans. Negative associations for the content of group 1 come exclusively from bots, highlighting their intent of bolstering conflict.

To enrich the results provided by our data-driven sentiment and network analyses, we perform human coding of 2,413 tweets posted by humans and social bots (*SI Appendix*). The analysis confirms the trends of sentiment polarities for human-to-human interactions, with shared content becoming increasingly pessimistic as a reaction to the violence registered on the onset of the referendum day. Moreover, human coding of the content of automated tweets confirms that bots mainly promoted news-media titles from hubs, mimicking the trend of human emotions and hence boosting sentiments of alarmism, fear, and reprobation during and after the vote.

Discussion

Through the synergy of cutting-edge techniques in bot detection, multilanguage sentiment analysis, network partitioning, and semantic network analysis, we find strong evidence of two opposing factions during a large-scale voting event. We provide quantitative findings that the captured online trends in the dataset mirror meaningful events in the real world concerning the voting timeline. Harnessing the structure and the semantic content of social actions within a large-scale dataset, we identify factions as groups of people having opposite stances during the Catalan referendum of October 1, 2017 (i.e., Independentists and Constitutionalists). Our results demonstrate that bots sustain each faction from the periphery of the online social network structure by mainly targeting human influencers. Bots tend to target human Independentists with messages evoking negative sentiments and associating hashtags with negative connotations. Importantly, we show that bots provide semantic associations, in messages directed to the Independentists, that inspire fight, violence, and shame against the government and the police. In addition to promoting target-specific content generated by human hubs, social bots achieved social contagion also by fabricating automated content within specific communities of humans. The negative associations highlighted in Fig. 4 were found only in endorsements relative to group 1 and were completely absent in messages within group 2. The specificity of such hatred-inspiring semantic associations provides evidence that bots achieved different social contagion across the groups also by forging artificial content.

While software-controlled agents might be beneficial to online networked systems, [e.g., by improving the collective performance of human groups (31)], their improper use can have dramatic effects. Our findings support the hypothesis that bots may influence information diffusion in social media systems (18, 19), specifically by accentuating the exposure to negative, hatred-inspiring, inflammatory content, thus exacerbating social conflict online. This concerning trend, coupled with the emerging ability to control time-varying networks such as online social systems (32), further motivates the crucial need for the development of quantitative techniques like the one proposed here for unmasking the social manipulation enacted by bots.

Materials and Methods

Data Collection. By following a consolidated strategy, we manually selected a set of hashtags and keywords to collect messages (tweets) posted to a microblogging platform (Twitter). The list contains various general Catalan issue-related terms: *#Catalunya*, *#Catalonia*, *#Catalogna*, *#1Oct*, *#votarem*, *#referendum*, and *#1O*. We monitored the Twitter stream and collected data by using the Twitter Search application programming interface (API), from September 22, 2017, to past the election day, on October 3, 2017: This allowed us to almost uninterruptedly collect all tweets containing any of the search terms. The data-collection infrastructure ran inside Fondazione Bruno Kessler servers to ensure resilience and scalability. We chose to use the Twitter Search API to make sure that we obtained all tweets that contain the search terms of interest posted during the data-collection period, rather than a sample of unfiltered tweets: This precaution avoided incurring known sampling issues related to collecting data by using the Twitter Stream API rather than the Search API. This procedure yielded a large dataset containing ∼3.6 million unique tweets, posted by 523,000 unique users.

Bot Detection. Various strategies exist to label social media users as bots or humans (19, 20). Here, we leveraged a scalable and accurate feature-based approach (33). Account metadata carry a highly predictive bot signature: We thus identified the top 10 most informative account metadata features (see *SI Appendix* for details). Off-the-shelf learning models were trained on multiple historical ground truth datasets and achieved high detection accuracy (>90%) on cross-validation benchmarks. Logistic regression (LR), our reference model for this study, was selected for its best trade-off between scalability and accuracy: The model is very precise at detecting human accounts—precision rate (PR) 98%, compared with bot accounts (PR: 92%), while detecting nearly all existing bots—recall rate (RR) 99%, compared with human users retrieval (RR: 88%). Furthermore, LR provides binary classifications, rather than continuous probabilistic scores—for example, like Botometer does (20)—simplifying the interpretability of resulting annotations without hampering classification accuracy. Finally, random samples of inferred bot and human labels were manually scrutinized for a sanity check. All bot-detection methods have some inherent limits (e.g., dependency on quality and size of training data and model generalizability) that we mitigated by using domain knowledge and state-of-the-art techniques (see *SI Appendix* for additional discussion).

Building the Twitter Network. People from the same faction tend to retweet each other as a form of social endorsement, as documented in the relevant literature (23), while cross-faction retweets are less likely. Considering that, only retweets would pose the question of how to get rid of spurious or infrequent interactions, possibly by identifying a given retweet threshold. Identifying a threshold would be problematic, as the final network structure might greatly vary with small perturbations on the considered threshold, as can happen on co-occurrence networks (34). We address this issue by considering strong social interactions: Twitter interactions where users perform at least one retweet but also at least another type of Twitter interaction, be it a mention or a reply during the considered time window. Notice that mentions and replies do not express the same social endorsement of retweets, but they can help in identifying the core interactions in the considered social system. The resulting Twitter Core Network (TCN) included 12,000 users, and 16,000 directed strong social interactions. Notice that the TCN aggregates interactions happening over the whole considered time window. However, the frequency of Twitter interactions strongly correlates with the indegree on the TCN (Kendall tau $\tau = 0.81$), thus indicating that the aggregated network topology is a valid proxy for investigating patterns of Twitter interactions.

PageRank Centrality of Humans and Bots. In the TCN, we used the average PageRank (25) as a measure of centrality of human and bot users quantifying how important individual nodes are for information flow in a given network topology. We computed PageRank centrality in Mathematica, which provides normalized values indicating the probability of a random walker to visit a given node. We used 0.85 as dampening factor, as in Google's PageRank. On average, human users displayed a PageRank of 8.1×10^{-4}, while bots displayed an average PageRank of 4.6×10^{-4}. Hence, on average, human users tended to be almost 1.8 times more central than bots in terms of information flow on the TCN.

Network Partitioning. To detect the two groups in the TCN, we used the Fiedler vector, a widely used heuristic in spectral graph partitioning (35). The Fiedler vector of a given graph is the eigenvector corresponding to the smallest nonzero eigenvalue (i.e., the algebraic connectivity) of the Laplacian matrix $L = D - A$ of the graph represented by the adjacency matrix A and by the diagonal matrix D. Negative and positive entries in the Fiedler vector partition the corresponding network nodes in two sets. One can prove analytically that this heuristic for graph partitioning is a valid approximation for solving the minimum cut problem on general

graphs (i.e., partitioning nodes in two groups so that the number of edges across groups is minimized). We applied spectral clustering on the undirected version of the TCN and then built randomized partitions. Through direct sampling, we show that the modularity of the Fiedler's partitioning is optimal compared with randomizations, even on the original TCN (*SI Appendix*).

Building the Hashtag Co-Occurrence Network. Hashtags are strings of characters starting with the hash (#) character and representing the main semantic content of a tweet (36). The literal meaning of hashtags is already considered in the sentiment analysis. Co-occurrence of different hashtags can provide important additional information on the semantic content of tweets, as it was recently shown (37). Analogously to other association networks in psycholinguistics (27, 28), networks of hashtag co-occurrences represent a powerful proxy of the cognitive profile of users (i.e., the way concepts are perceived and associated by users). From our Twitter dataset, we build semantic networks of hashtag co-occurrence where nodes represent hashtags and they are linked when co-occurring in at least one tweet. This network definition is in agreement with previous large-scale studies (37). We build one network of hashtag co-occurrences per group. The group 1 (group 2) co-occurrence network includes 8,451 (7,107) unique hashtags and 29,694 (23,644) links. The two networks overlap for 4,132 hashtags, on which the consistency analysis is performed (*SI Appendix*).

ACKNOWLEDGMENTS. E.F. was supported by Air Force Office of Scientific Research Award FA9550-17-1-0327.

1. Sawyer RK (2005) *Social Emergence: Societies as Complex Systems* (Cambridge Univ Press, Cambridge, UK).
2. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1:143–186.
3. Travers J, Milgram S (1969) An experimental study of the small world problem. *Sociometry* 32:425–443.
4. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329:1194–1197.
5. Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science* 334:1269–1272.
6. Shirado H, Fu F, Fowler JH, Christakis NA (2013) Quality versus quantity of social ties in experimental cooperative networks. *Nat Commun* 4:2814.
7. Lazer D, et al. (2009) Life in the network: The coming age of computational social science. *Science* 323:721.
8. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci USA* 106:21544–21549.
9. Onnela JP, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci USA* 107:18375–18380.
10. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337:337–341.
11. Bond RM, et al. (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489:295–298.
12. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci USA* 111:8788–8790.
13. Conover M, et al. (2011) Political polarization on Twitter. *The International Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), pp 89–96.
14. Lee JK, Choi J, Kim C, Kim Y (2014) Social media, network heterogeneity, and opinion polarization. *J Commun* 64:702–722.
15. Cruz FL, Troyano JA, Pontes B, Ortega FJ (2014) Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Syst Appl* 41:5984–5994.
16. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. *Sci Rep* 1:197.
17. Borge-Holthoefer J, et al. (2016) The dynamics of information-driven coordination phenomena: A transfer entropy analysis. *Sci Adv* 2:e1501158.
18. Bessi A, Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. *First Monday*, 10.5210/fm.v21i11.7090.
19. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Commun ACM* 59:96–104.
20. Varol O, Ferrara E, Davis C, Menczer F, Flammini A (2017) Online human-bot interactions: Detection, estimation, and characterization. *The International Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), pp 280–289.
21. Anderson ML (2003) Embodied cognition: A field guide. *Artif Intell* 149:91–130.
22. Taulé M, et al. (2017) Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017. *IberEval 2017*, Vol 1881, pp 157–177.
23. Metaxas PT, et al. (2015) What do retweets indicate? Results from user survey and meta-review of research. *International Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), pp 658–661.
24. Ding CH, He X, Zha H, Gu M, Simon HD (2001) A min-max cut algorithm for graph partitioning and data clustering. *IEEE International Conference on Data Mining* (IEEE, Piscataway, NJ), pp 107–114.
25. Brin S, Page L (2012) Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput Networks* 56:3825–3833.
26. Baronchelli A, Ferrer-i Cancho R, Pastor-Satorras R, Chater N, Christiansen MH (2013) Networks in cognitive science. *Trends Cognitive Sci* 17:348–360.
27. Kenett YN, Anaki D, Faust M (2014) Investigating the structure of semantic networks in low and high creative persons. *Front Hum Neurosci* 8:407.
28. Kenett YN, et al. (2018) Flexibility of thought in high creative individuals represented by percolation analysis. *Proc Natl Acad Sci USA* 115:867–872.
29. Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cogn Sci* 29:41–78.
30. Stella M, Beckage NM, Brede M (2017) Multiplex lexical networks reveal patterns in early word acquisition in children. *Sci Rep* 7:46730.
31. Shirado H, Christakis NA (2017) Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* 545:370–374.
32. Li A, Cornelius SP, Liu YY, Wang L, Barabási AL (2017) The fundamental advantages of temporal networks. *Science* 358:1042–1046.
33. Ferrara E (2017) Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 10.5210/fm.v22i8.8005.
34. Ninio A (2014) Syntactic networks, do they contribute valid information on syntactic development in children? *Phys Life Rev* 11:632–634.
35. Newman M (2010) *Networks: An Introduction* (Oxford Univ Press, Oxford).
36. Tsur O, Rappoport A (2012) What's in a hashtag?: Content based prediction of the spread of ideas in microblogging communities. *WSDM* (ACM, New York), pp 643–652.
37. Wang X, Wei F, Liu X, Zhou M, Zhang M (2011) Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. *ACM International Conference on Information and Knowledge Management CIKM* (ACM, New York), pp 1031–1040.