The Advantage of the Right in Social Media News Sharing

Sandra González-Bailón[1,*], Valeria d'Andrea [2], Deen Freelon [3], and Manlio De Domenico [2,*]

[1] Annenberg School for Communication, University of Pennsylvania, 19104 Philadelphia, PA, US

[2] Complex Multilayer Networks Lab, Center for Digital Society, Fondazione Bruno Kessler, 38123 Trento, Italy

[3] Hussman School of Journalism and Media, University of North Carolina at Chapel Hill, 27514 Chapel Hill, NC, US

* Corresponding authors: Sandra González-Bailón and Manlio De Domenico

Emails: sgonzalezbailon@asc.upenn.edu; mdedomenico@fbk.eu

Classification

Social Sciences, Computer Sciences.

Keywords

News sharing; misinformation; online networks; political mobilization; computational social science.

Author Contributions

Research design: SGB, VA, DF, MD. Data collection and pre-processing: SGB, VA, MD. Data analysis: SGB, VA, MD. Writing: SGB, VA, DF, MD.

This PDF file includes:

Main Text.
Figures 1 to 4.

Abstract

We analyze social media activity during one of the largest protest mobilizations in U.S. history to examine ideological asymmetries in the posting of news content. We extract the list of URLs shared on social media during the mobilization period and we characterize those web sites in terms of their audience reach and the ideological composition of that audience. We also analyze the reliability of the sites in terms of the credibility and transparency of the information they publish. We show that there is no evidence of unreliable sources having any prominent visibility during the protest period, but we do identify asymmetries in the ideological slant of the sources shared, with a clear bias towards right-leaning domains. These results support the “amplification of the right” thesis, which points to the structural conditions (social and technological) that lead to higher visibility of content with a partisan bent towards the right. Our results suggest that online networks are contested spaces where the activism of progressive movements coexists with the narratives of mainstream media, which gain visibility under the same stream of information (or hashtags) but whose reporting is not necessarily aligned with the activists’ goals.

Significance Statement

Existing research suggests that left- and right-wing activists use different media to achieve their political goals: the former operate on social media through hashtag activism; the latter partner with partisan outlets. However, legacy and digital media are not parallel universes. Sharing mainstream news in social media offers one prominent conduit for content spillover across channels. We analyze news sharing during a historically massive racial justice mobilization and show that misinformation posed no challenge to the coverage of these events. However, links to outlets with a partisan bent towards the right were shared more frequently, which suggests that right-leaning outlets have higher reach even within the confines of online activist networks built to enact change and oppose dominant ideologies.

Main Text

On June 6 2020, people across the U.S. joined one of the largest mobilizations in the country’s protest history (1). The immediate trigger was the killing of George Floyd while in police custody, but the scale of the mobilizations reflected years of organizing by a decentralized political movement seeking to end police brutality and advocate for criminal justice reform (2). The movement, initiated in 2014 around the #BlackLivesMatter (BLM) hashtag, has become a global symbol of social justice, and a prominent example of a new form of digital activism that defies old forms of organizing. Online networks, and the flows of information they allow, are the backbone of this type of mobilization: these networks help activists create alternative spaces in which to articulate discourses that are excluded for mainstream media (3). Of course, these tools are also available to other types of activism, including less progressive forces. Recent scholarship has started to pay attention to the asymmetries that characterize different forms of digital activism (4), but there are still many unknowns about how different actors use online tools as part of their repertoire, or how delimited different publics are online (5).

One idea that is receiving increasing empirical support is that the political left and the political right use media in different ways. Online media on the left and the right face different incentive structures and constraints that lead to different architectures and susceptibility to propaganda (6). This phenomenon, known as ideological asymmetry, manifests in a variety of ways. Adherents of the left, the argument goes, tend to consume more mainstream media (7), trust fact-checkers more (8), and curate more diverse personalized information environments (9) than right-wingers. The left engages heavily in hashtag activism --the use of social media hashtags to brand a political cause both on- and offline-- but the lack of research on right-wing hashtag activism makes ideological comparisons in this area difficult (4). People on the right tend to trust mainstream media less (10), spread more false news (6, 11), and tolerate the spreading of misinformation by politicians more than the left (12). Right-wing distrust of both the mainstream media and the set of internet platforms known as "Big Tech" has led followers to opt in to the "right-wing media ecosystem" of more congenial outlets (7) as well as "alt-tech" social media that offer more permissive terms of use (13).

While existing work on ideological asymmetry tells us much about how the left and right differ in terms of their distinctive styles of online political engagement, less is known about what happens when the two sets of tactics collide. Social media are sites of political contestation, and individuals on opposite sides of the same issue often clash directly and indirectly for attention, resources, and ideological converts. A key question here is: when left and right clash online, who is more successful in spreading their message? On the one hand, proponents of what might be called the "advantage of the right" thesis point to the structural conditions that allow better message production and dissemination – conditions that take the form of money and free time (14) or, in the digital realm, algorithmic amplification, which gives right-leaning content more visibility (15); and they also point to the greater audience susceptibility and engagement with moralized content (16). On the other hand, those predicting greater prevalence and circulation of left-wing messages can point to the fact that left-leaning users outnumber the right on Twitter (17, 18), and that left-wing hashtag activism campaigns such as #Blacklivesmatter and #MeToo have gained massive success (2, 3). These two realities co-exist online. For example, in the early days of the Black Lives Matter movement (2014-15), left-wing, anti-police brutality voices far outnumbered the pro-police right on Twitter (2). But on the issue of mass shootings, messages supporting gun rights on Twitter are more numerous than those advocating gun control (19). So far, little if any research has investigated the question of how such ideological contests shape the visibility of political messages on social media. Here we address this question in the context of the Black Lives Matter movement.

Our analyses consider three outstanding questions regarding ideological asymmetries in contested online spaces: First, whose messages reach the most people? Second, which ideological side consumes the most low-quality content (i.e., misinformation)? And third, how common in the conversation is low-quality content in terms of prevalence and consumption? Our results support the "amplification of the right" thesis, finding that content with a right-leaning partisan slant is viewed and shared substantially more than left-leaning content. This gives right-leaning outlets an advantage to the extent that they gain higher visibility and levels of engagement, both crucial in the attention economy. Unreliable content, on the other hand, is very infrequent and low in visibility overall: only a very small community of Twitter users share unreliable sources. But to the extent that right-leaning sources are, on the aggregate, more visible, our results also suggest that activist networks are seeded with messages that are not necessarily aligned with their framing of events, limiting the impact that online activism can have on public discourse.

Data

Our data come from four different sources: Twitter's API; records of offline protest events; web tracking data; and reliability ratings for news websites. We use Twitter data to reconstruct social media activity around the June 6 2020 mobilizations, and protest event data to build a benchmark to compare online activity with offline actions. Figure 1 shows a comparison of the timeline and location of protest events according to the Crowd Counting Consortium (20) with the timeline and location of the tweets we analyze. There are obvious parallels both in terms of volume (the peak in online activity coincides with the peak of protest events on June 6) and in terms of geographic distribution in the contiguous U.S. [See the SI for more details on the data and alternative counts of offline protest events]. To differentiate different types of Twitter users, we classify accounts in three groups: media accounts, bots, and human accounts. This classification relies on bot detection techniques (see Materials and Methods and the SI for technical details). Building on prior work (21, 22), we used the label "media" as a shorthand to refer to accounts with bot-like behavior that are also verified by Twitter. These include the accounts of public figures, journalists, or news organizations. We labelled as "bots" automated accounts that are not verified by Twitter, and the rest are classified as human accounts. As Figure 1C shows, most messages are generated by human accounts, but a large fraction of the total volume is generated by unverified bots. Less than 1% of all accounts fall in the "media" category. As we show in the SI, this categorization also reveals other expected differences (for instance, verified accounts post more reliable content, see figure SI15).

We parsed all the tweets to extract URLs, when present, and we identified their registered domain. This yielded a list of $N =$ 2,176 unique domains. We matched these domains with web browsing data collected during the same period from a representative panel of the U.S. population (see Materials and Methods). This web panel allowed us to obtain measures of audience reach (i.e., the fraction of the online population accessing the domains) and the ideological composition of these audiences (i.e., the fraction of users accessing the domain that self-identify as Republican or Democrat, or as Conservative or Liberal). We use party identification and political outlook to compute two variables of ideological slant, one derived from partisan sorting and the other from ideological differences in the audiences consuming those domains. Individuals who disclose the same partisan affiliation may have similar policy attitudes but different core political values, i.e., different predispositions to accept change (something that is well documented in the political science literature, e.g., 23). By looking at the composition of domain audiences in these two dimensions (party and outlook) we are differentiating domains that are favored by different types of people. Our assumption is that audience composition tells us something about the slant of the coverage provided by the domains (i.e., news sources). This is an assumption that we share with prior research (e.g., 24, 25-27) but that may yield different classifications compared to content-based or editorial measures of ideological bias (e, g., those provided by AdFontes or AllSides). With these measures we compute ideological scores that we assign to domains and to the Twitter users posting those domains (see Materials and Methods and the SI for more details).

We also matched the domains with reliability scores that rate the credibility of news and information websites. Sites receive a trust score on a 0-100 scale based on criteria related to the credibility of the information published and the transparency of the sources (see Materials and Methods and the SI for more details).

Results

Most URLs shared during the protest mobilizations lead to 'News/Information' domains. As Figure 2A shows, this is by far the most popular category (see table SI1 for examples of top domains within these groups). Within the news category, local news prevail (panel A inset). Panel B plots the distributions of the ideology measures for all domains. Most of the sites have audiences that include roughly the same number of Democrats and Republicans, but there is a clear bent towards more conservative audiences. Party and outlook scores are moderately correlated (rho ~ 0.5), which means that partisan sorting and ideological differences are far from a perfect match (see figure SI7 in the SI; the magnitude of the correlation falls in line with that drawn from survey data, 23). These distributions suggest that audiences that identify as Democrat identify, at the same time, with conservative values – which explains the different median values for the party and outlook scores: while the former is roughly centered around 0, the latter is shifted to the right. Determining which set of values are part of the conservative political outlook of these audiences goes beyond the scope of our paper, but we do find that most domains have audiences that lean conservative, and that more domains have extreme conservative scores than extreme liberal scores.

Panel C shows a moderate correlation between audience reach on the web and number of tweets pointing to the domains. Once we control for the audience reach of the websites and their classification in the 'News/Information' category, domains that lean Republican are more frequent both in terms of total counts and unique counts of URLs shared (panel D). Controlling for the audience partisanship of these domains, conservative slant actually has a negative impact on domain visibility. These results signal that the advantage of the right (in terms of visibility) has an upper limit that excludes the most conservative outlets (i.e., those favored by audiences who, according to their own placement in the conservative scale, are more extreme). Yet right-leaning outlets (in a partisan sense) are still the most visible, both in terms of total URLs and unique URLs count. This is a surprising finding given the liberal bias of the Twitter user base and the specific stream of information we analyze.

Figure 3 unpacks the ideological slant of the three more frequent categories within the 'News/Information' group: local news, general news, and politics (or partisan) domains (see figure SI3 in the SI for the top 30 domains within these categories). There is a clear shift towards the right of the distribution for both party and outlook as we move from local news, to general news, to political domains. There are no strong differences in ideological scores for the three types of accounts (media, bots, and humans), but verified media accounts tend to share political URLs that lean more clearly towards the right (see figure SI14 in the SI). This ideological asymmetry suggests that the stream of information around BLM hashtags (and the offline protest events) was punctuated by messages systematically drawing attention to content from right-leaning domains. Even if sharing URLs does not necessarily amount to endorsement, re-tweeting activity amplifies sources of information that have a clear ideological leaning.

Figure 4 shifts the focus of attention from domains as the unit of analysis to the users posting those domains. We built the retweet (RT) network and calculated centrality measures (see table SI2 and figure SI16 in the SI for descriptive statistics). Panels A and B show the results of regression models where the dependent variable is number of RTs received (or incoming strength centrality, log-transformed). Panel A shows that verified media accounts are the most central accounts [consistent with prior research (21)], but also that posting tweets that contain URLs is associated with an increase in centrality in the RT network. We assign ideological scores to users based on the domains they share. The most central amongst this subset of users (panel B) are again media accounts; controlling for this and other covariates (including number of URLs posted), users that post Republican-leaning URLs receive a higher number of RTs. Controlling for party identification, conservative-leaning URLs have, again, a negative impact on the probability of receiving RTs. We take this as additional evidence that the advantage of the right (in terms of visibility and engagement) has a limit: the most extreme domains (those accessed predominantly by the most conservative audiences) receive less traction. Yet the overall advantage of right-leaning sources (in a partisan sense and in terms of higher visibility) is still significant.

Panel 4C shows the RT network collapsed to its communities (the network has a high modularity score, Q = 0.88; see Materials and Methods for details on our community detection method). Each node represents a community, and the edges capture RTs among users in these communities. Node color encodes the average reliability score of the URLs shared within each community. As the plot shows, most content rates high in the reliability scale (60 or higher, see figure SI10 in the SI for the histogram of reliability scores and correlations with the two ideology scores). There is a cluster of communities that have lower average reliability, and one small community in this cluster clearly sharing unreliable sources (in red) but the users in these communities amount to a very small fraction of all users [the secluded nature of the users interacting with less reliable sources is also consistent with prior research (6)]. Panel D shows that this cluster of communities are on the right-leaning quadrangle of the ideological space.

A more detailed analysis of content shared in the top 10 communities in terms of size revels a division of the network in two sets of communities: the most popular user on one side of the divide is a conservative talk radio host; the most popular user on the other side of the divide is a civil rights attorney (see figures SI17 and SI18). The sparsity of RTs separating these two clusters is suggestive of a divide in the diffusion of #BLM messages in two distinct sets of communities; and yet, as we discuss in the SI, these two groups do not map onto the two halves of the ideological continuum: the average favorability scores are all above the 0 line (and, therefore, signaling a right-leaning slant).

Discussion

Social media have allowed users to create a public sphere where alternative voices can arise, and progressive movements organize as they frame political causes in their own terms. These spaces, however, also host opposing voices, including those of conservative actors and mainstream media gatekeepers. Here, we document the clash of perspectives that arose around the Black Lives Matter protests in 2020. Critically, we address the questions of who produces news coverage and how audiences respond to that content in the same social media stream used to organize the protests. Social media users consume much more content than they produce (28), which allows mainstream media and other content creators to have an influence on the platform.

Our analyses show that most of the news sources posted on social media as the massive street protests unfolded are produced by media with a right-leaning ideological slant (in partisan terms) and that this content generates more engagement in the form of RT activity, thus increasing its reach. Right-leaning media are unlikely to portray the protest events in the movement's terms, and their visibility in this stream of information is indicative of the ideological clash that characterize contested online spaces. At the same time, the visibility of content with a partisan bent towards the right has a limit that excludes the most conservative outlets (i.e., those favored by audiences that are, according to their own self-placement, more resistant to change). The most extreme outlets in the conservative dimension gain less visibility in this stream of protest mobilization. Put together, our results suggest that right-leaning domains do better (in terms of gaining visibility and engagement) than left-leaning domains. The right, in other words, has an advantage in the attention economy created by social media (with a limit that penalizes the most extreme conservative content).

Empirical results are always contingent on measurement. Future research should try to replicate our findings using different measures of ideology, based, for instance, on voter registration records for Twitter users (e.g., 6) or content-based measures of media bias (e.g., 29). A recent research paper using an experimental approach, different measures of ideology, and a different timeframe shows results that are consistent with our main claim that the right has an advantage in social media (30). This work suggests there is algorithmic amplification of the mainstream political right in 6 of 7 countries analyzed. Even though the main focus of analyses lies on elected legislators from major political parties, the paper also contains additional analysis for news content in the U.S. The results reported for news are less clear than the results reported for legislators, but the findings are still suggestive of amplification of right-leaning sources. Our approach to labeling the ideological slant of web domains is very different (we rely on the self-placement of audiences accessing those domains on the web, instead of content-based labels; and we use two ideology variables – partisanship and outlook -- instead of one) but our results showing the prevalence of right-leaning domains are consistent. This alignment suggests that the findings we report are not contingent on our data or research design choices. If anything, our empirical context (the BLM protests) offers a more stringent test to document the advantage (i.e., the increased visibility) of right-leaning outlets.

Together, these findings show that mainstream media can shape events even in the context of activist networks. Their prominence dilutes the power of activist networks, and it lengthens the shadow of what we call the "advantage of the right": a disproportionate tendency of right-wing media voices to gain visibility in ideologically diverse social media spaces. The dominance of right-leaning voices on social media has been documented anecdotally in journalistic accounts (e.g., 31, 32) but there is still a lack of systematic empirical evidence offering support to this claim. Here we provide that type of evidence, and proof that ideological asymmetries manifest even in the context of movements with progressive goals.

In sum, our work shows that on one of the most prominent political issues of the 21st century, the perspective of right-leaning outlets dominated on Twitter. As discussed, this is surprising in some respects, especially given the well-documented population advantage of liberals over conservatives on the platform (17, 18). Yet it accords with studies of the right-wing media ecosystem, which has developed as an alternative to more centrist mainstream media and regularly attracts mass attention on controversial issues (7, 33). The prevalence of right-wing media content about Black Lives Matter protests and protesters poses a challenge to the latter's attempts to set the media agenda and attract supporters. Future research should determine if the advantage of the right on display here also applies at other times and for other political issues; it should also aim to improve our measures and offer a more granular definition of what counts as misinformation or low-quality content (e.g., our domain-level measures mask heterogeneity in quality and bias at the news story level). Yet our results stand on their own as a demonstration of the prominence that right-leaning media have in the social media marketplace of ideas.

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 for the list of keywords). To benchmark online activity with the actions taking place on the streets across the country, we obtained protest event data from the Crowd Counting Consortium (20). The web-tracking data offering reach estimates and audience-based ideological scores comes from Comscore's Plan Metrix panel, and it covers the same period as the Twitter activity data (May-June 2020). Following prior work (24, 25), we assign ideology scores to these domains using the audience-based measures $fav_p(d) = \frac{(R-D)}{(R+D)}$ and $fav_o(d) = \frac{(C-L)}{(C+L)}$. These favorability scores equal -1 when a domain is visited exclusively by Democrats or Liberals and 1 when it is visited exclusively by Republicans or Conservatives. The calculations exclude panelists that self-identify as 'Independent' or 'Middle of the Road'. We used Comscore's classification of web domains when available (e.g., news/information, entertainment, etc), and we manually checked to solve inconsistencies, errors, and missing labels. See the SI for more details. Finally, the reliability scores come from NewsGuard, a journalism and technology company that rates the transparency and credibility of news and information websites. These scores are provided at the domain (source) level, which may mask unreliable information published in specific news stories. See SI for more details on the data, sources, and additional descriptive statistics.

Methods. We identify automated accounts using a bot classification technique trained and validated on publicly available datasets. Using 80% of the data for training and the remaining 20% for validation, the model achieves a classification accuracy of about 90%. When applied to an independent dataset to test out-of-domain performance, the classification accuracy decreases to 60%, which suggests the model can be generalized to new data but also that performance decreases with respect to training and validation sets (as is well known in the literature, see the SI for more details on the model and cross-validation checks). We build the weighted version of the retweet network and calculate centrality scores on the largest connected component (see table SI2 in the SI for descriptive statistics). We identify communities in the network using a random walk algorithm designed to identify dense subgraphs in sparse structures (34).

Models. The regression models have two main dependent variables (DV): the number of URLs shared for every domain (total count and unique count); and centrality in the RT network for every user (number of RTs received). For domains, the main control variables are audience reach on the web and category (binary attribute identifying the domains classified as "news"). The ideological scores of the domains are the main variables of interest. For users, the main control variables are number of followers and friends, RTs made, and account type (media, bot, or human). The main variable of interest is whether the user posted URLs and, if so, of which ideological slant. See SI for additional details and specifications.

Acknowledgments

Work on this paper was partly funded by NSF grants 1729412 and 2017655. The authors would like to thank Oriol Artime and Riccardo Gallotti for useful discussions, and Anna Bertani for technical support during data collection.

References

1. L. Buchanan, Q. Bui, J. K. Patel (2020) Black Lives Matter May Be the Largest Movement in U.S. History. in *The New York Times*.
2. D. Freelon, C. D. McIlwain, M. Clark (2016) Beyond the Hashtags: #Ferguson, #Blacklivesmatter, and the Online Struggle for Offline Justice. in *Center for Media & Social Impact* (American University).
3. S. J. Jackson, M. Bailey, B. Foucault Welles, *#HashtagActivism. Networks of Race and Gender Justice* (MIT Press, Cambridge, MA, 2020).
4. D. Freelon, A. Marwick, D. Kreiss, False equivalencies: Online activism from left to right. *Science* **369**, 1197 (2020).
5. S. Shugars *et al.*, Pandemics, Protests, and Publics: Demographic Activity and Engagement on Twitter in 2020. *Journal of Quantitative Description: Digital Media* **1** (2021).
6. 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).
7. Y. Benkler, R. Faris, H. Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press, Oxford, 2018).
8. J. Shin, K. Thorson, Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication* **67**, 233-255 (2017).
9. J. T. Jost, S. van der Linden, C. Panagopoulos, C. D. Hardin, Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology* **23**, 77-83 (2018).
10. R. R. Mourão, E. Thorson, W. Chen, S. M. Tham, Media Repertoires and News Trust During the Early Trump Administration. *Journalism Studies* **19**, 1945-1956 (2018).

11. A. Guess, J. Nagler, J. Tucker, Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* **5**, eaau4586 (2019).
12. J. De Keersmaecker, A. Roets, Is there an ideological asymmetry in the moral approval of spreading misinformation by politicians? *Personality and Individual Differences* **143**, 165-169 (2019).
13. R. Rogers, Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication* **35**, 213-229 (2020).
14. J. Schradie, *The Revolution That Wasn't: How Digital Activism Favors Conservatives* (Harvard University Press, 2019).
15. F. Huszár *et al.*, Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences* **119**, e2025334119 (2022).
16. W. J. Brady, J. A. Wills, D. Burkart, J. T. Jost, J. J. Van Bavel, An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General* **148**, 1802-1813 (2019).
17. D. Freelon (2019) Tweeting left, right, & center: How users and attention are distributed across Twitter. (John S. & James L. Knight Foundation, Miami, FL).
18. S. Wojcik, A. Hughes (2019) Sizing Up Twitter Users. (Pew Research Center).
19. Y. Zhang *et al.*, Whose Lives Matter? Mass Shootings and Social Media Discourses of Sympathy and Policy, 2012–2014. *Journal of Computer-Mediated Communication* **24**, 182-202 (2019).
20. J. Pressman, E. Chenoweth, Crowd Counting Consortium, https://sites.google.com/view/crowdcountingconsortium/view-download-the-data.
21. S. González-Bailón, M. De Domenico, Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences* **118**, e2013443118 (2021).
22. M. Stella, E. Ferrara, M. De Domenico, Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 10.1073/pnas.1803470115 (2018).
23. R. N. Lupton, S. M. Smallpage, A. M. Enders, Values and Political Predispositions in the Age of Polarization: Examining the Relationship between Partisanship and Ideology in the United States, 1988–2012. *British Journal of Political Science* 10.1017/S0007123417000370, 1-20 (2017).
24. M. Tyler, J. Grimmer, S. Iyengar, Partisan Enclaves and Information Bazaars: Mapping Selective Exposure to Online News. *Journal of Politics* **forthcoming** (2021).
25. T. Yang, S. Majó-Vázquez, R. K. Nielsen, S. González-Bailón, Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences* **117**, 28678-28683 (2020).
26. M. Gentzkow, J. M. Shapiro, Ideological Segregation Online and Offline. *The Quarterly Journal of Economics* **126**, 1799-1839 (2011).
27. A. M. Guess, (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science* **65**, 1007-1022 (2021).
28. A. Antelmi, D. Malandrino, V. Scarano (2019) Characterizing the Behavioral Evolution of Twitter Users and The Truth Behind the 90-9-1 Rule. in *Companion Proceedings of The 2019 World Wide Web Conference* (Association for Computing Machinery, San Francisco, USA), pp 1035–1038.

29. C. Budak, S. Goel, J. M. Rao, Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly* **80**, 250-271 (2016).
30. F. Huszár *et al.*, Algorithmic Amplification of Politics on Twitter. *arXiv* **arXiv:2110.11010** (2021).
31. A. Thompson (2020) Why the right wing has a massive advantage on Facebook. in *Politico*.
32. K. Roose (2020) What if Facebook Is the Real 'Silent Majority'? in *The New York Times*.
33. C. Pak, K. Cotter, J. DeCook, Intermedia Reliance and Sustainability of Emergent Media: A Large-Scale Analysis of American News Outlets' External Linking Behaviors. *International Journal of Communication* **14** (2020).
34. P. Pons, M. Latapy, Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications* **10**, 191-218 (2006).

Figure Captions

Figure 1. Temporal and Spatial Distribution of Protest Events and Tweets. The upper row shows the number (A) and location (B) of protests organized in the contiguous United States. The lower row shows the count (C) and location (D) of the tweets in our data set. Overall, Twitter activity reflects a similar temporal and spatial distribution to offline protest events, with June 6 being the day of greatest activity. Most of the tweets are generated by human accounts (~ 58% of all accounts), but unverified bots (~ 42%) generate a very large fraction of the total volume. Verified media accounts (< 1%) generate a very small fraction of messages. See Materials and Methods and the SI for more details on data sources and classifications.

Figure 2. URLs Shared on Twitter during the Protests. Most of the URLs go to news/information domains (panel A); within this category local news prevail (A inset). We assign ideological scores to these domains based on their audience composition in terms of party affiliation and political outlook (panel B, see Materials and Methods for details on calculations). The score equals -1 when a domain is visited exclusively by Democrats or Liberals and 1 when it is visited exclusively by Republicans or Conservatives (0 means that Republicans/Conservatives and Democrats/Liberals are equally likely to visit the domain). These distributions suggest that audiences that identify as Democrat identify, at the same time, with conservative values – which explains the different median values for the party and outlook scores. Panel C looks at the association between the audience reach of these domains on the web (i.e., the fraction of the U.S. online population accessing the domains during this period) and the number of tweets that contain URLs to those domains (note that only a few labels are shown to improve legibility). Domains in blue have a favorability score below the median (e.g., their audiences lean democrat/liberal) and domains in red have a favorability score equal or above the median (e.g., they lean republican/conservative). Panel D shows the results of linear models predicting domain visibility (measured as total URLs shared and unique URLs shared, both log-transformed, 99% CI). Web audience reach is the most important predictor of visibility on Twitter, and URLs pointing to News/Information sources are also more salient than non-news URLs. Controlling for these two variables, Republican-leaning URLs appear in more tweets, but Conservative-leaning URLs appear in less, suggesting an upper ceiling for the most extreme outlets on the Conservative right (see SI for regression tables and other specifications).

Figure 3. Ideology Distributions by Domain Sub-Category. The panels in this figure show ideology distributions for the three most frequent categories within the 'News/Information' group: 'local news' (A); 'general news' (B); and 'politics' (C) [See figure SI3 in the SI for the top 30 domains within these sub-categories]. The 'general news' and 'political news' domains shared during the protests have a clear right-leaning slant, both on terms of party and political outlook. The shift to the right tail of the ideological distributions is particularly clear for political domains.

Figure 4. Ideology and Reliability in the RT Network. Panels A and B show the results of linear models predicting the number of RTs received by user accounts (log-transformed, 99% CI). Accounts posting URLs have a higher centrality in the RT network. Accounts posting URLs to Republican-leaning domains receive more RTs, but those posting URLs to conservative-leaning domains receive less (see SI for regression tables and other specifications). Panel C shows the RT network collapsed to the ~280 communities identified by a random-walk algorithm (34). The network is very modular (Q = 0.88) and each community represents a group of accounts that

retweet each other more frequently than other accounts. Color encodes the reliability score assigned to each community based on the URLs shared. Panel D shows the same communities as they fall in the two-dimensional space defined by the ideological scores (the red, dashed lines mark the medians of the distributions in Figure 2A). Most communities share URLs to reliable content, even those on the extremes of the ideological distributions. Communities with users sharing less reliable sources are in the right-leaning quadrants of the ideology distributions.

Figures

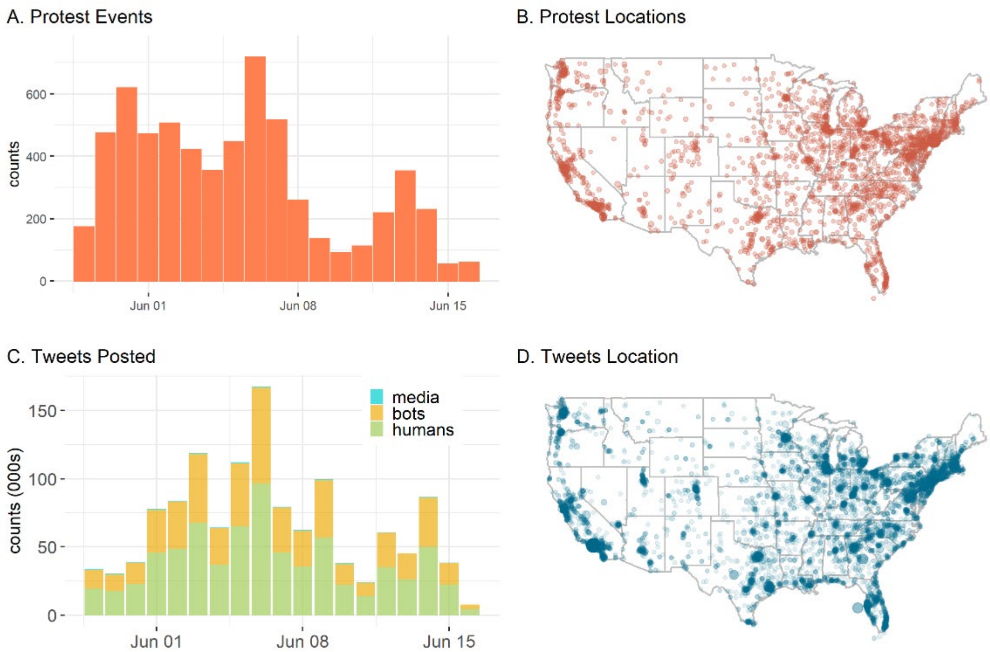

Figure 1. Temporal and Spatial Distribution of Protest Events and Tweets. The upper row shows the number (A) and location (B) of protests organized in the contiguous United States. The lower row shows the count (C) and location (D) of the tweets in our data set. Overall, Twitter activity reflects a similar temporal and spatial distribution to offline protest events, with June 6 being the day of greatest activity. Most of the tweets are generated by human accounts (~ 58% of all accounts), but unverified bots (~ 42%) generate a very large fraction of the total volume. Verified media accounts (< 1%) generate a very small fraction of messages. See Materials and Methods and the SI for more details on data sources and classifications.

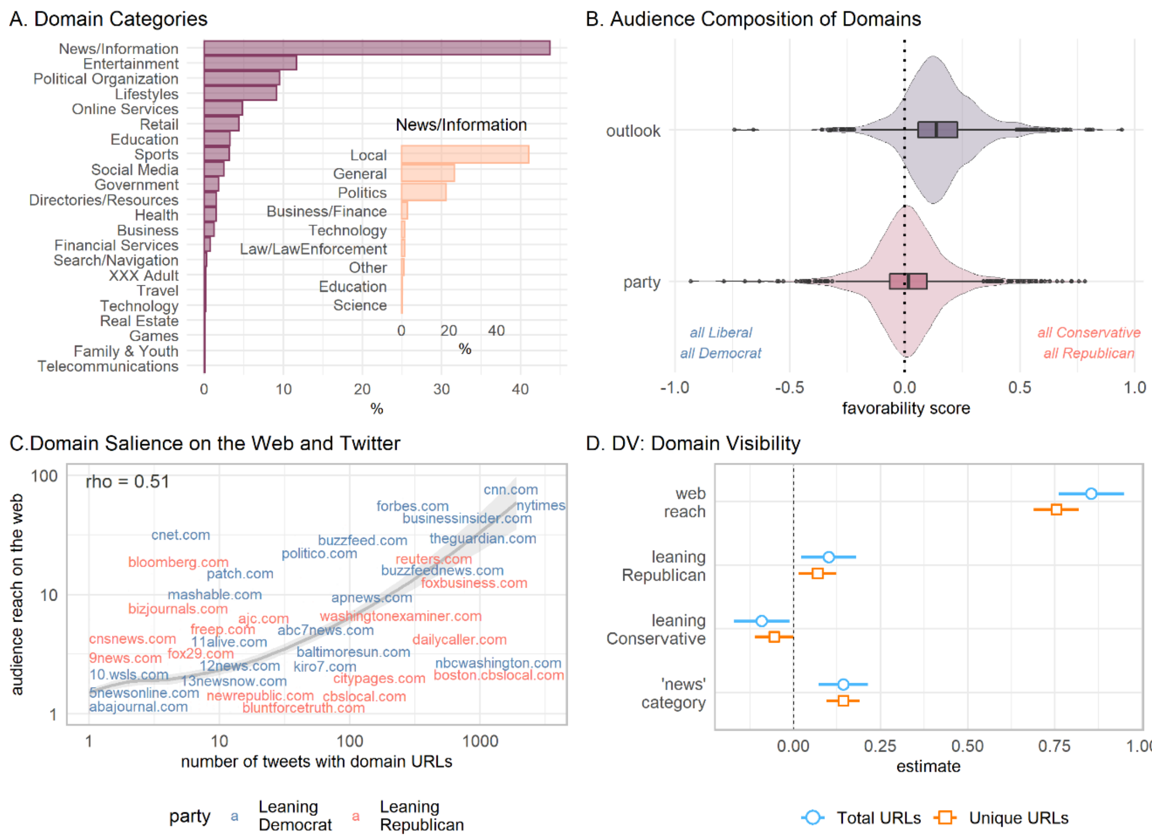

Figure 2. URLs Shared on Twitter during the Protests. Most of the URLs go to news/information domains (panel A); within this category local news prevail (A inset). We assign ideological scores to these domains based on their audience composition in terms of party affiliation and political outlook (panel B, see Materials and Methods for details on calculations). The score equals -1 when a domain is visited exclusively by Democrats or Liberals and 1 when it is visited exclusively by Republicans or Conservatives (0 means that Republicans/Conservatives and Democrats/Liberals are equally likely to visit the domain). These distributions suggest that audiences that identify as Democrat identify, at the same time, with conservative values – which explains the different median values for the party and outlook scores. Panel C looks at the association between the audience reach of these domains on the web (i.e., the fraction of the U.S. online population accessing the domains during this period) and the number of tweets that contain URLs to those domains (note that only a few labels are shown to improve legibility). Domains in blue have a favorability score below the median (e.g., their audiences lean democrat/liberal) and domains in red have a favorability score equal or above the median (e.g., they lean republican/conservative). Panel D shows the results of linear models predicting domain visibility (measured as total URLs shared and unique URLs shared, both log-transformed, 99% CI). Web audience reach is the most important predictor of visibility on Twitter, and URLs pointing to News/Information sources are also more salient than non-news URLs. Controlling for these two variables, Republican-leaning URLs appear in more tweets, but Conservative-leaning URLs appear in less, suggesting an upper ceiling for the most extreme outlets on the Conservative right (see SI for regression tables and other specifications).

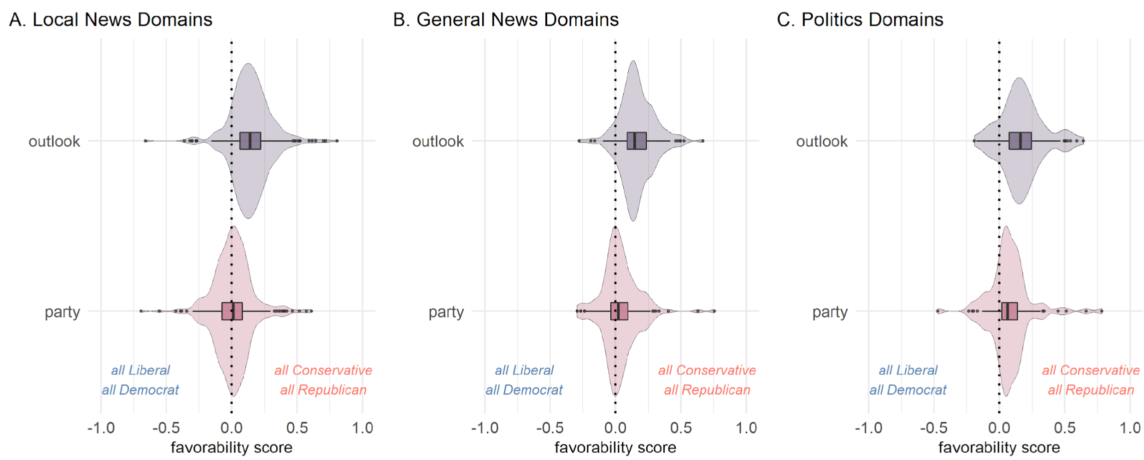

Figure 3. Ideology Distributions by Domain Sub-Category. The panels in this figure show ideology distributions for the three most frequent categories within the 'News/Information' group: 'local news' (A); 'general news' (B); and 'politics' (C) [See figure SI3 in the SI for the top 30 domains within these sub-categories]. The 'general news' and 'political news' domains shared during the protests have a clear right-leaning slant, both on terms of party and political outlook. The shift to the right tail of the ideological distributions is particularly clear for political domains.

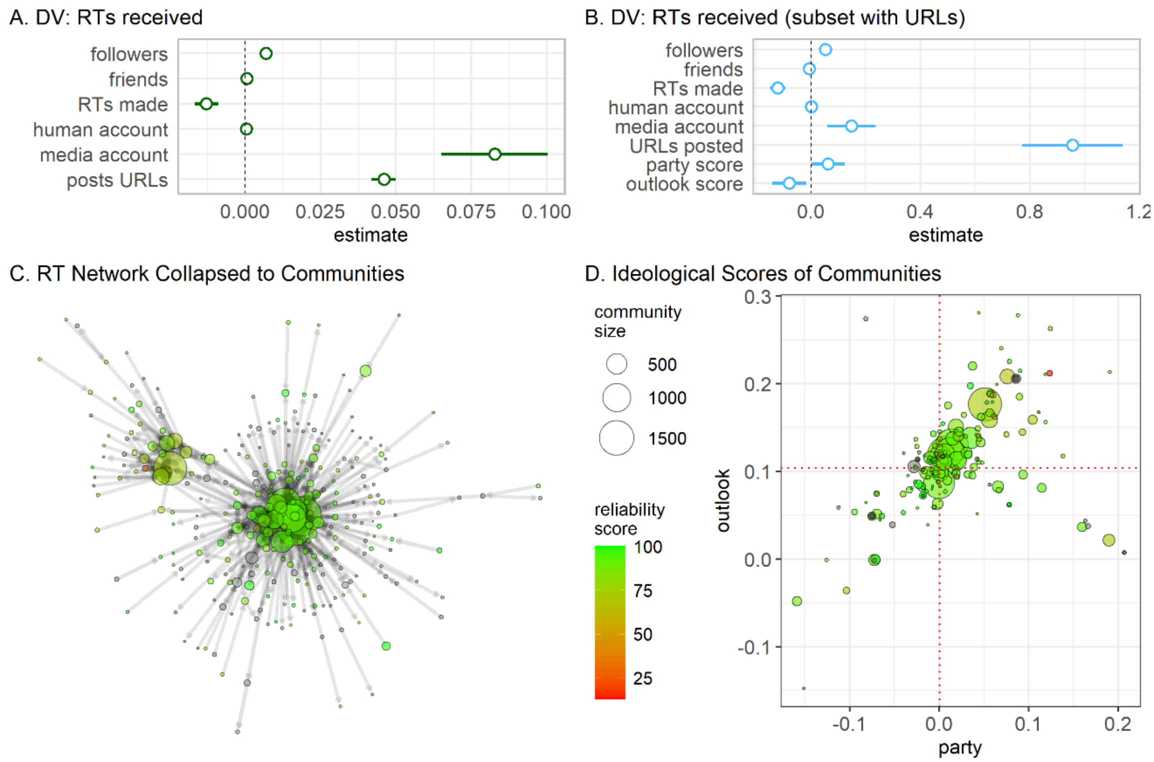

Figure 4. Ideology and Reliability in the RT Network. Panels A and B show the results of linear models predicting the number of RTs received by user accounts (log-transformed, 99% CI). Accounts posting URLs have a higher centrality in the RT network. Accounts posting URLs to Republican-leaning domains receive more RTs, but those posting URLs to conservative-leaning domains receive less (see SI for regression tables and other specifications). Panel C shows the RT network collapsed to the ~280 communities identified by a random-walk algorithm (34). The network is very modular (*Q* = 0.88) and each community represents a group of accounts that retweet each other more frequently than other accounts. Color encodes the reliability score assigned to each community based on the URLs shared. Panel D shows the same communities as they fall in the two-dimensional space defined by the ideological scores (the red, dashed lines mark the medians of the distributions in Figure 2A). Most communities share URLs to reliable content, even those on the extremes of the ideological distributions. Communities with users sharing less reliable sources are in the right-leaning quadrants of the ideology distributions.

Supplementary Information

The Advantage of the Right in Social Media News Sharing

Sandra González-Bailón[1, *], Valeria d'Andrea [2], Deen Freelon [3], and Manlio De Domenico [2, *]

[1] Annenberg School for Communication, University of Pennsylvania, 19104 Philadelphia, PA, US

[2] Complex Multilayer Networks Lab, Center for Digital Society, Fondazione Bruno Kessler, 38123 Trento, Italy

[3] Hussman School of Journalism and Media, University of North Carolina at Chapel Hill, 27514 Chapel Hill, NC, US

* Corresponding authors: Sandra González-Bailón and Manlio De Domenico

Emails: sgonzalezbailon@asc.upenn.edu; mdedomenico@fbk.eu

This PDF file includes:

Sections

Figures

Tables

1. Data

We analyze and combine data from four sources: (1) Twitter; (2) records of offline protest events; (3) web tracking (including information on the ideological composition of audiences accessing websites); and (4) reliability ratings for news websites.

(1) The Twitter data was collected using both the search and stream APIs for keywords related to the protests (e.g., #Black_Lives_Matter, #BlackLivesMatter, #GeorgeFloydProtests, #GeorgeFloyd). In total, the data consists of ~ 52 million tweets (all languages and locations). We restricted our analyses to tweets written in English and produced from the U.S. during the time period May 28 2020 to June 16 2020 (20 days). Figure SI1 plots the volume of Tweets worldwide, excluding the U.S. The geographic filter relies on the location and lat/long information contained in the tweets metadata. We focus on tweets published in the U.S. to be able to match URL sharing activity with web browsing data, which is based on a panel that is representative of the internet population in the U.S. (a more thorough description of the web log data can be found below). Overall, the filtered dataset we analyze is formed by *N* ~ 1.3 million unique tweets. In section 2 we show that this filter does not substantially distort patterns of URL sharing, compared to domain sharing activity that includes all English tweets (regardless of location).

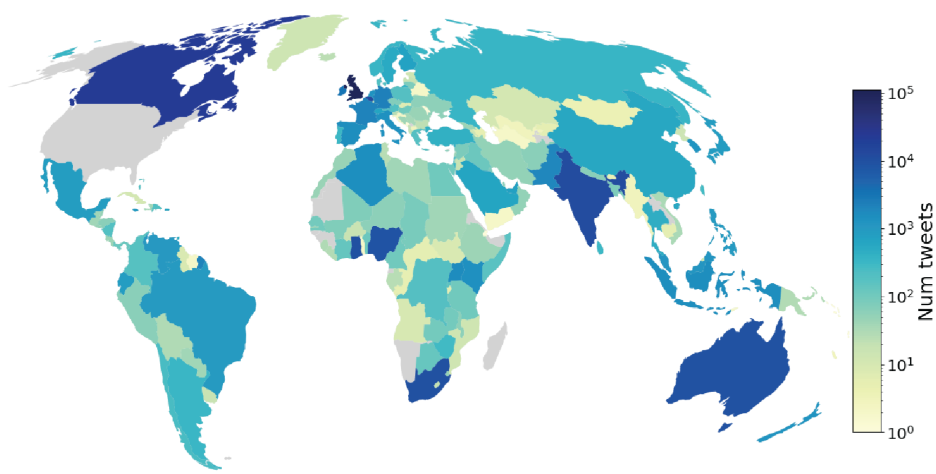

Figure SI1. Worldwide Spatial Distribution of BLMs Twitter Activity. This map excludes activity in the U.S. (which we analyze in the main paper). The audience-based ideological attributes assigned to the domains shared on Twitter derive from a web panel that is representative of the U.S. online population; the measures of ideology and partisanship we use are not meaningful in political contexts other than the U.S. -- hence our decision to apply the geographical filter to our main analyses. See figure SI12 for tests that show this filter does not substantially distort observed patterns.

(2) We use two publicly available datasets of offline protest events to compare with online activity. The first dataset was compiled by the Crowd Counting Consortium (1) and it encompasses data on political crowds reported in the U.S. through news stories as well as individual reports, including marches, protests, strikes, demonstrations, riots, and other actions. The second dataset was compiled by Count Love (2) through the automated daily crawl of local newspaper and television sites and the count of public displays of protest. Both data sources have been used in journalistic coverage of the protests (3, 4). Figure SI2 shows the temporal and spatial patterns of protest activity according to these two sources.

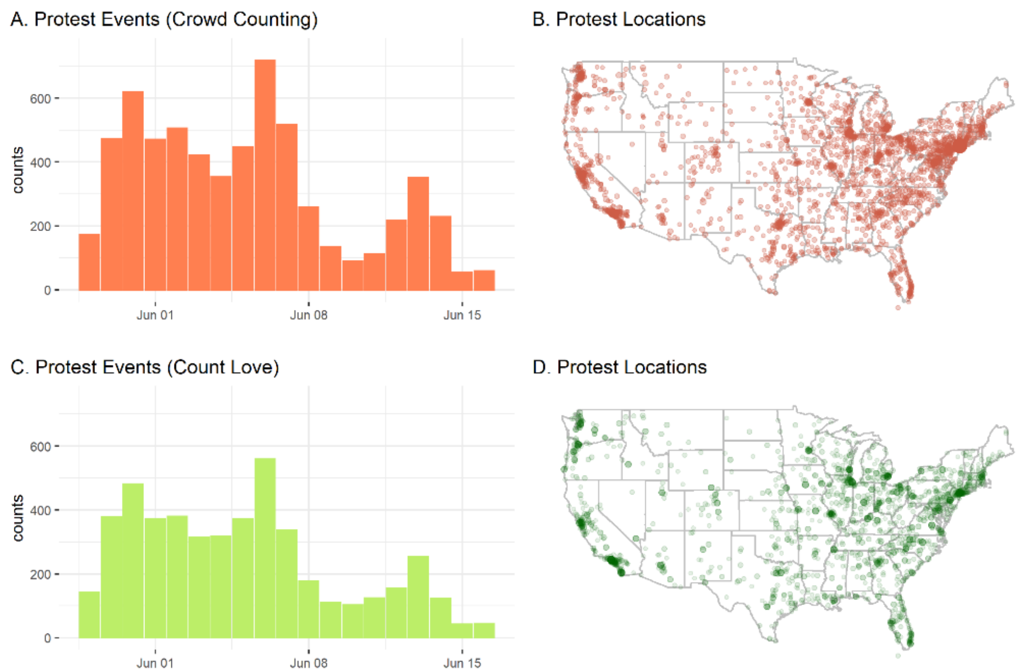

Figure SI2. Temporal and Spatial Distribution of Protest Events. The upper row shows count (A) and location (B) data according to the Crowd Counting Consortium (1). The lower row shows count (C) and location (D) data according to Count Love (2). The two sources of data are consistent in the overall patterns. June 6 is the day with more intense protest activity. We find similar patterns in the Twitter data we analyze (see Figure 1 in the main text).

(3) We use Comscore's Plan Metrix data (multi-platform key measures reports) to obtain audience reach and ideological composition for news domains. Plan Metrix combines web tracking and survey responses for N ~ 12,000 panelists and it is provided monthly. This panel is representative of the U.S. population and it has been used in prior research to examine the presence of ideological segregation online (6) and selective exposure in the consumption of digital news (7). We use this data to obtain web audience reach estimates for the domains shared on Twitter during the protest period and estimates of the partisanship and political outlook of the audiences of these domains. Audience reach measures the fraction of the online population that accessed a given web domain during

the period we consider (May and June 2020). To assign ideology scores to these domains, we use the audience-based measures $fav_p(d) = \frac{(R-D)}{(R+D)}$ and $fav_o(d) = \frac{(C-L)}{(C+L)}$, following prior work (5, 7). The score is 1 when a domain is visited exclusively by Republicans or Conservatives, and -1 when it is visited exclusively by Democrats or Liberals. For Political Party, the calculation excludes panelists that self-identify as 'Independent' or disclose no affiliation. For Political Outlook we count as 'Conservative' the fraction of panelists that self-identify as 'Very Conservative' or 'Somewhat Conservative', and we count as 'Liberal' the fraction that self-identify as 'Very Liberal' or 'Somewhat Liberal'. The calculation excludes 'Middle of the Road' panelists. The sections that follow also present analyses for a second version of Political Outlook (v2) that uses only the extreme categories 'Very Conservative' and 'Very Liberal' to binarize the variable. See figures SI5, SI6, SI7, SI8, SI10, SI14, SI17, SI18, SI19 and SI20 for a comparison of v1 (used in the main analyses) and v2 (focused on the extremes) of political outlook.

(4) We obtain reliability scores for the domains shared on Twitter from NewsGuard, a journalism and technology company that rates the credibility of news and information websites. Each site receives a trust score on a 0-100 scale based on nine criteria, five related to credibility (the site does not repeatedly publish false content; gathers and presents information responsibly; regularly corrects or clarifies errors; handles the difference between news and opinion responsibly; avoids deceptive headlines) and four related to transparency (website discloses ownership and financing; clearly labels advertising; reveals who's in charge, including possible conflicts of interest; and the site provides names of content creators, along with either contact or biographical information). Websites with a score of 60 points or higher receive a green rating (i.e., the website generally adheres to basic standards of credibility and transparency); below 60 points, websites receive a red rating (i.e., the website generally fails to meet basic standards of credibility and transparency). The criteria are evaluated and applied by trained journalists. More information on the rating process and methodology can be found in (8). We use this data to obtain reliability scores for the domains shared on Twitter during the protest period. We average these reliability scores for May and June 2020, the two months we cover with the Twitter and protest events data sources (see figure SI10 in section 2 for plots of the reliability distributions).

2. URLs and Domains

We extracted all URLs present in the tweets we analyze (N ~ 1.3 million unique tweets, 4% of these containing URLs) and we pruned them to identify unique registered domains. URL shortening created some duplicates, for instance, chng.it and change.org for the petition website. We consolidated these duplicates aggregating the counts under the main domain (in this example, change.org). This process resulted in *N* = 2,176 unique domains. We then classified these domains in the categories shown in figure 2A of the main manuscript. We used Comscore's classification of web domains when available (e.g., news/information, entertainment, etc), and we manually checked to solve inconsistencies, errors, and missing labels. The most common category is, by far, 'News/Information' (*N* = 925 domains belong to this group); within this category, ‘Local News’ is the most common (N = 496). Table SI1 shows the top 5 domains according to total URL count in the top 5 categories; the last column of the table shows the number of unique URLs for each domain.

Figure SI3 shows the top 30 domains within the “News/Information” category, ranked according to total counts (A) and unique counts (B). General news prevails in the top of the rankings, followed by local news. The insets show the rankings within the “politics” sub-category.

Rank	Category	Sub-rank	Domain	URL count	Unique URLs
1	News/Information				
		1	abc13.com	2114	73
		2	nytimes.com	1902	263
		3	cnn.com	1062	298
		4	foxnews.com	758	144
		5	washingtonpost.com	630	135
2	Entertainment				
		1	pscp.tv	734	123
		2	beyonce.com	409	4
		3	variety.com	331	21
		4	hollywoodreporter.com	137	4
		5	tmz.com	120	52
3	Political Organization				
		1	change.org	6139	4245
		2	blacklivesmatters.carrd.co	263	11
		3	minnesotafreedomfund.org	247	5
		4	go.theactionpac.com	186	130
		5	sign.moveon.org	120	101
4	Lifestyles				
		1	thegospelcoalition.org	783	1
		2	refinery29.com	329	13
		3	christianitytoday.com	94	6
		4	vogue.com	49	12
		5	thecut.com	33	3
5	Online Services				
		1	bit.ly	1257	383
		2	dlvr.it	837	69
		3	trib.al	375	138
		4	cameo.com	174	3
		5	newsbreakapp.com	103	97

Table SI1. Top 5 Domains within the Top 5 Categories. We show counts for total URLs shared (the criterion for the ranking) and for unique URLs shared.

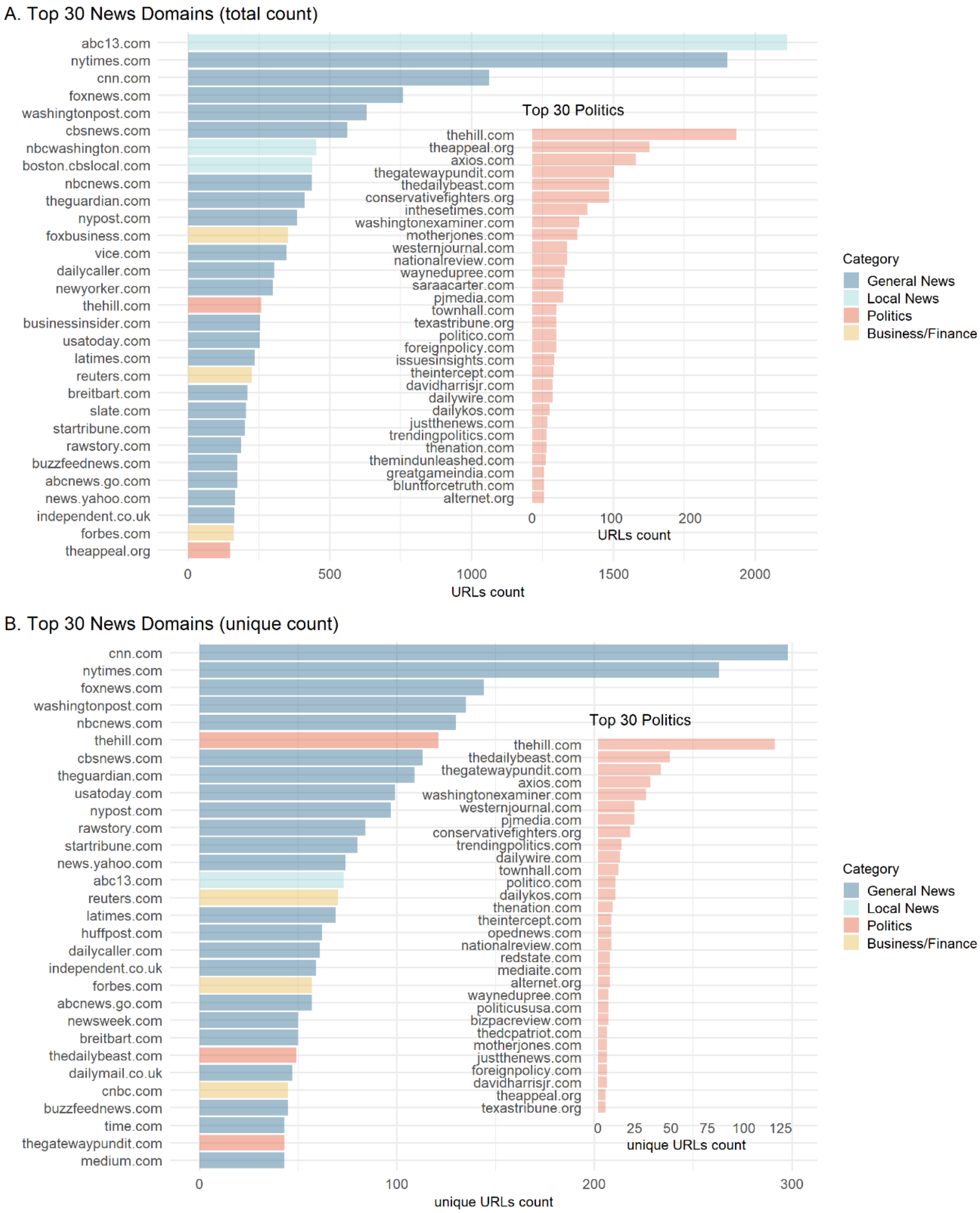

Figure SI3. Top 30 'News/Information' Domains. Panel A shows the top domains classified under the "News/Information" category according to total URL counts; panel B shows the top domains according to the number of unique URLs shared. General news prevail in the top of the rankings, followed by local news. The insets show the top 30 domains within the 'politics' subcategory.

We matched the full list of URLs with the web tracking data, and we obtained reach and audience-based ideology estimates for 52% of them (*N* = 1,135). The domains shared on Twitter during this protest period that do not appear on the web tracking data are likely to have very low reach (which means that very few of Comscore's panelists visited them, so there are no estimates) or they are shortened/deleted URLs (e.g., bit.ly or broken links) that we could not connect to a specific domain. Figure SI4 shows the reach distribution for the domains we could match.

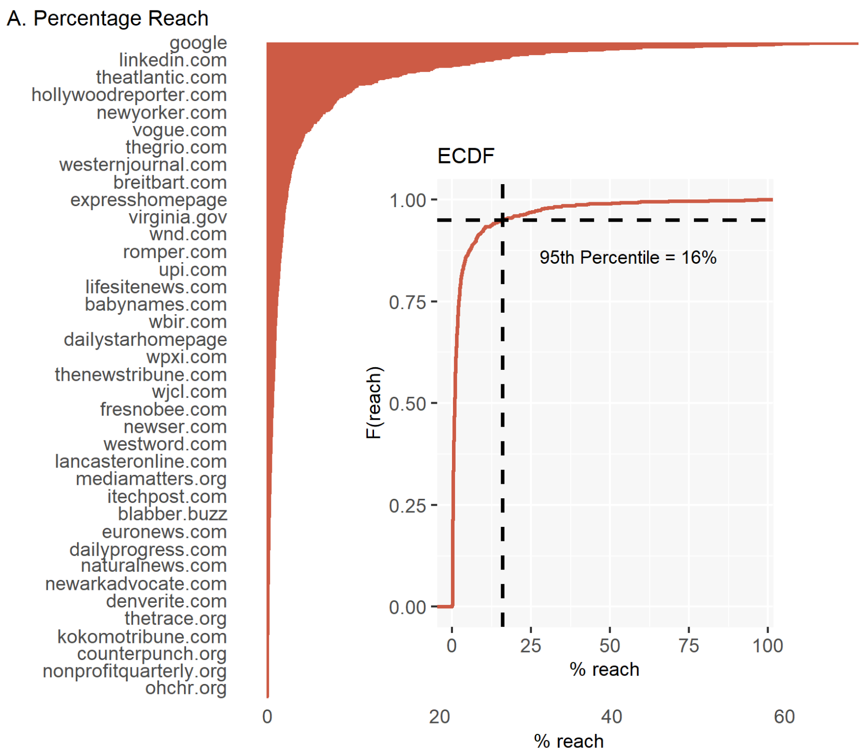

Figure SI4. Audience Reach for Web Domains. The histogram shows the percentage reach of all domains shared on Twitter during the protest period (to improve legibility, labels are only visible for a few domains, the full list contains *N* ~ 1,100). The inset shows the empirical cumulative distribution function (ECDF), which suggests that only 5% of the domains have a percentage reach of 16% or higher.

Figure SI5 shows the distribution of ideology scores for all domains (panel A) and the 'News/Information' (B) and 'Politics' (C) categories. This figure visualizes the same information shown in Figure 3 of the main text but adding the distribution of Political Outlook scores according to v2. This version produces a more extreme shift to the right of the ideological scale. Figure SI6 shows the correlation between the audience reach and the ideology scores for all domains (panels A-C) and for the subset classified as 'News/information' (panels D-F). The scatterplots show a lack of association.

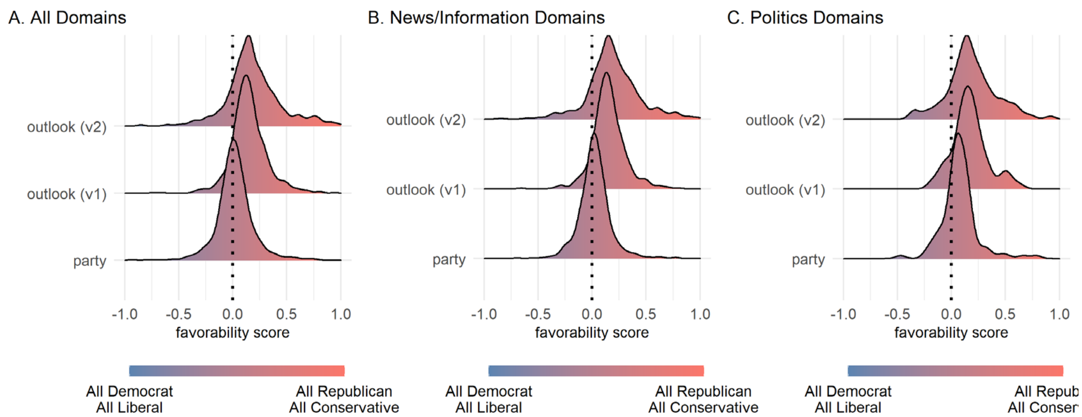

Figure SI5. Ideology Scores Distributions. These density plots summarize the relative position of the domains in our data along the favorability score continuum. Most domains have scores close to zero, suggesting they attract audiences from both sides of the ideological divide. However, the right-leaning bias is clearly visible, especially in the political domains, both in terms of party identification and political outlook (regardless of the version of political outlook used).

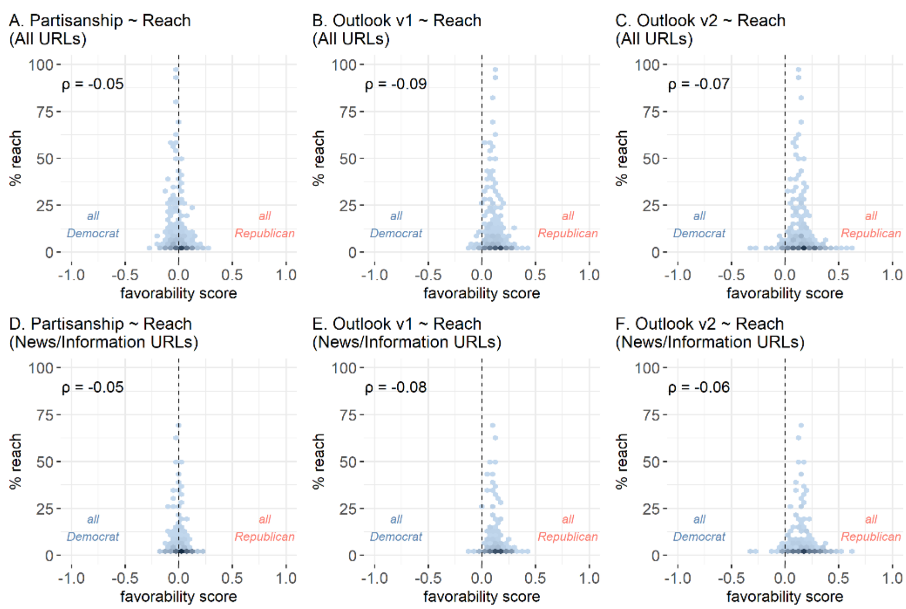

Figure SI6. Correlation of Audience Reach and Ideological Composition of Domains. These scatterplots (binned in hexagons that are colored in proportion to density) show that domains on the extremes of the ideological distribution have lower reach (although, as already shown in Figure SI4, most domains have very low reach, regardless of the ideological composition of their audience). The reach and the ideology scores of domains are largely uncorrelated.

Figure SI7 shows the correlation of the two ideology scores for domains classified in the 'News/Information' category: panel A uses the version of Political Outlook used in the main text (v1), and panel B shows v2. The magnitude of the correlation identified with v1 is closer to the correlation inferred with survey data collected through the American National Election Studies, estimated to be between 0.5 and 0.6 (e.g., 18).

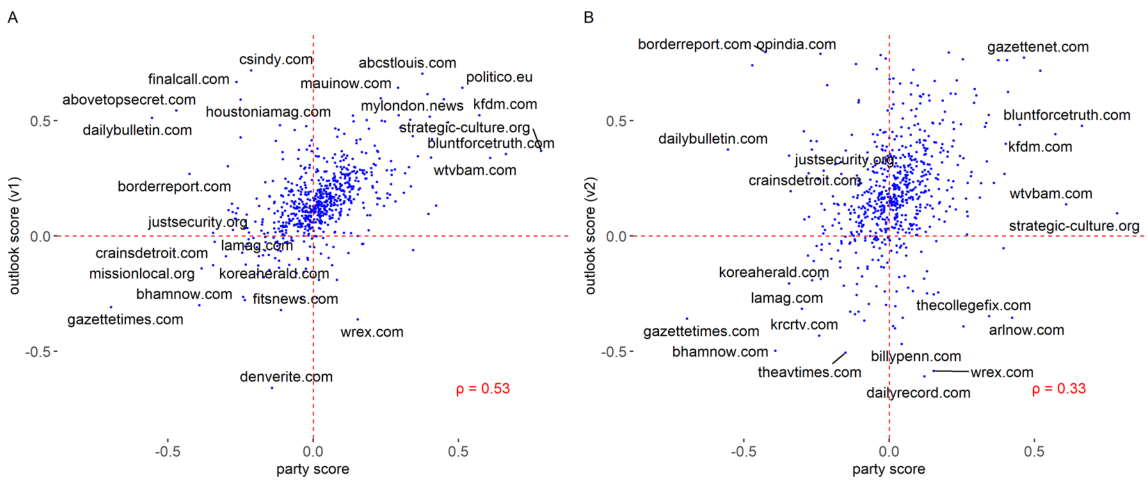

Figure SI7. Correlation of Ideology Scores. These scatterplots show the association between the two favorability scores (partisanship and political outlook, both for v1, used in the main text, and v2) for 'News/Information' domains. The association is moderate, especially for v2, which only focuses on the most extreme audiences to calculate the outlook score (rho ~ 0.3). The magnitude of the correlation shown in panel A is more closely aligned with that inferred using survey data, estimated to be between 0.5 and 0.6 (see, for instance, 18). In this visualization only extreme values are labelled (these are the domains at the tails of the ideology distributions shown in figures SI5 and SI6, which also have lower reach).

Figure SI8 show the scores for the top 30 in terms of total counts and unique counts and figure SI9 maps the location of messages sharing URLs. Tweets containing URLs are only a subset of all tweets in our data but, as shown in Figure 4A in the main text (and in figures SI19 and SI20 below), they are more likely to be posted and retweeted.

To obtain measures of source reliability, we also matched the domains with the sites rated by NewsGuard. Figure SI10 summarizes the distribution of the reliability scores and their lack of association with the ideology scores (including v2 of Political Outlook). To give a few examples, a site that ranks low with a score of 25 is 'neonnettle.com'; 'breitbart.com' has a score of 62; 'foxnews.com' has a score of 69.5; 'cnn.com' a score of 87.5; and 'nytimes.com' has the maximum score 100.

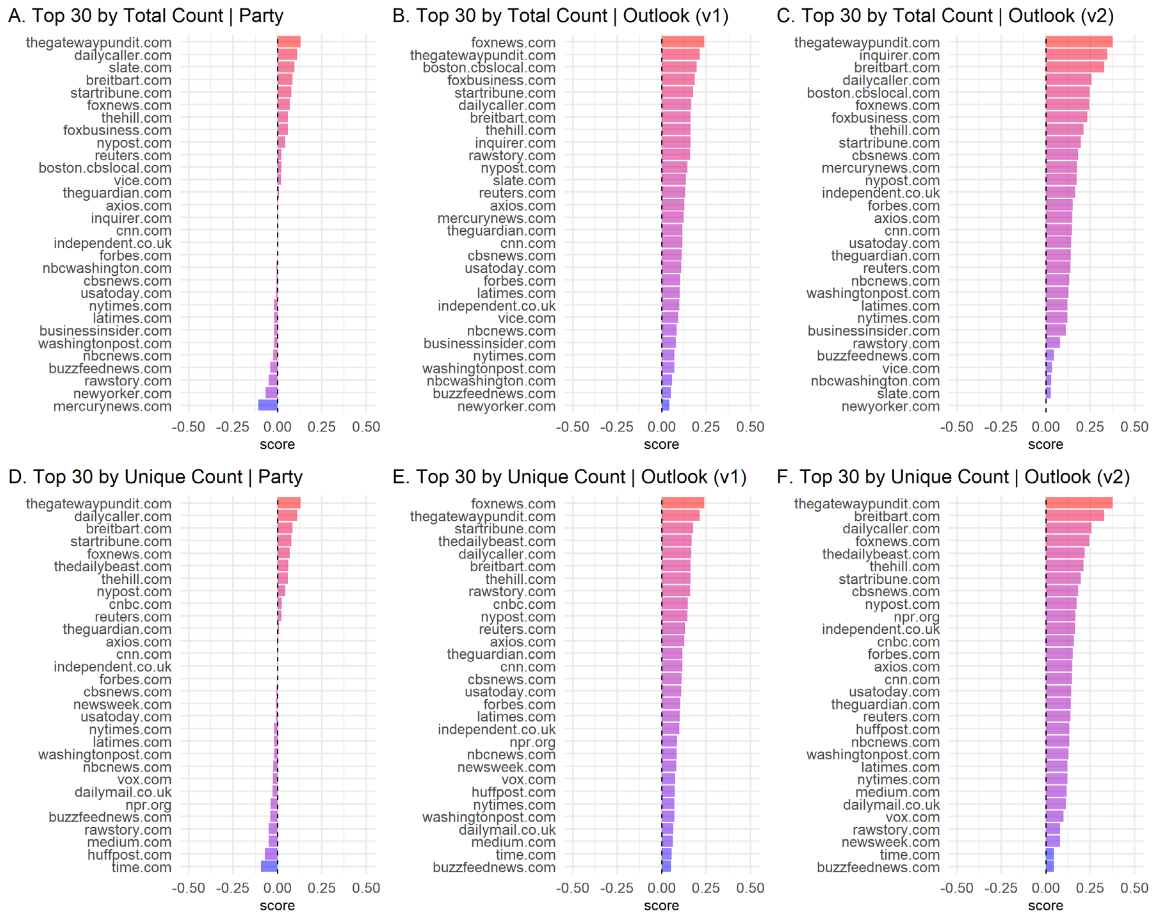

Figure SI8. Ideology Scores for Top 30 Domains. The upper row shows the ideology scores for the top 30 domains in terms of total count; the lower row shows the top 30 in terms of unique count.

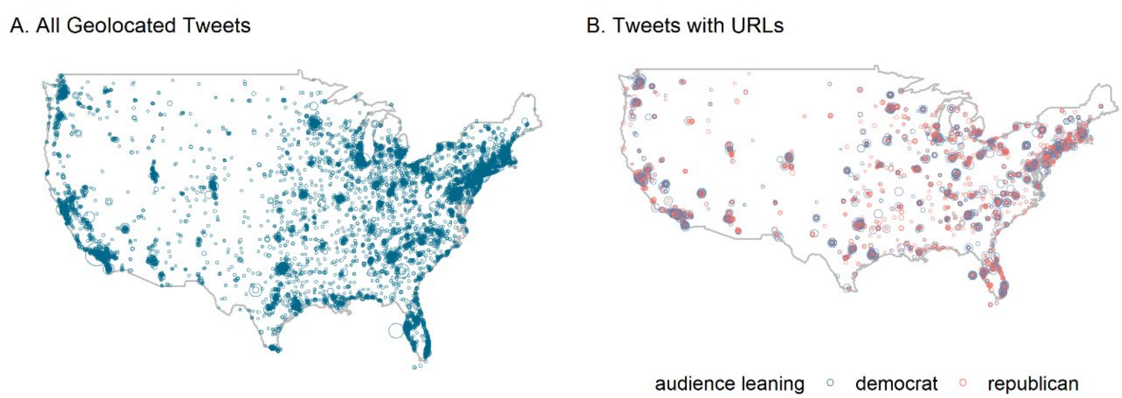

Figure SI9. Spatial Distribution of Tweets with URLs. (A) Map of tweets with geolocation data. (B) Map of tweets with geolocation data that include URLs, with data points color-coded by ideological leaning.

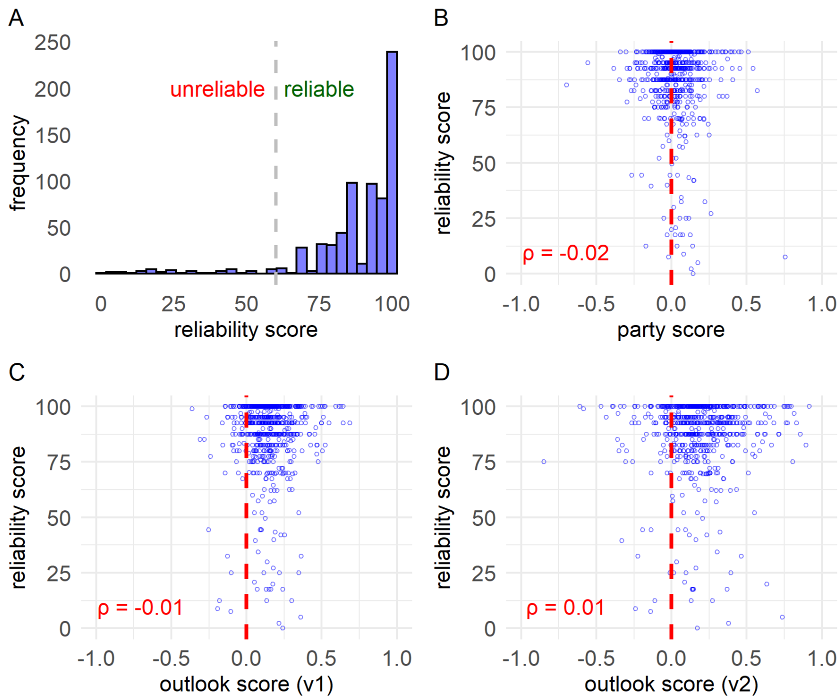

Figure SI10. Reliability Scores of Web Domains. (A) Most domains shared on Twitter are classified by NewsGuard as reliable (they have a reliability score of 60 or higher). (B-D) There is no association between reliability scores and ideology scores.

In addition to the reliability scores from NewsGuard, we also examined overlap with the list of fake and non-fake news websites used in past research (i.e., 9), where the authors analyzed exposure to political content on Twitter during the 2016 U.S. presidential election. Only 44 sites out of our total *N* = 2,176 domains appear in this list -- these sites are shown in Figure SI11. The domains were classified as "black" when identified as spreading 'fake news; "red" when associated with frequent falsehoods; "orange" when spreading occasional falsehoods; "yellow", when some reporting was identified as inaccurate; and "green" when the source was a large news agency with factual and sourced reporting. According to this classification, our data contains 2 black sources, 7 red, 9 orange, 13 yellow, 12 green, and 1 non-news source with satirical content.

We run two additional analyses to determine whether the URL sharing patterns we identify are specific to communication around these BLM protests or they reveal more general patterns characteristic of Twitter activity. First, we compared the visibility of domains in our data set (i.e., English tweets published in the U.S.) with the visibility of the same domains in the full data set of all English tweets. The goal is to determine if our geographic filter is introducing a bias -- for instance, users who disclose their location may be more likely to post links to certain types of sources. Figure SI12 shows the correlation of domain visibility (i.e., number of times URLs to that domain were shared) according to the two sets of Twitter data. The correlation (*rho* ~ 0.7) is moderate-high – higher for URLs classified in the 'News/Information' category and for the domains in the upper tail of the distribution (i.e., those shared more frequently).

Second, we analyzed URL sharing activity in a random sample collected from Twitter during a 20-day timeframe leading up to the mobilizations we analyze. Specifically, we used the Twitter API v2 to gather historical tweets posted in the U.S. in English during the period April 1 to April 20, 2020. This data collection resulted in a total of 12.3 million tweets, of which *N* ~ 5.3 million contain URLs (*N* ~ 282,000 are unique). In figure SI13 we show the results of the comparison of this random dataset with the BLM data. Less than half (42%) of the URLs present in the BLM data also appear in the random sample (the percentage of overlapping URLs is slightly higher for the subcategory 'News/Information', panels A-B). This means that a very different set of sources were being shared on the Twitter stream a month prior to the mobilizations. The correlation in the visibility of these subset of overlapping URLs in the two datasets is weak (*rho* ~ 0.4) regardless of whether we use total count or unique count as a measure of domain salience (panels C-F).

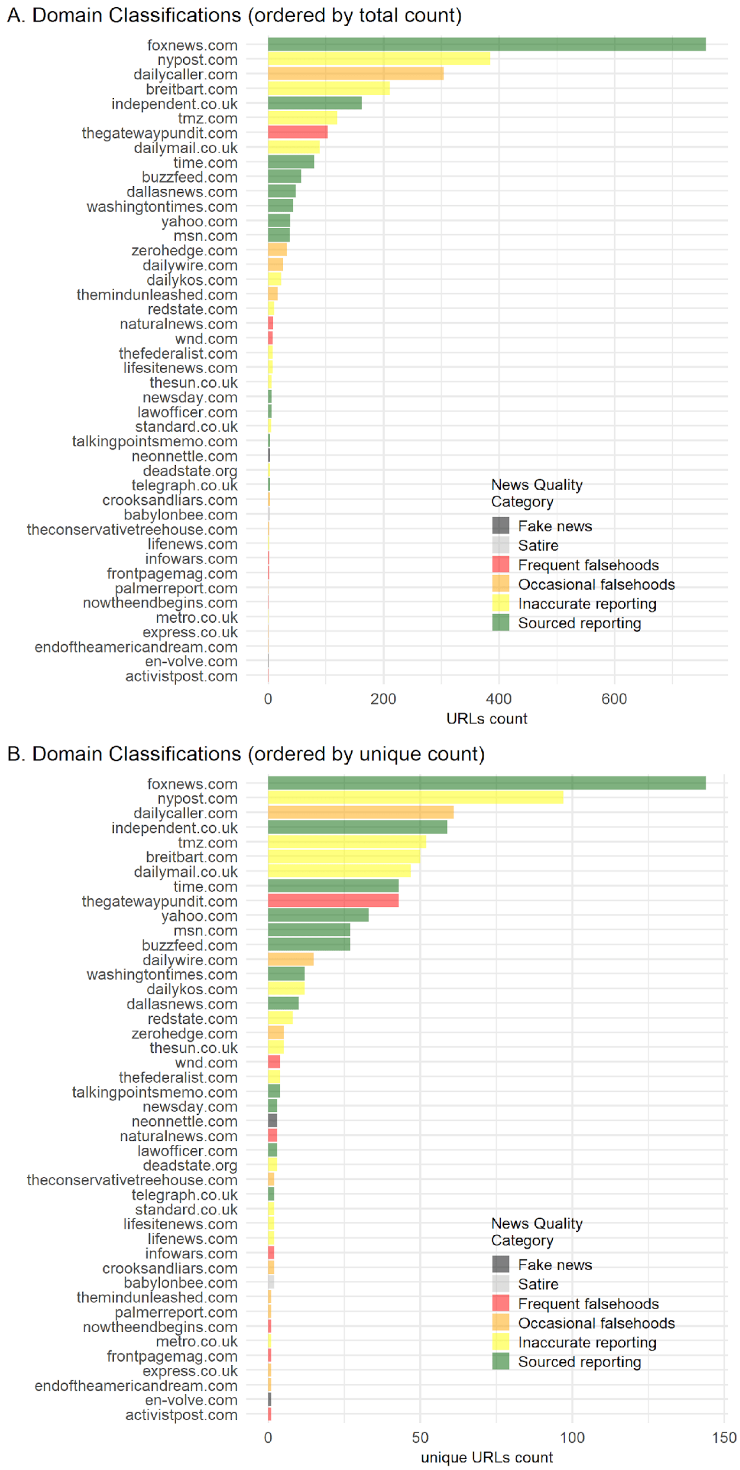

Figure SI11. Domain Quality Classification. The overlap of domains in our data with those found in (9) is very small (44 domains out of 2,176). Only two of these sources are classified as “fake news” (‘neonnettle.com’ and ‘en-volve.com’), and they have very low visibility in terms of URL counts.

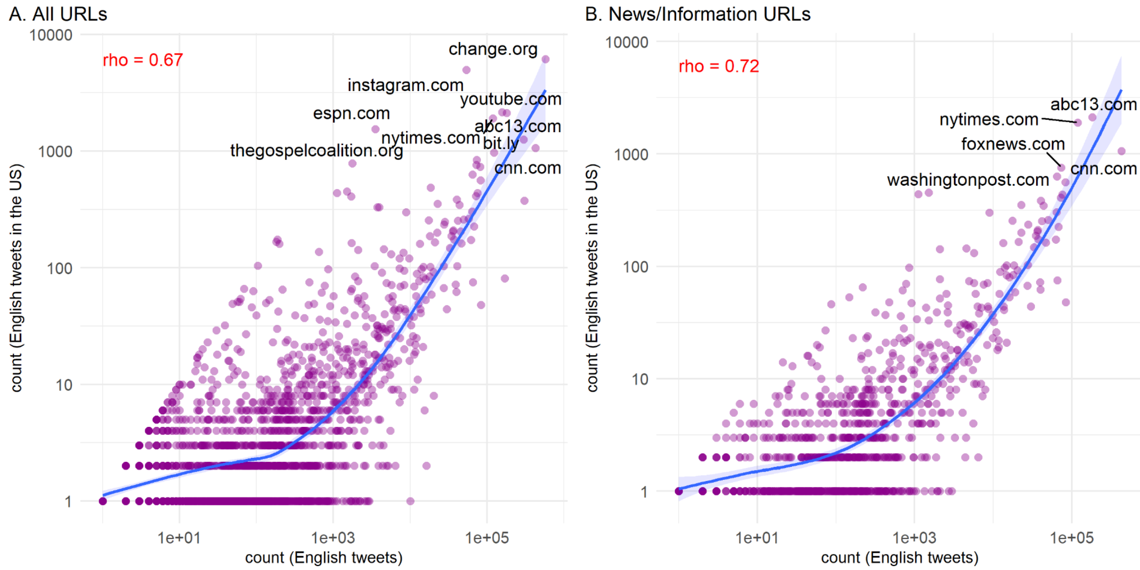

Figure SI12. Domain Visibility in all English Tweets (Regardless of Location). The horizontal axes measure the number of times a given domain was shared in the set of all English tweets, and the vertical axes measure the number of times the domain was shared in all English tweets with location in the U.S. The correlations are high (rho ~ 0.7), especially for URLs in the 'News/Information' category. English tweets are published all over the world, so some differences in domain salience are to be expected. Blue lines and shaded areas trace smoothed regression lines with the standard error.

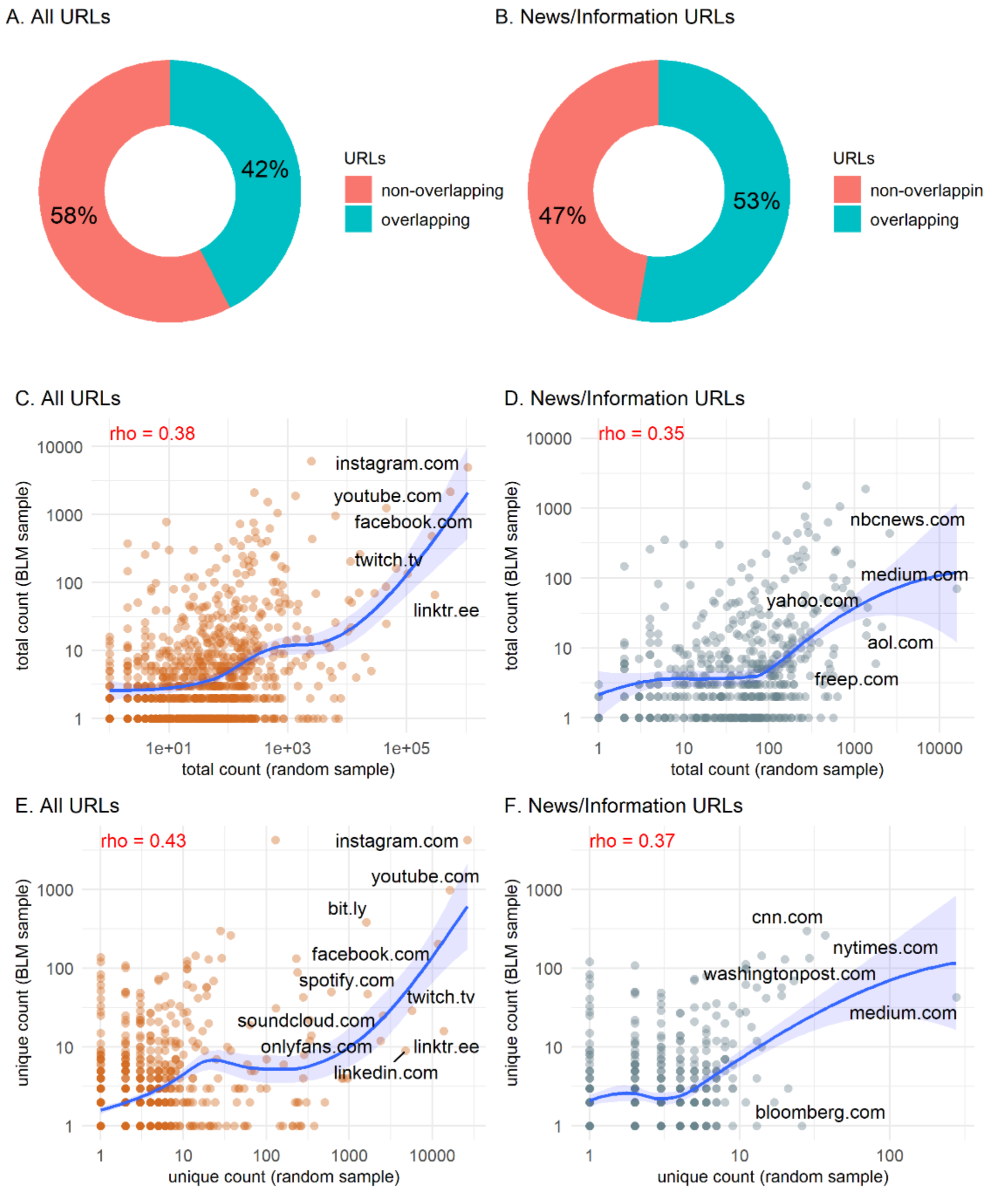

Figure SI13. Domain Visibility in BLM Tweets and a Random Sample of Twitter Activity. Panels A-B show the percentage of unique URLs in the BLM data that also appear in the random sample (spanning activity from April 1 to April 20 2020). Less than half of all URLs (about half for URLs in the 'News/Information' subcategory) appear in the random sample data. The scatterplots show the association in visibility for this subset of overlapping URLs, according to total counts (panels C-D) and unique counts (panels E-F). The correlations are weak (*rho* ~ 0.4). Blue lines and shaded areas trace smoothed regression lines with the standard error.

3. Account Classification

We use bot detection techniques and Twitter's verified feature to classify accounts as media (automated and verified), bot (automated and unverified) and human (non-automated). In line with what has been found in prior work [e.g., (10)], most accounts (~58%) fall in the 'human' category, although unverified bots are close in terms of numbers (~42%). Less than 1% of all accounts are in the verified media category.

We identify automated accounts using a machine learning procedure designed to classify Twitter accounts as humans or bots and used in previous research (10, 11). The procedure trains and validates the classifier on publicly available data sets (10, 12, 13). Overall, our framework relies on information about 22,993 users, consisting of 14,218 bots and 8,775 humans, using 80% of the data set for training and the remaining 20% for validation, while controlling the balance between bots and humans present at the level of the original datasets to have different types of bots in the training and validation stages. The parameters of the model are obtained by means of three-fold cross-validation on the training data set.

Our classifier follows the same prescriptions of recent studies (10, 11, 14, 15) to achieve the maximum accuracy, including 10 account features that can be obtained through Twitter's API as publicly available information: (1) statuses count; (2) followers count; (3) friends count; (4) favorites count; (5) listed count; (6) default profile; (7) geo enabled; (8) profile use background image; (9) protected; and (10) verified.

To better understand which machine learning model achieves the highest accuracy, we compare different state-of-the-art algorithms, including logistic regression (LOGR), ada-boost classifier (AB), random forest (RNF), stochastic gradient descent (SGD), and deep learning (DL). The DL model consists of four fully-connected layers of 2 x Nfeats, 4 x Nfeats, Nfeats and 2 hidden nodes respectively. For all layers we use a rectified linear activation unit (or ReLU) function, with the exception of the last layer, for which we use a sigmoid function. A dropout of 0.2 was also applied between the fully-connected layers in order to prevent overfitting, as in (11). The implementation is based on the pytorch framework (http://pytorch.org/). For all the other models we rely on the scikit-learn library (http://scikit-learn.org).

For each model, we calculate a battery of statistical descriptors to compare accuracy, specificity, sensitivity, balanced accuracy, etc. Overall, we achieve a classification accuracy of about 90% when DL, RNF and AB are considered, with comparable performances also in balanced accuracy (~90%), F1-score (~90%), sensitivity (~90%) and specificity (~90%). Therefore, we opted to use DL in all subsequent analyses.

To test the ability of the algorithm to generalize the classification out of the data sample used for training and validation, we applied the DL on an independent data set built during the 2018 U.S. midterm elections (16), which consist of labeled information about 8,092 humans and 42,446 bots. The choice of this data set is motivated by the geographic relevance of the event for our study, which is focused on the U.S., as well as the fact that it is publicly available and adopted in the literature.

The results of the classifier are satisfactory, with a balanced accuracy close to 60%, an F1-score higher than 70%, and a sensitivity (i.e., recall) of 58%. We estimate that the rate at which a human is erroneously labeled as bot is 7%, while the rate at which a bot is erroneously labeled as human is 42%. While our DL model generalizes fairly well to new data, we note that performance is expected to decrease with respect to training and validation sets, as is well known in the literature (17). Enhancing the classification of bots is not the primary goal of our study, and it is worth noting that out-of-domain performance is still an open challenge for a broad variety of online machine learning systems. Future research will only improve current benchmarks but, for the time being, we have to rely on state-of-the-art classifiers if we want to parse massive datasets (i.e., analyzing millions of accounts, as our study does, cannot rely on manual annotations, but when such a validation is manually performed on a sub-set of accounts, our model has been shown to perform extremely well, correctly classifying 90% of news accounts of interests, (10)).

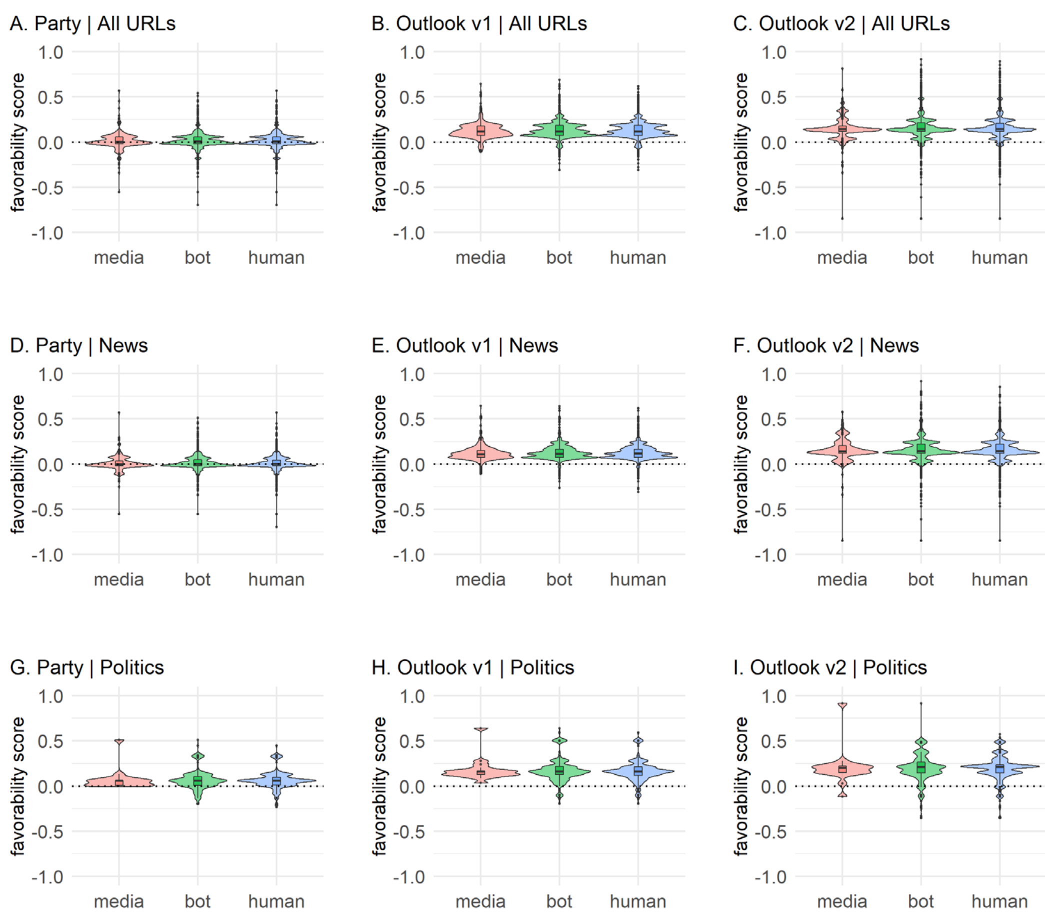

Figure SI14. Ideology Scores by Account Type. The top row (panels A-C) summarizes party and outlook (v1 and v2) distributions for all URLs. The middle row (panels D-F) focuses on the subset of URLs classified as ‘News’, and the last row (panels G-I) focus on URLs classified as ‘Politics’. There are no strong differences in ideology scores for the three types of accounts, although verified media accounts tend to share political URLs that lean more clearly towards the right (e.g., scores above the 0 line), especially for v2 of Political Outlook.

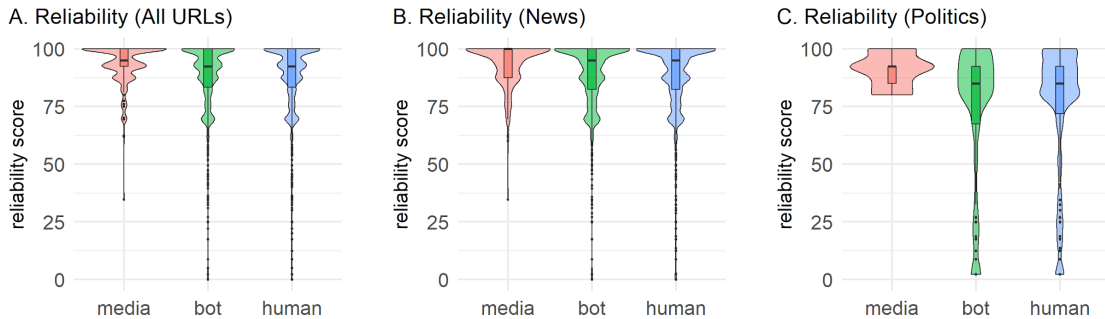

Figure SI15. Reliability Scores by Account Type. Verified media accounts post URLs that are more reliable across types of domains, but especially so for domains classified as “politics”. Bot and human accounts do not exhibit different behaviors when it comes to sharing (un)reliable sources.

4. Retweet Network

Retweets (RTs) are the main mechanism for information diffusion on Twitter. Table SI2 shows descriptive statistics for the largest connected component (LCC) of the weighted RTs network built with our data.

	RTs
size	837485
number of edges	1079598
mean degree	3
mean indegree	1
maximum indegree	63803
mean outdegree	1
maximum outdegree	268
mean strength	3
mean instrength	1
maximum instrength	64263
mean outstrength	1
maximum outstrength	341
reciprocity	0.0002
clustering	0.035
degree correlation	-0.075

Table SI2. Network Statistics for the RTs Network. The table summarizes the statistical properties of the weighted retweet network (LCC). The network is very sparse, with low clustering and very low reciprocity. It is also very heterogeneous in the distribution of centrality, with a minority of accounts attracting most of the RTs.

Figure SI16 visualizes the distribution of centrality measures for the three types of accounts (media, bots, humans) in terms of followers and friends (upper row) and RTs received and made (lower row).

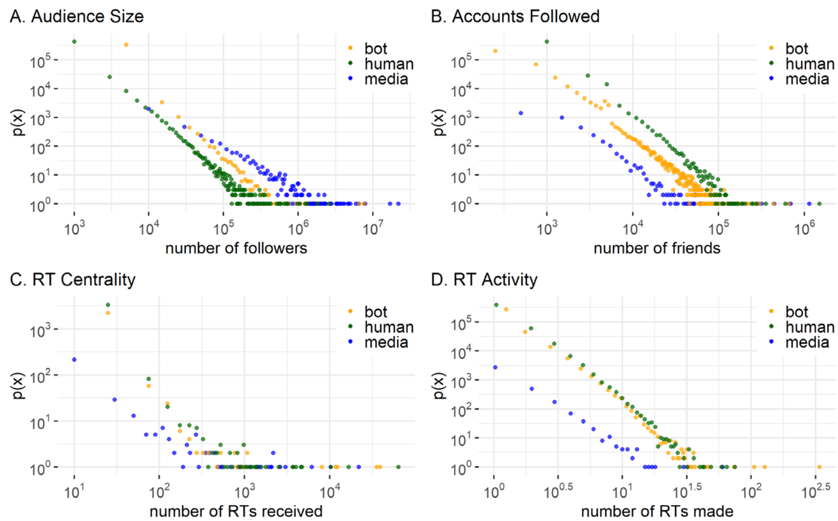

Figure SI16. Centrality Distribution for the Three Types of Accounts. Verified media accounts have larger audiences (A) and they are less active following other accounts (B). No clear patterns appear in terms of RTs received (C) but they are also less likely to RT other accounts (D).

User accounts in the RT network were assigned ideological and reliability scores derived from the URLs they posted (or reposted). We averaged those scores or assigned missing values if no URLs were shared by a given account. In the network collapsed to communities (which, as explained in the Materials and Methods section of the main text, we identified using a random walk algorithm), we averaged the scores of individual accounts classified in each community. Figure SI17 reproduces Figure 4C in the main text and extracts the top 10 communities in terms of size (i.e., the number of user accounts classified in each community). The lower row zooms onto these top 10 communities, where node color encodes mean ideology scores. The two largest communities are #1 (N ~ 1900, most popular domain ‘cnn.com’) and #20 (N ~ 1400, most retweeted domain ‘foxbusiness.com’). The headline of the most popular news URL shared in community #1 reads “Minnesota police arrest CNN reporter and camera crew as they report from protests in Minneapolis”. The headline for the most popular news URL shared in community #20 is “Manufacturing company in Minneapolis since 1987 leaving city after violent protests”. The sparsity of RTs between communities #1 and #20 (and the clusters around them) indicates there is a divide in the diffusion of #BLM messages that separates two distinct sets of communities;

however, these two groups do not map onto the two halves of the ideological continuum, as indicated by the favorability scores, which in some cases are very close to 0 but always above the 0 line.

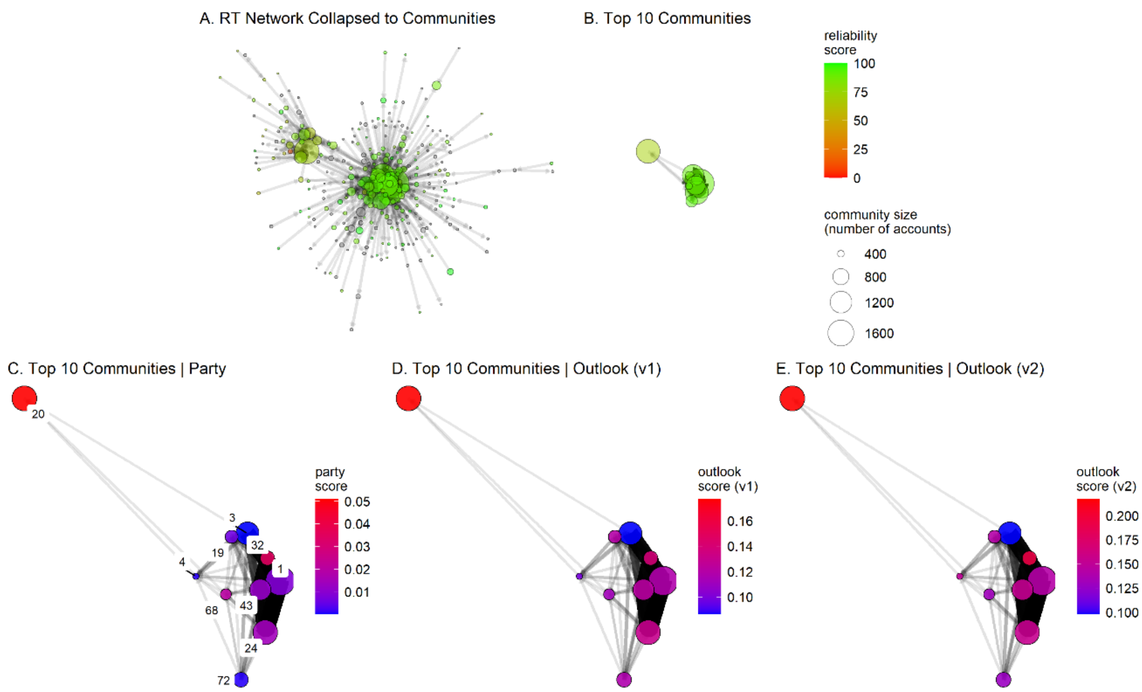

Figure SI17. Top 10 Communities in the RT Network. Panel A reproduces Figure 4C in the main text. Panel B extracts the top 10 communities, in terms of size (or number of unique accounts classified in each). Panels C to E assign average ideology scores to these communities; edge width is proportional to the number of retweets across communities. The two largest communities are #1 (N ~ 1900) and #20 (N ~ 1400). Users in these two communities do not retweet each other, although they are not far apart in their average ideology scores (derived from the URLs shared within): mean party scores are 0.009 and 0.05, respectively; mean outlook scores are 0.12 and 0.28 (v1), and 0.14 and 0.22 (v2). The most central user in community #1 is an American attorney who specializes in civil rights; the most popular news domain in this community is cnn.com. The most central user in community #20 is a conservative talk radio host; the most popular news domain is foxbusiness.com. The structural hole, or sparsity of RTs, separating community #20 (and the cluster around it, as depicted in panel A) and community #1 (and surrounding cluster) is suggestive of a divide in the diffusion of #BLM messages in two distinct sets of communities; however, these two groups do not map onto the two halves of the ideological continuum, as indicated by the favorability scores, all above the 0 line (and, therefore, with a right-leaning slant).

Figure SI18 locates the top 10 communities in the two-dimensional space created by the two ideology scores, including v2 of Political Outlook. The association does not change substantially when using v2 of political outlook, but communities receive more extreme scores.

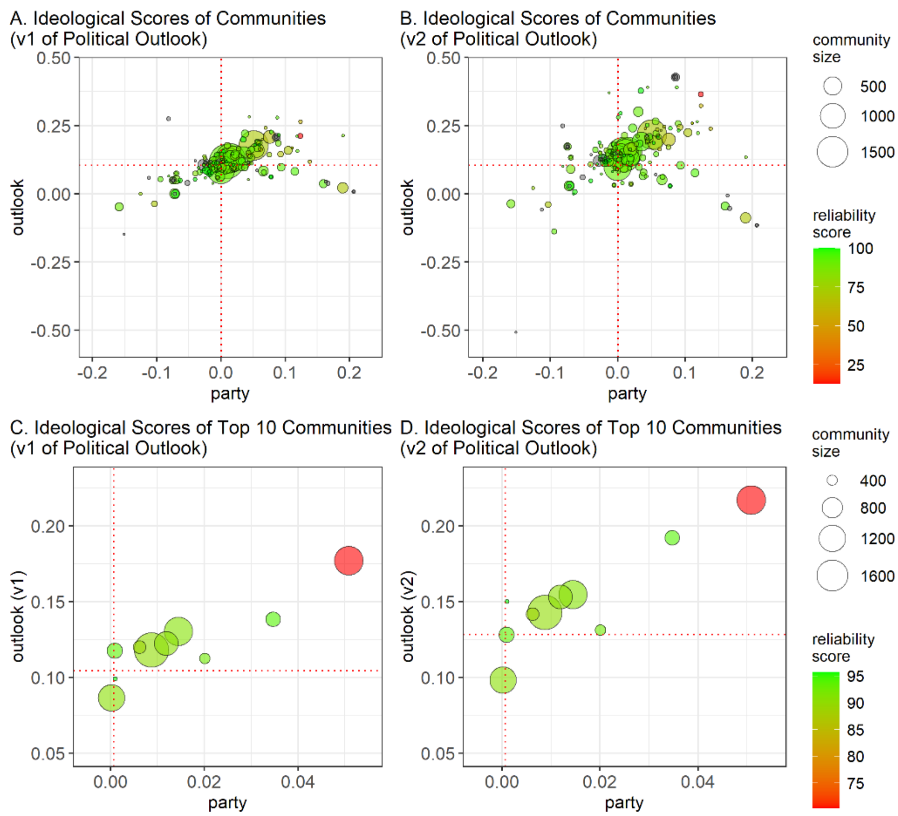

Figure SI18. Ideology Scores for the Top 10 Communities. Panels A and C show the association of party identification and v1 of political outlook; panels B and D use v2 of political outlook. Dashed red lines mark the median values of the corresponding distributions (note that for political outlook, the median value is above zero). The association does not change drastically when using v2 of political outlook, but communities receive more extreme scores. Communities in the lower right and upper left quadrants of panels A and B are smaller and with less tweets embedding URLs, so the average ideology scores for these communities are less accurate.

5. Regression Models

We report results from two sets of linear regression models. The first use domain visibility as the dependent variable, operationalized as total URLs and unique URLs shared (in both cases, log-transformed). In these models the main unit of analysis are the domains shared. Figure SI19 compares the outputs reported in the main paper (Figure 2D) with the same models using v2 of

Political Outlook. Using this operationalization renders the ideology variables statistically insiginificant. Tables SI3 and SI4 show the full regression tables.

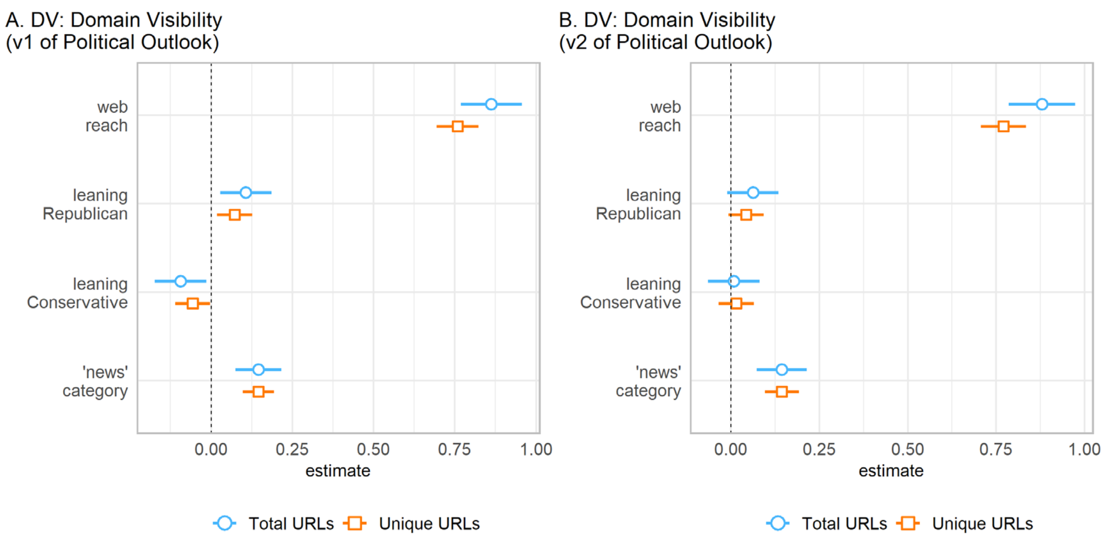

Figure SI19. Regression Models Explaining Domain Visibility. Panel A reproduces Figure 2D in the main text. Panel B shows the same outputs using v2 of political outlook (using only the extreme categories 'Very Liberal'/'Very Conservative' to define the conservative slant of domains). With this operationalization, the ideology variables do not have statistically significant effects.

The second set of regression models employ user centrality in the RT network, opearationalized as number of RTs received (weighted version of centrality, reported in the main text) and indegree centrality (unweighted version, shown in SI20 below). Using v2 of Political Outlook again renders the ideology scores statistically insignificant. Results comparing weighted and unweighted definitions of network centrality are indistinguishable. Table SI5 shows the regression table for the effects summarized in figures 4A and 4B in the main text, where the DV is number of RTs received (user-level). The model summarized in the last column includes the effect of the reliability scores (averaged at the user level), which is negative but indistinguishable from 0.

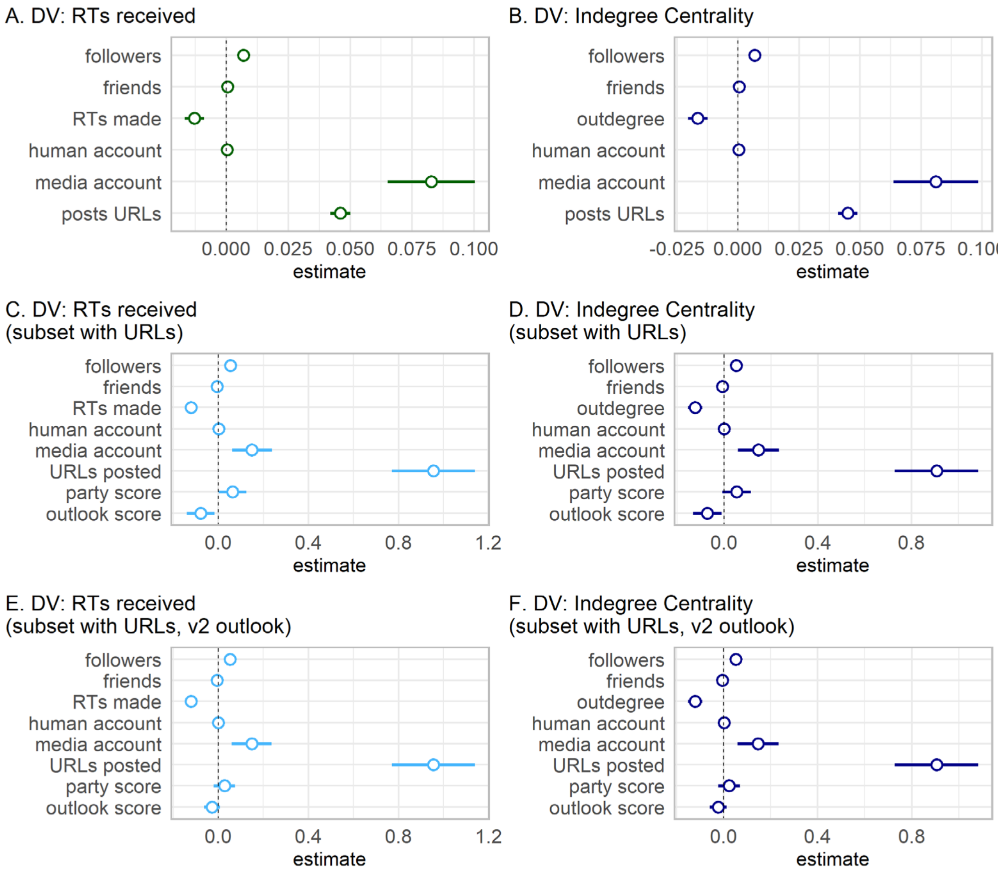

Figure SI20. Regression Models Explaining RTs Received. The first column shows regression results using number of RTs received (user level) as DV. Panel A reproduces Figure 4A and panel C reproduces Figure 4B as shown in the main text. Panel E shows the same model using v2 of Political Outlook, which again renders the ideology scores statistically insignificant. The second column uses indegree centrality (user level) as the DV. Results are indistinguishable compared to the weighted version of network centrality.

	DV: total URLs (v1 outlook)	DV: total URLs (v2 outlook)	DV: total URLs (v1 outlook)	DV: total URLs (v2 outlook)
web reach	0.86 ***	0.88 ***	1.19 ***	1.20 ***
	[0.73, 1.00]	[0.74, 1.02]	[1.05, 1.33]	[1.06, 1.34]
leaning Republican	0.11 **	0.06	0.04	0.02
	[0.04, 0.18]	[-0.01, 0.13]	[-0.04, 0.13]	[-0.06, 0.10]
leaning Conservative	-0.09 *	0.01	-0.05	0.02
	[-0.17, -0.02]	[-0.06, 0.08]	[-0.13, 0.04]	[-0.06, 0.10]
'news' category	0.15 ***	0.14 ***	0.18 **	0.18 **
	[0.07, 0.22]	[0.07, 0.22]	[0.06, 0.29]	[0.06, 0.29]
reliability score			-0.00 **	-0.00 **
			[-0.01, -0.00]	[-0.01, -0.00]
N	1107	1107	697	697
R2	0.25	0.24	0.38	0.37

Standard errors are heteroskedasticity robust. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table SI3. Linear Regressions (OLS) Explaining Domain Visibility (Total URLs). This table uses total count as DV and different specifications (the last two models include domain reliability scores, which reduce the number of observations -- many news domains do not have a reliability score assigned). Controlling for reach and the category of the domain, domains whose audiences lean republican have more visibility and, controlling for party identification, domains leaning conservative have less. The effects of ideology stop being significant under v2 of Political Outlook. After reliability is controlled for, the effects of ideology also cease being significant. Since the effects of the reliability scores are effectively 0 (and many observations are dropped) we chose to focus on model 1.

	DV: unique URLs (v1 outlook)	DV: unique URLs (v2 outlook)	DV: unique URLs (v1 outlook)	DV: unique URLs (v2 outlook)
web reach	0.76 ***	0.77 ***	0.98 ***	0.99 ***
	[0.65, 0.87]	[0.66, 0.88]	[0.88, 1.07]	[0.89, 1.08]
leaning Republican	0.07 **	0.04	0.04	0.01
	[0.02, 0.12]	[-0.00, 0.09]	[-0.02, 0.09]	[-0.04, 0.07]
leaning Conservative	-0.06 *	0.01	-0.03	0.03
	[-0.11, -0.01]	[-0.03, 0.06]	[-0.09, 0.02]	[-0.03, 0.08]
'news' category	0.15 ***	0.14 ***	0.19 ***	0.19 ***
	[0.10, 0.19]	[0.09, 0.19]	[0.12, 0.25]	[0.12, 0.25]
reliability score			-0.00 **	-0.00 **
			[-0.00, -0.00]	[-0.00, -0.00]
N	1107	1107	697	697
R2	0.35	0.35	0.49	0.49

Standard errors are heteroskedasticity robust. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table SI4. Linear Regressions (OLS) Explaining Domain Visibility (Unique URLs). This table uses unique count as DV and different specifications (the last two models include domain reliability scores, which reduces the number of observations -- many news domains do not have a reliability score assigned). Controlling for reach and the category of the domain, domains whose audiences lean republican have more visibility and, controlling for party identification, domains leaning conservative have less. The effects of ideology stop being significant under v2 of Political Outlook. After reliability is controlled for, the effects of ideology also cease being significant. Since the effects of the reliability scores are effectively 0 (and many observations are dropped) we chose to focus on model 1.

	DV: RTs received	DV: RTs received (subset with URLs)	DV: RTs received (subset with URLs, v2 outlook)	DV: RTs received (subset with URLs)	DV: RTs received (subset with URLs, v2 outlook)
followers	0.01 ***	0.05 ***	0.05 ***	0.04 ***	0.04 ***
	[0.01, 0.01]	[0.04, 0.06]	[0.04, 0.06]	[0.03, 0.06]	[0.03, 0.06]
friends	0.00	-0.01	-0.01	-0.00	-0.00
	[-0.00, 0.00]	[-0.02, 0.01]	[-0.02, 0.01]	[-0.01, 0.01]	[-0.01, 0.01]
RTs made	-0.01 ***	-0.12 ***	-0.12 ***	-0.12 ***	-0.12 ***
	[-0.02, -0.01]	[-0.14, -0.10]	[-0.14, -0.10]	[-0.15, -0.10]	[-0.15, -0.10]
human account	0.00 *	0.00	0.00	-0.00	-0.00
	[0.00, 0.00]	[-0.00, 0.01]	[-0.01, 0.01]	[-0.01, 0.01]	[-0.01, 0.01]
media account	0.08 ***	0.15 ***	0.15 ***	0.11 ***	0.11 ***
	[0.07, 0.10]	[0.08, 0.22]	[0.08, 0.22]	[0.05, 0.17]	[0.05, 0.17]
posts URLs	0.05 ***				
	[0.04, 0.05]				
URLs posted		0.96 ***	0.96 ***	1.00 ***	0.99 ***
		[0.82, 1.10]	[0.82, 1.10]	[0.80, 1.19]	[0.80, 1.19]
party score		0.06 **	0.03	0.03	-0.03
		[0.02, 0.11]	[-0.01, 0.06]	[-0.04, 0.11]	[-0.09, 0.03]
outlook score (v1)		-0.08 ***		-0.06	
		[-0.13, -0.03]		[-0.13, 0.01]	
outlook score (v2)			-0.03 *		0.02
			[-0.06, -0.00]		[-0.02, 0.07]
reliability				0.00	0.00
				[-0.00, 0.00]	[-0.00, 0.00]
N	824438	19177	19177	11682	11682
R2	0.02	0.11	0.11	0.11	0.11

Standard errors are heteroskedasticity robust. *** p < 0.001; ** p < 0.01; * p < 0.05.

Table SI5. Linear Regressions (OLS) Explaining Centrality in the RT Network. Posting URLs increases centrality and, within the subset of users who post these URLs, those who post Republican-leaning links are also more central (controlling for partisanship, those posting URLs that lean conservative are less central). The model using v2 of Political Outlook reduces the magnitude and statistical significance of the ideology effects. Models including reliability scores (averaged at the user level) also eliminate the significance of the ideology scores but, again, the use of this variable reduces the number of observations (since many domains do not have a reliability score assigned). Because the effect of the reliability scores is again effectively 0 and the inclusion of this variable does not increase the amount of variance explained, we decided to focus on model 2.

6. References

1. Pressman J & Chenoweth E (2021) Crowd Counting Consortium, https://sites.google.com/view/crowdcountingconsortium/view-download-the-data.
2. Leung T & Perkins N (2021) Count Love, https://countlove.org/.
3. Buchanan L, Bui Q, & Patel JK (July 3 2020) Black Lives Matter May Be the Largest Movement in U.S. History. The New York Times.
4. Burch ADS, Gianordoli G, McCarthy M, & Patel JK (June 13 2020) How Black Lives Matter Reached Every Corner of America. The New York Times.
5. Tyler M, Grimmer J, & Iyengar S (2021) Partisan Enclaves and Information Bazaars: Mapping Selective Exposure to Online News. *Journal of Politics* forthcoming.
6. Gentzkow M & Shapiro JM (2011) Ideological Segregation Online and Offline. *The Quarterly Journal of Economics* 126:1799-1839.
7. Yang T, Majó-Vázquez S, Nielsen RK, & González-Bailón S (2020) Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*:202006089.
8. NewsGuard (2021) Rating Process and Criteria, https://www.newsguardtech.com/ratings/rating-process-criteria/.
9. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, & Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425):374-378.
10. González-Bailón S & De Domenico M (2021) Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences* 118(11):e2013443118.
11. Stella M, Cristoforetti M, & De Domenico M (2019) Influence of augmented humans in online interactions during voting events. *PLOS ONE* 14(5):e0214210.
12. Cresci S, Di Pietro R, Petrocchi M, Spognardi A, & Tesconi M (2015) Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems* 80:56-71.
13. Cresci S, Pietro RD, Petrocchi M, Spognardi A, & Tesconi M (2017) The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. in *Proceedings of the 26th International Conference on World Wide Web Companion* (International World Wide Web Conferences Steering Committee, Perth, Australia), pp 963–972.
14. Stella M, Ferrara E, & De Domenico M (2018) Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*.
15. Ferrara E (2017) Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. *First Monday* 22(8).
16. Yang K-C, Varol O, Hui P-M, & Menczer F (2020) Scalable and Generalizable Social Bot Detection through Data Selection. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(01):1096-1103.
17. Echeverria J*, et al.* (2018) LOBO: Evaluation of Generalization Deficiencies in Twitter Bot Classifiers. in *Proceedings of the 34th Annual Computer Security Applications Conference* (Association for Computing Machinery, San Juan, PR, USA), pp 137–146.
18. Lupton, R. N., Smallpage, S. M., & Enders, A. M. (2017). Values and Political Predispositions in the Age of Polarization: Examining the Relationship between Partisanship and Ideology in the United States, 1988–2012. *British Journal of Political Science*, 1-20.