The spread of low-credibility content by social bots

Chengcheng Shao,[1,2] Giovanni Luca Ciampaglia,[3] Onur Varol,[1]
Kaicheng Yang,[1] Alessandro Flammini,[1,3] Filippo Menczer,[1,3*]

[1]School of Informatics, Computing, and Engineering, Indiana University, Bloomington, USA
[2]College of Computer, National University of Defense Technology, Changsha, Hunan, China
[3]Indiana University Network Science Institute, USA

[*]To whom correspondence should be addressed; E-mail: fil@iu.edu.

The massive spread of digital misinformation has been identified as a major global risk and has been alleged to influence elections and threaten democracies. Communication, cognitive, social, and computer scientists are engaged in efforts to study the complex causes for the viral diffusion of misinformation online and to develop solutions, while search and social media platforms are beginning to deploy countermeasures. With few exceptions, these efforts have been mainly informed by anecdotal evidence rather than systematic data. Here we analyze 14 million messages spreading 400 thousand articles on Twitter during and following the 2016 U.S. presidential campaign and election. We find evidence that social bots played a disproportionate role in amplifying low-credibility content. Accounts that actively spread articles from low-credibility sources are significantly more likely to be bots. Automated accounts are particularly active in amplifying content in the very early spreading moments, before an article goes viral. Bots also target users with many followers through replies and mentions. Humans are vulnerable to this manipulation, retweet-

ing bots who post links to low-credibility content. Successful low-credibility sources are heavily supported by social bots. These results suggest that curbing social bots may be an effective strategy for mitigating the spread of online misinformation.

Introduction

If you get your news from social media, as most Americans do (*1*), you are exposed to a daily dose of false or misleading content — hoaxes, conspiracy theories, fabricated reports, click-bait headlines, and even satire. We refer to such content collectively as "misinformation." The incentives are well understood: traffic to fake news sites is easily monetized through ads (*2*), but political motives can be equally or more powerful (*3, 4*). The massive spread of digital misinformation has been identified as a major global risk (*5*). Claims that fake news can influence elections and threaten democracies (*6*), however, are hard to prove (*7*). Yet we have witnessed abundant demonstrations of real harm caused by misinformation and disinformation spreading on social media, from dangerous health decisions (*8*) to manipulations of the stock market (*9*).

A complex mix of cognitive, social, and algorithmic biases contribute to our vulnerability to manipulation by online misinformation (*10*). These include information overload and finite attention (*11*), novelty of false news (*12*), the selective exposure (*13–15*) caused by polarized and segregated online social networks (*16, 17*), algorithmic popularity bias (*18–20*), and other cognitive vulnerabilities such as confirmation bias and motivated reasoning (*21–23*).

Abuse of online information ecosystems can both exploit and reinforce these vulnerabilites. While fake news are not a new phenomenon (*24*), the ease with which social media can be manipulated (*4*) creates novel challenges and particularly fertile grounds for sowing disinformation. Public opinion can be influenced thanks to the low cost of producing fraudulent websites and high volumes of software-controlled profiles or pages, known as *social bots* (*9, 25*). These

fake accounts can post content and interact with each other and with legitimate users via social connections, just like real people (*26*). Bots can tailor misinformation and target those who are most likely to believe it, taking advantage of our tendencies to attend to what appears popular, to trust information in a social setting (*27*), and to trust social contacts (*28*). Bots alone may not entirely explain the success of false news, but they do contribute to it (*12*). Since earliest manifestations uncovered in 2010 (*3, 4*), we have seen influential bots affect online debates about vaccination policies (*9*) and participate actively in political campaigns, both in the U.S. (*29*) and other countries (*30, 31*).

The fight against online misinformation requires a grounded assessment of the relative impact of different mechanism by which it spreads. If the problem is mainly driven by cognitive limitations, we need to invest in news literacy education; if social media platforms are fostering the creation of echo chambers, algorithms can be tweaked to broaden exposure to diverse views; and if malicious bots are responsible for many of the falsehoods, we can focus attention on detecting this kind of abuse. Here we focus on gauging the latter effect. Most of the literature about the role played by social bots in the spread of misinformation is based on anecdotal or limited evidence; a quantitative understanding of the effectiveness of misinformation-spreading attacks based on social bots is still missing. A large-scale, systematic analysis of the spread of low-credibility content by social bots is now feasible thanks to two tools developed in our lab: the *Hoaxy* platform to track the online spread of claims (*32*) and the *Botometer* machine learning algorithm to detect social bots (*25*). Here we examine how social bots promoted hundreds of thousands of false and misleading articles spreading through millions of Twitter posts during and following the 2016 U.S. presidential campaign.

Results

Our analysis is based on the spread of content from low-credibility sources rather than focusing on individual stories that are labeled as misinformation. There are two reasons for this approach (*10*). First, sources have intent and processes for the deception and manipulation of public opinion. Second, fact-checking millions of individual articles is unfeasible. The links to the articles considered here were crawled from 120 *low-credibility sources* that, according to lists compiled by reputable third-party news and fact-checking organizations, routinely publish various types of false and/or misleading news. Our own analysis of a sample of articles from low-credibility sources confirms that the vast majority of this content is some type of misinformation (see Supplementary Material). We also crawled the articles published by seven independent fact-checking organizations. The present analysis focuses on the period from mid-May 2016 to the end of March 2017. During this time, we collected 389,569 articles from low-credibility sources and 15,053 articles from fact-checking sources. We further collected *all* of the public posts that included links to these articles: 13,617,425 tweets linking to low-credibility stories and 1,133,674 linking to fact checks. See Methods and Supplementary Materials for details.

Low-credibility sources each produced approximately 100 articles per week, on average. By the end of the study period, the mean popularity of these articles was approximately 30 tweets per article per week (see Supplementary Materials). However, as shown in Fig. 1, success is extremely heterogeneous across articles. Whether we measure success by number of accounts sharing an article or number of posts containing a link, we find a very broad distribution of popularity spanning several orders of magnitude: while the majority of articles go unnoticed, a significant fraction go viral. Unfortunately, and consistent with prior analysis using Facebook data (*11*), we observe that the popularity distribution of low-credibility articles is indistinguish-

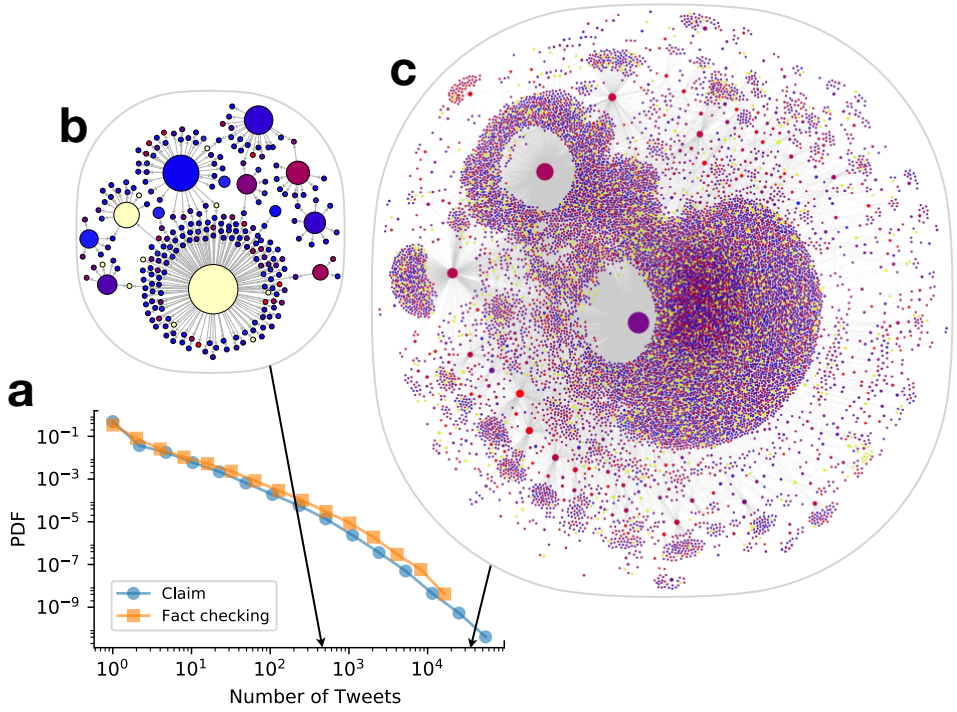

Figure 1: Online virality of content. (a) Probability distribution (density function) of the number of tweets per link, for articles from both low-credibility and fact-checking sources. The distributions of accounts sharing an article are very similar (see Supplementary Materials). As illustrations, the diffusion networks of two stories are shown: (b) a medium-virality misleading article titled *FBI just released the Anthony Weiner warrant, and it proves they stole election,* published a month after the 2016 U.S. election and shared in over 400 tweets; and (c) a highly viral fake news article titled *"Spirit cooking": Clinton campaign chairman practices bizarre occult ritual,* published four days before the 2016 U.S. election and shared in over 30 thousand tweets. In both cases, only the largest connected component of the network is shown. Nodes and links represent Twitter accounts and retweets of the article, respectively. Node size indicates account influence, measured by the number of times an account is retweeted. Node color represents bot score, from blue (likely human) to red (likely bot); yellow nodes cannot be evaluated because they have either been suspended or deleted all their tweets. An interactive version of the larger network is available online (`iunetsci.github.io/HoaxyBots/`). Note that Twitter does not provide data to reconstruct a retweet tree; all retweets point to the original tweet. The retweet networks shown here combine multiple cascades (each a "star network" originating from a different tweet) that all share the same article link.

able from that of fact-checking articles. This result is not as bleak as that of a similar analysis based on only fact-checked claims, which found false news to be even more viral than real news (*12*). However, the qualitative conclusion is the same: massive numbers of people are exposed to low-credibility content.

We observe some anomalous patterns in the spread of low-credibility content. First, although content from low-credibility and fact-checking sources has similar popularity distributions, this is the result of different diffusion mechanisms. Most low-credibility articles spread through original tweets and retweets, while few are shared in replies (Fig. 2(a)); this is different from fact-checking articles, which are shared mainly via retweets but also replies (Fig. 2(b)). In other words, the spreading patterns of low-credibility content are less "conversational." Second, for the most popular articles from low-credibility sources, much of the spreading activity is concentrated around a small portion of accounts (Fig. 2(c)), even though one would expect organic spread among many human users for viral articles. This suggests that the spread is amplified artificially. In fact, a single account can post the same low-credibility article hundreds or even thousands of times (see Supplementary Materials).

We suspect that the "super-spreaders" of low-credibility content are social bots that automatically post links to articles, retweet other accounts, or perform more sophisticated autonomous tasks, like following and replying to other users. To test this hypothesis, we used the Botometer service to evaluate the Twitter accounts that posted links to articles from low-credibility sources. For each account we computed a bot score, which can be interpreted as the level of automation of that account. We classified an accounts as likely bot or human by comparing its bot score to a threshold of 0.5. Details about the Botometer system and the threshold can be found in Methods. We first considered a random sample of accounts that shared at least one link to a low-credibility article. Only 8% of accounts in the sample are labeled as likely bots using this method, but they are responsible for spreading 33% of all tweets linking to low-credibility con-

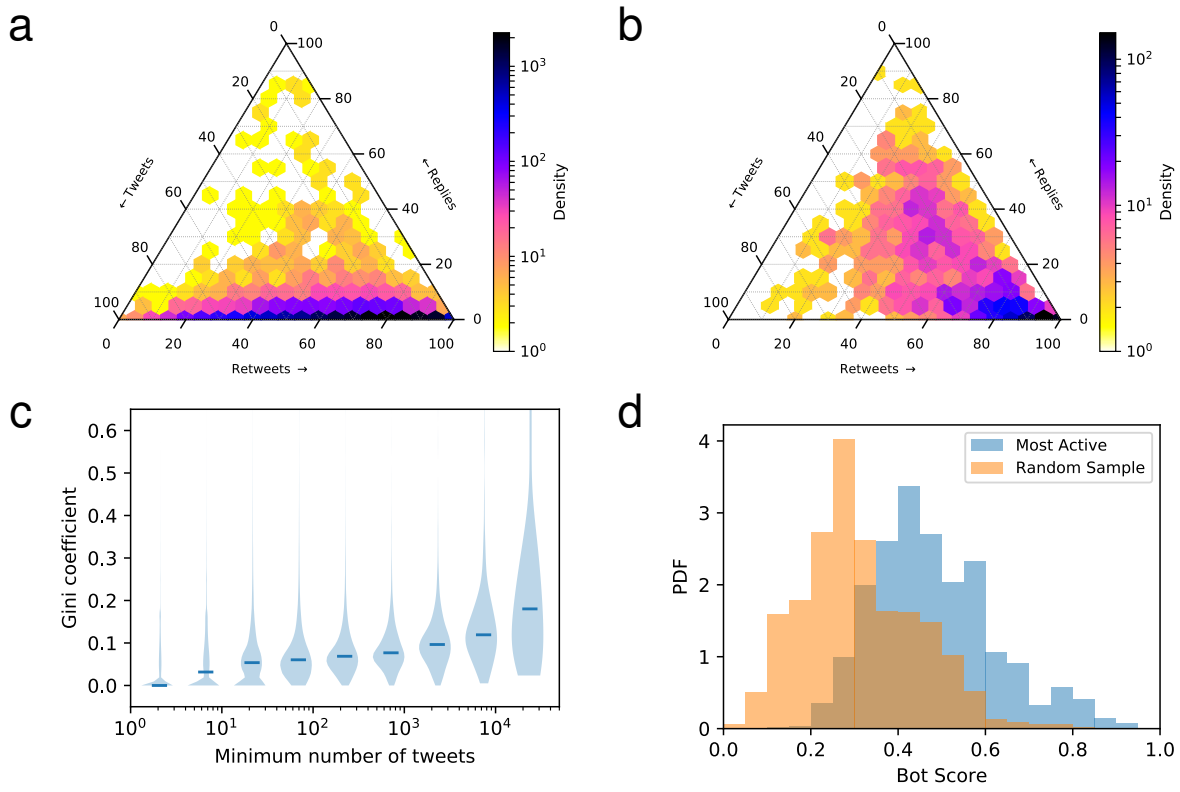

Figure 2: Anomalies. The distribution of types of tweet spreading articles from (a) low-credibility and (b) fact-checking sources are quite different. Each article is mapped along three axes representing the percentages of different types of messages that share it: original tweets, retweets, and replies. When user Alice retweets a tweet by user Bob, the tweet is rebroadcast to all of Alice's followers, whereas when she replies to Bob's tweet, the reply is only seen by Bob. Color represents the number of articles in each bin, on a log-scale. (c) Correlation between popularity of articles from low-credibility sources and concentration of posting activity. We consider a collection of articles shared by a minimum number of tweets as a popularity group. For articles in each popularity group, a violin plot shows the distribution of Gini coefficients, which measure concentration of posts by few accounts (see Supplementary Materials). In violin plots, the width of a contour represents the probability of the corresponding value, and the median is marked by a colored line. (d) Bot score distributions for a random sample of 915 accounts who posted at least one link to a low-credibility source, and for the 961 "super-spreaders" that most actively shared content from low-credibility sources. The two groups have significantly different scores ($p < 10^{-4}$ according to a Mann-Whitney U test): super-spreaders are more likely bots.

tent, and 36% of all articles from low-credibility sources. We then compared this group with a sample of super-spreaders, 38% of which have bot score above 0.5 — almost five times as many (details in Supplementary Materials). Fig. 2(d) presents a further comparison between the two groups, confirming that the super-spreaders are significantly more likely to be bots compared to the general population of users who share low-credibility content. It is important to note that the higher bot scores cannot be attributed to a bias of the Botometer machine learning model based on account activity (see Supplementary Materials).

We hypothesize that these bots play a critical role in driving the viral spread of low-credibility content. To test this conjecture, we examined the different spreading phases of viral articles. In each of these phases we examined the accounts posting these articles. As shown in Fig. 3(a), bots actively share links in the first few seconds after they are first posted. This early intervention exposes many users to low-credibility articles, effectively boosting their viral diffusion.

Another strategy often used by bots is to mention influential users in tweets that link to low-credibility content. Bots seem to employ this targeting strategy repetitively; for example, a single account mentioned `@realDonaldTrump` in 18 tweets linking to a false claim about millions of votes by illegal immigrants (see details in Supplementary Materials). For a systematic investigation, let us consider all tweets in our corpus that mention or reply to a user and include a link to a viral low-credibility story. The number of followers of a Twitter user is often used as a proxy for their influence. Tweets tend to mention popular people, of course. However, Figs. 3(b) shows that when accounts with the highest bot scores share these links, they tend to target users with a higher number of followers (median and average). In this way bots expose influential people, such as journalists and politicians, to an article, creating the appearance that it is widely shared and the chance that the targeted users will spread it.

We examined whether bots (or rather their programmers) tended to target voters in certain states by creating the appearance of users posting articles from those locations. We did find

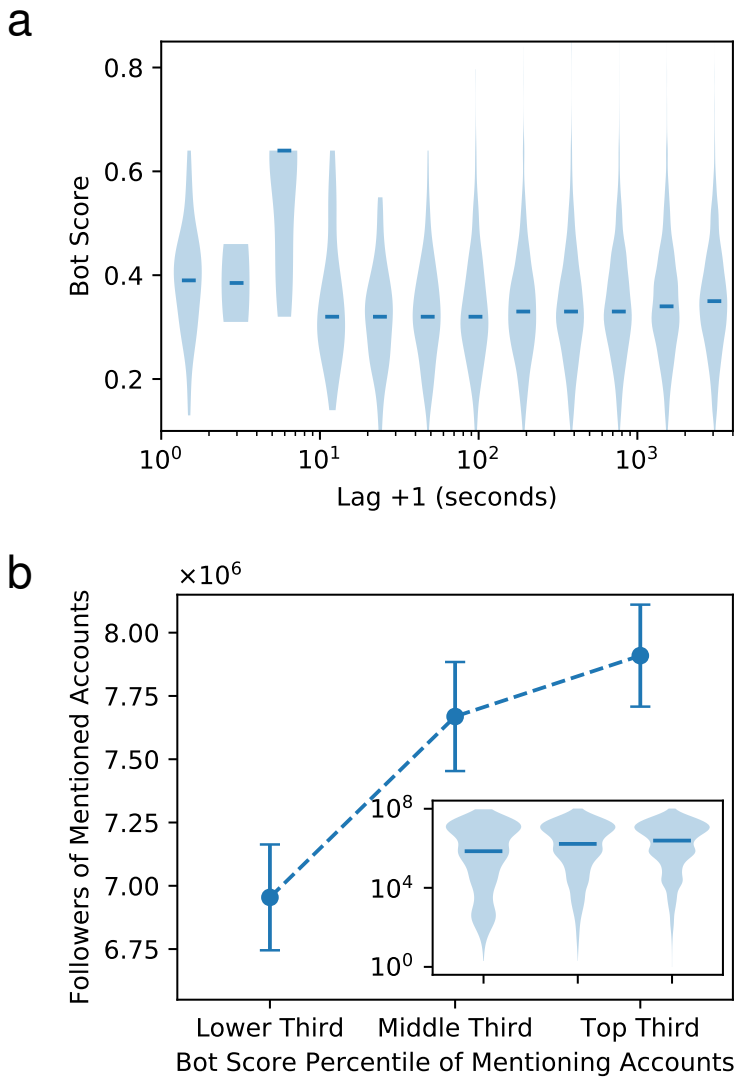

Figure 3: Bot strategies. (a) Early bot support after a viral low-credibility article is first shared. We consider a sample of 60,000 accounts that participate in the spread of the 1,000 most viral stories from low-credibility sources. We align the times when each article first appears. We focus on a one-hour early spreading phase following each of these events, and divide it into logarithmic lag intervals. The plot shows the bot score distribution for accounts sharing the articles during each of these lag intervals. (b) Targeting influentials by likely bots. We plot the average number of followers of Twitter users who are mentioned (or replied to) by accounts that link to the most viral 1000 stories. The mentioning accounts are aggregated into three groups by bot score percentile. Error bars indicate standard errors. Inset: Distributions of follower counts for users mentioned by accounts in each percentile group.

that the location patterns produced by bots are inconsistent with the geographic distribution of conversations on Twitter. However, we did not find evidence that bots sharing articles from low-credibility sources targeted swing states. Details of this analysis can be found in Supplementary Materials.

Having found that bots are employed to drive the viral spread of low-credibility articles, let us explore how humans interact with the content shared by bots, which may provide insight into whether and how bots are able to affect public opinion. Fig. 4 shows that humans do most of the retweeting (upper panel), and they retweet articles posted by bots as much as by other humans (left panel). This suggests that collectively, people do not discriminate between low-credibility content shared by humans versus social bots. It also means that when we observe many accounts exposed to low-credibility information, these are not just bots (re)tweeting. In fact, we specifically looked at likely bots vs. likely humans tweeting links to low-credibility sources, and found that the human reach of these articles is *amplified* by social bots: the number of tweets by likely humans linking to stories from low-credibility sources grows super-linearly with the number of tweets by likely bots linking to the same articles. The same amplification effect is not observed for articles from fact-checking sources. Details are presented in Supplementary Materials.

Finally, we compared the extent to which social bots successfully manipulate the information ecosystem in support of different low-credibility sources. We considered the most popular sources in terms of median and aggregate article posts, and measured the bot scores of the accounts that most actively spread their content. As shown in Fig. 5, one site (`beforeitsnews.com`) stands out in terms of manipulation, but other popular low-credibility sources also have many bots among their promoters. Satire sites like *The Onion* and fact-checking websites do not display the same level of bot support.

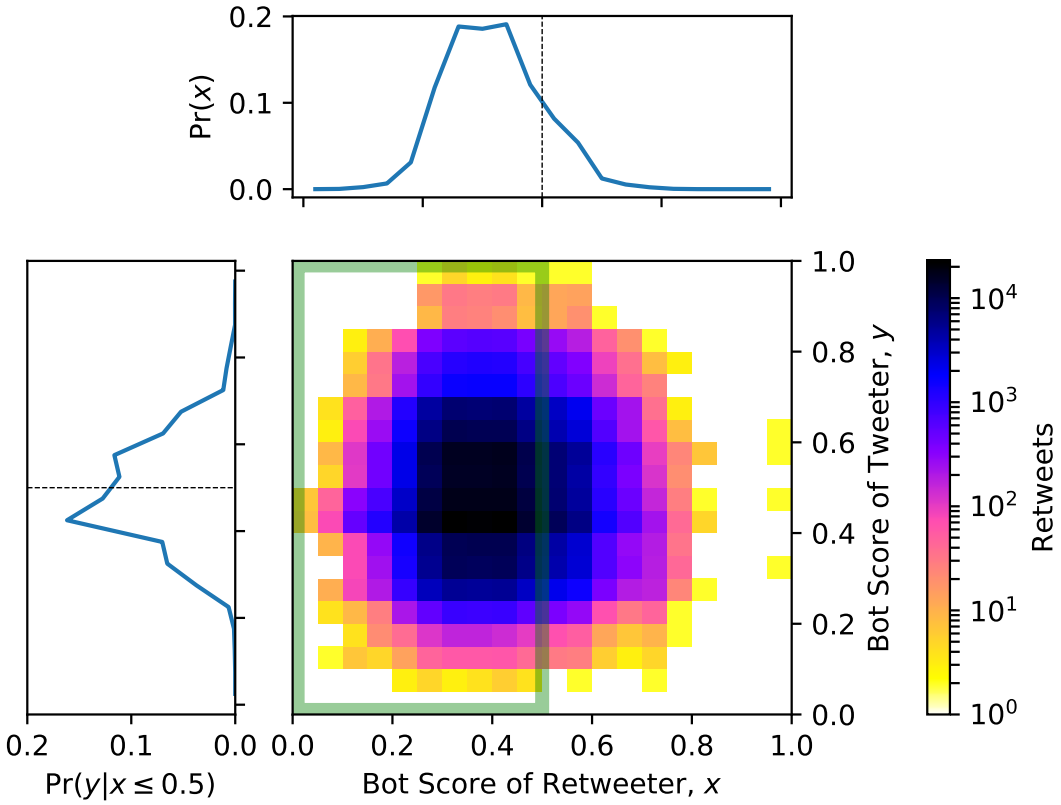

Figure 4: Impact of bots. We plot the joint distribution of the bot scores of accounts that retweeted links to low-credibility articles and accounts that had originally posted the links. Color represents the number of retweeted messages in each bin, on a log scale. The top projection shows the distributions of bot scores for retweeters, who are mostly human. The left projection shows the distributions of bot scores for accounts retweeted by likely humans, with a significant portion of likely bots.

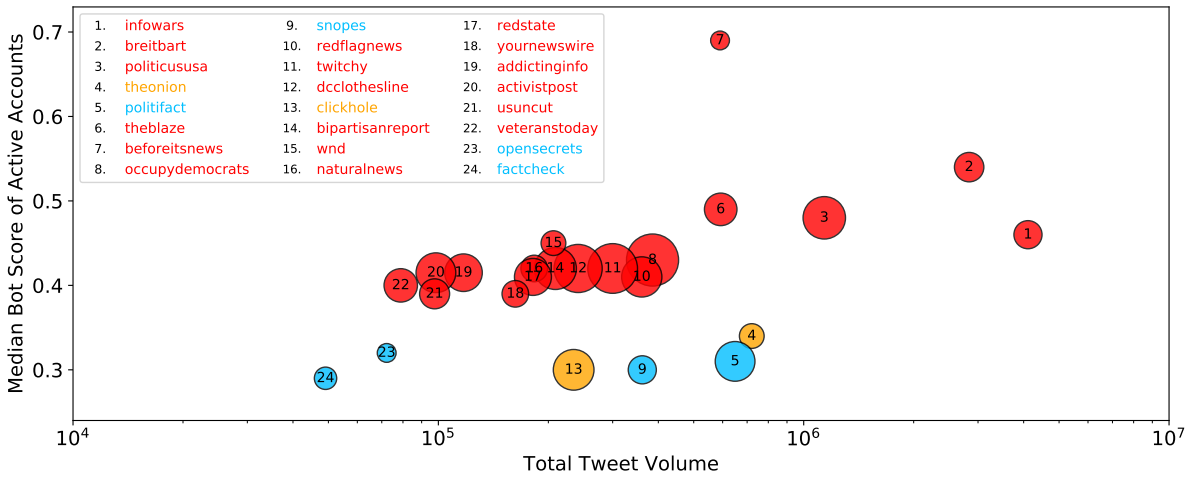

Figure 5: Popularity and bot support for the top sources. Satire websites are shown in orange, fact-checking sites in blue, and low-credibility sources in red. Popularity is measured by total tweet volume (horizontal axis) and median number of tweets per article (circle area). Bot support is gauged by the median bot score of the 100 most active accounts posting links to articles from each source (vertical axis). Low-credibility sources have greater support by bots, as well as greater median and/or total volume in many cases.

Discussion

Our analysis provides quantitative empirical evidence of the key role played by social bots in the viral spread of low-credibility content. Relatively few accounts are responsible for a large share of the traffic that carries misinformation. These accounts are likely bots, and we uncovered two manipulation strategies they use. First, bots are particularly active in amplifying content in the very early spreading moments, before an article goes viral. Second, bots target influential users through replies and mentions. People are vulnerable to these kinds of manipulation, retweeting bots who post low-credibility content just as much as they retweet other humans. As a result, bots amplify the reach of low-credibility content, to the point that it is statistically indistinguishable from that of fact-checking articles. Successful low-credibility sources in the U.S., including those on both ends of the political spectrum, are heavily supported by social bots. Social media

platforms are beginning to acknowledge these problems and deploy countermeasures, although their effectiveness is hard to evaluate (*10, 33, 34*).

The present findings are not in contradiction with the recent work by Vosoughi *et al.* (*12*). Their analysis is based on a small subset of articles that are fact-checked, whereas the present work considers a much broader set of articles from low-credibility sources, most of which are not fact-checked. In addition, the analysis of Vosoughi *et al.* neglected an important mechanism by which bots can amplify the spread of an article, namely, by resharing links originally posted by human accounts. Because of these two methodological differences, the present analysis provides new evidence about the role played by bots.

Our results are *robust* with respect to various choices. First, using a more restrictive criterion for selecting low-credibility sources, based on a consensus among several news and fact-checking organizations (see Methods), yields qualitatively similar results, leading to the same conclusions. Second, our analysis about active spreaders of low-credibility content being likely bots is robust with respect to the activity threshold used to identify the most active spreaders. Furthermore, activity and bot scores are uncorrelated with account activity volume. Third, the conclusions are not affected by the use of different bot-score thresholds to separate social bots and human accounts. Details about all of these robustness analyses can be found in Supplementary Materials.

Our findings demonstrate that social bots are an effective tool to manipulate social media. While the present study focuses on the spread of low-credibility content, such as false news, conspiracy theories, and junk science, similar bot strategies may be used to spread other types of malicious content, such as malware. And although our spreading data is collected from Twitter, there is no reason to believe that the same kind of abuse is not taking place on other digital platforms as well. In fact, viral conspiracy theories spread on Facebook (*35*) among the followers of pages that, like social bots, can easily be managed automatically and anonymously.

Furthermore, just like on Twitter, false claims on Facebook are as likely to go viral as reliable news (*11*). While the difficulty to access spreading data on platforms like Facebook is a concern, the growing popularity of ephemeral social media like Snapchat may make future studies of this abuse all but impossible.

The results presented here suggest that curbing social bots may be an effective strategy for mitigating the spread of online misinformation. Progress in this direction may be accelerated through partnerships between social media platforms and academic research (*10*). For example, our lab and others are developing machine learning algorithms to detect social bots (*9, 25, 26*). The deployment of such tools is fraught with peril, however. While platforms have the right to enforce their terms of service, which forbid impersonation and deception, algorithms do make mistakes. Even a single false-positive error leading to the suspension of a legitimate account may foster valid concerns about censorship. This justifies current human-in-the-loop solutions, which unfortunately do not scale with the volume of abuse that is enabled by software. It is therefore imperative to support research both on improved abuse detection algorithms and on countermeasures that take into account the complex interplay between cognitive and technological factors that favor the spread of misinformation (*36*).

An alternative strategy would be to employ CAPTCHAs (*37*), challenge-response tests to determine whether a user is human. CAPTCHAs have been deployed widely and successfully to combat email spam and other types of online abuse. Their use to limit automatic posting or resharing of news links could help stem bot abuse by increasing its cost, but also add undesirable friction to benign applications of automation by legitimate entities, such as news media and emergency response coordinators. These are hard trade-offs that must be studied carefully as we contemplate ways to address the fake news epidemics.

Methods

The online article-sharing data was collected through Hoaxy, an open platform developed at Indiana University to track the spread of claims and fact checking on Twitter (*32*). A search engine, interactive visualizations, and open-source software are freely available (`hoaxy.iuni.iu.edu`). The data is accessible through a public API. Further details are presented in Supplementary Materials.

We started the collection in mid-May 2016 with 70 low-credibility sites and added 50 more in mid-December 2016. The collection period for the present analysis extends until the end of March 2017. During this time, we collected 389,569 articles from these 120 sites. We also tracked 15,053 stories published by independent fact-checking organizations, such as `snopes.com`, `politifact.com`, and `factcheck.org`. We did not exclude satire because many fake-news sources label their content as satirical, and viral satire is often mistaken for real news. *The Onion* is the satirical source with the highest total volume of shares. We repeated our analyses of most viral articles (e.g., Fig. 3(a)) with articles from `theonion.com` excluded and the results were not affected.

The full list of 120 sources is reported in Supplementary Materials. We also repeated the analysis using a more restrictive criterion for selecting low-credibility sources, based on a consensus among three or more news and fact-checking organizations. This yields 327,840 articles (86% of the total) from 65 low-credibility sources, also listed in Supplementary Materials, where we show that the results are robust with respect to these different source selection criteria.

Our analysis does not require a complete list of low-credibility sources, but does rely on the assumption that many articles published by these sources can be classified as some kind of misinformation or unsubstantiated information. To validate this assumption, we checked the content of a random sample of articles. For the purpose of this verification, we adopted

a definition of "misinformation" that follows industry convention and includes the following classes: fabricated content, manipulated content, imposter content, false context, misleading content, false connection, and satire (*38*). To these seven categories we also added articles whose claims could not be verified. We found that fewer that 15% of articles could be verified. More details are available in Supplementary Materials.

Using the filtering endpoint of Twitter's public streaming API, we collected 13,617,425 public posts that included links to articles from low-credibility sources and 1,133,674 public posts linking to fact checks. This is the *complete* set of tweets linking to these articles in the study period, rather than a sample (see Supplementary Materials for details). We extracted metadata about the source of each link, the account that shared it, the original poster in case of retweet or quoted tweet, and any users mentioned or replied to in the tweet.

We transformed links to canonical URLs to merge different links referring to the same article. This happens mainly due to shortening services (44% links are redirected) and extra parameters (34% of URLs contain analytics tracking parameters), but we also found websites that use duplicate domains and snapshot services. Canonical URLs were obtained by resolving redirection and removing analytics parameters.

The bot score of Twitter accounts is computed using the Botometer service, which evaluates the extent to which an account exhibits similarity to the characteristics of social bots (*25*). The system is based on a supervised machine learning algorithm leveraging more than a thousand features extracted from public data and meta-data about Twitter accounts. These features include various descriptors of information diffusion networks, user metadata, friend statistics, temporal patterns of activity, part-of-speech and sentiment analysis. The classifier is trained using publicly available datasets of tens of thousands of Twitter users that include both humans and bots of varying sophistication. The Botometer system is available through a public API (`botometer.iuni.iu.edu`). It has also been employed in other studies (*12, 39*) and is

widely adopted, serving hundreds of thousand requests daily.

For the present analysis, we use the Twitter Search API to collect up to 200 of an account's most recent tweets and up to 100 of the most recent tweets mentioning the account. From this data we extract the features used by the Botometer classifier.

There is no crisp, binary classification of accounts as human or bot, as there are many types of bots and humans using different levels of automation. Accordingly, Botometer provides a score (percentage) rather than a binary classification. Nevertheless, the model can effectively discriminating between human and bot accounts of different nature; five-fold cross-validation yields an area under the ROC curve of 94% (*25*). A value of 50% indicates random accuracy and 100% means perfect accuracy. When a binary classification is needed, we use a threshold of 0.5, which maximizes accuracy (*25*). See Supplementary Materials for further details about bot classification and its robustness.

In the targeting analysis (Fig. 3(b)), we exclude mentions of sources using the pattern "via @screen_name."

References

1. J. Gottfried, E. Shearer, News use across social media platforms 2016, *White paper*, Pew Research Center (2016).

2. B. Markines, C. Cattuto, F. Menczer, *Proc. 5th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)* (2009).

3. E. Mustafaraj, P. T. Metaxas, *Proc. Web Science Conference: Extending the Frontiers of Society On-Line* (2010).

4. J. Ratkiewicz, *et al.*, *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2011).

5. L. Howell, *et al.*, *Global Risks* (World Economic Forum, 2013).

6. L. Gu, V. Kropotov, F. Yarochkin, The fake news machine: How propagandists abuse the internet and manipulate the public, *Trendlabs research paper*, Trend Micro (2017).

7. H. Allcott, M. Gentzkow, *Journal of Economic Perspectives* **31**, 211 (2017).

8. P. J. Hotez, *PLOS Medicine* **13**, 1 (2016).

9. E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, *Comm. ACM* **59**, 96 (2016).

10. D. Lazer, *et al.*, *Science* **359**, 1094 (2018).

11. X. Qiu, D. F. M. Oliveira, A. S. Shirazi, A. Flammini, F. Menczer, *Nature Human Behavior* **1**, 0132 (2017).

12. S. Vosoughi, D. Roy, S. Aral, *Science* **359**, 1146 (2018).

13. C. R. Sunstein, *Going to Extremes: How Like Minds Unite and Divide* (Oxford University Press, 2009).

14. E. Pariser, *The filter bubble: How the new personalized Web is changing what we read and how we think* (Penguin, 2011).

15. D. Nikolov, D. F. M. Oliveira, A. Flammini, F. Menczer, *PeerJ Computer Science* **1** (2015).

16. M. D. Conover, B. Gonçalves, A. Flammini, F. Menczer, *EPJ Data Science* **1**, 6 (2012).

17. M. Conover, *et al.*, *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)* (2011).

18. M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854 (2006).

19. N. O. Hodas, K. Lerman, *Proc. ASE/IEEE International Conference on Social Computing* (2012).

20. A. Nematzadeh, G. L. Ciampaglia, F. Menczer, A. Flammini, How algorithmic popularity bias hinders or promotes quality, *Preprint 1707.00574*, arXiv (2017).

21. N. Stroud, *Niche News: The Politics of News Choice* (Oxford University Press, 2011).

22. D. M. Kahan, *Judgment and Decision Making* **8**, 407 (2013).

23. M. S. Levendusky, *American Journal of Political Science* **57**, 611 (2013).

24. W. Lippmann, *Public Opinion* (Harcourt, Brace and Company, 1922).

25. O. Varol, E. Ferrara, C. A. Davis, F. Menczer, A. Flammini, *Proc. Intl. AAAI Conf. on Web and Social Media (ICWSM)* (2017).

26. V. Subrahmanian, *et al.*, *IEEE Computer* **49**, 38 (2016).

27. Y. Jun, R. Meng, G. V. Johar, *Proceedings of the National Academy of Sciences* **114**, 5976 (2017).

28. T. Jagatic, N. Johnson, M. Jakobsson, F. Menczer, *Communications of the ACM* **50**, 94 (2007).

29. A. Bessi, E. Ferrara, *First Monday* **21** (2016).

30. S. C. Woolley, P. N. Howard, Computational propaganda worldwide: Executive summary, *Working Paper 2017.11*, Oxford Internet Institute (2017).

31. E. Ferrara, *First Monday* **22** (2017).

32. C. Shao, *et al.*, *PLoS ONE* **13**, e0196087 (2018).

33. J. Weedon, W. Nuland, A. Stamos, Information operations and facebook, *white paper*, Facebook (2017).

34. A. Mosseri, News feed fyi: Showing more informative links in news feed, *press release*, Facebook (2017).

35. M. Del Vicario, *et al.*, *Proc. National Academy of Sciences* **113**, 554 (2016).

36. S. Lewandowsky, U. K. Ecker, J. Cook, *Journal of Applied Research in Memory and Cognition* **6**, 353 (2017).

37. L. von Ahn, M. Blum, N. J. Hopper, J. Langford, *Advances in Cryptology — Proceedings of EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques*, E. Biham, ed. (Springer, 2003), pp. 294–311.

38. C. Wardle, Fake news. It's complicated., *White paper*, First Draft News (2017).

39. S. Wojcik, S. Messing, A. Smith, L. Rainie, P. Hitlin, Bots in the twittersphere, *White paper*, Pew Research Center (2018).

Acknowledgments

We are grateful to Ben Serrette and Valentin Pentchev of the Indiana University Network Science Institute (`iuni.iu.edu`), as well as Lei Wang for supporting the development of the Hoaxy platform. Clayton A. Davis developed the Botometer API. Nic Dias provided assistance with claim verification. We are also indebted to Twitter for providing data through their API. C.S. thanks the Center for Complex Networks and Systems Research (`cnets.indiana.edu`) for the hospitality during his visit at the Indiana University School of Informatics and Computing. He was supported by the China Scholarship Council. G.L.C. was supported by

IUNI. The development of the Botometer platform was supported in part by DARPA (grant W911NF-12-1-0037). The development of the Hoaxy platform was supported in part by the Democracy Fund. A.F. and F.M. were supported in part by the James S. McDonnell Foundation (grant 220020274) and the National Science Foundation (award CCF-1101743). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.