**Abstract**

The massive spread of digital misinformation has been identified as a major threat to democracies. Communication, cognitive, social, and computer scientists are studying the complex causes for the viral diffusion of misinformation, while online platforms are beginning to deploy countermeasures. Little systematic, data-based evidence has been published to guide these efforts. Here we analyze 14 million messages spreading 400 thousand articles on Twitter during ten months in 2016 and 2017. We find evidence that social bots played a disproportionate role in spreading articles from low-credibility sources. Bots amplify such content in the early spreading moments, before an article goes viral. They also target users with many followers through replies and mentions. Humans are vulnerable to this manipulation, resharing content posted by bots. 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**

As we access news from social media[1](https://www.nature.com/articles/s41467-018-06930-7#ref-CR1), we are exposed to a daily dose of false or misleading news reports, hoaxes, conspiracy theories, click-bait headlines, junk science, and even satire. We refer to such content collectively as “misinformation.” The financial incentives through advertising are well understood, but political motives can be equally powerful. The massive spread of digital misinformation has been identified as a major global risk and alleged to influence elections and threaten democracies. While such claims are hard to prove, real harm of disinformation has been demonstrated in health and finance.

Social and computer scientists are engaged in efforts to study the complex mix of cognitive, social, and algorithmic biases that make us vulnerable to manipulation by online misinformation. These include information overload and finite attention, novelty of false news, the selective exposure caused by polarized and segregated online social networks, algorithmic popularity bias, and other cognitive vulnerabilities such as confirmation bias and motivated reasoning.

Abuse of online information ecosystems can both exploit and reinforce these vulnerabilities. While fabricated news are not a new phenomenon, the ease with which social media can be manipulated 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, known as social bots. These fake accounts can post content and interact with each other and with legitimate users via social connections, just like real people. 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, and to trust social contacts. Since earliest manifestations uncovered in 2010 , we have seen influential bots affect online debates about vaccination policies and participate actively in political campaigns, both in the United Statesand other countries.

The fight against online misinformation requires a grounded assessment of the relative impact of different mechanisms 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. With few exception, the literature about the role played by social bots in the spread of misinformation is largely 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 misinformation by social bots is now feasible thanks to two tools developed in our lab: the Hoaxy platform to track the online spread of claims and the Botometer machine learning algorithm to detect social bots. Here we examine social bots and how they promote the spread of misinformation through millions of Twitter posts during and following the 2016 US presidential campaign. We find that social bots amplify the spread of misinformation by exposing humans to this content and inducing them to share it.

**Low-credibility content**

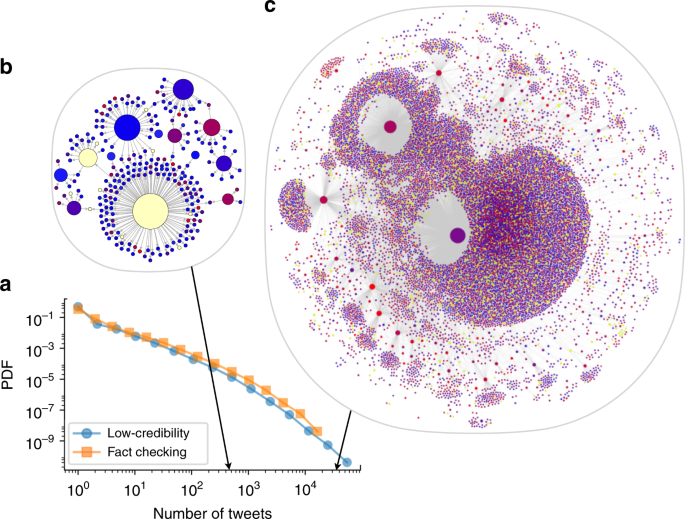
Our analysis is based on a large corpus of news stories posted on Twitter. Operationally, rather than focusing on individual stories that have been debunked by fact-checkers, we consider low-credibility content, i.e., content from low-credibility sources. Such sources are websites that have been identified by reputable third-party news and fact-checking organizations as routinely publishing various types of low-credibility information (see Methods). There are two reasons for this approach. First, these sources have processes for the publication of disinformation: they mimic news media outlets without adhering to the professional standards of journalistic integrity. Second, fact-checking millions of individual articles is unfeasible. As a result, this approach is widely adopted in the literature (see Supplementary Discussion).

We track the complete production of 120 low-credibility sources by crawling their websites and extracting all public tweets with links to their stories. Our own analysis of a sample of these articles confirms that the vast majority of low-credibility content is some type of misinformation (see Methods). We also crawled and tracked 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 from Twitter all of the public posts linking to these articles: 13,617,425 tweets linked to low-credibility sources and 1,133,674 linked to fact-checking sources. See Methods and Supplementary Methods for details.

**Spreading patterns and actors**

On average, a low-credibility source published approximately 100 articles per week. By the end of the study period, the mean popularity of those articles was approximately 30 tweets per article per week (see Supplementary Fig. [1](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)). However, as shown in Fig. [1](https://www.nature.com/articles/s41467-018-06930-7#Fig1), success is extremely heterogeneous across articles. Whether we measure success by number of posts containing a link (Fig. [1a](https://www.nature.com/articles/s41467-018-06930-7#Fig1)) or by number of accounts sharing an article (Supplementary Fig. [2](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)), we find a very broad distribution of popularity spanning several orders of magnitude: while the majority of articles goes unnoticed, a significant fraction goes “viral.” We observe that the popularity distribution of low-credibility articles is almost indistinguishable from that of fact-checking articles, meaning that low-credibility content is equally or more likely to spread virally. This result is similar to that of an analysis based on only fact-checked claims, which found false news to be even more viral than real news[2](https://www.nature.com/articles/s41467-018-06930-7#ref-CR2). The qualitative conclusion is the same: links to low-credibility content reach massive exposure.

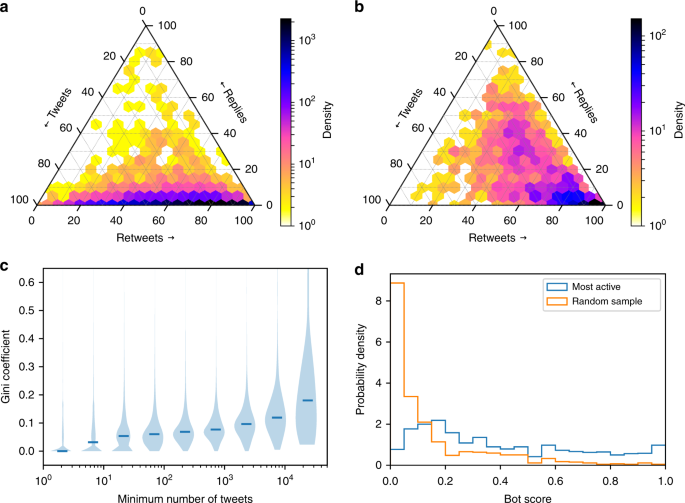
**Fig. 1**

[](https://www.nature.com/articles/s41467-018-06930-7/figures/1)

Online virality of content. **a** Probability distribution (density function) of the number of tweets for articles from both low-credibility (blue circles) and fact-checking (orange squares) sources. The distributions of the number of accounts sharing an article are very similar (see Supplementary Fig. [2](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)). 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 US election and shared in over 400 tweets; and **c** a highly viral fabricated news report titled “Spirit cooking”: Clinton campaign chairman practices bizarre occult ritual, published 4 days before the 2016 US election and shared in over 30,000 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 was 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/](http://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

Even though low-credibility and fact-checking sources show similar popularity distributions, we observe some distinctive patterns in the spread of low-credibility content. First, most articles by low-credibility sources spread through original tweets and retweets, while few are shared in replies (Fig. [2a](https://www.nature.com/articles/s41467-018-06930-7#Fig2)); this is different from articles by fact-checking sources, which are shared mainly via retweets but also replies (Fig. [2b](https://www.nature.com/articles/s41467-018-06930-7#Fig2)). In other words, the spreading patterns of low-credibility content are less “conversational.” Second, the more a story was tweeted, the more the tweets were concentrated in the hands of few accounts, who act as “super-spreaders” (Fig. [2c](https://www.nature.com/articles/s41467-018-06930-7#Fig2)). This goes against the intuition that, as a story reaches a broader audience organically, the contribution of any individual account or group of accounts should matter less. In fact, a single account can post the same low-credibility article hundreds or even thousands of times (see Supplementary Fig. [6](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)). This could suggest that the spread is amplified through automated means.

**Fig. 2**

[](https://www.nature.com/articles/s41467-018-06930-7/figures/2)

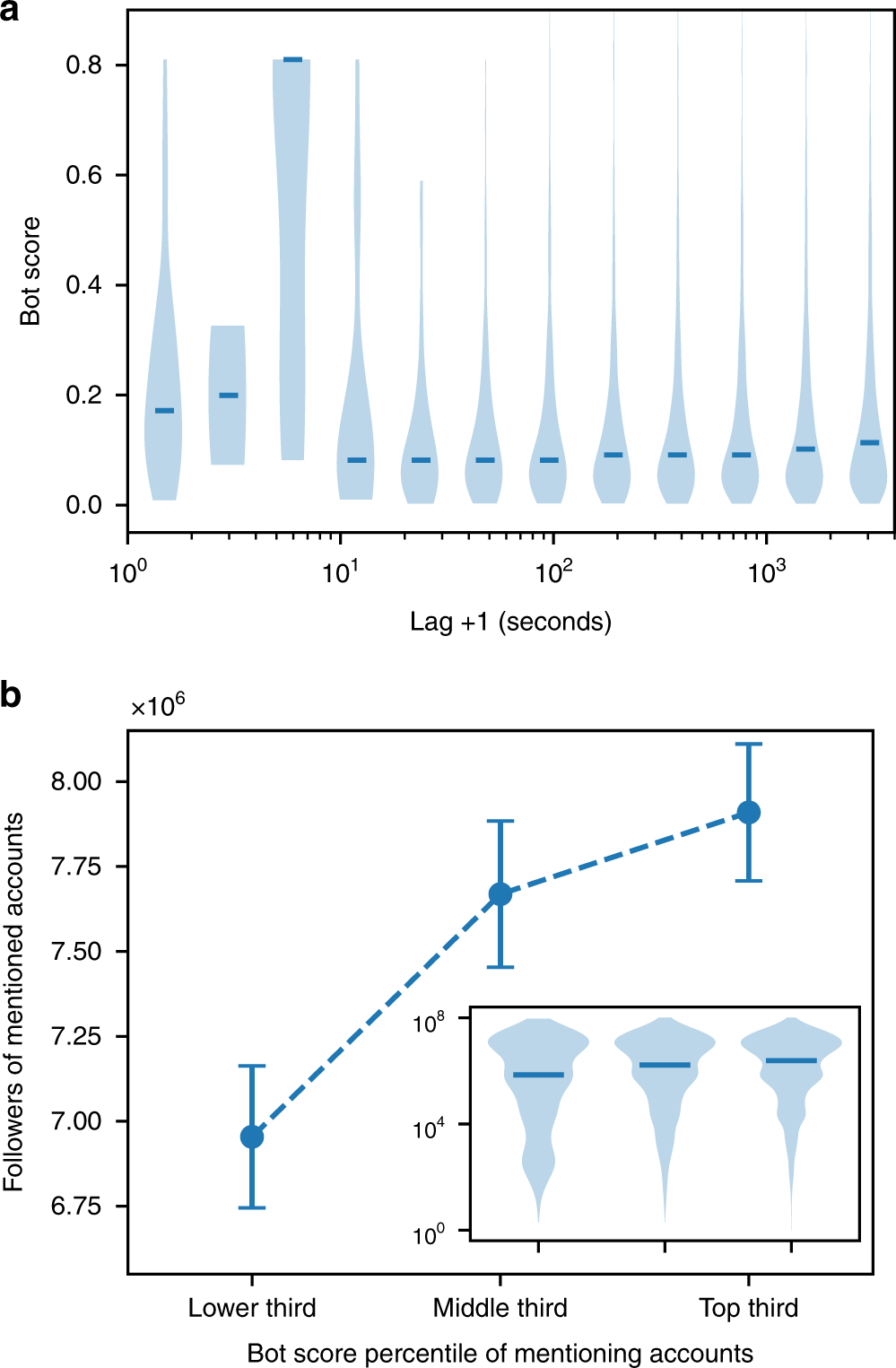
Anomalies. The distribution of types of tweet spreading articles from **a** low-credibility and **b** fact-checking sources are quite different. Each article is mapped along three axes representing the percentages of different types of messages that share it: original tweets, retweets, and replies. When 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 and users who follow them both. 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 Methods). 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 (orange), and for the 961 “super-spreaders” that most actively shared content from low-credibility sources (blue). The two groups have significantly different scores (*p* < 10−4 according to the Mann–Whitney *U* test): super-spreaders are more likely bots

We hypothesize that the “super-spreaders” of low-credibility content are social bots which are automatically posting links to articles, retweeting other accounts, or performing more sophisticated autonomous tasks, like following and replying to other users. To test this hypothesis, we used Botometer to evaluate the Twitter accounts that posted links to articles from low-credibility sources. For each account we computed a bot score (a number in the unit interval) which can be interpreted as the level of automation of that account. We used a threshold of 0.5 to classify an account as bot or human. Details about the Botometer system and the threshold can be found in Methods. We first considered a random sample of the general population of accounts that shared at least one link to a low-credibility article. Only 6% of accounts in the sample are labeled as bots using this method, but they are responsible for spreading 31% of all tweets linking to low-credibility content, and 34% of all articles from low-credibility sources (Supplementary Table [2](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)). We then compared this group with a sample of the top most active accounts (“super-spreaders”), 33% of which have been labeled as bot—over five times as many (details in Supplementary Methods). Figure [2d](https://www.nature.com/articles/s41467-018-06930-7#Fig2) confirms that the super-spreaders are significantly more likely to be bots compared to the general population of accounts who share low-credibility content. Because these results are based on a classification model, it is important to make sure that what we see in Fig. [2d](https://www.nature.com/articles/s41467-018-06930-7#Fig2) is not due to bias in the way Botometer was trained—that the model did not simply learn to assign higher scores to more active accounts. We rule out this competing explanation by showing that higher bot scores cannot be attributed to this kind of bias in the learning model (see Supplementary Fig. [16](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)).

**Bot strategies**

Given this evidence, we submit that bots may play a critical role in driving the viral spread of content from low-credibility sources. To test this question, we examined whether bots tend to get involved at particular times in the spread of popular articles. As shown in Fig. [3a](https://www.nature.com/articles/s41467-018-06930-7#Fig3), likely bots are more prevalent in the first few seconds after an article is first published on Twitter than at later times. We conjecture that this early intervention exposes many users to low-credibility articles, increasing the chances than an article goes “viral.”

**Fig. 3**



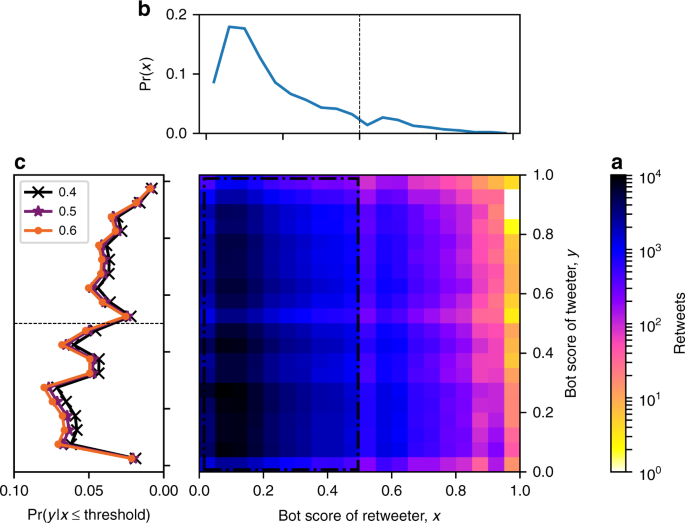
Bot strategies. **a** Early bot support after a viral low-credibility article is first shared. We consider a sample of 60,000 accounts that participate in the spread of the 1000 most viral stories from low-credibility sources. We align the times when each article first appears. We focus on the 1 h early 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 of influentials. 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

We find that 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 in 19 tweets, each linking to the same false claim about millions of votes by illegal immigrants (see details in Supplementary Discussion and Supplementary Fig. [7](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)). For a systematic investigation, let us consider all tweets that mention or reply to a user and include a link to a viral article from a low-credibility source in our corpus. The number of followers is often used as a proxy for the influence of a Twitter user. As shown in Fig. [3b](https://www.nature.com/articles/s41467-018-06930-7#Fig3), in general tweets tend to mention popular people. However, accounts with the largest bot scores tend to mention users with a larger number of followers (median and average). A possible explanation for this strategy is that bots (or rather, their operators) target influential users with content from low-credibility sources, creating the appearance that it is widely shared. The hope is that these targets will then reshare the content to their followers, thus boosting its credibility.

**Bot impact**

Having found that automated accounts are employed in ways that appear 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. Figure [4a](https://www.nature.com/articles/s41467-018-06930-7#Fig4) shows who retweets whom: humans do most of the retweeting (Fig. [4b](https://www.nature.com/articles/s41467-018-06930-7#Fig4)), and they retweet articles posted by likely bots almost as much as those by other humans (Fig. [4c](https://www.nature.com/articles/s41467-018-06930-7#Fig4)). This result, which is robust to the choice of threshold used to identify likely humans, 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 it. In fact, we find that the volume of tweets by likely humans scales super-linearly with the volume by likely bots, suggesting that the reach of these articles among humans is amplified by social bots. In other words, each amount of sharing activity by likely bots tends to trigger a disproportionate amount of human engagement. The same amplification effect is not observed for articles from fact-checking sources. Details are presented in Supplementary Discussion (Supplementary Figs. [8](https://www.nature.com/articles/s41467-018-06930-7#MOESM1), [9](https://www.nature.com/articles/s41467-018-06930-7#MOESM1)).

**Fig. 4**

[](https://www.nature.com/articles/s41467-018-06930-7/figures/4)

Impact of bots on humans.

Joint distribution of 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 distribution of bot scores for retweeters, who are mostly human.

The left projection shows the distribution of bot scores for accounts retweeted by likely humans who are identified by scores below a threshold of 0.4 (black crosses), 0.5 (purple stars), or 0.6 (orange circles). Irrespective of the threshold, we observe a significant portion of likely bots retweeted by likely humans