

RESEARCH ARTICLE

Anatomy of an online misinformation network

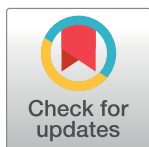
Chengcheng Shao^{1,2✉*}, Pik-Mai Hui^{2✉}, Lei Wang², Xinwen Jiang³, Alessandro Flammini^{2‡}, Filippo Menczer^{2‡}, Giovanni Luca Ciampaglia^{4‡}

1 College of Computer, National University of Defense Technology, Changsha, Hunan, China, **2** School of Informatics, Computing, and Engineering, Indiana University, Bloomington, United States of America, **3** The MOE Key Laboratory of Intelligent Computing and Information Processing, Xiangtan University, Xiangtan, Hunan, China, **4** Indiana University Network Science Institute, Bloomington, United States of America

✉ These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

* sccotte@gmail.com



OPEN ACCESS

Citation: Shao C, Hui P-M, Wang L, Jiang X, Flammini A, Menczer F, et al. (2018) Anatomy of an online misinformation network. PLoS ONE 13(4): e0196087. <https://doi.org/10.1371/journal.pone.0196087>

Editor: Alain Barrat, Centre de physique theorique, FRANCE

Received: January 20, 2018

Accepted: April 5, 2018

Published: April 27, 2018

Copyright: © 2018 Shao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the analyses presented in this paper can be replicated by collecting data through the Hoaxy API (<https://market.mashape.com/truthy/hoaxy>) or downloading the network dataset at doi.org/10.5281/zenodo.1154840.

Funding: C.S. was supported by the China Scholarship Council. X.J. was supported in part by the National Natural Science Foundation of China (No. 61272010). G.L.C. was supported by Indiana University Network Science Institute. The development of the Botometer platform was

Abstract

Massive amounts of fake news and conspiratorial content have spread over social media before and after the 2016 US Presidential Elections despite intense fact-checking efforts. How do the spread of misinformation and fact-checking compete? What are the structural and dynamic characteristics of the core of the misinformation diffusion network, and who are its main purveyors? How to reduce the overall amount of misinformation? To explore these questions we built Hoaxy, an open platform that enables large-scale, systematic studies of how misinformation and fact-checking spread and compete on Twitter. Hoaxy captures public tweets that include links to articles from low-credibility and fact-checking sources. We perform *k*-core decomposition on a diffusion network obtained from two million retweets produced by several hundred thousand accounts over the six months before the election. As we move from the periphery to the core of the network, fact-checking nearly disappears, while social bots proliferate. The number of users in the main core reaches equilibrium around the time of the election, with limited churn and increasingly dense connections. We conclude by quantifying how effectively the network can be disrupted by penalizing the most central nodes. These findings provide a first look at the anatomy of a massive online misinformation diffusion network.

Introduction

The viral spread of online misinformation is emerging as a major threat to the free exchange of opinions, and consequently to democracy. Recent Pew Research Center surveys found that 63% of Americans do not trust the news coming from social media, even though an increasing majority of respondents uses social media to get the news on a regular basis (67% in 2017, up from 62% in 2016). Even more disturbing, 64% of Americans say that fake news have left them with a great deal of confusion about current events, and 23% also admit to passing on fake news stories to their social media contacts, either intentionally or unintentionally [1, 2, 3].

supported in part by DARPA (grant W911NF-12-1-0037) and 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.

Competing interests: The authors have declared that no competing interests exist.

Misinformation is an instance of the broader issue of abuse of social media platforms, which has received a lot of attention in the recent literature [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. The traditional method to cope with misinformation is to fact-check claims. Even though some are pessimistic about the effectiveness of fact-checking, the evidence is still conflicting on the issue [16, 17]. In experimental settings, perceived social presence reduces the propensity to fact-check [18]. An open question is whether this finding translates to the online setting, which would affect the competition between low- and high-quality information. This question is especially pressing. Even though algorithmic recommendation may promote quality under certain conditions [19], models and empirical data show that high-quality information does not have a significant advantage over low-quality information in online social networks [20, 15].

Technology platforms, journalists, fact checkers, and policymakers are debating how to combat the threat of misinformation [21]. A number of systems, tools, and datasets have been proposed to support research efforts about misinformation. Mitra and Gilbert, for example, proposed CRED BANK, a dataset of tweets with associated credibility annotations [22]. Hassan *et al.* [23] built a corpus of political statements worthy of fact-checking using a machine learning approach. Some systems let users visualize the spread of rumors online. The most notable are TwitterTrails [24] and RumorLens [25]. These systems, however, lack monitoring capabilities. The Emergent site [26] detected unverified claims on the Web, tracking whether they were subsequently verified, and how much they were shared. The approach was based on manual curation, and thus did not scale.

The development of effective countermeasures requires an accurate understanding of the problem, as well as an assessment of its magnitude [27, 28]. To date, the debate on these issues has been informed by limited evidence. Online social network data provides a way to investigate how human behaviors, and in particular patterns of social interaction, are influenced by newsworthy events [29]. Studies of news consumption on Facebook reveal that users tend to confine their attention on a limited set of pages [30, 31]. Starbird demonstrates how alternative news sites propagate and shape narratives around mass-shooting events [32].

Articles in the press have been among the earliest reports to raise the issue of fake news [33]. Many of these analyses, however, are hampered by the quality of available data—subjective, anecdotal, or narrow in scope. In comparison, the internal investigations conducted by the platforms themselves appear to be based on comprehensive disaggregated datasets [34, 35], but lack transparency, owing to the two-fold risk of jeopardizing the privacy of users and of disclosing internal information that could be potentially exploited for malicious purposes [36].

Motivated by these limitations, in previous work we presented a prototype of Hoaxy, an open platform for the study of the diffusion of misinformation and its competition with fact-checking [37]. Here we build upon this prior effort, contributing to the debate on how to combat digital misinformation in two ways:

- We describe the implementation and deployment of the Hoaxy system, which was first introduced in a 2016 demo [37]. The system has been collecting data on the spread of misinformation and fact checking from the public Twitter stream since June of 2016. It is now publicly available (hoaxy.iuni.iu.edu). Users can query the tool to search instances of claims and relative fact checking about any topic and visualize how these two types of content spread on Twitter.
- We leverage the data collected by Hoaxy to analyze the diffusion of articles from low-credibility sources and fact-checks on Twitter in the run up to and wake of the 2016 US Presidential Election. This analysis provides a first characterization of the anatomy of a large-scale online misinformation diffusion network.

When studying misinformation, the first challenge is to assess the truthfulness of a claim. This presents several difficulties. The most important is scalability: it is impossible to manually evaluate a very large number of claims, even for professional fact-checking organizations with dedicated staff. Here we mitigate these issues by relying on a list of low-credibility sources compiled by trusted third-party organizations. In the run-up to and wake of the 2016 US Presidential Elections, several reputable media and fact checking organizations have compiled lists of popular sources that routinely publish unverified content such as hoaxes, conspiracy theories, fabricated news, click bait, and biased, misleading content. We manually assess that the great majority of the articles published by these sources, considered here, contain some form of misinformation or cannot be verified (see [Methods](#)). For brevity, in the remainder of the paper we refer to articles from low-credibility sources simply as “articles.”

Hoaxy retrieves the full and comprehensive set of tweets that share (i.e., include a link to) articles and fact-checks. These tweets are important because, by tracking them, we can observe how a particular piece of content spreads over the social network. It is important to note that Hoaxy collects 100% of these tweets, not a sample. This lets us obtain, for any given piece of misinformation in our corpus, the full picture of how it spreads and competes with subsequent fact-checking, if any.

In this paper we address three research questions:

- **RQ1:** How do the spread of misinformation and fact-checking compete?
- **RQ2:** What are the structural and dynamic characteristics of the core of the misinformation diffusion network, and who are its main purveyors?
- **RQ3:** How to reduce the overall amount of misinformation?

We pose our first question (RQ1) to investigate whether those who are responsible for spreading articles are also exposed to corrections of those articles. Regretfully, only 5.8% of the tweets in our dataset share links to fact-checking content—a 1:17 ratio with misinformation tweets. We analyze the diffusion network in the run up to the election, and find a strong core-periphery structure. Fact-checking almost disappears as we move closer to the inner core of the network, but surprisingly we find that *some* fact-checking content is being shared even inside the main core. Unfortunately, we discover that these instances are not associated with interest in accurate information. Rather, links to Snopes or Politifact are shared either to mock said publications, or to mislead other users (e.g., by falsely claiming that the fact-checkers found a claim to be true). This finding is consistent with surveys on the trust of fact-checking organizations, which find strong polarization of opinions [38].

Our second question (RQ2) is about characterizing the core of the article diffusion network. We find the main core to grow in size initially and then become stable in both size and membership, while its density continues to increase. We analyze the accounts in the core of the network to identify those users who play an important role in the diffusion of misinformation. The use of Botometer, a state-of-the-art social bot detection tool [12], reveals a higher presence of social bots in the main core. We also consider a host of centrality measures (in-strength, out-strength, betweenness, and PageRank) to characterize and rank the accounts that belong in the main core. Each metric emphasizes different subsets of core users, but interestingly the most central nodes according to different metrics are found to be similar in their partisan slant.

Our last question (RQ3) addresses possible countermeasures. Specifically we ask what actions platforms could take to reduce the overall exposure to misinformation. Platforms have already taken some steps this direction, by prioritizing high-quality over low-quality content [35, 39]. Here we consider a further step by investigating whether penalizing the main

purveyors of misinformation, as identified by RQ2, yields an effective mitigation strategy. We find that a simple greedy solution would reduce the overall amount of misinformation significantly.

Methods and data

Network core analysis

The k -core of a graph is formally defined as the maximal subgraph with nodes of at least degree k . In practice, k -core decomposition uses a recursive procedure, given the k -core, for extracting the $(k + 1)$ -core by recursively removing all nodes with degree k . The nodes that have been removed constitute the k -shell. The k -core decomposition is the sequence of k -cores of increasing values of k . Finally, the non-empty graph with maximum value of k is called the *main core*. Prior work has used k -core decomposition to probe the structure of complex networks [40, 41]. In the case of social networks, k -cores can be used to identify influential users [42], and to characterize the efficiency of information spreading [43].

We convert the weighted, directed diffusion networks analyzed in this paper into unweighted, undirected networks when applying k -core decomposition. We also used s -core decomposition, which takes weights into account by using the *strength* s (a.k.a. weighted degree) of a node in place of the degree k [44]. The results are qualitatively similar, so we omit them here and present the analysis based on k -core decomposition.

Bot detection

Social bots play an important role in the spread of misinformation [15]. Researchers have built supervised learning tools to detect such automated accounts with high accuracy. We leverage such a tool, called Botometer [12], to evaluate Twitter accounts.

Botometer performs classification over a large set of features that include temporal, network, language, and sentiment signals. The classifier is trained in supervised fashion from a set of labeled examples. The set includes examples discovered with a honeypot and by human raters. Two classifiers are available, a standard one, which includes English-based language features, and a ‘universal’ one, which does not include language features, and is thus applicable beyond English-speaking contexts. We use the standard classifier through a public API [45, 46].

Article verification

Our analysis considers content published by a set of websites flagged as sources of misinformation by third-party journalistic and fact-checking organizations. We merged several lists of low-credibility sources compiled by such organizations. It should be noted that these lists were compiled independently of each other, and as a result they have uneven coverage. However, there is some overlap between them. The full list is available online [47].

The source-based approach relies on the assumption that most of the articles published by our compilation of low-credibility sources are some type of misinformation, as we cannot fact-check each individual claim. To validate this assumption, we manually verified a sample of 50 articles drawn uniformly at random from our corpus of articles—all those tweeted at least once by any of the sources in our compilation, during the period of interest. Each article was evaluated independently by two reviewers, with ties broken by a third reviewer. We applied a broadly used rubric based on seven types of misinformation: fabricated content, manipulated content, imposter content, false context, misleading content, false connection, and satire [21]. We also added articles that could not be verified (inconclusive). Satire was not excluded



Fig 1. Verification based on a sample of 50 articles. We excluded six articles with no factual claim. Articles that could not be verified are grouped with misinformation.

<https://doi.org/10.1371/journal.pone.0196087.g001>

because fake-news sites often label their content as satirical, and viral satire is often mistaken for real news. Further details about the verification procedure can be found in a technical report [15]. Fig 1 shows that only a minority of articles in the collection ($27 \pm 7\%$) can be verified. The sampling method biases the analysis toward more prolific sources, some of which simply copy and paste large numbers of articles from other sources. The fraction of verifiable articles is cut in half when sampling articles by tweets, thus biasing the sample toward popular rather than prolific sources [15].

We also tracked the websites of several independent fact-checking organizations: politifact.com, snopes.com, factcheck.org, badsatiretoday.com, hoax-slayer.com, opensecrets.org, and truthorfiction.com. In April 2017 we added climatefeedback.org, which does not affect the present analysis.

Hoaxy system

Fig 2 shows the architecture of the Hoaxy system. The system is composed of a back-end and a front-end. Next we describe some of the technical aspects that went into the design and implementation of these components.

Back-end. The back-end provides data collection, processing, storage, and indexing capabilities. We start from the list of sources discussed earlier. Data are collected from two realms:

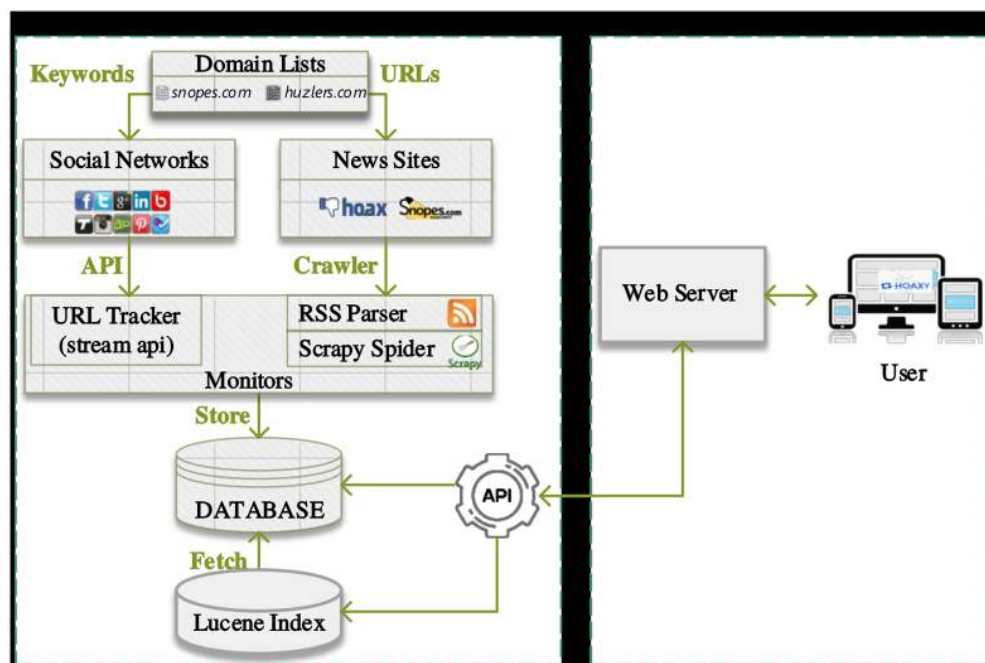


Fig 2. Hoaxy system architecture.

<https://doi.org/10.1371/journal.pone.0196087.g002>

social media (i.e., Twitter) and the news source sites in the list. To collect data from Twitter, Hoaxy filters the real-time stream for tweets matching our list of domain keywords [48]. Matches are performed server-side against the complete text of the tweet. This means that for each delivered tweet we further make sure that the match is actually a hyperlink. Tweets that simply mention our sources but do not link to them (e.g., “I read this on snopes.com!”) are discarded.

All matching link URLs are then extracted from the tweet and fetched directly from the source. To get a complete snapshot of all content produced by the sources, Hoaxy also regularly crawls their websites in a separate process. We use a mix of RSS and direct crawling to do so. Regardless of the way it is collected, from each fetched document Hoaxy extracts title, metadata, and body information.

All collected data (tweets and fetched documents) are saved in a relational database. Documents are further indexed using Lucene [49], to enable full-text search from the front-end.

Content duplication and document text extraction are two critical aspects of this data collection pipeline. Because we are crawling data from the Web, we expect to observe several different variants of the same URLs. This is especially true for the resources obtained from the social media stream, for which duplication may occur due to marketing campaign and other tracking parameters, shortening (e.g., bit.ly) and snapshotting (e.g., archive.is) services, and domain aliasing (e.g., dcmagazine.com and thedcmagazine.com).

While acknowledging that principled solutions to deal with the problem of Web content duplication have been around for decades [50], we found that a set of few, simple heuristics gave satisfactory results. For example, we found that focusing on the most common tracking parameters (i.e. UTM parameters) we can canonicalize about 30% of all URLs. Similarly, by following all types of HTTP redirect responses, we resolve shortened URLs for about 45% of the URLs extracted from tweets. Snapshotting and domain aliases account instead for only five duplicates, and we simply ignore them.

We also had the problem of extracting the actual text of the fetched documents. There is a lot of extraneous content in the body of documents due to the presence of ads, comment threads, and personalization. All this ‘noise’ poses a problem for indexing the corpus efficiently. Algorithms for document text extraction have been around for several years [51]. We tested several implementations and eventually settled for the one offered by a third-party API [52].

Having collected, processed, stored, and indexed all the data, the final component of the back-end is the API, a small piece of middleware that enables programmatic access to both the relational database and the full-text Lucene index for the purposes of search and visualization.

Front-end. Hoaxy provides an intuitive Web interface to search and visualize the spread of articles contained in our database, and the competition with subsequent fact-checking verifications (see Fig 3). The user first specifies a query (Fig 3(a)). Users can choose to retrieve either the most relevant or the most recent results. To do so, we first send the query to Lucene, which returns a list of most relevant/recent articles from low-credibility and fact-checking sources. In practice, because there are many more articles from low-credibility sources than fact-checking sources, we rank these articles separately, and then merge the top results from the two rankings into a single list. Finally we re-rank the results based on the number of tweets in the database.

After selecting the results that match their query (Fig 3(b)), the user can finally visualize the results. Hoaxy provides two types of visualization: a timeline plot (not shown in the figure) that displays the growth in the number of tweets for articles from both low-credibility and fact-checking sources, and an interactive visualization of the diffusion network (Fig 3(c)). In

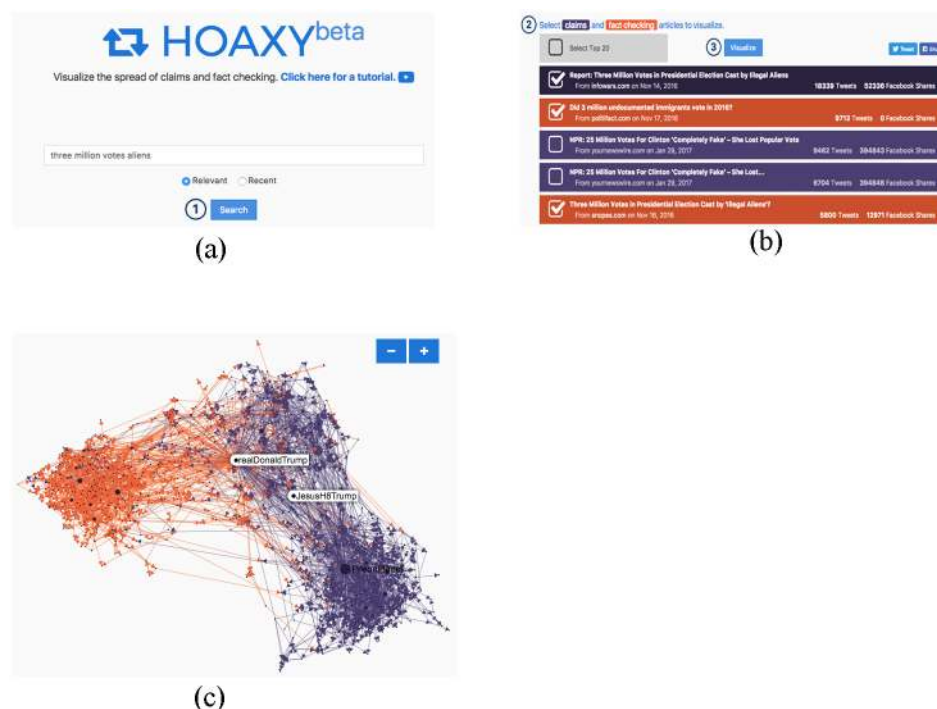


Fig 3. Screen shots from the user interface of Hoaxy: (a) the user enters a query in the search engine interface; (b) from the list of results, the user selects articles to visualize from low-credibility (purple) and/or fact-checking (orange) sources; (c) a detail from the interactive network diffusion visualization for the query “three million votes aliens”. Edge colors represent the type of information exchanged. The network shown here displays strong polarization between articles from low-credibility and fact-checking sources, which is typical.

<https://doi.org/10.1371/journal.pone.0196087.g003>

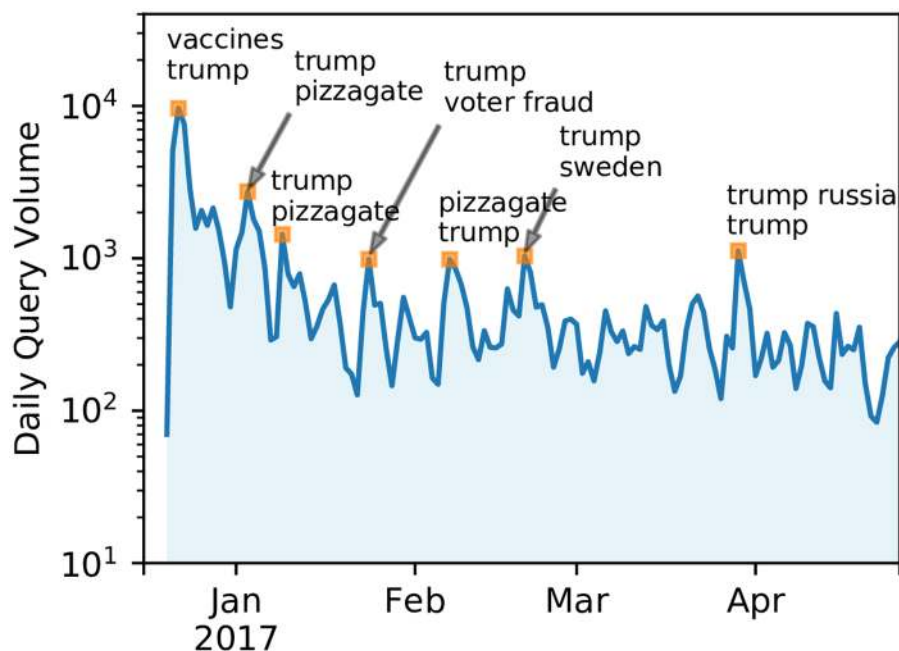


Fig 4. Usage of Hoaxy in terms of daily volume of queries since the launch of the public Web tool in December 2016. The two most frequent search terms are shown in correspondence to some of the main peaks of user activity.

<https://doi.org/10.1371/journal.pone.0196087.g004>

the network, nodes represent Twitter accounts and edges connect any two users that exchanged information by retweet, reply, mention, or quoted retweet. Edge directionality represents the flow of information, e.g., from the retweeted to the retweeter account or from the mentioning to the mentioned account.

Deployment. We started collecting data with Hoaxy from 69 low-credibility and 7 fact-checking sources—76 in total—in June 2016. In December 2016, 50 more low-credibility sources were added. The system has collected data continuously ever since. As of October 2017, Hoaxy has collected a total of 29,351,187 tweets—27,648,423 with links to articles from low-credibility sources and 1,705,576 with links to fact-checking sources. The total number of documents collected so far is 653,911—628,350 by low-credibility sources and 25,561 by fact-checking ones.

The public Web interface of Hoaxy was launched on December 20, 2016. Fig 4 plots the daily query volume and some of the most popular topics queried by users over the course of the first 6 months of operation. Unsurprisingly, the term ‘Trump’ is among the most popular search terms, but we also see substantial churn in user interest, with topics following closely the most popular pieces of controversial information of the moment, e.g., ‘vaccines,’ ‘pizzagate,’ ‘voter fraud’ and ‘Trump Russia.’

Datasets

To explore our research questions, we focus on the retweet network (including quoted retweets) for links to articles from either low-credibility or fact-checking sources. A retweet provides information about the primary spreader (retweeted account) and secondary spreader (retweeting account). To be sure, Hoaxy does collect any kind of tweet, as long as a URL, whose Web domain matches our list of sources, is included in the tweet. To give an idea of the full scope of the Hoaxy dataset, retweets and quoted retweets occur 66.9% of the times;

approximately 1 in 10 retweets is a quoted retweet. Of the remaining types of tweets, replies (i.e. tweets forming a conversation thread and including an @-mention of another users) account for 2.1% of the total. The remaining tweets are neither retweets nor replies; they are *original* tweets.

The network is a graph defined as follows: we include a node for each Twitter user account in the database. Edges are directed (as explained earlier) and weighted. The weight of an edge represents the number of retweets from one account to another. That is, we increase by one the weight on a directed edge from user a to user b every time we observe that b retweets a . Edges are labeled by the type of content being retweeted. To do so, we split the total weight $w(e)$ of edge e in two separate counts, one that keeps track of retweets of articles ($w_c(e)$) and one of fact-checks ($w_f(e)$), respectively. That is, $w(e) = w_c(e) + w_f(e)$ for all $e \in E$. We observe $w_c(e) \cdot w_f(e) > 0$ in only a small minority of edges, meaning that we can easily label each edge as an ‘article’ or ‘fact-check’ edge with a simple majority rule (ties are broken at random).

Because prior work shows that collective attention patterns change dramatically in the wake of highly anticipated events, like elections [53], we split our analysis in two periods, pre- and post-Election Day (Nov. 8, 2016). Table 1 provides a summary of the three networks analyzed in this paper. We explore the overall spread of content on the full network spanning six months before Election Day, including articles from both low-credibility and fact-checking sources. We decompose this network into its k -core shells to uncover the functional roles of the most densely connected sub-graph. Row 1 of Table 1 shows summary statistics for the network used at this stage.

The second dataset is used to study the diffusion of the sole misinformation, therefore we ignore all edges labeled as ‘fact-check.’ To characterize the long-term evolution of the main core, we extend the period of observation to October 9, 2017. Recall that 50 additional low-credibility sources were added to Hoaxy in December 2016. To keep our analysis consistent across the pre- and post-Election Day periods, we do not include data from these sites in the present work. The network in row 2 of Table 1 is considered at this stage.

The third dataset (row 3 of Table 1) includes only articles from low-credibility sources but goes back to the pre-Election-Day period to characterize the most central users in the core and the robustness of the network.

All the analyses presented in this paper can be replicated by collecting data through the Hoaxy API [54] or downloading the network datasets at doi.org/10.5281/zenodo.1154840.

Ethics statement

Hoaxy collects data about public tweets obtained through the Twitter API, in full compliance with Twitter’s Terms of Service. The Indiana University Human Subjects Committee has deemed the protocol exempt from review on the basis that the data is public and poses no risk for the subjects.

Table 1. Summary of the data used in the network analysis. E_f is the set of edges labeled as ‘fact-check’.

	Network	Period	$ V $	$ E $	$ E_f $
1	Full (articles + fact-checks)	pre-Election ^a	346, 573	1, 091, 552	279, 283
2	Misinformation (articles only)	pre- ^a & post-Election ^b	630, 368	2, 236, 041	0
3	Misinformation (articles only)	pre-Election ^a	227, 363	816, 453	0

^aMay 16, 2016—Nov. 07, 2016

^bNov. 08, 2016 (Election Day)—Oct. 09, 2017

<https://doi.org/10.1371/journal.pone.0196087.t001>

The reviewers who performed article verification and annotated Twitter profiles were researchers in our lab who were recruited and provided informed consent via email.

Results

Having described in the prior section how Hoaxy collects data, let us now analyze the misinformation diffusion networks. To the best of our knowledge, the following is the first in-depth analysis of the diffusion network of online misinformation and fact-checking in the period of the 2016 US Presidential Election.

Low-credibility articles vs. fact-checking

We performed k -core decomposition of the entire network (row 1 of Table 1). The main core is obtained at $\max\{k\} = 50$. We start with a simple exercise of visualizing the fraction of fact-checking retweets among all retweets that forms each core in Fig 5. If nodes contributed to articles and fact-checking in the same proportions irrespective of core number, one would expect this fraction to remain constant as peripheral nodes are removed. On the contrary, we see a drastic decrease as k increases. Fact-checking activity almost disappears as we move toward the innermost, densest portion of the network.

To understand the details, Fig 6 visualizes different k -core networks for increasing values of k . We can draw several insights from these visualizations. The force-directed layout algorithm splits the network in two communities. There is substantial content segregation across these two communities, which we denote as the ‘fact-checkers’ and (misinformation) ‘spreaders,’ respectively. The edges across the two groups appear to be mostly colored in orange, suggesting some exposure of misinformation spreaders to fact-checks.

However, it is still possible to appreciate *some* retweeting of fact-checking content involving spreaders even in the main core (Fig 6(d)). To understand in more quantitative terms the role of fact-checking in the spread of information in the core, we characterize users according to

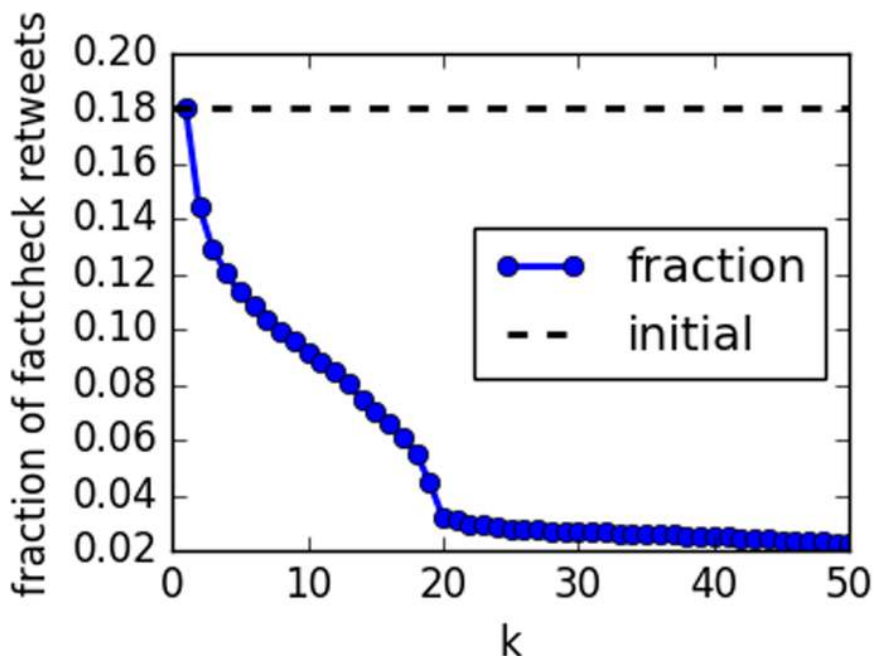


Fig 5. Fraction of retweets in k -core graph that link to fact-checking vs. core number k .

<https://doi.org/10.1371/journal.pone.0196087.g005>

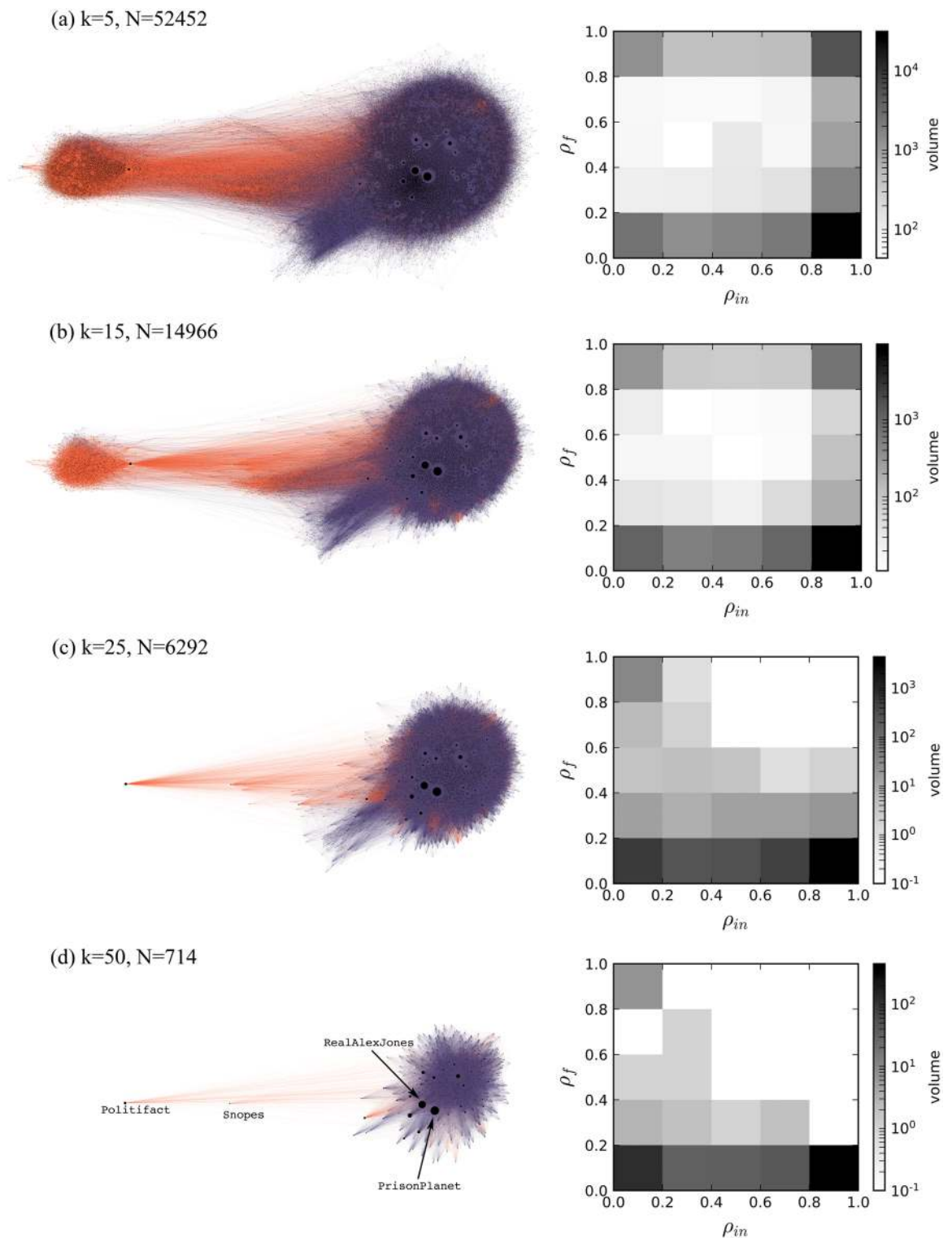


Fig 6. k -core decomposition of the pre-Election retweet network collected by Hoaxy. Panels (a)–(d) show four different cores for values of $k = 5, 15, 25, 50$ respectively. Networks are visualized using a force-directed layout. Edge colors represent the type of article source: orange for fact-checking and purple for low-credibility. The innermost sub-graph (d), where each node has degree $k \geq 50$, corresponds to the main core. The heat maps show, for each core, the distribution of accounts in the space represented by two coordinates: the retweet ratio ρ_{in} and the fact-checking ratio ρ_f (see text).

<https://doi.org/10.1371/journal.pone.0196087.g006>

two simple metrics. Recall that in a weighted, undirected network the *strength* of a node is the sum of all the weights of all its incident edges, $s(v) = \sum_{e \in v} w(e)$. In a directed network one can likewise define the *in-strength* s_{in} and the *out-strength* s_{out} , by taking the sum only on the incoming and outgoing edges, respectively. We further consider edge labels and distinguish between article (s_c) and fact-check (s_f) strength. For each node $v \in V$ let us define two ratios, the *fact-checking ratio* ρ_f and the *retweet ratio* ρ_{in} :

$$\rho_f(v) = \frac{s_f(v)}{s_f(v) + s_c(v)} = \frac{s_f(v)}{s(v)} \quad (1)$$

$$\rho_{in}(v) = \frac{s_{in}(v)}{s_{in}(v) + s_{out}(v)} = \frac{s_{in}(v)}{s(v)}. \quad (2)$$

Intuitively, a user with a value of ρ_f close to unity is going to be a fact-checker (as opposed to misinformation spreader), whereas an account with a value of ρ_{in} close to unity is going to be a secondary spreader of information, i.e., to amplify messages through retweets rather than post original messages. The right-hand side of Fig 6 shows the joint distributions of (ρ_f, ρ_{in}) for different values of k . We observe that for small values of k , most users fall close to the four corners of the space, meaning that they take on exactly one of the four possible combinations of roles ('primary spreader of articles from low-credibility sources', 'secondary spreader of articles from fact-checking sources', etc.).

For larger values of k (Fig 6(c) and 6(d)), we observe a shift away from secondary spreaders of fact-checking. In other words, the fact-checking links in the network core are retweeted by accounts who mainly spread misinformation. Fig 7 confirms that the drop in the spread of fact-checking is precipitous. For $k > 20$ there is only a small, stable residual activity.

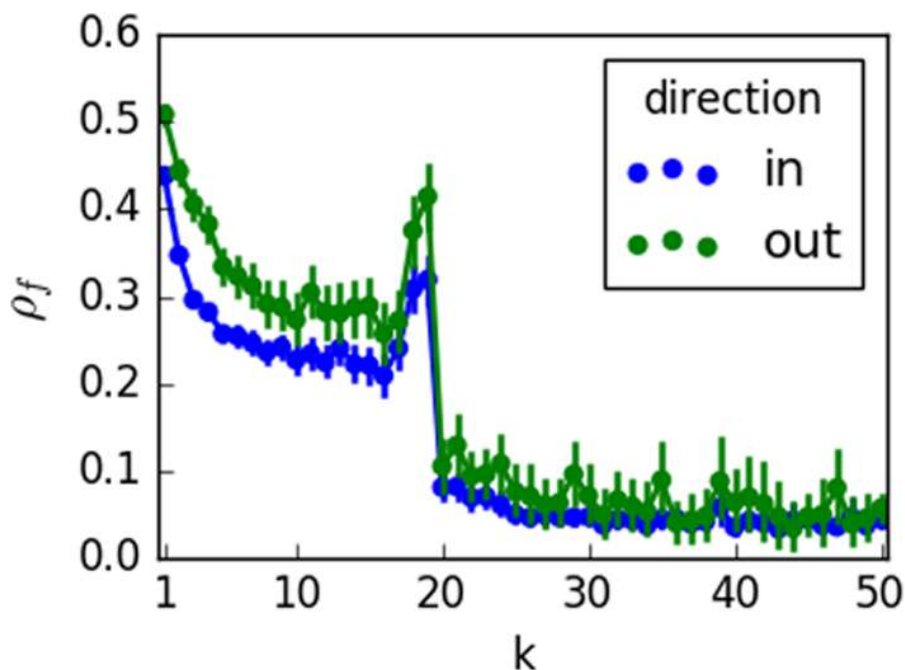


Fig 7. Average fact-checking ratio as a function of the shell number k for activities of both primary spreading ('out') and secondary spreading ('in'). Error bars represent standard error.

<https://doi.org/10.1371/journal.pone.0196087.g007>

Table 2. Sample of tweets with fact-checking content published by accounts in the main core of the misinformation network.

BIASED REPETITION
BREAKING NEWS! RINO GOP #NeverTrump Leader PAID \$294K RINO TRAITORS prefer #Hillary; Marxist SCOTUS Click https://www.opensecrets.org/politicians/contrib.php?cid=N00035544&cycle=2016&type=I
HRC PRAISED HER KKK FRIEND MENTOR, BYRD! HRC IS RACIST! Hillary Kissed by Former Klan Member http://www.snopes.com/clinton-byrd-photo-klan/
ATTACKS ON FACT-CHECKING
Newsflash: Snopes itself is a biased left-wing Clinton mouthpiece She knew he fooled the polygraph, was guilty http://www.snopes.com/hillary-clinton-freed-child-rapist-laughed-about-it/
Lying Politifact caught telling another objective lie. CNN Is Hitler. http://www.politifact.com/truth-o-meter/statements/2016/aug/23/donald-trump/donald-trump-fundraising-email-takes-cnn-anchors-c/
INCONSISTENCY WITH FACT-CHECKING STANCE
13 Hours of HELL in Benghazi! No HELP was Sent?? Her E-Mails Show SHE KNEW THE TRUTH! #LIAR http://www.politifact.com/truth-o-meter/article/2016/feb/09/what-did-hillary-clinton-tell-families-people-who-/
Machado had sex on camera while filing a reality show. The media is lying about there being no sex tape. http://www.snopes.com/alicia-machado-adult-star/
https://doi.org/10.1371/journal.pone.0196087.t002

The fact that fact-checking still spreads in the main core is a somewhat surprising observation. Therefore we search for patterns that explain how misinformation spreaders interact with fact-checking. Manual inspection of the data let us identify three key characteristics of these retweets of fact-checking content made by spreaders in the main core: (1) they link to fact-checking articles with biased, misleading wording; (2) they attack fact-checking sites; or (3) they use language that is inconsistent with the stance of a fact-checking article, for example implying that a claim is true even though the linked fact-checking article states that it is false. A sample of tweets with each of the aforementioned characteristics is shown in Table 2. Similar patterns of citing mainstream media to challenge them have been observed by Starbird [32].

Anatomy of the main core

Core dynamics. Although we observe that the main core is dominated by misinformation spreaders, it is unclear if this has always been the case. From this point and in the subsequent analysis we discard all edges labeled as ‘fact-check’ and focus only on the spread of misinformation. We start by investigating the long-term growth of the network. To do so we consider a network based on the Hoaxy dataset that extends post-Election Day; see row 2 of Table 1.

In particular we consider all retweets in our dataset in chronological order. At any given point in time, we consider a snapshot of the network formed by all retweets up to that point. We perform k -core decomposition on this cumulative network. We extract two pieces of information: the maximum value of k and the size of the main core. The left panel of Fig 8 shows how these two quantities change over time. We observe that the core gets both larger and denser. To characterize the extent to which the increasing density is just a byproduct of considering snapshots of growing size, we also plot the k of the main core for a shuffled version of each snapshot (i.e., the configuration model of each snapshot). In rewiring edges, we preserve the degree sequence of the network. While both the actual and the rewired network grow denser, the actual network does so at a higher rate, and the difference is statistically significant.

The main core reaches an equilibrium size of approximately 800 accounts around Election Day (see inset of left panel of Fig 8). This observations prompts the question of who are the users in the main core, and whether there is substantial churn in this group over time. By considering the intersection between two consecutive main cores in the sequence, we find a low churn rate after a peak in August (right panel of Fig 8), implying a stable set of users who

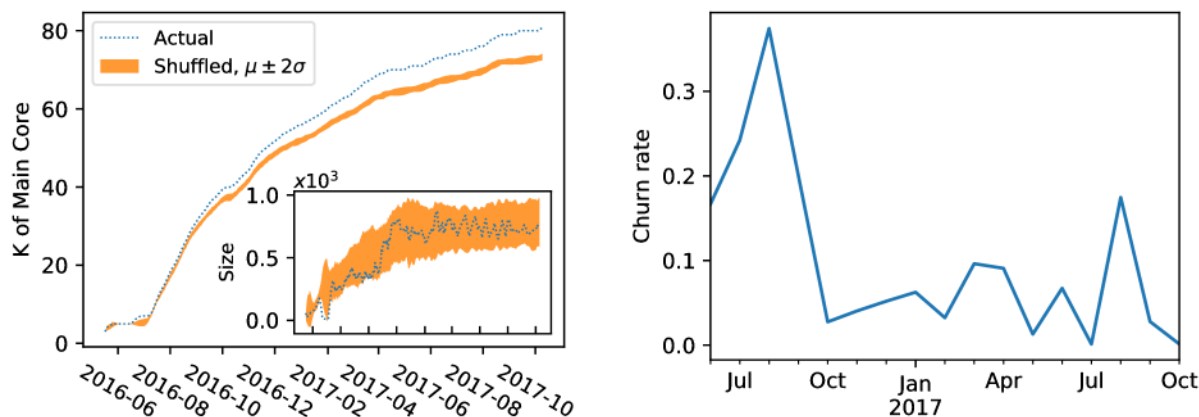


Fig 8. *Left:* Change of main core number k with the evolution of the network. A rolling window of one week is applied to filter fluctuations. The shuffled version is obtained by sampling from the configuration model. This is repeated many times to obtain the 95% confidence interval shown in orange. The inset shows the size of the main core over time. *Right:* Churn rate (relative monthly change) of accounts in the main core.

<https://doi.org/10.1371/journal.pone.0196087.g008>

consistently drive network activities. In fact, there are 321 users who remain in the main core for the whole duration of the observation period. The retweet network of a subset of them is shown in Fig 9.

Core membership. We now return to the period before the election. For this analysis we consider the pre-Election network, whose statistics are described in row 3 of Table 1. We use this network to characterize several aspects of how misinformation spreads before the election, starting with the presence of automated accounts. We do expect to find evidence of automation, given recent findings that show that a sizable fraction of the broader Twitter conversation about the elections was due to bots [55]. Open questions are whether bots were successful at spreading misinformation, which would be reflected by their tendency to locate in the core of the network, as opposed to the periphery; and whether bots interact with humans.

To determine the likelihood that the observed patterns of activity in the core of the network are the result of the deployment of social bots, we perform bot detection on a sample of accounts. After k -core decomposition, we sample 2000 accounts at random from each k -shell. If the size of a shell is smaller, we include the whole shell. To estimate the likelihood that each of these account is automated we compute a bot score by querying Botometer, a state-of-the-art bot detection tool (see Methods). Fig 10 shows a sharp increase in bot scores as we move toward the core of the network. We observe many human-like accounts in the main core as well, indicated by the bot scores of accounts in the main core. If we define bot accounts as those having bot scores above 0.5, about 75% of the accounts in the main core are classified as human. In addition, about 25% of the retweets in the main core are classified as humans retweeting bots. This suggests a significant impact of bots on human perception in information diffusion.

For at least some of the most important individuals in the network core, it would be useful to have a behavioral description at a more granular level than just group averages. To this end, we need first to define a subset of important accounts. The network science toolbox provides us with several methods to identify salient nodes. We consider four centrality metrics: in-strength s_{in} , out-strength s_{out} , PageRank [56], and betweenness [57]. The first two are local measures of primary (s_{out}) and secondary (s_{in}) spreading activity. Intuitively, s_{out} captures influence, as measured by the number of times that an account is retweeted. On the other hand, s_{in} captures prolific accounts that retweet a lot. The distribution of s_{in} and s_{out} is shown

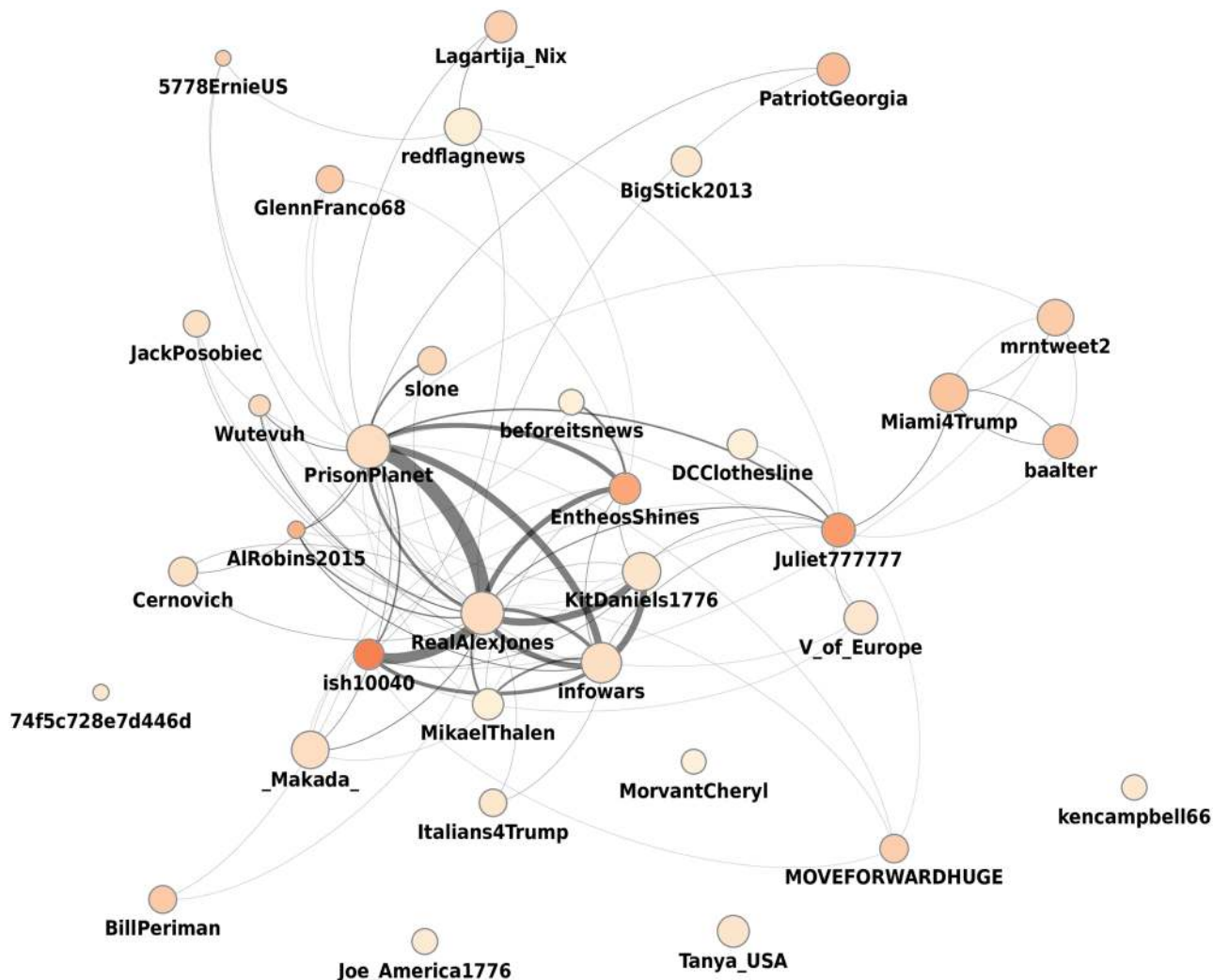


Fig 9. Retweet network of the stable main core of spreaders of article from low-credibility sources. Filtering by in-degree was applied to focus on the 34 accounts that retweet the most other accounts in the core. Node size represents out-degree (number of retweeters) and node color represents in-degree.

<https://doi.org/10.1371/journal.pone.0196087.g009>

in Fig 11 (left panel). Both distributions are broad, and the range of out-strength is broader than that of in-strength, due to the simple fact that, in networks of this size, the rate at which one can be retweeted is generally larger than that at which one can retweet others. The third and fourth measures are instead global notions of centrality, based on random walks and shortest paths, respectively. Given that these metrics capture fundamentally different notions of centrality, we expect them to produce different rankings of the nodes in the network. The right panel of Fig 11 shows strong variation in the average rank of accounts in the main core, confirming this intuition. The metric that seems to best capture the main core is the in-strength, indicating that a majority of core accounts are secondary spreaders (prolific accounts).

For each measure, we rank the accounts in the main core and consider the top ten users. This exercise yields four lists of accounts, whose screen names are shown in Table 3. There is little or no overlap between these lists, and their union yields 34 unique accounts. Having identified a subset of moderate size that includes main core members of potential interest, we

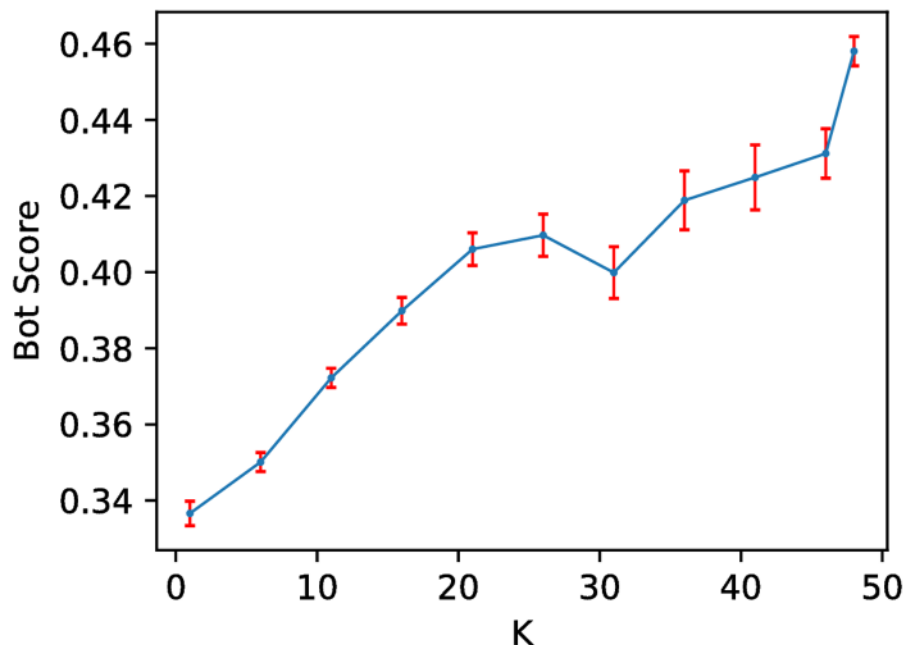


Fig 10. Average bot score for a random sample of accounts drawn from different k -shells of the pre-Election Day retweet network, as a function of k . Only retweets including links to sources of misinformation are considered. Error bars represent standard errors.

<https://doi.org/10.1371/journal.pone.0196087.g010>

performed a qualitative analysis of these accounts. Three human raters were asked to inspect the Twitter profile of each user independently, and to provide answers to the following questions:

1. Bot or Human?
2. Partisanship?
3. Personal or organizational account?
4. How often does it share articles from low-credibility sources?

For questions 1–3, whose answer is a categorical variable, the raters could also choose ‘neither’ or ‘not sure’. After the annotators coded each account, for each question we applied a

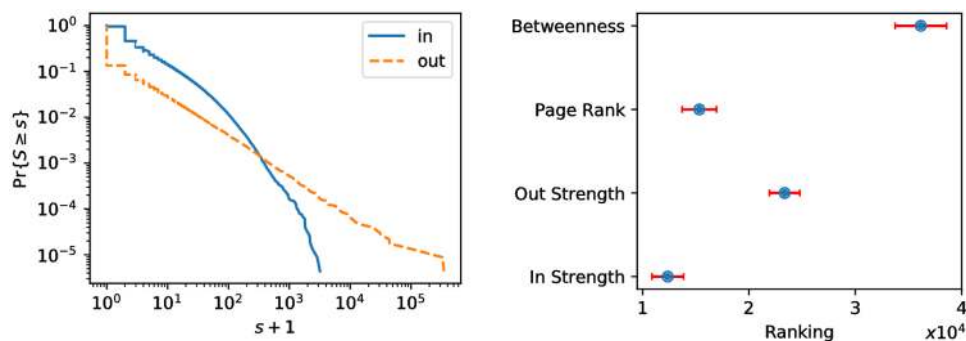


Fig 11. Left: Distribution of s_{in} and s_{out} . Right: The average rank of users in the main core according to each centrality metric. Error bars represent standard errors.

<https://doi.org/10.1371/journal.pone.0196087.g011>

Table 3. The top ten central accounts, ranked in descending order of centrality, in the article network before the 2016 Election. Rankings are based on four different centrality metrics.

Betweenness	PageRank	s_{in}	s_{out}
PrisonPlanet	ImmoralReport	PATROIT73	RealAlexJones
AlternativViewz	BillPeriman	BadCompany709	PrisonPlanet
RealAlexJones	alllibertynews	LovToRideMyTrek	infowars
libertytarian	eavesdropann	PhilDeCarolus	redflagnews
eavesdropann	Lagartija_Nix	Roostwoodstock	Miami4Trump
BillPeriman	MMUSA_PPN	Skinner147	beforeitsnews
wvufanagent99	retireleo	MrNoahItALL	KitDaniels1776
Miami4Trump	Nuevomedio	RESPECTPUNK434	V_of_Europe
Juliet777777	ish10040	Rubenz1133	_Makada_
ish10040	EntheosShines	Cecil3695Cecil	Tanya_USA

<https://doi.org/10.1371/journal.pone.0196087.t003>

majority rule to identify the consensus label. The few cases in which a consensus could not be reached were broken by a fourth rater (one of the authors). The results are shown in Table 4. We report results for 32 of the 34 original accounts, since two accounts had been suspended by Twitter, and thus could not be annotated. Many of the central accounts appear to be automated and display a high degree of partisanship, all in support of the same candidate.

Network robustness

Our last question is about the overall robustness of the network (row 3 of Table 1). We ask: *How much does the efficient spread of articles from low-credibility sources rely on the activity of the most central nodes?* To explore this question we apply node disconnection, a standard procedure for estimating robustness in network science [58]. The idea is to remove one node at a time, and analyze how two simple metrics are curtailed as a result: total volume of article retweets, and total number of unique article links. The more these quantities can be reduced by removing a small number of nodes, the more the efficiency of the misinformation network is disrupted.

We wish to measure the fraction of retweets remaining after simulating the scenario in which a certain number of accounts are disconnected, by removing all edges to and from those accounts. There is one caveat. The retweet metadata from Twitter does not allow to reconstruct

Table 4. Annotation of central users. For categorical questions (1–3), the top most frequent label, and its frequency, are reported. The question about article sharing frequency (4) was on a 5-point Likert scale; we report the mean and standard deviation of the answers.

	Betweenness	PageRank	s_{in}	s_{out}
	BOT/HUMAN			
Top Freq.	Bot	Bot	Bot	Bot
	6	6	4	4
	PARTISANSHIP			
Top Freq.	Partisan	Partisan	Partisan	Partisan
	10	7	7	8
	PERSONAL/ORGANIZATIONAL			
Top Freq.	Personal	Personal	Personal	Personal
	9	5	8	6
	HOW OFTEN DOES IT SHARE ARTICLES FROM LOW-CREDIBILITY SOURCES?			
Mean	2.9 ± 0.8	2.4 ± 0.9	2.6 ± 0.7	3 ± 1

<https://doi.org/10.1371/journal.pone.0196087.t004>

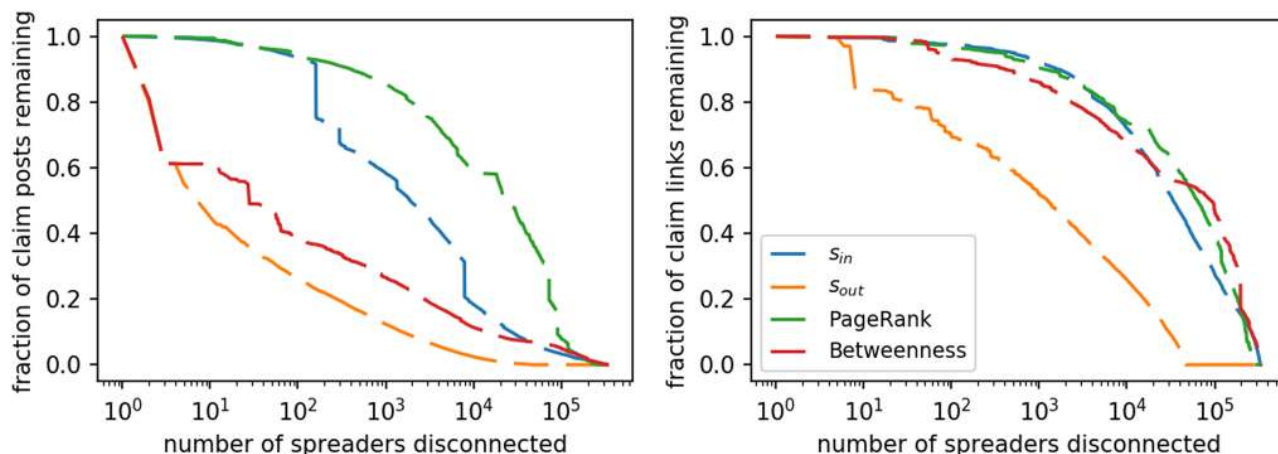


Fig 12. Left: Fraction of the retweets remaining vs. number of spreaders disconnected in the network. Right: Fraction of unique article links remaining vs. number of spreaders disconnected in the network. The priority of disconnected users is determined by ranking on the basis of different centrality metrics.

<https://doi.org/10.1371/journal.pone.0196087.g012>

the full retweet cascades. Instead, a retweet connects directly to the root of the cascade tree, so that the cascade tree is projected onto a star network. This means that when we disconnect a retweeting node, only the single leaf node is removed from the star, rather than the subtree rooted at that node in the actual cascade tree.

We prioritize accounts to disconnect based on the four centrality metrics discussed before (s_{in} , s_{out} , betweenness, and PageRank). Fig 12 shows the result of the simulation. The greedy strategy that ranks users by decreasing out-strength achieves the best reduction of both metrics. The efficiency of the network is greatly impaired even after disconnecting as few as 10 most influential accounts (i.e., with greatest s_{out}). Surprisingly, disconnecting nodes with the highest s_{in} is not as efficient a strategy for reducing misinformation; the network is robust with respect to the removal of prolific accounts in the core. Betweenness, in comparison, seems to give good results on the total number of retweets (left panel of Fig 12), but does not produce better results than PageRank and in-strength when considering unique links (right panel).

From a policy perspective, we are not proposing that a social media platform should suspend accounts whose posts are highly retweeted. Of course, platforms must take great care in minimizing the chances that a legitimate account is suspended. However, platforms do use various signals to identify and penalize low-quality information [35, 39]. The present analysis suggests that the use of s_{out} in the article spreading network might provide a useful signal to prioritize further review, with the goal of mitigating the spread of misinformation. Such an approach assumes the availability of a list of low-quality sources, which can be readily compiled.

Discussion

The rise of digital misinformation is calling into question the integrity of our information ecosystem. Here we made two contributions to the ongoing debate on how to best combat this threat. First, we presented Hoaxy, an open platform that enables large-scale, systematic studies of how misinformation and fact-checking spread and compete on Twitter. We described key aspects of its design and implementation. All Hoaxy data is available through an open API. Second, using data from Hoaxy, we presented an in-depth analysis of the misinformation diffusion network in the run up to and wake of the 2016 US Presidential Election. We found that

the network is strongly segregated along the two types of information circulating in it, and that a dense, stable core emerged after the election. We characterized the main core in terms of multiple centrality measures and proposed an efficient strategies to reduce the circulation of information by penalizing key nodes in this network. The networks used in the present analysis are available on an institutional repository (see [Methods](#)).

Recall that Hoaxy collects 100% of the tweets carrying each piece of misinformation in our collection, not a sample. As a result, our analysis provides a complete picture of the anatomy of the misinformation network. Of course, our methodology has some unavoidable limitations. First of all, Hoaxy only tracks a fixed, limited set of sources, due to data volume restrictions in the public Twitter API. Of these sources, it only tracks how their content spreads on Twitter, ignoring other social media platforms. Facebook, by far the largest social media platform, does not provide access to data on shares, ostensibly for privacy reasons, even though a significant fraction of misinformation spreads via its pages [30], which are understood to be public. Thus we acknowledge that coverage of our corpus of misinformation is incomplete. Nonetheless, by focusing on low-credibility sources that have come to the attention of large media and fact-checking organizations, and that have been flagged as the most popular purveyors of unverified claims, Hoaxy captures a broad snapshot of misinformation circulating online.

Second, Hoaxy does not track the spread of unsubstantiated claims in the professional mainstream press. News websites do report unverified claims, in a manner and with a frequency dictated by their own editorial standards. For example, hedging language is often used to express degrees of uncertainty [59]. While most claims reported in the mainstream media are eventually verified, many remain unverified, and some even turn out to be false. Some instances of misinformation may see their spread boosted as a result of additional exposure on mainstream news outlets. Understanding the dynamics of the broader media and information ecosystem is therefore needed to fully comprehend the phenomenon of digital misinformation, but it is outside the scope of the present work.

Third, we consider only US-based sources publishing English content. This is an unavoidable consequence of our reliance on lists produced by US-based media organizations. Different sources will be of course active in different countries. Worrisome amounts of misinformation, for example, have been observed in the run-up to the general elections in France [14]. To foster the study of misinformation in non-US contexts, we have released the code of Hoaxy under an open-source license, so that other groups can build upon our work [60, 61].

Last but not least, it is important to reiterate that the articles collected by Hoaxy are in general not verified. Inspection of our corpus confirms that not all articles collected by Hoaxy are completely inaccurate. As far as the present analysis is concerned, we provide an assessment of the rate of confirmed articles in our dataset (see [Methods](#)). When used as a search engine for misinformation, Hoaxy addresses this limitation by showing the most relevant fact-checking articles matching the input query, thereby facilitating claim verification. We hope that the data, software, and visualizations offered by the Hoaxy platform will be useful to researchers, reporters, policymakers, and, last but not least, ordinary Internet users as they learn to cope with online misinformation.

Acknowledgments

Thanks to Ben Serrette and Valentin Pentchev of the Indiana University Network Science Institute (iuni.iu.edu) for supporting the development of the Hoaxy platform. Clayton A. Davis developed the Botometer API. Kaicheng Yang provided bot scores for the robustness analysis. Nic Dias was instrumental in developing the rubric for article verification. We are grateful to Onur Varol, Mihai Avram, Zackary Dunivin, Gregory Maus, and Vincent Wong

who assisted in verification of articles and annotation of central accounts. 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, Computing, and Engineering.

Author Contributions

Conceptualization: Chengcheng Shao, Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia.

Data curation: Chengcheng Shao, Pik-Mai Hui, Lei Wang.

Formal analysis: Chengcheng Shao, Pik-Mai Hui, Filippo Menczer, Giovanni Luca Ciampaglia.

Funding acquisition: Xinwen Jiang, Filippo Menczer.

Methodology: Chengcheng Shao, Filippo Menczer, Giovanni Luca Ciampaglia.

Project administration: Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia.

Resources: Lei Wang, Filippo Menczer.

Supervision: Alessandro Flammini, Filippo Menczer.

Visualization: Chengcheng Shao, Pik-Mai Hui.

Writing – original draft: Chengcheng Shao, Pik-Mai Hui.

Writing – review & editing: Xinwen Jiang, Alessandro Flammini, Filippo Menczer, Giovanni Luca Ciampaglia.

References

1. Barthel M, Mitchell A, Holcomb J. Many Americans Believe Fake News Is Sowing Confusion; 2016. Available from: <http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>.
2. Gottfried J, Shearer E. News Use Across Social Media Platforms 2017; 2017. Available from: <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>.
3. Barthel M, Mitchell A. Americans' Attitudes About the News Media Deeply Divided Along Partisan Lines; 2017. Available from: <http://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/>.
4. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, et al. Truthy: Mapping the Spread of Astroturf in Microblog Streams. In: Proceedings of the 20th International Conference Companion on World Wide Web. WWW'11. New York, NY, USA: ACM; 2011. p. 249–252. Available from: <http://doi.acm.org/10.1145/1963192.1963301>.
5. Xiang W, Zhilin Z, Xiang Y, Yan J, Bin Z, Shasha L. Finding the hidden hands: a case study of detecting organized posters and promoters in SINA weibo. China Communications. 2015; 12(11):1–13. <https://doi.org/10.1109/CC.2015.7366237>
6. Ratkiewicz J, Conover M, Meiss M, Goncalves B, Flammini A, Menczer F. Detecting and Tracking Political Abuse in Social Media. In: Proc. International AAAI Conference on Web and Social Media. Palo Alto, CA: AAAI; 2011. p. 297–304. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850>.
7. Sampson J, Morstatter F, Wu L, Liu H. Leveraging the Implicit Structure Within Social Media for Emergent Rumor Detection. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. CIKM'16. New York, NY, USA: ACM; 2016. p. 2377–2382. Available from: <http://doi.acm.org/10.1145/2983323.2983697>.
8. Wu L, Morstatter F, Hu X, Liu H. Mining Misinformation in Social Media. In: Thai MT, Wu W, Xiong H, editors. Big Data in Complex and Social Networks. Business & Economics. Boca Raton, FL: CRC Press; 2016. p. 125–152.

9. Declerck T, Osenova P, Georgiev G, Lendvai P. Ontological Modelling of Rumors. In: TrandabăT D, Gifu D, editors. *Linguistic Linked Open Data: 12th EUROLAN 2015 Summer School and RUMOUR 2015 Workshop*, Sibiu, Romania, July 13–25, 2015, Revised Selected Papers. Berlin/Heidelberg, Germany: Springer International Publishing; 2016. p. 3–17. Available from: https://doi.org/10.1007/978-3-319-32942-0_1.
10. Kumar S, West R, Leskovec J. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. In: *Proceedings of the 25th International Conference on World Wide Web. WWW'16*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2016. p. 591–602. Available from: <https://doi.org/10.1145/2872427.2883085>.
11. Varol O, Ferrara E, Menczer F, Flammini A. Early detection of promoted campaigns on social media. *EPJ Data Science*. 2017; 6(1):13. <https://doi.org/10.1140/epjds/s13688-017-0111-y>
12. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In: *Proc. International AAAI Conference on Web and Social Media*. Palo Alto, CA: AAAI; 2017. p. 280–289. Available from: <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587>.
13. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The Rise of Social Bots. *Commun ACM*. 2016; 59(7):96–104. <https://doi.org/10.1145/2818717>
14. Ferrara E. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*. 2017; 22(8). <https://doi.org/10.5210/fm.v22i8.8005>
15. Shao C, Ciampaglia GL, Varol O, Flammini A, Menczer F. The spread of misinformation by social bots. *CoRR*; 2017. arXiv:1707.07592.
16. Ecker UKH, Hogan JL, Lewandowsky S. Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction? *Journal of Applied Research in Memory and Cognition*. 2017; 6(2):185–192. <https://doi.org/10.1016/j.jarmac.2017.01.014>
17. Nyhan B, Reifler J. Estimating Fact-checking's Effects; 2016. Available from: <https://www.americanpressinstitute.org/wp-content/uploads/2016/09/Estimating-Fact-Checkings-Effect.pdf>.
18. Jun Y, Meng R, Johar GV. Perceived social presence reduces fact-checking. *Proceedings of the National Academy of Sciences*. 2017; 114(23):5976–5981. <https://doi.org/10.1073/pnas.1700175114>
19. Nematzadeh A, Ciampaglia GL, Menczer F, Flammini A. How algorithmic popularity bias hinders or promotes quality. *CoRR*; 2017. arXiv:1707.00574.
20. Qiu X, F M Oliveira D, Sahami Shirazi A, Flammini A, Menczer F. Limited individual attention and online virality of low-quality information. *Nature Human Behavior*. 2017; 1:0132–. <https://doi.org/10.1038/s41562-017-0132>
21. Wardle C. Fake news. It's complicated. *First Draft News*; 2017. Available from: <https://firstdraftnews.com/fake-news-complicated/>.
22. Mitra T, Gilbert E. CRED BANK: A Large-Scale Social Media Corpus With Associated Credibility Annotations. In: *Proc. International AAAI Conference on Web and Social Media*. Palo Alto, CA: AAAI; 2015. p. 258–267. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10582>.
23. Hassan N, Li C, Tremayne M. Detecting Check-worthy Factual Claims in Presidential Debates. In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management. CIKM'15*. New York, NY, USA: ACM; 2015. p. 1835–1838. Available from: <http://doi.acm.org/10.1145/2806416.2806652>.
24. Metaxas PT, Finn S, Mustafaraj E. Using TwitterTrails.com to Investigate Rumor Propagation. In: *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. CSCW'15 Companion*. New York, NY, USA: ACM; 2015. p. 69–72. Available from: <http://doi.acm.org/10.1145/2685553.2702691>.
25. Carton S, Park S, Zeffer N, Adar E, Mei Q, Resnick P. Audience Analysis for Competing Memes in Social Media. In: *Proc. International AAAI Conference on Web and Social Media*. Palo Alto, CA: AAAI; 2015. p. 41–50. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10592>.
26. Silverman C. Emergent; 2015. Available from: <http://www.emergent.info/>.
27. Ciampaglia GL, Mantzarlis A, Maus G, Menczer F. Research Challenges of Digital Misinformation: Toward a Trustworthy Web. *AI Magazine*. 2018; in press.
28. Lazer D, Baum M, Grinberg N, Friedland L, Joseph K, Hobbs W, et al. Combating Fake News: An Agenda for Research and Action; 2017. Available from: <https://shorensteincenter.org/combating-fake-news-agenda-for-research/>.
29. Lu X, Brelsford C. Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. *Scientific Reports*. 2014; 4:6773. <https://doi.org/10.1038/srep06773> PMID: 25346468

30. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. *Proc National Academy of Sciences*. 2016; 113(3):554–559. <https://doi.org/10.1073/pnas.1517441113>
31. Schmidt AL, Zollo F, Del Vicario M, Bessi A, Scala A, Caldarelli G, et al. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*. 2017; 114(12):3035–3039. <https://doi.org/10.1073/pnas.1617052114>
32. Starbird K. Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In: *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*; 2017. p. 230–239. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15603>.
33. Silverman C. Viral Fake Election News Outperformed Real News On Facebook In Final Months Of The US Election; 2016. Available from: <https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>.
34. Weedon J, Nuland W, Stamos A. Information Operations and Facebook; 2017. Available from: <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>.
35. Mosseri A. News Feed FYI: Showing More Informative Links in News Feed; 2017. Available from: <https://newsroom.fb.com/news/2017/06/news-feed-fyi-showing-more-informative-links-in-news-feed/>.
36. Crowell C. Our Approach to Bots & Misinformation; 2017. Available from: https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html.
37. Shao C, Ciampaglia GL, Flammini A, Menczer F. Hoaxy: A Platform for Tracking Online Misinformation. In: *Proceedings of the 25th International Conference Companion on World Wide Web. WWW'16 Companion*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2016. p. 745–750. Available from: <https://doi.org/10.1145/2872518.2890098>.
38. Brandtzaeg PB, Følstad A. Trust and Distrust in Online Fact-checking Services. *Commun ACM*. 2017; 60(9):65–71. <https://doi.org/10.1145/3122803>
39. Gomes B. Our latest quality improvements for Search; 2017. Available from: <https://blog.google/products/search/our-latest-quality-improvements-search/>.
40. Dorogovtsev SN, Goltsev AV, Mendes JFF. *k*-Core Organization of Complex Networks. *Phys Rev Lett*. 2006; 96:040601 <https://doi.org/10.1103/PhysRevLett.96.040601> PMID: 16486798
41. Alvarez-Hamelin JL, Dall'Asta L, Barrat A, Vespignani A. K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases. *Networks and Heterogeneous Media*. 2008; 3(2):371–393. <https://doi.org/10.3934/nhm.2008.3.371>
42. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. *Nature Physics*. 2010; 6:888–. <https://doi.org/10.1038/nphys1746>
43. Conover MD, Gonçalves B, Flammini A, Menczer F. Partisan asymmetries in online political activity. *EPJ Data Science*. 2012; 1(1):6 <https://doi.org/10.1140/epjds6>
44. Eidsaa M, Almaas E. *s*-core network decomposition: A generalization of *k*-core analysis to weighted networks. *Phys Rev E*. 2013; 88:062819. <https://doi.org/10.1103/PhysRevE.88.062819>
45. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A System to Evaluate Social Bots. In: *Proceedings of the 25th International Conference Companion on World Wide Web. WWW'16 Companion*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2016. p. 273–274. Available from: <https://doi.org/10.1145/2872518.2889302>.
46. Davis CA. Botometer API; 2017. Available from: <https://botometer.iuni.iu.edu/#/api>.
47. Shao C, Menczer F, Ciampaglia GL. Hoaxy FAQ; 2017. Available from: <http://hoaxy.iuni.iu.edu/faq.html>.
48. Twitter. Filter realtime Tweets; 2017. Available from: <https://developer.twitter.com/en/docs/tweets/filter-realtime/overview/statuses-filter>.
49. Apache Software Foundation. Apache Lucene; 2005. Available from: <http://lucene.apache.org/>.
50. Broder AZ, Glassman SC, Manasse MS, Zweig G. Syntactic clustering of the Web. *Computer Networks and ISDN Systems*. 1997; 29(8):1157–1166. [https://doi.org/10.1016/S0169-7552\(97\)00031-7](https://doi.org/10.1016/S0169-7552(97)00031-7)
51. Gupta S, Kaiser G, Neistadt D, Grimm P. DOM-based Content Extraction of HTML Documents. In: *Proceedings of the 12th International Conference on World Wide Web. WWW'03*. New York, NY, USA: ACM; 2003. p. 207–214. Available from: <http://doi.acm.org/10.1145/775152.775182>.
52. LLC PL. Mercury Web Parser by Postlight; 2017. Available from: <https://mercury.postlight.com/web-parser/>.
53. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. Dynamical Classes of Collective Attention in Twitter. In: *Proceedings of the 21st International Conference on World Wide Web. WWW'12*. New York, NY, USA: ACM; 2012. p. 251–260. Available from: <http://doi.acm.org/10.1145/2187836.2187871>.

54. Shao C, Menczer F, Ciampaglia GL. Hoaxy API Documentation; 2017. Available from: <https://market.mashape.com/truthy/hoaxy>.
55. Bessi A, Ferrara E. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*. 2016; 21(11). <https://doi.org/10.5210/fm.v21i11.7090>
56. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab; 1999.
57. Freeman LC. A Set of Measures of Centrality Based on Betweenness. *Sociometry*. 1977; 40(1):35–41. <https://doi.org/10.2307/3033543>
58. Callaway DS, Newman MEJ, Strogatz SH, Watts DJ. Network Robustness and Fragility: Percolation on Random Graphs. *Phys Rev Lett*. 2000; 85:5468–5471. <https://doi.org/10.1103/PhysRevLett.85.5468> PMID: [11136023](https://pubmed.ncbi.nlm.nih.gov/11136023/)
59. Silverman C. Lies, Damn Lies and Viral Content: How News Websites Spread (and Debunk) Online Rumors, Unverified Claims and Misinformation. Tow Center for Digital Journalism; 2015. Available from: <https://towcenter.org/research/lies-damn-lies-and-viral-content/>.
60. Shao C, Menczer F, Ciampaglia GL. Hoaxy Backend; 2017. Available from: <https://github.com/IUNetSci/hoaxy-backend>.
61. Shao C, Wang L, Serrette B, Pentchev V, Menczer F, Ciampaglia GL. Hoaxy Frontend; 2017. Available from: <https://github.com/IUNetSci/hoaxy-frontend>.