

Hoaxy: A Platform for Tracking Online Misinformation

Chengcheng Shao^{*}
School of Computer
National University of Defense Technology,
China

Giovanni Luca Ciampaglia¹,
Alessandro Flammini^{1,2},
Filippo Menczer^{1,2}
¹ Indiana University Network Science Institute
² School of Informatics and Computing
Indiana University, Bloomington, USA

ABSTRACT

Massive amounts of misinformation have been observed to spread in uncontrolled fashion across social media. Examples include rumors, hoaxes, fake news, and conspiracy theories. At the same time, several journalistic organizations devote significant efforts to high-quality fact checking of online claims. The resulting information cascades contain instances of both accurate and inaccurate information, unfold over multiple time scales, and often reach audiences of considerable size. All these factors pose challenges for the study of the social dynamics of online news sharing. Here we introduce *Hoaxy*, a platform for the collection, detection, and analysis of online misinformation and its related fact-checking efforts. We discuss the design of the platform and present a preliminary analysis of a sample of public tweets containing both fake news and fact checking. We find that, in the aggregate, the sharing of fact-checking content typically lags that of misinformation by 10–20 hours. Moreover, fake news are dominated by very active users, while fact checking is a more grass-roots activity. With the increasing risks connected to massive online misinformation, social news observatories have the potential to help researchers, journalists, and the general public understand the dynamics of real and fake news sharing.

Keywords

Misinformation; hoaxes; fake news; rumor tracking; fact checking; Twitter

1. INTRODUCTION

The recent rise of social media has radically changed the way people consume and produce information online [21]. Approximately 65% of the US adult population accesses the news through social media [2, 30], and more than a billion

^{*}Work performed as a visiting scholar at Indiana University.
Email: shaoc@indiana.edu

people worldwide are active on a daily basis on Facebook alone [16].

The possibility for normal consumers to produce content on social media creates new economies of attention [10] and has changed the way companies relate to their customers [35]. Social media allow users to participate in the propagation of the news. For example, in Twitter, users can rebroadcast, or *retweet*, any piece of content to their social circles, creating a competition among posts for our limited attention [37]. This has the implication that no individual authority can dictate what kind of information is distributed on the whole network. While such platforms have brought about a more egalitarian model of information access according to some [4], the inevitable lack of oversight from expert journalists makes social media vulnerable to the unintentional spread of false or inaccurate information, or *misinformation*.

Large amounts of misinformation have been indeed observed to spread online in viral fashion, oftentimes with worrying consequences in the offline world [8, 22, 29, 23, 27, 7]; examples include rumors [18], false news [8, 23], hoaxes, and even elaborate conspiracy theories [1, 13, 19].

Due to the magnitude of the phenomenon, media organizations are devoting increasing efforts to produce accurate verifications in a timely manner. For example, during Hurricane Sandy, false reports that the New York Stocks Exchange had been flooded were corrected in less than an hour [36]. These *fact-checking* assessments are consumed and broadcast by social media users like any other type of news content, leading to a complex interplay between ‘memes’ that vie for the attention of users [37]. Examples of such organizations include Snopes.com, PolitiFact, and FactCheck.org.

Structural features of the information exchange networks underlying social media create peculiar patterns of information access. Online social networks are characterized by homophily [25], polarization [12], algorithmic ranking [3], and social bubbles [28] — information environments with low content diversity and strong social reinforcement.

All of these factors, coupled with the fast news life cycle [14, 10], create considerable challenges for the study of the dynamics of social news sharing. To address some of these challenges, here we present *Hoaxy*, an upcoming Web platform for the tracking of social news sharing. Its goal is to let researchers, journalists, and the general public monitor the production of online misinformation and its related fact checking.

As a simple proof of concept for the capabilities of this kind of systems, here we present the results of a prelimi-

nary analysis on a dataset of public tweets collected over the course of several months. We focus on two aspects: the temporal relation between the spread of misinformation and fact checking, and the differences in how users share them. We find that, in absolute terms, misinformation is produced in much larger quantity than fact-checking content. Fact checks obviously lag misinformation, and we present evidence that there exists a characteristic lag of approximately 13 hours between the production of misinformation and that of fact checking. Finally, we show preliminary evidence that fact-checking information is spread by a broader plurality of users compared to fake news.

2. RELATED WORK

Tracking abuse of social media has been a topic of intense research in recent years. Beginning with the detection of simple instances of political abuse like *astroturfing* [31], researchers noted the need for automated tools for monitoring social media streams. Several such systems have been proposed in recent years, each with a particular focus or a different approach. The Truthy system [32], which relies on network analysis techniques, is among the best known of such platforms. The TweetCred system [9] focuses instead on content-based features and other kind of metadata, and distills a measure of overall information credibility.

More recently, specific systems have been proposed to detect rumors. These include RumorLens [33], TwitterTrails [26], and FactWatcher [20]. The fact-checking capabilities of these systems range from completely automatic (TweetCred), to semi-automatic (TwitterTrails, RumorLens). In addition, some of them let the user explore the propagation of a rumor with an interactive dashboard (TwitterTrails, RumorLens). However, they do not monitor the social media stream automatically, but require the user to input a specific rumor to investigate. Compared with these, the objective of the Hoaxy system is to track both accurate and inaccurate information in automatic fashion.

Automatic attempts to perform fact checking have been recently proposed for simple statements [11], and for multimedia content [6]. At this initial stage, because our focus is on automatic tracking of news sharing, the Hoaxy system does not perform any kind of fact checking. Instead, we focus on tracking news shares from sources whose accuracy has been determined independently. There have been investigations on the related problems of finding reliable information sources [15] and news curators [24].

3. SYSTEM ARCHITECTURE

Our main objective is to build a uniform and extensible platform to collect and track misinformation and fact checking. Fig. 1 shows the architecture of our system. Currently our efforts have been focused on the ‘Monitors’ part of the system. We have implemented a tracker for the Twitter API, and a set of crawlers for both fake news and fact checking websites, as well as a database.

We begin by describing the origin of our data. The system collects data from two main sources: news websites and social media. From the first group we can obtain data about the origin and evolution of both fake news stories and their fact checking. From the second group we collect instances of these news stories (i.e., URLs) that are being shared online.

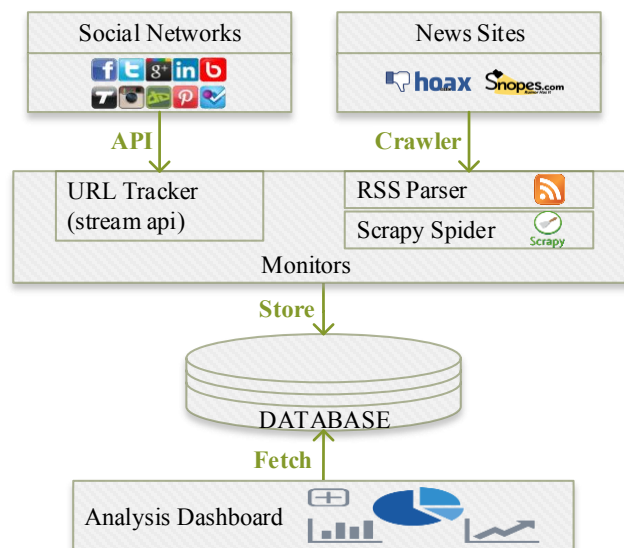


Figure 1: Architecture of the Hoaxy system.

To collect data from such disparate sources, we make use of a number of technologies: Web scraping, Web syndication, and, where available, APIs of social networking platforms. For example, we use the Twitter streaming API to do real-time tracking of news sharing. Because tweets are limited to 140 characters, the most common method to share a news story on Twitter is to include directly a link to its Web article. This means that we can focus only on tweets containing links to specific domains (websites), a task that is performed efficiently by the *filter* endpoint of the Twitter streaming API.¹

To collect data on news stories we rely on RSS, which allows us to use a unified protocol instead of manually adapting our scraper to the multitude of Web authoring systems used on the Web. Moreover, RSS feeds contain information about updates made to news stories, which let us track the evolution of news articles. We collect data from news sites using the following two steps: when a new website is added to our list of monitored sources, we perform a ‘deep’ crawl of its link structure using a custom Python spider written with the Scrapy framework²; at this stage, we also identify the URL of the RSS feed, if available. Once all existing stories have been acquired, we perform every two hours a ‘light’ crawl by checking its RSS feed only. To perform the ‘deep’ crawl, we use a depth-first strategy. The ‘light’ crawl is instead performed using a breadth-first approach. We store all these structured data into a database; this allows convenient retrieval for future analysis, which we plan to implement as an interactive Web dashboard. Unfortunately, URLs do not make for good, unique identifiers, since URLs with different protocol schema, query parameters, or fragments may all refer to the same page. We rely on canonical URLs where possible, and adopt a simple URL canonization technique in other cases (see below).

¹dev.twitter.com/streaming/reference/post/statuses/filter

²scrapy.org

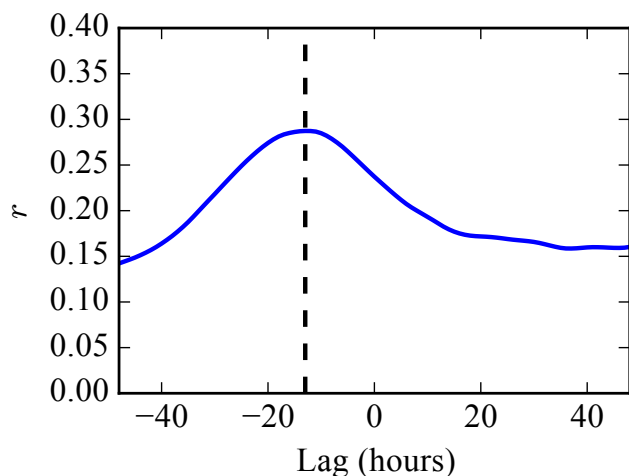


Figure 2: Lagged cross correlation (Pearson’s r) between news sharing activity of misinformation and fact-checking, with peak value at lag = -13 hours.

Table 1: Summary statistics of tweet data.

source	N_{sites}	N_{tweets}	N_{users}	N_{URLs}
fake news	71	1,287,769	171,035	96,400
fact checking	6	154,526	78,624	11,183

4. PRELIMINARY ANALYSIS

In this section we report results from a preliminary analysis performed on a large set of public tweets collected over the course of several months. Since we are interested in characterizing the relation between the overall social sharing activity of misinformation and fact checking, we begin our analysis by focusing on the overall aggregate volume of tweets, without breaking activity down to the level of an individual story or set of stories. We take aggregate volume as a proxy for the overall social sharing activity of news stories.

4.1 Data

We collect tweets containing URLs from two lists of Web domains: the first, fake news, covers 71 domains and was taken from a comprehensive resource on online misinformation.³ We manually removed known satirical sources like *The Onion*. The second list is composed of the six most popular fact-checking websites: Snopes.com, PolitiFact.com, FactCheck.org, OpenSecrets.org, TruthOrFiction.com, and HoaxSlayer.com. The keywords we used to collect all these tweets correspond to the domain names of these websites.

To convert the URLs to canonical form we perform the following steps: first, we transform all text into lower case; then we remove the protocol schema (e.g. ‘http://’); then we remove, if present, any prefix instance of the strings ‘www.’ or ‘m.’; finally, we remove all URL query parameters.

We collected about 3 months of filtered tweets traffic from Oct 14, 2015 to Jan 24, 2016. The summary statistics for the numbers of tweets, unique users, and unique canonical

³fakenewswatch.com

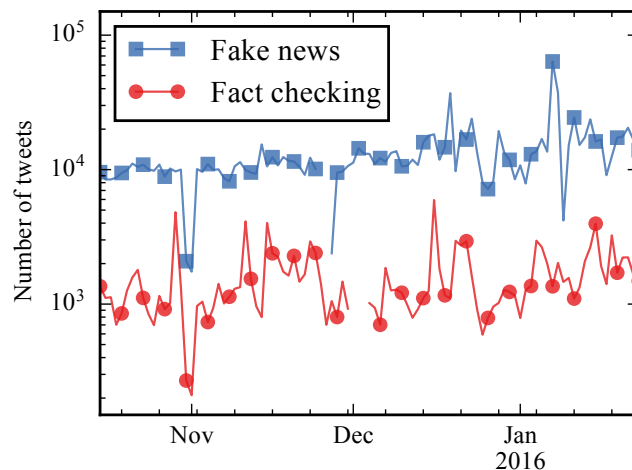


Figure 3: Daily volume of tweets. The gaps correspond to two windows with missing data when our collection script crashed.

URLs (Table 1) illustrate the imbalance between the sets of fake news and fact-checking sites.

4.2 Tweet Volume

Fig. 3 plots the daily volume of tweets. As described before, we track more fake sites than fact checking ones, so the volume of fake news tweets is larger than that of fact checking ones by approximately one order of magnitude.

While both time series display significant fluctuations, the presence of aligned peaks (Nov 16) and valleys (Nov 2) suggests the presence of cross-correlated activity. To better understand this, we perform a lagged cross-correlation analysis, which measures the similarity between two time series signals as a function of the lag.

Fig. 2 shows the results of the cross correlation analysis, with lags ranging from -48 hours to +48 hours. A higher correlation at a *negative* lag indicates that the sharing of fake news *precedes* that of fact checking. To eliminate circadian fluctuations, we use a simple moving average method with centered window of size equal to 24 hours. The results suggest that, in the limited number of examples at our disposal, there is a characteristic time lag between fake news and fact checking of approximately 13 hours. Because the moving average cleaning could only remove circadian fluctuations, we do not exclude the presence of correlations at larger lags (e.g., weekly).

While this cross-correlation is suggestive of a temporal relation between misinformation and fact-checking, it is important to understand that it is based on aggregate data. We selected a subset of URLs from both data sets to see if these correlations also hold at the level of individual events. We followed two strategies: (1) we selected a single URL from our pool of fake news stories and a matching URL from that of fact-checking stories; (2) we considered a small set of keywords and used them to perform pattern matching on the lists of URLs.

Table 2 displays the two URLs (A1, A2) used in the first strategy. The reported story focuses on the current Syrian conflict, and contains several inaccuracies that were debunked in a piece by Snopes.com. For the second strategy,

Table 2: A1: An example of inaccurate news story. A2: Corresponding fact checking page. B1: News articles reporting inaccurate information about the death of actor Alan Rickman. B2: Corresponding fact-checking pages.

A1	www.infowars.com/white-house-gave-isis-45-minute-warning-before-bombing-oil-tankers/
A2	www.snopes.com/2015/11/23/obama-dropped-leaflets-warning-isis-airstrikes/
B1	en.mediamass.net/people/alan-rickman/deathhoax.html www.disclose.tv/forum/david-bowie-alan-rickman-death-hoax-100-staged-t108254.html worldtruth.tv/david-bowie-and-alan-rickman-death-hoax-100-staged/ beforeitsnews.com/alternative/2016/01/alan-rickman-the-curse-of-the-69-takes-another-victim-january-man-predicts-his-death-video-3277444.html beforeitsnews.com/celebrities/2016/01/david-bowie-alan-rickman-death-hoax-100-staged-both-69-died-from-cancer-2474208.html age-69.beforeitsnews.com/alternative/2016/01/harry-potter-star-alan-rickman-dead-at-age-69-3277454.html from-cancer.beforeitsnews.com/celebrities/2016/01/david-bowie-alan-rickman-death-hoax-100-staged-both-69-died-from-cancer-2474208.html
B2	www.snopes.com/2016/01/14/alan-rickman-dies-at-69/ www.snopes.com/alan-rickman-potter-meme/

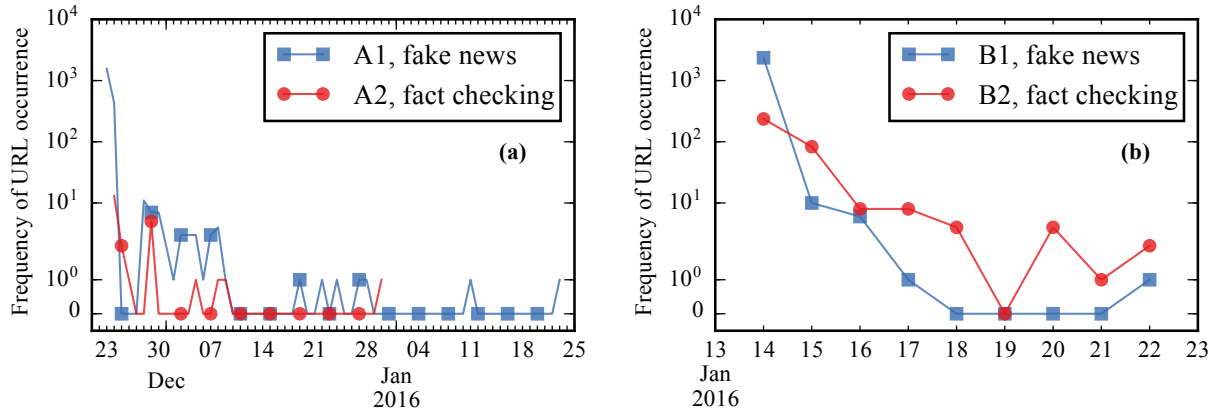


Figure 4: Daily volume of tweets for (a) A1 and A2 (cf. Table 2); (b) B1 and B2. Semi-log scale was used for the plot.

we focused on a recent event: the death of famous actor Alan Rickman on January 14, 2016. We used the keywords ‘alan’ and ‘rickman’ to match URLs from our database, and found 15 matches (B1) among fake news sources and two from fact-checking ones (B2). The fake news, in particular, were spreading the rumor that the actor had *not* died. Fig. 4 plots the volume of tweets containing URLs from both strategies. Despite low data volumes, the spikes of activity and the successive decay show fairly strong alignment.

4.3 User Activity and URL Popularity

We measure the activity a of users by counting the number of tweets they posted, and the popularity of a given story (either fake news or fact checking) by counting either the total number n of times its URL was tweeted, or the total number p of people who tweeted it. Fig. 5 shows that these quantities display heavy-tailed, power-law distributions $P(x) \sim x^{-\gamma}$. We estimated the power-law decay exponents, obtaining the following results: for user activity $\gamma_a^{\text{fn}} = 2.3$, $\gamma_a^{\text{fc}} = 2.7$; for URL popularity by tweets (tail fit for $n \geq 200$) $\gamma_n^{\text{fn}} = 2.7$, $\gamma_n^{\text{fc}} = 2.5$; and for URL popularity by users (tail fit for $p \geq 200$) $\gamma_p^{\text{fn}} = 2.9$, $\gamma_p^{\text{fc}} = 2.5$. These observations suggest that fake news and fact checking have

similar popularity profiles, with fake news being spread by accounts that, in some cases, can generate huge numbers of tweets. While it is expected that active users are responsible for producing a majority of news shares, the strong difference between fake news and fact checking deserves further scrutiny.

In Twitter there are four types of content: *original* tweets, *retweets*, *quotes*, and *replies*. In our data, original tweets and retweets were the most common category (80–90%) while quotes and replies correspond to only 10–20% of the total, usually with slightly more replies than quotes. However, we observe differences between fact checking and misinformation tweets. The first is that there are more replies and quotes among fact checking tweets (> 20%) than misinformation ($\approx 10\%$), suggesting that fact checking is a more *conversational* task.

The second difference has to do with how content generation is shared among the top active users and the remaining user base. To investigate this difference, we select for both fake news and fact checking the tweets generated by the top active users, which we define as the 1% most active user by number of tweets. In Fig. 6 we plot the ratio ρ between original tweets and retweets for all users and top active ones. For

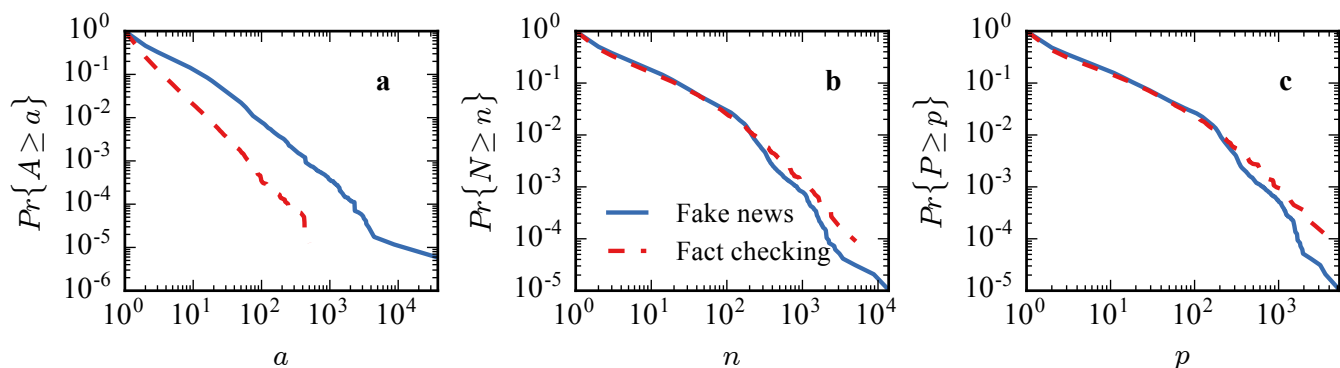


Figure 5: Complementary cumulative distribution function (CCDF) of (a) user activity a (tweets per user) (b) URL popularity n (tweets per URL) and (c) URL popularity p (users per URL).

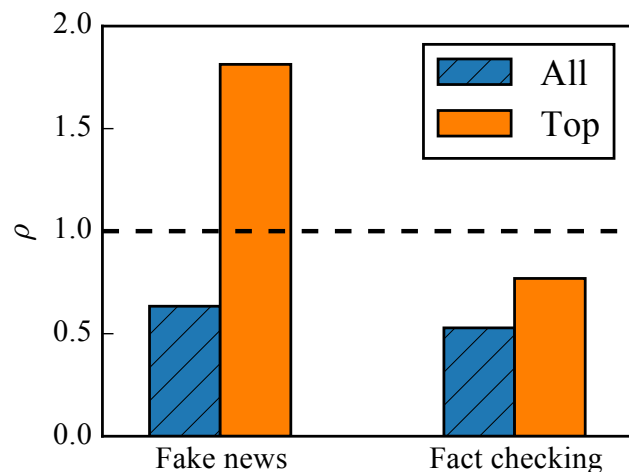


Figure 6: Ratio of original tweets to retweets for all vs. top active users.

all users, the ratio is similar; there are more retweets than original tweets. This is also the case for the top spreaders of fact checking. However, for top spreaders of fake news, this ratio is much higher: these users do not retweet as much but post many original messages promoting the misinformation.

Taken together, these observations strongly suggest that rumor-mongering is dominated by few very active accounts that bear the brunt of the promotion and spreading of misinformation, whereas the propagation of fact checking is a more distributed, grass-roots activity.

5. CONCLUSIONS & FUTURE WORK

Social media provide excellent examples of marketplaces of attention where different memes vie for the limited time of users [10]. A scientific understanding of the dynamics of the Web is increasingly critical [5], and the dynamics of online news consumption exemplify this need, as the risk of massive uncontrolled misinformation grows. Our upcoming Hoaxy platform for the automatic tracking of online misinformation may provide an important tool for the study of these phenomena. Our preliminary results suggest an interesting interplay between fake news promoted by few very

active accounts, and grass-roots responses that spread fact checking information several hours later.

In the future we plan to study the active spreaders of fake news to see if they are likely social bots [17, 34]. We will also expand our analysis to a larger set of news stories and investigate how the lag between misinformation and fact checks varies for different types of news.

Acknowledgments

CS was supported by the China Scholarship Council while visiting the Center for Complex Networks and Systems Research at the Indiana University School of Informatics and Computing. GLC acknowledges support from the Indiana University Network Science Institute (iuni.iu.edu) and from the Swiss National Science Foundation (PBTIP2-142353). This work was supported in part by the NSF (award CCF-1101743) and the J.S. McDonnell Foundation.

6. REFERENCES

- [1] A. Anagnostopoulos, A. Bessi, G. Caldarelli, M. Del Vicario, F. Petroni, A. Scala, F. Zollo, and W. Quattrociocchi. Viral misinformation: the role of homophily and polarization. *arXiv preprint arXiv:1411.2893*, 2014.
- [2] M. Anderson and A. Caumont. How social media is reshaping news. <http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/>, 2014. [Online; accessed December 2015].
- [3] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.
- [4] Y. Benkler. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006.
- [5] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner. Creating a science of the web. *Science*, 313(5788):769–771, 2006.
- [6] C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schiffrer, and N. Newman. Challenges of computational verification in social multimedia. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pages 743–748, 2014.

- [7] A. M. Buttenheim, K. Sethuraman, S. B. Omer, A. L. Hanlon, M. Z. Levy, and D. Salmon. {MMR} vaccination status of children exempted from school-entry immunization mandates. *Vaccine*, 33(46):6250 – 6256, 2015.
- [8] C. Carvalho, N. Klagge, and E. Moench. The persistent effects of a false news shock. *Journal of Empirical Finance*, 18(4):597 – 615, 2011.
- [9] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*, page 675, 2011.
- [10] G. L. Ciampaglia, A. Flammini, and F. Menczer. The production of information in the attention economy. *Scientific Reports*, 5:9452, 2015.
- [11] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini. Computational fact checking from knowledge networks. *PLoS ONE*, 10(6):e0128193, 06 2015.
- [12] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer. Political polarization on Twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [13] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.
- [14] Z. Dezső, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A.-L. Barabási. Dynamics of information access on the web. *Phys. Rev. E*, 73:066132, Jun 2006.
- [15] N. Diakopoulos, M. De Choudhury, and M. Naaman. Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2451–2460, 2012.
- [16] Facebook Newsroom. Company info. <https://web.archive.org/web/20151222043012/http://newsroom.fb.com/company-info/>. Online; accessed December 2015.
- [17] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *Comm. ACM*, Forthcoming. preprint arXiv:1407.5225.
- [18] A. Friggeri, L. A. Adamic, D. Eckles, and J. Cheng. Rumor cascades. In *Proc. Eighth Intl. AAAI Conf. on Weblogs and Social Media (ICWSM)*, 2014.
- [19] S. Galam. Modelling rumors: the no plane pentagon french hoax case. *Physica A: Statistical Mechanics and Its Applications*, 320:571–580, 2003.
- [20] N. Hassan, A. Sultana, Y. Wu, G. Zhang, C. Li, J. Yang, and C. Yu. Data in, fact out: Automated monitoring of facts by factwatcher. *Proc. VLDB Endow.*, 7(13):1557–1560, Aug. 2014.
- [21] A. M. Kaplan and M. Haenlein. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1):59 – 68, 2010.
- [22] A. Kata. Anti-vaccine activists, web 2.0, and the postmodern paradigm—an overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, 30(25):3778–3789, 2012.
- [23] T. Lauricella, C. S. Stewart, and S. Ovide. Twitter hoax sparks swift stock swoon. *The Wall Street Journal*, 23, 2013.
- [24] J. Lehmann, C. Castillo, M. Lalmas, and E. Zuckerman. Finding news curators in twitter. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, pages 863–870, 2013.
- [25] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.
- [26] P. T. Metaxas, S. Finn, and E. Mustafaraj. 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, pages 69–72, 2015.
- [27] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi. Collective attention in the age of (mis)information. *Computers in Human Behavior*, 51, Part B:1198–1204, 2015.
- [28] D. Nikolov, D. Fregolente, A. Flammini, and F. Menczer. Measuring online social bubbles. *PeerJ Computer Science*, 1:e38, 2015.
- [29] B. Nyhan, J. Reifler, and P. A. Ubel. The hazards of correcting myths about health care reform. *Medical Care*, 51(2):127–132, 2013.
- [30] A. Perrin. Social media usage: 2005-2015. <http://www.pewinternet.org/2015/10/08/2015/Social-Networking-Usage-2005-2015/>.
- [31] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [32] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 249–252, 2011.
- [33] P. Resnick, S. Carton, S. Park, Y. Shen, and N. Zeffer. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In *Proc. Computational Journalism Conference*, 2014.
- [34] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, et al. The DARPA Twitter Bot Challenge. *arXiv preprint arXiv:1601.05140*, 2016.
- [35] D. Tapscott and A. D. Williams. *Wikinomics: How mass collaboration changes everything*. Penguin, 2008.
- [36] E. Wemple. Hurricane Sandy: NYSE NOT flooded! <http://wapo.st/1QiG16A>, October 2012. Last accessed 2016-02-04.
- [37] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2, 2012.