”What Draws Your Attention?”: Analyzing the Impact of Duplicate Hoaxes

Sameera Horawalavithana
University of South Florida
Tampa, FL

Sreeja Nair
University of South Florida
Tampa, FL

Essa Alhazmi
University of South Florida
Tampa, FL

Introduction

Information dis-ordering is a long-standing problem in the history of web. More recently, this forgery has attracted a significant media attention in the political periphery with the buzzword of ”fake news”. According to (Lazer et al. 2018) fake news overlay a common ground between misinformation and dis-information. Many fact checking efforts have been invested to mitigate the impact of fake news spread in social networks.

An important question in the process of fact checking is to build methods for resolving the duplicates of fake claims. We observe such duplicate claims that published under the same or different publishers in subsequent time intervals [1] [2]. Further, such duplicate claims help to spread unverified contents much faster than the competing fact checking process.

In general, we also address the continued influence effect of misinformation (Ecker, Hogan, and Lewandowsky 2017). In such phenomena, a false story may be retracted after the verification of fact-checking process. However, the claim behind this story could be repeated with a modified outline. Such that, a similar claim does spread again, making it more familiar to the parties who already exposed in prior. This level of familiarity does strengthen the spread of such false claims. To combat, we need to design more robust re-correction methods in the fact-checking process.

In this study, we first try to detect the existence of duplicate claims published by low-credibility sources. Then, we propose an approach to analyze the characteristics of such groups by constructing a set of *fact networks*.

Related Work

The main objective of our study is to analyze the impact of duplicate hoaxes in Twitter. More recently, Shao et. al. (Shao et al. 2018) studied the mis-information diffusion network in Twitter using Hoaxy (i.e., a web-tool to monitor the spread of low-credible claims). In this work, we also used Hoaxy to collect such low-credible claims, and extract the retweet networks.

There has been many researches on how people respond to corrections of misinformation which is initially considered to be true but later found to be false, and recorrected (Ecker, Hogan, and Lewandowsky 2017). Many such studies conclude corrections rarely change people's mind or perception towards the initial exposure to misinformation: that is, despite being corrected and acknowledged, the people continue to rely at least partially on information they know to be false (Lewandowsky et al. 2012). This phenomena is known as *continued influence* which make the people to believe the misinformation that is repeating, hence familiar.

Further, (Friggeri et al. 2014) show how the variants of rumours repeatedly deceive individuals in Facebook. In this study, we also have similar objectives, but we use Twitter to analyze the impact of duplicate hoaxes available in the Web.

Methodology

In this work we used Hoaxy to collect the misinformation such as rumors, fake news, and hoaxes. Hoaxy is a platform to collect, detect and track misinformation and its related fact checking (Shao et al. 2016).

Near-duplicate Detection Using Shingling To detect the duplicate web pages, the most common approach is to *shingle* the document. Shingling is a technique in which a set of consecutive words from a document is grouped as a single object. A *k-shingle* consists of a consecutive set of k words. For example, consider a document d_i with the text *the cat is an animal*. The (k = 4)-shingles of d_i are *the cat is an*, *cat is an animal*. These objects or tokens are compared with the document corpus to ascertain the similarity between documents. Let us say $S(d_j)$ denotes the set of shingles of document d_j. Now, we calculate Jaccard Coefficient, which measures the degree of overlap between the sets $S(d_1)$ and $S(d_2)$ as $|S(d_1) \cap S(d_2)|/|S(d_1) \cup S(d_2)|$; denote this by $J(S(d_1), S(d_2))$ If the coefficient exceeds a preset threshold, we group them as near duplicates.

In this study, we did perform shingling to the contents of the list of URL collected from Hoaxy dataset. The Jaccard coefficient threshold is set to range 0.7 - 0.9.

[1] https://www.infowars.com/trump-says-he-doesnt-trust-vicious-people-in-the-white-house/

[2] https://www.prisonplanet.com/trump-says-he-doesnt-trust-vicious-people-in-the-white-house.html

Build Fact URL Network Once the duplicate-detection is done, we build the Fact-URL network. An edge is drawn between pairs of URL if its' content similarity score is above the preset threshold (between 0.7 - 0.9). The similarity score of two documents is set as the edge weight. Likewise, we construct a weighted, undirected network with URLs as nodes and Jaccard Coefficient as the edge weight.

Extract Connected Components We extract all the connected components from the Fact-URL network where each component represents a group of URLs that have similar content. Each component is associated with a particular fact, news or hoax. For example, when the documents d_1 and d_2 have high Jaccard coefficient and the documents d_2 and d_3 also have high Jaccard coefficient, then the documents d_1 and d_3 should be near duplicates. That is, d_1, d_2, and d_3 form a group of near-duplicates around a shared fact.

Build Fact Retweet Network We used Hoaxy to collect retweet networks given a URL. Then, we represent each component in the Fact-URL network by merging different retweet networks under the URLs present in the component. We construct Fact-RT networks to the maximum number of the components in Fact-URL network. Note that, some URLs don't have any reshares in Twitter.

We visualize the retweet networks to better understand the structural patterns of the cascades. Further, we analyze the temporal growth of such networks with the support of article meta-data (e.g., publication date)

Data Characteristics

In this section, we present network characteristics of two fact networks; Fact URL and Fact Retweet (RT) networks.

Fact URL Network

Fact URL network represents nodes as URLs, and an edge reflects whether two documents under particular URLs are similar. The network is weighted, where the edge strength represents the similarity score. As we described earlier, we used a near-duplicate detection algorithm based on shingling to measure this similarity between text documents.

It is important mention that the Fact-URL network is sensitive to the pre-set threshold of similarity. The size of Fact-URL network is dis-proportionate to the similarity threshold. The correlation is non-linear, where a unit change of similarity threshold may change the size of Fact-URL network significantly. The standard similarity threshold is 0.5 in web documents (Schütze, Manning, and Raghavan 2008). The selection of this threshold is specific to the domain, and agnostic to the dataset.

Empirically, we observe some documents share the same promotional text. This leads the near-duplicate detection algorithm to output many *false negatives*. We try to mitigate such scenarios by setting up a strict similarity threshold of 0.7. However, such threshold is entirely specific to this dataset, which might have consequences. As an example, we could have been skipped some true positives in the similarity threshold between 0.5 - 0.7.

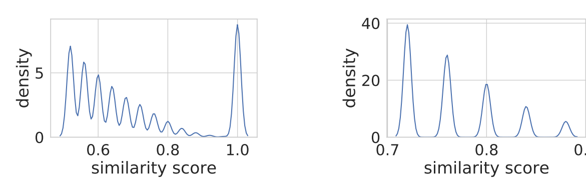
Table 1 describes the structural properties of Fact-URL network. Note that, the network is reduced to 14,799 nodes

Network Metric	Value
Number of nodes	14,799
Number of edges	402,661
Density	0.0036
Transitivity	0.7065
Assortativity	0.55
Number of connected components	3269

Table 1: Basic characteristics of Fact-URL network

(URLs) from the initial corpus of 100,000 documents for two reasons. First, we clean the 15% URLs not to have referenced tags (e.g., starts with ?), and make them uniform to have "https://" instead of "http://". We also remove URLs which don't include any text content. Then, we feed the cleansed document corpus to the duplicate detection algorithm, and extract $2.4M$ document pairs associated with the standard similarity threshold of 0.5.

Second, the size of the corpus is reduced significantly when we control the similarity threshold. With the standard similarity threshold of 0.5, there are $32,045$ nodes in the Fact-URL network, which is reduced into $14,799$ nodes when the similarity threshold is between 0.7 and 0.9. We increase the similarity threshold to 0.7 due to the appearance of promotional text in the documents. We understand such mechanism is not robust, but our objective was to get a dataset that's qualitative enough to proceed with the rest of our work. We also don't include perfect duplicates (where the similarity threshold is above 0.9). Majority of such perfect duplicates are from the same website which are coupled with either a page in the sub-domain or their own accelerated mobile page (amp). Table 1 presents the distribution of similarity scores in the duplicates.



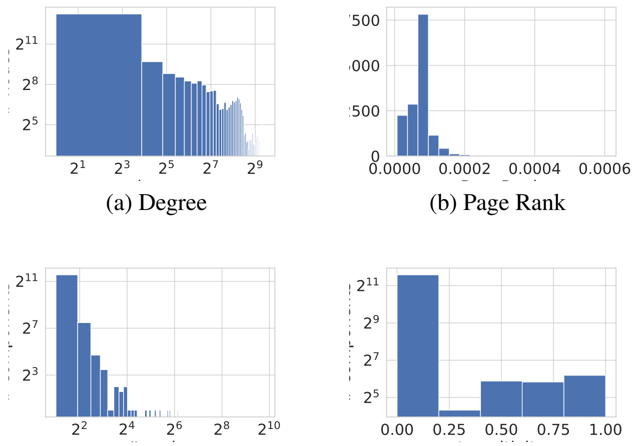
(a) Duplicate Pairs ($t_x \geq 0.5$) (b) Duplicate Pairs ($0.7 \geq t_x \geq 0.9$)

Figure 1: Kernel Density Estimate of the similarity scores of document pairs extracted from near-duplicate detection algorithm

As presented in Table 1, the Fact-URL network is very sparse, where we observed 3269 connected components in total. However, it does show significant level of clustering. This is biased towards few larger cliques that we observed in the Fact-URL components as shown in Figure 2d.

We observe a positive value 0.55 of assortativity (i.e., degree-based assortativity coefficient) in the Fact-URL network (Table 1). This is due to many components of single-edge, where two URLs are found to be duplicates (Figure 2c). In such cases, the joint probability of observing an edge between a similar degree pair is increased.

Figure 2b presents the distribution of page-rank scores. We observe a node with the maximum page-rank score of 6×10^{-4}.



(a) Degree (b) Page Rank

(c) Component Size (d) Component Transitivity

Figure 2: Structural Properties of Fact URL Network

Figure 3 visualizes three connected components of Fact-URL network that we found interesting. Each component represents a set of URLs which has an identical claim (e.g., engagement of a celebrity, death of a celebrity etc.). More interestingly, all the documents are published under a same web-domain, *mediamass*. We had a closer look into the domain, and found it as a popular web-site producing unverified content for mass consumption [3]. Specifically, they modify claims found in popular gossip articles to reflect all the celebrities in their database. As an example, they would say a popular actor Harrison Ford got engaged [4] and the same news appeared for the popular singer Shawn Mendes [5]. Figure 3a visualizes the structure of network under this engagement hoax.

More often, they placed these hoaxes strategically, where they refer an external source as a verification. As an example, they referred "American Sun-times" as the source who report such news of Harrison Ford's engagement hoax. In Shawn Mendes's case, it was "Canadian Sun-times" since the artist is Canadian. Further, the web-site is denied to accept the ownership of the claim, where they end the article asking whether the "Sun-times" starts a rumour. Such deceptive strategies would keep these articles alive and active until fact-checker's response.

Similarly, Figure 3b and Figure 3c visualize the Fact-URL network components under the pregnancy claims of celebrities and surgery claims of celebrities' dogs. When compared with other two networks, the component network of the surgery claims looks different. As shown in Figure 3c, it shows two sub-groups that are connected. In this component, there are 376 nodes (or similar URLs) and 13412 edges (or duplicate pairs). As we discovered, there are two bridges in this component, connecting articles related with English, Jamaican and French celebrities. Such that, sub-groups contextualize the claims according to the celebrities' nationality.

Fact Retweet Network

We use Twitter to show the reactions on duplicate hoax articles. Usually, people share such hoax article URLs in Twitter, and others react to such by *re-tweeting* - a way to share original content to a broader community. As we discussed earlier, we used Hoaxy (Shao et al. 2018) to collect the retweet network given a URL. Such retweet network is in the pattern of a star, where Twitter don't share actual retweet cascades to the open public. (Note that, there are existing methods to estimate such retweet cascades (Vosoughi, Roy, and Aral 2018))

We merge multiple retweet networks of the URLs observed in the same component in the Fact-URL network. Such that, we construct 3269 Fact retweet (RT) networks, which is the same number of components appeared in the Fact-URL network.

Figure 4 presents the structural properties of Fact-RT networks. As expected, many RT networks are small, where 50% of such networks have at most 50 nodes.

The largest RT network has 10873 nodes, and 13016 edges. There are 60 connected components in this network, where the largest component covers 98% of the nodes. This network shows a negative assortativity value of -0.33, and the transitivity is 8.6×10^{-5}. We note the largest RT network was under the story of Hillary Clinton's elite pedophile scandal [6]. The duplicate documents are found in accelerated mobile pages (amp). However, one of the articles contain a description about the writer, which leads to have a similarity score of 0.88. This is a typical example that our method of avoiding amp pages from the similarity threshold does not work as expected.

There are many Fact-RT networks remain almost connected, where the largest connected component covers more than 90% of the network (Figure 4e). As shown in Figure 4f, majority of the networks have a small number of components.

Many Fact-RT networks shows negative assortativity, one reason for this behavior is the existence of star networks for retweets. We observe a larger fraction of networks has an assortativity value of 0.5. When there are multiple star networks overlapped, we no longer observe a single hub dominated in the network.

Figure 5 visualizes three example Fact-RT networks along with the details of the spatio-temporal dynamics of the evolving network. We have three interesting observations.

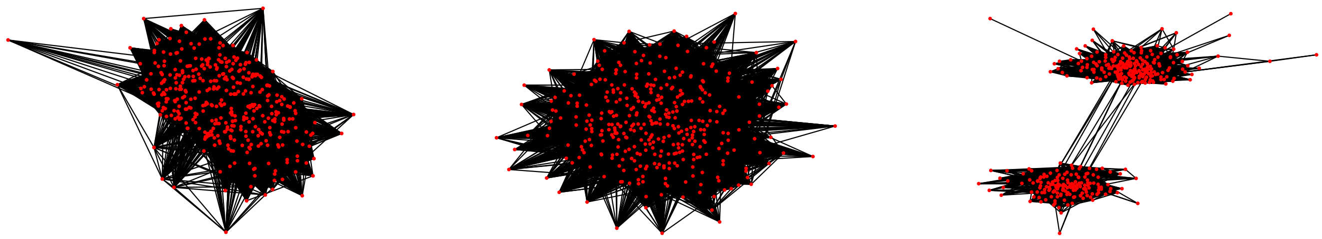
First, the reaction in Twitter is most frequently originated by the first published article in the group of duplicates. However, the amount of such reactions achieves a peak during

[3]https://en.mediamass.net/blog/mediamass-project

[4]https://en.mediamass.net/people/harrison-ford/engaged.html

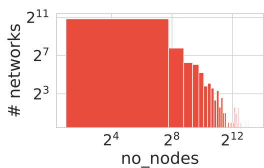
[5]https://en.mediamass.net/people/shawn-mendes/engaged.html

[6]https://newspunch.com/hillary-clinton-pedophile-ring-state-department/

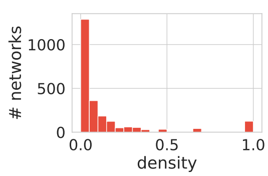


(a) Engagement Hoax (b) Pregnancy Hoax (c) Dog Surgery Hoax

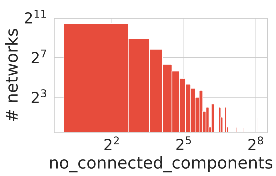
Figure 3: Visualization of Fact URL Networks on three examples of hoax news chains



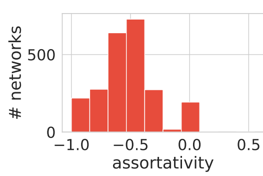
(a) Number of nodes



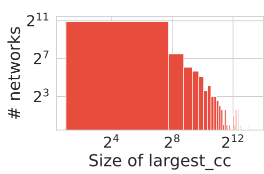
(b) Density



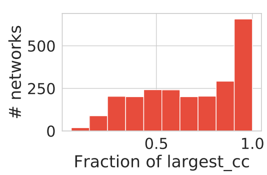
(c) No. of connected components



(d) Assortativity



(e) Size of largest connected component



(f) Fraction of largest connected component

Figure 4: Structural Properties of Fact RT Networks

their initial burst, and drops rapidly after the attention. The ”attention seekers” could take advantage from this behavior. Figure 5b demonstrates such example, where the first article published in `ActivistPost` [7], and the duplicate article published in `DC ClothesLine` [8] a day after when the subject attention achieves a peak in Twitter.

Second, a group of similar articles could act simultaneously to spread a claim much faster in a community. Such articles might be written by the same author, but published in different places to maximize the breadth of attention. Figures 5c and 5d shows such example, where the similar articles published in `Dangerous` and `InfoWars` are written by the same author (we note the the nick-name under the article writer is same). As shown in Figure 5d, this coupled activity leads to have a burst of reactions in Twitter which is much faster than the earlier example. In the network visualized in Figure 5c has a transitivity value of 1.2×10^{-2}, which is three magnitude higher than the clustering observed in the largest Fact-RT network. This is a good example of the competing contagion process where individuals exposed to multiple cascades. We also note that the fact-checkers verify the credibility of the above claim much later at the end of initial burst period (3 days after the article published date) [9].

Finally, we observe that duplicate articles are always not successful on supporting the spread of unverified claims. As Figure 5f shows, three similar articles are published in three consecutive dates. First published article (`InfoWars`) initiated the attention while second published article (`YourNewsWire`) was the beneficial party in the peak reaction burst in Twitter. However, the last article published in `TheFreeThoughtProject` at the end of initial burst period could not take the same attention forward, but leads to have a very short burst.

Discussion

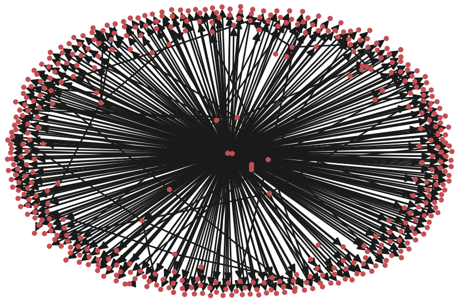
In our study, we used the claims published in low-credible web-sites as identified by Hoaxy. Similar to Shao et. al. (Shao et al. 2018), we also analyze the spread of such claims in Twitter. More specifically, we characterize the misinformation network in Twitter with respect to duplicate hoaxes.

Our approach is to build and characterize the Fact Retweet (RT) networks with the identification of URLs that have duplicate contents. We mainly identify the ”attention seekers”

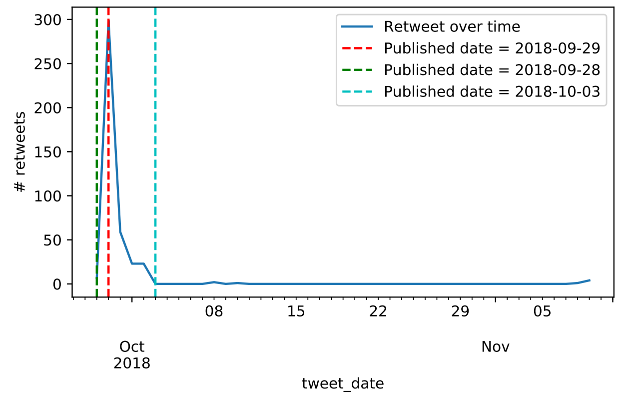
[7]https://www.activistpost.com/2018/09/reddit-now-quarantining-users-who-question-9-11-direct-users-to-govt-site-instead.html

[8]https://www.dcclothesline.com/2018/09/29/reddit-now-quarantining-users-who-question-9-11-direct-users-to-govt-site-instead/

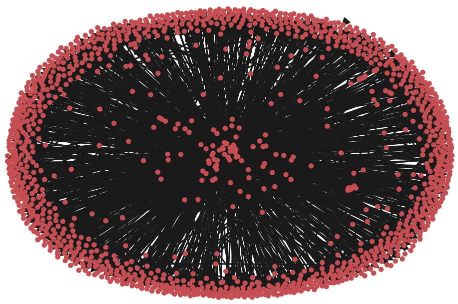
[9]https://www.snopes.com/fact-check/christine-blasey-ford-psychologist/



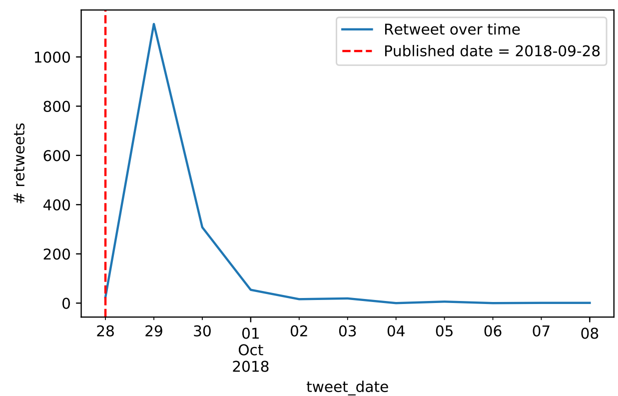
(a) Reddit 9/11 Quarantined (Fact-RT Network)



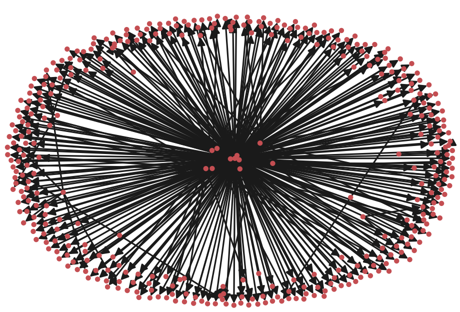
(b) Reddit 9/11 Quarantined (No. retweets over time)



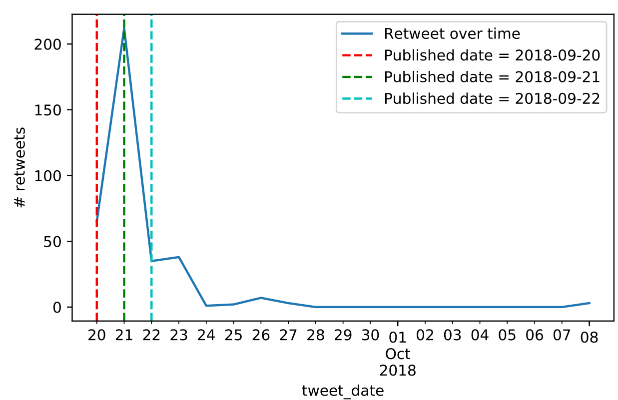
(c) Dr. Ford's commited perjury (Fact-RT Network)



(d) Dr. Ford's commited perjury (No. retweets over time)



(e) WikiLeaks last free generation (Fact-RT Network)



(f) WikiLeaks last free generation (No. retweets over time)

Figure 5: Visualization of Fact RT Networks on three examples of low-credible claims. RT networks represents the spread of multiple URLs that have similar content published in the claims of,(a,b) Reddit 9/11 Quarantined at `ActivistPost`,`DC ClothesLine`,`YourNewsWire` and `TheFreeThoughtProject` (c,d) Dr. Ford's committed perjury at `Dangerous` and `InfoWars`, (e,f) WikiLeaks on last free generation at `InfoWars`,`TheFreeThoughtProject`,`YourNewsWire` and `WorldTruth`.

who're in the process of spreading low-credible claims via replicating original content. We characterize the temporal patterns of such attention seekers, and detect such "attention seekers" could act as a group to strengthen the spread of *false claims*. Further, we observe a repeated claim has a fixed life-time, such that it could not remain active longer than the verification delay of the fact-checkers. We also analyze the structural properties of Fact-RT networks to quantify the impact of such attention seekers. Our code is available to download in Github (Horawalavithana, Nair, and Alhazmi 2018).

References

Ecker, U. K.; Hogan, J. L.; and Lewandowsky, S. 2017. Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition* 6(2):185–192.

Friggeri, A.; Adamic, L. A.; Eckles, D.; and Cheng, J. 2014. Rumor cascades. In *ICWSM*.

Horawalavithana, S.; Nair, S.; and Alhazmi, E. 2018. Duplicate detection on hoaxy dataset. `https://github.com/SamTube405/D-Hoaxy`.

Lazer, D. M.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. 2018. The science of fake news. *Science* 359(6380):1094–1096.

Lewandowsky, S.; Ecker, U. K. H.; Seifert, C. M.; Schwarz, N.; and Cook, J. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest* 13(3):106–131. PMID: 26173286.

Schütze, H.; Manning, C. D.; and Raghavan, P. 2008. *Introduction to information retrieval*, volume 39. Cambridge University Press.

Shao, C.; Ciampaglia, G. L.; Flammini, A.; and Menczer, F. 2016. Hoaxy: A platform for tracking online misinformation. *CoRR* abs/1603.01511.

Shao, C.; Hui, P.-M.; Wang, L.; Jiang, X.; Flammini, A.; Menczer, F.; and Ciampaglia, G. L. 2018. Anatomy of an online misinformation network. *PloS one* 13(4):e0196087.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science* 359(6380):1146–1151.