

3

TRUTH AND THE DYNAMICS OF NEWS DIFFUSION ON TWITTER

Robert Ackland and Karl Gwynn

Introduction

This chapter investigates two aspects of misinformation: how to determine whether information (such as a news story) is true, and how the truthfulness of information affects its diffusion or spread. The chapter has a particular focus on the significance of social media for misinformation (in particular fake news): its prevalence, impact, and methods for identifying and studying the phenomenon. We review recent literature on how computational methods and “big data” sources (e.g., social media) are being used for identifying misinformation and understanding how people engage with and spread misinformation.

Our empirical application involves a new approach for manually checking the truthfulness of news stories, and we apply this method to a sample of Australian political news stories from 2017. We then explore how the veracity of news affects its diffusion (via retweets) on Twitter, focusing on the following key measures of diffusion: reach (how many people are involved in the diffusion), speed, and breadth (how far into the network does the news spread, and how diverse are the actors involved in the diffusion).

Background

In this section, we first review existing definitions of misinformation and fake news and then summarize the reasons why people contribute to spreading misinformation. We then summarize research on how social media potentially exacerbates the problem of fake news. Approaches to checking the veracity of news are then outlined and assessed in their ability to accurately determine measurements of truthfulness, with a particular focus on manual approaches (truthfulness checked by domain experts) and computational approaches. We also summarize

some key research that uses computational methods and big data to understand the phenomenon of misinformation.

Definition of misinformation and fake news

Misinformation refers to false information that has the capacity to spread through society and influence public opinion. Examples of misinformation are satire news (designed for humor and not intended to deceive), rumors (pieces of information that have not yet been confirmed as true or false), conspiracy theories (which by definition are not verifiable, and tend to be spread by people who believe them to be true), and hoaxes (which are designed to deceive, may be humorous or malicious, and often involve citing a trusted source). Misinformation, like all information, is *piecemeal* and subject to revision when newer knowledge becomes available.

Contemporary usage of the term “misinformation” has become politicized, following misinformation scandals such as the “Obama’s birth certificate” and “Clinton’s child sex ring” incidents. While misinformation can refer to any publicly accessible erroneous information, the use of the term today generally implies malintent and deception.¹ For the purpose of this chapter, misinformation will refer to all instances where information can be verified as containing clear falsehoods.

Fake news is a type of misinformation where the information relates to a news event, and malintent is present on behalf of the person(s) creating the news, but not necessarily on behalf of the person(s) spreading the news. Allcott and Gentzkow (2017) define fake news as news articles that are “intentionally and verifiably false”. Production of fake news is generally motivated by financial or ideological gain as it can mislead readers into believing that false news content is true. However, the term fake news is also sometimes used in reference to legitimate publications as a method of discreditation and defamation (this is a notable tactic by the US president Donald Trump); this is suggestive of the sensitivity of the topic and capacity for the word to be misapplied.²

Motivations for spreading misinformation

Why do people spread inaccurate information? Often, they believe information to be truthful due to heuristic biases. Heuristics are the tendencies of individuals to rely on simplistic patterns to reduce the expenditure of critical thought. This is evident in the reliance on prior beliefs and opinions: if the information confirms these priors, it is more likely to be believed (confirmation bias) and hence potentially spread (e.g., Ecker, Lewandowsky, Fenton, & Martin, 2014). Lewandowsky, Ecker, Seifert, Schwarz, and Cook (2012) have identified four factors that influence whether a person believes information: consistency of message (is it consistent with prior beliefs?), coherency of message (is it internally coherent and plausible?), credibility of source, and general acceptability (how many other people appear to believe it?). Another reason for spreading misinformation

relates to normative pressures whereby people spread misinformation in order to gain social affirmation and acceptance: this relates to social identity theory (e.g., Tajfel & Turner, 2001, 2004). Once someone believes misinformation, it is difficult to change these beliefs (Lewandowsky et al., 2012), and attempts to correct falsifications may even perpetuate misinformation spread, particularly within ideological groups (Nyhan & Reifler, 2010). This supports the need for accurate and timely detection of false information, and has motivated the building of systems for detecting misinformation.

Fake news in the digital age

Although fake news is not a new phenomenon, there are several reasons why it is of growing importance and concern in the digital age (see, e.g., Allcott & Gentzkow, 2017; Shu, Sliva, Wang, Tang, & Liu, 2017).

First, barriers to entry in news media have dropped significantly as websites can be easily set up and monetized via advertising. Regarding the process of spreading of fake news, the fixed costs associated with getting on social media are very small – this increases the viability of short-term strategies involving establishing a social media presence for a particular fake news campaign, and reduces incentive for establishing long-term presence associated with quality journalism. Second, social media are well suited to dissemination of fake news; the format of social media is such that information tends to be distributed in short snippets of text, which makes it harder for users to assess veracity. Third, there has been a continued decline in public trust and confidence in mainstream media. Fourth, in many western countries there has been a rise in political polarization (degree of negative feelings oriented to the other side of the political spectrum) and this can increase the likelihood of fake news being believed.

The increase in political polarization is related to another important aspect of social media that may be affecting the extent to which people are exposed to fake news. In early research into political communication, Katz and Lazarsfeld (1955) contended that media-savvy individuals (“opinion leaders”) were intermediaries between mass media and the public (the “two-step flow of communication”). However, social media has led to a reduction in the presence or importance of intermediaries between producers and consumers of information (this is referred to as “disintermediation”) with people now able to create and share information via online social networks. It is much easier to have fine-grained control over particular information sources (e.g., follow users on Twitter who share your political beliefs) and this “narrowcasting” can lead to the creation of so-called echo chambers: groups of like-minded users who are not subject to outside views, which can lead to greater polarization (difference in attitudes). Related to this is the phenomenon of “filter bubbles”: algorithms used by social media companies select new content for users based on their previous engagement with content, thus reinforcing information consumption patterns and making it less likely that users are exposed to new information.³

Echo chambers and filter bubbles are relevant to the fake news problem because they can affect the likelihood of a person transmitting fake news in the following ways. First, one is more likely to be connected to other people who evidently believe the news (this leads to social credibility and reinforcement). Second, one is more likely to be exposed to the fake news story, and increased exposure has been found to increase the likelihood of belief (Hasher, Goldstein, & Toppino, 1977), and less likely to be exposed to information that would counter an ideologically aligned but fake news story. Finally, there is increased normative pressure to spread fake news (even if one does not believe it).

Allcot and Gentzkow (2017) found that fake news was heavily tilted in favor of Donald Trump in 2016; their database contains 115 pro-Trump fake news stories that were shared on Facebook 30 million times, and 41 pro-Clinton fake news stories that were shared 7.6 million times on Facebook. Guess, Nyhan, and Reifler (2018) study how political identity affected consumption of fake news during the 2016 US presidential election. The authors found that Trump supporters were disproportionately inclined to visit websites hosting fake news, and that this was due to the fact that fake news was largely targeted at Trump supporters, and hence was attitude-consistent (and thus likely to be consumed).⁴

The dangers of fake news are that it can lower trust in democratic institutions, reduce social cohesion, and contribute to the rise of populist leaders (some commentators have attributed a significant role to fake news in the election of Donald Trump in 2016). However, Vargo, Guo, and Amazeen (2018) look at another potential impact of fake news – its potential for shaping the online news landscape. The authors used the Network Agenda-Setting (NAS) conceptual framework, which posits that the news media can influence how the public connects or relates issues to one another (e.g., energy crisis is related to foreign-relation problems) and also the salience or popularity of issues. Fake news could have an agenda-setting impact on news media (particularly partisan news media) simply by generating misinformation that needs to be responded to by journalists (e.g., through fact checking). They test the agenda-setting potential of fake news using the GDELT Global Knowledge Graph, a network of people, locations, themes, and events computationally constructed from global news (Leetaru & Schrodt, 2013). Using Granger causality tests,⁵ the authors found that fake news was successful in transferring issue salience to online media for particular issues (e.g., international relations) during the period 2014–2016. They also distinguished partisan news media outlets and found that conservative media transferred issue salience to fake news media, and this in turn drove the agenda of liberal media (who were responding to fake news).

Overview of approaches for verifying news

There are two broad approaches to establishing the veracity of news.⁶ The first approach involves the news content itself – features of the text (and images, if

present) of the news item, and also its source. The news content approach to assessing the veracity of news can be further delineated by knowledge-based and style-based approaches.

Knowledge-based approaches involve using external sources to check the claims made in the news item (“fact checking”), and there are three main variants of fact checking.

- (1) Fact checking by domain experts (the present chapter involves an implementation of expert-oriented fact checking and so we provide a summary of relevant literature in the next section).
- (2) Crowdsourcing is fact checking involving an aggregated consensus of members of the general public. Applications such as Fiskkit and the LINE account “For real” allow user suggestion and comment to provide an indication of news truthfulness, and there is considerable interest in the use of distributed ledger technology (blockchain) for crowdsourcing fact checking.
- (3) Knowledge graphs such as DBpedia or Google Knowledge Graph are networks showing the connection between real-world entities (people, places, things). Fact checking can be automatically performed by comparing the content of news stories (and in particular, relationship between entities mentioned in the news story) with content existing within the knowledge graph. The primary issue with this method is that while it is fast and accurate, it relies on information sources that cannot incorporate all knowledge.

Style-based approaches to checking news veracity involves linguistic analysis of text content in the news itself. This includes lexical features (e.g., number of words, average word length, number of unique words) and syntactic features (e.g., n-grams, parts of speech). Domain-specific linguistic features, such as external links and the presence of tables and graphs, may also be useful. Images can also be used to evoke particular emotional responses (anger, shock), which increase the likelihood of believing the news, and so features extracted from images are also used in detecting fake news.

The second approach for checking veracity of news involves using data on the social context of news, that is, how it is consumed and shared. This is where social media has had a major impact, because it allows for fine-grained data on the social context in which news is being consumed and spread. First, features relating to the users who have engaged with news may be used in detecting fake news. In the case of Twitter, such features may include the number of friends/followers, age of user account, and the number of tweets authored. Second, with social media it is also possible to collect post-level data – the reactions that people have to news items (e.g., quoted retweets when forwarding news URLs, comments on Facebook posts) – and these can provide useful information for detecting fake news. These reactions can be mined for linguistic features as discussed previously, but there is also the possibility of making use of social media data on the debate or contestation surrounding particular news items. Third, there are

network features that can be used to detect fake news, for example, networks of the diffusion of news on Twitter via retweets (retweet cascades).

Fact checking by domain experts

Fact checking is frequently employed as a method of determining the reliability of statements and news, especially when there is a perceived risk of misinformation. Fact checking has risen to prominence over the past decade as websites such as Politifact.com and Factcheck.org have become established political fact verifiers. The recent inclusion of fact checking during media coverage of the 2016 US presidential election and the Trump presidency highlights the ability of fact checking to confront misinformation and fake news. The Australian political system has its own fact-checking outlets, such as RMIT University-ABC Fact Check and The Conversation's FactCheck.

The fact-checking industry has grown from a minor attraction during election cycles to a prominent element of the political sphere. Young, Jamieson, Poulsen, and Goldring (2018) identify the 1988 presidential election as the first example of contemporary fact checking being used to monitor the behavior of potential presidential candidates. This was framed as “adwatches” where fact checking was televised drawing on the combined knowledge of academics and media professionals. What started as a footnote gained traction as the American political environment adopted rhetorical techniques designed to mislead and misinform the public.

The spread of political misinformation was a prominent area of academic debate during the 2000s, particularly regarding the US invasion of Iraq and supposed stockpiling of weapons of mass destruction (WMDs). The revelation of government-coordinated misinformation spurred research into how individuals determine true from false and the ease in which politicians deceive the public.

Nyhan and Reifler (2010) make a binary distinction between the “uninformed” and “misinformed” subject: uninformed subjects react through heuristic biases whereas misinformed subjects have received false information through a political affiliation. Reversing misinformation through “corrective information” is generally unsuccessful because subjects have embedded knowledge within their political identity. This is a far more challenging issue than an uninformed public whose lack of knowledge is not embedded in conceptions of political and social identity.

Coinciding with the rapid growth of the internet, fact checking became increasingly prevalent online as it could quickly produce responses to political misinformation. Factcheck.org was launched in 2003 by the Annenberg Public Policy Center with the goal of confronting misinformation through the combined ability of journalism and scholarship. Following the success of Factcheck.org, websites like Politifact and Fact Checker (*Washington Post*) were established with similar methods of detecting misinformation.

Up until this point, fact checking had remained a relatively well-regarded source of information that increased accountability in politics. While the

effectiveness of fact checking in terms of audience engagement has been looked at by authors such as Nyhan and Reifler (2010), the epistemological and methodological foundations of fact checking have not been explored extensively. Uscinski and Butler (2013) provide a critique of fact-checking practices, stating that “These practices share the tacit presupposition that there cannot be genuine political debate about facts” (p. 163).

Uscinski and Butler (2013) contend that fact checking takes a complex inter-relationship between politics, policy, society, economics, and history, reducing it to the most simplistic of metrics that assigns truth across a unidimensional spectrum. If a radically interpretative epistemology is adopted, truth cannot be dictated by a politician or a fact checker when a complex reality allows for varying conceptions of the truth. The authors outline key methodological failings of fact checking that are defined by a simplistic objectivist epistemology. These include selection effects (where the selection of facts is usually based on sensationalist or headlining stories), multiple fact/part of a fact concerns (defining the measurable parameters of a single fact), casual claims (assertion of an unknown relationship between facts), future predictions, and selection criteria (What constitutes truth? How true does something need to be?). The authors refrain from providing a simple answer to the failures of the fact-checking methodology. Instead they re-emphasize the interpretative nature of facts and the complex nature of political rhetoric and information.

Amazeen (2015) is critical of the generalizations and sampling used in Uscinski and Butler’s (2013) critique of fact checking. While acknowledging the interpretative nature of fact checking, Amazeen (2015) suggests that Uscinski and Butler (2013) selected examples of fact checks that most clearly support their opinion. Amazeen (2015) claims that many facts are beyond debate and that interpretative facts are actually a small minority of the overall facts analyzed, and highlights the consistency across separate agencies in their fact check of common stories. However, for Uscinski (2015) the consistency across fact-checking agencies is simply a reflection of fact checkers sharing political biases and access to information. Uscinski (2015) argues that fact-checker consistency merely indicates their collective approach to fact checking rather than an empirical validation of inter-agency consistency. Uscinski (2015) further questions the very role of the “fact checker”, as such actors lack qualifications, or are “epistemologically naïve” to believe that no qualification are needed and that the truth is easily accessible (p. 247).

Wu, Agarwal, Li, Yang, and Yu (2017) attempt to overcome some of the methodological failings of fact checking by introducing computational methods. The authors propose “query perturbations” as a way of avoiding issues of “cherry-picking” within fact checking – a problem highlighted by Uscinski (2015) as a key failing of fact checking given its reliance on unjustified selection criteria. What query perturbations aim to do is extend the parameters of the fact to see if the claimed truth still holds up under different levels of measurement. The example used by Wu et al. (2017) is the claim by New York’s former mayor

Rudy Giuliani that adoption rates grew by 65–70% during his time in power. This is true given a very particular measurement (1990–1995, 1996–2001 as two grouped sets of data) but when adjusted for the actual time Giuliani was mayor (1994–2001), there was in fact a 1% decrease in adoptions during his term. This indicates the ways in which data can be manipulated to present truths that are statistically correct but contextually inaccurate.

Computational approaches to studying misinformation

This section summarizes recent studies that use computational approaches to either identify misinformation or study its spread and impact.

Most computational approaches for detection of misinformation such as fake news on social media use machine learning classification to predict whether a news article is true or not. The first automated approaches to detecting misinformation on the internet were in the context of detection of problematic emails and website text content (e.g., spam and hoaxes). These approaches generally involved applying supervised machine learning approaches to text. For detection of spam emails, for example, this involves constructing a training dataset consisting of emails that have been manually coded as spam and non-spam, and a classifier (e.g., logistic regression, neural network) is used to predict the likelihood of an email being problematic, based on extracted features, for example, keywords, or patterns in sentence structure.

Castillo, Mendoza, and Poblete (2011) investigate the use of automated methods to assess the credibility (“offering reasonable grounds to be believed”, p. 675) of news-related tweets. Human coders first identified a set of tweets relating to news events and a second set of coders labeled the tweets as to whether they believed they were true or false. The authors extracted a set of features from tweets that are related to the preceding social context approach: *post-based features* (e.g., length of tweet, use of punctuation such as exclamation marks, positive/negative sentiment of the tweet, whether it is a retweet, whether it contains a hashtag); *user-based features* (age of account, number of followers and following, number of tweets authored); and *propagation-based features* (number of retweets). It was found that credible information was more likely to be spread by users with newer accounts, who are more active (in terms of tweets) and with many followers and followees. Positive sentiment, as well as the presence of question marks and smiley emoticons, in the tweets spreading the story were associated with less credible news stories, while the presence of a URL in the tweet was associated with higher credibility. Regarding propagation-related features, tweets having many retweets are more likely to be judged as credible.

Tacchini, Ballarin, Della Vedova, Moret, and Alfaro (2017) developed an approach to classify posts on public Facebook pages as containing valid science or conspiracy theory science, using data on which Facebook users had “liked” the different posts.⁷ Their approach relies on an assumption of assortativity in behavior of Facebook users; the users will tend to group or cluster by liking

similar posts, based on their preferences toward consuming information containing scientific fact or conspiracy theory. However, as noted by the authors, this approach requires as input some information from the content of the post since *a priori* it is not known whether a set of Facebook users liking the same post are doing it because they value the science or the conspiracy.

Vosoughi, Mohsenvand, and Roy (2017) investigated supervised learning approaches to automatically verify rumors on Twitter. Their dataset consisted of 209 manually selected and annotated rumors relating to real-world events, and they focused on features related to linguistic style of the tweets, the users spreading the rumors, and the dynamics of propagation. They were motivated to build a system that could be used for real-time rumor identification, and their system correctly predicted the veracity of 75% of rumors faster than trusted public sources (e.g., journalists, law enforcement officials). To measure propagation, they devised an approach to reconstruct the retweet cascade networks (the Twitter API does not provide retweet cascades, instead connecting all retweet events to the original tweet, regardless of the actual chain of retweets) using data on timestamps of retweets and following relationships between people who retweeted.⁸ They found that the temporal dynamics of features had significant predictive power, and since these dynamics are generally invariant to the size of a rumor (e.g., the total amount of attention it garnered), this allows their approach to generalize to events and rumors of various sizes.

A key finding was that propagation features did most of the “heavy lifting” in terms of prediction of veracity of rumor, and they found that rumors that were eventually found to be true tended to propagate by high-influence users (in terms of number of followers) retweeting low-influence users. A justification for this finding was that a high influence person would not risk retweeting a tweet from a lesser-known person, unless confident that the rumor was in fact true. Another key finding was that tweets of rumors subsequently found to be false tended to exhibit a bi-modal distribution in terms of language sophistication (compared to other tweets in their corpus) and the authors related this to intent of the spreaders; spreaders of malicious rumors tend to use sophisticated language to make the rumor more legitimate and believable, whereas spreaders of rumors that were non-malicious tended to be careless in their language and hence lacking in linguistic sophistication. Finally, they found that false rumors tend to be spread by users who are influential and controversial (where the controversiality of users is computed by first measuring the sentiment of up to 1,000 replies to the user and then constructing a score whereby users with many replies and an even mix of negative and positive replies are rated as more controversial), whereas spreaders of rumors subsequently found to be true tend to be influential and less controversial.

Vosoughi, Roy, and Aral (2018) studied the propagation behavior of true and false news on Twitter (as noted previously, they eschewed the use of the term “fake news”), using the approach for reconstructing retweet cascade trees used in Vosoughi et al. (2017). They found that false news spreads deeper (measured as

the length of the longest sub-tree within a retweet cascade), farther (the unique number of retweeters of a story, which is a measure of how many people consumed the news), and more broadly (the maximum width of a sub-tree). They also found that false news diffused faster, with the truth taking about six times as long as falsehood to reach 1,500 people. These patterns were especially pronounced for fake political news, compared with news regarding, for example, terrorism, natural disasters, or other topics.

The authors found that network and user characteristics did not explain the markedly different diffusion profiles of true and false news. The people spreading false news were less connected (fewer followers and followees), less active on Twitter, and had been on Twitter for shorter periods of time; falsehood was 70% more likely to be retweeted, compared with truth, even after controlling for these factors. They therefore investigated the impact (on diffusion) of a particular characteristic of the news itself: novelty. They measured novelty of news by comparing the topic distribution (identified using a Latent Dirichlet Allocation topic model) of the news stories in their sample with the tweets that users were exposed to in the 60 days prior to their retweeting of sample news stories. They found that false news is more novel than true news, and suggested that the differential diffusion patterns could be due to people preferring to share novel information (regardless of its veracity) because novelty attracts attention (it allows us to update our understanding of the world) and also there may be social benefits to spreading novel information (one is “in the know”).

Shin, Jian, Driscoll, and Bar (2018) studied the dynamic process of misinformation on social media by tracking the life cycle of 17 political rumors that circulated on Twitter during the 2012 US presidential election. Consistent with the findings of Vosoughi et al. (2018), Shin et al. (2018) found that false rumors tended to re-emerge on Twitter (often with “mutations”, i.e., textual changes) exhibiting multiple spikes in attention over time, whereas true rumors did not, tending to have a single prominent pike of attention. The authors proposed three potential reasons for these different temporal patterns. First, rumor spreaders may feel that false rumors needed more “help” in gaining wider acceptance and hence require repeated attempts at spreading. Second, since true rumors tend to originate from mainstream media outlets, rumor spreaders may feel that they’ve “exhausted” their potential readership and hence not allocate resources to their further spread. Finally, spreading false rumors may be for the purpose of identity signaling, rather than persuasion, and repeated attempts at spreading false rumors reflects people “participating in a common epistemological sphere” (p. 285) thus promoting group bonding.

In addition to examining propagation features as signatures for identifying fake news, researchers have also modeled the spread of misinformation using formal models of contagion processes. For example, Tambuscio, Ruffo, Flammini, and Menczer (2015) use stochastic epidemic models to model the spread of a hoax as a virus. This approach implicitly views the spread of misinformation as a simple contagion process (only requires one contact between “adopter/infected”

and “non-adopter/non-infected” person for there to be adoption or transmission), but Törnberg (2018) instead draws on the concept of “complex contagion” (multiple sources of reinforcement may be required to induce adoption), which is useful for describing the spread of behaviors such as social movements and avant-garde fashion (Centola & Macy, 2007). Törnberg (2018) contends that the spread of fake news involves complex contagion since a person’s decision whether to spread or not can involve group identity processes. The simulation models directly address the potential contribution of echo chambers (which as noted earlier can reinforce group identity) to the spread of fake news.

Some researchers have focused specifically on the role of social bots in spreading misinformation. For example, Ferrara (2017) analyzed activity of social bots during a disinformation campaign relating to the 2017 French presidential election. Contrary to expectations, Vosoughi et al. (2018) found that Twitter social bots and humans exhibit similar behavior in terms of sharing true and false news. However, even if social bots do not have a greater direct impact on the spread of fake news (compared with humans), their contribution to the spread of fake news can indirectly influence more humans to believe and hence spread the fake news. By propagating falsehoods, social bots make it more likely that people will encounter this fake news, contributing to the perception that the fake news is endorsed by many people, and thus promoting further circulation of fake news.

Application: how does truthfulness of news affect the dynamics of its diffusion?

This section presents an analysis of the diffusion on Twitter (via retweets) of a sample of Australian news stories, with an objective of assessing how the truthfulness of a news story affects the dynamics of its diffusion.

Twitter retweet and following network data

The dataset is a collection of retweets of news stories published by three Australian media sources on three randomly selected days during 2017, from Ackland, O’Neil, and Park (2019).⁹

The steps for collecting the data were:

- (1) Randomly select one weekday from three consecutive months in 2017 (we checked the sampled day was not dominated by a particular news event): May 22, June 16, and July 4.
- (2) Collect all of the news stories tweeted by the brands on the sampled days.
- (3) Collect all of the retweets of the news stories, over the next seven days.
- (4) Collect the following edges (or ties) among the retweeters of political stories (we label these people “political retweeters”).
- (5) Collect the following edges from political retweeters to Australian federal politicians.

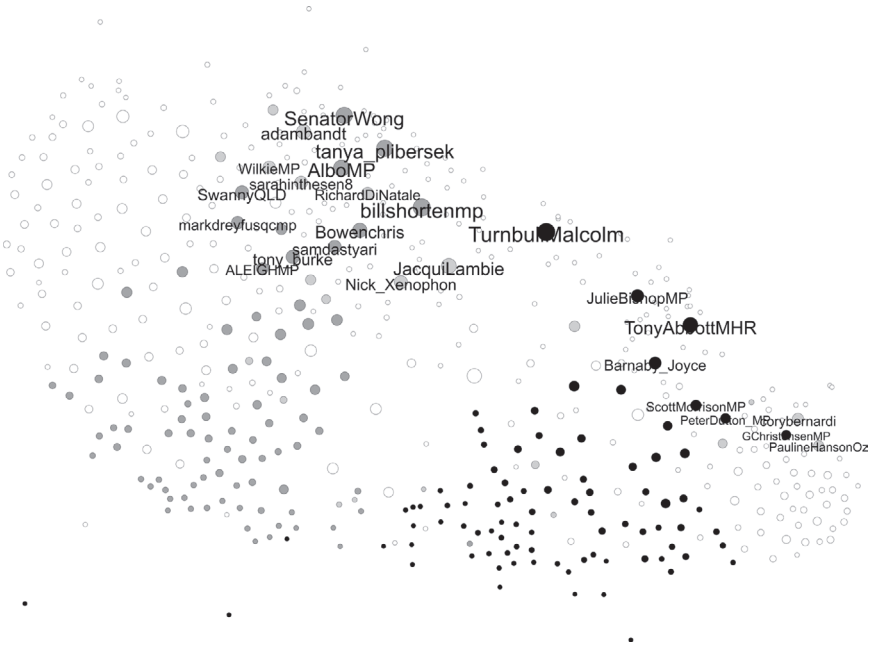


FIGURE 3.1 Twitter following network: political news story retweeters and Australian federal politicians

Note: Dark gray – Labor (political left), black – Liberal-National Coalition (political right), gray – other parties, white – political retweeters. To improve the readability of the network, edges are not displayed.

The giant component of the Twitter following network (political retweeters and Australian federal politicians) is shown in Figure 3.1; this is the network over which diffusion (via retweets) of news stories is measured. The number of nodes (edges) is 495 (12,467) and node size is proportional to indegree. The nodes in the network have been positioned using a force-directed algorithm whereby nodes that are connected (via following edges) are drawn closer together (as if the edges were springs), and nodes that are not connected are pushed apart (as if the nodes have electrostatic charges and repulse one another). Force-directed algorithms can reveal clustering in networks, and there is a marked “Divided They Follow” phenomenon displayed here with strong evidence of clustering on the basis of political ideology.

Selection of news stories

We initially planned to select news stories that contained at least one fact or statistic that is capable of being verified. However, a pilot test of the fact-checking approach indicated that we needed a further selection criteria because traditional news coverage relies heavily on quotes and observed facts, and the capacity to

include misinformation is therefore relatively low. Impartiality in news content is a product of twentieth-century norm shifts following the Second World War. In response to propaganda during the war, media outlets and government sought to realign the medium of news presentation. The Commission on the Freedom of the Press (1947) developed a norm of analytical news with opinion pieces forming a separate area in newspapers (Metzgar & Hornaday, 2013). We therefore decided to focus our fact checking on opinion pieces, since these are spaces in which individualism and partiality are expected. Finally, another factor influenced our selection of news stories for fact checking: since our intention was to analyze the diffusion (via retweets) of news stories on Twitter, we also chose articles with a large number of retweets.

Fourteen articles were analyzed in total, with five from the *Sydney Morning Herald*, five from the *Huffington Post*, and four from *The Australian* (Table 3.1). Opinion pieces were the primary focus, although traditional news analysis was included to expand the sample size, as well as for comparative purposes: eight

TABLE 3.1 Sampled news stories with truth score and diffusion measures

<i>Story id</i>	<i>Article</i>	<i>Brand</i>	<i>Date</i>	<i>Format</i>	<i>Truth score</i>	<i>Reach</i>	<i>Speed</i>	<i>Breadth</i>
1	ALP Gender Equality	Aus	16/6/2017	Opinion	4	1	n.a.	n.a.
2	Turnbull Midwinter Ball	Aus	16/6/2017	Opinion	4	10	2.5	2.67
3	Too Late for Turnbull	Aus	22/5/2017	Opinion	5	13	3.0	2.52
4	Medicare	Aus	16/6/2017	Opinion	4	1	n.a.	n.a.
5	Scott Morrison Economy	SMH	22/5/2017	Opinion	5	1	n.a.	n.a.
6	Off the Record	SMH	16/6/2017	Opinion	4	1	n.a.	n.a.
7	Citizenship Test	SMH	16/6/2017	Opinion	2	8	64.3	2.26
8	Tony Abbott Manifesto	SMH	4/7/2017	Opinion	5	4	3.8	1.57
9	Manus Island	HP	22/5/2017	Opinion/ Traditional	4	28	6.3	2.79
10	James Ashby Denial	HP	22/5/2017	Opinion/ Traditional	4	12	1.8	2.63
11	Illegal Medicare Machines	HP	4/7/2017	Opinion/ Traditional	4	7	1.8	2.45
12	Fake Refugees	HP	22/5/2017	Opinion/ Traditional	5	14	4.0	2.00
13	Adani Mine Investment	SMH	22/5/2017	Traditional	5	47	3.5	2.90
14	Manus Settlement	HP	16/6/2017	Traditional	4	4	5.0	1.59

Note: Aus – *The Australian*, SMH – *Sydney Morning Herald*, HP – *Huffington Post*.

articles were classified as “opinion” and six articles classified as “traditional news” (the *Huffington Post* articles were classified as a hybrid).

Fact checking of news stories

We devised and implemented an original fact-checking methodology that attempts to code news stories according to their adherence to the truth. The primary consideration of the methodology was how to fit an interpretative model of analysis within a largely quantitative framework. If fact checking was truly objective or at least had proven consistency, then this would be less of a concern. However, as indicated by Uscinski and Butler (2013), this is far from reality. To avoid, or at least mitigate, some of the inherent problems with current fact-checking approaches, we used the following work flow (Figure 3.2) to assign each news article a score on a Likert scale ranging from 1 (very untruthful) to 5 (very truthful):

- (1) Are any of the statistics or facts not supported by evidence or information from reliable sources?¹⁰
- (2) What is the significance of the statistic or fact that was not verified? Does it have minor/moderate/high impact on the validity of the article and/or the overarching discourses the article engages in?

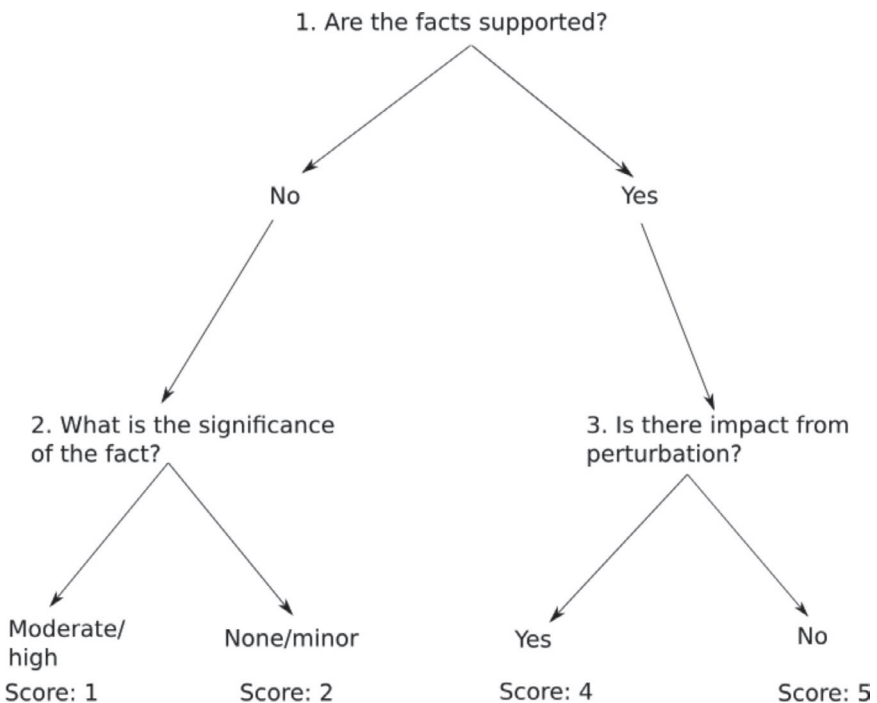


FIGURE 3.2 Fact checking workflow

- (3) In the case where a fact or statistic has been verified, do reasonable perturbations to the parameters of the fact or statistic modify or compromise the underlying intent or framing of the story, in a way that is reliable and sensible?

The preceding method includes some of the foundational elements of a fact check, such as verifiability and adherence to the truth (step 1). Fact checking cannot be assumed to be an objective account of truth, and framing it in this way diminishes its legitimacy and value. Uscinski and Butler (2013) are primarily critical of fact checking because it tends to promote a voice of authority and certainty. The perspective of the fact checker is underemphasized, which implicitly suggests an empiricism in fact checking. While these points are sometimes viewed as strengths of fact checking, in the context of this research it would have diminished the applicability of the fact checks. The preceding fact-checking method highlights its interpretative limitations and avoids implying empirical truths.

Steps 2 and 3 attempt to incorporate a simplified version of query perturbation proposed by Wu et al. (2017). Step 2 assesses the impact of untrue claims by classifying facts based on their contextual relevance: if the argument relies heavily on an untrue claim then it receives a reduction in score. Step 3 perturbs the parameters of a fact to see if truthfulness is a result of statistical manipulation or cherry-picking. This is done through researcher analysis rather than computational methods so it cannot be directly compared with query perturbations proposed by Wu et al. (2017), however it is in the same spirit.¹¹

The workflow can be briefly demonstrated using the “ALP Gender Equality” article that was published in *The Australian* on June 16, 2017. The statement “Bill Shorten’s support for gender equality within Labor . . . is floundering at the national level of the party organisation” must pass step 1 to be considered more than an unjustified critique. The article then mentioned the existence of an “affirmative action report card obtained by *The Weekend Australian*”: as this claim was verified, the article therefore passed step 1 in the workflow. The article then claimed “Most of the senior roles in Mr Shorten’s office are filled by men”. Only when considering step 3 does this article’s utilization of facts become less convincing. The preceding facts have been used in a way that can be impacted by justifiable perturbations. When comparing the Labor Party to its competitor the Liberal/National Coalition, gender imbalances are relatively minor. The article bases its argumentation on a normative critique of gender levels in a party whose representation is better than its closest comparison. For this reason, the article fails step 3 and receives an overall score of 4.

Table 3.1 shows the “truthfulness” rating for the 14 news articles (the complete fact checks are available from the authors on request). There is a reasonably strong adherence to the truth, with an average score of 4.2. As anticipated, traditional news stories adhered closer to the truth, compared with opinion pieces; the latter are often written by politicians, who self-monitor misinformation to

a lesser degree than media, whose reputations lie in their ability to provide the truth. The average score was quite similar across the news platforms: *The Australian* had an average of 4.3, while *The Sydney Morning Herald* and *Huffington Post* both had an average score of 4.2. Given the interpretative nature of analysis, a close look at the outliers can give an indication of what news containing substantial misinformation looks like. The “Citizenship Test” article received the lowest fact check score of 2 out of 5. This is due to a clear manipulation of facts and context, and sensationalistic rhetoric. Based on this sample of news articles, we conclude that although there are some examples of misinformation spread by new media outlets, it is not common, at least among these news brands.

Diffusion of news stories on Twitter

As noted previously, Vosoughi et al. (2018) reconstructed retweet cascade trees and then measured diffusion of true and false news stories across these trees. We take a different approach here by instead looking at how the news stories diffused (via retweets) across the network of following edges shown in Figure 3.1. For a given story, we do not know the exact pathway of diffusion (because we have not reconstructed the retweet cascade trees), but we do know exactly who in the following network retweeted the story, and when. Figure 3.3 shows the following network, with black nodes indicating those Twitter users who retweeted

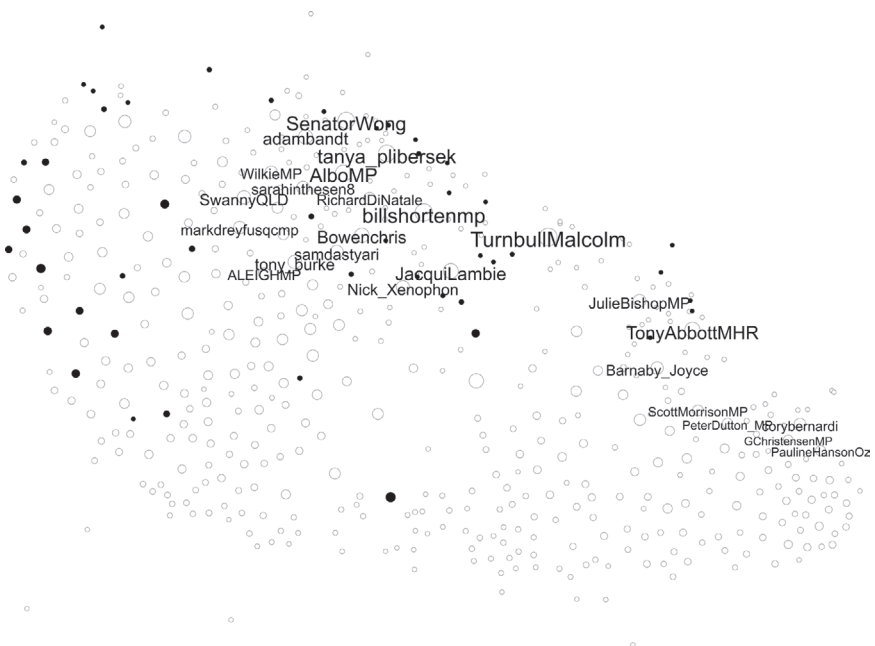


FIGURE 3.3 Twitter following network – black nodes indicate users who retweeted the “Adani Mine Investment” story

the “Adani Mine Investment” story (this was the story with the highest number of tweets – 47). It is apparent that this story was mainly retweeted by Twitter users who are located in the political left part of the map (they predominantly following Labor Party politicians), although there is a small number of retweeters of the story who are located close to prominent Liberal-National Coalition (right-wing) politicians.

We assessed how the truthfulness of the story affects its diffusion in terms of the following measures:

- (1) Reach: Total number of retweets of the story
- (2) Speed: Number of hours taken for 50% of total retweets to be achieved
- (3) Breadth: Mean geodesic distance, or average shortest path length calculated across all of the pairs of retweeters of the story

A challenge for our analysis is the fact that we do not have many data points: we only coded 14 news stories because of the time-consuming nature of fact-checking, but also because we only collected news stories from three sampled days, there was a limited supply of potential stories that fit our selection criteria stated earlier. Further, there was not much variation in the truthfulness rating of the stories: all but one story were coded 4 or 5.

The three diffusion measures (reach, speed, and breadth) are reported in Table 3.1, and it is immediately apparent that despite our best intentions in selecting news stories that had received significant retweets, four stories were retweeted only once by Twitter users in our sample. In reality, these stories were retweeted more than this, but our sample only includes those Twitter users who have (1) retweeted at least one of the political stories and (2) are connected by a following edge to at least one other political retweeter or else to an Australian federal politician. These four stories were excluded from the following analysis.

Based on the research summarized previously (especially Vosoughi et al., 2018) our expectation was that false stories would have greater diffusion (more retweets), diffuse faster, and have greater breadth of diffusion. The preceding data issues mean we cannot provide a rigorous analysis of the diffusion patterns of true versus false news stories; the intention here is to illustrate the approach, rather than provide conclusive evidence. The one story that was given a score of 2 (untruthful) – “Citizenship Test” – had a reach of eight retweets (lower than the average for all stories of 15 retweets), took 64 hours to achieve 50% of the total number of retweets (much higher than the average of ten hours), and by the end of the period the average length of the shortest path (in the following network) between those people who had retweeted the story was 2.26 (lower than the average breadth for all stories of 2.34). Thus, on all three measures, the pattern of diffusion for the one untruthful story in the sample, in comparison to the average, was the opposite of what we expected. However, as already noted, data limitations are such that we are presenting our application

as a methodological contribution rather than a definitive test of the diffusion behavior of true and false news.

Conclusion

In this chapter, we reviewed the state-of-the-art research in the area of misinformation and social media, focusing on approaches for identifying misinformation (and its most topical or newsworthy form today, fake news), and also research that aims to further understanding about how misinformation spreads and its consequences. Although our emphasis has been on computational approaches involving social media data, we also reviewed current research into manual fact checking of news stories.

We then provided an empirical application aimed at characterizing the diffusion patterns of true and false Australian political news stories. The application demonstrated a new approach for fact checking news stories and we also attempted to demonstrate that it is possible to study diffusion of news stories even in the absence of retweet cascade trees (which are not provided by Twitter, and are computationally challenging to reconstruct). However, our empirical application was challenged by the fact that we had few data points (coded stories), and there was not a lot of variation in the truthfulness ratings. We offer our empirical application as a demonstration of a new approach to studying fake news using Twitter data.

Notes

- 1 Some authors distinguish misinformation and disinformation, where the former may involve an actor spreading inaccurate information that they believe is true, and the latter involves a conscious attempt to deceive. In practical applications, disinformation is generally treated as a subset of misinformation, as it is more difficult to ascertain the motives of actors transmitting inaccurate information.
- 2 Noting the current politicization of the term “fake news”, Vosoughi et al. (2018) eschew its use in favor of the term “false news”.
- 3 Bakshy, Messing, and Adamic (2015) compare the relative effects of echo chambers and filter bubbles on a person’s exposure to politically diverse content on Facebook.
- 4 To our knowledge, it has not been tested whether Trump supporters were on a per story basis more inclined to consume fake news, or whether it was simply that they were exposed to a greater volume of attitude-consistent fake news.
- 5 A time series X is said to “Granger cause” time series Y if lagged values of X are statistically significant in a regression predicting future values of Y .
- 6 The following draws from Shu et al. (2017).
- 7 Vosoughi et al. (2018) argue that research into diffusion of science and conspiracy science stories does not allow understanding of factors affecting spread of true versus false news, since by definition, a conspiracy story cannot be verified as false or true.
- 8 The authors note that both Castillo et al. (2011) and Kwon, Cha, Jung, Chen, and Wang (2013) investigated the use of propagation features in predicting veracity of tweets, but these authors did not impute the retweet cascade tree and so were unable to use propagation features as sophisticated as those used by Vosoughi et al. (2017).
- 9 The three Australia media sources were *The Australian*, the *Sydney Morning Herald*, and the *Huffington Post*. These sources aim to cover a broad range of political affiliation, with *The*

- Australian* being on the political right, the *Sydney Morning Herald* being on the left, and the *Huffington Post* potentially on either side.
- 10 Reliable sources refer to evidence that comes from either a primary source or a secondary source whose integrity can be determined based on historical evidence of their truthfulness. We assessed this based on the standing of the source, as determined through external indicators such as historical veracity, accountability, and retrospective corrections. The undertaking of fact checking with this methodology must avoid all agendas and biases as much as possible. This is impossible to completely eradicate, so the verified facts must be considered an interpretive account of true and false.
 - 11 The coding was conducted by one of the co-authors, but in a larger-scale study it would be preferable to have multiple fact checkers, and to compute intercoder reliability.

References

- Ackland, R., O'Neil, M., & Park, S. (2019). Engagement with news on Twitter: Insights from Australia and Korea. *Asian Journal of Communication*, 29(3), 235–251.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–235.
- Amazeen, M. A. (2015). Revisiting the epistemology of fact-checking. *Critical Review*, 27(1), 1–22.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). *Information credibility on twitter*. Proceedings of the 20th International Conference on World Wide Web (WWW), Hyderabad, India.
- Centola, D., & Macy, M. W. (2007). Complex contagion and the weakness of long ties. *American Journal of Sociology*, 113(3), 702–734.
- Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, 42(2), 292–304.
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22(8).
- Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign. Retrieved from www.ask-force.org/web/Fundamentalists/Guess-Selective-Exposure-to-Misinformation-Evidence-Presidential-Campaign-2018.pdf
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, 16, 107–112.
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Glencoe, IL: Free Press.
- Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). *Prominent features of rumor propagation in online social media*. Proceedings of the 13th International Conference on Data Mining (ICDM), IEEE, 1103–1108.
- Leetaru, K., & Schrodt, P. A. (2013). *GDELT: Global data on events, location, and tone, 1979–2012*. International Studies Association Meetings, San Francisco.
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Metzgar, E. T., & Hornaday, B. W. (2013). Leaving it there? The Hutchins commission and modern American journalism. *Journal of Mass Media Ethics*, 28(4), 255–270.

- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.
- Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83, 278–287.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22–36.
- Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & Alfaro, L. (2017). *Some like it Hoax: Automated fake news detection in social networks*. Proceedings of the Second Workshop on Data Science for Social Good, Skopje, FYR Macedonia. CEUR Workshop Proceedings, Volume 1960.
- Tajfel, H., & Turner, J. C. (2001). An integrative theory of intergroup conflict. In M. A. Hogg & D. Abrams (Eds.), *Key readings in social psychology: Intergroup relations: Essential readings* (pp. 94–109). New York, NY, US: Psychology Press.
- Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Key readings in social psychology: Political psychology: Key readings* (pp. 276–293). New York, NY, US: Psychology Press.
- Tambuscio, M., Ruffo, G., Flammini, A., & Menczer, F. (2015). *Fact-checking effect on viral Hoaxes: A model of misinformation spread in social networks*, WWW 2015 Companion.
- Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS One*, 13(9).
- Uscinski, J. E. (2015). The epistemology of fact checking (is still naive): Rejoinder to Amazeen. *Critical Review*, 27(2), 243–252.
- Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact checking. *Critical Review*, 25(2), 162–180.
- Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 20(5), 2028–2049.
- Vosoughi, S., Mohsenvand, M. N., & Roy, D. (2017). Rumor Gauge: Predicting the veracity of rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11(4), 50.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wu, Y., Agarwal, P. K., Li, C., Yang, J., & Yu, C. (2017). Computational fact checking through query perturbations'. *ACM Transactions on Database Systems*, 42(1), 1–41.
- Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, 95(1), 49–75.