Niko Giantzaklidis

Fake News In The Real World

Bachelor's Thesis in Mathematical Information Technology

July 10, 2019

University of Jyväskylä

Faculty of Information Technology

Author: Niko Giantzaklidis

Contact information: nikrgian@student.jyu.fi

Title: Fake News In The Real World

Työn nimi: Valeuutiset oikeassa maailmassa

Project: Bachelor's Thesis

Page count: 33+0

Abstract: The phenomenon of fake news and its role in spreading misinformation has dominated the public discourse in the United States for several years now. This paper examines fake news and finds consistent and predictable patterns that can be used by machine learning algorithms. However, despite the display of consistent, predictable behaviour, fully automated pre-emptive solutions appear to be unviable in the current sociopolitical climate.

Keywords: fake news, social media, misinformation

Suomenkielinen tiivistelmä: Valeuutiset ovat ilmiönä dominioineet poliittista keskustelua Yhdysvalloissa jo muutaman vuoden ajan. Tämä tutkielma tutkii valeuutisia ja havaitsee toistuvia elementtejä, joita voisi käyttää koneoppimismenetelmissä. Tutkielman johtopäätös on, että täysin automatisoidut, ennalta ehkäisevät menetelmät eivät ole toteuttamiskelpoisia tämän hetkisessä sosiopoliittisessa ympäristössä.

Avainsanat: valeuutiset, sosiaalinen media, misinformaatio

List of Figures

List of Tables

Contents

1 Introduction

In 2017 the American Dialect Society named *fake news* as its 28th annual word of the year (American Dialect Society 2018). The phenomenon itself, although not particularly new, took social media by storm during the 2016 U.S. Presidential election and the phrase has been a frequent punchline in President Trump's Twitter diatribes. Since then, fake news has been used in information warfare to subjugate protests in Sudan (Lister, Shukla, and Elbagir 2019), incite violence towards an ethnic minority in Myanmar (Mozur 2018), and has even led to a shooting in the United States (Goldman and Adam 2016).

This paper examines the problem of fake news and explores possible solutions by using the 2016 presidential election as a point of reference. Chapter 2 establishes a definition for fake news and discusses the sociopolitical climate and norms that need to be taken into account when dealing with the phenomenon. Chapter 3 focuses on the data that is either collected or accessible to social networks and can potentially be utilizable to tackle the issue. Chapter 4 highlights several propositions on how to tackle the issue by using the data considered in chapter 3 while taking into account the challenges described in chapter 2.

This paper utilizes a list of fake news sites (See Appendix A, Table 1) which was made by combining sites found in the papers by Allcott and Gentzkow 2017 and Guess, Nagler, and Tucker 2019. The list contains 29 formerly and currently active sites that shared fake news[1] and some information collected about them which will be discussed in later chapters.

1. Out of the 29 sites shared fake news, 20 were retrieved from Guess, Nagler, and Tucker 2019, 8 from Allcott and Gentzkow 2017, and 1 was shared between both papers. See Table 1 for more information.

2 Background

2.1 Defining and distinguishing fake news

During the last few years the phrase "fake news" has been used as a buzzword to describe various phenomena ranging from satirical pieces all the way to political propaganda (Edson et al. 2018, p. 138). The frequency by which "fake news" appear in news, discussions and political discourse underline the need to study the phenomenon.

Every definition has its own limitations and grey areas. For instance, one of the more widely cited definitions of fake news is: "..news articles that are intentionally and verifiably false, and could mislead readers" (Allcott and Gentzkow 2017, p. 213). It is important to emphasize that Allcott and Gentzkow, like other similar definitions, use a definition that highlights the fact that these stories should be verifiably false. The problem with this caveat is that it implies that the burden of proof lies on the party disputing the claim rather than origin of the claim. A claim that is not verifiably false shouldn't be interpreted as being more true than a claim that is. That said, it is important to note that Alcott and Gentzkow use their definition within the context of the 2016 U.S. presidential election. Within that context the definition is a lot more applicable when filtering out other types of misleading or false statements.[1] For the scope of this paper, I define fake news as: *the intentional creation or circulation of a false sentiment under the guise of being factually accurate news for social or financial benefit*. What makes this definition preferable is that it removes the emphasis on whether the story is verifiable. For instance, a false headline: "BREAKING NEWS: TRUMP received an illegal donation of $400000 FROM PUTIN" is not demonstrably false. Had this claim been correct, it could have been verified via proof of a financial transaction but there is no evidence that can be used to outright mark the claim as being false. While, admittedly, most false news stories have been demonstrably false thus far (e.g. The Pope endorsing Donald Trump), we cannot rule out the possibility of fake news taking a direction which would make it a lot more difficult for the layman to make an informed value judgement.

1. Allcott and Gentzkow specifically try to rule out "close cousins of fake news" such as conspiracy theories. p. 214

2.2 The utility of fake news

The utility, or instrumental value, of fake news can be divided into three areas. There is the utility gained by the initial creator, the utility gained by someone who shares fake news and the utility of the reader. While there is certainly some crossover between these three areas, it is important that they remain distinctive.

For the purpose of this paper, fake news are split into two categories. Firstly, we have *fake news sites*, these are external websites that imitate the look and behavior of a news organization. Secondly, we have *fake news in social media*. These are accounts that pretend to be news organizations or real people who are merely sharing news stories. Both of these types of fake news have distinctive features that can be utilized for different purposes.

Fake news sites have more editorial control over the type of content that they publish and how said content is displayed. Having a separate site also makes its content more monetizable as the administrator can freely add advertisements that require executable Javascript or fully customizable CSS. What is beneficial to the site owner can raise issues to the reader, however. One such example is that the administrators can very well use the lack of restrictions for more nefarious purposes such as Clickjack attacks. One such example would be a now defunct fake news site [2] which compromised the user's display integrity (Huang et al. 2012, pp. 2-3) in order to trick users into pressing the site's Facebook page's 'Like' button. This was done by positioning the Facebook page's iframe on top of a welcome message's close button with the following CSS:

```
overflow: hidden; position: absolute; display: block;
z-index: 99999999; opacity: 0;
line-height: 1; right: 10px; top: -32px;
height: 23px; width: 24px; transform: scale(3);
transform-origin: 0px 0px 0px; border: 0px;
```

Figure 1. A clickjacking attack's CSS properties.

The second line displays the most integral part of this attack. A fully transparent object with

2. http://lastdeplorables.com - Screenshot and code retrieved on 30.5.2017.

a high z-index value will always be displayed on top and thus clicked by visitors trying to close the site's welcome message. While this behavior doesn't seem to be too prevalent, it along other methods of bolstering social network presence need to be taken into account when assessing the popularity of fake news sites or groups.

This leads us to fake news in social media. The primary motivator to use social media is to make the most of the wide exposure that it offers with relative ease. This means that social media can be used as an effective medium to *disinform*, as in spreading misinformation by intent. Furthermore, the owners of fake news sites can use social media to spread links to external sites for financial gains.

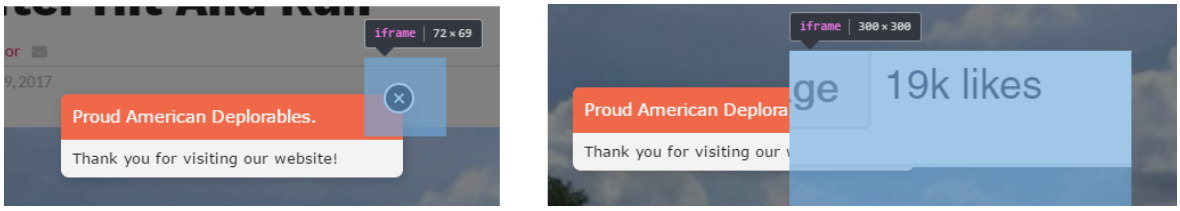

Figure 2. A clickjacking attack before and after the iframe's CSS values are altered.

The former represented examples of the benefits for creators and sharers of fake news. The reader's utility plays a different role. Allcott and Gentzkow characterize in their model the reader's utility as a trade-off where the end user needs to decide whether they consume news stories that are inaccurate but affirmatory for the psychological benefit or to consume unbiased and accurate news (Allcott and Gentzkow 2017, p. 218). While *confirmation bias* seems to be a natural tendency (Nickerson 1998, p. 211) which affects us all, we need to re-evaluate the scale and role that it plays when we take into account the fact that social networks are designed to connect us with other like minded people and show us content that already appeals to our interests and biases. In other words, do we interact with like minded people and read confirmatory news on social media because of our natural tendencies or do we act so because that is what social networks are designed to recommend?

2.3 Fake news and conspiracies

While fake news appears to be a fairly recent phenomenon, there is evidence to suggest that it shares similar traits to conspiracy theories. The power of fake news stories and conspiracy theories appears to stem from distrust. For fake news that distrust is partially targeted at the mainstream media, or at the very least correlates with an incredibly high perception of bias and misinformation within the media (Jones 2018). For conspiracy theories the distrust is usually targeted at the political status quo or its members. There is evidence that suggests that despite the effects of confirmation bias, if people trust the world around them then they're less apt to believing in conspiracies regarding political rivals (Miller, Saunders, and Farhart 2016, p. 838).

An important contrast to note here, however, is that fake news *are* created by people who certainly don't believe in them whereas conspiracy theories *might* be created by people who do believe in them. Regardless of the intentions of the authors, both can be used by a third party as an instrument to *disinform*.

Admittedly, more study is needed to fully determine how much overlap there is between the people and personalities that believe in fake news and conspiracies. When we take into account recent events combined with the very blurry line that separates the two, there is indicia that the overlap may very well be noteworthy.

2.4 The sociopolitical environment in the United States

Much of the discussion regarding fake news and politics has revolved around the 2016 U.S. Presidential election. Some high ranking Democrats, for political reasons, have painted a bleak picture, claiming that Donald Trump's campaign colluded with the Russian government which, among other practices, used social media to muddy the public discourse (Demirjian 2019). Republicans, for their own political reasons, deny such allegations and the President in turn has even gone as far as to claim that it was the Democrats that colluded with the Russian government (Trump 2018).

In early 2017 the Office of the Director of National Intelligence released a declassified ver-

sion of a more thorough report which detailed some of the alleged practices used by the Russian government in an attempt to influence the election. The report mentions that, among other strategies, the Russian government utilized paid social media "trolls" in an attempt to sway voters into supporting Donald Trump (*Assessing Russian activities and intentions in recent US elections* 2017, pp. ii, 2).[3] While critics may point out that the report was released merely days before Trump's inauguration, by the then Democratic administration and therefore might be biased, its conclusions are corroborated by the recent reports that the U.S. Cyber Command disrupted the internet access of the infamous Russian "troll farm" during the 2018 midterm elections (Nakashima 2019). This demonstrates that the U.S. considers Russia's disinformation campaigns as a threat regardless of which political party controls the executive branch of government.

When discussing social media "trolls" there seems to be some confusion regarding fake news. Some media outlets have implied that the existence of these "trolls" is to primarily spread fake news. For instance, right after the 2016 presidential election several articles strongly emphasized the false nature of the stories being spread by the "trolls". These articles indirectly exaggerated the scale that these false news stories played in a larger propaganda effort.[4] To clarify, the "trolls" partially utilized fake news within their disinformation campaigns. When mainstream news outlets posted partisan editorials with a corresponding message, the trolls gladly shared it on social media. The trolls treated fake news as an instrument, not as a goal. A more recent example of the activities of the Russian "troll farm" include the meddling in the vaccine debate, where according to researchers, the trolls amplified both sides of the debate (Broniatowski et al. 2018, p. 1378). What makes these disinformation campaigns particularly worrisome is that they do not necessarily exploit social networks on a technical level, they merely use the available tools in dubious schemes. As the details released by special counsel Robert Mueller elucidate, the ability to reach 126 million people is not an exploit, it is a feature. The same can be said about the ability to purchase approximately 3500 advertisements on Facebook totalling 100000 USD, as Russian intelligence did in 2016 (Mueller 2019, pp. 24-26). This dynamic makes it clear that there is not perfect

3. According to the report: "All three agencies agree with this judgment. CIA and FBI have high confidence in this judgment; NSA has moderate confidence.", p. ii

4. Some examples include Gregory 2016 and Dougherty 2016

fix for the problem. Any solution to combat any form of misinformation online will likely involve trading-off some of the functionality of the service in favour of limiting the impact of abuse. A recent example of such a fix includes WhatsApp's decision to limit the amount of times a message can be forwarded in order to hinder the reach of fake news (Hern and Safi 2019). Given that false news spread further than real news (Vosoughi, Roy, and Aral 2018, p. 1), limiting the ability to forward links is certainly not misguided but may not be foolproof.

2.5 Demographics

A recent paper which analyzed the sharing history of users found that approximately 8.5% of its respondents shared at least one fake news article (Guess, Nagler, and Tucker 2019, pp. 1-2). The paper also found that sharing fake news stories was more popular among Republicans compared to Democrats. Anecdotal evidence seems to support the hypothesis that there is more right-leaning fake news, either due to higher supply or because of a stronger driving force that spreads such stories. However, to imply that this phenomenon exclusively affects Republicans would be naive. A survey which asked respondents to categorize statements as fact or opinion found that "Republicans and Democrats are more likely to think news statements are factual when they appeal to their side – even if they are opinions" (Mitchell et al. 2018). This does not mean that fake news is equally an issue on both sides of the political spectrum, but rather to typify that just because this phenomenon has not affected the left as much in the past, it doesn't mean that it won't in the future.

The most notable finding by Guess, Nagler and Tucker is that age was the most consistent feature that seemed to affect susceptibility to fake news. According to the paper: "On average, users over 65 shared nearly seven times as many articles from fake news domains as the youngest age group" (Guess, Nagler, and Tucker 2019, p. 1). The authors discussed possible explanations, including media illiteracy (idem, p. 4).

3 Features of interest

In machine learning features refer to quantifiable characteristics that can be used in a machine learning model. In order to identify fake news outlets, it is imperative to collect data that portrays the characteristics of such outlets. The best way to collect general data about social media accounts and fake news sites, is to collect and analyze the metadata. For social networks some of the potential data would include the user's device info and country of origin. Other valuable metrics that aren't at the disposal of third-party developers but certainly collected by social networks would be metrics on user engagement. For example, "...real users spend comparatively more time messaging and looking at other users' contents (such as photos and videos)" (Ferrara et al. 2016, p. 102). For fake news sites metadata can be collected from the website itself and external sources. For instance, SSL certificate providers, keywords and certain serverside software metadata can be collected by authoring HTTP/HTTPS requests to the server and analyzing the response. Other information can be collected from third parties such as WHOIS records or ALEXA site traffic details.

Many papers have focused on analyzing the relationship between social media and fake news sites rather than fake news sites by themselves. Even when fake news sites are taken into account, it is primarily done to examine the underlying sentiment rather than the overall behavior of the outlet. There are a few potential reasons as to why this is the case. For instance, social networks are easier to harvest for consistent data, as long as researchers are provided with the proper API access. Additionally, the fact that fake news sites get roughly 41.8% of their traffic from social media, as opposed to 10.1% for real news outlets (Allcott and Gentzkow 2017, pp. 221-222), highlights fake news' dependency on social media. This means that consistent data that can be collected from fake news sites themselves, such as RSS feeds, becomes somewhat overlooked. A lot of fake news sites appear to be operated by people with limited technical skills and as a result administrators are more likely to resort to using easily installable CMS software such as Wordpress. Wordpress, like other content management systems, by default enables RSS feeds. In fact, the easiest way to access these feeds is by simply ammending "/rss" to the end of the top level domain. Additionally, in Wordpress, for instance, individual RSS feeds exist for tags, comments, categories, and

authors (*Documentation: WordPress Feeds*). What makes these feeds so invaluable is that they can be used to retrieve headlines and article content in an XML format. The results, especially the headlines, can subsequently be used for semantic NLP analysis.

When deciding what features we want to take into account while analyzing fake news sites we must first assess the importance of a specific feature and the means used to measure said feature. Understanding how important a specific feature is not necessarily problematic. Some machine learning algorithms, such as Random Forests, can be used for both classification and to make estimates about variable importance (Breiman 2001, p. 10). Defining a way to measure the features themselves is where complications arise.

Regarding what variables should be left out of the classification problem, there is often a temptation to leave out variables that are useless by themselves in order to avoid overfitting. While this logic sounds reasonable, having two variables that provide little value by themselves can actually be useful together (Guyon and Elisseeff 2003, pp. 1165-1166). Additionally, we can apply a similar logic during the data gathering process. For example, let's pretend that social networks would send a crawler to quickly analyze a "news" site whenever a link is posted. Some of the information that this crawler could hypothetically collect is "how many local pages does the site have?" and "how many advertisements does this site have?". These two data points might not be all that interesting by themselves. If anything, we can almost assume that a real news site would have more pages and, by extension, more ads. Collecting the number of pages that a site has allows us, however, to infer additional knowledge such as how many advertisements are there per page by simply dividing the two together. Variables such as "ads per page" also allow us to get a better understanding of the content that a page would display without the need to construct an interpreter that analyzes the site's client-side code. Moreover, such a variable could be used for more than just classification purposes. In chapter 2 we discussed how fake news sites are usually operated for financial gain, and thus one can assume that the operators of such sites would try to monetize the content as aggressively as possible. On the other hand, the recent crackdown on deceptive sites by advertisers could mean that people operating fake news sites struggle to monetize the pages as aggressively as the digital media is able to. Apart from classification, an "ads per page" variable could provide some sense as to how lucrative the clampdown against fake

news has been.

Other interesting inferences that we could make using a site's content, for instance, could include analyzing headlines for *clickbait*. Given that previous papers have shown that methods of identifying clickbait certainly exist (Chen, Conroy, and Rubin 2015 and Chakraborty et al. 2016), and the fact that fake news sites rely heavily on clicks from social networks it is easy to presuppose that fake news outlets would resort to using clickbait. The underlying problem here is that while fake news outlets might resort to using clickbait titles, it is not behaviour that is distinctive enough to accurately use it to distinguish real news outlets from fake ones. Such behavior is presumably caused by the fact that many established media outlets have resorted to more questionable methods to increase readership in the age of digital media. This highlights a point that warrants more attention in the discussion regarding fake news. The conventional wisdom up to this point has been that fake news is to some indistinguishable from real news because of its ability to emulate the practices of real news outlets. What is often ignored in this discussion is the fact that the legitimate news outlets have degraded their standards the the point that the practices can be easily mimicked. In other words, clickbait and other questionable practices such as native advertising don't appear to be considered as a token of mendacity but rather as a "necessary evil" in the age of digital media (Schauster, Ferrucci, and Neill 2016, p. 1419).

When discussing the characteristics and areas of interest, we have largely neglected in this chapter the fact that fake news can be presented in more forms than mere plain text. Despite the increased availability of third-party optical character recognition (OCR) tools and libraries exist, it is unclear whether the potential results warrant the additional computation time. This is why any partially automated system that is implemented needs to take into account metadata and cannot solely rely on content for sentiment analysis.

4 Methods of combating fake news

When discussing methods of combating fake news, we need to assess our underlying goal. Do we want to limit the amount of misinformation in social networks, or limit the capability to quickly spread misinformation? While idealistically speaking there is a temptation to answer both, every method has its strengths and weaknesses. Fake news, like rumours and conspiracies, are a constant, never ending arms race between social networks and their adversaries. This means that solutions to the problem are never foolproof, but rather aim to increase the cost of an adversary's continuous attacks while minimizing the cost to fend off such attacks. This means that, for social networks, some form of a partially automated system that involves machine learning is a necessity. Additionally, we also need to take into account where such automated systems can be implemented. Analyzing all content for being potentially fake news is not viable, yet only analyzing content that is labeled as 'news' is easily circumvented by changing a label. For this chapter we will assume that the best place for an automated analysis to take place is once it gets reported. This allows the automated processes to add a weight to a report, making more plausible reports processed earlier. The following subsections discuss actions that can be taken to contest the effects of fake news and potentially utilized in a weighted reporting system.

4.1 Partially automated systems

We will begin this chapter by discussing some of the more automated solutions that could be utilized in a broader effort to combat fake news starting with the role of the URL itself. As previously observed, fake news sites and social media accounts tend to utilize names that would mimic a real newspaper or organization (Allcott and Gentzkow 2017, p. 217). This pattern suggests that lexical analysis of a given URL could be used to potentially detect fake news sites, similarly to how URL analysis has been utilized to detect phishing links (See Blum et al. 2010). What makes this approach particularly viable is the fact that fake news sites try to utilize the terminology as a sign of credibility, therefore URLs with abundant political or news-related terminology are frequent. Figure 3 illustrates the structure of such

domains.[1]

Figure 3. An example of the syntactic structure for hyper-partisan sites that frequently share fake news. See Table 2 for the full list of domains and their words.

While there is plenty of room to discuss how certain words should be classified (e.g. should 'world' be classified as a news-related term or as a location?), the bottom line is that most terms can be categorized into three distinct categories: news, politics and location. Even if lexical analysis wouldn't yield high accuracy results in distinguishing a fake news site from a real one, it still has a lot of potential to be used in distinguishing politics or news-related sites from others. This could be particularly useful when Facebook pages that consistently share links to fake news sites categorize themselves as "Entertainment" or any label other than "News" to circumvent detection. For instance, a Facebook page that is categorized as a sports group but exclusively shares links to what seems like a political site could warrant further investigation.

We can also extend our partially automated analysis to the posted content itself using sentiment analysis. For instance, in a study where fake negative reviews were analyzed, results showed that deceptive reviews were identifiable due to the exaggerations that they made (Ott, Cardie, and Hancock 2013, p. 500). Conroy, Rubin and Chen underline in their paper various approaches to identify fake news. Their proposals can be split into two: linguistic approaches and network approaches (Conroy, Rubin, and Chen 2015, p. 1).

1. See Appendix 1 for the list of domains.

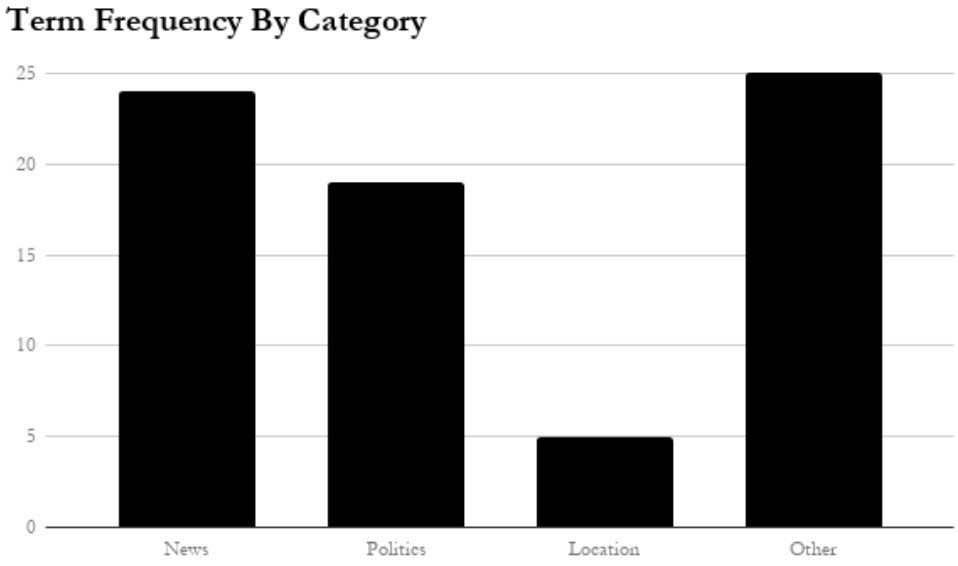

Figure 4. Term frequencies organized by each category. This chart illustrates that a majority of terms found in fake news domains can be categorized into three distinct themes: News, Politics or Location. See Appendix C, Table 3 for each word and its frequency found in the domain names

Instead of focusing solely on what fake news outlets do, we can also aim our attention at the emotional responses that fake news invokes. According to Vosoughi, Roy, and Aral 2018: ".. false news stories inspired fear, disgust and surprise in replies, true stories inspired anticipation, sadness, joy and trust". When we take into account the fact that sentiment analysis has been used with reasonable accuracy rates for plain text comments (Greaves et al. 2013), as have emojis (Felbo et al. 2017), it is clear that user responses provide an additional dimension that is at least worth looking into. Data collection for such analysis is not particularly difficult for social networks that have full API access to user comments. Such analysis gets more arduous when we try to extract comments from external fake news sites. As of this writing, a disproportionately large percentage of sites that spread fake news run on WordPress.[2] This detail is particularly relevant because Wordpress, by default, enables RSS feeds for comments (See chapter 3).[3] If for some reason fake news sites collectively decide to migrate to another CMS, collecting comments for analysis would become less consistent

2. Based on Table 1, at least 26/29 sites were running WordPress when they were in a functional state.
3. The URL to a WordPress site's comment RSS feed is: http://example.com/comments/feed/

or outright impractical. This means that sentiment analysis on comments can at best only be used with data collected by or from social networks.

4.2 Policies

Service policies can also be appropriated when combating fake news. Obviously a broad policy of banning fake news is self evident but not as enforceable as one can imagine, hence the need to address the issue. In addition, a smaller set of guidelines that limit the reach of fake news while minimizing the loss of the service's integrity should be put into place. Even if such guidelines seem obvious, we must always take into account the potential outliers.

For example, in late 2017 Google updated its policies for Google News, barring news sites that "misrepresent or conceal their country of origin or that direct content at users in another country under false premises" (*Google Content policies - Publisher Center Help*). The easiest way to obtain information about a website's owner and origin is to send a WHOIS query as documented in *RFC 3912: WHOIS Protocol Specification* 2004. A WHOIS query can easily be sent by using the 'whois' command on many modern UNIX based operating systems, making the process by extension easy to automate. This would make WHOIS queries a potentially attractive source of data for feature selection. But how good of a feature is this overall? Common sense would suggest that no legitimate news site would take active measures to hide its country of origin, but exceptions always exist. One such exception is the *Republican-American*, a family-owned local paper. The public WHOIS records for the *Republican American*,[4] as seen in Figure 5, show that the registrant's details are hidden thanks to WhoisGuard, a privacy protection service.

In our example, a paper which is the combination of two local papers (the Waterbury Republican and the Waterbury American) and which were awarded the Pulitzer Prize in 1940 for Public Service, would have been excluded from Google's news services, if such detection was fully automated. While such behaviour in legitimate news outlets is very much the exception and not the rule, this example exhibits that a policy's enforcement is equally as important as the policy itself. Furthermore, in this example it was assumed that the best (or

4. WHOIS lookup performed on 19.3.2019 - https://www.rep-am.com/

```
[.. Lines omitted ..]
    Registry Registrant ID:
Registrant Name: WhoisGuard Protected
Registrant Organization: WhoisGuard, Inc.
Registrant Street: P.O. Box 0823-03411
Registrant City: Panama
Registrant State/Province: Panama
Registrant Postal Code: 00000
Registrant Country: PA
[.. Lines omitted ..]
```

Figure 5. The Registrant's WHOIS records for The Republican-American

only) way to identify a site's country of origin is through a WHOIS query. This is just one of the many issues that needs to be taken into account when measuring features. Conroy, Rubin and Chen discuss about how "hybrid systems" could be used for classification within the context of deception detection but that similar principle can be used in a hybrid system for data collection as well. In other words, collecting information for the same feature from different sources could be used to compose a potentially more accurate dataset.

Another real world case which illustrates the challenges of policy enforcement is the recent report which suggests that Twitter is not proactively banning white supremacists not because it doesn't have the capacity to do so, but because it would end up banning several right wing politicians as well (Cox and Koebler 2019). This case highlights the role of viability. Even if the politicians deserved the suspension, violating the social norms by limiting the free speech of political figures or accidentally triggering a false positive could harm the network's public image. A similar reaction could occur if social networks started proactively banning fake news sites or groups incorrectly. To quote Conroy, Rubin, and Chen 2015, p. 4: "Tools should be designed to augment human judgement, not replace it".

This paper proposes that the aforementioned tools are utilized in a weighted, user reporting system in social networks. What makes a weighted reporting system beneficial is that it provides a good opportunity to analyze potentially questionable posts without having to

analyze all new content, which is costly. A weighted reporting system also allows room for company policies to be enforced. For instance, if a company policy states that all user reports should be read within a week, the weight that is added to the date of submission could be made to grow exponentially so that reports that are closing in on a week will almost always be processed before others. The main drawback of a report system is that it doesn't provide a pre-emptive solution to the problem. While the concept of a pre-emptive solution is particularly attractive, the main issues that obstruct the feasibility of such a solution are more political than technical.

5 Conclusion

This paper has discussed some of the elements that characterize fake news and its illicit dependency on social media. Fake news behaves in predictable and repetitive patterns, both within the scope of metadata and linguistics. However, this paper advocates for reactive solutions, not fully automated pre-emptive ones. This judgement is not necessarily made out of scepticism towards technical capacity, but is a conclusion which bears in mind the sociopolitical climate and norms. Social networks have already struggled in dealing with controversial and hateful views, putting the proposition of letting social networks become the kingmakers of what is news and what is not in hot water. Reactive approaches might not fix this issue but they can stop the public relations nightmare from worsening. What has fundamentally been ignored in the discussion regarding fake news is the responsibility of individuals. Regardless of how social networks decide to fine-tune their reporting process or detection systems, none of that will do much good if people continue to be susceptible to fake news.

Bibliography

Allcott, Hunt, and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election”. *Journal of Economic Perspectives* 31 (2): 211–236. doi:10.3386/w23089.

American Dialect Society. 2018. *"Fake news" is 2017 American Dialect Society word of the year.* https://www.americandialect.org/fake-news-is-2017-american-dialect-society-word-of-the-year.

Assessing Russian activities and intentions in recent US elections. 2017. Office of the Director of National Intelligence, National Intelligence Council. https://www.dni.gov/files/documents/ICA_2017_01.pdf.

Blum, Aaron, Brad Wardman, Thamar Solorio, and Gary Warner. 2010. “Lexical Feature Based Phishing URL Detection Using Online Learning”. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security,* 54–60. AISec ’10. Chicago, Illinois, USA: ACM. ISBN: 978-1-4503-0088-9. doi:10.1145/1866423.1866434. http://doi.acm.org/10.1145/1866423.1866434.

Breiman, Leo. 2001. “Random Forests”. *Machine learning* 45 (1): 5–32. ISSN: 1573-0565. doi:10.1023/A:1010933404324.

Broniatowski, David A., Amelia M. Jamison, Sihua Qi, Lulwah Alkulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 2018. “Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate”. *American Journal of Public Health* 108 (10): 1378–1384. doi:10.2105/ajph.2018.304567.

Chakraborty, Abhijnan, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. “Stop Clickbait: Detecting and preventing clickbaits in online news media”. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).* doi:10.1109/asonam.2016.7752207.

Chen, Yimin, Niall J. Conroy, and Victoria L. Rubin. 2015. “Misleading Online Content”. *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection - WMDD 15.* doi:10.1145/2823465.2823467.

Conroy, Niall J., Victoria L. Rubin, and Yimin Chen. 2015. "Automatic deception detection: Methods for finding fake news". *Proceedings of the Association for Information Science and Technology* 52 (1): 1–4. doi:10.1002/pra2.2015.145052010082.

Cox, Joseph, and Jason Koebler. 2019. *Twitter Won't Treat White Supremacy Like ISIS Because It'd Have to Ban Some GOP Politicians Too.* https://motherboard.vice.com/en_us/article/a3xgq5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too.

Demirjian, Karoun. 2019. *'Undoubtedly there is collusion': Trump antagonist Adam Schiff doubles down after Mueller finds no conspiracy.* https://www.washingtonpost.com/powerpost/undoubtedly-there-is-collusion-trump-antagonist-adam-schiff-doubles-down-after-mueller-finds-no-conspiracy/2019/03/26/e972d9e8-4fdd-11e9-a3f7-78b7525a8d5f_story.html.

Documentation: WordPress Feeds. https://wordpress.org/support/article/wordpress-feeds/.

Dougherty, Jill. 2016. *The reality behind Russia's fake news.* https://edition.cnn.com/2016/12/02/politics/russia-fake-news-reality/index.html.

Edson, C., Jr. Tandoc, Zheng Wei Lim, and Richard Ling. 2018. "Defining "Fake News"". *Digital Journalism* 6 (2): 137–153. doi:10.1080/21670811.2017.1360143.

Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* 1615–1625. Copenhagen, Denmark: Association for Computational Linguistics. doi:10.18653/v1/D17-1169.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. "The rise of social bots". *Communications of the ACM* 59 (7): 96–104. doi:10.1145/2818717.

Goldman, Cecilia Kang, and Adam. 2016. *In Washington Pizzeria Attack, Fake News Brought Real Guns.* https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html?module=inline.

Google Content policies - Publisher Center Help. https://support.google.com/news/publisher-center/answer/6204050?hl=en%5Cu0026ref_topic=9010378.

Greaves, Felix, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi, and Liam Donaldson. 2013. "Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online". *J Med Internet Res* 15, number 11 (): e239. ISSN: 14388871. doi:10.2196/jmir.2721. http://www.ncbi.nlm.nih.gov/pubmed/24184993.

Gregory, Paul Roderick. 2016. *Media Wakes Up To Russia's 'Fake News' Only After It Is Applied Against Hillary.* https://www.forbes.com/sites/paulroderickgregory/2016/11/29/media-wakes-up-to-russias-fake-news-only-after-it-is-applied-against-hillary/.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". *Science Advances* 5 (1). doi:10.1126/sciadv.aau4586.

Guyon, Isabelle, and André Elisseeff. 2003. "An introduction to variable and feature selection". *Journal of machine learning research* 3 (Mar): 1157–1182.

Hern, Alex, and Michael Safi. 2019. *WhatsApp puts limit on message forwarding to fight fake news.* https://www.theguardian.com/technology/2019/jan/21/whatsapp-limits-message-forwarding-fight-fake-news.

Huang, Lin-Shung, Alex Moshchuk, Helen J. Wang, Stuart Schecter, and Collin Jackson. 2012. "Clickjacking: Attacks and Defenses". In *Presented as part of the 21st USENIX Security Symposium (USENIX Security 12),* 413–428. Bellevue, WA: USENIX. ISBN: 978-931971-95-9. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/huang.

RFC 3912: WHOIS Protocol Specification. 2004. https://tools.ietf.org/html/rfc3912.

Jones, Jeffrey M. 2018. *Americans: Much Misinformation, Bias, Inaccuracy in News.* https://news.gallup.com/opinion/gallup/235796/americans-misinformation-bias-inaccuracy-news.aspx.

Lister, Tim, Sebastian Shukla, and Nima Elbagir. 2019. *A Russian company's secret plan to quell protests in Sudan.* https://edition.cnn.com/2019/04/25/africa/russia-sudan-minvest-plan-to-quell-protests-intl/index.html.

Miller, Joanne M., Kyle L. Saunders, and Christina E. Farhart. 2016. "Conspiracy Endorsement as Motivated Reasoning: The Moderating Roles of Political Knowledge and Trust". *American Journal of Political Science* 60 (4): 824–844. doi:10.1111/ajps.12234.

Mitchell, Amy, Jeffrey Gottfried, Michael Barthel, and Nami Sumida. 2018. *Can Americans Tell Factual From Opinion Statements in the News?* http://www.journalism.org/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/.

Mozur, Paul. 2018. *A Genocide Incited on Facebook, With Posts From Myanmar's Military.* https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Mueller, Robert S. 2019. *Report on the Investigation Into Russian Interference in the 2016 Presidential Election.* U.S. Department of Justice. https://www.justice.gov/storage/report.pdf.

Nakashima, Ellen. 2019. *U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms.* https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html?utm_term=.2fc67d765292.

Nickerson, Raymond S. 1998. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises”. *Review of General Psychology* 2 (2): 175–220. doi:10.1037/1089-2680.2.2.175.

Ott, Myle, Claire Cardie, and Jeffrey T Hancock. 2013. “Negative deceptive opinion spam”. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies,* 497–501.

Schauster, Erin E., Patrick Ferrucci, and Marlene S. Neill. 2016. “Native Advertising Is the New Journalism: How Deception Affects Social Responsibility”. *American Behavioral Scientist* 60 (12): 1408–1424. doi:10.1177/0002764216660135.

Trump, Donald J. 2018. https://twitter.com/realDonaldTrump/status/1026471244949061632.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. “The spread of true and false news online”. *Science* 359 (6380): 1146–1151. ISSN: 0036-8075. doi:10.1126/science.aap9559.

A List of sites that spread fake news

Table 1 contains domains retrieved from the papers by Allcott and Gentzkow 2017 and Guess, Nagler, and Tucker 2019. The table also contains the current status of each domain and whether the site was running on the WordPress platform while it was functional. This was done by using the Wayback machine (https://archive.org/web/), an archiving tool that stores the page's client-side source code in its archives. Only 18 out of the 29 sites were fully functional as of 22.4.2019. Additionally, What is left out of this table is that 4 out of the 18 operating sites have a separate disclaimer/about page which states that the site produced "satirical" content. Out of the 11 defunct sites, only one had such a page.

Key: PD = Parked Domain, RP = Repurposed Domain, ED = Expired Domain, F = Functional, PF = Partially Functional, NF = Not Functional (and/or other).

Table 1. Table of domains that spread fake news.

Site URL	Source	Site Status	Wordpress
http://abcnews.com.co/	Allcott & Gentzkow & Guess et al.	PD	True (Dec 01 2016)
http://usuncut.com/	Allcott & Gentzkow	RP	True (Oct 14 2016)
https://www.thedailysheeple.com/	Allcott & Gentzkow	F	True (April 22 2019)
http://occupydemocrats.com/	Allcott & Gentzkow	F	True (April 22 2019)
http://nationalreport.net/	Allcott & Gentzkow	F	True (April 22 2019)
https://governmentslaves.news/	Allcott & Gentzkow	F	True (April 22 2019)
beforeitsnews.com	Allcott & Gentzkow	F	False
veteranstoday.com	Allcott & Gentzkow	F	True (April 22 2019)
bluenationreview.com	Allcott & Gentzkow	F	True (April 22 2019)
usanewsflash.com	Guess et al.	RP	True (Nov 21 2016)
denverguardian.com	Guess et al.	RP	True (Jul 14 2016)
https://rickwells.us/	Guess et al.	F	True (April 22 2019)
truepundit.com	Guess et al.	F	True (April 22 2019)
http://redstatewatcher.com/	Guess et al.	F	False
https://worldpoliticus.com/	Guess et al.	PF	True (Mar 22 2019)
https://www.subjectpolitics.com/	Guess et al.	F	True (April 22 2019)
conservativestate.com	Guess et al.	PD	True (Dec 11 2016)
conservativedailypost.com	Guess et al.	F	True (April 22 2019)
libertywritersnews.com	Guess et al.	NF	N/A
worldnewsdailyreport.com	Guess et al.	F	True (April 22 2019)
endingthefed.com	Guess et al.	RP	True (Dec 04 2016)
donaldtrumpnews.co	Guess et al.	F	True (April 22 2019)
yesimright.com	Guess et al.	F	True (April 22 2019)
burrardstreetjournal.com	Guess et al.	F	True (April 22 2019)
bizstandardnews.com	Guess et al.	F	True (April 22 2019)
everynewshere.com	Guess et al.	PD	True (Nov 18 2016)
departed.co	Guess et al.	PD	True (Nov 15 2016)
americanmilitarynews.com	Guess et al.	F	True (April 22 2019)
tmzhiphop.com	Guess et al.	RP	True (Oct 11 2016)

B Site domains and term frequency

Table 2 and Table 3 illustrate that domains can be dissected into smaller words that can be categorized. For instance, the domain name libertywritersnews.com can be dissected into "liberty", "writers" and "news". The first term could be classified as political terminology and the two terms after that could be classified as terminology related to news. The minutiae is not particularly important here, but this example illustrates that these domains feature words like "news" (9 occurrences) or "daily" (3 occurrences) frequently when we take into account the sample size.

Table 2. Every retrieved domain name split into words.

URL	Words
http://abcnews.com.co/	abc,news
http://usuncut.com/	us,uncut
https://www.thedailysheeple.com/	the,daily,sheeple
http://occupydemocrats.com/	occupy,democrats
http://nationalreport.net/	national,report
https://governmentslaves.news/	government,slaves
beforeitsnews.com	before,its,news
veteranstoday.com	veterans,today
bluenationreview.com	blue,nation,review
usanewsflash.com	usa,flash,news
denverguardian.com	denver,guardian
https://rickwells.us/	rick,wells
truepundit.com	true,pundit
http://redstatewatcher.com/	red,state,watcher
https://worldpoliticus.com/	world,politicus
https://www.subjectpolitics.com/	subject,politics
conservativestate.com	conservative,state
conservativedailypost.com	conservative,daily,post
libertywritersnews.com	liberty,writers,news
worldnewsdailyreport.com	world,news,daily,report
endingthefed.com	ending,the,fed
donaldtrumpnews.co	donald,trump,news
yesimright.com	yes,im,right
burrardstreetjournal.com	burrard,street,journal
bizstandardnews.com	biz,standard,news
everynewshere.com	every,news,here
departed.co	departed
americanmilitarynews.com	american,military,news
tmzhiphop.com	tmz,hiphop

C Term Frequency and Classification

Frequency	Word	Category
9	news	News
3	daily	News
2	world	News
2	the	Other
2	state	Politics
2	report	News
2	conservative	Politics
1	yes	Other
1	writers	News
1	wells	Other
1	watcher	Other
1	veterans	Politics
1	usa	Location
1	us	Location
1	uncut	Other
1	trump	Politics
1	true	Other
1	today	News
1	tmz	Other
1	subject	Other
1	street	Other
1	standard	News
1	slaves	Other
1	sheeple	Other
1	right	Other
1	rick	Other
1	review	News

Table 3 continued from previous page

Frequency	Word	Category
1	red	Politics
1	pundit	News
1	post	Politics
1	politicus	Politics
1	politics	Politics
1	occupy	Politics
1	national	News
1	nation	Politics
1	military	Politics
1	liberty	Politics
1	journal	News
1	its	Other
1	im	Other
1	hiphop	Other
1	here	Other
1	guardian	News
1	government	Politics
1	flash	Other
1	fed	Politics
1	every	Other
1	ending	Other
1	donald	Politics
1	departed	Other
1	denver	Location
1	democrats	Politics
1	burrard	Location
1	blue	Politics
1	biz	Other

Table 3 continued from previous page

Frequency	Word	Category
1	before	Other
1	american	Location
1	abc	Other

Table 3: Each word occurance in the domain names organized by frequency and classified into one of four categories: News, Politics, Location and Other.