Current State of the Art to Detect Fake News in Social Media: Global Trendings and Next Challenges

Alvaro Figueira[1], Nuno Guimaraes[1] and Luis Torgo[2]
[1]*CRACS / INESCTEC and University of Porto, Rua do Campo Alegre, Porto, Portugal*
[2]*Faculty of Computer Science, Dalhousie University, Halifax, Canada*

Keywords: Fake News, Detection Systems, Survey, Next Challenges.

Abstract: Nowadays, false news can be created and disseminated easily through the many social media platforms, resulting in a widespread real-world impact. Modeling and characterizing how false information proliferates on social platforms and why it succeeds in deceiving readers are critical to develop efficient algorithms and tools for their early detection. A recent surge of researching in this area has aimed to address the key issues using methods based on machine learning, deep learning, feature engineering, graph mining, image and video analysis, together with newly created data sets and web services to identify deceiving content. Majority of the research has been targeting fake reviews, biased messages, and against-facts information (false news and hoaxes). In this work, we present a survey on the state of the art concerning types of fake news and the solutions that are being proposed. We focus our survey on content analysis, network propagation, fact-checking and fake news analysis and emerging detection systems. We also discuss the rationale behind successfully deceiving readers. Finally, we highlight important challenges that these solutions bring.

1 INTRODUCTION

The large increase of social media users in the past few years has led to an overwhelming quantity of information available in daily (or even hourly) basis. In addition, the easy accessibility to these platforms whether it's by a computer, tablet or mobile, allows the consumption of information at a distance of a click. Therefore, traditional and independent news media urge to adopt social media to reach a broader audience and gain new clients/consumers.

The ease of creating and disseminating content in social networks like Twitter and Facebook has contributed to the emergence of malicious users. In particular, users that infect the network with the propagation of misinformation or rumours. This actions combined with the fact that 67% of adults consume some type of news in social media (20% on a frequent basis) (Gottfried and Shearer, 2017) have already caused real-world consequences (Snopes, 2016).

However, unreliable content or, how it is now referred – "fake news" –, is not a recent problem. Although the term gained popularity in the 2016 US presidential election, throughout the years newspapers and televisions have shared false content resulting in severe consequences for the real world. For example, in 1924 a forged document known as "*The Zinoviev Letter*" was published on a well known British newspaper four days before the general elections. The goal was to destabilize the elections in favour of the conservative party with a directive from Moscow to British communists referring an Anglo-Soviet treaty and inspiring "agitation-propaganda" in the armed forces (Norton-Taylor, 1999). Another example happened after the "Hillsborough accident", where 96 people died crushed due to overcrowding and lack of security. Reports from an illustrious newspaper claimed that, as people were dying, some fellow drunk supporters stole from them and beat police officers that were trying to help. Later, such claims were proven false (Conn, 2016).

The verified impact of fake news in society throughout the years and the influence that social networks currently have today forced high reputation companies, such as Google and Facebook, to start working on a method to mitigate the problem (Hern, 2017; Hern, 2018). The scientific community has also been increasing the activity on the topic. In fact, if we search in Google Scholar [1] for "fake news", we will find a significantly high number of results that have

[1]https://scholar.google.com

Figueira, A., Guimaraes, N. and Torgo, L.
Current State of the Art to Detect Fake News in Social Media: Global Trendings and Next Challenges.
DOI: 10.5220/0007188503320339
In *Proceedings of the 14th International Conference on Web Information Systems and Technologies (WEBIST 2018)*, pages 332-339
ISBN: 978-989-758-324-7

an increase of 7K publications, when compared with the number obtained in the previous year.

Nevertheless, the problem of fake news is still a reality since the solution is anything but trivial. Moreover, research on the detection of such content, in the context of social networks, is still recent. Therefore, in this work we attempt to summarize the different and most promising branches of the problem as well as the preliminary proposed solutions in the current literature.

In addition, we present a perspective on the next steps in the research with a focus on the need to evaluate the current systems/proposals in a real-world environment.

2 LITERATURE REVIEW

There are several approaches to tackle the problem of unreliable content on social media. Some authors opt by analyzing the patterns of propagation, others by creating supervised systems to classify unreliable content, and others by focusing on the characteristics of the accounts that share this type of content. In addition, some works also focus on developing techniques for fact-checking claims or focus on specific case studies.

2.1 Account Analysis

Regarding the analysis of social media accounts, the current state of the art has being focusing on trying to identify bot or spammer accounts.

Castillo et al. (Castillo et al., 2011) target the detection of credibility in Twitter events. The authors created a dataset of tweets regarding specific trending topics. Then, using a crowd sourcing approach, they annotated the dataset regarding the credibility of each tweet. Finally, they used four different sets of features (Message, User, Topic, and Propagation) on a Decision Tree Model that achieved an accuracy of 86% from a balanced dataset. A more recent work (Erahin et al., 2017) used an Entropy Minimization-Discretization technique that combines numerical features with assessing fake accounts on Twitter.

Benevuto et al.(Benevenuto et al., 2010) developed a model to detect spammers by building a manual annotated dataset of 1K rcors of spam and non-spam accounts. Then, they extracted attributes regarding content and user behaviour. The system was capable of detecting correctly 70% of the spam accounts and 96% of non-spam.

A similar problem is the detection of bot accounts. Chu et al. (Chu et al., 2012) distinguished accounts into three different groups: humans, bots and cyborgs. Using a human-labelled dataset of 6K users, they built a system with four distinct areas of analysis (entropy measures, spam detection, account properties, and decision making). The performance of the system was evaluated using accuracy, which reached 96% in the "Human" class. Another similar work (Dickerson et al., 2014), introduced a methodology to differentiate accounts into two classes: humans and bots.

Gilani et al. (Gilani et al., 2017) presented a solution to a similar goal: to distinguish automated accounts from human ones. However, they introduced the notion that "automated" is not necessarily bad. Using a dataset containing a large quantity of user accounts, the authors divided and categorized each entry into 4 different groups regarding the popularity (followers) of the account. The evaluation was conducted using the F1-measure. The results obtained fall between 77% and 91% .

2.2 Content Analysis

A work by Antoniadis et al. (Antoniadis et al., 2015) tried to identify misinformation on Twitter. The authors annotated a large dataset of tweets and developed a model using the features from the Twitter text, the users, and the social feedback it got (number of retweets, number of favourites, number of replies). Finally, they assessed the capability of the model in detecting misinformation in real time, i.e. in *a priori* way (when the social feedback is not yet available). Evaluations on real-time only decay in 3% when compared with the model that uses all available features. An approach also using social feedback was presented by Tacchini et al. (Tacchini et al., 2017). The authors claim that by analyzing the users who liked a small set of posts containing false and true information, they can obtain a model with an accuracy near 80%.

Perez-Rosas (Pérez-Rosas et al., 2017) created a crowd-sourced fake news dataset in addition to fake news available online. The dataset was built based on real news. In other words, crowd-source workers were provided with a real news story and were asked to write a similar one, but false. Furthermore, they were asked to simulate journalistic writing. The best model obtained a 78% accuracy in the crowd-sourced dataset and only less 5% in a dataset obtained by fake news on the web.

Another example is the work of Potthast (Potthast et al., 2017) which analyses the writing style of hyperpartisan (extremely biased) news. The authors adopt a meta-learning approach ("unmasking") from the authorship verification problem. The results obtained

show models capable of reaching 78% in F1-measure in the task of classifying hyper-partisan and mainstream news, and 81% in distinguishing satire from the hyper-partisan and mainstream news. However, we must note that using only style-based features does not seem to be enough to distinguish fake news since the authors best result was 46%.

2.3 Network Propagation

In Shao (Shao et al., 2016) the authors expose a method describing the extraction of posts that contained links to fake news and fact-checking web pages. Then, they analyzed the popularity and patterns of the activity of the users that published these type of posts. The authors concluded that the users that propagate fake news are much more active on social media than the users that refute the claims (by spreading fact-checking links). The authors' findings also suggest that there is a small set of accounts that generate large quantities of fake news in posts.

Another work by Tambuscio et al. (Tambuscio et al., 2015) describes the relations between fake news believers and fact-checkers. The study modifies and resorts to a model commonly used in the analysis of disease spreading, where the misinformation is analyzed as a virus. Nodes on the network can be believers of fake news, fact-checkers or susceptible (neutral) users. Susceptible nodes can be infected by fake news believers although they can "recover" when confronted with fact-checking nodes. By testing their approach in 3 different networks, the authors concluded that fact-checking can actually cancel a hoax even for users that believe, with a high probability, in the message.

A similar approach is proposed in (Litou et al., 2016) where a Dynamic Linear Model is developed to timely limit the propagation of misinformation. The model differs from other works since it relies on the ability for the user's susceptibility to change over time and how it affects its dissemination of information. The model categorizes users in 3 groups: infected, protected and inactive, and validates the effectiveness of the approach on a real-world dataset.

2.4 Fact-Checking

Another way to tackle the problem of false information is through fact-checking. Due to the enormous quantity of information spread through social networks, the necessity to automatize this task has become crucial. Automated fact-checking aims to verify claims automatically through consultations and extraction of data from different sources. Then, based on the strength and stance of reputable sources regarding the claim, a classification is assigned (Cohen et al., 2011). This methodology, despite being in development is very promising.

2.4.1 Stance Detection

In earlier research, stance detection has been defined as the task of a given fragment of text agrees, disagrees or is unrelated to a specific target topic. However, in the context of fake news detection, stance detection has been adopted as a primary step to detect the veracity of a news piece. Simply putting it, to determine the veracity of a news article, one can look to what well-reputed news organizations are writing about that topic. Therefore, stance detection can be applied to understand if a news written from an unknown reputation source is agreeing or disagreeing with the majority of the media outlets. A conceptually similar task to stance detection is textual entailment (Pfohl et al., 2016; Sholar et al., 2017)

The fake news challenge [2] promotes the identification of fake news through the used of stance detection. More specifically, given a headline and a body of text (not necessarily from different articles), the task consists in identifying if the body of text agrees, disagrees, discusses or its unrelated with the headline. Several approaches were presented using the dataset provided. The authors in (Mrowca and Wang, 2017) present several approaches using a conditioned bidirectional LSTM (Long Short Term Memory) and the baseline model (GradientBoosted Classifier provided by the authors of the challenge) with an additional variation of features. As for the features, Bag of Words and GloVe vectors were used. In addition, global features like binary co-occurrence in words from the headline and the text, polarity words and word grams were used. The best result achieved was using bidirectional LSTM with the inclusion of the global features mentioned. The improvement regarding the baseline was 9.7%. Other works with similar approaches were proposed (Pfohl et al., 2016; Sholar et al., 2017) however, results do not vary significantly.

Stance detection is an important step towards the problem of fake news detection. The fake news challenge seems to be a good starting point to test possible approaches to the problem. Furthermore, the addition of source reputation regarding topics (p.e. politics) can provide useful insight to detect the veracity of a news.

[2] http://www.fakenewschallenge.org/

2.4.2 Fact-checking as a Network Problem

The authors in (Ciampaglia et al., 2015) tackle fact-checking as a network problem. By using the Wikipedia infoboxes to extract facts in a structured way, the authors proposed an automatic fact-checking system which relies on the path length and specificity of the terms of the claim in the Wikipedia Knowledge Graph. The evaluation is conducted in statements (both true and false) from the entertainment history and geography domains (for example "x was marry to y", "d directed f" and "c is the capital of r") and an independent corpus with novel statements annotated by human raters. The results of the first evaluation showed that true statements have higher truth values than false. In the second evaluation, the values from human annotators and the ones predicted by the system are correlated.

Another work by the same authors (Shiralkar et al., 2017) use an unsupervised approach to the problem. The Knowledge Stream methodology adapts the Knowledge Network to a flow network since multiple paths may provide more context than a single path and reusing edges and limiting the paths where they can participate may limit the path search space. This technique, when evaluated in multiple datasets, achieves results similar to the state of the art. However, in various cases, it provides additional evidence to support the fact-checking of claims.

2.5 Fake News Analysis

Another major area of study is the analysis of large quantities of fake news spread through social networks. Vosoughi et al. (Vosoughi et al., 2018) presented a study of the differences between propagation of true and false news. The work focused on the retweet propagation of false, true, and mixed news stories for a period of 11 years. The findings were several. First, false news stories peaks were at the end of 2013, 2015 and 2016. Then, through the analysis of retweets of false news stories, the authors concluded that falsehood reaches a significantly larger audience than the truthful. In addition, tweets containing false news stories are spread by users with fewer followers and friends, and that are less active than users who spread true news stories. Another work (Vargo et al., 2017) studied the agenda-setting relationships between online news media, fake news, and fact checkers. In other words, if each type of content is influenced by the agenda of others. The authors found out that certain issues were transferred to news media due to fake news (more frequently in fake stories about international relations). Furthermore, fake news also predicted the issue agenda of partisan media (more in the liberal side than the conservative). Other relevant findings are the reactive approach of fake news media to traditional and emerging media and the autonomy of fact-checking websites regarding online media agendas.

2.6 Case Studies

Some works focus on analyzing the dissemination of false information regarding a particular event. One of those related to the Boston Marathon in 2013, where two homemade bombs were detonated near the finish of the race (CNN, 2013). For example, in (Gupta et al., 2013) the authors performed an analysis on 7.9 million tweets regarding the bombing. The main conclusions were that 20% of the tweets were true facts whether 29% were false information (the remaining were opinions or comments), it was possible to predict the virality of fake content based on the attributes of the users that disseminate it, and accounts created with the sole purpose of disseminating fake content often opt by names similar with official accounts or names that explore the sympathy of people (by using words like "pray" or "victim"). Another work has the analysis focused on the the main rumours spread on Twitter after the bombings occurred (Starbird et al., 2014).

A different event tackled was the US Presidential Election in 2016. For example, the authors in (Allcot and Gentzkow, 2017) combined online surveys with information extracted from fact-checking websites to perceive the impact of fake news in social media and how it influenced the elections. Findings suggest that articles containing fake news pro-Trump were shared three times more than articles pro-Clinton and the average American adult has seen at least one fake news stories on the month around the election. Another work (Bovet and Makse, 2018) studied the influence of fake news and well know news outlets on Twitter during the election. The authors collected approximately 171 million tweets in the 5 months prior to the elections and showed that bots diffusing unreliable news are more active than the ones spreading other types of news (similar to what was found in (Allcot and Gentzkow, 2017)). In addition, the network diffusing fake and extreme bias news is denser than the network diffusing center and left-leaning news. Other works regarding this event are presented in (Kollanyi et al., 2016; Shao et al., 2017).

Other works address similar events such as Hurricane Sandy (Antoniadis et al., 2015), the Fukushima Disaster (Thomson et al., 2012) and the Mumbai Blasts in 2011 (Gupta, 2011).

2.7 Fake News Detection Systems

The majority of the implementations to detect fake news comes in the form of a browser add-on. For example, the bs-detector (The Self Agency, 2016) flags content in social media in different categories such as clickbait, bias, conspiracy theory and junk science. To make this evaluation, the add-on uses OpenSources [3] which is a curated list of dubious websites. A more advanced example is the Fake News Detector (Chaves, 2018). This add-on uses machine learning techniques in a ground truth dataset combined with the "wisdom of the crowd" to be constantly learning and improving the detection of fake news. An interesting system that also took the shape of an add-on was the one developed by four colleges students during a hackathon at Princeton University (Anant Goel, 2017). Their methodology combined two different approaches: the first makes a real-time analysis of the content in user's feed. The other notifies the user when they are posting or sharing doubtful content. The system is capable of analyzing keywords, recognizes images and verified sources to accurately detect fake content online. With confidence we can say that new systems are being created with a frequency of more than a dozen a year. Most of then uses the add-on approach, but many are not yet system to be usable by the normal people as they are yet proof of concept prototypes.

3 DISCUSSION

Fake news is nothing new. It has been shown that even before the term has become trending, the concept has been active in different occasions. We might say that fake news is only a threatening problem these days because of the mass distribution an dissemination capabilities that current digital social networks have. Due to these problems, and particularly to the problems that consequently emerge from it for the society, the scientific community started tackling the problem, taking an approach of addressing first its different subproblems.

The detection of bot/spam accounts, the machine learning and deep learning approaches to the detection of fake content or even the network analysis to understand how this type of content can be identified is diffuse and generally yet quite difficult to be understood by a general public.

Regarding the bot/spam detection, we do believe that even they play an important role on the diffusion of fake news and misinformation, they do not represent all the accounts that spread this type of content. In some cases, the spreaders of misinformation are cyborg accounts (humanly operated but that also include some automatic procedures), as the authors in (Chu et al., 2012) refer. Another case which has been less discussed in the current literature are the human operated accounts that spread misinformation. Common examples are users who are highly influenced by extreme biased news and that spread that information intentionally to their followers. One could argue that this is the effect of the propagation of unreliable content through the network. However, the probability of having misinformation in our feed through the propagation of close nodes in our network is higher than from the original node that spread the content. Therefore, the evaluation of this accounts can be of major importance when implementing a solid system for an everyday use.

Another important aspect in adapting the current theory, to a system that informs users about which news are credible and which are misinformation, is the effect that such system may have on more skeptic or biased users. In fact, the "hostile media phenomenon" can affect the use of these systems if these are not equipped with the capability of justifying the credibility of the content. Hostile media phenomenon states that users who already have an opinion on a given subject can interpret the same content (regarding that subject) in different ways. The concept was first studied in (Abdulla et al., 2002) with news regarding the Beirut massacre. Consequently, just like news media, such systems can be criticized by classifying a piece of news as fake by users who are in favor of the content for analysis. This leads us to the problem of the current detection approaches in the literature. For example, deep learning approaches and some machine learning algorithms are black-box systems that, given an input (in this case, a social media post), they output a score or a label (in this case a credibility score or a fake/true label). Therefore, explain to a common user why the algorithm predicted such label/score can be a hard task. Furthermore, without some type of justification, such systems can be discredited. To tackle this problem in a real-world environment, the major focus after developing an accurate system must be to be capable to explain how it got to the result.

3.1 Future Guidelines to Tackle Unreliable Content

We do believe that the analysis and effect of detection systems on the perception and beliefs of users towards

[3] http://www.opensources.co/

fake news and all sorts of misinformation should be the next important step to be studied by the scientific community. Accordingly, we suggest some guidelines to approach the problem.

Recently, we have observed an increasing distrust on the press by the common citizen. Several reasons can be pointed such as the president of the United States calling fake news to mainstream news media such as CNN or NBC [4] or even the mainstream media itself retracting a large number of stories and fail to highlight the importance of certain entities or events such as Donald Trump, Jeremy Corbin or Brexit.

Therefore, a misinformation detection system that only scores the credibility of a social media post, justifying it by the absence/presence of similar information on traditional news media outlets may not be enough to convince the majority of users that consume this type of information, and to change their beliefs and acknowledge the veracity or falsehood of the content. In addition, if it is used, the selection of mainstream news media sources to perform such comparison, then is must be balanced (at least with some normalization of its intensity/frequency) regarding some possible bias. Moreover, it is necessary to add features to the system to allow more information to be provided alongside the credibility score. Such examples could include the reputation of the original source (that created the post), and analyze and present the social feedback of the users that voice it, again in a weighted manner.

The credibility of the original source can be analyzed and explained through the history of previous published posts that include misinformation and were debunked by fact-checking entities such as Snopes [5] or PolitiFact [6]. In addition, information should be provided by the system on the possibility of an original source be a bot or cyborg. This can be done by the analysis of the posting frequency and near accounts (followers and friends). Such analysis can have a more impact on changing user beliefs than a simple score or label.

Regarding the social feedback, one can look at the propagation of the content through shares/retweets, comments/replies and favorites/likes. However, an important factor must come into play which is the "echo chamber" effect. This refers to the problem of users with the same interests are aggregated together in social circles, and the opposing ideas are rejected and disapproved by the majority. In Facebook, for example, when looking at comments/replies to a post inside one of this echo chambers, the majority of comments are in agreement with the post. This factor combined with the number of likes, shares, etc, may lead the user to a false impression that the information is true. Moreover, Facebook by default increases the ranking of the comments based on the number of replies and likes that the comments got as well as if the comments come from the user's friends. This scenario is extremely propitious to influence the opinion of a user, therefore contributing to the formation and expansion of these "echo chambers".

Therefore, in a real-world fake news detection system, information put on social feedback should be ranked and presented to the user based on the "bias" of the social media accounts that are engaging with the post being analyzed. In addition, social feedback from users who are more neutral or equally active on both sides of the political spectrum should be prioritized since they could present a more careful and unbiased view of the subject or even act as fact-checkers of more suspicious claims. It is our hypothesis that the way social feedback is presented may influence users' beliefs regarding the credibility score (even not affecting the score directly).

For the sake of clarity, let us consider the domain of politics where there are two main groups: conservatives (c) and liberals (l). Let us also consider a social media post which is false and favours the liberal side. A user that likes, spreads, replies, and is connected to liberal content accounts should have a lower rank on posts that favour his political views. In the same way, an opposing user which displays the same behaviour regarding conservative content must also have a smaller rank with respect to the content that opposites their political views. However, users who are capable of engaging with content from both political views in an equal and neutral position should have a higher rank. A simple function f that can be used to score the ranking of an user u may be given by:

$$f_u = \frac{1}{1 - |p_{u,l} - p_{u,c}|}$$

where p refers to an intermediate score of user u engagement in liberal (l) or conservative content (c). This probability can be computed using the type of posts that users propagate, like or reply/comment. In addition (and expanding what is done in (Tacchini et al., 2017)), the compliance between the post and the user comments/replies can also determine the tendency of a user. This can be analyzed recurring, for example, to sentiment analysis tools.

Analyzing the source account of the content and ranking the presentation of social feedback (prioriti-

[4] https://twitter.com/realDonaldTrump/status/1006891643985854464

[5] https://www.snopes.com/

[6] http://www.politifact.com/

zing neutral but active users) might increase skeptic to trust in the fake news detection system.

4 CONCLUSION

Fake news is nowadays of major concern. With more and more users consuming news from their social networks, such as Facebook and Twitter, and with an ever-increasing frequency of content available, the ability to question the content instead of instinctively sharing or liking it is becoming rare. The search for instant gratification by posting/sharing content that will allow social feedback from peers has reached a status where the veracity of information is left to the background.

Industry and scientific communities are trying to fix the problem, either by taking measures directly into the platforms that are spreading this type of content (Facebook, Twitter, Google, for example), developing analysis and systems capable of detecting fake content using machine and deep learning approaches, or even by developing software to leverage social network users in distinguishing what is fake from what is real.

However, observing the various approaches taken so far, mainly by the scientific community but also some rumours about actions taken by Facebook and Google, we might say that mitigating or removing fake news comes with a cost (Figueira and Oliveira, 2017): there is the danger to having someone establishing the limits of reality, if the not the reality itself.

The trend to design and develop systems that are based on open source resources, frameworks or APIs which facilitate entity recognition, sentiment analysis, emotion recognition, bias recognition, relevance identification (to name just a few), and which may be freely available, or available at a small price, gives an escalating power to those service-providers. That power consists on their internal independent control to choose their machine learning algorithms, their pretrained data and, ultimately, in a control over the intelligence that is built on the service provided by their systems.

Therefore, the saying "the key to one problem usually leads to another problem" is again true. However, we have not many choices at the moment. It is a too much important endeavor to create systems that hamper of stop the proliferation of fake news and give back to the people not only real information, but also a sentiment of trust in what they are reading. Meanwhile, we need to be prepared to the next challenge, which will be for the definition of what is real, what is important, or even, what is real.

ACKNOWLEDGEMENTS

Project "TEC4Growth - Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

REFERENCES

Abdulla, R. A., Garrison, B., Salwen, M., Driscoll, P., Casey, D., Gables, C., and Division, S. (2002). The credibility of newspapers, television news, and online news.

Allcot, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election.

Anant Goel, Nabanita De, Q. C. M. C. (2017). Fib - lets stop living a lie. https://devpost.com/software/fib. Accessed: 2018-06-18.

Antoniadis, S., Litou, I., and Kalogeraki, V. (2015). A Model for Identifying Misinformation in Online Social Networks. 9415:473–482.

Benevenuto, F., Magno, G., Rodrigues, T., and Almeida, V. (2010). Detecting spammers on twitter. *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, 6:12.

Bovet, A. and Makse, H. A. (2018). Influence of fake news in Twitter during the 2016 US presidential election. pages 1–23.

Castillo, C., Mendoza, M., and Poblete, B. (2011). Information Credibility on Twitter.

Chaves, R. (2018). Fake news detector. https://fakenewsdetector.org/en. Accessed: 2018-06-18.

Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2012). Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, 9(6):811–824.

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. *PLoS ONE*, 10(6):1–13.

CNN (2013). What we know about the boston bombing and its aftermath. https://edition.cnn.com/2013/04/18/us/boston-marathon-things-we-know. Acessed: 2018-06-12.

Cohen, S., Li, C., Yang, J., and Yu, C. (2011). Computational Journalism: a call to arms to database researchers. *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research (CIDR 2011) Asilomar, California, USA.*, (January):148–151.

Conn, D. (2016). How the sun's 'truth' about hillsborough unravelled. https://www.theguardian.com/football/2016/apr/26/how-the-suns-truth-about-hillsborough-unravelled. Acessed: 2018-06-07.

Dickerson, J. P., Kagan, V., and Subrahmanian, V. S. (2014). Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? *ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (Asonam):620–627.

Erahin, B., Akta, Ö., Kilmç, D., and Akyol, C. (2017). Twitter fake account detection. *2nd International Conference on Computer Science and Engineering, UBMK 2017*, pages 388–392.

Figueira, Á. and Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121(December):817–825.

Gilani, Z., Kochmar, E., and Crowcroft, J. (2017). Classification of Twitter Accounts into Automated Agents and Human Users. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17*, pages 489–496.

Gottfried, B. Y. J. and Shearer, E. (2017). News Use Across Social Media Platforms 2017. *Pew Research Center*, Sept 2017(News Use Across Social Media Platforms 2017):17.

Gupta, A. (2011). Twitter Explodes with Activity in Mumbai Blasts! A Lifeline or an Unmonitored Daemon in the Lurking? *Precog.Iiitd.Edu.in*, (September 2017):1–17.

Gupta, A., Lamba, H., and Kumaraguru, P. (2013). $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on twitter. *eCrime Researchers Summit, eCrime*.

Hern, A. (2017). Google acts against fake news on search engine. https://www.theguardian.com/technology/2017/apr/25/google-launches-major-offensive-against-fake-news. Accessed: 2018-04-13.

Hern, A. (2018). New facebook controls aim to regulate political ads and fight fake news. https://www.theguardian.com/technology/2018/apr/06/facebook-launches-controls-regulate-ads-publishers. Accessed: 2018-04-13.

Kollanyi, B., Howard, P. N., and Woolley, S. C. (2016). Bots and Automation over Twitter during the First U.S. Election. *Data Memo*, (4):1–5.

Litou, I., Kalogeraki, V., Katakis, I., and Gunopulos, D. (2016). Real-time and cost-effective limitation of misinformation propagation. *Proceedings - IEEE International Conference on Mobile Data Management*, 2016-July:158–163.

Mrowca, D. and Wang, E. (2017). Stance Detection for Fake News Identification. pages 1–12.

Norton-Taylor, R. (1999). Zinoviev letter was dirty trick by mi6. https://www.theguardian.com/politics/1999/feb/04/uk.politicalnews6. Acessed: 2018-06-07.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2017). Automatic Detection of Fake News.

Pfohl, S., Triebe, O., and Legros, F. (2016). Stance Detection for the Fake News Challenge with Attention and Conditional Encoding. pages 1–14.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., and Stein, B. (2017). A Stylometric Inquiry into Hyperpartisan and Fake News.

Shao, C., Ciampaglia, G. L., Flammini, A., and Menczer, F. (2016). Hoaxy: A Platform for Tracking Online Misinformation. pages 745–750.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K., Flammini, A., and Menczer, F. (2017). The spread of low-credibility content by social bots.

Shiralkar, P., Flammini, A., Menczer, F., and Ciampaglia, G. L. (2017). Finding Streams in Knowledge Graphs to Support Fact Checking.

Sholar, J. M., Chopra, S., and Jain, S. (2017). Towards Automatic Identification of Fake News : Headline-Article Stance Detection with LSTM Attention Models. 1:1–15.

Snopes (2016). Fact-check: Comet ping pong pizzeria home to child abuse ring led by hillary clinton. https://www.snopes.com/fact-check/pizzagate-conspiracy/. Accessed: 2018-04-13.

Starbird, K., Maddock, J., Orand, M., Achterman, P., and Mason, R. M. (2014). Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. *iConference 2014 Proceedings*.

Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., and de Alfaro, L. (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. pages 1–12.

Tambuscio, M., Ruffo, G., Flammini, A., and Menczer, F. (2015). Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. pages 977–982.

The Self Agency, L. (2016). B.s. detector - a browser extension that alerts users to unreliable news sources. http://bsdetector.tech/. Accessed: 2018-06-18.

Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., Isochi, R., and Wang, Z. (2012). Trusting Tweets : The Fukushima Disaster and Information Source Credibility on Twitter. *Iscram*, (April):1–10.

Vargo, C. J., Guo, L., and Amazeen, M. A. (2017). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, page 146144481771208.

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