On the Origin, Proliferation and Tone of Fake News

Shivam B. Parikh, Vikram Patil and Pradeep K. Atrey
Albany Lab for Privacy and Security, College of Engineering and Applied Sciences
University at Albany, State University of New York, Albany, NY, USA
Email: {sparikh, vpatil, patrey}@albany.edu

***Abstract*—Since the advent of social media, we have turned towards consuming news from stand-alone websites to popular social media sites (i.e. Facebook, Twitter, Reddit, etc.); we have noticed a growing number of fake news articles spread across the internet. Existing methods for fake news detection mainly focus on natural language processing and machine learning models to analyze the legitimacy of the news content in order to detect whether it is legit or fake. Currently, there are not many approaches aimed at testing, validating, and ideally refining the findings from traditional fake news detection literature as obtained via surveys and understanding how fake news originate and spread in first place. This paper presents three crucial hypotheses studies that are derived from analyses like, 1) media outlets that publish fake news (origin), 2) social media users who post or share fake news (proliferation), and 3) linguistic (tone) in which fake news are written. The hypotheses are tested on two real-world datasets and results are provided. We envision that this study paves the way to design and develop new multifarious fusion models to detect fake news.**

***Keywords*-Fake news; detection; publishing; proliferation; tone**

I. Introduction

Fake news is any form of visual material that is designed and written to change people's opinion about an individual, an organization, or a belief. With the advent of multimedia editing tools, fake news can be easily made to look like a real one. Further, fakes news is prone to abrupt dissemination through increasing accessibility of the internet and online social media outlets.

Fake news can be very detrimental in many aspects, be it political, economic or knowledge-based [1]. An example of a trending political topic [2] "Did Congress Set Aside $50 Billion in 2006 for the Construction of Border Fencing"; after some fact checking truth reveals that the Secure Fence Act did not authorize any funding for fencing. In early 2018, celebrity Kylie Jenner tweeted that "sooo does anyone else not open Snapchat anymore Or is it just me... ugh this is so sad."; such tweet caused social media company Snapchat about $1.3 billion in a single day; however, it can be argued that it's her opinion and she had no motivation of hurting Snapchat economically, but this just shows the impact of fake news for the economy. Another fake news story [3] "Will Dangerous Cosmic Rays Pass Close to Earth 'Tonight"' is a perfect example of knowledge-based fake news and fact-checking reveals that it has no adverse effect on the surface of the earth. Therefore, it is important to not only devise automated methods to identify fake news with sufficient credibility but also limit their dissemination to minimize the impact.

In the past, multiple research efforts have looked at the problem of fake news from varying perspectives. While social science researchers have studied the impact of fake news on society [4], [5], [6], [1], computer scientists have developed automated methods to detect fake news, largely by analyzing the content using natural language processing (NLP) and multimedia forensics techniques [7], [8], [9], [10]. However, since the accuracy of the content analysis based methods remains limited, it is paramount to understand the other characteristics of fake news that can augment the state-of-the-art content analysis based methods. For example, it is crucial to know i) whether the origin (publishing media outlet) of a given fake news is well-known or not-so-known, ii) whether the author who created the fake news or a social media user who proliferated it is verified or unverified, and furthermore, iii) whether the fake news is crafted with a specific linguistic tone (negative, neutral or positive).

In this paper, we define three hypotheses focusing on learning and understanding the origin, spread and tone of fake news. We evaluate these hypotheses on the FakeNewsNet dataset, which contains 422 news stories. The considered dataset is cleaned appropriately to make it suitable to test our hypotheses (the details are provided in Section IV). After testing the hypotheses, we observed that all three hypotheses hold true.

To summarize, our key contribution is that we studied different characteristics of fake news problem from three different perspectives origin, proliferation, and tone. We believe that the result of our hypotheses tests will give an edge to researchers working in this area to better design the fake news detection algorithm with the incorporation of our hypotheses.

The organization of the rest of this paper is as follows. Section II surveys the related work. In Section III, we discuss various characteristics of fake news and present research questions that form the basis of a set of hypotheses. Section IV describes the dataset used to test the hypotheses. Further, methodologies to test the hypotheses and results from the tests are described in Section V. Finally, Section VI concludes the paper with suggestions for future work.

II. Existing Work

Fake news problem has been studied widely by traditional social science as well as computational social science researchers. Below, we discuss the past works from these two domains.

The social media companies such as Facebook have been criticized for allowing fake news stories to make an impact on people's lives [11]. In fact, in the past, multiple impact studies have focused on citizens' knowledge, opinion, and political trust being influenced by [4], [12], [5], [6]. Also, there are studies that not only go over the impact of fake news on society but also discuss how these impacts are being made using popular social media sites and platforms. For instance, Balmas [13] discussed associations between viewing fake news and altering opinion on a political candidate. The author did that by drawing out certain scenarios about opinions being changed before and after reading the fake news. Also, Brewer et al. [14] described the concept of "intertextuality", which stands for how meaning is derived from synergistic dynamics among multiple messages. This study also uncovers how exposure to certain news coverage could influence knowledge and opinion about a particular subject. The authors in [15] discussed cases where the impact of fake news was severe. Also, Silverman and Singer-Vine in [4] surveyed 3,015 US adults and learned that about 75% of US adults fell for fake news headlines. In another work [5], Kucharski et al. discussed the impact of health-related fake news. Further, Lazer et al. [6] talked about fake news in general sense, history of journalism, impact on individuals and promotes interdisciplinary research to reduce the spread of fake news.

There are a few survey-based studies focused on comparing and drawing out the pros and cons of state-of-the-art fake news detection approaches. For example, in [16], Orellana et al. provided a review about how Twitter is being used for news reporting and studied the impact on readers' attention and engagement of news consumption using Twitter. There are also studies that compare handful datasets available for fake news detection algorithms to compare their results on. For instance, Shu et al. [17] introduced a comprehensive review of fake news detection approaches mainly focused on social media. Their work is broken down into two phases: characterization and detection of fake news.

There is a notable amount of work that has used computational tools like machine learning and NLP techniques. For instance, Rubin et al. [1] explained a theoretical approach to detect fake news by separating them into three types of fake (i.e., serious fabrications, large-scale hoaxes, and humorous fakes). In another work, Hosseini and Bromiatowki [7] proposed a combined approach that leverages machine learning and NLP to train their model and detect fake news. This is performed by training linguistic models to investigate the psychological factors that trigger humans to react to fake news. Also, Perez-Rosas et al. [8] introduced an approach which involves exploratory analyses on fake news dataset to identify linguistic properties that are predominantly present in fake content and they used that knowledge to detect fake news. Further, Chen et al. [9] presented a survey; they cover approaches that can potentially detect clickbait as a form of false news. Subsequently, Horne et al. [18] described characteristics of fake news stories such as title, and physiological and stylistic features to separate fake news from real news. The authors report their results using one-way ANOVA and Wilcoxon rank sum tests on BuzzFeed election and their proprietary political dataset. Also, Rubin et al. [19] proposed a method to verify if given a news article is truthful or deceptive. Their proposed model achieves approximately 63% of accuracy.

Furthermore, Karimi et al. [20] proposed a multi-class fake news detection framework. This framework integrates the following three components: automated feature extraction, interpretable multi-source fusion, and fakeness discrimination. The authors validated the framework using LIAR dataset and achieved the accuracy of 31.81%. Also, Shu et al. [21] presented a tool called FakeNewsTracker that can help users and researchers to collect, detect, and visualize fake news. In addition, they proposed a model to detect fake news using linguistic features and social-context of the news story.

Compared to our previous work [22], in which we reviewed literature and pointed out open research challenges in the area of fake news detection; in this paper, we specifically focus on understanding the phenomena of origination, spread and linguistic tone of fake news by testing hypotheses that we have identified based on existing social science literature [14], [15], [16]. There are a few works such as [14] and [23] that are closely related to the work presented in this paper. While Brewer et al. [14] tested a number of hypotheses in the context of a particular news story and how readers react based on the consumption of that story, Bovet and Makse [23] studied dynamic and influence of fake news on Twitter during 2016 US presidential election. The research presented in this paper, though inspired by these works, particularly [23], is different in the following aspects. In this work, we test two new hypotheses related to the origin and linguistic tone of fake news and revalidate the finding from [23] by testing a hypothesis related to the proliferation of fake news. Furthermore, the study presented in [23] is mostly in the context of the US presidential election, whereas our work doesn't restrict to a particular event, hence is more generalized.

III. Characteristics of Fake News

There exists significant social science research [14], [15], [16] that suggests that structural features of the fake news problem surrounding the original host (URL) and the social network users that spread (not every news gets spread through social media) may provide vital clues to fake news detection. This motivates us to put forward the following three research questions (RQs):

RQ1: Are fake news stories not hosted on popular news websites?
After reviewing the existing approaches and their limitations, we firmly believe that the answer to this research question serves the great need for the researchers looking to leverage site traffic as a deciding factor in their fake news detection method. The sites that have more traffic are considered popular, while the sites that are not much accessed can be put into not-so-popular sites. Particularly, in the case of news publishing sites, popularity can be an important clue. It is generally assumed that the popular news sites publish any news after proper fact checking. However, it is important to verify this supposition with data.

RQ2: Are fake news stories proliferated by unverified users?
The adversaries and vested interest groups who want to spread fake news with the intention of gaining political, economic or other benefits, change people's perception, stock market manipulations, etc. do not want to get caught, since these activities have potential to be investigated by authorities. Hence, adversaries are more likely to create fake accounts. Fortunately, many social media platforms, like Twitter, have a provision of designating users' accounts as verified or unverified by checking other credentials that can establish the authenticity of the user. Therefore, it is imperative to test in our dataset whether fake news are proliferated more by verified or unverified users.

RQ3: Are fake news stories written with a specific linguistic tone?
Fake news stories are intended to be emotional, outraging, and filled with negative emotions in order to change the opinion of the user. Hence, adversaries are more likely to write fake news stories with negative emotions. However, as this is just our conjecture, it should be validated or negated with data.

Based on these three questions, we form three hypotheses which are described in Section V.

IV. Dataset

This study uses the FakeNewsNet [24] dataset which is a combination of PolitiFact [25] and BuzzFeed [26] datasets. Various social contextual information of fake news articles from Twitter like user_profile, user_contents, user_followers, and user_following are also included for fake news content for relevant users. The total size of the dataset is 422 stories and after cleaning data the final size is 347, and further breakdown is shown in Table I. The cleaning procedure is explained in the individual hypothesis sections.

Table I: The details of FakeNewsNet dataset

	FakeNewsNet (Combined)		BuzzFeed		PolitiFact	
	Fake	**Legit**	Fake	Legit	Fake	Legit
Actual dataset	**211**	**211**	91	91	120	120
Cleaned dataset	**139**	**208**	83	91	56	117

V. Hypotheses Testing

In this section, we present three hypotheses based on the research questions posed earlier (in Section III), and describe the methodologies and test results of these hypotheses.

A. H1 Fake News Stories and Popular News Websites

In order to measure the popularity of the web pages, we utilize Amazon's Alexa Web Information Service [27]. This state-of-the-art service provides detailed information about a particular URL and also allows us to calculate the web pages' popularity ranking.

1) Dataset for H1: In the following two subsections, we describe how we cleaned the original data for hypotheses testing.

Dataset of popular news web sites: In order to get a list of top 100 US-based news websites, we referred FeedSpot [28] based on the index using search and social metric. We calculated the popularity of these websites using the rank provided by Alexa Web Information Service. The mean popularity rank for the complete dataset before removing anomalies is 36,968.9. Since maximum 7 entries were significantly different (4 times higher) than the mean value and rest of the dataset, we have considered them as anomalies and ignored them. The least popular rank of the 93rd popular website (after ignoring 7 anomalies from 100 news sites) has the page rank of 89,885 as shown in Table II (as accessed in December 2018).

Fake news dataset: In the PolitiFact dataset, there are 120 fake news stories but some of them are data dumps, so we could not calculate their Alexa global page rank. Hence, we ignored those news stories and created a cleaned version of the dataset, and performed experiments on the cleaned version, which led us to 56 news stories. In BuzzFeed dataset, we have 91 fake news stories, similarly, we ignored 8 news stories whose Alexa global page rank was not derivable and ended up with 83 news stories. The combination of the above datasets yields a total 139 fake news stories, highest ranked site and lowest ranked news sites are shown in Table III.

2) t-test for H2: For this hypothesis, we use t-test for the metrics of popular websites and the dataset of the fake news stories.

We define the null and alternate hypotheses as follows:

Null Hypothesis (H_0): There is no significant difference between means of the fake news stories Alexa global page ranks and the popular websites Alexa global page ranks.

Table II: Sample of top popular news sites

Rank	Popular News Sites	Global Rank
1	yahoo.com	7
2	espn.com	76
3	cnn.com	103
. . .	. . .	. . .
92	villagevoice.com	88743
93	minnpost.com	89885

Table III: Sample of fake news stories sites (high to low)

Rank	Fake News Sites	Global Rank	Dataset
1	londonwebnews.com	18890626	PolitiFact
2	spinzon.com	17818495	PolitiFact
3	uspoln.com	17818405	PolitiFact
. . .	. . .	. . .	. . .
137	usherald.com	49094	BuzzFeed
138	rightwingnews.com	2076	BuzzFeed
139	author.addictinginfo.org	618	BuzzFeed

Alternate Hypothesis (H_1): There is a significant difference between means of the fake news stories Alexa global page ranks and popular websites Alexa global page rank.

The t-test results are provided in Table IV. As can be observed from the table, the mean (μ) and standard deviation (σ) of the two categories are quite apart. The p-value is 7.8167e-15 $\approx$ 0.00, which indicates that we can reject the null hypothesis and accept the alternate hypothesis.

We also visualized the page rank of fake news stories with logarithmic scale in Figure 1 and we clearly observe that page ranks of the news stories are greater than the threshold τ = 89,885, which is represented by the horizontal line in the figure. This also further supports the hypothesis. Since we can only calculate the popularity score of the limited number of websites and not all, this can not be treated as sole criteria.

B. H2. Unverified Users and Proliferation

1) Dataset for H2: Creating a new account on any social media platforms like Twitter is effortless. However, there are only less than 10% verified accounts on Twitter. The combined dataset of fake news stories contains 139 entries (refer to Table I). We divide this dataset into two subsets of verified and unverified Twitter users. As shown in Table VI,

Table IV: Result of t-test for $H1$

Test Parameters	Fake News Websites	Popular Websites
Mean μ	4990381.94	21745.24
Standard deviation σ	5755230.85	22751.18
Sample size n	139	93
p-value	7.8167e-15 $\approx$ 0.00	
Result	Accept	

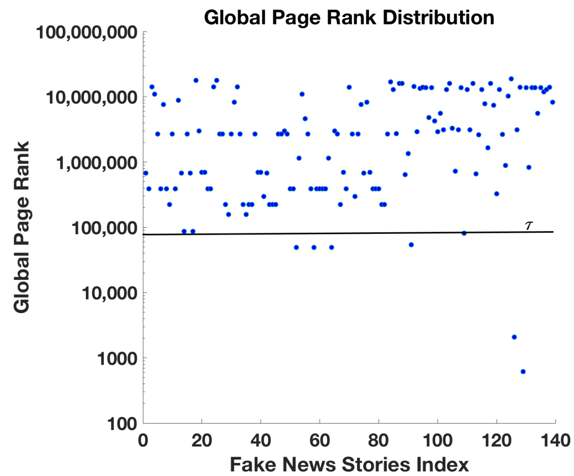

Figure 1: Web page rank for fake news stories. The horizontal line depicts the threshold $\tau = 89,885$.

Table V: Distribution of combined fake news stories between verified and unverified users

Dataset	Actual (Ratio) Verified Users	Unverified Users
BuzzFeed	9 (10.8%)	74 (89.2%)
PolitiFact	3 (5.4%)	53 (94.6%)
Combined	12 (8.6%)	127 (91.4%)

fake news stories posted by verified and unverified users are 12 and 127, respectively.

2) χ^2-test for H2: For testing this hypothesis, we define the null and alternate hypotheses as follows:

Null Hypothesis (H_0): There is no significant difference between verified and unverified users who spread fake news stories.

Alternate Hypothesis (H_1): There is a significant difference between verified and unverified users who spread fake news stories.

In order to test this hypothesis, we used χ^2-test. We used the expected proportion of 0.5 for the χ^2-test and obtained χ^2 value = 95.14 for 1 degree of freedom. Hence, we reject our null hypothesis and accept the alternate hypothesis. With this, we can conclude that unverified Twitter users spread fake news stories. This also reconfirms the similar finding from [23], but with different methodology and dataset. Table VI elaborates this in brief.

It is also important to mention that sometimes the verified users may also believe the fake news to be true and retweet it. In some cases, when the fake news concerns the verified users, e.g. people have a tendency of retweeting a story if it favors them or their ideology, which limits the results of this hypothesis.

Table VI: Result of χ^2-test of *H2*, with expected proportion 0.5 and degree of freedom 1

Dataset	Verified Users	Unverified Users	Total
Observed (expected)	12 (69.5)	127 (69.5)	139
χ^2-value	95.14		
Level of significance α	0.05		
Critical value	3.84		
p-value	$0.0045 \approx 0$		
Result	Accept		

C. H3. Fake News Stories and Sentiments

It is widely assumed that most of the fake news articles have a negative sentiment, e.g., outrageous, deceitful stories, etc. Hence, we decided to examine if this is indeed true. To test this hypothesis, we utilized Microsoft Azure's text analysis service [29]. This is an NLP service that estimates whether the given story is written with positive, negative, or neutral sentiments. Using this service, we determined the sentiment of the articles in the given two datasets.

1) Dataset for H3: In the PolitiFact dataset, there were 2 news stories that did not have text data (only had multimedia data). Similarly, in the BuzzFeed dataset, there were 2 news stories of such kind. Hence, we ignored a total 4 news stories from the 211 combined fake news stories (refer to Table I). Thus, after ignoring the 4 news stories we have accumulated 207 total fake news stories, as mentioned in Table VII.

We calculated the sentiment of these 207 news stories using Azure's text analysis service, which returns the sentiment score in the range of 0.0 to 1.0. The scores between 0.0 to 0.3 represent negative sentiment, the score between 0.3 to 0.7 represents neutral sentiment, and the score between 0.7 and 1.0 represent the positive sentiment. We grouped and calculated the total number of sentiments for a specific category and reported them in Table VII.

2) χ^2-test for H3: For testing this hypothesis, we define the null and alternate hypotheses as follows:

Null Hypothesis (H_0): There is no relation between tone (sentiments) and fakeness of a news story.

Alternate Hypothesis (H_1): There is a relation between tone (sentiments) and fakeness of a news story.

We use χ^2-test with three groups of stories (stories with negative sentiments, stories with neutral sentiments, and stories with positive sentiments). For this test, we used the expected proportion of 0.33 for each category, with 2 degree of freedom. We obtained χ^2 as 57.04 (see Table VIII), which suggests that the null hypothesis can be rejected and the alternate hypothesis can be accepted.

Based on the results, we can say that a large percentage of them (114 out of 207) have a neutral sentiment. However, further experimentation is required on a larger dataset to buttress our observations.

Table VII: Fake news stories with sentiment ranges: positive (0.7-1.0), neutral (0.3-0.7) and negative (0.0-0.3)

Dataset	Sentiment	Fake news stories
BuzzFeed	Positive	20
	Neutral	63
	Negative	35
PolitiFact	Positive	5
	Neutral	51
	Negative	33
Combined	Positive	25
	Neutral	114
	Negative	68

Table VIII: Result of χ^2-test of *H3*, with expected proportion 0.33 and degree of freedom 2

Dataset	Positive	Negative	Neutral	Total
Observed (expected) fake news stories	25 (69)	68 (69)	114 (69)	207
χ^2-value	57.04			
α	0.05			
Critical-value	5.99			
p-value	$0.00001 \approx 0$			
Result	Accept			

VI. Conclusion and Future Work

Fake news is an important challenge that takes place over social and computational infrastructure. In this paper, we have proposed multiple hypotheses related to three different characteristics of fake news: origin, proliferation and linguistic tone. The hypotheses are tested using statistical methods on the FakeNewsNet dataset collected via Twitter. The results of these hypotheses suggest the following: 1) fake news is not published on popular websites, rather they are published by lesser known media outlets or websites; 2) fake news are proliferated more by unverified users compared to verified users; and 3) the fake news stories are written in a specific linguistic tone, though it is inconclusive to say which one (negative, positive or neutral). The results expand the understanding of fake news as a phenomenon and motivate future work, which includes: 1) expanding the study on additional hypotheses, and 2) designing and developing a multifarious fusion model to better detect given news for fakeness or legitimacy.

References

[1] V. Rubin, Y. Chen, and N. Conroy, "Deception detection for news: three types of fakes," in *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, St. Louis, MO, USA, 2015, p. 83.

[2] "Did congress set aside $50 billion in 2006 for the construction of border fencing" https://www.snopes.com/fact-check/secure-fence-act-50-billion-border/, accessed 01/2019.

[3] "Will dangerous cosmic rays pass close to earth tonight" https://www.snopes.com/fact-check/dangerous-cosmic-gamma-rays/, accessed 01/2019.

[4] C. Silverman and J. Singer-Vine, "Most americans who see fake news believe it, new survey says," *BuzzFeed News*, 2016.

[5] A. Kucharski, "Post-truth: Study epidemiology of fake news," *Nature*, vol. 540, no. 7634, p. 525, 2016.

[6] D. Lazer, M. Baum, Y. Benkler, A. Berinsky, K. Greenhill, F. Menczer, M. Metzger, B. Nyhan, G. Pennycook, D. Rothschild *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[7] P. Hosseini, M. Diab, and D. Broniatowski, "False news on social media," *arXiv preprint arXiv:10.13140*.

[8] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," *arXiv preprint arXiv:1708.07104*, 2017.

[9] Y. Chen, N. Conroy, and V. Rubin, "Misleading online content: Recognizing clickbait as false news," in *Proceedings of the ACM Workshop on Multimodal Deception Detection*, Seattle, WA, USA, 2015, pp. 15–19.

[10] Y. Li, M.-C. Chang, H. Farid, and S. Lyu, "In ictu oculi: Exposing AI generated fake face videos by detecting eye blinking," *arXiv preprint arXiv:1806.02877*, 2018.

[11] "Facebook and Cambridge Analytica: What you need to know as fallout widens," accessed 01/2019.

[12] L. Sydell, "We tracked down a fake-news creator in the suburbs. heres what we learned," *National Public Radio*, vol. 23, 2016.

[13] M. Balmas, "When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism," *Communication Research*, vol. 41, no. 3, pp. 430–454, 2014.

[14] P. Brewer, D. Young, and M. Morreale, "The impact of real news about fake news: Intertextual processes and political satire," *International Journal of Public Opinion Research*, vol. 25, no. 3, pp. 323–343, 2013.

[15] Y. Papanastasiou, "Fake news propagation and detection: A sequential model," *Available at SSRN Electronic Journal 3028354*, 2018.

[16] C. Orellana-Rodriguez and M. Keane, "Attention to news and its dissemination on Twitter: A survey," *Elsevier Computer Science Review*, vol. 29, pp. 74–94, 2018.

[17] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[18] B. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," *arXiv preprint arXiv:1703.09398*, 2017.

[19] V. Rubin, N. Conroy, and Y. Chen, "Towards news verification: Deception detection methods for news discourse," in *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, Kauai, HI, USA, 2015, pp. 5–8.

[20] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, NM, USA, 2018, pp. 1546–1557.

[21] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: A tool for fake news collection, detection, and visualization," *Computational and Mathematical Organization Theory*.

[22] S. Parikh and P. K. Atrey, "Media-rich fake news detection: A survey," in *The 1st IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, USA, 2018, pp. 436–441.

[23] A. Bovet and H. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," *Nature communications*, vol. 10, no. 1, p. 7, 2019.

[24] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media," *arXiv preprint arXiv:1809.01286*, 2018.

[25] "PolitiFact," https://www.politifact.com/truth-o-meter/, accessed 11/2018.

[26] "BuzzFeedNews: 2017-12-fake-news-top-50," https://github.com/BuzzFeedNews/2017-12-fake-news-top-50, accessed 11/2018.

[27] "Alexa web information service," https://docs.aws.amazon.com/AlexaWebInfoService/latest/index.html, accessed 01/2019.

[28] "Top 100 usa news websites on the web," https://blog.feedspot.com/usa_news_websites/, accessed 12/2018.

[29] "Azure text analytics," https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/overview, accessed 01/2019.