Towards Impact Scoring of Fake News

Shivam B. Parikh, Vikram Patil, Ravi Makawana 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, rmakawana, patrey}@albany.edu

***Abstract*—Tackling fake news has become one of the highest priority issues in recent years. Its effects are damaging, and urgent attention needs to be paid to find a solution. Although several approaches have been proposed to understand the characteristics of fake news, not much attention has been given to determining its impact. Fake news can also be casual opinions or sarcasm, entertainment-related media stories, which pose no risk. Alternatively, there are some high impact news stories which causes panic, confusion, rumors and is responsible for real damage. Hence, in this paper, we propose a model to calculate the impact score of a given fake news story. The proposed model integrates various factors related to fake news, such as the scope of the news, the reputation of the publishing site, and the popularity of the proliferator. We validate the results of the proposed impact computation model with user surveys.**

***Keywords*-Fake News; Impact Computation Model; Social Media**

I. Introduction

In many instances, information is broadcasted by news/media outlets, social media platforms, and blogs, etc. This information could be authentic, verified news or false, unverified news. Sometimes the unverified news is also referred to as rumors, industry grapevine, propaganda and so on, based on the domain. In most cases, unverified or false news is harmless news, e.g., mimicry, sarcasm or jokes regarding a person or incident which is primarily intended for entertainment purposes. We come across such social media posts every day. However, in few cases, the news is spread with malicious intentions such as political gains, instigating or spreading violence or hatred between communities, and in some cases, manipulation of stock prices by purposefully spreading fake news around a specific company. In these instances, the fake news is not only harmful, but it can be fatal. Interesting fact about fake news is, the damage is not done by the fake news itself but by the users who interpret it as real and act on it in panic. Often, proliferators use fake accounts, and bots to broadcast the news story to millions of users across various social media platforms.

When a user comes across a news story, his reaction is based on certain characteristics of the news story. For instance, the user might enjoy reading about his favorite sport, like a post of from his friends, share a news story which he thinks is interesting, sign up for an event or a service which interests him. Subconsciously, the user may ponder about its authenticity and then either believe or ignore it. On the other hand, if the user is stressed or the same information is repeated on his or her social media feed or sees his or her connection(s) reacting to it, the user perceives that the news is probably right and acts on it in anxiety.

Fake news can be classified into two categories as impactful fake news and impactless fake news, as depicted in Figure 1. The impactful fake news is those articles which have the potential to cause panic and anger, and the readers feel like taking actions after reading those news stories. These actions can include selling a stock, changing the perception of an event, riots, etc. These are very dangerous, and more attention needs to be paid in this area. The impactless fake news is entertainment, parody, rumors, etc. which do not have much impact and are not dangerous.

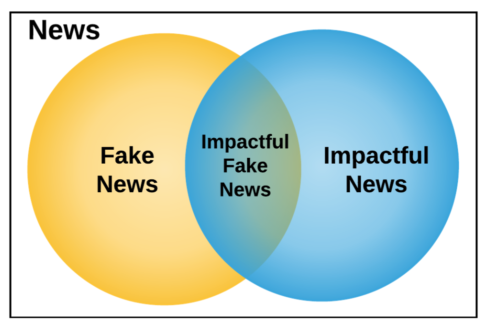

Figure 1: An illustration of impactful fake news.

There are numerous approaches aimed to detect the fake news articles in the literature as we have described in our previous work [1]. However, to the best of our knowledge, this is the first paper, which provides a mathematical model to compute the impact score of the fake news. The proposed impact computation model is based on the weighted fusion of different factors such as the scope of the news (political, economic, etc.), the credibility of the news publishing site (URL), and proliferator's popularity (in terms of his/her followers). With this model, researchers can focus on stopping the proliferation of the impactful fake news.

The rest of the paper is organized as follows. Section II provides an overview of the existing approaches from the literature. Section III describes the proposed impact computation model and impacting factors. In Section IV, we validate the model using a real dataset and discuss the results. Finally, Section V provides conclusion and future work.

II. Related Work

There have been several approaches proposed to detect fake news and bots. Most of these approaches are generic and may produce false positives, i.e., stop fake news which may not be harmful or meant for entertainment, etc. However, there has been a surprisingly limited amount of work done in the area of detecting the impact of fake news and its characteristics.

Majority of social media companies have their approaches to deal with spams accounts. Facebook's approach involves detecting inauthentic behavior [2]. It detects accounts that are operated by bots. Further, Facebook's approach for preventing impactful spams focuses on the behavior of an account rather than the content [2]. Also, Twitter's initiative on Election Integrity [3] is focused on giving users more transparency on U.S. federal political campaigning ads. Twitter does this by revealing billing information, ad spend, impression data, etc., which allows users to make an educated decision [4].

There is a selective amount of work on fake news detection that has some elements that can be used to detect impactness of the news story in the subject, but these approaches do not leverage impactful scoring before detection of fake news. In [5], Gupta et al. performed classification analysis using URLs of images, follower-friend ratio, user verified or not, the number of followers, etc. Their work was mainly focused on detecting tweets with fake images based on certain characteristics. Also, Anger et al. [6] discussed an interesting approach to measuring user's influence on Twitter. Authors discussed how reactions, retweets, mentions, and other ranking services could be used, but out of all of them, the author's number of followers and following is a very impactful and can be partially leveraged for scoring impact of news in question. Further, in [7], Aldwairi et al. proposed a simple approach to detect fake news in social media networks using certain information from posts such as title, use of numbers in title, capital words in title, user behavior on site, occurrence of title words in content, and keywords from title. Similarly, in [8] authors introduced an algorithm that can be utilized to crowdsource the reporting mechanism. Once users have registered a certain number of negative reports, manual fact-checking can be performed on the news in the subject.

Note that there has been a significant amount of work in the area of bot detection [9], [10] and [11]. However, bot detection approaches do not consider the content of the posts proliferated by such accounts. It is worth mentioning that approaches for fake news detection focus on classifying a news into real or fake category, rather than determining its impact.

In [12], Piotrkowicz et al. studied the impact of the linguistic features of headlines of the news stories on their popularity. Similarly, in [13], the authors discussed the impact of fake news consumption on the US presidential election. Compared to existing work, this work presents a computational model to determine the impact of a fake news story based on various factors such as scope, publishing site's reputation, and proliferator's popularity. To the best of our knowledge, this is the first attempt to develop a model to compute the impact score for fake news.

III. Proposed Work

In this section, we describe the impact computation model and the three impacting factors that are used in this work.

A. Impact Computation Model

Let there be n factors that determine the impact score of a given fake news story and $x_i \in [0, 1]$ be the impact score of the i^{th} factor, $1 \leq i \leq n$. Also, let $w_i \in [0, 1]$ be the normalized weight for the i^{th} factor, i.e. $\sum_{i=1}^{n} w_i = 1$. The objective is to determine the impact score $I \in [0, 1]$ using a weighted linear fusion function, as follows:

$$I = \sum_{i=1}^{n} w_i \times x_i \tag{1}$$

B. Impacting Factors

Although the proposed impact model is generic to include any number of factors, in this paper, we consider three factors (i.e., $n = 3$) for determining the impact of the fake news stories. These are 1) scope of the fake news, 2) reputation of the publishing site, and 3) proliferator's popularity. In the following sections, we describe these in detail.

1) Scope: The scope or the category of the news article can be defined as a set S = {*Politics, Economics, Crime, Science*}. A given fake news story N is said to be impactful if its scope or category falls within S. Let $f()$ be a function that returns the category or scope of the given news story.

We calculate the impact score x_1 for the first factor (scope), as follows:

$$x_1 = \begin{cases} 1 & \text{if } f(N) \in S \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In this work, we used the Text Razer service [14] to determine the scope for a given news story.

2) Publishing site's reputation: In order to determine the reputation of the website that published the news in question, we extract its keywords and run it using the Google search API. Further, we compare the search results with the top 100 news websites (in the context of US). If the top m search results contain at least one popular news website, we conclude that the news story is also published by the popular news websites. This increases the impactness of the news story.

Let $U = \{U_1, U_2, U_3, \ldots, U_{100}\}$ be the set of top 100 popular news stories, where U_j represents the URL of the j^{th} news website. Also, let $R = \{URL_1, URL_2, URL_3, \ldots, URL_m\}$ be the set of m search results obtained from the Google search API for the given keywords of a news story.

Next, using Equation (3), we calculate $m' \leq m$, which is the total number of websites in the search results (i.e. R) that matched with the popular websites of rank lower than τ (the lower the rank, the higher the popularity).

$$m' = \{U_j \mid \phi(U_j) \leq \tau\} \tag{3}$$

where $\phi()$ is the function to calculate the popularity rank of the URL and τ is the highest popularity rank from the top 100 news websites.

Finally, we calculate the second impacting factor x_2 using the following equation:

$$x_2 = 1 - e^{-(m'+\delta)\times\alpha} \tag{4}$$

where δ and α are configurable parameters; δ is a very small value which accounts for the fact that there could a minimal impact of a news irrespective of whether it is published on popular site or not-so-popular site; and α determines the growth of impact when a news is published on the increasing number of popular sites.

3) Proliferator's popularity: The proliferator of the post is a crucial characteristic that can affect the impact of the fake news story. The popular and trustworthy social media users have not only a huge number of followers, but their posts also receive many likes and shares. For the sake of simplicity, in this work, we restrict it to the number of followers. Such information can be retrieved using the APIs provided by the particular social media platforms; for instance, Twitter API. In [6], Anger et al. discussed follower-following ratio and other characteristics to classify tweets with fake images. This motivates us to use the followers feature for determining the proliferator's popularity. We define the third impacting factor x_3 (i.e., proliferator's popularity score) as:

$$x_3 = 0.5 \odot \lambda \tag{5}$$

where λ and $\odot$ are defined by the following two equations:

$$\lambda = \frac{1}{2} \times \frac{current - mean}{limit - mean} \tag{6}$$

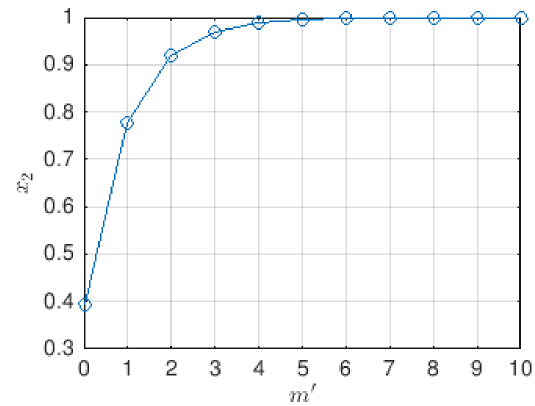

Figure 2: An illustration of growth of impact factor (x_2) with respect to the number of popular sites in the search result that published the story (m') for $\delta = 0.6$ and $\alpha = 0.8$.

$$\odot = \begin{cases} + & \text{if } current > mean \\ - & \text{otherwise} \end{cases} \tag{7}$$

where the term $limit$ is determined as follows:

$$limit = \begin{cases} max & \text{if } current > mean \\ min & \text{otherwise} \end{cases} \tag{8}$$

In the above three equations, $current$ is the number of followers of the proliferator of the fake news story in question, and $mean$, max and min are the average, maximum and minimum of the number of followers of the proliferators in general, respectively. In this work, we calculate these statistics from the dataset (refer to Section IV) that is used to validate the proposed model.

IV. Experiments and Results

In this section, we describe the dataset and present the experimental results for validating the proposed model. Also, we also present experiments to determine the appropriate values of the parameters used in the model.

A. Dataset

In this work, we used publicly available FakeNewsNet dataset [15], which is a combination of two datasets, PolitiFact [16] and BuzzFeed [17]. The FakeNewsNet dataset consisted of 211 fake news stories. Since for the proposed model, we require categorical data, URL information, and proliferator's number of followers; only 54 out of 211 news stories have all the information we need for the proposed model, and therefore we used the 54 news stories in our experiments.

B. Calculating Impacting Factors

Here we show how we calculate the values of impacting factors x_1, x_2 and x_3. For x_1, out of 54 fake news stories there were 48 stories that belonged to the specified domains;

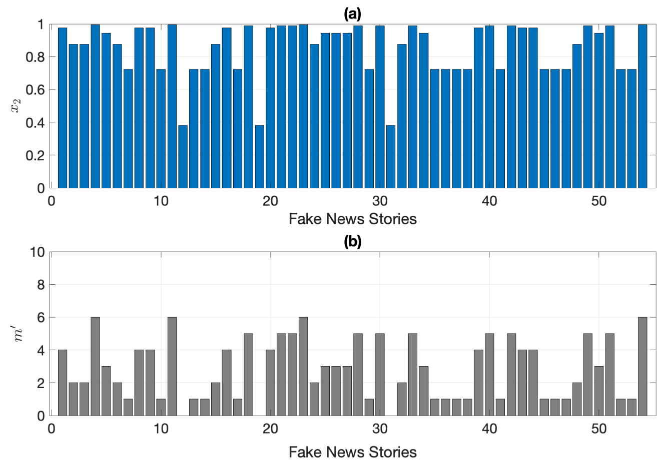

Figure 3: The values of x_2 with respect to m' for 54 fake news stories: (a) the value of x_2, and (b) the value of m'.

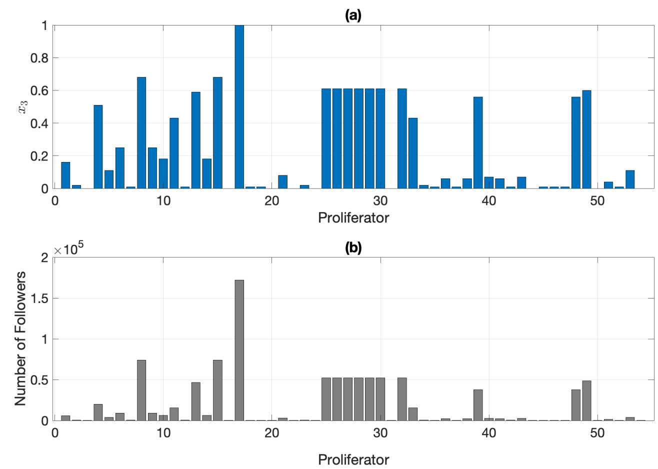

Figure 4: The values of x_3 with respect to the number of followers for the proliferators of 54 fake news stories: (a) the value of x_3, and (b) the number of followers.

hence $x_1 = 1$ for these 48 stories, leaving $x_1 = 0$ for only remaining 6 stories.

For x_2, we used $\delta = 0.6$ and $\alpha = 0.8$ in the model given in Equation (4). We chose these values because it provided us the maximum similarity in the results obtained via the proposed model and the user study, however it cannot be generalized and these parameters can be learned empirically for different datasets. With these values, x_2 increases as m' increases as shown in Figure 2. As can be observed from the figure, x_2 reaches close to 1 with $m' = 6$, which suggests that a news story when published by 6 or more popular sites would have almost 100% impact. Further, in Figure 3(a) and Figure 3(b) we show the value of m' and x_2, respectively, for all 54 stories.

The x_3 is calculated using the model described by Equation (5) through Equation (8). The value of x_3 is directly proptional to the number of proliferators' followers, which is depicted in Figure 4.

C. *Model Validation with User Study*

Initially we provided equal weights to all three factors, i.e. $w_1 = 0.33$ or 33% for the scope factor (x_1), $w_2 = 0.33$ for the publishing site's reputation score (x_2), and $w_3 = 0.33$ for the proliferator's popularity score (x_3). With these weights and the scores (x_1, x_2 and x_3) for the three impact factors, we calculated the overall impact score I for all 54 stories.

To gauge the correctness of the impact scores obtained through the proposed model, we conducted a user study. This study involved three users and was anonymous. Users were shown the 54 news stories from the dataset and were asked to rate the impact on a scale of 0 to 10, 0 being the least impact and 10 being the most impact. We calculated the Mean Opinion Score (MOS), denoted by M and divided it by 10 to bring it to the scale of 0 to 1. The comparison of impact score I computed using the proposed model and

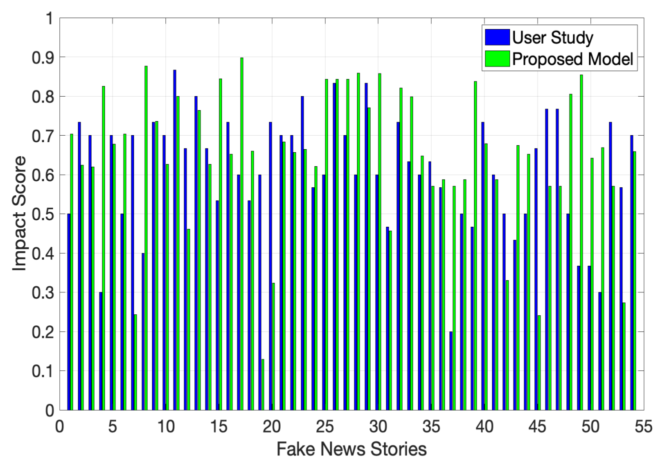

Figure 5: The comparison between the impact score (I) calculated using the proposed model and the MOS (M) obtained via user study, with equal weights to all three factors and $m = 6$.

M obtained from the user study for 54 fake news stories are shown in Figure 5. It can be observed from the figure that I is mostly similar to MOS within a reasonable range.

To further examine the results of the proposed model as compared to user study, we used the following two distance metrics, Mean Average Error (MAE) and Mean Square Error (MSE), between I and M. Let ϵ_{MAE} be the MAE, and ϵ_{MSE} be the MSE. The values of ϵ_{MAE} and ϵ_{MSE} for the data shown in Figure 5 are 0.1856 and 0.0566, respectively. ϵ_{MAE} = 0.1856 suggests that the average difference between the impact score, I, computed using the proposed model and the MOS, M, obtained via user study is reasonably small, i.e. $\approx$ 18%. However, since in our data there were some

Table I: ϵ_{MSE} based on different weight criterion for $m = 6$. Criteria 1: Equal weights; Criteria 2: Higher weights to the two factors than the third one; and Criteria 3: Higher weight to one factor than the other two factors

Weight Criteria	w_1	w_2	w_3	ϵ_{MSE}
Criteria 1	0.33	0.33	0.33	0.0566
Criteria 2	0.20	0.40	0.40	0.0521
	0.40	0.20	0.40	0.0632
	0.40	0.40	0.20	0.0736
Criteria 3	0.50	0.25	0.25	0.0758
	0.25	0.50	0.25	0.0586
	0.25	0.25	0.50	0.0631

unexpected differences between I and M, we preferred to use ϵ_{MSE} for the rest of the analyses in order to compensate for unexpected values.

D. Determining Appropriate Weights

In Table I, we show the comparison of ϵ_{MSE} for different weight options. We started with giving equal weights (i.e., criteria 1) to all three factors and then tested with a few different combinations. We experimented with giving higher weights to two of the factors than the third one (i.e., criteria 2), followed by giving higher weight to one specific factor than the other two factors (criteria 3). From the presented results, we can say that having $w_1 = 0.20$, $w_2 = 0.40$, and $w_3 = 0.40$ gives the least ϵ_{MSE} of 0.0521, which suggests that the publishing site's popularity and the proliferator's reputation are more dominating factors than the scope of the news. Note that this result is specific to the dataset we used, and further investigation is needed to generalize this finding.

V. Conclusion and Future Work

In this paper, we introduced a model to analyze the impact of the fake news story and provide its impact score. The proposed model considers three impacting factors, and it is tested on public fake news dataset and validated with a user study. We presented our preliminary ideas in this paper; future work includes expansion of the proposed impact computation model with the other factors, learning of the model parameters using a large dataset, deploying the model to identify the fake news of high impact in real-life scenarios and inventing mechanisms to limit the spread of such news.

References

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

[2] "Removing coordinated inauthentic behavior from Russia," https://newsroom.fb.com/news/2019/01/removing-cib-from-russia/, accessed January, 2019.

[3] "Twitters focus is on a healthy public conversation." https://about.twitter.com/en_us/values/elections-integrity.htm, accessed January, 2019.

[4] "Increasing transparency for political campaigning ads on Twitter," https://blog.twitter.com/official/en_us/topics/company/2018/Increasing-Transparency-for-Political-Campaigning-Ads-on-\Twitter.html, accessed January, 2019.

[5] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: characterizing and identifying fake images on Twitter during hurricane sandy," in *The 22nd International Conference on World Wide Web*, Rio de Janeiro, Brazil, 2013.

[6] I. Anger and C. Kittl, "Measuring influence on Twitter," in *The 11th International Conference on Knowledge Management and Knowledge Technologies*, Graz, Austria, 2011.

[7] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," *Procedia Computer Science*, vol. 141, pp. 215–222, 2018.

[8] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez-Rodriguez, "Leveraging the crowd to detect and reduce the spread of fake news and misinformation," in *The Eleventh ACM International Conference on Web Search and Data Mining*, Los Angeles, CA, USA, 2018.

[9] C. Xiao, D. Freeman, and T. Hwa, "Detecting clusters of fake accounts in online social networks," in *The 8th ACM Workshop on Artificial Intelligence and Security*, Denver, CO, USA, 2015.

[10] J. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?" in *The 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Beijing, China, 2014.

[11] F. Morstatter, L. Wu, T. Nazer, K. Carley, and H. Liu, "A new approach to bot detection: striking the balance between precision and recall," in *The 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, San Francisco, CA, USA, 2016.

[12] A. Piotrkowicz, V. Dimitrova, J. Otterbacher, and K. Markert, "The impact of news values and linguistic style on the popularity of headlines on Twitter and Facebook," in *The Second International Workshop on News and Public Opinion*, Montreal, QC, Canada, 2017.

[13] A. Guess, B. Nyhan, and J. Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign," *European Research Council*, 2018.

[14] "Text Razer," https://www.textrazor.com/, accessed 01/2019.

[15] 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.

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

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