Detecting Fake News Spreaders on Twitter Through Follower Networks

Smita Ghosh[1], Juan Manuel Zuluaga Fernandez[1], Isabel Zuluaga González[2], Andres Mauricio Calle[1], and Navid Shaghaghi[1](✉)

[1] Ethical, Pragmatic, and Intelligent Computing (EPIC) Research Lab, Santa Clara University (SCU), Santa Clara, CA 95053, USA
{sghosh3,jzuluga,aclle,nshaghaghi}@scu.edu

[2] Universidad Pontificia Bolivariana, Medellín, Antioquia, Colombia
isabel.zuluaga@upb.edu.co

https://www.scu.edu, https://www.upb.edu.co/es/home

Abstract. Obtaining news from social media platforms has become increasingly popular due to their ease of access and high speed of information dissemination. These same factors have, however, also increased the range and speed at which misinformation and fake news spread. While machine-run accounts (bots) contribute significantly to the spread of misinformation, human users on these platforms also play a key role in contributing to the spread. Thus, there is a need for an in-depth understanding of the relationship between users and the spread of fake news. This paper proposes a new data-driven metric called *User Impact Factor (UIF)* aims to show the importance of user content analysis and neighbourhood influence to profile a fake news spreader on Twitter. Tweets and retweets of each user are collected and classified as 'fake' or 'not fake' using Natural Language Processing (NLP). These labeled posts are combined with data on the number of the user's followers and retweet potential in order to generate the user's impact factor. Experiments are performed using data collected from Twitter and the results show the effectiveness of the proposed approach in identifying fake news spreaders.

Keywords: Bidirectional Encoder Representations from Transformers (BERT) · Fake News Detection · Misinformation Spread · Natural Language Processing (NLP) · Social Media · Twitter · User Impact Factor (UIF)

1 Introduction

The past decade has seen an increase in the public's reliance on social media as a source not only of entertainment, but also of news and commentary on current affairs. Unlike traditional media sources, social media provides extremely low barriers of entry to anyone seeking to disseminate information. These new sources of news also benefit from their platforms being not specifically news focused, allowing them to capture the attention of a wider audience who might

R. Hou et al. (Eds.): BDTA 2022, LNICST 480, pp. 181–195, 2023.
https://doi.org/10.1007/978-3-031-33614-0_13

have originally logged on for a different reason but were recommended the post by an advertising algorithm. In January 2019, Pew Research Center announced that 59% of Twitter's users get their news from the social media platform regularly [6]. Despite it's advantages, the quality of news on social media is however considered lower than that of traditional media [13].

The same low barriers of entry to disseminating news also result in the ability to spread much lower quality and even fake information to billions of users. Spreading of fake news has a detrimental effect on society. For example, during the initial outbreak of the COVID-19 pandemic, social media applications were heavily used in order to spread false information on the pandemic, prevention measures, and treatments [2]. During the 2016 presidential elections, a popular fake news story known as 'Pizzagate' was retweeted over 1 million times in November 2016 [7]. Thus it is not only important to detect fake news but also be able to find ways to anticipate their spreaders.

Detecting fake news spreaders involves detecting fake news and analyzing the pattern of the spread of such information. The former is achieved through analyzing the content of misinformation posts, which tend to invoke sensationalist tropes and often partisan language to attract and mislead their audience. The latter involves understanding the role of users in the propagation of misinformation on social media platforms [12,15,28].

This paper focuses on new solutions for detecting fake news spreaders through follower networks. This is done by understanding the relationship between users and their influence on not only their followers, but the followers of their followers. A new data-driven metric called *User's Impact Factor (UIF)* is proposed which combines a user's activity of posting and the influence of people they follow. These two properties help profile fake news spreaders. The aim of this novel metric is to highlight the interconnected nature of Twitter, and how a user's influence can extend past their immediate followers to have a greater impact on the spread of misinformation within a larger network.

The main contributions of this paper are:

1. Creating a repository of user's data by collecting tweets and follower data from Twitter.
2. Training a highly accurate and reliable fake news classifier to label the tweets and retweets as 'fake' or 'not fake'.
3. Proposing and formulating a new data-driven metric called *User's Impact Factor* to help profile fake news spreaders on Twitter.

This paper is organized as follows: Sect. 2 describes the background and related work. Section 3 describes the phases of the proposed approach and Sect. 4 talks in detail about the first two phases. Sections 5 and 6 describe the formulation of the new proposed UIF metric and show the results of utilizing it on gathered Twitter data. Section 7 provides concluding remarks and lastly, Sect. 8 describes the future directions for this work.

2 Background and Related Work

2.1 Fake News

The COVID-19 pandemic has served to amplify concerns regarding misinformation and fake news, as it has suddenly become capable of significantly effecting the health outcomes of a nation. Exposure to fake news regarding the pandemic and vaccines among people with poor skills at detecting misinformation unsurprisingly lead to increased vaccine hesitancy [21] and subsequently has increased mortality during the COVID-19 pandemic [11].

In addition to purely health related negative impacts, fake news has also resulted in many negative societal outcomes. While the scapegoating of minority groups during a pandemic has no end of examples in human history, social media allows for fake news in this vein to travel much further and faster than it would have been possible before. Previous studies have shown a connection between the consumption of misinformation and the physical assault and harassment of people of Asian origin as well as healthcare professionals [26]. Fake News Detection is a very popular area of research. Recent literature focuses on deep learning and machine learning methods for the detection of fake news [18,20].

2.2 Fake News Spreaders

Current literature focuses on analysing the content of the posts to detect fake news [19,22,24,25]. In order to combat the spread of misinformation we must not only look into the content of the tweet, but also the users who are retweeting them. Identifying 'influential' users is a very popular research area. It has also been studied that 'influential' accounts on different social media platforms tend to have an impact in information propagation. Authors of [1] study that the influence of users on their friends can increase or decrease sales, so businesses are interested in finding influential people and encouraging them to create positive influence. They propose a method that uses interaction between social network users to detect the most influential among them. Literature also shows that user's in general contribute towards the spread of misinformation [5,16].

The reach of misinformation can vary significantly based on what user is spreading it, showing the importance of the users themselves in the spread of information on social media [3]. User profile analysis has become significantly important in identifying spreaders. Given this, it becomes critical to not only be able to detect fake news, but to also be able to identify super-spreader users.

Most of the existing literature focuses on profiling fake news spreaders by analyzing the content that they post [24,28,30]. Recent works have started focusing on analyzing the network structure and the neighbourhood of the social media users to profile spreaders. For instance, in [15], the author's proposed a novel machine learning based approach for automatic identification of the users spreading rumors based on the trust measures of users in Twitter's retweet network. Authors of [32] analyzed different features for such as bot usage, patterns and emotions in tweets posted by bots, heterogeneity among the spreaders, and geographic as well as demographic characteristics to profile fake news spreaders. In [9] the authors analyze how news spreads in social networks, simulating a simple information-spreading process in various network topologies and demonstrating that news spreads much more quickly in existing social-network topologies than in other network topologies. Another paper [34], addresses the complications as well as challenges encountered when measuring message propagation and social influence on OSNs. These works indicate that social-interaction among user's on the platform play a pivotal role in spreading information.

This paper introduces a data-driven metric called User's Impact Factor(UIF) which draws inspiration from [9,15] and combines two important features that profile a fake news spreader 1) their own probability of tweeting fake news 2) Retweeting fake news by being influenced by the people they follow who tweet high number of fake news tweets. This metric aims to show the importance of content analysis and neighbourhood influence, to help profile a fake news spreader.

2.3 Twitter

Among different social media platforms, Twitter was chosen for this study due to it's inherent structure. All activity on Twitter is publicly available and consists of four basic actions: A post, a retweet or quote tweet of a post, a reply to a post, and a like of a post. All activity of a user is publicly available on their account, and unlike other social media platforms like Facebook, on Twitter a user can only either be public or protected. A user cannot customize the level of privacy their account has, it is either entirely public, or private and viewable only by approved followers. This simple structure allows for much simpler data collection and analysis, as well as making the construction of a Twitter ecosystem much more feasible.

2.4 Influence/Trust Network

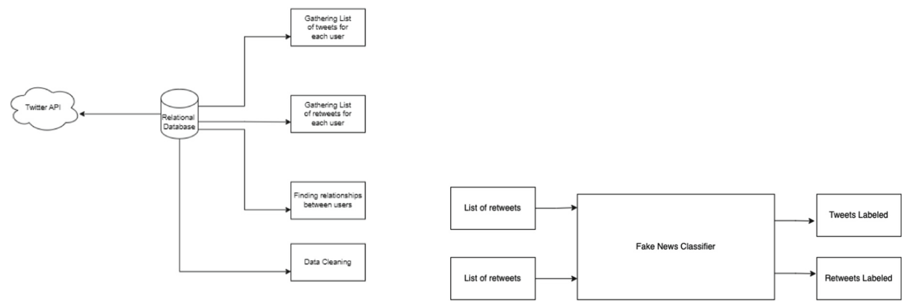
Trust networks comprise of influential accounts that have built trust with their audiences over time [8]. Studies have found that Trust networks play an important role in impacting the spread of information on social media [4]. Trust networks can be used to examine how a particular user spreading fake news may have a larger impact than another [27]. For example, during the COVID-19 pandemic, a famous Bollywood actor had tweeted about houseflies playing a role in the spread of the virus, resulting in a sharp spike in the number of web searches for the term 'houseflies' [10]. Users with a strong trust network have more impact in the diffusion of information. In this paper, a trust network is represented as a follower network which is built by scraping data from Twitter. Section 4.2 explains this process in detail.

2.5 BERT Transformer

The model chosen for fake news detection was the Bidirectional Encoder Representations from Transformers (BERT), which was developed by Google in 2018. While experiments were also conducted using bidirectional long short term memory networks and non-Deep Learning based classifiers, the BERT model was found to significantly out-preform these models. While previously, the best attempts to accomplish this classification have used bidirectional long short term memory networks [33], and the authors have had experience with the use of these models [29], the development of the BERT model provided a new tool for research. BERT-based fake news classification models provided the critical step forward in providing a robust, reliable model for fake news detection. Whether trained on detecting fake headlines or specifically attempting to identify fake news relating to the COVID-19 pandemic, the BERT models have consistently produced results at or above 90% accuracy in determining the trustworthiness of the news [14,17]. As a transformer based model, BERT consists of stacks of encoder and decoder layers. The model is capable of performing different Natural Language Processing (NLP) tasks, including being trained on a large corpus to classify misinformation tweets. One of the basic advantages that BERT provides compared to other models is bidirectionally. Instead of simply reading the tweet sequentially, from left to right in English, BERT reads the input from both directions to provide it with a better comprehension of the input.

3 Proposed Approach

The proposed approach in this paper is divided into the following four phases:



(a) Phase 1: Data Collection and Cleaning (b) Phase 2: Data Labelling

Fig. 1. Phases

- **Phase 1: Data scraping and collection from Twitter** - In this phase, a Twitter API is used to scrape and collect data from Twitte. Two lists of information is gathered, one list being a collection of tweets and retweets of a user and the other being a list of each user's followersr (Fig. 1a).

- **Phase 2: Building the fake news classifier** - In this phase a BERT Transformer is trained to label the tweets and retweets from phase 1, as fake news or real newsr (Fig. 1b).
- **Phase 3: Building the graph** - In this phase, a graph is built to represent a Twitter follower network based on the list of followers of each user obtained in Phase 1. This also represents a trust network among users. The follower network is represented as a graph $G(V, E)$ where V represents the users scraped from Twitter and E represents the set of edges depicting the followers of that user who retweet. So if node A follows node B and retweets a tweet authored B, then there is an edge between them. Section 4.2 describes this process in detail.
- **Phase 4: Formulating and Calculating UIF** - After the tweet and retweets are labelled in phase 2, the UIF is calculated for each user by traversing the graph built in Phase 3. Section 5 describes the formulation in detail.

4 Data Gathering and Pre-processing

4.1 Twitter Scraping

There are four primary ways to obtain Twitter data: Retrieve the data using the Twitter public Application Programming Interface (API), find an existing dataset, purchase access from Twitter, or purchase access from a Twitter service provider.

Twitter provides APIs in order to search and post tweets, get a list of users and their likes and see relationships between users. However, this public twitter API has a number of limitations. The standard API only allows retrieval of tweets up to 7 days old and is rate limited, only allowing the scraping of 18,000 tweets per a 15 min window. The dataset must contain relationships between accounts to identify the followers along with the list of tweets, retweets, and who they retweeted from.

Purchasing the datasets from either Twitter or a Twitter service provider like PowerTrack requires a monthly subscription and are mostly oriented to businesses. Other services like DiscoverText have access to old tweets that cannot be scraped using the public Twitter API, and the cost will depend on requirements like the number of tweets and the periods of time gathered from. Taking all of these methods in consideration the one that best fits the needs and resources of the project is the public twitter API.

The first step was retrieving the tweets and some of the followers of four seed accounts. Based on the importance of influential users described Sect. 2.2, the initial users were chosen because of their popularity and the role they play in politics. After getting seed nodes and their group of followers, it was identified which ones among them are common retweeters, i.e, they re-post many tweets from the account they follow. A list of tweets, retweets for each user, and who they follow were recorded. The process was then repeated on the followers of the seed accounts, creating a new level of users with data associated to their accounts who also had followers to be recursively scraped from. A total of 222 user's accounts were scraped and a combined total of 393 followers were recorded. The dataset collected consists of the user's id, the user handle, the text of their tweets, and a list of their followers.

It was important to identify the users relationships and the post that are involved in a retweet chain. Retweeting is the fastest way in twitter to spread a post and was used in order to measure the influence of a user on the network. The more a user's tweets tended to be retweeted by their followers, the more influence they would have on the nodes following them. The Retweet Probability(RTP) in Sect. 5 describes this in detail.

4.2 Building the Twitter Follower Network

In this paper, the follower network was built by scraping information from Twitter using the Twitter API. The resulting dataset is a list of users who are related by a follower relationship who retweet from the person they follow. In the network each user is a node and the relationships between two nodes is represented as edges. A user 'A' has an edge with user 'B' if 'B' follows 'A' and have retweeted a tweet whose author was 'A'. Figure 2 shows an example of this graph. The seed nodes were selected from a list of the influential people in politics. Section 2.2 cites literature that indicate that influential nodes tend to play a vital role in the spreading of fake news. Of the selected seed users a list of followers were obtained and for this list it was critical to identify which users are not only followers but also retweeters. Each follower who was also a retweeter, was added as a node to the graph by adding an edge between the two nodes. For the users in the rest of the levels the same classification was done. The retweets posted were also recorded for each user. This was needed to calculate the influence between the nodes.

The Twitter follower network was represented as a directed graph $G = (V, E)$, where V represents the users on Twitter and the edges represented users who follow other users and retweet their authored posts. In Fig. 2, the list of followers is represented as a Node −¿ List of Nodes where the List of Node represents the list of followers of the Node. For example, directed edges are built between Node A and Nodes B,C and D, because the latter follow Node A and retweet posts authored by Node A. Figure 2 elaborates the graph building process from the list of followers. A total of 222 users were collected with a combined total of 392 followers. Thus the follower network comprised of 222 nodes and 392 edges.

4.3 Fake News Data Labeling

While many tweets were collected as a part of this project, the fake news identifier was not trained off of any of the user tweets that were gathered. In part, this was due to the fact that there were simply not enough tweets in that pool to properly train a NLP model, however, that was not the only reason. It was important to the creation of the model that all data points within it be new to the BERT fake news classifier, in order to better represent the reality of its use. To have the fake news detection model assign probabilities of fake news generation on users

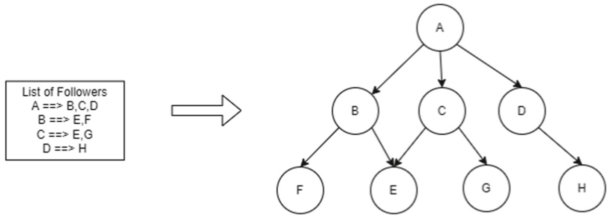


Fig. 2. Building the Twitter Follower Network from a List of Followers

who's tweets it was trained on would result in accuracy that could not properly be verified.

In order to train the fake news detection model, tweets from two major datasets were acquired. These two major datasets each focused on a different speciality, one consisting of true or false labels of political content, and the second focusing on COVID-19 related fake news. The political content dataset was aquired from the previously published FakeNewsNet which labeled its data with assistance from politifact [31]. The COVID-19 dataset was compiled and submitted at the CONSTRAINT 2021 workshop, and while the dataset included both English and Hindi language entries, only the English language tweets were used [23]. Together, the datasets accounted for 9615 unique items of true and fake news, off of which the model was trained on 90% of the dataset, with another 5% serving as a validation set and the last 5% serving as the testing set.

5 User's Impact Factor Formulation

In this paper, a new data-driven metric called User's Impact Factor (UIF) is proposed to identify potentially fake news spreaders. Current literature focuses on analyzing the content of the tweets that a user posts to classify them as spreaders. In this paper, UIF aims to highlight the importance of neighborhood influence in the dissemination of information in a network. It integrates the two characteristics of the user that form what content appears on their Twitter page: original content they tweeted, and content by others that they have chosen to retweet. By measuring both their activity of posting fake news content, and the fake news content of others they have retweeted, the total level of fake news spread by their account can be obtained, which is indicated by UIF. UIF helps in identifying nodes in the network that have a high impact in spreading fake news content. A high value of UIF indicates a high impact on the user in the propagation of fake news. Every user with a non-zero UIF value is considered as a potential fake news spreader. The value of UIF indicates the degree of their potential, instead of juts categorising users on a binary scale of 'spreaders' or 'non-spreaders'. One major significance of this metric is identifying fake news spreaders who do not post fake news content themselves but tend to retweet a lot of fake news due to influence from people they follow, thus highlighting the importance of influential neighborhood of users. UIF is formulated in the following steps:

1. **Fake Tweet Probability(FTP)** calculates the user's probability of posting fake news content. It is defined as the ratio of the number of tweets a user posts that are fake to the total number of posts that the user makes.

$$\text{FTP(U)\&} = \frac{\text{—Fake\_News\_Tweets(U)}}{\text{—Tweets(U)—}} \tag{1}$$

   where U represents a user on the network. A high value of FTP(U) indicates that the user U has a high potential of posting fake news content themselves.
2. **Retweet Probability(RTP)** calculates the user's probability of re-posting fake news content posted originally by people they follow.

$$\text{RTP(U,V)\&} = \frac{\text{—Retweets(U,V)—}}{\text{—Retweets(U)—}} \tag{2}$$

   where V represent a neighbour of U, i.e, U follows V. $|Retweets(U,V)|$ is the count of retweets made by U whose author was V and $|Retweets(U)|$ is the total number of retweets made by U. A high value of RTP(U,V) indicates V has a high influence on U.
3. **Fake Retweet Influence(FRI)** is a value assigned to a user that combines the RTP between them and all the people they follow on the network and the FTP of the people they follow.

$$\text{FRI(U)} = (\frac{\sum_{V \in Neighbors(U)} RTP(U,V) * FTP(V)}{\text{—Neighbors(U)—}}) \tag{3}$$

4. **User's Impact Factor (UIF)** is calculated as the sum of two metrics of each user 1) the FTP of the user (their own probability of tweeting fake news) and FRI (the retweets that they make as a result of following users who have a high probability of tweeting fake news.

$$\text{UIF(U)} = (\frac{FRI(U) + FTP(U)}{2}) \tag{4}$$

6 Results

6.1 BERT Classifier

The BERT Classifier was selected due to the reasons mentioned in Sect. 2.5. The performance of the BERT model can be seen in Fig. 3 which shows the final confusion matrix generated by the model.

With an accuracy of **96.26%**, the BERT model was able to provide the highly reliable predictions it was reputed to have. The weights for the best performing version of the model were saved, and could be then loaded in whenever predictions would need to be made on the twitter users.

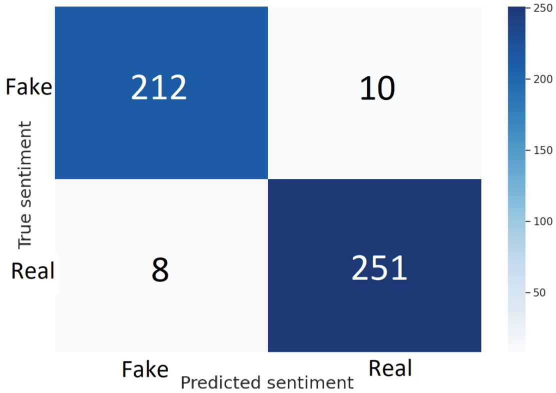


Fig. 3. Fake News Confusion Matrix

6.2 User's Impact Factor

For each user in the graph built in Sect. 4.2, their FTP, RTP and FRI were calculated to find their UIF by following the formulation mentioned in Sect. 5. Figure 4 shows the FTP and the UIF of 10% of users in the dataset. FTP indicates the probability of the user posting false tweets themselves and UIF is the probability that incorporates the neighbourhood influence of the user in retweeting. The comparison between these two metrics indicates the importance of influential neighborhood of a user.

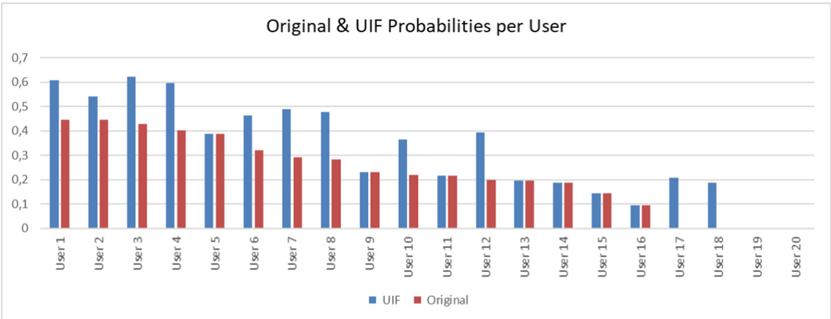


Fig. 4. Fake Tweet Probability (Original) VS User's Impact Factor

For privacy purposes, the names and handles of the users have been anonymized by replacing them with unique numbers. As seen in the Fig. 4, the UIF score analysis can be categorised as following

- **Case 1 (Increase in UIF from a zero FTP value):** This case indicates that the user originally did not post any fake news content by themselves, but were following and retweeting people who have a high tendency of posting fake news content. This increase in value indicated the user had a high impact in fake news propagation due to the high influence from their neighbors. In Fig. 4, users 17 and 18 are examples of this case.

- **Case 2 (Increase in UIF from a non-zero FTP value):** This case indicates that the user posts some fake news content themselves and additionally retweets fake news content posted by people they follow. Depending on the degree of fake news content posted by the user's neighbors, the increase would be small in some situations (User 2 in Fig. 4, and larger for others (User 1 in Fig. 4).
- **Case 3 (No change in UIF and FTP values):** This case indicates the user is not influenced by anybody they follow. However a non-zero value does indicate that the user still contributes to fake news propagation by posting fake news content themselves. Users 9, 11, 13 through 16 are some examples of this case.
- **Case 4 (Both UIF and FTP values are zero):** This case indicates that the user does not post any fake news content themselves and also does not retweet any fake news content from people they follow. User 19 and 20 is an example of this case.

The objective of the UIF metric was to highlight the impact of influential users on the fake news spreading potential of their followers, such as in User 4. Originally they had a FTP of 0.4, indicating that 40% of the tweets they made were classified as fake. However, User 4 followed popular nodes on the network and would retweet a high number of posts that were authored by the influential node. This influential node also had a high probability of making fake news posts. Thus, UIF(User 4) increased to a value of 0.6, indicating a high impact in spreading fake news. Also the aim of UIF is to weed out users who have an FTP value of '0' (i.e., they do not tweet any fake news content) however, they have a non-zero UIF value, indicating that they retweet fake news significantly. Among the 222 users in the dataset, the following table shows the categorisation of people into the different cases as mention above.

Table 1. UIF Score Categorization of Users

UIF Analysis	#Users
Case 1: FTP = 0, UIF != 0	87
Case 2: FTP != 0, UIF > FTP	32
Case 3: FTP = UIF	24
Case 4: FTP = 0, UIF = 0	79

From the Table 1, it is observed that 65% of users are categorised as potential fake news spreaders and 35% of users are classified as 'non-spreaders'. They are also not classified on a binary scale of 'spreader' and 'non-spreader', instead the

UIF value indicates the impact of the user in spreading fake news on a scale of 0 to 1. Among the 65%, almost 40% of the users were those who did not tweet any fake news content but would follow people who had a high influence on them and would also post fake tweets that user would be influenced to retweet. Without the UIF metric, chances are that the 40% of the users would not be profiled as fake news spreaders. The UIF metric weeds out these user's who would have otherwise gone unnoticed. 15% of users have an increase in the value from FTP to UIF. In addition to posting fake news content themselves, these users are also influenced by people who tweet fake news that these users retweet. 10% of users were those who were not influenced by their neighbors, but still posted fake news content themselves.

7 Conclusion

In this paper, a new data-driven metric called 'User's Impact Factor (UIF)' was proposed to help identify fake news spreaders. This metric aimed to highlight the impact of influential users on the dissemination of misinformation by their followers. This metric also representing the categorisation of users on a scale of 0 to 1, instead of a binary scale of 'spreader' and 'non-spreader'.

A BERT Transformer was trained to classify fake news tweets with 96% accuracy. The results show that UIF helps in identifying users that have a high impact in spreading fake news. Experimentation was performed on real world data and 65% of the users in the dataset was found to be potential fake news spreaders based on the UIF score. This metric helped in weeding out nodes that appear 'non-threatening' as they do not post any fake news tweets themselves, but tend to be highly influenced by the people they follow. The UIF scores show that analyzing content of the tweets posted by a user isn't enough and neighbourhood influence needs to be incorporated when analyzing patterns of fake news dissemination.

8 Future Steps

For future steps, many of the project phases can be explored further. For phase 1, more data is intended to be collected to expand the Twitter network of users and their followers. In phase 2, more state-of-the-art fake news detection models will be explored. For phase 4, variations of UIF like time-based or content-based UIF will be explored. Content-based UIF would analyse the change in retweet probability of a user based on the topic of the tweet. Other social media platforms like Reddit and Facebook are also intended to be explored. Another future direction is to detect and model the spread of different types of malicious content like hate speech, click-bait, scam links etc.

Acknowledgement. Many Thanks are due to the department of Mathematics & Computer Science (MCS) of the College of Arts & Sciences (CAS), the department of Computer Science & Engineering (CSEN) and the Frugal Innovation Hub (FIH) of the

School of Engineering (SoE), and the department of Information Systems & Analytics (ISA) of the School of Business at Santa Clara University in California, USA; as well as, the Computer Engineering department of the school of Engineering at Universidad Pontificia Bolivariana in Medellín, Colombia for their continued support of the project. Also thanks to Mubashir Hussain another member of the research team who worked on UIF graphs.

References

1. Afrasiabi Rad, A., Benyoucef, M.: Towards detecting influential users in social networks. In: Babin, G., Stanoevska-Slabeva, K., Kropf, P. (eds.) MCETECH 2011. LNBIP, vol. 78, pp. 227–240. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20862-1_16
2. de Barcelos, T., Muniz, L.N., Dantas, D.M., Cotrim Junior, D.F., Cavalcante, J.R., Faerstein, E.: Analysis of fake news disseminated during the covid-19 pandemic in brazilanálisis de las noticias falsas divulgadas durante la pandemia de covid-19 en brasil. Revista Panamericana de Salud Publica= Pan Am. J. Public Health **45**, e65–e65 (2021)
3. Bodaghi, A., Oliveira, J.: The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. Expert Syst. Appl. **189**, 116110 (2022)
4. Buskens, V.: Spreading information and developing trust in social networks to accelerate diffusion of innovations. Trends Food Sci. Technol. **106**, 485–488 (2020)
5. Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G., Previti, M.: Mutual influence of users credibility and news spreading in online social networks. Future Internet **13**(5), 107 (2021)
6. Center PR: News use across social media platforms in 2020 (2021). https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/
7. Courier-Tribune: Religious zeal drives N.C. man in ‘pizzagate’ (2016). https://www.courier-tribune.com/story/news/state/2016/12/06/religious-zeal-drives-nc-man-in-8216pizzagate8217/24407347007/
8. Data, Society: Why influence matters in the spread of misinformation (2018). https://points.datasociety.net/why-influence-matters-in-the-spread-of-misinformation-fc99ee69040e
9. Doerr, B., Fouz, M., Friedrich, T.: Why rumors spread so quickly in social networks. Commun. ACM **55**(6), 70–75 (2012). https://doi.org/10.1145/2184319.2184338
10. First Draft: Online influencers have become powerful vectors in promoting false information and conspiracy theories (2020). https://firstdraftnews.org/articles/influencers-vectors-misinformation/
11. Dyer, O.: Covid-19: Moderna and Pfizer vaccines prevent infections as well as symptoms, CDC study finds. BMJ: Br. Med. J. (Online) **373** (2021)
12. Fernández, J.L., Ramírez, J.A.L.: Approaches to the profiling fake news spreaders on Twitter task in English and Spanish. In: CLEF (Working Notes) (2020)
13. Flick, D.: Combatting fake news: alternatives to limiting social media misinformation and rehabilitating quality journalism. SMU Sci. Tech. L. Rev. **20**, 375 (2017)
14. Heidari, M., et al.: BERT model for fake news detection based on social bot activities in the Covid-19 pandemic. In: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0103–0109. IEEE (2021)

15. Huang, Q., Zhou, C., Wu, J., Wang, M., Wang, B.: Deep structure learning for rumor detection on Twitter. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2019)
16. Imaduwage, S., Kumara, P., Samaraweera, W.: Importance of user representation in propagation network-based fake news detection: a critical review and potential improvements. In: 2022 2nd International Conference on Advanced Research in Computing (ICARC), pp. 90–95. IEEE (2022)
17. Kula, S., Choraś, M., Kozik, R.: Application of the BERT-based architecture in fake news detection. In: Herrero, Á., Cambra, C., Urda, D., Sedano, J., Quintián, H., Corchado, E. (eds.) CISIS 2019. AISC, vol. 1267, pp. 239–249. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-57805-3_23
18. Liu, Y., Wu, Y.F.B.: FNED: a deep network for fake news early detection on social media. ACM Trans. Inf. Syst (TOIS) **38**(3), 1–33 (2020)
19. Luo, M., Hancock, J.T., Markowitz, D.M.: Credibility perceptions and detection accuracy of fake news headlines on social media: effects of truth-bias and endorsement cues. Commun. Res. **49**(2), 171–195 (2022)
20. Mishra, S., Shukla, P., Agarwal, R.: Analyzing machine learning enabled fake news detection techniques for diversified datasets. Wirel. Commun. Mob. Comput. **2022** (2022)
21. Montagni, I., et al.: Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy. J. Public Health **43**(4), 695–702 (2021)
22. Nassif, A.B., Elnagar, A., Elgendy, O., Afadar, Y.: Arabic fake news detection based on deep contextualized embedding models. Neural Comput. Appl. **34**, 16019–16032 (2022). https://doi.org/10.1007/s00521-022-07206-4
23. Patwa, P., et al.: Overview of CONSTRAINT 2021 shared tasks: detecting English COVID-19 fake news and Hindi hostile posts. In: Chakraborty, T., Shu, K., Bernard, H.R., Liu, H., Akhtar, M.S. (eds.) CONSTRAINT 2021. CCIS, vol. 1402, pp. 42–53. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-73696-5_5
24. Pizarro, J.: Using N-grams to detect fake news spreaders on Twitter. In: CLEF (Working Notes) (2020)
25. Raza, S., Ding, C.: Fake news detection based on news content and social contexts: a transformer-based approach. Int. J. Data Sci. Anal. **13**, 335–362 (2022)
26. Rocha, Y.M., et al.: The impact of fake news on social media and its influence on health during the Covid-19 pandemic: a systematic review. J. Public Health 1–10 (2021)
27. Ruan, Y., Durresi, A., Alfantoukh, L.: Using Twitter trust network for stock market analysis. Knowl.-Based Syst. **145**, 207–218 (2018)
28. Saeed, U., Fahim, H., Shirazi, F.: Profiling fake news spreaders on Twitter. In: CLEF (Working notes) (2020)
29. Shaghaghi, N., Calle, A.M., Manuel Zuluaga Fernandez, J., Hussain, M., Kamdar, Y., Ghosh, S.: Twitter sentiment analysis and political approval ratings for situational awareness. In: 2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), pp. 59–65 (2021). https://doi.org/10.1109/CogSIMA51574.2021.9475935
30. Shahid, W., Li, Y., Staples, D., Amin, G., Hakak, S., Ghorbani, A.: Are you a cyborg, bot or human?—a survey on detecting fake news spreaders. IEEE Access **10**, 27069–27083 (2022)
31. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: FakeNewsNet: a data repository with news content, social context and spatialtemporal information for studying fake news on social media. arXiv preprint arXiv:1809.01286 (2018)

32. Singh, M., Kaur, R., Iyengar, S.R.S.: Multidimensional analysis of fake news spreaders on Twitter. In: Chellappan, S., Choo, K.-K.R., Phan, N.H. (eds.) CSoNet 2020. LNCS, vol. 12575, pp. 354–365. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66046-8_29
33. Wang, W.Y.: "Liar, liar pants on fire": a new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017)
34. Ye, S., Wu, S.F.: Measuring message propagation and social influence on Twitter.com. In: Bolc, L., Makowski, M., Wierzbicki, A. (eds.) SocInfo 2010. LNCS, vol. 6430, pp. 216–231. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16567-2_16