2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Are You Influenced? Modeling the Diffusion of Fake News in Social Media

Abishai Joy
Computer Science Department
Boise State University (USA)
abishaijoy@u.boisestate.edu

Anu Shrestha
Computer Science Department
Boise State University (USA)
anushrestha@u.boisestate.edu

Francesca Spezzano
Computer Science Department
Boise State University (USA)
francescaspezzano@boisestate.edu

***Abstract*—We propose an approach inspired by the diffusion of innovations theory to model and characterize fake news sharing in social media through the lens of the different levels of influential factors (users, networks, and news). We address the problem of predicting fake news sharing as a classification task and demonstrate the potentials of the proposed features by achieving an AUROC of 0.97 and an average precision of 0.88, consistently outperforming baseline models with a higher margin (about 30% of AUROC). Also, we show that news-based features are the most effective at predicting real and fake news sharing, followed by the user- and network-based features.**

***Index Terms*—Misinformation, fake news sharing, information diffusion, diffusion of innovations theory**

I. Introduction

Social media have been increasingly used as a go-to resource for any information and daily news diet. The popularity of such platforms has dramatically transformed the news ecosystem and information flow in the past decade. Users in social media can easily access any kind of information, and further spread them intentionally or unintentionally through their social activities like tweets/retweets on Twitter without any friction. Consequently, making social media users equally responsible for the surge of fake news spread. Moreover, malicious individuals are capitalizing on such platforms to create misinformation, spread to a wider audience, and influence public opinion on important topics through information diffusion. Therefore, understanding the motivating factors that influence users' decision to share is important to understand the information diffusion phenomenon in social media.

Classical models for information diffusion, such as the Independent Cascade and Linear Threshold models, assume that a user will share the news with some probability only according to the fact that some of their friends have previously shared the same news [1]. However, recent works on fake news sharing in the social science domain have shown that a user decision of sharing or not sharing a piece of given news does not only depend on the influence of their friends but also on specific characteristics of the users (e.g., demographics, profile properties, behavior, and activity, etc.), the news received (e.g., title and content characteristics, etc.), and the social context (e.g., number of followers and following, tie strength, etc.) [2]. All these aspects align with what is theorized by the diffusion of innovations theory to explain how an innovation (which in our case is news) diffuses in a social network [3].

Thus, in this paper, we propose an approach based on the diffusion of innovations theory to model and characterize how fake news is shared in social media. Specifically, we address the following problem: *given that a user u is influenced on some given (real or fake) news item n by at least one of their followees v (i.e., u is following v and v has shared some news item n among their followers), predict whether the user u will also share news item n among their followers.*

We model the problem as a binary classification task and propose a set of features that takes into account user, news, and social network characteristics to better predict real and fake news sharing in social media. Our user-based features consider demographics, profile information, personality, emotions, user interest, and behavior, while our news-based features encode style, complexity, and psychological aspects of news headline and body. Network-based features consider, instead, the user following network to measure tie strength and quantify opinion leadership. All these factors have never been combined into a unique predictive model or tested on a large scale before. We tested our approach on a Twitter dataset of 1,557 users built upon the FakeNewsNet[1] data and containing 7,661 user-news item sharing and not sharing instances of 169 fake news items. Our results show that our proposed set of features outperform the results of both baseline approaches, i.e., independent cascade and linear threshold models. Specifically, we show that our proposed features can predict fake news sharing with an AUROC of 97.34 and an average precision of 88.43 (vs. an AUROC of 67.70 and an average precision of 87.45 achieved by the best baseline). Among the proposed features, we observed that news-based features are more effective in predicting fake news sharing, followed by the user-based and network-based features.

ASONAM '21, November 8–11, 2021, Virtual Event, Netherlands

ACM ISBN 978-1-4503-9128-3/21/11...$15.00
https://doi.org/10.1145/3487351.3488345

II. Related Work

News sharing has been studied both in computer science and social science. Vosoughi et al. [4] revealed that the fake news spreaders had, on average, significantly fewer followers, followed significantly fewer people, and were significantly less active on Twitter. Moreover, the dissemination of fake news on Twitter is mainly caused by human (and not bot) activity. Shu et al. [5] found that, on average, users who share fake news tend to be registered for a shorter time than the ones who share real news and that bots are more likely to post a piece of fake news than a real one, even though users who spread fake news are still more likely to be humans than bots. They also show that older people and females are more likely to spread fake news. Guess et al. [6] found out that political orientation, age, and social media usage to be relevant predictors of fake news sharing. Specifically, people are more likely to share articles they agree with, seniors tend to share more fake news probably because they lack digital media literacy skills, and the more people are familiar with the platform sharing features, the less they are likely to share fake news.

The author profiling shared task at PAN 2020 focused on determining whether or not the author of a Twitter feed is keen to spread fake news [7]. Participants proposed different linguistic features to address the problem, including (i) n-grams, (ii) style, (iii) personality and emotions, and iv) embeddings. Giachanou et al. [8] proposed an approach based on a convolutional neural network to process the user Twitter feed in combination with features representing user personality traits and linguistic patterns used in their tweets to address the problem of discriminating between fake news spreaders and fact-checkers.

News spread in social media is usually modeled by considering the underlined network among its users. Existing models differ according to network observability: Independent Cascade and Linear Threshold models assume the user connections to be explicit, while epidemic models or Hawkes processes work with an implicit network to predict the number of infected people over time [9]–[11]. Some works went beyond considering just the network to explain news diffusion and tested hypotheses inspired by the diffusion of innovations theory, which also considers user and news characteristics as important factors to explain news sharing behavior [12]. Ma et al. [13] found opinion leadership, news preference, and tie strength to be the most important factors at predicting news sharing, while homophily hampered news sharing in users' local networks. Also, people driven by gratifications of information seeking, socializing, and status-seeking were more likely to share news on social media platforms [14].

III. Dataset

We used the PolitiFact dataset from the FakeNewsNet data repository[1] to carry out our experiments.

The dataset contains details about 992 news articles content, social engagement, and user network. News ground truth labels (fake or not) have been collected from the fact-checking website `politifact.com`. For the purpose of testing our proposed method to predict news sharing by influenced users, we computed the labels for user sharing or not sharing a given piece of news, as explained here below. First, we determined the influencer and influenced user pairs. An influencer is a user who tweeted a given piece of news, and an influenced user is a follower of that influencer. We selected only those users (influencers) who have shared at least one news and have at least one follower (influenced user). Next, for each follower, we considered at least five instances of news they shared and ordered them in chronological order of shared time. Among them, we used two most recently shared news to annotate a news item as **Shared** if the piece of news that is shared/tweeted by a user (influencer) is then retweeted by their follower (influenced user), and **Not Shared** if the piece of news that is shared/tweeted by a user (influencer) is not retweeted by their follower. The remaining three or more news shared by followers were concatenated to compute features to analyze user's interest similarity with news, as discussed in IV-A5. In addition, for each follower, we crawled all tweets posted one month prior to the publish date of the shared news to compute remaining user-based features. In this paper, we considered only the sharing instances of fake news articles. The size of this dataset is shown in Table I.

TABLE I
Size of the dataset used in our experiments.

# Users	# Fake News	# Shared	# Not Shared
1,557	169	527	7,134

IV. Features

This section describes the features we used to implement the diffusion of innovations theory for modeling news sharing. Specifically, we leveraged three different categories of features, namely user-based, network-based, and news-based features.

A. User-based Features

1) Demographics: Those features are often not explicitly available on social media platforms. Therefore, we utilized m3inference [15], a deep-learning-based system trained on Twitter data, to infer user demographic characteristics. This tool predicts the *gender* of the user as male or female, *age* of the user grouped in four categories ($\leq$18, 19$-$29, 30$-$39 and $\geq$40) and whether the given account is handled by an *organization* or not. For our analysis, we utilized only age and gender. We used the #polar score [16] to compute the political ideology of each user from the hashtags in their tweets. The model is trained on a dataset of tweets by U.S. Congress members provided by Chamberlain et al. [17].

2) Explicit Features and Activity: We consider available user profile information as explicit features such as: Protected (whether a user has chosen to protect their tweets or not), Verified (whether a user is a verified user), Register Time (the

[1]https://github.com/KaiDMML/FakeNewsNet

number of days passed since the registration of the account), Status count (the number of tweets and retweets by a user), and Favor count (the number of tweets a user has liked).

In addition, we analyzed the user tweeting behavior within the day (24 hours) by computing the user *insomnia index*, i.e., the difference between the number of night and day posts upon the total number of user posts.

3) Personality: To compute personality features, we leveraged IBM Watson Personality Insights service that uses linguistic analytics to infer individuals' intrinsic personality characteristics from digital communications such as social media posts, and includes Big Five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism), Needs (excitement, harmony, curiosity, ideal, closeness, self-expression, liberty, love, practicality, stability, challenge, and structure), and Values (elf-transcendence, conservation, hedonism, self-enhancement, and openness to change). For this, we concatenated all the user tweets in a unique document to compute their personality characteristics.

4) Emotion: We capture the emotion of users from their tweets. To compute these features, we concatenated all the tweets by each user to form a single document per user. To determine the intensity of emotions such as anger, joy, sadness, fear, disgust, anticipation, surprise, and trust, we leveraged the Emotion Intensity Lexicon (NRC-EIL) [18]. Next, after removing all stopwords, each lemmatized word in the text is looked up in the emotion intensity dictionary, and intensity scores of matching words are averaged element-wise to generate an emotion vector representation of the text. Along with these emotions, we also consider a stress feature computed using the lexical dictionary created by Wang et al. [19] for LIWC. Moreover, we used VADER [20] for sentiment analysis. We measured the average sentiment (positive, negative, and neutral) across all their tweets for each user.

5) User Interest in News: To compute this feature, we trained an LDA model [21] with 100 topics on Wikipedia data to infer topics of the text in our dataset. In particular, we utilized this model for retrieving topical similarity between user's interest and shared news item by using the following two approaches: *Similarity 1*, i.e., the cosine similarity between the topics extracted from the news item to be shared and the concatenation of the influenced user's previously shared three or more news documents and *Similarity 2*, i.e., the cosine similarity between the topics extracted from the news item to share and timeline tweet document formed by concatenating all tweets from the timeline of the influenced user.

B. Network-based Features

1) User Centrality: In this category of features, we considered centrality measures such as degree centrality (both in-degree and out-degree) and PageRank as well as the Twitter follower to following (TFF) ratio as in [5], which is computed as $TFF = \frac{\#Follower+1}{\#Followee+1}$ and indicates the ratio of the number of followers to the number of followees of the user. The greater the ratio, the higher the popularity of the user.

2) Weak and Strong ties: Tie strength in online social networks is positively associated with news sharing intention in social media [13], [22]. We have used the following two approaches to determine tie strength between the influencer and influenced user pairs: (i) *receiver's perspective*, where for a given influencer-influenced user pair in the dataset, we computed, out of all the retweets by the influenced user what percentage of them were tweeted by the given influencer, and (ii) *time-based tie strength*, i.e., the average time taken by the influenced user to share/retweet news tweeted by the given influencer.

C. News-based Features

As news-based features, we considered the ones proposed in our previous work [23] to detect fake news on the same PolitiFact dataset used in this paper. These features include stylistic, psychological, and complexity features that are computed for both title and body text of news items.

1) Stylistic Features: This set of features captures the user's writing style. We used the subset of LIWC features that represent the functionality of text, including word count (WC), words per sentence (WPS), number of personal (I, we, you, she/he – one feature each) and impersonal pronouns, number of exclamation marks (exlam), number of punctuation symbols (allPunc), number of quotes (quote). Also, we considered part of speech (POS) features computed by the Python Natural Language Toolkit part of speech tagger.

2) Psychology Features: We computed the positive (pos) and negative (neg) sentiment metrics using the LIWC tool. In addition, we calculated emotion features, such as anger, joy, sadness, fear, disgust, anticipation, surprise, and trust by using the Emotion Intensity Lexicon (NRC-EIL) [18].

3) Complexity Features: The complexity of text in natural language processing depends on how easily the reader can read and understand a text. We used the Simple Measure of Gobbledygook Index (SMOG) readability measure as a complexity feature in our analysis. Higher scores of readability indicate that the text is easier to read. This group of features also includes lexical diversity or Type-Token Ratio (TTR) and the average length of each word (avg wlen).

V. Experiments

We addressed the problem of automatically identifying whether a user will share a news item or not as a binary classification task. Specifically, we compared various machine learning algorithms, namely Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, and Extra Trees Classifier. We used the features described in Section IV as input to these algorithms. We used class weighting to deal with the class imbalance and performed 10-fold cross-validation. As evaluation metrics, we considered the Area Under the ROC Curve (AUROC) and Average Precision (AvgP), which are well-suited for unbalanced data.

As our model is predicting which users will become infected (as opposed to the number of infected users), we compared with two well-known information diffusion models, namely

TABLE II
PERFORMANCE OF OUR PROPOSED FEATURES ACCORDING TO DIFFERENT CLASSIFIERS AND COMPARISON WITH BASELINES. BEST VALUES ARE IN BOLD.

		AUROC	AvgP.
Our approach	Logistic Regression)	93.78	65.22
	SVM	89.08	36.58
	Random Forest	**97.86**	85.11
	Extra Trees	96.82	76.75
	XGBoost	97.34	**88.43**
Baselines	ICM	67.67	56.72
	LTM	63.78	87.45

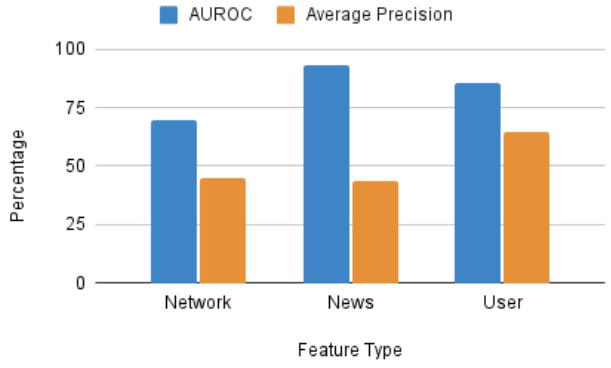

Fig. 1. AUROC and average precision per feature group.

the Independent Cascade Model (ICM) and the Linear Threshold Model (LTM) [9], as they work with the explicit network. Parameters are set according to the heuristics proposed by Goyal et al. [24] for the ICM and by Li et al. [25] for LTM.

The results are reported in Table II according to AUROC and average precision. As we can see, XGBoost classifier comparatively outperformed other classifiers that we considered and achieved the best results with an AUROC of 97.34 and average precision of 88.43. Moreover, our model with proposed features consistently outperformed both baseline models with a higher margin (30% of AUROC, approximately).

In addition, we measured the performances of each group of features (network-based, news-based, and user-based) using the best classifier (i.e., XGBoost) to measure their contribution to the fake news sharing model. As shown in Figure 1, we observed that a significant amount of contribution in decision-making is from news-based features, followed by user-based features and network-based features according to AUROC; user-based features are the most important group of features followed by network-based features and news-based features according to average precision. Thus, in general, there are other features that are more important than network features when predicting fake news sharing.

VI. CONCLUSION

In this paper, we addressed the problem of modeling fake news sharing in social media. We considered the problem of predicting whether a user will share a fake news item among their followers, given that the news item was shared by one of their followees (the user was influenced by that given news item). For this, we proposed three main sets of features, namely news-based features computed from news headline and body, user-based features computed from user Twitter feed and profile, and network-based features computed from the user following network, and performed a comprehensive analysis on a Twitter dataset with ground truth created as news *Shared* and *Not Shared*. Our experiments showed the potential of the proposed features in predicting fake news sharing, which achieves an AUROC of 97.34 and average precision of 88.43 and outperforms the considered baseline approaches. Further, our analysis revealed that other features beyond classical network-related features need to be considered to effectively model fake news sharing.

ACKNOWLEDGEMENTS

This work has been supported by the National Science Foundation under award no. 1943370. We thank Hongmin (Steven) Kim for collecting additional Twitter data.

REFERENCES

[1] P. Shakarian, A. Bhatnagar, A. Aleali, E. Shaabani, R. Guo *et al.*, *Diffusion in social networks*. Springer, 2015.
[2] A. S. Kümpel, V. Karnowski, and T. Keyling, "News sharing in social media: A review of current research on news sharing users, content, and networks," *Social media+ society*, vol. 1, no. 2, p. 2056305115610141, 2015.
[3] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.
[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
[5] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *1st IEEE International Workshop on Fake MultiMedia (FakeMM 2018)*, 2018.
[6] A. Guess, J. Nagler, and J. Tucker, "Less than you think: Prevalence and predictors of fake news dissemination on facebook," *Science advances*, vol. 5, no. 1, p. eaau4586, 2019.
[7] F. Rangel, A. Giachanou, B. Ghanem, and P. Rosso, "Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on twitter," in *CLEF*, 2020.
[8] A. Giachanou, E. A. Rissola, B. Ghanem, F. Crestani, and P. Rosso, "The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers," in *NLDB 2020*, p. 181.
[9] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *SIGKDD*, 2003, pp. 137–146.
[10] M. Maleki, E. Mead, M. Arani, and N. Agarwal, "Using an epidemiological model to study the spread of misinformation during the black lives matter movement," *arXiv preprint arXiv:2103.12191*, 2021.
[11] T. Murayama, S. Wakamiya, E. Aramaki, and R. Kobayashi, "Modeling the spread of fake news on twitter," *Plos one*, vol. 16, no. 4.
[12] R. Zafarani, M. A. Abbasi, and H. Liu, *Social media mining: an introduction*. Cambridge University Press, 2014.
[13] L. Ma, C. S. Lee, and D. H. Goh, "Understanding news sharing in social media from the diffusion of innovations perspective," in *GREENCOM-ITHINGS-CPSCOM*. IEEE, 2013, pp. 1013–1020.
[14] C. S. Lee and L. Ma, "News sharing in social media: The effect of gratifications and prior experience," *Computers in human behavior*, vol. 28, no. 2, pp. 331–339, 2012.
[15] Z. Wang, S. Hale, D. I. Adelani, P. Grabowicz, T. Hartman, F. Flöck, and D. Jurgens, "Demographic inference and representative population estimates from multilingual social media data," in *The World Wide Web Conference*, 2019, pp. 2056–2067.
[16] L. Hemphill, A. Culotta, and M. Heston, "Polar scores: Measuring partisanship using social media content," *Journal of Information Technology & Politics*, vol. 13, no. 4, pp. 365–377, 2016.

[17] J. M. Chamberlain, F. Spezzano, J. J. Kettler, and B. Dit, "A network analysis of twitter interactions by members of the us congress," *ACM Transactions on Social Computing*, vol. 4, no. 1, pp. 1–22, 2021.

[18] S. M. Mohammad, "Word affect intensities," *arXiv preprint arXiv:1704.08798*, 2017.

[19] W. Wang, I. Hernandez, D. A. Newman, J. He, and J. Bian, "Twitter analysis: Studying us weekly trends in work stress and emotion," *Applied Psychology*, vol. 65, no. 2, pp. 355–378, 2016.

[20] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *AAAI ICWSM*, 2014.

[21] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, 2010, pp. 45–50.

[22] O. Zorzi, "Granovetter (1983): The strength of weak ties: A network theory revisited," in *Schlüsselwerke der Netzwerkforschung*. Springer, 2019, pp. 243–246.

[23] A. Shrestha and F. Spezzano, "Textual characteristics of news title and body to detect fake news: A reproducibility study," in *ECIR, Proceedings, Part II*, ser. Lecture Notes in Computer Science, vol. 12657. Springer, 2021, pp. 120–133.

[24] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks," in *ACM WSDM*, 2010, pp. 241–250.

[25] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 10, pp. 1852–1872, 2018.