Catch Me If You Can: A Participant-Level Rumor Detection Framework via Deep Multi-View Learning

Xueqin Chen
School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China 610054 chenxueqin@std.uestc.edu.cn

Yi Yang
Department of Information Systems, Business Statistics and Operations Management (ISOM)
Hong Kong University of Science and Technology
Sai Kung, HK imyiyang@ust.hk

Kunpeng Zhang
Department of Decision Operations & Information Technologies
Robert H. Smith School of Business University of Maryland, College park
College park, MD 20742 kzhang@rhsmith.umd.edu

Fan Zhou
School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China 610054 fan.zhou@uestc.edu.cn

Electronic copy available at: https://ssrn.com/abstract=3623452

ABSTRACT

Researchers have exerted tremendous effort in designing ways to automatically detect and identify rumors. Traditional approaches focus on feature engineering. They require lots of human efforts and are difficult to generalize. Deep learning solutions come to help. However, they usually fail to capture the underlying structure of the rumor propagation and the influence of all participants involved in the spreading chain. In this study, we propose a novel deep multi-view learning-based rumor detection approach. It explicitly models and integrates the event content, user characteristics, social network structures, and temporal influence of all participants. Experiments conducted on two real-world datasets demonstrate a significant accuracy improvement of our approach, up to 22.1% higher than baselines. Theoretically, we contribute to the effective usage of data science and analytics for social information diffusion design, particularly rumor detection. Practically, our results can be used to improve the quality of rumor detection services for social platforms.

Keywords: Rumor detection, participant-level, deep multi-view learning, structure, temporal.

Electronic copy available at: https://ssrn.com/abstract=3623452

1. INTRODUCTION

The rapid development of Internet technology has democratized the exchange of information. Online social platforms, such as Twitter, Yelp, Reddit, etc., have emerged and gradually become the main source of information to guide individuals' everyday decisions (Shearer and Matsa 2018; Kumar et al. 2018). According to Pew Research[1], as of 2017 approximately 88% of American adults have either free or paid Internet access at home, and about 81% obtain news from online platforms (e.g., news websites/apps, social media, or both). While Internet technology and social media facilitate free information creation and sharing, the proliferation of fake news, rumors and false information has had strong and negative societal and economic consequences. The explosive spread of false information can pose a threat to the credibility of legitimate online platforms and resources and has a serious negative impact on both individuals and society (Shu et al. 2017), with the potential consequences to destabilize nations, affect the fairness of competition (Allcott and Gentzkow 2017), and shock the stock market (DiFonzo and Bordia 1997). For example, in the global effort to contain the COVID-19 pandemic, misinformation abounds and flourishes on the Internet, and people have been led to believe that COVD-19 can be cured by ingesting fish tank cleaning products or that 5G networks generate radiation that triggers the virus. Such misinformation not only causes panic among citizens but could potentially undercut collective efforts to control the pandemic.

Detecting fake news on online social platforms as early as possible is therefore a necessary, urgent and socially beneficial task. However, the definition of fake news is vague, and there is a widespread disagreement over what constitutes fake news. Simply labeling some news as fake news can itself be considered a propagandistic act that is not credible. It is also possible that what

[1] https://www.journalism.org/2016/07/07/pathways-to-news/

Electronic copy available at: https://ssrn.com/abstract=3623452

starts out as fake news may later be proven true when more evidence is available. Therefore, we first give a clear formal categorization of fake news, as defined in Table 1 and supported by the examples in Appendix. In this categorization system, our approach is more nuanced than most existing fake news detection algorithms, which focus on binary classification: fake vs. non-fake. To make our method more precise, we follow a previous study to divide news into four separate categories based on degree of authenticity which, as in some existing fact-checking systems, is categorized as either true rumor (mostly true), false rumor (false), unverified rumor (unproven) or non-rumor (true). Since this four-type categorization better captures the nuance and nature of news, rather than calling our approach "fake news detection" with its implied binary, we call it "rumor detection." An effective rumor detection method is expected to identify different types of rumors with high accuracy. This is necessary because, for example, poorly classifying non-rumor news as true rumor news and blocking it from a social media platform may backfire, as such an inappropriate action undermines the fundamental information-sharing purpose of the social media platform.

Table 1: Categorization of rumors.

category	**definition**
true rumor	The primary elements of news are demonstrably true.
false rumor	The primary elements of news are demonstrably false.
unverified rumor	Typical news for which there is little or no affirmative evidence. However, declaring it false would require proving a negative or accurately discerning an actor's thoughts and motivations.
non rumor	None of above.

To develop our approach, we considered the life cycle of real and fake news on social media, which plays a key role in information diffusion. When news is produced by the content creator, it starts its journey on the social media platform; people are exposed to the news, becoming content

Electronic copy available at: https://ssrn.com/abstract=3623452

consumers. According to the confirmation bias theory, people tend to favor, interpret and share information in a way that confirms or strengthens their prior beliefs or ideologies. As a result, if a news item confirms the consumer's prior ideology, they may share it within their social networks in the role of content distributor. Since fake news is intentionally written to mislead readers into believing and propagating false information (e.g., 5G networks trigger COVID-19), it is plausible that fake news is more easily distributed among its believers than real news, which is neutral in its beliefs and ideology. This idea is supported by prior studies, which have noted that false information tends to spread significantly faster, further, deeper and more broadly than real information (Vosoughi et al. 2018). Therefore, considering all participants, including content creators and content distributors, in the news diffusion chain may improve the overall rumor detection performance of our approach.

Another design consideration is the lack of effective methods to represent all participants in the news diffusion chain. Prior predictive studies have simply used aggregated statistics, such as the total number of content distributors (retweets) and the average time of information distribution, to quantify the diffusion process. This inevitably results in information loss and suboptimal performance. Other example of such aggregated statistics include network-level attributes (e.g., density) to represent diffusion networks, the final hidden representation from recurrent neural networks to model the temporal spreading sequence, and overall descriptive statistics of user characteristics (e.g., mean user tenure) to describe users in the diffusion (Castillo et al. 2011). While such data may be helpful in modeling, they are not quite specific enough to provide a clear picture of by whom, when, why and how news is diffused. Therefore, the key question motivating our study is how to design an effective predictive method that represents fine-grained all-participant patterns throughout the whole diffusion process, including

Electronic copy available at: https://ssrn.com/abstract=3623452

structural patterns (who distributes the news and how it is distributed in the social network), temporal patterns (how fast the news is distributed in the social network), all-participant profile patterns, and content patterns.

Our approach is rooted in predictive analytics, which can add theoretical and practical value to Information Systems (IS) research (Shmueli and Koppius 2011). From this foundation, following the guidelines of design science paradigm in IS research (Hevner et al. 2004; Peffers et al. 2007; Gregor and Hevner 2013; Grover et al. 2018) and prior literature (Abbasi et al. 2015; Guo et al. 2018; Kumar et al. 2019), we propose a novel framework based on deep multi-view learning for rumor detection, named DMV-RD (Deep Multi-View Rumor Detection). In view of theories on propagation and social influence, DMV-RD models rumor behaviors at a fine-grained participant level, meaning that it incorporates characteristics and propagation activities of all users who participate in the diffusion process of posting content (e.g., a tweet) in order to predict the content's credibility (e.g., classify it as one of the four rumor types). DMV-RD fully utilizes relevant information extracted from the diffusion process. This information it uses to detect rumors includes the diffusion structure of an implicit social network (i.e., spreading capability, connection degree, structural similarity and distance among nodes), diffusion temporality (i.e., the sequential time order of diffusion), user profiles (i.e., demographics of all participants involved in diffusion), and linguistic characteristics of the posted content.

In this way, DMV-RD models and captures these patterns, a task that is challenging for traditional methods. Yet it is also critical to learn the deep and latent inter- and intra-dependencies among these patterns. To do this, DMV-RD employs the power of deep representation learning to improve rumor detection performance: DMV-RD designs a diffusion graphical convolution network, a time-decay long short-term memory network, a user

Electronic copy available at: https://ssrn.com/abstract=3623452

characteristics fusion, and a text embedding to learn high-level representation of diffusion. Figure 1 illustrates the overall framework. User A initializes an event (e.g., a tweet). Users B, C, D, and E retweet the event sequentially after certain time periods. DMV-RD must then predict whether A's original tweet is a rumor, and if so, precisely what type of rumor. Based on the observed data, DMV-RD can build a retweet network, form a sequential path of diffusion (both on the right pane of Figure 1), and integrate characteristics of all participants thus far (i.e., users A-E). This information is all used to learn structural, temporal, and personal representations of diffusion. The learned representations along with the textual content of the tweet is further fed into a non-linear classifier for rumor prediction.

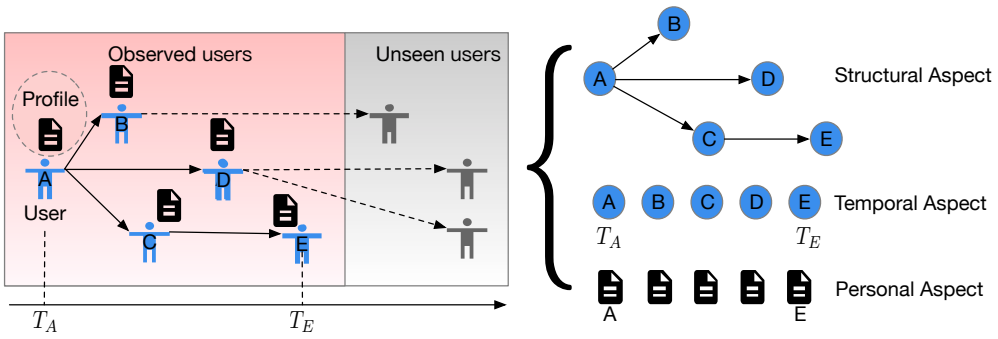


Figure 1: Illustration of DMV-RD framework.

Our design science work on rumor detection makes two main contributions to the literature in this field. First, our design is rooted in social influence and propagation theory, from which we derive various constructs in our model. To the best of our knowledge, our approach is among the first to detect rumors at a very fine-grained participant level. It is very different from prior works that also combine information from various sources (e.g., diffusion networks, diffusion sequence, and attributes of spreading users), but while previous studies integrate each aspect as one single representation, our approach makes multiple individual representations. Second, we make a methodological contribution by proposing a deep representation learning approach that

Electronic copy available at: https://ssrn.com/abstract=3623452

effectively captures all participant information in a diffusion chain. This information includes diffusion structural patterns, diffusion temporal patterns, all-participants profile patterns and news content patterns. Experimental results using real-world datasets confirm the effectiveness of our approach over prior rumor detection methods. Our approach has direct implications for social media platforms that are vulnerable to rumor spreading, since it can be deployed to identify original users who initiate rumors and those who spread rumors. Overall, the proposed rumor-detection algorithms can help improve the user experience, ultimately opening up more revenue-generating opportunities for digital platforms and benefiting society by helping individuals to obtain healthy and genuine information.

The rest of the paper is organized as follows. Section 2 reviews related work, then Section 3 introduces the theoretical foundations of our study, as well as the data and some preliminary background. Section 4 presents the major aspects of DMV-RD methodology in detail. Experiment evaluations quantifying the benefit of our approach are described in Section 5. Section 6 concludes the paper and outlines directions for future work.

2. LITERATURE REVIEW

We review prior predictive studies on rumor detection, which generally fall into two categories: those that use a feature engineering-based approach and those that use a deep learning-based approach.

2.1 Feature engineering in rumor detection

The IS community is now well accustomed to leveraging various social media features to improve performance of predictive analytics. For instance, researchers have extracted sentiments, emotions, topics, lexical characteristics, and social network features from social media data for

Electronic copy available at: https://ssrn.com/abstract=3623452

corporate fraud detection (Dong et al. 2018), and they have combined geographical and social influences for personalized point-of-interest recommendation (Guo et al. 2018). In the area of rumor detection, handcrafted feature engineering-based approaches also prevail and are used to extract various relevant features from raw data, features which typically fall into two major categories: (1) content features extracted from text (e.g., characters, words, sentences and documents) (Castillo et al. 2011; Jin et al. 2017; Kwon et al. 2013; Ma et al. 2017; Wu et al. 2015; Zhao et al. 2015) and (2) social context features (Castillo et al. 2011; Kwon et al. 2013; Liu et al. 2015; Yang et al. 2012) extracted from the user behavior and the diffusion network, which reflect the relationships among users and describe the diffusion process of a rumor, such as user demographics, propagation activities, and diffusion temporality. Both content features and social context features are fed into discriminative machine learning algorithms to perform rumor detection.

Rumors aim to arouse intense attention and stimulate the public mood, and their textual content often tends to have certain patterns in contrast to non-rumors. One study (Zhao et al. 2015) describes two types of language patterns (inquiry and correction patterns) in rumors, then detects the patterns of rumor messages through supervised feature selection on a set of labeled messages. Wu et al. (2015) define a set of topic features to summarize semantics and train a Latent Dirichlet Allocation (LDA) model for detecting rumors on Weibo. Moving towards a more comprehensive understanding of text on social media, other studies have derived non-general textual features, such as source links (Zhang et al. 2015) and emotions (Castillo et al. 2011), from social media platforms.

Social context features are derived from the social connection characteristics of social media users. Rumors are usually created by a few users but spread by many. User profiles can be used

Electronic copy available at: https://ssrn.com/abstract=3623452

to measure the user's characteristics and credibility (Castillo et al. 2011). Kwon et al. (2013) extended this work and further proposed 15 structural features extracted from the diffusion network as well as the user friendship network.

We note that the performance of the above approaches heavily depends on the handcrafted features, for which there is no standard and systematic design protocol. In fact, the conclusions of existing works usually contradict each other, largely due to the different types of datasets used to inform feature design. In our work, instead of manually creating features, we leverage deep learning to represent the fine-grained diffusion process.

2.2 Deep learning techniques

Due to the recent success of deep learning in different fields, such as NLP (Natural Language Processing) and CV (Computer Vision), much attention has been paid to developing a deep neural network-based method for rumor detection. Various studies have already shown that such methods offer a performance improvement due to their enhanced ability to extract relevant features and represent data. The first study to apply recurrent neural networks (RNN) to model rumors as varied length time series aimed to learn both temporal and textual features from raw data and thus detect rumors (Ma et al. 2016). Other researchers built a tree-structured RNN to catch hidden representations from both propagation structures and text content. For example, Yu et al. proposed a convolutional neural network-based approach to handle the problem occurring in RNN-based methods; this was done because RNN is not suitable to perform early detection tasks with limited inputting data, and it has a bias towards the latest elements of the input sequence (Yu et al. 2017). Simultaneously, plenty of methods use a combination of CNN and RNN. For example, Yang et al. built a time series classifier with both RNN and CNN to predict whether a given news story is fake at an early stage, which takes common users' characteristics

Electronic copy available at: https://ssrn.com/abstract=3623452

and propagation paths into consideration (Yang et al. 2019). Other studies use different approaches altogether. For instance, Bian et al. (2020) proposed a GCN-based model to learn global structural relationships of rumor dispersion (Bian et al. 2020). Ma et al., enlightened by the multi-task learning scheme, proposed two multi-task architectures based on RNNs, joint training the task of stance classification and rumor detection (Ma et al. 2018).

While the aforementioned models can achieve state-of-the-art performance and demonstrate that fusing different kinds of features can improve rumor detection performance, they still exhibit several drawbacks. Specifically, these models still heavily depend on linguistic-based content features, and they are inefficient in learning the propagation structure and user temporal influence. Furthermore, no existing method has modeled rumor detection at a fine-grained all-participant level, which motivates our present study.

We summarize prior work on analytics-driven rumor detection in Table 2.

Table 2: Summary of differences among extant works on rumor detection. Agg.: aggregate-level; Ind.: all participant-level.

Method	**Study**	**Information used**							
		User profile				**Propagation activity**			
		Text	**Source tweet user**	**All participants**		**Temporal**		**Structural**	
				Agg.	**Ind.**	**Agg.**	**Ind.**	**Agg.**	**Ind.**
Not deep learning-based	[Castillo et al., 2011]	✓	✓	✓					
	[Yang et al., 2012]	✓	✓	✓					
	[Kwon et al., 2013]	✓	✓			✓		✓	
	[Zhao et al., 2015]	✓							
	[Wu et al., 2015]	✓						✓	
	[Ma et al., 2015]	✓		✓		✓			
	[Jin et al., 2016]								
	[Ma et al., 2017]							✓	
Deep learning-based	[Ma et al., 2016]	✓							
	[Jin et al., 2017]	✓							
	[Ma et al., 2018]	✓						✓	

Electronic copy available at: https://ssrn.com/abstract=3623452

[Liu et al., 2018]			✓		✓			
[Ma et al., 2018]	✓				✓			
[Khattat et al., 2019]	✓							
Our study	✓	✓		✓		✓		✓

3. BIRDS OF A FEATHER FLOCK TOGETHER: THE PERSPECTIVE OF ALL PARTICIPANTS

Rumor detection has long been a subject of interdisciplinary research. Various theories have been proposed and validated. In this section, we discuss several major theories that can guide us to derive relevant constructs in our model for better rumor detection.

3.1 Theory

Users play major roles in the dissemination of rumor or fake news. A set of user-based and propagation-based theories were developed to study how rumor spreads, how users engage with rumor, and the role users play in rumor creation, propagation, or intervention. For example, in the echo chamber effect, individuals tend to believe information is correct after repeated exposures (Shore et al. 2018). Confirmation bias theory tells us that individuals tend to trust information that confirms their preexisting beliefs or hypotheses, which they perceive to surpass that of others (Nickerson 1998). People choose to interact with those who share similar opinions and avoid those with whom they profoundly disagree. Both indicate that people may react to and process information differently based on information type (e.g., rumor vs. non-rumor). On the other hand, homophily theory says that individuals in homophilic relationships share common characteristics (beliefs, demographics, etc.) that make communication easier (McPherson et al. 2001). Meanwhile, social identity theory shows that individuals do something primarily because others are doing it and to conform in order to be liked and accepted by others. Such social influence and homophilic atmospheres also exist in and are commonly seen in online social networks (Kelman 1958).

Electronic copy available at: https://ssrn.com/abstract=3623452

In sum, all the above considerations suggest that structural information in user networks and user attributes likely has an impact on rumor detection. This supposition is also proven by our data exploration (see below) and computational experiments (see the Evaluation section). We therefore hypothesize that

Hypothesis 1. *Combining various information at a fine-grained all-participant level in a diffusion chain will improve the performance of rumor discovery.*

One advantage of deep learning is that it can be used to model multiple levels of representation by simply composing corresponding modules in a nonlinear way. Each module abstracts higher-level representations from lower-level data. For example, in computer vision, a picture can be transformed from lower-level pixels into higher-level features representing color, texture, and morphology. While a simple concatenation strategy using traditional machine learning methods is unsuitable when combining different features from the rumor diffusion process, deep learning can implicitly and effectively deal with heterogeneous features. For example, deep learning has been proven effective in multimodal learning where data are from different sources such as text, images, and network. Hence, the deep learning model is very good at discovering intrinsic high-level features, which we believe makes it useful for rumor detection tasks. Thus, we hypothesize that

Hypothesis 2. *Deep learning-based methods will improve the performance of rumor discovery in comparison to using shallow machine learning methods.*

3.2 Data

To support the above hypotheses, we first explore our data and provide some model-free evidence. We use two standard real-word testbeds: *Twitter15* and *Twitter16*. Both were collected

Electronic copy available at: https://ssrn.com/abstract=3623452

from Twitter, the most popular social media platform in the U.S, and released by Ma et al. (2016). The two datasets share the same exact structure. Both contain the tweets (events) and retweets from a thousand news articles published in 2015 and 2016. For each article, the data contains the first tweet shared on Twitter and a sequence of retweets following this initial post. These datasets were originally constructed for a binary (rumor vs. non-rumor) classification. Later, as research demands in this area increased, the authors decided to change the label of each event from binary to quaternary (see Table 1) according to the veracity tag of each article in rumor debunking websites (e.g., snopes.com, Emergent.info). The labeling details can be found in Ma et al. (2017).

For each dataset, we first construct a diffusion-based user-user network and a temporal path of diffusion for each source tweet based on its propagation activities. In the original datasets, due constraints set out in Twitter's terms of service, no user information is provided. We crawled all related user profiles via Twitter API[2] using user IDs. From the crawled user profiles, we use eight important user attributes (length of username, user tenure (time since account created), length of user description, number of followers, number of friends, number of status updates, verified status (yes/no), and geo enabled status (yes/no)) to construct a user characteristics matrix. Note that "length of username" and "length of user description" are the number of characters for the username and for their self-description text, respectively. If a user does not have any descriptions, the value is set to 0. The descriptive statistics of datasets are shown in Table 3.

Table 3: Descriptive statistics of datasets.

	Twitter15	Twitter16
# of tweets	1,490	818
# of users	276,663	173,487

[2] https://dev.twitter.com/rest/public.

Electronic copy available at: https://ssrn.com/abstract=3623452

# of posts	331,612	204,820
# of true-rumors	372	205
# of false-rumors	370	205
# of unverified-rumors	374	203
# of non-rumors	374	205
Max. # of retweets	1,768	2,765
Min. # of retweets	55	81
Avg. # of retweets	401	251
Avg. time between a post and its last retweet	1,268 hours	848 hours

3.3 Model-free evidence

By further investigating two datasets, we obtain several interesting findings. Following from our earlier hypothesis that utilizing the information of all users who participated in a diffusion chain might improve rumor detection, we first check for any patterns or differences among involved participants in terms of their overall attributes across the four types of rumors.

- **Average participant tenure.** Prior studies find that the longer a user's social media platform tenure, the more likely that account is to spread fake news, while newer accounts have the opposite pattern. However, these studies only considered the instigator, the one who initially spread the message (Kwon et al. 2013). From our theoretical foundation, then, we predict that it will be useful to consider the tenure of all participants in the diffusion chain. As shown in Figure 2, the average user tenure of participants involved in true/false rumors is higher than the user tenure of participants involved in non-rumors.

Electronic copy available at: https://ssrn.com/abstract=3623452

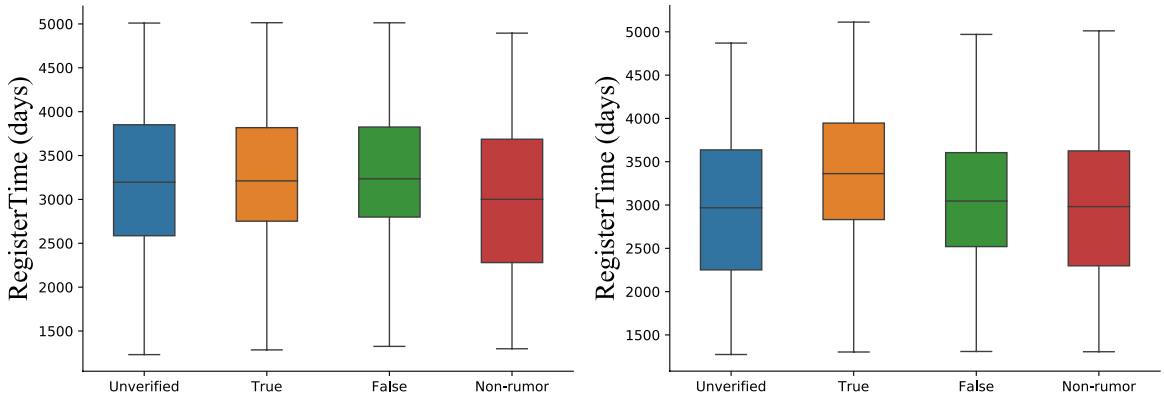


Figure 2: Average participant tenure for Twitter15 (left) and Twitter16 (right).

- **Social influence: followers/friends**. Social influence may affect the speed and the depth of diffusion. On Twitter, a user's social influence can be measured by the size of their social circle in relation to two factors: the number of friends (i.e., users the specific user is following) and the number of followers (i.e., users following the specific user). We calculate the average followers/friends across all participants and find that this average in the rumors group is different from that in the non-rumors group. To compute the social influence of all participants, we define a new metric, *TFF,* which combines followers and friends: $TFF = \frac{\#followers}{\#friends}$. Users with *TFF<1* are less influential since they have fewer followers than friends, and extreme cases are fake users. In contrast, users with TFF>1 are more influential, e.g., celebrity accounts. Figure 3 shows that a higher percentage of users with TFF<1 are involved in rumors, while more influential users participate in non-rumors.

Electronic copy available at: https://ssrn.com/abstract=3623452

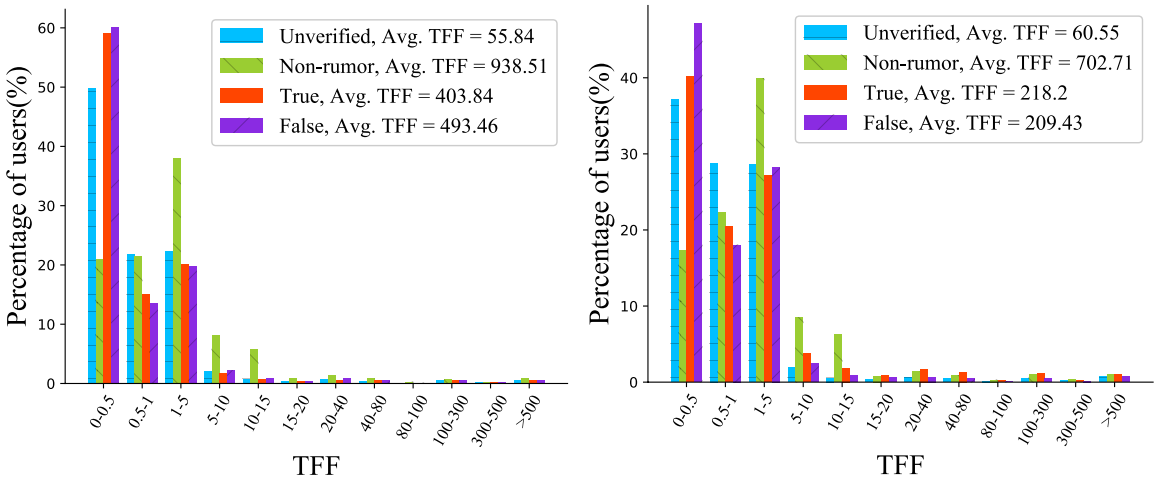


Figure 3: *TFF* distribution of participants for Twitter15 (left) and Twitter16 (right)

- **Structural and temporal impact of diffusion**. Many prior studies have demonstrated that falsehood diffuses significantly farther, faster, deeper, and more broadly than the truth in all categories of information (Vosoughi et al. 2018). This being the case, we expect to see that the spreading pattern, in terms of network structures and the temporal sequence of retweeting, should vary based on rumor type. To observe this, we randomly pick one tweet in each rumor category and show their spreading patterns in different time periods (see Figure 4). Rumor spreaders ((a)-(d) and (e-h)) more easily propagated information to both direct neighbors and those from a higher degree of connection, while users that spread genuine news ((m)-(p)) are more direct neighbors of the rumor creator. The number of affected users is larger for rumors than for non-rumors. This observation substantiates our belief that incorporating such structural and temporal information into the model can improve the performance of rumor detection.

Electronic copy available at: https://ssrn.com/abstract=3623452

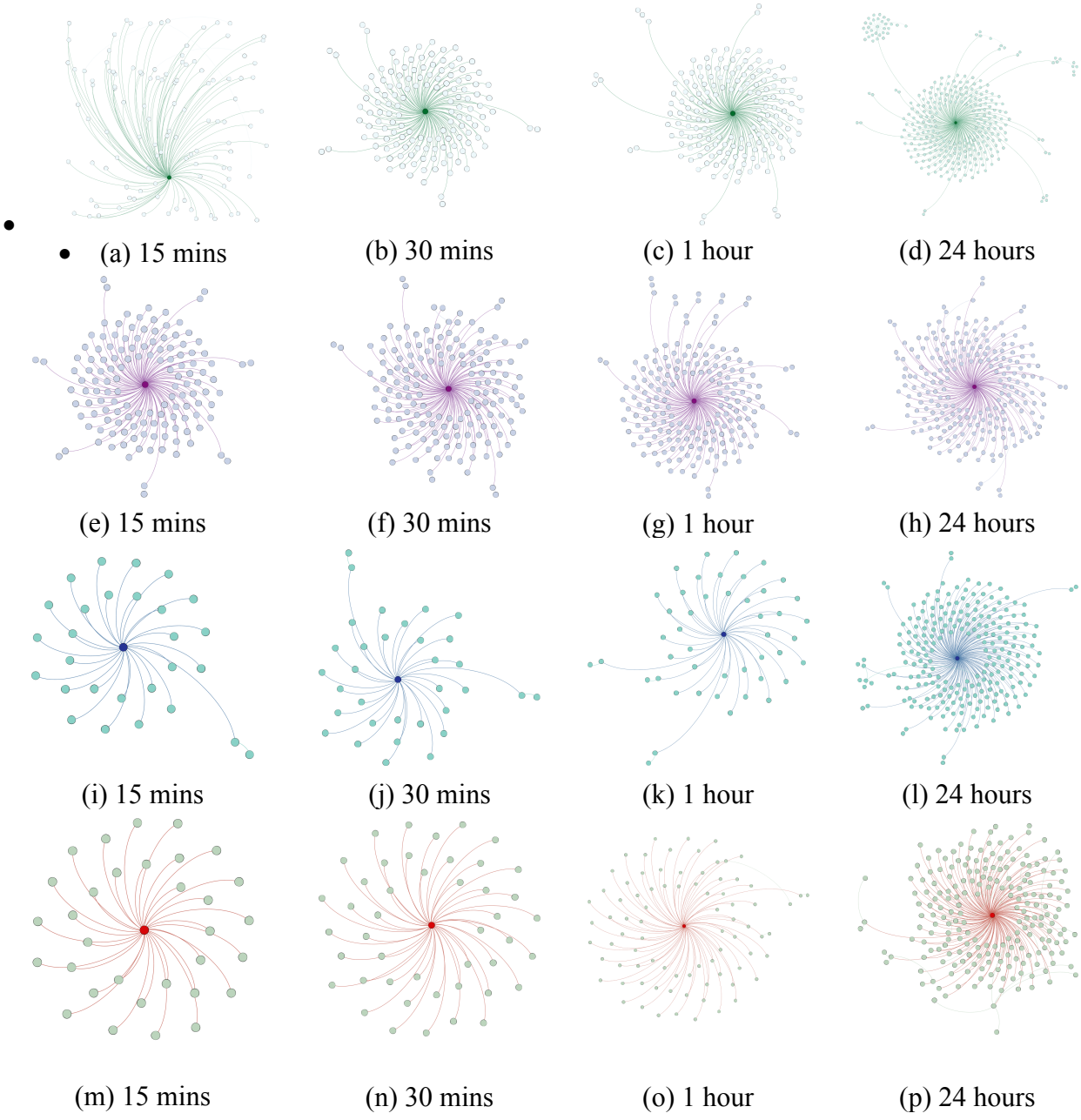


Figure 4: Diffusion patterns for a tweet: (a)-(d): True tweet; (e)-(h): False rumor tweet; (i)-(l): Unverified rumor; (m)-(p): Non-rumor rumor.

4. METHODOLOGY

In this section, we first present a preliminary overview of rumor detection in our context and describe the overall framework of the proposed DMV-RD method, followed by details of each

Electronic copy available at: https://ssrn.com/abstract=3623452

component in the model. As discussed above, two challenges need to be addressed when designing an effective rumor detection system: (1) how to incorporate fine-grained all-participant information in one model, not simply aggregating or concatenating, to capture rumor patterns and (2) how to effectively learn latent structural and temporal representations of propagation activities in a diffusion chain to capture inter- and intra-relationships among all participants. To answer these two questions, we first formally define our problem and describe its context.

4.1 Preliminary definitions

In this study, we formalize our rumor detection problem as a supervised multi-class classification task. Suppose the input of the task is from a rumor detection dataset (e.g., Twitter) consisting of a set of posts (e.g., tweets) denoted as $M = \{M_i, i \in [1, |M|]\}$. Each M_i corresponds to its own diffusion process and all participating users. M_i can be represented by $\{G_i, P_i, \mathcal{U}_i, C_i\}$, where G_i, P_i, $\mathcal{U}_i$, and C_i are the diffusion network, the diffusion path, the user characteristics matrix and the post content, respectively. See their formal definitions below.

Definition 1. ***Diffusion Network*** A diffusion network for a tweet M_i is denoted as $G_i = \{U_i, E_i\}$, where U_i is the node set comprising *t* nodes (i.e., all users who are involved in the diffusion process of M_i) and $E_i = \{(u_i, u_j) | u_i, u_j \in U_i\}$ represents a set of edges connecting pairs of users when u_i retweets u_j about M_i. Note that the diffusion network is a directed acyclic graph.

Definition 2. ***Diffusion Path*** A diffusion path of a given tweet M_i is defined as a variable-length multivariate time series $P_i = \{(u_1, T_1), (u_2, T_2), \cdots, (u_t, T_t)\}$, where each tuple (u_k, T_k) denotes that user u_k retweets the source tweet M_i at time T_k, u_1 is the source user (who initiates the tweet), and $k \in [2, t]$are corresponding retweeting users. Here, all users are in chronological order according to their retweeting time.

Electronic copy available at: https://ssrn.com/abstract=3623452

Definition 3. ***User Characteristics Matrix*** Each user $u_i \in U_i$ is associated with a user vector $d_i \in \mathbb{R}^{d_{user}}$, which represents all relevant user profile information, such as number of followers, whether the user is verified, etc. We concatenate vectors of all users who retweeted the tweet M_i to form a user characteristics matrix $D_i \in \mathbb{R}^{t \times d_{user}}$. The concatenation order follows the diffusion time.

Definition 4. ***Rumor Detection*** Given a tweet $M_i = \{G_i, P_i, \mathcal{U}_i, C_i\}$ within an observation window T (T can be time or the total number of retweets), the goal of rumor detection is to learn a function $f(\mathcal{L}(M_i)|G_i, P_i, \mathcal{U}_i, C_i; T)$ based on all existing retweeting activities to classify the source tweet M_i into one of four rumor categories, where $\mathcal{L}(M_i)$ represents either false rumor, true rumor, unverified rumor or non-rumor.

Table 4 summarizes all notations used throughout the paper.

Table 4: List of notations.

Symbol	Description
M	A set of tweets / posts
M_i	The i^{th} tweet / post
G_i	Diffusion network of tweet M_i
P_i	Diffusion path of tweet M_i
$\mathcal{U}_i$	User demographics matrix of tweet M_i
C_i	The textual content of tweet M_i
U_i, E_i	The set of participating users for tweet M and edges in G_i
T, T_i	The observation window, timestamp of retweet from user u_i
t	The total number of participating users in P_i
$H_{Participant}, H_{Stru}$ H_{Temp}, H_{Text}	The personal aspect, structural aspect, temporal aspect, and text aspect of all participants
d_{user}, d_{stru} d_{temp}, d_{text}	The dimensionality of personal aspect, structural aspect, temporal aspect, and text aspect
$\hat{Y}/\hat{y}, Y/y$	The predicted label and the ground truth
K	The maximum hops from the central node, i.e., K^{th}-hop neighborhood or Chebyshev coefficients

Electronic copy available at: https://ssrn.com/abstract=3623452

4.2 Overall framework of DMV-RD

In this section, we describe our proposed DMV-RD rumor detection system. It consists of the following components (see Figure 5): (a) inputs, including (1) an observed diffusion network, (2) a diffusion path, (3) social media post content, and (4) a user characteristics matrix; (b) representation learning, which involves a diffusion graph convolutional network, a time-decay LSTM, and text embedding; and (c) rumor classification, in which we fuse the multi-view representations and feed them into a rumor classifier. We use several fully connected feedforward layers and a *softmax* output layer to generate a rumor prediction. Below, we explain each of the above components in detail.

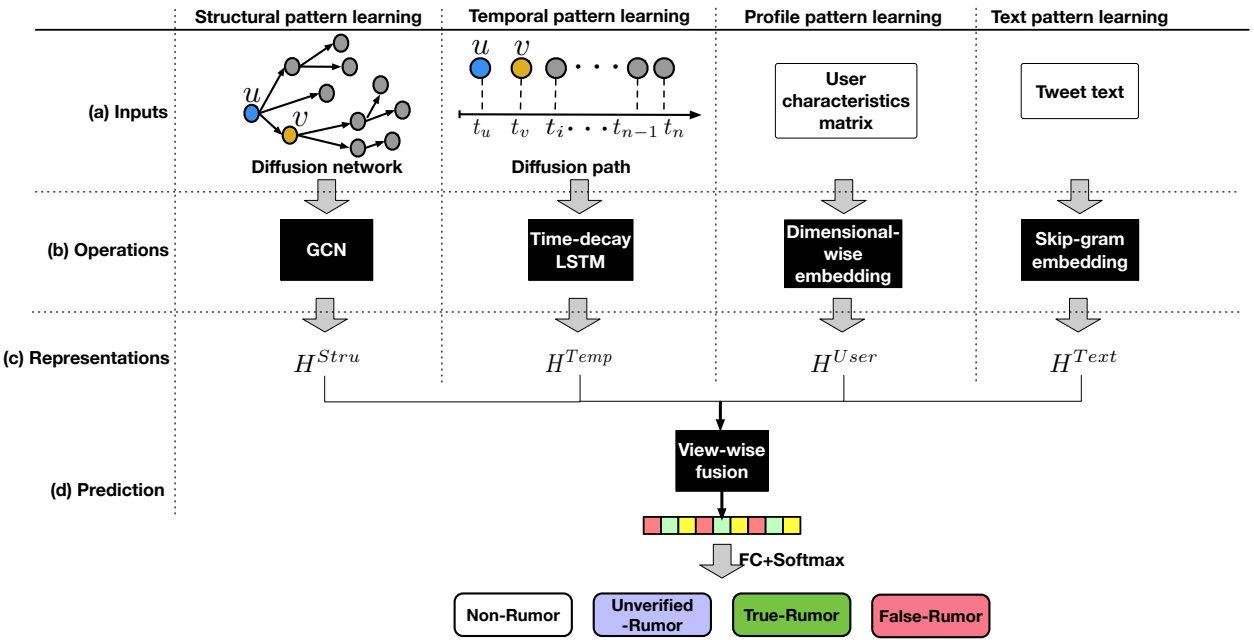


Figure 5: Overview of DMV-RD.

4.3 Participants' structural information

Due to the complex nature of information diffusion, explicitly crafting a set of network-based features that cover all possible structural information, such as the degree of connection, similarity, distance and even community among nodes is difficult. It is even more challenging to

Electronic copy available at: https://ssrn.com/abstract=3623452

represent all of this multi-dimensional information in one variable (for the sake of efficiency). This can be done to some extent using a simple aggregation strategy, but this inevitably results in information loss.

In order to better capture the structural information of all participants involved in the process of rumor propagation within a diffusion network, we employ a multi-layer diffusion convolutional network (Li et al. 2017) to encode multi-dimensional structural information into one vector representation for each participant. The structural information related to one news item is then represented as a matrix formed by concatenating the vectors of all participants. This structural information may include many elements, such as what types of users are spreading news to each other. This is informative, as users who spread news early might affect more users who are similar to themselves, or rumors may spread deeper than non-rumors. Overall, when two news items have similar structural information at the all-participant level, this probably suggests that their spreading patterns are similar. The idea of this diffusion convolutional network is to update each node by combining the information from the node itself and information from other nodes via propagation.

The rule of information propagation within the diffusion convolutional network is defined as

$$H^{(l+1)} = \sigma(\sum_{k=1}^{K}(\theta_{k,1}(D_O^{-1}W)^k + \theta_{k,2}(D_I^{-1}W^T)^k)H^{(l)})$$

where k is the finite diffusion step, $\theta \in \mathbb{R}^{K\times 2}$ are the filter parameters and $D_O^{-1}W$, $D_I^{-1}W^T$ are transition matrices, with D_O and D_I representing the out-degree diagonal matrix and the in-degree diagonal matrix, respectively. Activation function is denoted by $\sigma(\cdot)$, i.e., $ReLU(\cdot)$. $H^{(l)} \in \mathbb{R}^{t\times F}$ is the matrix of activation in the l^{th} layer, where t is the number of nodes in diffusion

Electronic copy available at: https://ssrn.com/abstract=3623452

network and F is the dimension of the output. In the proposed model, the initial input $H^{(0)}$ is obtained from a pre-trained network embedding layer that maps a user u^i to its D-dimensional representation $g_i \in \mathbb{R}^D$, which allows for the learning of varying-size diffusion networks.

To reduce over-fitting for diffusion convolutional networks, in each training epoch we randomly drop out edges from the input diffusion networks to generate different copies with a certain ratio. Formally, suppose the total number of edges in a diffusion network is $|E_i|$ and the dropping rate is m, then the adjacency matrix after dropout is computed as

$$\hat{A} = A - A_{drop}$$

where A_{drop} is the matrix constructed using $|E_i| \times m$ edges randomly sampled from the original edge set E_i. After the diffusion convolutional layer, the diffusion network G_i is represented as a matrix $H^{Stru} \in \mathbb{R}^{t \times d_{stru}}$, where d_{stru} is the representation dimension and can be tuned.

4.4 Participants' temporal information

The goal of learning temporal representation is to capture sequential patterns of user engagement with a specific tweet. If the order in which a tweet spreads from one user to another is known, this might aid in rumor detection performance because fake news spreads much faster than genuine news and affects more users in the early stage. For simplicity, we leverage recurrent neural networks (RNNs) to model this temporal dependencies of diffusion – in particular, we use a stable and powerful variant of RNN called Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997). However, the influence of retweeting users will decay as time elapses, and this cannot be captured by the traditional LSTM. Therefore, we borrow the idea of time-decay LSTM (see Zhu et al. (2017)) to introduce a time gate into the LSTM. The representation of the

Electronic copy available at: https://ssrn.com/abstract=3623452

temporal aspect for all *t* participants is denoted as $H^{Temp} = \{h_1^{d_{temp}}, h_2^{d_{temp}}, \cdots, h_t^{d_{temp}}\}$ where d_{temp} is the temporal representation dimension for each participant.

4.5 Participants' personal information

To obtain a single user profile representation that covers all participants who are involved in the diffusion of tweet M_i, we first combine the eight aforementioned important profile attributes to form a feature vector for each participant u_i in the observed diffusion network; we represent this vector as $p_i \in \mathbb{R}^{d_{user}}$. Then we concatenate vectors of all participants into one user characteristics matrix $\mathcal{U}_i \in \mathbb{R}^{t \times d_{user}}$, where t is the total number of users who retweet the original tweet M_i and d_{user} is the dimension of the user feature vector (i.e., eight here).

To provide further guidance on possible interdependencies among users, we build a dimensional-wise embedding to explicitly model their latent relations. This embedding includes three operations. First, it expands the dimension of the user's personal aspect $\mathcal{U}_i$ into a three-dimensional tensor with a shape of $t \times 1 \times d_{user}$; this last dimension can be regarded as a channel, similar to that in a colorful image. Second, the embedding utilizes global average pooling to generate implicit dimensional-wise statistics $z \in \mathbb{R}^{d_{user}}$ to tackle the issue of exploiting dimension dependencies. To accomplish its objective of fully capturing dimensional-wise dependencies, it employs two fully connected (FC) layers along with non-linear activation functions, i.e., a dimension-reduction layer and a dimension-increasing layer

$$f^{(1)} = \delta(W_1 z + b_1)$$

$$f^{(2)} = \sigma(W_2 f^{(1)} + b_2)$$

Electronic copy available at: https://ssrn.com/abstract=3623452

where δ and σ refer to *tanh* and *softmax* functions, respectively. Weight parameters associated with neurons are calculated as $W_1 \in \mathbb{R}^{\frac{d_{user}}{r} \times d_{user}}$ and $W_2 \in \mathbb{R}^{d_{user} \times \frac{d_{user}}{r}}$, $b_1 \in \mathbb{R}^{\frac{d_{user}}{r}}$ and $b_2 \in \mathbb{R}^{d_{user}}$ are biases, and *r* is the ratio for dimension reduction. Lastly, dimensional-wise multiplication is applied to obtain the final output for all *t* users: $H^{Participant} \in \mathbb{R}^{t \times d_{user}}$.

4.6 Source tweet text

Building on the literature of deception detection using linguistic features, prior research has shown that fake news articles have different linguistic styles than corresponding legitimate news articles (Clarke et al. 2019). However, this study used one-hot encoding to represent news articles, and this strategy usually ignores the rich latent and deep semantics among words and particularly overlooks the order or the contextual information among words. Since the goal of our predictive task is to boost rumor detection performance, we use word embedding to represent words as well as documents. Due to a breakthrough prompted by recent advances, neural word representation learning, also known as word embedding (Mikolov et al. 2013), has become an increasingly important building block in text analysis. Word embedding represents each word as a multi-dimensional vector, where not only the individual semantic characteristics of words but also the relationships among words are well captured. In this study, we use pre-trained embedding for each word in all tweets and obtain a tweet representation using a simple yet effective weighted strategy where the weight of each word in the tweet is the term frequency-inverse document frequency (*tf-idf*) across all tweets in our dataset. Note, this textual representation is denoted as $H^{Text} \in \mathbb{R}^{d_{text}}$ for a tweet, where d_{text} is the dimension of word embedding and can be tuned.

Electronic copy available at: https://ssrn.com/abstract=3623452

4.7 Rumor detection

The above components represent different aspects of a news diffusion chain, including its structural pattern, temporal pattern, profile pattern and content pattern. We finally fuse all of these features together and feed them into a deep neural network in order to learn the interdependencies among the individual patterns.

Our ultimate goal is to predict the rumor label $\widehat{y_i}$ of tweet M_i. We calculate this through several fully connected layers and a *softmax* output layer $y_i = softmax(FC(S))$, where $y_i \in \mathbb{R}^C$ is a vector of predicted probabilities of all rumor categories (e.g., $C = 4$) for the tweet M_i, and S is the multi-view fusion of H^{Text}, $H^{Participants}$, H^{Stru}, and H^{Temp}. Specifically, we train *DMV-RD* to estimate all parameters by minimizing cross-entropy of the predictions $\hat{Y}$ and the ground truth labels Y. The well-known stochastic gradient descent is applied to update parameters. The overall objective loss function is

$$\mathcal{L}(Y, \hat{Y}) = -\sum_{i \in |M|} y_i log y_i + \lambda ||\Theta||_2^2$$

where $||\Theta||_2^2$ is the L_2 regularizer applied over all model parameters Θ, and λ is a hyper-parameter whose value is optimized for better results. In this work, we use the adaptive learning rate optimization algorithm Adam during training. All hyper-parameters are tuned using the standard grid search for optimal results. The next section provides the details of the computational experiments.

5. EVALUATION

Electronic copy available at: https://ssrn.com/abstract=3623452

Following the design science paradigm, we rigorously evaluate our proposed DMV-RD framework and demonstrate its practical utility through quantitative experiments. Note that all results are reported based on five-fold cross-validation and both datasets are well balanced.

Before we prepare the training set, we specify the observation window (i.e., *T* can be time or the current total number of retweets) for rumor prediction. Figure 6 shows that the diffusion is saturated in approximately 24 hours, and by that time a tweet can gain about 370 retweets on average, which is close to the overall average number of retweets in each dataset. Our objective is to design an artifact that can not only achieve better performance over state-of-the-art baselines but is also able to detect rumors as early as possible in order to minimize their harmful effects. Therefore, we set the observation window from the perspectives of both time and the current number of retweets.

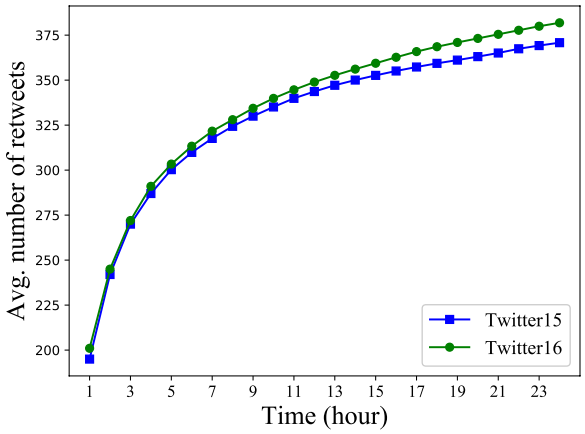


Figure 6: Average diffusion speed within the first 24 hours.

In sum, we present empirical evaluations of our proposed DMV-RD method by comparing its performance with several state-of-art baselines and several variants on the rumor detection task.

Electronic copy available at: https://ssrn.com/abstract=3623452

5.1 Evaluation metrics and baselines

In this paper, we use two standard evaluation metrics, accuracy and *F1* score, to measure the performance of rumor detection models, which are common in most classification tasks. Specifically, accuracy (ACC) is the most intuitive performance measure, and it is simply a ratio of correctly predicted observation to the total observations: $ACC = \frac{\#\ of\ correctly\ detected\ tweets}{\#\ of\ tweets}$.

The $F1$ score is the harmonic mean of the precision and recall. Here, we calculate $F1$ for each rumor class (i.e., non-rumor, true-rumor, false-rumor, and unverified rumor) as $F1 = \frac{2 \times P^* \times R^*}{P^* + R^*}$ where P^* and R^* are precision and recall, respectively.

We rigorously compare our method with a battery of baselines. To demonstrate the importance of each component in our model, we design four benchmarks, each with one model component excluded: (i) ***RD-NP*** excludes user attributes, making this variant useful for privacy preservation; (ii) ***RD-NN*** excludes structural information of diffusion networks, and (iii) ***RD-NT*** excludes modeling of the temporal effect of diffusion. Since one of our major contributions lies in the representation and incorporation of all participants in the diffusion process, we purposely implement a baseline, ***RD-Aggregate***, that extracts aggregate-level information from the diffusion network and the diffusion sequence. Specifically, this baseline only leverages the hidden representation of the last time step, rather than all the time steps, as the temporal aspect. For the structural aspect, ***RD-Aggregate*** averages all node representations into one vector (e.g., average pair-wise similarity) and also adds several network-level properties, such as average degree, network diameter, and density. To demonstrate our deep representation-based method outperforms non-deep learning methods, we design ***Random-Forest***, which is another random forest-based ensemble method. Note that all algorithms are implemented using a desktop Linux

Electronic copy available at: https://ssrn.com/abstract=3623452

machine with 32GB of memory. To facilitate reproducibility of our results, we will make the source code publicly available.

5.2 Overall performance

Table 5 summarizes the performance comparison between DMV-RD and the above baselines for the case where the observation window is 24 hours for both Twitter15 and Twitter16. The last column shows the performance of the complete version of our model DMV-RD, which basically yields much better performance than the other baseline methods across all metrics.

We make the following additional observations. Model performance when all information is at the aggregate level (column 3) is relatively poor, varying between 0.317 and 0.643 for *F1*, indicating that the fine-grained information of all the participants involved in the diffusion process is vital to rumor prediction. The models with one component missing (columns 4-6) all underperform our model for various reasons: (a) RD-NP performs the worst because it neglects specific participant information, which, as seen in our model-free evidence, is likely to have a big impact on rumor detection performance; (b) RD-NN does not utilize complex structural information in the diffusion network; and (c) RD-NT incorporates the diffusion network and learns high-level representation from the propagation structure, but it ignores the temporal aspect of the diffusion path. We also review performance of a model that excludes textual content; this model performs comparably, thus indicating that the content of a short tweet is not that important in rumor detection. We believe that this is because a tweet differs from a fake news article, which is typically a long story from which many statements, facts, and evidence can be extracted.

Lastly, our proposed DMV-RD model, which takes different aspects (i.e., personal characteristics, temporal, and structural information) into consideration and leverages deep

Electronic copy available at: https://ssrn.com/abstract=3623452

representation learning, significantly outperforms the widely used non-deep learning ensemble-based method Random Forest, as well as all other baselines. DMV-RD yields good prediction results with accuracy of around 0.883 and 0.887, respectively, on the Twitter15 and Twitter16 datasets.

Table 5: Overall performance comparison for rumor detection on Twitter15 and Twitter16. ("UR": Unverified rumor; "NR": Non-rumor; "TR": True rumor; "FR": False rumor.) The * denotes statistical significance compared to the second best model results under a one-tailed *t*-test (*** for $p < 0.001$ and ** for $p < 0.01$).

Datasets	**Metric**		**RD-Aggregate**	**RD-NP**	**RD-NN**	**RD-NT**	**Random-Forest**	**DMV-RD**
Twitter15	ACC		0.454	0.544	0.646	0.723	0.697	**0.883*****
	UR	F1	0.415	0.483	0.592	0.654	0.689	**0.897*****
	NR	F1	0.733	0.796	0.792	0.683	0.760	**0.871*****
	TR	F1	0.317	0.404	0.608	0.821	0.696	**0.863*****
	FR	F1	0.355	0.472	0.574	0.758	0.645	**0.912*****
Twitter16	ACC		0.465	0.574	0.633	0.737	0.702	**0.887*****
	UR	F1	0.403	0.526	0.593	0.708	0.608	**0.920*****
	NR	F1	0.643	0.755	0.772	0.662	0.711	**0.877*****
	TR	F1	0.419	0.571	0.686	0.835	0.816	**0.868*****
	FR	F1	0.393	0.420	0.489	0.743	0.664	**0.925*****

In addition, we compare DMV-RD with RD-Aggregate and Random-Forest in terms of the ROC[3] curve on both datasets. A ROC curve or ROC graph is a two-dimensional graph that represents the relationship between the TPR (true positive rate) on the Y axis and the FPR (false positive rate) on the X axis for various decision threshold settings. The larger the ROC, the better the rumor detection performance. As shown in Figure 7, DMV-RD consistently outperforms all baselines.

[3] https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Electronic copy available at: https://ssrn.com/abstract=3623452

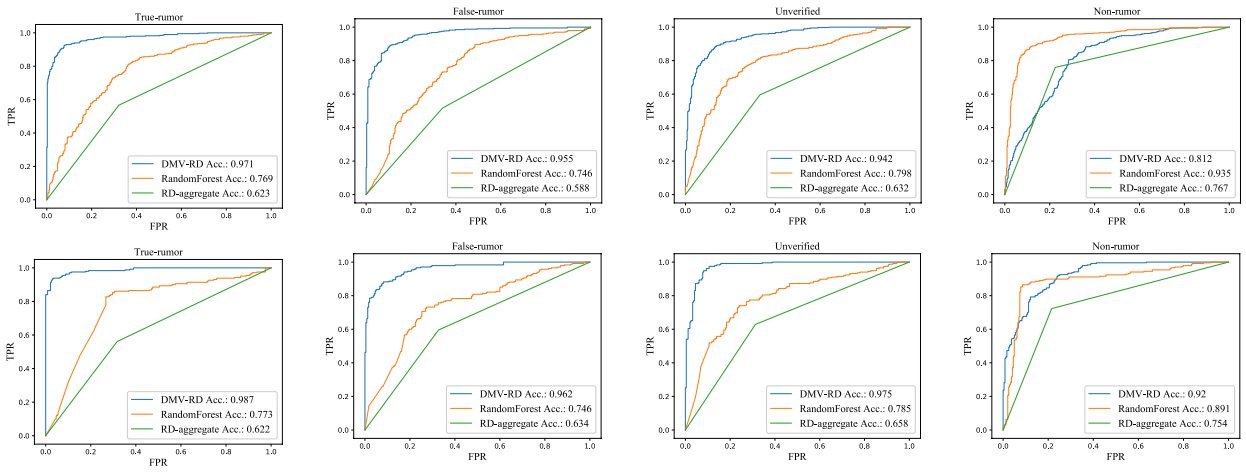


Figure 7: Comparison of ROC on Twitter 15 (upper row) and Twitter16 (lower row).

5.3 Ablation study

To demonstrate the effectiveness of each component in DMV-RD, we present two variants of DMV-RD, denoted as ***DMV-DW*** and ***DMV-Time***. These variants were built upon the original DMV-RD with some components changed, as detailed below. In ablation studies, we fix the observation window to 10 retweets.

- **DMV-DW**: In this model, we do not consider the dimension-wise embedding of user characteristic representation learning, which means the interdependencies among user attributes are ignored.

- **DMV-Time**: In this model, the time decay effect of retweeting is ignored, which suggests that the time-decay LSTM is replaced by the plain LSTM.

The results, shown in Table 6, indicate that the original DMV-RD outperforms these variants in terms of both metrics, accuracy and *F1* score. In the comparison to *DMV-Time*, we find that the time decay effect, i.e., the influence of retweet decay as time elapses, is indispensable for rumor detection. DMV-DW also provides lower detection performance than

Electronic copy available at: https://ssrn.com/abstract=3623452

DMV-RD in most cases, which indicates the effectiveness of the dimension-wise embedding of user characteristic representation learning in DMV-RD. This also corroborates our first hypothesis: that interdependencies among participants play a critical role in rumor detection.

Table 6: performance comparison between DMV-RD and its variants.

Datasets	**Metric**		**DMV-DW**	**DMV-Time**	**DMV-RD**
Twitter15	ACC		0.630	0.648	**0.714**
	UR	F1	**0.667**	0.608	0.657
	NR	F1	0.699	0.592	**0.733**
	TR	F1	0.638	0.792	**0.843**
	FR	F1	0.448	0.574	**0.642**
Twitter16	ACC		0.641	0.664	**0.725**
	UR	F1	0.667	0.643	**0.718**
	NR	F1	0.713	0.606	**0.768**
	TR	F1	0.775	**0.837**	0.823
	FR	F1	0.334	0.475	**0.674**

6 DISCUSSIONS AND CONCLUSION

The wide adoption and use of the Internet and social media facilitate free communication and information sharing. They also play a key role in the proliferation of online rumors, which have a significant negative societal and economic effect. Early detection of rumors provides means to moderate their diffusion. In this study, following the design science research paradigm and building on social influence theory, we propose a fine-grained all-participant level rumor detection method. It is an effective deep learning framework that leverages the information of all participants in a diffusion chain. This distinguishes our work from prior rumor detection studies, which focus on either content or user profile or which consider participants' structural and temporal features from an aggregate perspective. In particular, we formulate our problem from a supervised learning perspective and propose a deep multi-view learning framework DMV-RD.

Electronic copy available at: https://ssrn.com/abstract=3623452

DMV-RD is a theory-driven approach, incorporating various recently developed techniques such as graph convolution, time-decay LSTM networks, and text embedding. We evaluate DMV-RD on two widely used datasets, Twitter15 and Twitter16. Our extensive analysis shows that DMV-RD significantly outperforms baselines that utilize fewer participant aspects or that utilize the same aspects but at an aggregate level. This work is among the first to conduct rumor detection at a fine-grained all-participant level and to explore the effect of different aspects of all participants involved in the diffusion chain.

Our approach can effectively model each tweet in the diffusion process from a multi-view perspective, a more challenging task than considering aspects of a tweet separately. The modeling and representation we use in DMV-RD can be extended to many downstream applications, such as link prediction (Li et al. 2016) and information cascade prediction. In addition, rumor detection is very important for online platform stakeholders and has societal and economic impact.

Our study therefore provides key implications for academics, social media platforms and policy makers. First, predictive analytics can add theoretical and practical value to IS research (Shmueli and Koppius 2011). We first provide empirical evidence that all participants in the diffusion chains of rumors exhibit different patterns than participants in the diffusion chains of non-rumors. While prior studies have mostly focused on making predictions using content features or the rumor instigator's features, we believe, based on the social influence theory, that considering the information of all participants in a diffusion chain can improve overall rumor detection performance. Theoretically, our findings provide new perspectives on the traditional social influence theory as it operates within the context of online news diffusion.

Electronic copy available at: https://ssrn.com/abstract=3623452

Methodologically, our novel multi-view deep learning approach is effective in learning all participants' features without overfitting the machine learning model.

Second, social media platforms such as Facebook and Twitter have a vision of bringing people closer together using social media technologies that allow them to express themselves freely, fairly and safely. Yet since these technologies also create plenty of fake information, identifying and moderating misinformation plays a critical role in platforms' vision. However, misclassifying non-fake rumors as fake news and removing them from the platform may backfire, with negative consequences for the expression and sharing of free opinion. For this reason, our predictive framework with its solid theoretical foundation has direct and actionable insights for social media platforms.

Third, our empirical finding suggests that the fine-grained all-participant features in a news diffusion chain are predictive for rumor detection, which implies that the echo chambers of fake news and non-fake news are different. Online users are unlikely to seek out and distribute information that is counter to their viewpoints. Therefore, strategies aimed at exposing false information online may have limited applicability. As such, findings from this study suggest a need to consider alternative strategies apart from applying fact-checkers or developing "whitelists" of articles on the online platform. In the long run, the policy makers may adopt a proactive approach to cultivate news consumers' critical faculties so that they are able to distinguish credible sources and stories from non-credible ones.

Our work also has several limitations that can be addressed in the future. For example, we do not incorporate explicit social networks (e.g., a friendship network) into the model, and the higher-order relationships among participants in the diffusion structure are also ignored. In

Electronic copy available at: https://ssrn.com/abstract=3623452

addition, when we calculate the temporal effect, we only leverage the order of diffusion, rather than the speed of diffusion, among participants, which might be affect the rumor detection performance. As previous studies have shown, diffusion patterns, including speed, are significantly different for rumors and for genuine news. Lastly, from a business management standpoint, we always expect to understand the economic and social impact of rumor detection. In this study, since we did not collaborate with a real social media platform, we are unable to demonstrate the impact (e.g., user satisfaction or revenue change) of DMV-RD. Finding a collaborating platform that will deploy our system is well worth pursuing.

Electronic copy available at: https://ssrn.com/abstract=3623452

References

Abbasi, A., Zahedi, F. "Mariam," Zeng, D., Chen, Y., Chen, H., and Nunamaker, J. F. 2015. "Enhancing Predictive Analytics for Anti-Phishing by Exploiting Website Genre Information," *Journal of Management Information Systems* (31:4), pp. 109–157. (https://doi.org/10.1080/07421222.2014.1001260).

Allcott, H., and Gentzkow, M. 2017. "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* (31:2), Nashville: Amer Economic Assoc, pp. 211–235. (https://doi.org/10.1257/jep.31.2.211).

Amado, B. G., Arce, R., and Fariña, F. 2015. "Undeutsch Hypothesis and Criteria Based Content Analysis: A Meta-Analytic Review," *The European Journal of Psychology Applied to Legal Context* (7:1), Elsevier, pp. 3–12.

Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., and Huang, J. 2020. "Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks," *ArXiv:2001.06362 [Cs]*. (http://arxiv.org/abs/2001.06362).

Castillo, C., Mendoza, M., and Poblete, B. 2011. "Information Credibility on Twitter," in *Proceedings of the 20th International Conference on World Wide Web*, pp. 675–684.

Clarke, J., Chen, H., Du, D., and Hu, Y. J. 2019. "Fake News, Investor Attention, and Market Reaction," SSRN Scholarly Paper No. ID 3213024, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network, September 1. (https://doi.org/10.2139/ssrn.3213024).

DiFonzo, N., and Bordia, P. 1997. "Rumor and Prediction: Making Sense (but Losing Dollars) in the Stock Market," *Organizational Behavior and Human Decision Processes* (71:3), San Diego: Academic Press Inc Jnl-Comp Subscriptions, pp. 329–353. (https://doi.org/10.1006/obhd.1997.2724).

Dong, W., Liao, S., and Zhang, Z. 2018. "Leveraging Financial Social Media Data for Corporate Fraud Detection," *Journal of Management Information Systems* (35:2), pp. 461–487. (https://doi.org/10.1080/07421222.2018.1451954).

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly*, pp. 337–355.

Grover, V., Chiang, R. H. L., Liang, T.-P., and Zhang, D. 2018. "Creating Strategic Business Value from Big Data Analytics: A Research Framework," *Journal of Management Information Systems* (35:2), pp. 388–423. (https://doi.org/10.1080/07421222.2018.1451951).

Guo, J., Zhang, W., Fan, W., and Li, W. 2018. "Combining Geographical and Social Influences with Deep Learning for Personalized Point-of-Interest Recommendation," *Journal of Management Information Systems* (35:4), pp. 1121–1153. (https://doi.org/10.1080/07421222.2018.1523564).

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105. (https://doi.org/10.2307/25148625).

Electronic copy available at: https://ssrn.com/abstract=3623452

Hochreiter, S., and Schmidhuber, J. 1997. "Long Short-Term Memory," *Neural Computation* (9:8), pp. 1735–1780.

Jin, Z., Cao, J., Guo, H., Zhang, Y., and Luo, J. 2017. "Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs," in *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*, Mountain View, California, USA: ACM Press, pp. 795–816. (https://doi.org/10.1145/3123266.3123454).

Kelman, H. C. 1958. "Compliance, Identification, and Internalization Three Processes of Attitude Change," *Journal of Conflict Resolution* (2:1), Sage Publications Sage CA: Thousand Oaks, CA, pp. 51–60.

Kumar, N., Venugopal, D., Qiu, L., and Kumar, S. 2018. "Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning," *Journal of Management Information Systems* (35:1), pp. 350–380. (https://doi.org/10.1080/07421222.2018.1440758).

Kumar, N., Venugopal, D., Qiu, L., and Kumar, S. 2019. "Detecting Anomalous Online Reviewers: An Unsupervised Approach Using Mixture Models," *Journal of Management Information Systems* (36:4), pp. 1313–1346. (https://doi.org/10.1080/07421222.2019.1661089).

Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. 2013. "Prominent Features of Rumor Propagation in Online Social Media," in *2013 IEEE 13th International Conference on Data Mining*, IEEE, pp. 1103–1108.

Li, Y., Yu, R., Shahabi, C., and Liu, Y. 2017. "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," *ArXiv Preprint ArXiv:1707.01926*.

Li, Z., Fang, X., Bai, X., and Sheng, O. R. L. 2016. "Utility-Based Link Recommendation for Online Social Networks," *Management Science* (63:6), pp. 1938–1952.

Liu, X., Nourbakhsh, A., Li, Q., Fang, R., and Shah, S. 2015. "Real-Time Rumor Debunking on Twitter," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM '15*, Melbourne, Australia: ACM Press, pp. 1867–1870. (https://doi.org/10.1145/2806416.2806651).

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., and Cha, M. 2016. "Detecting Rumors from Microblogs with Recurrent Neural Networks," *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, p. 7.

Ma, J., Gao, W., and Wong, K.-F. 2017. "Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada: Association for Computational Linguistics, July, pp. 708–717. (https://doi.org/10.18653/v1/P17-1066).

Ma, J., Gao, W., and Wong, K.-F. 2018. "Rumor Detection on Twitter with Tree-Structured Recursive Neural Networks," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, July, pp. 1980–1989. (https://doi.org/10.18653/v1/P18-1184).

Electronic copy available at: https://ssrn.com/abstract=3623452

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* (27:1), pp. 415–444.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. 2013. "Distributed Representations of Words and Phrases and Their Compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (eds.), Curran Associates, Inc., pp. 3111–3119. (http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf).

Nickerson, R. S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* (2:2), SAGE Publications Sage CA: Los Angeles, CA, pp. 175–220.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45–77.

Shearer, E., and Matsa, K. E. 2018. "News Use Across Social Media Platforms 2018," *Pew Research Center's Journalism Project*, , September 10. (https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/, accessed April 13, 2020).

Shmueli, G., and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly* (35:3), pp. 553–572. (https://doi.org/10.2307/23042796).

Shore, J., Baek, J., and Dellarocas, C. 2018. "Network Structure and Patterns of Information Diversity on Twitter," *MIS Quarterly* (42:3), pp. 849–872.

Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. 2017. "Fake News Detection on Social Media: A Data Mining Perspective," *ArXiv:1708.01967 [Cs]*. (http://arxiv.org/abs/1708.01967).

Vosoughi, S., Roy, D., and Aral, S. 2018. "The Spread of True and False News Online," *Science* (359:6380), pp. 1146–1151. (https://doi.org/10.1126/science.aap9559).

Wu, K., Yang, S., and Zhu, K. Q. 2015. "False Rumors Detection on Sina Weibo by Propagation Structures," in *2015 IEEE 31st International Conference on Data Engineering*, IEEE, pp. 651–662.

Yang, F., Liu, Y., Yu, X., and Yang, M. 2012. "Automatic Detection of Rumor on Sina Weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics - MDS '12*, Beijing, China: ACM Press, pp. 1–7. (https://doi.org/10.1145/2350190.2350203).

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., and Liu, H. 2019. "Unsupervised Fake News Detection on Social Media: A Generative Approach," *Proceedings of the AAAI Conference on Artificial Intelligence* (33), pp. 5644–5651. (https://doi.org/10.1609/aaai.v33i01.33015644).

Yu, F., Liu, Q., Wu, S., Wang, L., and Tan, T. 2017. "A Convolutional Approach for Misinformation Identification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, August, pp. 3901–3907. (https://doi.org/10.24963/ijcai.2017/545).

Electronic copy available at: https://ssrn.com/abstract=3623452

Zhang, Zili, Zhang, Ziqiong, and Li, H. 2015. "Predictors of the Authenticity of Internet Health Rumours," *Health Information & Libraries Journal* (32:3), pp. 195–205.

Zhao, Z., Resnick, P., and Mei, Q. 2015. "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts," in *Proceedings of the 24th International Conference on World Wide Web*, pp. 1395–1405.

Zhu, Y., Li, H., Liao, Y., Wang, B., Guan, Z., Liu, H., and Cai, D. 2017. "What to Do Next: Modeling User Behaviors by Time-LSTM.," in *IJCAI* (Vol. 17), pp. 3602–3608.

Electronic copy available at: https://ssrn.com/abstract=3623452

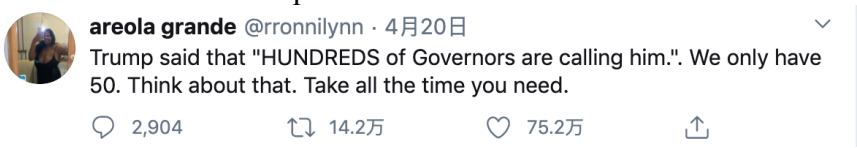
Appendix: Examples of four types of rumors.

Note that we do not use samples from our dataset because some news items that were demonstrably unverified might become true given new evidence. We pick the following four examples and use a fact checking system (snopes.com) to obtain their labels, as done by Ma et. al (2017).

True rumor:
Claim: Three previous U.S. Presidents were all younger than 2020 candidates.

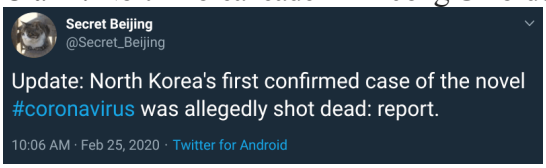
Fact-check: https://www.snopes.com/fact-check/presidents-younger-2020-candidates/

False rumor:
Claim: U.S. President Donald Trump said "hundreds" of governors are calling him amid the COVID-19 coronavirus pandemic.



Fact-check: https://www.snopes.com/fact-check/trump-governors-covid-19/

Unverified rumor:
Claim: North Korea leader Kim Jong Un ordered his country's first COVID-19 patient to be shot dead.



Fact-check: https://www.snopes.com/fact-check/kim-jong-shot-coronavirus-patient/

Non-rumor:
Claim: U.S. President Donald Trump suggested during a White House briefing that injecting disinfectants could treat COVID-19.

Fact-check: https://www.snopes.com/fact-check/trump-disinfectants-covid-19/

Electronic copy available at: https://ssrn.com/abstract=3623452