# Rumor Detection over Varying Time Windows

**Sejeong Kwon[1], Meeyoung Cha[1]\*, Kyomin Jung[2]**

**1** Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea, **2** Department of Electrical and Computer Engineering, Seoul National University, Seoul, Republic of Korea

\* meeyoungcha@kaist.ac.kr

## Abstract

This study determines the major difference between rumors and non-rumors and explores rumor classification performance levels over varying time windows—from the first three days to nearly two months. A comprehensive set of user, structural, linguistic, and temporal features was examined and their relative strength was compared from near-complete date of Twitter. Our contribution is at providing deep insight into the cumulative spreading patterns of rumors over time as well as at tracking the precise changes in predictive powers across rumor features. Statistical analysis finds that structural and temporal features distinguish rumors from non-rumors over a long-term window, yet they are not available during the initial propagation phase. In contrast, user and linguistic features are readily available and act as a good indicator during the initial propagation phase. Based on these findings, we suggest a new rumor classification algorithm that achieves competitive accuracy over both short and long time windows. These findings provide new insights for explaining rumor mechanism theories and for identifying features of early rumor detection.

## Introduction

Rumors are a powerful, pervasive, and persistent force that affects people and groups [1]. Interest in the psychology of rumors and their control has increased since World War II [2, 3], where these early studies relied on extensive yet manual data collection from books, newspapers, magazines, and interviews. Rumors have been described in numerous fashions, where the most well known definitions are 'public communications that are infused with private hypotheses about how the world works' [4] and 'ways of making sense to help us cope with our anxieties and uncertainties [5]. As these definitions suggest, rumors help members of a society learn about important issues by offering a collective problem-solving opportunity to individuals who participate.

The social media era has facilitated rumor propagation even further, as any piece of information now can be promoted by online users without censorship. False rumors or unverified information can spread more rapidly and widely because individuals receiving information through this medium lack the capacity to determine the veracity of the information they receive. Some rumors have mistakenly harmed reputation of individuals or organizations and this negative role has received much attention in both research and society [6]. With the

Internet and social media becoming a major information dissemination channel, rumor studies have started utilizing large-scale data [7–9] and to employ driven-intensive quantitative methods for rumor identification [10, 11]. User activity logs have been used to explain the evolutionary pattern of rumors under different conditions and to verify related social and psychological theories [12]. In addition to exploratory studies, researchers also have suggested various machine learning-based algorithms to develop rumor classifiers [13].

Most rumor classification studies are not free from two limitations. First is that the critical process of feature identification is driven by data rather than theoretical grounds, and hence there is no guarantee that findings apply for different data. To alleviate this problem, we investigated various rumor theories from social, psychological, and CMC studies and used them to derive features. While some of the rumor theories that we base on do not rely on data-driven analyses, they have been shown to be viable in explaining rumors on online communication platforms [14]. Second is that previous studies report findings based on a single snapshot of data, assuming that identified features are important throughout the rumor propagation phases. However, rumor propagation process is more complex and is known to show evolutionary patterns similar to group problem-solving activities [15]. Thus, it is more natural to assume that propagation mechanisms will change by observation periods—a characteristic that no previous data-driven study paid attention to. In the present study, we repeated statistical comparisons and classification experiments over different snapshots of 111 real rumor and non-rumor events to gain a comprehensive understanding.

Rumors in this study are defined such that a statement should (*i*) be unverified at the time of circulation and (*ii*) either remain unverified or be verified to be false after some time (i.e., at the time of this study). Hence a leak of information about later-confirmed facts (e.g., a merger or acquisition news that later gets confirmed) is not considered a rumor in this study, although its spreading pattern may look rumor-like during initial circulation. Our results provide new insight into understanding different rumor and non-rumor spreading mechanisms over time. Furthermore, we propose the best set of features and an algorithm for early rumor classification. This study makes use of longitudinal and near-complete data containing all public conversations of Twitter users over a 3.5-year period and their network structure. The data allow us to examine rumor propagation instances in the past. Rumor-related tweets were extracted from this data and examined over different time windows, form the initial 3 days to 7, 14, 28, and 56 days of circulation. Analyzing millions of rumor tweets with four sets of features from different perspectives help us quantify the relative effects of features, as follows:

The first set of features is about user themselves. Theory-driven rumor literature notes that 'rumor spreaders are persons who want to get attention and popularity' [16]. This hypothesis has inspired data-driven studies to assume that rumormongers have special characteristics in terms of their follower count, friend or following count, and the tweet count of users [17]. Expanding on these metrics, we consider a thorough set of the percentile, mean, variance, skewness, and kurtosis values in this research. The user-related features alone can differentiate rumors with F1 scores of 0.76 at an early stage of propagation (e.g., the first three days). Nevertheless, the classification performance decreases when the observation period becomes longer.

The second set of features is the linguistic characteristics of the rumor messages. Here, we compile an extensive list of linguistic features from previous studies and test them in rumor exploration and classification. Previous studies in the Natural Language Processing (NLP) domain have developed techniques to count the frequency of question marks, hashtags, dimensions of sentiment (e.g., positive, negative), and occurrences of arbitrary words [18, 19]. Past studies in the psycholinguistics domain on the other hand have examined the mapping between words and their semantic meanings such as social and psychological processes (e.g., tentative words such as 'maybe', 'perhaps', etc.) [20, 21]. We utilize all of these linguistic

features in the context of rumors to understand why and how rumors spread. The training result demonstrates that the linguistic features are robust throughout all observation periods with F1 scores of 0.8-0.86.

The third set of features consists of structural properties of the rumor diffusion network. Networks have often been considered a key data type in general cascade studies [22–24] as well as in rumor studies [25–28]. We consider three types of propagation structures—the extended network, the friendship network, and the diffusion network—and estimate key measures such as network density and clustering coefficient. Considering these various network structures is important for obtaining features of information propagation and many complex network studies have introduced new adequate structures from theory, observation, and numerical analysis such as Fast Fourier Transformation (FFT) and the Finite Element Method (FEM) [29–31]. The diffusion network that is of the smallest in size is simple yet effective in capturing rumor information flows among users. This network helps us obtain information about nonreciprocal interactions among users. Classification result demonstrates that the network features do not perform well during initial circulation period, whereas its F1 score improves to 0.79 for a longer time duration.

Finally, we consider temporal features extracted from daily time series of the number of tweets as parameters of an epidemic model. Epidemic models have been popularly used to describe information adoption process due to their similarity in mechanisms. Among them are practical models suitable to describe non-uniform transmission and disease interactions [32, 33]. Recent models further consider new features like awareness, homophily, and diverse activity patterns with multiplexes to explain cooperative actions [34–36]. In the same manner, we identified major temporal features of each rumor (such as the natural propagation cycle) based on the Periodic External Shock (PES) epidemic model proposed in earlier work [37]. While temporal features are found to be the most predictive for longitudinal data (with an F1 score 0.88 for the 56-day window), they become unavailable when observation window is limited to the initial few days of circulation.

## Materials and Methods

We begin by discussing the methodology to extract rumor and non-rumor events for analysis. The process of feature extraction based on data observations and psychological theories about rumor spreading are introduced.

### Rumors on Twitter

The massive data logged in social media provide an opportunity to study rumors quantitatively. One of the largest and near-complete repositories of Twitter is from [38], where the authors crawled the network and took a snapshot of the 54 million public profiles and 1.9 billion follow links among them in August 2009. The repository also contains the 3200 most recent tweets of every individual at the time (which nearly covers the complete tweet history of the individuals, since the launch of Twitter in 2006) accounting a total of 1.7 billion tweets. This longitudinal data were used to extract popular rumors between 2006 and 2009, which reveal various features such as the user-level, structural, temporal, and linguistic characteristics at the time of circulation.

To select rumor cases, we searched for two famous rumor archives, snopes.com and urban-legends.about.com and identified popular rumors at the time of the dataset. For non-rumor cases, we searched for notable events from news media outlets like times.com, nytimes.com, and cnn.com. We identified a total of 130 events (72 rumors and 58 non-rumors) from the time period covered by the Twitter dataset and extracted tweets corresponding to these events

based on two criteria: (1) a tweet should contain explicit keywords relevant to the event and (2) a tweet should have been posted during the time of circulation (i.e., within the first two months).

To test whether tweets identified in the above manner are indeed relevant to the event of interest, we hired four human labelers to evaluate the dataset. The labelers were asked to judge whether each event was rumor or non-rumor by examining four randomly selected tweets and URLs embedded in tweets for each event. Then, we selected the events that were evaluated by four participants and had the majority agreement for this study. The final set of rumor and non-rumor events had agreement among three or more labelers. The intra-class correlation coefficient (ICC), which measures the level of agreement, was 0.993 and the p-value was close to zero. As a result, 111 events were retained (60 rumors and 51 non-rumors). S1 and S2 Tables list the selected rumors and non-rumors, respectively.

The key contribution of this paper is at examining rumor characteristics over different observation time windows, from the initial 3, 7, 14, and 28 to 56 days of circulation. For each of these time windows, we repeatedly estimated features from user, linguistic, network, and temporal characteristics. Once all feature values were obtained, we traced the changes in the statistical differences and predictive powers of these features in distinguishing rumors from non-rumors.

## User Feature

The question of who spreads rumors has been investigated extensively in previous qualitative studies through interviews and user surveys. A particular focus has been on user influence [39]. Because rumors have the power to arouse curiosity, individuals who seek attention can exploit rumor rhetoric [16]. In contrast, the low credibility of rumors can hurt reputation individuals who participate in the discourse and hence highly influential individuals may try to avoid participating in rumor conversations [37]. These findings hence suggest that rumor spreaders typically involve individuals with small or moderate influence on social networks.

We consider the number of followers, friends, and tweets as a proxy of user influence. Fig 1 illustrates the distributions of these three quantities aggregated over rumor and non-rumor events, respectively. The figure shows that, across the majority of quantile values, user features for non-rumor events are higher than those of rumor events, suggesting that user features can be used for determining rumors from non-rumors. In order to represent this difference, we approximated the rumor distributions at the 25th, 50th, 75th, and 100th percentiles, standard
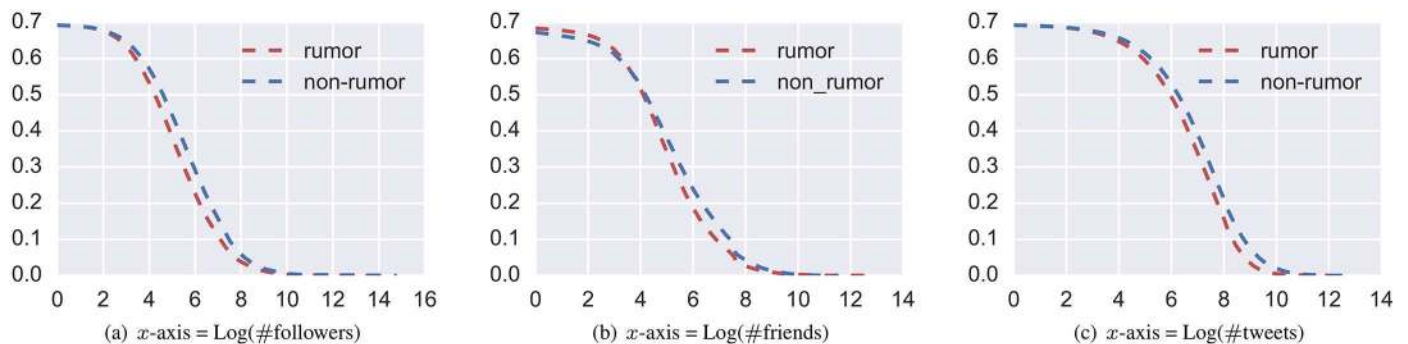


**Fig 1. Log-log scale of the CCDF.** Complementary Cumulative Distribution Function (CCDF) of aggregated user characteristics for rumor and non-rumor events in the 56-day observation.

doi:10.1371/journal.pone.0168344.g001

deviation, skewness, and kurtosis values and utilize these for rumor classification. S3 Table lists the 27 variables for describing the user characteristics for each event.

## Linguistic Feature

Previous studies that develop linguistic features have primarily focused on building classifiers of high accuracy but have not provided supportive theories regarding why the proposed features work [7, 17, 40]. In contrast, we start from theory-based rumor research to define relevant linguistic features that are missed in data-driven studies. Psychological theories can explain why and how people react to a given rumor [39, 41]. For example, upon receiving a rumor, a person will first raise a doubt, judge its meaning with her knowledge, then check with existing sources to verify its veracity [42]. This process of doubt ends when a person gathers sufficient evidence, at which point the user either accepts the rumor by propagating it further or rejects it by either ignoring it or expressing negating comments. We define features of doubt, negation, and guessing to describe this process.

Reputation is also known to give people a strong incentive to conform to social norms in this process [43]. Low credibility of rumors and doubts incurred by rumor audience will result in a different writing style of rumor conversations compared to conversations on non-rumors. For example, rumors are more likely to contain words related to cognition (e.g., skepticism, guessing) than non-rumors. Furthermore, because rumors are known to propagate among social peers via word-of-mouth (than involving authorities as in non-rumors) [5], they are more likely to referring to language on social relationships (e.g., family, mate, friend), and actions like hearing. We also define features related to social processes and cognition.

Below are example rumors on a fake celebrity death news that use such language:

*(Tweet 1) Emma Watson died in a car accident last night, but I'm not gonna believe it unless BBC reports it. It's all REALLY fishy.*

*(Tweet 2) Nope, it's supposedly a hoax because they were claiming that Harrison Ford had died as well. Not true.*

As seen in these examples, words related to skepticism ('hoax'), negation ('not true'), and doubt ('not gonna believe') appear. Rumor tweets are sometimes written in capital letters as shown above, which is a way of shouting or adding emphasis online. Below is another type of rumor regarding health issues, which also uses capital letters and a sign of doubt. The rumor below also confirms another theory about rumors, which is unclear information sources ('heard on').

*(Tweet 3) Just heard on the radio. . ... There is NO link between breast cancer and deodorants.*

We utilize a sentiment analysis tool, Linguistic Inquiry and Word Count (LIWC), which was built to identify a group of words that have basic emotional and cognitive dimensions that are often studied in social, health, and personality psychology. Its dictionary has approximately 4,500 words and word stems each of which is assigned to several predefined categories. In order to compute sentiment scores for given text, each word in the text is compared with the LIWC dictionary. Considering the example in Tweet 1, the program would find the word 'not' and verify whether this word is in the negation category. Then, the score of the negation category will increase. If the score of the negation category is 2.34, this means that 2.34% of all words in the given text matched with the negation category. There are five major categories and a number of subcategories about psychological processes (e.g., social, affect, cognitive,

perceptual, tentative, swearing) and writing patterns (negation, home, tense), which satisfy the requirements to examine the hypotheses in this study.

Prior to analysis, we sanitized the tweet data by removing '@username', hashtags, short URLs, as well as any emoticons and special characters that the LIWC tool cannot parse. The tool's guideline suggests to ensure at least 50 words for analysis. Given that individual tweet is short in length, we grouped tweets belonging to the same event into a single input file and measured the LIWC sentiment scores collectively for each event.

## Network Feature

Network representation enables the entire propagation pattern of information and interactions among users to be visualized [12, 44]. We defined three kinds of structures—the extended network, the friendship network, and the diffusion network.

The extended network is defined as an induced directed subgraph of users, who have either posted any tweet related to the event or who is following or is followed by other participating users, from the original follower-followee relationship on Twitter. The friendship network is a smaller induced subgraph of the extended network that only contains users who have tweeted content related to the event and the links among such users. The diffusion network is a subgraph of the friendship network, which depicts the direction of information flow among users. Information flows are defined such that a event is diffused from user A to user B, if and only if (1) B follows A on Twitter and (2) B posted about the given event by mentioning the appropriate keywords only after A did so. When a user has multiple possible sources of flows, we select the user who most recently posted about the event as the source. Then, we retain the links of the selected source and the target in a diffusion network, which hence captures information flows. Fig 2 presents the diffusion networks of a rumor and non-rumor events, respectively. These networks were extracted from the 56-day observation period.

Network measures were computed from the three kinds of networks explained above. In addition to the traditional measures such as density and clustering coefficient, we also considered measures that are specific to rumor propagation. For instance, rumors are known to spread without strong evidence and have low response rate, in particular when many rumor receivers decide not to participate in the conversation [43]. Such process can be captured by the count or the fraction of singletons in the friend network. Consequently, each network and its largest connected component have an identical set of features, such as the average clustering coefficient, density, the number of nodes, the number of edges, and the fraction of users with nonreciprocal interactions.

Not only standard metrics but also rumor-specific measures (inspired by theoretical studies on rumors) were defined for this study. For these measures we rely on the diffusion network, whose direct links represent inferred information flows. One example is the direction of information flow. In rumor propagation, influential users are known to restrain themselves from participating; hence initial rumor conversations will begin from individuals with relatively low or medium influence and then ultimately reach a wider set of the networks [37]. This process was captured by examining, for each information flow instance, whether the direction of flow started from a lower-degree user to a higher-degree user. S4 Table presents the comprehensive list of 43 network features that were analyzed.

## Temporal Feature

A rumor characteristic that has received recent attention is the temporal aspect. Fig 3 depicts several time series of rumor and non-rumor events, which shows multiple and periodic spikes for rumor events. In contrast, most non-rumor events appear with a single prominent spike.

(a) Deodorant can cause cancer. (rumor)          (b) John Hartson has brain cancer. (non-rumor)
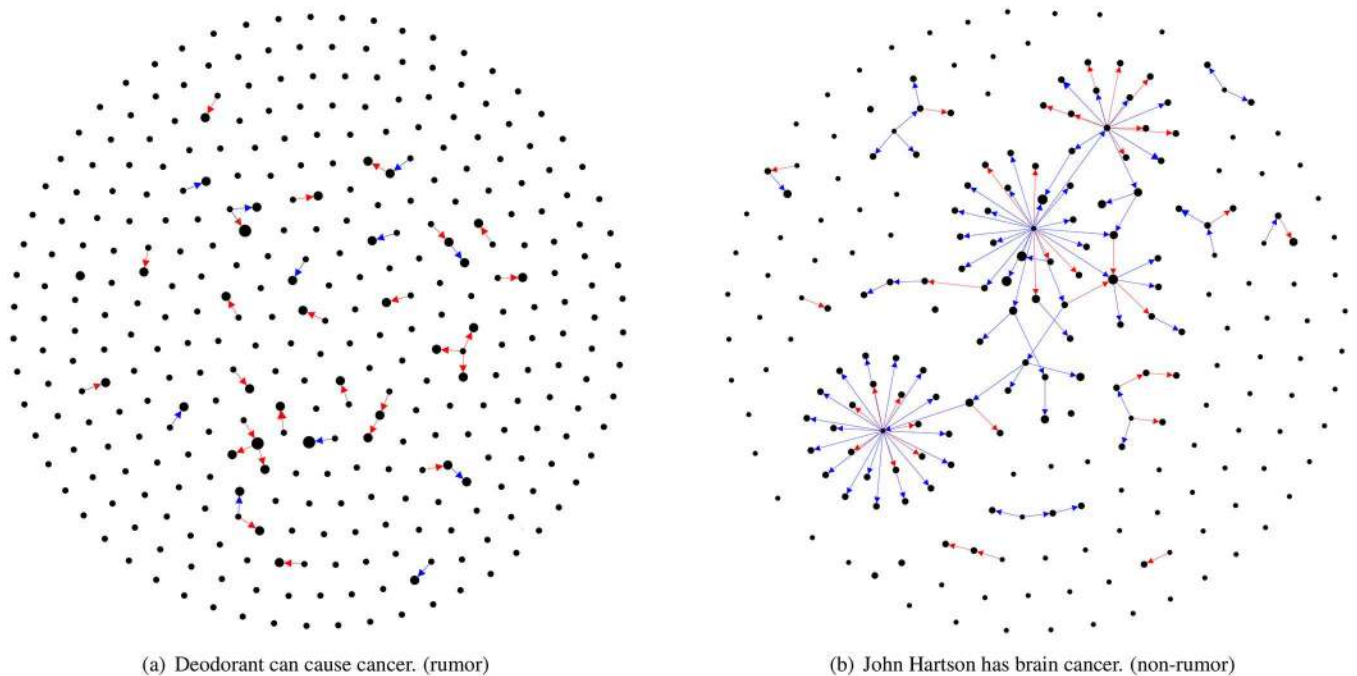
**Fig 2. Diffusion network examples.** The network visualization in (a) shows that rumors involve a larger fraction of singletons and smaller communities, resulting in a sporadic diffusion pattern. In contrast, the diffusion network of a non-rumor event in (b) is highly connected, forming a giant connected component and a smaller fraction of singletons. Edge colors represent the relative influence of the spreader and recipient, such that red (blue) means information propagated from a lower-degree (higher-degree) spreader to a higher-degree (lower-degree) recipient.
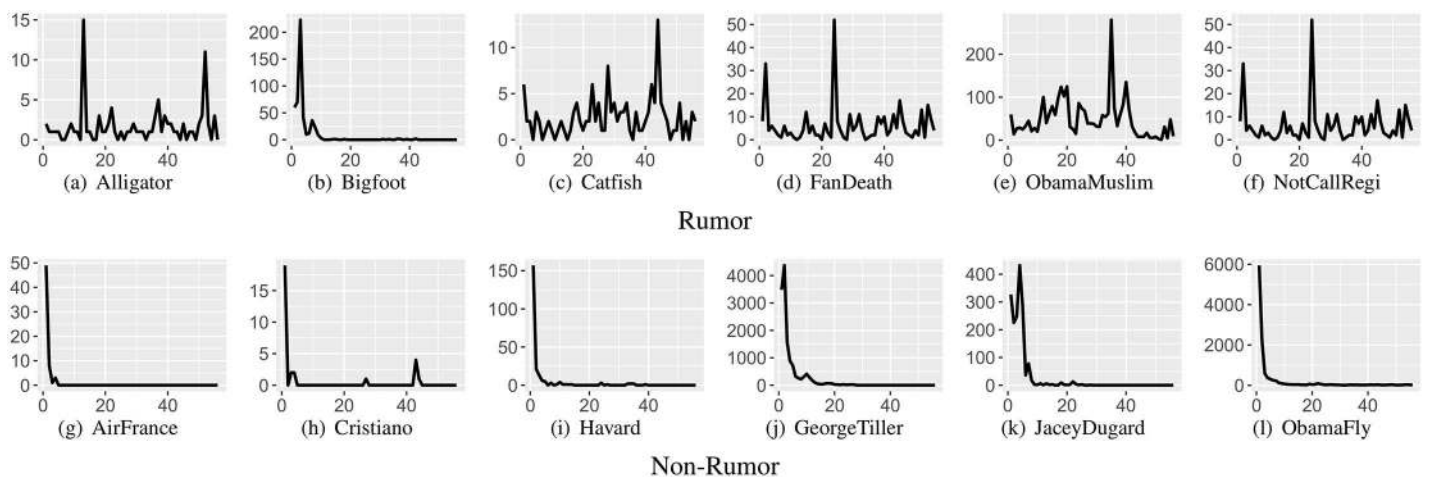
**Fig 3. Samples of extracted time series.** The time series are extracted from 56-day observation period($x$-axis = days; $y$-axis = number of tweets). Rumors typically have longer life spans and more fluctuations.

While non-rumor events typically show a large peak during the initial phase of circulation that decays quickly over time, rumor events are known to show repeated peaks over time that lead to a cyclic trend. Such cyclic behavior can be captured mathematically by the Periodic External Shock (PES) model [37], which is our previous work that is an extension of the time-series fitting model proposed in [45]. Here, we describe how we derived the PES model from the base model. The base model, called SpikeM, is from the Susceptible-Infected (SI) model in epidemiology and captures both the periodic behaviors and power-law decays that are popularly observed in real-world data. The SpikeM model is formally described as below.

*SpikeM with parameters* $\theta = \{N, \beta, n_b, S_b, \epsilon, p_p, p_a, p_s\}$:

$$U(n + 1) = U(n) - \Delta B(n + 1),$$

$$\Delta B(n + 1) = p(n + 1) \cdot \left[ \frac{\beta}{N} \cdot U(n) \cdot \right.$$

$$\left. \sum_{t=n_b}^{n} (\Delta B(t) + S(t)) \cdot (n + 1 - t)^{-1.5} + \epsilon \right] \tag{1}$$

$$p(n) = 1 - \frac{1}{2} p_a \left[ 1 + \sin \left( \frac{2\pi}{p_p} (n + p_s) \right) \right],$$

$$S(t) = S_b \text{ when } t = n_b; \text{ otherwise } 0.$$

$$U(0) = N, \Delta B(0) = 0.$$

In the model, $U(n)$ is the number of uninfected (or uninformed) nodes in the network at time step $n$; $\Delta B(n)$ is the newly infected (or informed) nodes at time $n$; $N$ is the total number of nodes involved in the diffusion process; $n_b$ is the time when the first external shock of the event occurs, and $S_b$ is the scale of the first external shock, i.e., the number of nodes that are infected at the beginning of the event at time $n_b$. Thus, the term $\sum_{t=n_b}^{n} (\Delta B(t) + S(t))$ represents the total number of infected nodes at time $n$. Then $\Delta B(n + 1) = \frac{\beta}{N} \cdot U(n) \cdot \sum_{t=n_b}^{n} (\Delta B(t) + S(t))$ is the standard SI model describing that at time $n + 1$ each uninfected node $u$ randomly selects another node $v$ from all nodes and if $v$ is already infected; $u$ becomes infected at time $n + 1$ with the probability $\beta$, which is a parameter for infection strength.

The SpikeM model extends the SI model through introducing (a) a power-law decay term $(n + 1 - t)^{-1.5}$ in Eq (1) so that the strength of the infection of previously infected nodes becomes weaker in a power-law decay pattern, and (b) a periodic interaction function $p(n)$ that reflects people's periodic interaction patterns (e.g. people may have more time to interact on Twitter in the evening than during the day when they are at work or school). Parameters $p_p$, $p_a$, and $p_s$ correspond to the period, amplitude, and phase shift of the periodic interaction function, respectively. Finally parameter $\epsilon$ represents a noise term in the model.

While the SpikeM model can adequately capture any periodic or seasonal patterns in real-world data, the model is not directly applicable for rumor analysis. This is because the multiple peaks seen in rumors are not due to people's natural tendency to repeat behaviors daily or weekly (shown as the periodic interaction function $p(n)$), but is more irregular and may be driven by a number of external factors. For example, peaks may appear when a rumor reaches a new set of audience who were not exposed to the same information previously or when a group of rumor mongers decide to spread information further.

We hence extended the SpikeM model and proposed the Periodic External Shocks (PES) model, which newly assumes there may be *multiple* external shocks and each shock will show a

power-law decay over time in [37]. For simplicity, external shocks are assumed to have a short periodic cycle. Then the PES model is described as below:

PES model with parameters $\theta' = \{N, \beta, n_b, S_b, \epsilon, p_p, p_a, p_s, q_p, q_a, q_s\}$:

$$\Delta B(n+1) = p(n+1) \cdot \left[ \frac{\beta}{N} \cdot U(n) \cdot \right.$$

$$\left. \sum_{t=n_b}^{n} (\Delta B(t) + \bar{S}(t)) \cdot (n+1-t)^{-1.5} + \epsilon \right], \tag{2}$$

$$\bar{S}(t) = S(t) + q(t),$$

$$q(t) = q_a \left[ 1 + \left( \sin \left( \frac{2\pi}{q_p} (t + q_s) \right) \right) \right],$$

All other terms are the same as in SpikeM.

The PES model has a periodic external shock function $q(t)$, which has three parameters $q_p$, $q_a$ and $q_s$ to represent the period, amplitude, and shift of the periodic external shock function, respectively. When $q_a = 0$, the PES model is reduced to SpikeM and hence can be considered a generalization of the SpikeM model. Initial shock function is termed $S(t)$. We consider the parameters of the PES model as temporal features of rumors. For parameter learning, the Levenberg-Marquard method [46] was used to minimize the sum of the squared errors: $D(X, \theta) = \sum_n (X(n) - \Delta B(n))^2$ with a given time series $X(n)$. S5 Table displays the final set of temporal features.

## Results

### Changes in significance over time

In order to identify the significant differences between rumors and non-rumors for the first 3, 7, 14, 28 and 56 days from the initiation, Mann-Whitney U test is used. Tables 1–4 list the user, network, linguistic and temporal features with constant tendencies over these observation periods.

Table 1 focuses on the significant features among the user characteristics. Considering the number of followers and tweets, non-rumors had higher values for all quantiles except the minimum. This can be interpreted that users participating in non-rumor spreading are more active tweeters and have larger audiences. Kurtosis and skewness of the number of followers are also higher for non-rumors, which indicates that non-rumors are more likely have users with extraordinarily larger audiences. These observations lead to a conclusion that users who get more attention on Twitter (i.e., have more followers) are less likely to participate in rumor spreading. This observation provides a meaningful insight that the high audience size of users might hinder a user from participating in rumor propagation in addition to the existing findings about effects of audience size from other information propagation studies [38].

The minimum number of friends ($m_{fr}$) is the only a feature, whose values turn out higher for rumors than non-rumors. For validation, we examined which individuals had the smallest number of friends for each event. We found non-rumor events to involve news media accounts like 'BBCNewsOnceADay', which have large followers and disproportionately small followings (i.e., friends). At the time of circulation, the existence of such media with a small number of friends and a large number of followers creates the particular propagation patterns that rumors have higher minimum number of friends but lower values for the others.

**Table 1. User features with invariant tendency during the observation periods.**

| Symbol | Description | Class | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 7 | 14 | 28 | 56 |
| $m_{fr}$ | Minimum of number of friends | R | R | R | R | R |
| $q75_{fr}$ | 75th percentile of number of friends | N | N | N | N | |
| $M_{fr}$ | Maximum of number of friends | N | | N | N | N |
| $\sigma_{fr}$ | Standard deviation of number of friends | N | N | N | N | N |
| $q25_{fo}$ | 25th percentile of number of followers | N | N | N | N | N |
| $med_{fo}$ | median of number of followers | N | N | N | N | N |
| $q75_{fo}$ | 75th percentile of number of followers | N | N | N | N | N |
| $M_{fo}$ | Maximum of number of followers | N | N | N | N | N |
| $\mu_{fo}$ | Average of number of followers | N | N | N | N | N |
| $\sigma_{fo}$ | Standard deviation of number of followers | N | N | N | N | N |
| $\kappa_{fo}$ | Kurtosis deviation of number of followers | N | N | N | N | N |
| $S_{fo}$ | Skewness of number of followers | N | N | N | N | N |
| $q25_{t}$ | 25th percentile of number of tweets | N | N | N | N | N |
| $med_{t}$ | median of number of tweets | N | N | N | N | N |
| $q75_{t}$ | 75th percentile of number of tweets | N | N | N | N | N |
| $M_{t}$ | Maximum of number of tweets | N | N | N | N | N |
| $\mu_{t}$ | Average of number of tweets | N | N | N | N | N |
| $\sigma_{t}$ | Standard deviation of number of tweets | N | N | N | N | N |

In the Class column, fields denoted with R (or N) indicate that the target feature over the specific observation period had higher value for rumors (or non-rumors) at a significant level of $p < 0.05$ based on the Mann-Whitney U test. A Blank field indicates that there was no significance in correlation. Results are shown for observation periods of 3, 7, 14, 28, and 56 days.

doi:10.1371/journal.pone.0168344.t001

Table 2 supports the assumptions about the different cognitive and writing styles for rumormongers described in the social and psychological theories. According to Table 2, it can be seen that rumors have more words related to a certain cognitive style such as skepticism. Unlike other types of information, users who are more likely spread not only the story but also their opinion and thinking about a rumor. New patterns began to appear after 7 days. Users began to mention social relationships (e.g, mate, friend) as they referred to information sources and hearing actions (e.g., heard) in rumor conversations. We also see features with unexpected patterns. Words in assent (e.g., agree, okay), swear (e.g., damn, piss), and certain (e.g., always, never) category become significant after 14 days.

Fig 4 explains why these changes occurred. As seen in Fig 4, the reactions to rumors became more confident. The first reaction is strong denial expressed by a high correlation between negation and certain. Next, simple agreement is described by the sole existence of assent category. This result can be one of the additional supportive result for Bordia's work [15] that people express fears and anxieties or hopes and wishes in the context of the rumors.

From Table 3, counting the measures of the introduced networks such as number of nodes, edges, and certain types of user exhibited significantly higher values for non-rumors. This can be explained from the previous observation that rumor spreaders apparently have fewer followers. For non-counting measures, density of the largest connected component (LCC), proportion of users without followers (i.e., singletons), friends, or both are higher for rumors. From these findings, three propositions can be inferred: i) a rumor is more likely to be obtained from an external source, ii) a rumor is more likely to be ignored on social media even

**Table 2. Linguistic features with invariant tendencies during the observation periods.**

| Symbol | Description | Class | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 7 | 14 | 28 | 56 |
| family | family (daughter, father, husband, aunt) | N | N | N | N | N |
| i | 1st person singular (I, me, mine) | R | R | R | R | R |
| you | 2nd person (you, your, thou) | R | R | R | R | R |
| conj | conjunctions (and, but, whereas, although) | R | R | R | R | R |
| present | present tense (is, does, hear) | R | R | R | R | R |
| auxverb | auxiliary verbs (am, will, have) | R | R | R | R | R |
| discrep | discrepancy (would, should, could) | R | R | R | R | R |
| adverb | adverb (very, really, quickly) | R | R | R | R | R |
| excl | exclusive (but, without, exclude) | R | R | R | R | R |
| cogmech | cognitive mechanism (cause, know, ought) | R | R | R | R | R |
| negate | negations (not, no, never) | R | R | R | R | R |
| tentat | tentative (maybe, perhaps, guess) | R | R | R | R | R |
| assent | assent (agree, OK, yes) | | R | R | R | R |
| certain | certain (always, never) | | R | R | R | R |
| social | social processes (mate, talk, they, child) | | | R | R | R |
| swear | swear words (damn, piss, f*ck) | | | R | R | R |
| hear | hear (listen, hearing) | | | R | R | R |

The 'Description' column lists example words for each symbol.

**Table 3. Network features with invariant tendencies during the observation periods.**

| Symbol | Description | Class | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 7 | 14 | 28 | 56 |
| $V_e$ | Number of nodes of extended network | N | N | N | N | N |
| $E_e$ | Number of edges of extended network | N | N | N | N | N |
| $NI_e$ | Number of nodes without incoming edge in extended network | N | N | N | N | N |
| $NO_e$ | Number of nodes without outgoing edge in extended network | N | N | N | N | N |
| $D_{el}$ | Density of LCC of extended network | R | R | R | R | R |
| $E_{el}$ | Number of edges of LCC of extended network | N | N | N | N | N |
| $V_f$ | Number of nodes of friendship network | N | N | N | N | N |
| $E_f$ | Number of edges of friendship network | N | N | N | N | N |
| $NO_f$ | Number of nodes without outgoing edge in friendship network | N | N | N | N | N |
| $D_{fl}$ | Density of LCC of friendship network | R | R | R | R | R |
| $E_{fl}$ | Number of edges of LCC of friendship network | N | N | N | N | N |
| $pNI_f$ | Proportion of nodes without incoming edges of friendship network | R | R | R | R | R |
| $pNO_f$ | Proportion of nodes without outgoing edges of friendship network | R | R | R | R | R |
| $pI_f$ | Proportion of isolated nodes of friendship network | R | R | R | R | R |
| $D_{dl}$ | Density of LCC of diffusion network | R | R | R | R | R |
| $V_{dl}$ | Number of nodes of LCC of diffusion network | N | N | N | N | N |
| $E_{dl}$ | Number of edges of LCC of diffusion network | N | N | N | N | N |
| $pNI_d$ | Proportion of nodes without incoming edges of diffusion network | R | R | R | R | R |
| $pNO_d$ | Proportion of nodes without outgoing edges of diffusion network | R | R | R | R | R |
| $pI_d$ | Proportion of isolated nodes of diffusion network | R | R | R | R | R |

**Table 4. Temporal features with invariant tendencies during the observation periods.**

| Symbol | Description | Class | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 7 | 14 | 28 | 56 |
| $q_p$ | Periodicity of external shock | | | | N | N |
| $q_s$ | External shock periodicity offset | | | N | N | N |
| $p_s$ | Interaction periodicity offset | | | N | N | N |

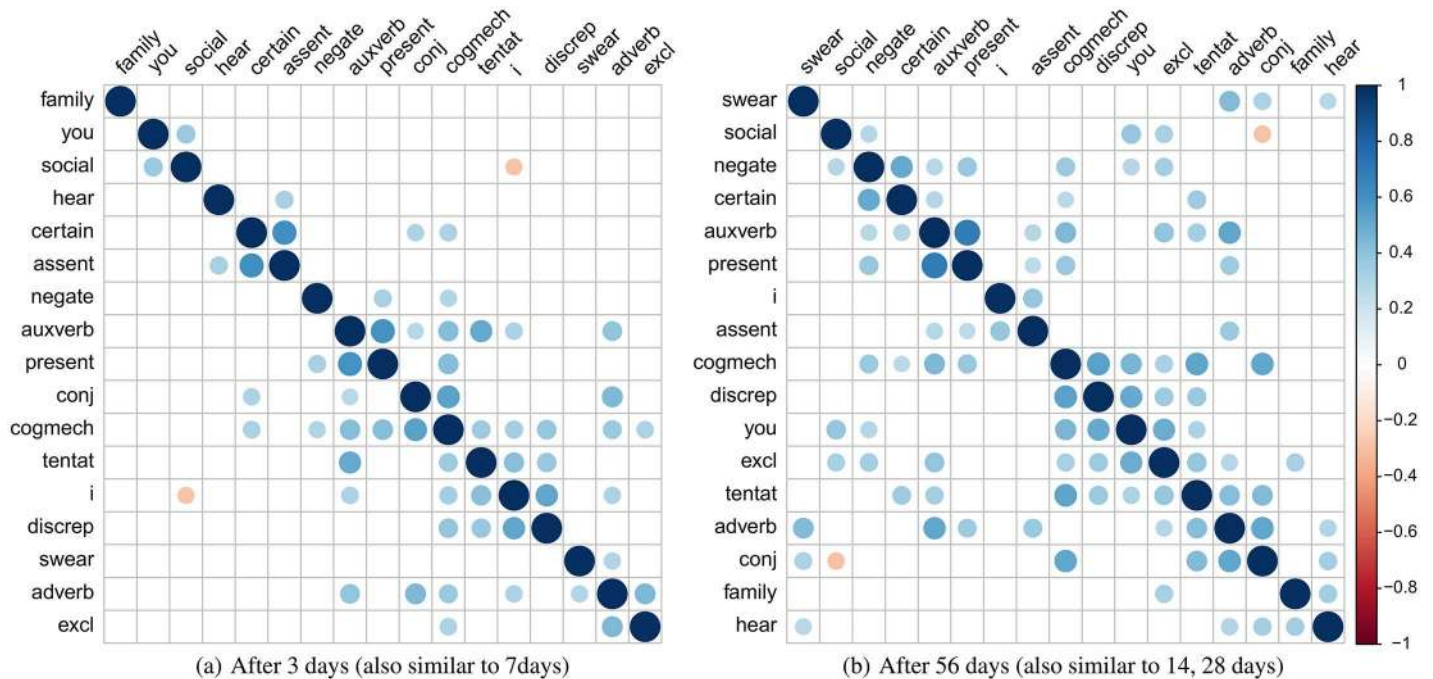doi:10.1371/journal.pone.0168344.t004



**Fig 4. Correlogram among the linguistic features.** These plots show correlations among linguistic features with significance for rumor events. Colors of circles represent correlation coefficients, where the dark blue (red) color indicates coefficients of 1 (-1). Size of circles represent the absolute values of correlation coefficients. Blank or no circle indicates non-significant cases where p-values $\geq$ 0.05. From (b), strong negation is expressed by a positive correlation between scores of 'negation' and 'certain'. Interestingly, the 'assent' category increases for rumors, which indicates that users who confirmed the rumor also appeared as time passed.

doi:10.1371/journal.pone.0168344.g004

when a friend mentions it, and iii) the community structure of rumor spreaders is smaller but denser in general than those of non-rumors.

Table 4 shows that there is no temporal feature with a significant difference at the initial time. Nonetheless, for 14 days and longer duration, temporal feature about periodicity becomes significant and exhibit very high prediction performance (especially for the 56-day period), which we will discuss in more detail. This phenomenon is expected since the proposed model contains two trigonometric functions to explain the information propagation and as a result short observation periods like 3 or 7 days are not sufficiently long enough to fit the temporal model.

## Predictive Features and Classification

Machine learning was used to test the significance in differences between the rumor and non-rumor events based on the theory-driven features discussed in the previous section. A three-

step variable selection process proposed in [47] was applied to identify predictive features and build a classifier. This process works as follows: first, eliminating the irrelevant variables with a small permutation importance index, second, selecting all variables related to the response for interpretation purpose (i.e., interpretation set), and third, refining the selection through eliminating redundancy in the set of variables selected in the second step, for prediction purpose (i.e., prediction set). We repeated the process 10 times and used features that appeared more than eight times in order to prevent inadvertent selection of irrelevant features due to the permutation. Hence, features that are robust discriminating factors can be selected for rumor classification.

The feature selection process was repeated for the user, linguistic, network, and temporal domains in order to compare the prediction power of each feature. With the features in the prediction sets, we applied a random forest with three-fold cross-validation. Given that the size of events (111 rumor and non-rumors) can be insufficient to conclude the prediction performance, the comparison process was repeated ten times. Fig 5 illustrates the average classification performance of each set of features according to time. Combining Fig 5 and Table 5, following observations are noted. The final outcome is summarized in Table 5, which lists the selected features for the interpretation and prediction task. We make the following observations.

First, user features are only predictive during the initial rumor circulation. The observation that only the standard deviation matters for the classification could indicate that the existence of users with extraordinarily high numbers of followers and friends, i.e. outliers, is the most important factor in identifying rumors. On the other hand, distributions for rumors and non-rumors related to the user characteristic become similar over time as more users participate in the information propagation process. For example, if influential users mention rumors about the death of a famous person or an urgent health issue, they later deny them with confirmatory
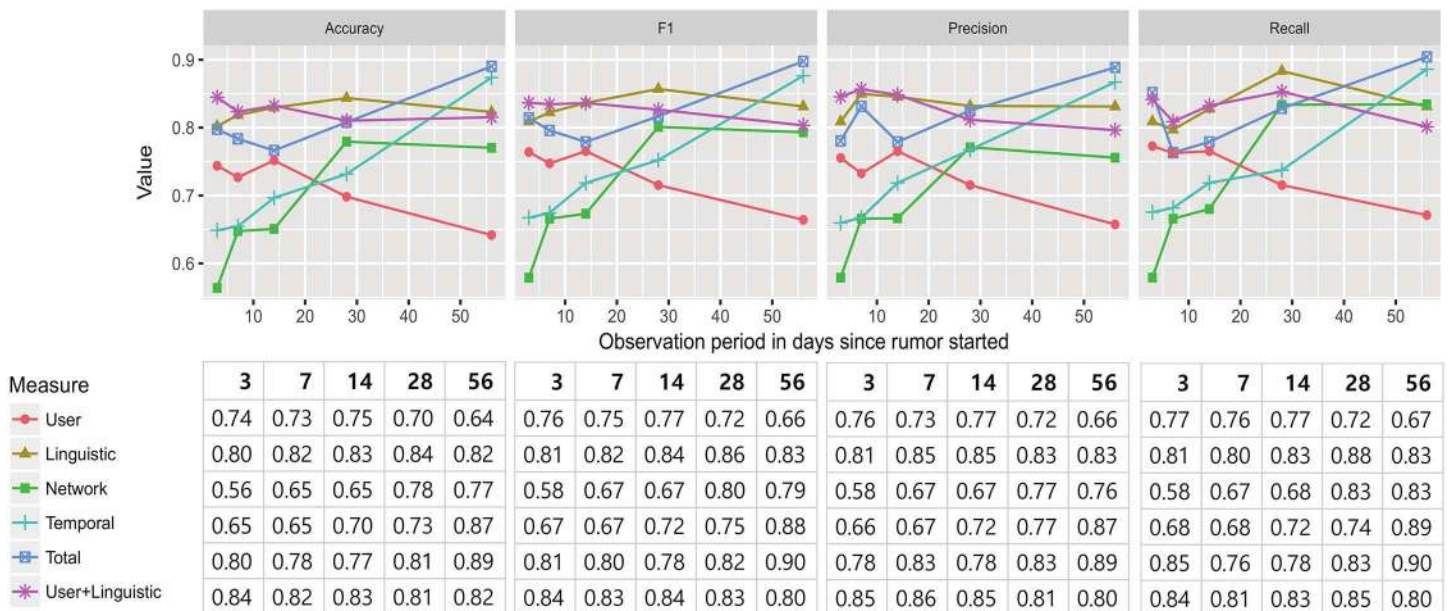


| Measure | Accuracy | | | | | F1 | | | | | Precision | | | | | Recall | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 7 | 14 | 28 | 56 | 3 | 7 | 14 | 28 | 56 | 3 | 7 | 14 | 28 | 56 | 3 | 7 | 14 | 28 | 56 |
| User | 0.74 | 0.73 | 0.75 | 0.70 | 0.64 | 0.76 | 0.75 | 0.77 | 0.72 | 0.66 | 0.76 | 0.73 | 0.77 | 0.72 | 0.66 | 0.77 | 0.76 | 0.77 | 0.72 | 0.67 |
| Linguistic | 0.80 | 0.82 | 0.83 | 0.84 | 0.82 | 0.81 | 0.82 | 0.84 | 0.86 | 0.83 | 0.81 | 0.85 | 0.85 | 0.83 | 0.83 | 0.81 | 0.80 | 0.83 | 0.88 | 0.83 |
| Network | 0.56 | 0.65 | 0.65 | 0.78 | 0.77 | 0.58 | 0.67 | 0.67 | 0.80 | 0.79 | 0.58 | 0.67 | 0.67 | 0.77 | 0.76 | 0.58 | 0.67 | 0.68 | 0.83 | 0.83 |
| Temporal | 0.65 | 0.65 | 0.70 | 0.73 | 0.87 | 0.67 | 0.67 | 0.72 | 0.75 | 0.88 | 0.66 | 0.67 | 0.72 | 0.77 | 0.87 | 0.68 | 0.68 | 0.72 | 0.74 | 0.89 |
| Total | 0.80 | 0.78 | 0.77 | 0.81 | 0.89 | 0.81 | 0.80 | 0.78 | 0.82 | 0.90 | 0.78 | 0.83 | 0.78 | 0.83 | 0.89 | 0.85 | 0.76 | 0.78 | 0.83 | 0.90 |
| User+Linguistic | 0.84 | 0.82 | 0.83 | 0.81 | 0.82 | 0.84 | 0.83 | 0.84 | 0.83 | 0.80 | 0.85 | 0.86 | 0.85 | 0.81 | 0.80 | 0.84 | 0.81 | 0.83 | 0.85 | 0.80 |

**Fig 5. Comparison of the strength of the features in determining rumors.** Total and User+Linguistic are the newly proposed rumor classification algorithms in this study.

doi:10.1371/journal.pone.0168344.g005

**Table 5. Selected features.**

| Symbol | Description | Selection | | | | |
|---|---|---|---|---|---|---|
| | | 3 | 7 | 14 | 28 | 56 |
| $\sigma_{fr}$ | Standard deviation of number of friends | ** | | | | |
| $\sigma_{fo}$ | Standard deviation of number of followers | ** | | | | |
| $\mu_{fo}$ | Average of number of followers | ** | | | | |
| i | 1st person singular (i, me, mine) | | | | * | * |
| conj | conjunctions (and, but, whereas, although) | | | ** | ** | ** |
| auxverb | auxiliary verbs (am, will, have) | | ** | | ** | ** |
| adverb | adverb (very, really, quickly) | | ** | * | ** | ** |
| excl | exclusive (but, without, exclude) | ** | * | ** | ** | ** |
| cogmech | cognitive mechanism (cause, know, ought) | | | * | ** | ** |
| affect | affective processes (happy, cried, abandon) | | ** | | * | |
| negate | negations (not, no, never) | ** | ** | ** | ** | ** |
| tentat | tentative (maybe, perhaps, guess) | ** | * | ** | ** | ** |
| certain | certain (always, never) | | | | ** | ** |
| hear | hear (listen, hearing) | | | | * | ** |
| $E_{el}$ | Number of edges of LCC of extended network | | | ** | ** | * |
| $AC_{fl}$ | Average clustering coefficients of LCC of friendship network | | | | * | * |
| $pI_d$ | Proportion of isolated nodes of diffusion network | | | | | ** |
| pLTH | Fraction of LTH among information diffusion | | | | ** | ** |
| pHTL | Fraction of HTL among information diffusion | | | | ** | ** |
| $S_c$ | Strength of external shock at birth | | | | * | ** |
| $q_p$ | Periodicity of external shock | | | | | ** |
| $q_s$ | External shock periodicity offset | | | | | ** |
| $p_s$ | Interaction periodicity offset | | | | ** | * |

This table lists the features that are selected as prominent differentiating ones among all features. In the Selection column, * and ** signs indicate that the target feature was selected for the interpretation set and prediction set. If a feature is selected for the prediction set, it is also in the interpretation set by definition. A blank field indicates that the corresponding feature was not selected.

doi:10.1371/journal.pone.0168344.t005

evidence. This demonstrates that certain rumor characteristics become prominent at different course of the propagation.

Second, linguistic features are powerful and stable predictors of rumors. During initial circulation, rumors can be identified by the use of words related to negation and skepticism. Over time, linguistic characteristics persist and additional patterns such as strong negation or hearing appear. Linguistic features consistently differentiate rumors with high performance (F1 score: 0.8-0.86). Furthermore, utilizing the linguistic features alone yields the higher predictive power even than using the full list of features, except for the 56-day observation window.

Third, the network features require a longer time period and further information on spreading to become predictive. The performance with short observation periods is comparable to random guessing. Only the features representing the specific direction of information flow, unique users whose source is outside the network and the rumor content being neglected are important for classification. From the results, tracing each information flow is significantly more important than knowing the entire relationship among users. Collecting the entire network structure has always been an issue in studies because it requires significant resources

compared with other types of information. This can be an indicator that these expensive features are not required in the early stages of rumor classification.

Considering the temporal features, the proposed PES model excelled in capturing the prominent differences between rumors and non-rumors for the 56-day observation window. Yet, the temporal features demonstrated poor differentiating power for shorter time windows. Because the PES model to fit temporal features utilize two trigonometric functions to explain user interaction and information propagation processes, all but the 56-day observation were not sufficiently long enough to define such features.

We propose two new algorithms for rumor classification. The first algorithm, which takes every feature into account, showed good performance with F1 score in the range from 0.77 to 0.90 after 28 days from the first date of rumor spreading. This algorithm however did not perform well for shorter observation periods. For shorter time duration, regarding network and temporal features rather decreased the performance of the classifier since the performances of these features are just slightly higher than random selection. Despite the insufficient performance at early rumor detection, this algorithm showed undoubtedly good performance when the network and temporal features become well available (i.e., after 14 days).

The second algorithm is geared toward early rumor classification and considers user and linguistic features. For short observation periods, this algorithm showed better performances compared to using a full set of features. For longer time windows, its performance slightly decreased to F1 score of 0.80. This algorithm nonetheless has two strengths. First, its classification performance is less sensitive to observation periods compared to the full model. Second, it does not require any temporal or network features that rely on near-complete data obtainable only via extensive gathering or explicit APIs. Given that most social networks do not grant researchers to obtain full data, the ability to design an algorithm that detects rumors with locally obtainable information is essential for practical use.

## Discussion

The present study examined an extensive set of features related to the user, linguistic, network and temporal characteristics of rumors and compared their relative strength in classifying rumors over time. This comprehensive study was conducted on real rumor cases extracted from near-complete data of Twitter. The defined features and experiments, which draws upon existing social and psychological theories as well as unique observations from data, provides a deep understanding of rumor mechanisms as follows. First, this study identifies an early marker of rumors that is visible from the first few days of circulation; rumor participants have fewer followers and express doubt and skepticism. Second, the network and temporal features became remarkable for over two months; rumors have low adoption rates and appear repeatedly over time, which results in smaller yet more densely-knit communities compared to non-rumor networks. Third, we proposed a practical algorithm that does not require full snapshot of the network nor complete historical records. While the most predictive features of rumors changed depending on the observation window, the combination of user and linguistic features functioned well consistently over short and long time windows.

Unlike the current study, many previous studies have used a single observation window that is often arbitrarily set [7]. Results obtained from one stationary viewpoint may be insufficient to represent the general rumor spreading patterns. Instead, our work was inspired by a few studies that have argued rumor spreading mechanisms change over time [13, 14], which is a theme that we explored deeply in this paper. Nonetheless, this study shares limitations of previous rumor studies in that statistically meaningful predictive features are observational and they cannot be generalized to all rumor cases. Toward handling this limitation, we have

carefully made the following efforts. First, we considered the pre-existing rumor theories from the social and psychological fields as important evidence in deriving the features and explaining the observations of the experiment. Hence, we have supportive reasons regarding why our findings are not inadvertent results. Second, we repeated the experiments for various observation periods and used only observations with invariant patterns. These efforts guide us to arriving at a more robust and reliable understanding of rumor spreading mechanisms in online social media.

The classification experiments remarkably showed that certain rumor characteristics become prominent at different course of the rumor circulation and hence appear with strong predictive power. The results provide valuable observations for identifying practical features for developing a rumor classifier. The introduced sets of features prove their validity as predictors with the minimum F1 score of 0.8. Although the time window over which each feature set showed the best performance differed, which makes it difficult to determine when a set of features becomes promising or unpromising due to the different speed of information spreading. Nonetheless, each set of features exhibited almost uniform increasing and decreasing patterns and it could be determined which feature sets are prominent for early stage rumor detection. At the rumor initiation, the linguistic and user characteristics are important for classifying rumors from streams of content.

Recent studies have begun to focus on rumor detection within only short time periods like 24 hours and they have achieved good classification results. One study highlighted the usefulness of content-based features (i.e, linguistic features) through demonstrating great performance (F1 score: 0.89) with a combination of TF-IDF (term frequency–inverse document frequency) of top-k vocabularies from data and deep learning [48]. Due to the nature of machine learning, however, there is always an issue of bias error. The linguistic features proposed in this study were drawn from a set of words that are the most common and widely used to express a person's ideas or thinking. Thus, we expect that these linguistic features are less likely to be affected by field-specific lexis that causes bias in typical NLP methods. Combining the user and linguistic features, we plan to develop a rumor classifier that is appropriate for only small and initial fraction of content and is less sensitive to the type of information.

## Supporting Information

**S1 Dataset. Tweets and user characteristics for rumor and non-rumor events.** See the respective attached file.
(7Z)

**S1 Table. Description of non-rumor events.**
(PDF)

**S2 Table. Description of rumor events.**
(PDF)

**S3 Table. User features.** Kurtosis is a measure of how much a given data is heavy or light-tailed compared to a normal distribution. Data with high kurtosis likely to have heavy tails or outliers. Skewness is a measure of symmetry. In definition column, a term 'SD' mean for standard deviation.
(PDF)

**S4 Table. Network features.** In definition column, *E*, *F*, and *D* mean for extended network, friendship network and diffusion network, respectively.
(PDF)

**S5 Table. Temporal features.** Symbols are for parameters of suggested PES model. (PDF)

## Author Contributions

**Conceptualization:** SK MC KJ.

**Data curation:** SK MC.

**Formal analysis:** SK.

**Funding acquisition:** MC KJ.

**Investigation:** SK MC KJ.

**Methodology:** SK MC KJ.

**Project administration:** MC.

**Resources:** MC SK.

**Software:** SK.

**Supervision:** SK MC.

**Validation:** SK KJ MC.

**Visualization:** SK.

**Writing – original draft:** SK MC KJ.

**Writing – review & editing:** SK MC.

## References

1. Prasad J. The psychology of rumour: A study relating to the great Indian earthquake of 1934. British Journal of Psychology General Section. 1935; 26(1):1–15. doi: 10.1111/j.2044-8295.1935.tb00770.x

2. Knapp RH. A psychology of rumor. Public Opinions Quarterly. 1944 Mar 20; 8(1):22–37.

3. Festinger L, Cartwright D, Barber K, Fleischl J, Gottsdanker J, Keysen A, Leavitt G. A study of a rumor: its origin and spread. Human Relations. Human Relations. 1948.

4. Rosnow RL. Inside rumor: A personal journey. American Psychologist. 1991 May; 46(5):484. doi: 10.1037/0003-066X.46.5.484

5. Rosnow RL. Rumor as communication: A contextualist approach. Journal of Communication. 1988 Mar 1; 38(1):12–28. doi: 10.1111/j.1460-2466.1988.tb02033.x

6. DiFonzo N, Bordia P. Rumor psychology: Social and organizational approaches. American Psychological Association; 2007.

7. Qazvinian V, Rosengren E, Radev DR, Mei Q. Rumor has it: Identifying misinformation in microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing 2011 Jul 27 (pp. 1589–1599). Association for Computational Linguistics.

8. Spiro ES, Fitzhugh S, Sutton J, Pierski N, Greczek M, Butts CT. Rumoring during extreme events: A case study of Deepwater Horizon 2010. In: Proceedings of the 4th Annual ACM Web Science Conference 2012 Jun 22 (pp. 275–283). ACM.

9. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F. Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th international conference companion on World wide web 2011 Mar 28 (pp. 249–252). ACM.

10. Oh O, Agrawal M, Rao HR. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. Mis Quarterly. 2013 Jun 1; 37(2):407–26.

11. Andrews C, Fichet E, Ding Y, Spiro ES, Starbird K. Keeping Up with the Tweet-dashians: The Impact of 'Official' Accounts on Online Rumoring. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing 2016 Feb 27 (pp. 452–465). ACM.

12. Lee J, Agrawal M, Rao HR. Message diffusion through social network service: The case of rumor and non-rumor related tweets during Boston bombing 2013. Information Systems Frontiers. 2015 Oct 1; 17 (5):997–1005. doi: 10.1007/s10796-015-9568-z

13. Friggeri A, Adamic LA, Eckles D, Cheng J. Rumor Cascades. In: ICWSM 2014 May 16.

14. Bordia P, DiFonzo N, Chang A. Rumor as Group Problem Solving: Development Patterns in Informal Computer-Mediated Groups. Small Group Research. 1999 Feb 1; 30(1):8–28. doi: 10.1177/104649649903000102

15. Bordia P, DiFonzo N. Problem solving in social interactions on the Internet: Rumor as social cognition. Social Psychology Quarterly. 2004 Mar 1; 67(1):33–49. doi: 10.1177/019027250406700105

16. Sunstein CR. On Rumours: How Falsehoods Spread, Why We Believe Them, What Can Be Done. Princeton University Press; 2014 Mar 9.

17. Castillo C, Mendoza M, Poblete B. Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web 2011 Mar 28 (pp. 675–684). ACM.

18. Yang F, Liu Y, Yu X, Yang M. Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics 2012 Aug 12 (p. 13). ACM.

19. Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H. Survey of review spam detection using machine learning techniques. Journal of Big Data. 2015 Oct 5; 2(1):1. doi: 10.1186/s40537-015-0029-9

20. Ezpeleta E, Zurutuza U, Hidalgo JM. Does Sentiment Analysis Help in Bayesian Spam Filtering? In: International Conference on Hybrid Artificial Intelligence Systems 2016 Apr 18 (pp. 79–90). Springer International Publishing.

21. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Journal of Management Information Systems. 2013 Apr 1; 29(4):217–248. doi: 10.2753/MIS0742-1222290408

22. Lim S, Jung I, Lee S, Jung K. Analysis of information diffusion for threshold models on arbitrary networks. The European Physical Journal B. 2015 Aug 1; 88(8):1–4. doi: 10.1140/epjb/e2015-60263-6

23. Gao ZK, Yang YX, Fang PC, Jin ND, Xia CY, Hu LD. Multi-frequency complex network from time series for uncovering oil-water flow structure. Scientific reports. 2015 Feb 4; 5:8222. doi: 10.1038/srep08222 PMID: 25649900

24. Takayasu M, Sato K, Sano Y, Yamada K, Miura W, Takayasu H. Rumor diffusion and convergence during the 3.11 earthquake: a Twitter case study. PLoS one. 2015 Apr 1; 10(4):e0121443. doi: 10.1371/journal.pone.0121443 PMID: 25831122

25. Wang Z, Dong W, Zhang W, Tan CW. Rumor source detection with multiple observations: fundamental limits and algorithms. In: ACM SIGMETRICS Performance Evaluation Review 2014 Jun 16 (Vol. 42, No. 1, pp. 1–13). ACM. doi: 10.1145/2637364.2591993

26. Karamchandani N, Franceschetti M. Rumor source detection under probabilistic sampling. In: Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on 2013 Jul 7 (pp. 2184–2188). IEEE.

27. Liu Y, Xu S, Tourassi G. Detecting Rumors Through Modeling Information Propagation Networks in a Social Media Environment. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction 2015 Mar 31 (pp. 121–130). Springer International Publishing.

28. Mendoza M, Poblete B, Castillo C. Twitter under crisis: can we trust what we RT? In: Proceedings of the first workshop on social media analytics 2010 Jul 25 (pp. 71–79). ACM.

29. Gao ZK, Yang YX, Fang PC, Zou Y, Xia CY, Du M. Multiscale complex network for analyzing experimental multivariate time series. EPL (Europhysics Letters). 2015 Feb 19; 109(3):30005. doi: 10.1209/0295-5075/109/30005

30. Gao ZK, Fang PC, Ding MS, Jin ND. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. Experimental Thermal and Fluid Science. 2015 Jan 31; 60:157–164. doi: 10.1016/j.expthermflusci.2014.09.008

31. Gao ZK, Yang YX, Zhai LS, Jin ND, Chen G. A four-sector conductance method for measuring and characterizing low-velocity oil-water two-phase flows. IEEE Transactions on Instrumentation and Measurement. 2016 Mar; 65(7):1690–1697 doi: 10.1109/TIM.2016.2540862

32. Sanz J, Xia CY, Meloni S, Moreno Y. Dynamics of interacting diseases. Physical Review X. 2014 Oct 8; 4(4):041005. doi: 10.1103/PhysRevX.4.041005

33. Xia CY, Wang Z, Sanz J, Meloni S, Moreno Y. Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks. Physica A: Statistical Mechanics and its Applications. 2013 Apr 1; 392(7):1577–85. doi: 10.1016/j.physa.2012.11.043

**34.** Granell C, Gómez S, Arenas A. Competing spreading processes on multiplex networks: awareness and epidemics. Physical Review E. 2014 Jul 14; 90(1):012808. doi: 10.1103/PhysRevE.90.012808

**35.** Xia CY, Meloni S, Perc M, Moreno Y. Dynamic instability of cooperation due to diverse activity patterns in evolutionary social dilemmas. EPL (Europhysics Letters). 2015 Mar 9; 109(5):58002. doi: 10.1209/0295-5075/109/58002

**36.** Di Stefano A, Scatà M, La Corte A, Liò P, Catania E, Guardo E, Pagano S. Quantifying the role of homophily in human cooperation using multiplex evolutionary game theory. PloS one. 2015 Oct 23; 10(10): e0140646. doi: 10.1371/journal.pone.0140646 PMID: 26496351

**37.** Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent Features of Rumor Propagation in Online Social Media. In: 2013 IEEE 13th International Conference on Data Mining 2013 Dec 7 (pp. 1103–1108). IEEE.

**38.** Cha M, Haddadi H, Benevenuto F, Gummadi PK. Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM. 2010 May 23; 10(10-17):30.

**39.** Nkpa NK. Rumor mongering in war time. The Journal of Social Psychology. 1975 Jun 1; 96(1):27–35. doi: 10.1080/00224545.1975.9923258

**40.** Ma J, Gao W, Wei Z, Lu Y, Wong KF. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management 2015 Oct 17 (pp. 1751–1754). ACM.

**41.** Cantril H. The invasion from Mars: A study in the psychology of panic.  Transaction Pub; 1940.

**42.** DiFonzo N, Bordia P. Corporate rumor activity, belief and accuracy. Public Relations Review. 2002 Feb 28; 28(1):1–9. doi: 10.1016/S0363-8111(02)00107-8

**43.** Solove DJ. The future of reputation: Gossip, rumor, and privacy on the Internet.  Yale University Press; 2007.

**44.** Spencer S, Srikant R. Maximum likelihood rumor source detection in a star network. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016 Mar 20 (pp. 2199–2203). IEEE.

**45.** Matsubara Y, Sakurai Y, Prakash BA, Li L, Faloutsos C. Rise and fall patterns of information diffusion: model and implications. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining 2012 Aug 12 (pp. 6–14). ACM.

**46.** Levenberg K. A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics. 1944 Jul 1; 2(2):164–8. doi: 10.1090/qam/10666

**47.** Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recognition Letters. 2010 Oct 15; 31(14):2225–36. doi: 10.1016/j.patrec.2010.03.014

**48.** Ma J, Gao W, Mitra P, Kwon S, Janse BJ, Wong KF, et al. Detect Rumors from Microblogs with Recurrent Neural Networks. In: Proceedings of IJCAI 2016.