Rumors at the Speed of Light? Modeling the Rate of Rumor Transmission during Crisis

Li Zeng
Information School
University of Washington
Seattle, WA, USA
lizeng@uw.edu

Kate Starbird
Department of Human Centered Design & Engineering
University of Washington
Seattle, WA, USA
kstarbi@uw.edu

Emma S. Spiro
Information School
University of Washington
Seattle, WA, USA
espiro@uw.edu

Abstract

Social media have become an established feature of the dynamic information space that emerges during crisis events. Both emergency responders and the public use these platforms to search for, disseminate, challenge, and make sense of information during crises. In these situations rumors also proliferate, but just how fast such information can spread is an open question. We address this gap, modeling the speed of information transmission to compare retransmission times across content and context features. We specifically contrast rumor-affirming messages with rumor-correcting messages on Twitter during a notable hostage crisis to reveal differences in transmission speed. Our work has important implications for the growing field of crisis informatics.

1. Introduction

Social media are vital information sources during non-routine circumstances such as crisis events [1][2][3]. The public increasingly turns to social media platforms to search for, disseminate, challenge, and make sense of uncertain information in non-routine settings such as disasters and crisis events [4][5][6]. Emergency responders and mass media outlets also participate in this emerging conversation online, resulting in a complex and dynamic, event-related information environment contained within a larger stream of informal communication on these platforms [7].

One natural consequence of such behavior is the emergence of *rumors* – stories of unknown validity at the time of communication, which may eventually turn out to be true or false (as evaluated after the conclusion of the event). Rumoring is often characterized as a sense-making process through which individuals and groups grapple with a lack of information during times of uncertainty [8][9]. A key mechanism through which sense-making can occur is the diffusion or transmission of information from one individual to the next (or many others in the case of broadcast content) via social relationships. As information diffuses through a population individuals interact with each other and the content, processing and commenting on the validity and importance of the information [10][7]. Serial transmission of information is also vital for increasing the reach, and thus exposure, of the information.

Serial transmission of content is especially important during crisis events, because information can be time-sensitive and critical for decision making - information could mean the difference between life and death. While a growing number of studies have examined information transmission during crisis events, less work has attempted to quantify the speed of this process, which could vary in promptness of dissemination and persistence of attention [11]. For example, in the context of a crisis, there might be important consequences in different transmission speeds between substantiated information, i.e. crisis-related facts or true rumors, and information that has not been verified or been falsified by the crowd. Moreover, certain indicators of the collective sense-making process (e.g. authors expressing uncertainty in their messages) may be relevant.

Prior work has looked at the ways in which content and user characteristics are associated with serial transmission and public attention on social media platforms [12]. This work reveals that content (e.g. topics, URLs) features and user characteristics (e.g. number of social ties) play a role in determining the number of reposts a message receives [12][13][14]. Building from this foundation, our work seeks to fill a gap in the current understanding of rumoring behavior on social media during crisis events by quantifying features of information content and context that are associated with the *speed* at which information is transmitted.

Our work has important implications. First, exploring the dynamics of rumoring behavior has the potential to change our understanding of how collective sense-making occurs online. In particular this analysis could reveal how individuals differentially respond to various kinds of rumor-related messages with respect to their decision to pass content along to social contacts, changing our understanding of how collective negotiations of information veracity play out online. Importantly, this work could also offer important practical recommendations for crisis management by discovering the features that contribute to popularity and persistence of false rumor content (i.e. misinformation) in social media spaces – a topic of growing interest within emergency management communities, as well as the growing field of crisis informatics [15]. Findings could, for example, be used to help emergency responders to rapidly disseminate information and regulate rumor propagation during crisis events.

In this work we measure waiting time (i.e. the time span between information production and redistribution) [11] to explore the relationship between waiting time and features about the author, post content and external context of messages posted on Twitter. We compare empirical differences in waiting time for rumor-related crisis messages with randomly sampled, more general posts on Twitter (i.e. tweets). We model the median waiting time (capturing average attention span) as explained by various author, content, and context features to address the following research questions: *Do we observe that crisis-related content increases the public's retransmission response speed on Twitter, when compared with random content? Does the average retransmission waiting time of content pertaining to crisis rumors vary by the content's expressed stance towards the rumor (i.e. affirming or denying rumor claims)?*

In what follows we situate our work within the growing field - crisis informatics, and more generally to core questions about rumoring behavior from the social sciences. We describe the crisis case to be analyzed and the dataset used. Our analysis looks first at the empirical distribution of waiting times for crisis-related content, compared with randomly sampled messages. We then present a model for median waiting times which we use to evaluate the relative influence of author, content and context features – focusing specifically on rumor dimensions.

2. Related Work

Social media are changing the ways in which information sharing occurs in social networks [22][23][24]. These technologies allow individuals to reach a larger number of social contacts across greater distances than was previously possible, for example if community members are restricted to face-to-face information exchange. Encouraging users to articulate their social connections, social media often leverage social networks to facilitate information sharing with friends, family and strangers. Social media platforms actively facilitate not only information sharing but also retransmission of content through reposting (i.e. re-sharing of content posted by other users). Twitter, for example, allows users to "retweet" content, resending a message posted by another user to one's own social network. It is through these kinds of transmission mechanisms that information can spread, or diffuse, through the underlying social network, reaching larger audiences.

The study of information diffusion in social networks is not new; it has a long history in the social sciences [16][17][18]. Many early studies in the field were motivated by exploration and interest in the role of rumors during war [19]. Rumoring, one example of an information diffusion process, was thought to be a natural consequence of circumstances characterized by high levels of uncertainty and a lack of official information [9][8]. Rumors, as defined in this line of work, are stories pertaining to facts or events of topical interest that do not occur as part of a formal, institutionalized communication process. Rumors then become a form of "improvised news" for individuals to discuss and process [9].

Early work on rumoring looked at crisis contexts because such non-routine circumstances were conducive to the growth and propagation of rumors. Prior work attempted to identify features related to increases in transmission, as well as how retransmission was related to information distortion [20]. In one of the classic studies exemplifying this tradition, Allport and Postman [21] propose three primary factors that influence rumor propagation: (1) perceived importance of the information, (2) degree of uncertainty or cognitive unclarity surrounding the information, and (3) relevance of the information to behavior. Indeed, crisis-related rumors satisfy all three of these conditions and are therefore ripe for transmission. Recently, many scholars have extended classic theories of information diffusion and rumoring into the context of social media platforms, exploring factors that are hypothesized to affect the magnitude of retransmission or size of the resulting information cascade [25][26]. These studies point to content themes, user characteristics, and context as important in determining the attention a particular message receives [27][12][28].

Despite notable work that explores information diffusion, rumor propagation in particular, on social media during crisis events, very few studies have looked at the *speed* of this process. How long does a message typically "wait" in the public stream before it is reposted or shared by other users? What is the distribution of these waiting times and how might they differ based on the content contained in the message? These are key questions to answer in order to gain a more thorough understanding of crisis-related rumoring behavior online. Indeed, while commonly held notions would have one believing that rumors "spread like wildfires" – rapidly without concern for obstacles in their path – research indicates such metaphors can be misleading [29].

Speed of information transmission is essential to understand in the crisis context. In emergency settings, time-sensitive information must reach target populations before it becomes outdated or irrelevant, or worse ineffective at warning of imminent danger. For example, suppose a flash flood warning takes days to diffuse through social media spaces from warning/alert organizations to those affected. In such a case, social media platforms would not be practically viable as the sole outlet for such warning information. On the other hand, if information can rapidly be transmitted online, emergency responders can make use of this preexisting "soft infrastructure" for emergency preparedness, warning, response, and recovery. It is important to note that social media are not isolated information environments; though outside the scope of the current work information transmission also occurs across online platforms and moves from online to offline spaces as well.

Rumor transmission may rely on a complex set of circumstances, from basic behavioral seasonality patterns known to affect online communication to rich crowd interaction and sense-making processes unfolding over time. It may even be driven by the participation of a few highly influential users. In this paper, we explore variations in the speed of information transmission along dimensions of sense-making and other context features. In the following sections we discuss the model formulation and implications, linking results back to foundational work from sociological studies of rumoring during crisis.

3. Data Collection

To understand rumor transmission dynamics during crisis events, we utilize two datasets of social media posts from the microblogging site Twitter. These two datasets were collected by our research team as part of a larger project to explore rumoring during crisis [38][39][40]. The first dataset is comprised of tweets collected during a hostage crisis that occurred in Sydney, Australia at the end of 2014. On 15 December 2014 at 9:45 AM a gunman took 18 people hostage in a café in central Sydney. The 16-hour standoff that followed included formal negotiations with the gunman, various police responses and reactions, successful hostage escapes, and more, leading to an environment ripe for the emergence of rumors. The crisis ended after police stormed the building; two hostages and the gunman were killed.

Utilizing custom python scripts to access the Twitter Streaming API, we collected and archived content (i.e. tweets) containing the following search terms: *sydneysiege*, *martinplace*, *sydney*, *lindt*, and *chocolate shop*. These terms were selected to capture event-related content and as such refer to the event and locale of the crisis. Data collection began on December 15th, 2014 at 11:06am AEDT (local time), immediately following breaking news coverage of the event and ended two weeks later. While the majority of event-related content was posted in the first 48 hours of the event, we extend data collection for a two-week period to prevent data loss due to unanticipated subsequent event discussion. We captured 5,429,345 tweets.

The second dataset used here is comprised of a sample of tweets collected via the random sample Twitter makes available through their streaming API; these tweets provide a baseline case to which we can compare crisis-related content. Data collected are a small random sample of all public tweets. Tweets were monitored over the five-month period from January to May, 2015. We captured millions of tweets, and use a sample of these in our analysis below.

4. Methods

Our goal in this work is to explore the timing and speed of information transmission during crisis events; we aim to quantify the association between speed of retransmission and rumor-related features of the messages.

4.1. Rumor Identification and Content Coding

Our analysis makes use of a complex and rich dataset that has been curated through an extensive process of data cleaning and manual coding. Identifying rumors in a large corpus of unstructured text is a challenging problem. Our research team continues to develop mixed-methods approaches, leveraging visual exploratory data analysis as well as external sources (e.g. media reports). After rumor stories are identified, we need to determine which

Table 1. Descriptive statistics for rumor tweets sets and randomly sampled tweets

	# of tweets	% of retweets	# of rumor-related retweets	Min. waiting time (min)	Med. waiting time (min)	Max. waiting time (min)
Random Sample	NA	225539	NA	0.00	52.13	4686181.17
Hadley Rumor	4094	0.665	1983	0.02	4.78	10037.61
Lakemba Rumor	7912	0.920	1207	0.22	4.48	245.70
Suicide Rumor	8134	0.762	2172	0.00	3.58	186701.60
Flag Rumor	23165	0.641	8104	0.00	13.64	8386.32
Airspace Rumor	7403	0.721	5293	0.00	4.92	15278.22

tweets are related to each rumor. In our approach, we refine the set of tweets related to each rumor via an iterative process, relying heavily on the ability of subject matter experts to generate keyword-based search queries that produce a comprehensive, low noise corpus for each identified rumor. In the case of the Sydney Siege event, research team members identified five salient rumors, constituting the rumor dataset (50,708 tweets) that we utilize in the analysis below. Descriptive statistics of each rumor-related set of tweets, along with the randomly sampled dataset are seen in Table 1. We briefly provide background on each of the rumors.

The first rumor centers on claims that an Australian radio host Ray Hadley spoke with one of the hostages on the phone, a story that turned out to be true (hereafter referred to as *Hadley*). The second rumor claims the Australian Federal Police were conducting home raids in Lakemba (a predominantly Muslin suburb of the city) at the same time as the Sydney Siege event; the story was later denied by authorities (hereafter referred to as *Lakemba*). The third rumor claims wearable suicide belts/vests were present; media reports suggest this claim was false (hereafter referred to as *Suicide*). The fourth rumor describes hostages holding an ISIL (Islamic State of Iraq and Levant) flag, which turned out later to be false (hereafter referred to as *Flag*). The last rumor reported that the airspace over Sydney had been closed due to the siege; while the story was thought to have been supported by the Sydney Airport, it was later disconfirmed by Airservices Australia (hereafter referred to as *Airspace*).

To capture various dimensions of rumoring activity the research team coded each unique tweet along two dimensions. The first, which is designed to identify crowd support or correction, consists of five mutually exclusive categories*: Affirm, Deny, Neutral, Unrelated*, and *Uncodable*. *Affirm* tweets support or help to spread the story, while *Deny* tweets challenge all or part of the story. *Neutral* tweets neither affirm nor deny the rumor story, expressing an exact neutral stance on rumor belief. *Unrelated* tweets do not match with the rumor description but are still related to the crisis event. Tweets that cannot be coded are coded as *Uncodable* (e.g. tweets that contain non-English words that impede comprehension). Next, we code for other interesting signals of collective sense-making within the information space – expressed uncertainty is a second level code that can be applied to any tweet, suggesting some level of uncertainty about rumor story. Three trained coders manually code every distinct tweet (removing retweets and very close matches so as to narrow the scale of this task). We apply a "majority rules" decision process for assigning final codes [1]. Rumor identification and rumor coding tasks provide the foundation of our study, however, detailed elaboration of these processes is beyond the scope of the paper. Interested readers can refer to related work for more details [38][39][40].

4.2. Measuring the Speed of Transmission

The primary quantity of interest in this analysis is the speed of message retransmission. We define the waiting time for a tweet-retweet pair as the length of time between when the original message was first posted and the later point in time it was subsequently redistributed by another Twitter user. We measure waiting times by focusing on retweets contained in the dataset; each retweet contains its metadata attributes describing its own post time as well as the post time of the original authored tweet. Each retweet record contains data about the original tweet itself including a set of attributes about the original author.

In the Sydney Siege dataset, 37,555 (74.06%) of tweets are retweets.[2] We then group together retweets of the same original tweet; if an original authored tweet was reposted multiple times, it has multiple retweets and hence multiple associated waiting times. We calculate average waiting times for every original tweet,

[1] An inter-rater reliability analysis using the Fleiss' Kappa statistic [41] was performed to evaluate agreement among raters and the result of Kappa is 0.892 (p<0.001) [39].

[2] We can only identify "official" retweets, those that users posted by clicking the Twitter "Retweet" button. Even though users could manually retweet by copying a tweet and reposting it, Twitter fails to automatically recognize these posts as retweets; subsequently data about the original message is not included as part of the record. We do not include manual retweets in our retweet dataset.

using the median of a set of waiting times per tweet as the outcome measure of interest. In the randomly sampled tweet data, we utilized a simple random sampling strategy to avoid an abundance of retweets from the same tweet. We randomly sampled 225,539 retweets to serve as the control group. We calculate waiting time of each random retweet as the interval between posting time of a randomly sampled retweet and its original tweet. Table 1 displays descriptive analysis of waiting times for the five rumor datasets and the random sample dataset.

4.3. Tweet Feature Extraction

One of the primary goals here is to evaluate which tweet features that are associated with longer/shorter median waiting time by building a predictive model for crisis-related content waiting time. We propose five categories of tweet features of interest to characterize different dimensions of these posts: tweet-element features, rumor features, interest features, exposure features and seasonality features. These categories of features are motivated by prior work, as discussed above, and extended into the rumor domain. Within each category, we come up with multiple features and hypothesize their effects on waiting times. Table 2 contains a listing of the potential features to be used in the model.

Table 2. Potential features

Tweet element	The presence of URLs, the presence of hashtags, the presence of user mentions
Rumor	Rumor stance, uncertainty, true or false rumor, sentiment
Interest	Retweet counts, favorite counts
Exposure	Average exposure degree (in-degree), average attention degree (out-degree), popularity of original poster, outgoingness of original poster
Seasonality	Time of day when original tweet was posted

Tweet-element features: URLs, hashtags and user mentions have been shown to have a high correlation with retweetability and distribution of waiting time of hazard-related tweets [11][31][32]. Tweet-element features might also be related to waiting time of tweets with rumor content. *URLs* may provide external evidence to support a rumor or challenge a rumor, which may lead Twitter users to view external information before they respond to tweets, thereby lengthening waiting time. *Hashtags* often suggest themes or content of tweets; these tags also allow tweets to be gathered together to form channels or topics. Rumor-related tweets with hashtags might be easier and faster for Twitter users to search, consume and redistribute. Twitter posters can explicitly mention other Twitter users in their tweets using "@username" (known as user mentions). In the context of rumor propagation, *user mentions* may lead to different attractiveness to users who were mentioned and others who were not.

Rumor features: This feature category characterizes elements related to rumor-specific content. *Rumor stance*, looking at tweets coded as *affirm*, *deny* or *neutral*, indicates the role that an original tweet may play during the process of rumor propagation and sense-making. Original tweets that support versus challenge a rumor serve as protagonists, and likewise skeptics, of the rumor story. Tweets can also take a neutral stance serving as messengers of a story. We explore the potential implications of these roles on waiting time, capturing differential tendencies for people who tend to have different attitudes towards rumor-related information.

Uncertainty (captured by our second-level codes) is used to indicate whether a rumor-related tweet author expresses uncertainty in the post. This uncertainty might further impact others' retweeting behaviors if users are wary of reposting uncertain information. *True or false rumor* measures whether a rumor turns out to be truth or untruth at the conclusion of the event. The true/false categorization is made based on researcher judgment after the event. Previous research suggests that tweets with extreme sentiment tend to be retweeted more often and faster compared to emotionally neutral ones [35][36][37]. Based on this hypothesis, we implemented a simple measure of content sentiment [3] based on textual properties of tweets content; these methods provide three classes of *sentiment*: positive, neutral and negative. While a rough measure of sentiment, it nonetheless provides a simple estimate of emotional affect.

Interest features: Existing theories suggest that rumor propagation can be affected by people's interest in and perceived importance of the topic or subject of a particular piece of information. In this analysis, we utilize *retweet counts* and *favorite counts* of the original tweets to quantify how much Twitter users are "interested" in a certain original tweet. *Retweet counts* capture how many times an original tweet has been reposted at the time of observation. *Favorite counts* indicate the total number of times that a tweet has been "favorited" or liked by Twitter users.

Exposure features: Exposure features are designed to capture the idea that once a tweet is posted it may be noticed, read, liked or reposted by others in the author's network (to whom it is automatically delivered). We use follower counts and friend counts

[3] We applied Sentiment API of MetaMind to our rumor-related dataset. The claimed accuracy of this method is 81.73%. https://www.metamind.io/classifiers/155

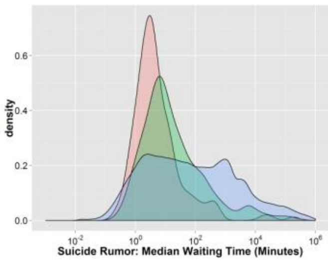


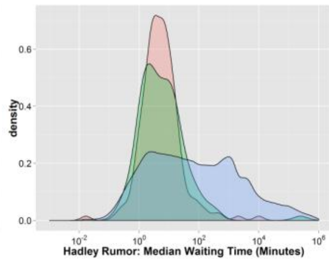
Figure 1 (a)


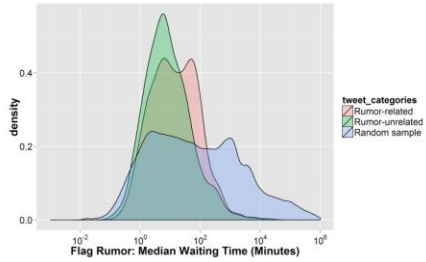
Figure 1 (b)


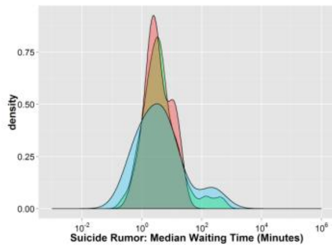
Figure 1 (c)


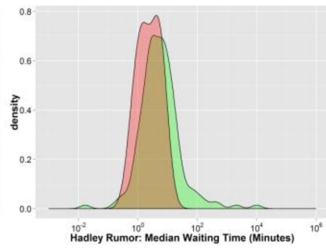
Figure 1 (d)


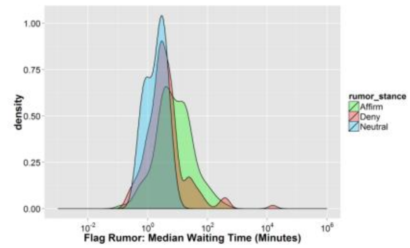
Figure 1 (e)


Figure 1 (f)

Figure 1 (a-f). Density plot for rumor-related, rumor-unrelated and randomly sampled tweets

of the authors of original tweets to quantify *popularity of original poster* and *outgoingness of original poster*. The *average degree of exposure* of an original tweet is characterized by the median of follower counts of all authors that repeated a same original tweet (i.e. duplicate content). The *average degree of attention* measures the median of friend counts of all authors that reposted the same original tweet.

Seasonality features: Previous research suggests that seasonality has strong association with users' posting behavior. We consider the posting time in the local time zone of the authors of original tweets to study whether posting rumor-related tweets at different times of the day correlates to different waiting times. We extract the hour of the day during which the original tweet was posted as the feature of seasonality.

4.4. Modeling Waiting Times

To model waiting times of rumor-related tweets, we first fit the observed data using a simple linear model. Even though the linear model is simplistic, it is able to suggest explanatory power of potential features, and to evaluate the relative impact of each of the features. To extend this simple modeling framework, we also fit the data using a regression tree model enabling a non-linear regression of waiting time on the set of potential features. In both models, waiting times are transformed to be on a log scale due to the skewness of the empirical distribution (as seen in Table 1).

To build a linear model of waiting times, we take the dependent variable as the waiting times, and model the expected waiting time for each rumor-related tweet as a linear combination of features:

$$E\left[\log wt_i\right] = \alpha + \beta_j \cdot feature_{ij}$$

wt_i is the median waiting time of the i^{th} rumor-related tweet;β_j is the coefficient of the j^{th} feature of the model and $feature_{ij}$ represents the value of the j^{th} feature of the i^{th} rumor-related tweet. Since this simple linear regression model is restrictive - a single predictive formula holds over for the entire data-space - we extend our analysis by applying a method of non-linear regression to partition our feature space into smaller regions, in order to get more interpretable predictions.

5. Results

5.1. Empirical Waiting Times

To begin we consider the empirical distribution of waiting times, comparing different rumors to control content. Figure 1 shows the observed waiting time distributions for three illustrative rumors, comparing the waiting times for rumor-related (tweets coded as *affirm*, *deny* and *neutral*), rumor-unrelated (tweets coded as *unrelated*), and control content – tweets from the randomly sampled dataset.

Three different characteristics of these waiting time distributions are shown from Figure 1 (a) to 1 (c). Figure 1 (a) shows the most common trend in the empirical study with three among the four false rumors in our dataset being consistent with this trend. The distribution of waiting times for rumor-unrelated and random tweets have fatter tails than those of rumor-related tweets. Figure 1 (b) shows the distribution for the one true rumor in our dataset. The distribution of rumor-related tweets presents a small bump at the range of very short waiting times and immediately after that, the bump dies away. The difference between the distribution of rumor-related tweets and the distribution of rumor-unrelated tweets is not as noticeable as in Figure 1 (a). One explanation for this difference might be that once the rumor was confirmed as truth (which occurred relatively early) it fails to maintain enough excitement or uncertainty to motivate redistribution. Figure 1 (c) presents the comparison of the distributions for the *Flag* rumor. Though this is a false rumor, the distribution show that overall, rumor-related tweets have longer waiting time than rumor-unrelated tweets. The average waiting time of event-related tweets is shorter than control content across all rumor cases. Therefore, this empirical analysis suggests that content pertaining to crises and crisis-related rumors is likely to be particularly salient and worthy of redistribution, leading to shorter waiting times.

Figure 1 (d) to 1 (f) compare the distributions of waiting times for rumor-affirming tweets, rumor-denying tweets and neutral tweets for three illustrative rumors. Here, we see a consistent pattern across all three illustrative rumors – rumor-affirming tweets tend to have longer waiting time than rumor-denying tweets. Moreover, waiting times of rumor-affirming tweets are also longer than those of neutral tweets for *Suicide* and *Flag rumors*. Since the *Hadley* rumor only contains one original neutral tweet, the density curve of neutral *Hadley* tweets is not visible.

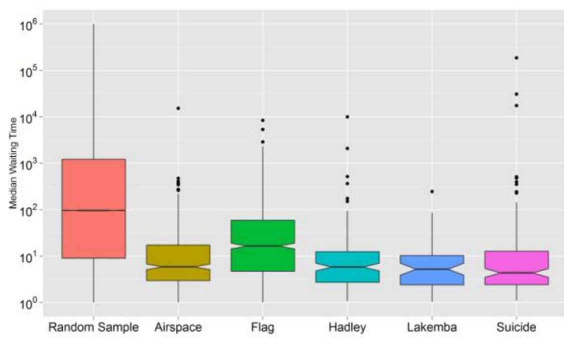


Figure 2. Median waiting time for rumor-related and randomly sampled content

We also show a comparison of median waiting time across all rumors, restricting the data to rumor-related tweets, in relation to median waiting time for randomly sampled tweets, in Figure 2. Median waiting time for control content is much larger than all rumor-related tweets across each different rumor. The variability of rumor-related waiting times is much smaller, indicating rumor-related tweets have consistently shorter waiting times than control content. Apart from the *Flag* rumor, each of the other four rumors has similar variability and median values. Further applications of this research, in particular across new crisis event cases might be able to reveal why some rumor waiting times take on different distributional characteristics than others.

5.2. Modeling Waiting Times

In modeling tweet waiting times during crisis events, we experiment with different combinations of feature sets in order to test the relative effect of each feature set on waiting times. This process of features or model selection is a standard approach for comparing the fit of different models, allowing the researchers to choose the best fit, most parsimonious model. To select the best performing models, we use standard information criteria, AIC and BIC, as indicators of model fit [33][34]. Table 3 presents model fit results, demonstrating that each set of features has additional explanatory power, evident in the decrease in both AIC and BIC as more features are added. Therefore, the best performing model includes all features. We use this model in subsequent analyses.

Table 3. Model selection

Model	AIC	BIC
tweet	6158.335	6185.286
rumor	6148.258	6191.38
interest	6133.767	6155.327
exposure	6131.456	6163.797
seasonality	6015.012	6149.735
tweet + rumor	6132.614	6191.906
tweet + rumor + interest	6093.509	6163.582
tweet + rumor + interest + exposure	6063.656	6155.289
tweet + rumor + interest + exposure + seasonality	**5933.285**	**6148.843**

5.3. Factors Associated with Waiting Times

Figure 3 shows the model coefficient estimates for each feature when all tweet-element, rumor, interest and exposure features are included in the model. The adjusted R-squared for this combined model is 0.1571. For tweet-element features, results suggest that the presence of URLs has a strong positive association with waiting times, indicating tweets with URLs tend to wait longer to be reposted. This may indicate that Twitter users need more time to read and comprehend external content provided by URL links. The presence of hashtags has a strong negative association with waiting time, suggesting these tags may help other Twitter users find the tweet easier and faster. A tweet containing a hashtag(s) is perhaps more likely to be seen by others earlier because it is easily accessible and searchable. In this model, the presence of user mentions does not have significant association with waiting times, which may suggest inclusion of @mentions does not affect redistribution of tweets.

For rumor features, our rumor stance codes have a strong association with waiting times. The codes of *affirm*, *deny* and *neutral* categories are included in the model as dummy variables, using *affirm* as reference class. Results show that rumor-denying tweets and neutral tweets are associated with shorter waiting times. Among these three classes, neutral tweets have the shortest waiting times, which may suggest that whether protagonists or challengers of a rumor require more time to collect evidence to support their point of views. Neutral stance tweets may present relatively low risk for reposting – they may not harm the poster's reputation (whereas spreading misinformation might be harmful). Interestingly, rumor-denying tweets are likely to have shorter waiting times than rumor-affirming tweets. One possible explanation of this effect might be that people are more interested in challenging a viewpoint than in supporting one. Thus, people respond to rumor-denying tweets more rapidly. Alternatively, denials might be perceived as important to curbing or stopping the spread of a false rumor and therefore important to pass along quickly.

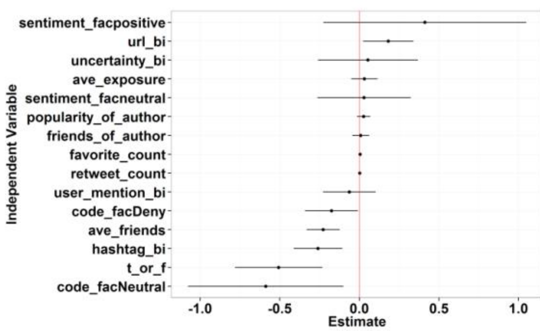


Figure 3. Coefficient plots for rumor, tweet-element, interest and exposure features

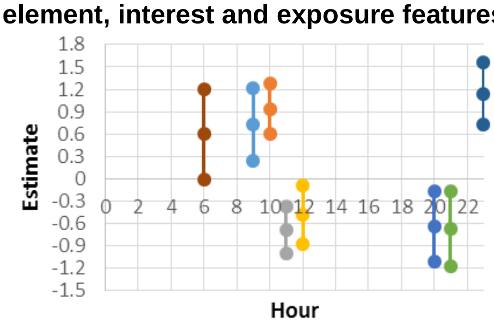


Figure 4. Seasonality

The feature of true/false rumor suggests that true rumors are associated with shorter waiting times, however, this requires further work (our dataset is limited because it contains only one true rumor). We need to identify more true rumors in future work to explore relationship between trustworthiness of rumors and human behaviors of redistributing rumor-related content. Even though the feature of content sentiment is not statistically significant in the model, it is still interesting to see that the estimated coefficient of dummy variable of positive sentiment is positive, indicating that rumors with positive sentiment spread slower than rumors with negative sentiment. Figure 4 shows that selected seasonality features do have an association with waiting times. Tweets posted during the hours of 6, 9 and 10 in the morning, correlate to higher waiting times. However, later during the noon break around 11 a.m. to 12 p.m. tweets are associated with shorter waiting times. During the night around 20 p.m. and 21 p.m., waiting times of tweets tend to become shorter as well. Seasonality results appear to match general usage patterns of Twitter.

Applying a more flexible model allows us to confirm many of these results. In the regression tree model we see similar patterns of association, as seen in Figure 5. The feature of seasonality plays a significant role in waiting times. Thus, we have a stronger confidence in the association between patterns of usage of Twitter users and waiting times. An interesting finding in the regression tree is that the features - *favorite counts* and *retweet counts* of original tweets serve as important splits in this model. Tweets that have been favorited by more Twitter users are likely to have longer median waiting times. Similarly, more retweet counts (greater than 2.5) are correlated to longer waiting times. Since more retweet counts and favorite counts might suggest a longer time span from the time point when original tweets were posted to the last time point when the last tweet, median waiting times for tweets might be longer.

In many ways, this split is easily recognized as a partition between tweets that receive no recognition from other users, versus those that do. Further, on the other side of the tree we see tweets that have large numbers of retweets. While these content features are important to note, we focus on the rumor codes. Here again we find support for rumor-affirming tweets having the longest waiting times, compared to rumor-denying or neutral tweets.

7. Discussion

Our results quantifying the speed of rumor transmission during a crisis event, demonstrate significant variability in the speed that messages propagate in social media spaces. While there are many factors associated with transmission speed, such as URLs and hashtags, we focus on rumor dimensions, calling attention to the consequences of differential spread. Grounded in prior sociological studies of rumor, we consider the differences between content that affirms a particular rumor versus content that denies that rumor. Our analysis demonstrates that rumor-affirming content tends to have longer waiting times than rumor-denying content, across five different

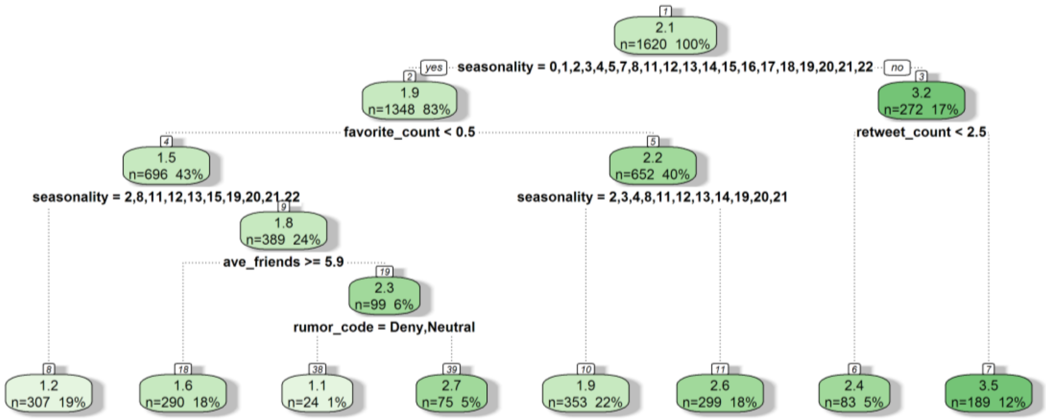


Figure 5. Model of regression tree

rumors that spread during a hostage crisis event. Implications for rumoring during crisis are widespread. First, our analysis suggests that rapid, coordinated denials of rumors could outpace rumor affirmers since denials tend to be spread more rapidly. In the context of false rumors, i.e. misinformation, this implies that crowd corrections may be viable in preventing misinformation cascades as the public may respond to denial more quickly.

Beyond implications for social studies of rumoring behavior, this research offers insight for emergency responders and practitioners. Demonstrating that the public responds differently to different rumor content could be encouraging for emergency responders who cite the potential for widespread proliferation of misinformation as one of the primary reasons for questioning the value of social media in the crisis space.

While the work presented here is a first look at measuring and modeling the speed of information transmission during one crisis event and across multiple rumors, it also aims to provide a foundation for future work in this area. Importantly, we recognize the need to bring together information transmission size and speed. In addition, we continue to replicate these results across additional case studies.

8. Conclusion

Social media continue to play a prominent and important role in information dissemination during crisis events. Both the public and government officials and organizations charged with crisis management can utilize social media platforms to disseminate event-related information. While social media offer many affordances in this setting – namely rapid, real-time communication – it is important to understand how rumor spreads in these settings. Here we fill notable gaps in our understanding of rumoring on social media during crisis, by quantifying and modeling the speed of transmission. More specifically, we contrast rumor-affirming posts with rumor-correcting posts on Twitter during a notable hostage crisis to reveal differences in transmission speed. This work extends understanding of social behavior in online environments.

9. References

[1] J. Sutton, L. Palen, and I. Shklovski, “Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires”, In Proc. of the 5th ISCRAM, Washington, D.C., May 2008, pp. 624-632.
[2] L. Palen, S. Vieweg, J. Sutton, S.B. Liu, and A.L. Hughes, “Crisis Informatics: Studying Crisis in a Networked World”, In. Proc. of the Third International Conference on E-Social Science, 2007.
[3] X. Guan, and C. Chen, “Using Social Media Data to Understand and Assess Disasters”, Natural Hazards, Vol. 74 (#2), November 2014, pp. 837-850.
[4] R. Chen, and Y. Sakamoto, “Perspective Matters: Sharing of Crisis Information in Social Media”, In Proc. of the 46th Hawaii International Conference, IEEE, Hawaii, January 2013, pp. 2033-2041.
[5] S. Vieweg, A.L. Hughes, K. Starbird, and L. Palen. “Microblogging during Two Natural Hazards events: What Twitter May Contribute to Situational Awareness”, In Proc. of the CHI, ACM, Atlanta, USA, April 2010, pp. 1079-1088.
[6] K. Starbird, and L. Palen, “Voluntweeters: Self-organizing by Digital Volunteers in Times of Crisis”, In Proc. of CHI, ACM, Vancouver, Canada, May 2011, pp. 1071-1080.
[7] M. Mendoza, B. Poblete, and C. Castillo, “Twitter Under Crisis: Can We Trust What We RT?”, In Proc. of the First Workshop on Social Media Analytics, ACM, Washington D.C., July 2010, 71-79.
[8] S. Anthony, “Anxiety and rumor”, The Journal of Social Psychology, Vol. 89(1), 1973, pp. 91-98.
[9] S. Tamotsu, “Improvised News. A Sociological Study of Rumor”, Bobbs-Merrill Company, Inc, Indianapolis, 1966.
[10] S. Counts, and K. Fisher, "Taking It All In? Visual Attention in Microblog Consumption." In Proc. of AAAI ICWSM, Barcelona, Spain, 2011, pp. 97-104.

[11] E.S. Spiro, C.L. DuBois, and C.T. Butts, “Waiting for a Retweet: Modeling Waiting Times in Information Propagation”, 2012 NIPS Workshop of Social Networks and Social Media Conference, Nevada, USA, December 2012, pp. 1-8.
[12] J. Sutton, E.S. Spiro, B. Johnson, S. Fitzhugh, and B. Gibson, “Warning Tweets: Serial Transmission of Messages during the Warning Phase of a Disaster Event”, Information, Communication & Society, Vol. 17 (6), 2014, pp. 765-787.
[13] S. Petrovic, M. Osborne, and V. Lavrenko, “RT to Win! Predicting Message Propagation in Twitter”, In Proc. of AAAI ICWSM, Barcelona, Spain, July 2011, 586-589.
[14] N. Naveed, T. Gottron, J. Kunegis, and A.C. Alhadi, “Bad News Travel Fast: A Content-Based Analysis of Interestingness on Twitter”, In Proc. of the 3rd International WebScience Conference, ACM, Koblenz, Germany, June 2011, pp. 8.
[15] K. Starbird, and L. Palen, “Pass it on?: Retweeting in mass emergency”, In Proc. of ISCRAM, Seattle Washington USA, May 2010.
[16] S.C. Dodd, “Testing message diffusion in controlled experiments: charting the distance and time factors in the interactance hypothesis”, American Sociological Review, 1953, pp. 410-416.
[17] B. Greenberg, “Diffusion of News of the Kennedy Assassination”, Public Opinion Quarterly, Vol. 28 (2), 1964, pp. 225-232.
[18] Noymer, A. “The transmission and persistence of ‘urban legends’: Sociological application of age-structured epidemic models”, Journal of Mathematical Sociology, Vol. 25(3), 2001, pp. 299-323.
[19] T. Caplow, “Rumors in war”, Social Forces, Vol. 25(3), 1947.
[20] R.L. Rosnow, J.L. Esposito, and L. Gibney, “Factors influencing rumor spreading: Replication and extension”, Language & Communication, Vol. 8(1), 1988, pp. 29-42.
[21] G.W. Allport, and L. Postman, L, The psychology of rumor, Henry Holt and Company, New York, 1947.
[22] K. Lerman, and R. Ghosh, “Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks”, In Proc of AAAI ICWSM, Washington D.C., 2010.
[23] N.B. Ellison, “Social network sites: Definition, history, and scholarship”, Journal of Computer - Mediated Communication, Vol. 13(1), 2007, pp. 210-230.
[24] C.S. Lee, and L. Ma, “News sharing in social media: The effect of gratifications and prior experience”, Computers in Human Behavior, Vol. 28(2), 2012, pp. 331-339.
[25] C. Hui, Y. Tyshchuk, W.A. Wallace, M. Magdon-Ismail, and M. Goldberg, “Information cascades in social media in response to a crisis: a preliminary model and a case study”, In Proc. of the 21st ACM WWW Companion, Lyon, France, April 2012, pp. 653-656.
[26] V. Qazvinian, E. Rosengren, D.R. Radev, & Q. Mei, “Rumor has it: Identifying misinformation in microblogs”, In Proc. of the Conference on Empirical Methods in Natural Language Processing, ACL, Edinburgh, UK, July 2011, pp. 1589-1599.
[27] E.S. Spiro, S. Fitzhugh, J. Sutton, N. Pierski, M. Greczek, and C.T. Butts, “Rumoring during extreme events: A case study of Deepwater Horizon 2010”, In Proc. of the 4th WebScience Conference, ACM, Evanston, IL, USA, June 2012, pp. 275-283.
[28] J. Maddock, K. Starbird, H.J. Al-Hassani, D.E. Sandoval, M. Orand, and R.M. Mason, “Characterizing online rumoring behavior using multi-dimensional signatures”, In Proc. of the 18th ACM CSCW, Vancouver, BC, Canada, February 2015, pp. 228-241.
[29] A. Friggeri, L.A. Adamic, D. Eckles, and J. Cheng, “Rumor cascades”, In Proc. of the 8th AAAI ICWSM. Ann Arbor, Michigan, USA, May 2014, pp. 101-110.
[30] S. Goel, D.J. Watts, and D.G. Goldstein, “The structure of online diffusion networks”, In Proc. of the 13th Conference on Electronic Commerce, ACM, New York, June 2012, pp. 623-638.
[31] J. Yang, and S. Counts, “Comparing Information Diffusion Structure in Weblogs and Microblogs”, In Proc. of AAAI ICWSM, May 2010, pp. 351-354.
[32] B. Suh, L. Hong, P. Pirolli, and E.H. Chi, “Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network.” In Proc. of IEEE SocialCom, Minneapolis, 20-22 Aug. 2010, pp. 177-184.
[33] H. Bozdogan, “Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions.” Psychometrika, Vol. 52(3), 1987, pp. 345-370.
[34] G. Schwarz, “Estimating the dimension of a model”, The annals of statistics, Vol. 6(2), The Institute of Mathematical Statistics, Lake Tahoe, Nevada, USA, 1978, pp. 461-464.
[35] S. Stieglitz, and L. Dang-Xuan, “Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior”, Journal of Management Information Systems, Vol. 29(4), M.E. Sharpe, Armonk, USA, April 2013, pp. 217-248.
[36] P. Bordia, and N. DiFonzo, “Problem solving in social interactions on the Internet: Rumor as social cognition”. Social Psychology Quarterly, Vol. 67(1), 2004, pp. 33-49.
[37] J. Walker, and B. Blaine, “The virulence of dread rumors: A field experiment”, Language & Communication, Vol. 11, Issue 4, 1991, pp. 291-297.
[38] A. Cynthia, E. Fichet, S. Ding, E. Spiro, and K. Starbird, “Keeping Up with the Tweet-Dashians: The Impact of ‘Official’ Accounts on Online Rumoring”, In Proc. of the ACM CSCW 2016.
[39] A. Ahmer, K. Shanahan, F. Chou, Y. Dosouto, K. Starbird, and E. Spiro, “How Information Snowballs: Exploring the Role of Exposure in Online Rumor Propagation”, In Proc. of the ACM CSCW 2016.
[40] K. Starbird, E. Spiro, A. Arif, F. Chou, S. Narisimhan, J. Maddock, K. Shanahan, and J. Robinson, “Expressed Uncertainty and Denials as Signals of Online Rumoring”, In Proc. of Collective Intelligence 2015, San Francisco, CA.
[41] L.J. Fleiss, “Measuring nominal scale agreement among many raters”, Psychological Bulletin, American Psychological Association, USA, Vol. 76(5), 1971, pp. 378-382.