Waiting for a Retweet: Modeling Waiting Times in Information Propagation

Emma S. Spiro
Department of Sociology
University of California
Irvine, CA
`eespiro@uci.edu`

Christopher L. DuBois
Department of Statistics
University of California
Irvine, CA
`duboisc@uci.edu`

Carter T. Butts
Department of Sociology
University of California
Irvine, CA
`buttsc@uci.edu`

Abstract

Social media are increasingly used to disseminate emergency warnings, alerts, and other hazard-related information. In this context, the timing of information propagation is of immediate interest. Time-sensitive information must reach members of the general public before the pertinence of the information expires. In this research we build a preliminary model for the time between information dissemination and redistribution on Twitter, i.e. the *waiting time* of a retweet. We explore the relationship between this time and features about the users involved, the external context, and the message itself. Results suggest shorter waiting times for hazard-related content.

1 Introduction

Recent developments in social media technologies and mobile devices have transformed informal communication by allowing individuals to reach a larger number of individuals across greater distances. As a result, computer-mediated forms of communication, such at microblogs, are increasingly being used for disseminating emergency warnings, alerts, and other hazard-related information [19]. In emergency settings, time-sensitive information must reach members of the general public before a specific time, after which the content may no longer be relevant or useful. For example, flood evacuation alerts must reach the target population before flooding occurs; failure to do so can have severe consequences. Moreover, collecting and organizing information about events in a timely and efficient way can aid in reducing deaths, economic losses and social disruption [23].

In the context of extreme events, *serial transmission* of content from one individual to the next is vital for information dissemination. Typically a small number of individuals are exposed to the initial warning and others hear through a process of information exchange [1, 14]. Indeed, previous social science research has shown that information exchange in emergencies primarily occurs via pre-existing social ties [12, 30]. While early work in this area focused on face-to-face communication, the phenomenon extends to computer-mediated forms of communication as well. Online social networks thus provide "soft infrastructure" for information dissemination which can be utilized in the event of a crisis situation. In addition, they provide viable platforms for collective problem solving and efficient coordination in response efforts.

The ability to broadcast, redistribute, and organize information quickly makes Twitter, a popular microblogging service, a viable emergency warning dissemination system. However, questions arise about the efficiency and practicality of this system. Does information get lost among the myriad content making it hard to find thus delaying or reducing exposure? Does hazard-related information posted to Twitter actually reach individuals faster than traditional media outlets? Information propagation has been studied in many online contexts, however, less attention has been paid to the timing of this phenomena. In particular, the waiting time between the first posting of content and instances of serial transmission is the quantity of interest for this research. *How do waiting times differ based on the social features of the poster, the message's content, and the external context at the time of posting?* The properties of these waiting times have important consequences for the speed and reach of information propagation, especially in the context of hazard-related content.

2 Background

Information exchange via chains of informal social ties has been a topic of longstanding interest to social scientists [21, 16, 38]. One of the most well-known demonstrations of information propagation [1] illustrates the process of serial transmission and its susceptibility to information distortion. Since early studies social theorists have proposed many other factors thought to influence information transmission [8, 6, 32]; notable examples include anxiety levels and the relevance of the message content to behavior [3, 40]. The question of how individuals evaluate content and decide to pass along information continues to be a popular theme in modern research [22, 5].

Prior work shows that social ties are pathways for information exchange in the event of a disaster or emergency situation. In many cases, pre-existing social ties are the *primary* means by which people obtain time-sensitive information, especially when official sources are slow to release updates or are unavailable [12, 30, 29]. This informal communication process plays a critial role in collective problem solving [32]. Moroever, in the context of extreme events, individuals are likely to pass along time-sensitive information *without* first evaluating content due to the high potential cost of failing to transfer critical information to others [12].

When a disaster or crisis event occurs, informal collection and dissemination of event-relevant information is increasingly conducted via social media [36, 34]. This has not gone unrecognized by emergency management practitioners, who are beginning to recognize the power of this "soft infrastructure" in a response context. In fact, the integration of social media and mobile devices enables of collective sense-making despite a chaotic environment in part because it utilizes pre-existing social ties. The success of such this technology has been demonstrated in recent years via the responses to major episodes of rioting in Mumbai and London, large earthquakes in Haiti, Japan, and New Zealand, and other world events [24, 27]. While this progress is encouraging, utilization of social media to facilitate response to extreme events continues to be constrained by our lack of knowledge regarding how such services are typically used for information diffusion and how such use changes in the context of an extreme event.

Retweeting – passing along messages on Twitter – offers a direct means of tracking information diffusion through social ties [25, 5, 22, 4, 41, 11]. Retweeting behavior has been explored specifically in the context of disaster or hazard events. [33] demonstrate that individuals are more likely to retweet information originally distributed through Twitter accounts run by local media and traditional service organizations. Though not specifically about retweeting, [37] demonstrate how Twitter allows for distributed users to share and collectively organize information about local disaster events. Our contribution looks directly at modeling the timing of retweets based on social features of the users involved, the content of the message, and the external context during which it was posted.

3 Dataset

Tweets containing a pre-specified list of hazard-related keywords were collected using the Twitter Search API from the period of January 1, 2011 until September 23, 2011 [7]. Tweets were also collected for a series of "control" keywords – words chosen at random from Ogden's English word list. Over 250 million tweets were collected.

In this work we use a sample of 20,000 non-retweet posts for keyword streams related to mudslides, earthquakes, tornadoes, and the control group. For each tweet we use the Twitter Rest API to obtain a list of its direct retweets (this does not include higher-order retweets, i.e. retweets of retweets).[1] Each message has an associated posting time, along with a set of attributes for the poster. The proportion of original posts that are retweeted varies across keywords. Differences also exists in the

[1] This query is censored at 100 results. This limit was not reached for any of the sampled data.

group	keyword	proportion of use hazard-related	mean waiting time	median waiting time	standard error of mean waiting time
control words	chalk	0.00	1036.11	15.40	139.10
	collar	0.00	1050.81	14.07	217.14
	form	0.00	885.99	19.05	224.52
	secretary	0.00	583.17	19.60	123.14
	trouble	0.00	731.23	7.16	135.69
earthquakes	aftershock	0.61	1429.73	9.85	308.24
	earthquake	0.75	376.85	12.20	69.01
	magnitude	0.76	507.42	14.93	137.91
	quake	0.75	1205.83	16.73	196.51
	seismic	0.28	302.94	20.38	42.48
mudslides	buried	0.04	2961.13	14.68	518.60
	debris	0.68	1496.78	14.02	363.94
	landslide	0.21	1260.95	11.01	322.09
	mudflow	0.84	638.49	19.27	318.40
	mudslide	0.81	309.76	14.76	66.62
tornados	funnel+cloud	0.87	54.03	5.47	5.79
	shelter	0.06	966.11	24.97	180.18
	tornado	0.38	421.93	10.28	110.62
	twister	0.01	1089.15	8.58	317.51
	wind	0.60	408.09	12.82	77.76

Table 1: Descriptive statistics of the retweets waiting times. Shown are estimated proportion that a keyword is used in the content of hazard-related content, the mean waiting time, the median waiting time, and the standard error of the mean waiting time.

proportion of unique individuals in the sample. While retweeters tend to be unique, many of the original tweeters on a given keyword are captured more than once. This may indicate some topics are dominated users who tend to post frequently with the same keywords.

Waiting times can be calculated by computing the difference between the time of the original tweet and each of its retweets. In turn we obtain a distribution of waiting times for each keyword. It is important to note that not all tweets are retweeted. In fact, typically only around 10% of posts have one or more retweets. Users may manually specify a retweet by copying text and reposting manually. In this case the retweet may be missing because it is not identified as such by Twitter, and is therefore not included in our dataset.

In Table 1 we show basic descriptives about the waiting times for each keyword. In Figure 1 we compare the mean waiting time in minutes for each keyword in the dataset. Keywords are grouped according to their topic e.g. mudslides, control, etc. Control keyword have consistent mean waiting times; we find more variability in the hazard keywords. Keyword such as "funnel cloud" have small waiting times while "twister" has longer waiting times on average. These differences may result from the differential tendency for words to be used in the context of hazard-related conversation.

Keywords are used for many different purposes: *mudslide* could be used to talk about a hazard event or to refer to a drink. In order to understand which keywords are more representative of hazard-related conversation on Twitter. We estimate for the proportion of times a keyword is used in the context of a hazard event, as seen in Table 1. These estimates were obtained by manually coding 100 randomly sampled tweets for each keyword from the data collected.[2] This feature gives an estimate of the "hazardousness" of the keyword in question. In the case of Figure 1 the fact that 87% of uses of "funnel cloud" are hazard-related while only 1% of uses of "twister" refer to hazard events, may explain some of the difference in waiting time distributions.

To illustrate potential differences in the waiting times we compare the observed waiting time distributions for "funnel cloud" and "chalk" in Figure 2. "Funnel cloud" has a more peaked waiting time distribution, with shorter waiting times on average. "Chalk", on the other hand, has a flatter distribution with longer waiting times. This exploratory analysis suggests a possible difference between hazard and control related content.

[2]In ongoing work, we use Amazon Mechanical Turk to get a better estimate of this proportion.

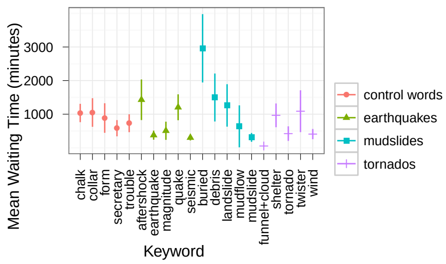

Figure 1: Mean and 95% confidence interval of the mean for observed retweet waiting times for each keyword.

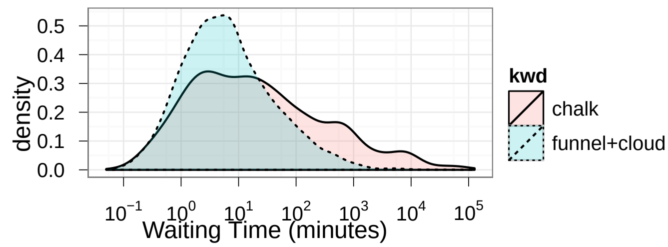

Figure 2: Waiting time distributions. Illustrative example of potential differences between hazard and control related content.

4 Statistical Models for Waiting Times

The quantity of interest in this research is the time between a tweet and each of its direct retweets. Our goal is to characterize the relationship between waiting time and known quantities about the tweet, such as its content, the user who created it, the user who retweeted it, and the external context at the time of posting.

A number of different factors affect this waiting time. We study each stage in turn: the *exposure* of a given tweet, the *attentiveness* of the exposed user, and *propensity to retweet*. Exposure occurs when a user follows the original poster, follows a list of interest or searches for a relevant keyword. Attentiveness refers to the differences in how users perceive the tweets they are exposed to, i.e. whether they notice them, comprehend them, or ignore them [10]. For example, some users may consistently see, read, and decide to retweet within a minute of the original post, while others do not see the tweet until hours later, if ever. In many cases this delay may be determined by the services individuals use to manage their Twitter accounts; use of web-based versus mobile clients, for example. Finally, the propensity to retweet reflects a decision that will vary by user, the subject matter of the tweet, and the context in which the decision is made. During extreme events, the hazard context may be particularly important as previous research indicates that actors are less likely to evaluate the content before passing it to others when it refers to warnings or alerts [12].

To model retweet times, we assume a linear model on the expected value of the log waiting time of the ith retweet $\text{E}[\log y_i] = \mathbf{x}_i'\beta$ where $y_i \in \mathbb{R}^+$ is the waiting time of retweet i, $\mathbf{x}_i \in \mathbb{R}^P$ is a vector of P covariates, and β is a vector of model parameters. The likelihood of N retweets is given by $\mathcal{L}(y|\theta) = \prod_{i=1}^{N} \mathbf{x}_i'\beta$, and we compute maximum likelihood estimates by optimizing the log likelihood using the `lm` function in the R statistical computing environment [28].

Since waiting times are strictly positive, it is reasonable to model the log transform of the response as we have done above. Due to the log transform, the interpretation of each parameter is multiplicative: two groups of tweets with a unit difference for some covariate x_{ip} will have an expected ratio of median waiting times of e^{β_p}, all else held constant. This is a simplistic model, but it provides an initial characterization of the waiting time. Future work will explore more sophisticated waiting time models.

In the following sections we discuss in detail various known quantities that we have collected which may have a relationship with the waiting time. A summary of these features in shown in Table **??**.

4.1 Social Features

As information exchange on Twitter is automated through a series of following relationships, the structure of this network impacts the communication dynamics and, in turn, observed waiting times between tweets and retweets. After all, a necessary condition for serial transmission is *exposure*. One cannot pass along information without first being privy to it. Differential exposure would have consequences for the distribution of waiting times. For example, if more people were on average initially exposed to a message containing keyword k than a message containing keyword j, there is greater opportunity for smaller waiting times to be observed for tweets containing keyword k.

Exposure can occur via several different mechanisms. Individuals may follow lists of interest, or search for relevant keywords. Unfortunately, getting reliable estimates for these exposure cases is difficult. We can, however, easily obtain data on the initial exposure based on following relationships. In this case we approximate exposure based on the number of followers of the original poster. Numbers of followers and friends (i.e. the number of initiated following relations) of the original poster have consistently been shown to be correlated with retweetability [25, 35]. Social features may also impact the propensity of the exposed users to retweet. Arguably, Twitter users are able to judge authoritativeness or credibility based on the number of followers of a given account [15, 9, 26]. These node level properties of the Twitter follower networks will likely impact waiting times as they influence exposure and propensity to retweet a given message.

Attentiveness of the recipient is important factor in waiting times because it affects the time of exposure. One way in which we might estimate attentiveness is to consider the activity of the retweeting user. One such measure of activity is the total number of statuses posted by the user. We consider this quantity for both the tweeter and retweeter. Additional user features considered here include whether the user is "verified." Verification is used to confirm the authenticity of user accounts. In previous work, verification has been shown to have significantly impact of retweeting behaviors [25]. Users with verification may be considered influential or authoritative sources, which in turn may affect people's attentiveness to content posted by these users or the propensity for others to retweet messages posted by these users.

4.2 Content Features

There is a strong tradition of research on how information or message content affects its subsequent transmission. Studies have shown that dread rumors tend to be more prevalent in the population [40]. Work on urban legends also supports the claim that dreadful or disgusting content in rumors increases the chance of transmission [17, 39]. Other influential factors in the content itself may include references to authoritative or credible sources [18, 15].

The keywords in each tweet may themselves impact the waiting time. Each keyword varies in the extent to which it is actually used to describe hazard-related events on Twitter. As previously discussed, we include a estimate of the proportion of use which is hazard-related for each keyword of interest. Content features such as these are important factors that may influence the propensity for users to retweet as well as the attentiveness of exposed users.

Research on predicting retweet occurrence explores a number of these features [25, 35]. This work suggests that the number of hashtags, URLs, and mentions of other users are highly correlated with retweetability. Hashtags allow users to add metadata to message, in a sense specifying content channels of interest. This convention aids in indexing content and subsequently facilitates search. If users can easily and quickly find content of interest, waiting times for retweets may be smaller due to faster exposure times. Inclusion of external URLs is another method for sharing information. In many cases these links point to new articles, photos, or even content that exceeds the 140 character limit. Additional references may impact credibility and therefore retweetability. We also consider mentions of other users, measured by the number of @ mentions in the text. Again, third party mentions may add additional context for the information and impact its reception with users.

4.3 External Context

The context of interaction is known to structure the way individuals response of others [2]. Acts of communication taken place within a social, economic and political environment. Indeed, tweets reflect current events, newsworthy stories, gossip, etc. In the case of hazard-related content, the context remains important to the saliency and relevance of the information.

The time a tweet is posted to that individuals stream likely impacts the chance that others are exposed to that message. Tweets posted during the morning in one country may be received in the middle of the night in a different location. These differences impact exposure and attentiveness. Since Twitter has a global user base we consider the time of posting in the time zone of the retweeter. We further consider basic seasonality indicators for day of week and time of day of the tweet.

Information is more likely to be important, time-sensitive, and relevant to potential behavior during an hazard-related event than during period of no events. We utilize data from the National Oceanic

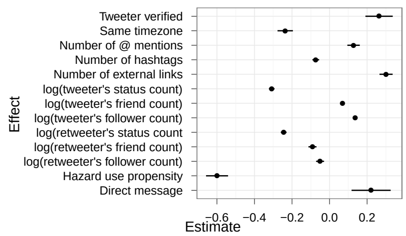

Figure 3: Plot of maximum likelihood estimates of social and content features. (Intercept $\beta_0 = 11$ left off for visualization purposes.)

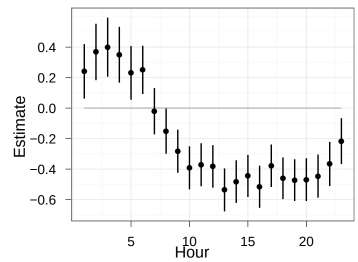

Figure 4: Plot of maximum likelihood estimates of hour of day seasonality.

and Atmospheric Administration[3] on tornado warnings and data from the U.S. Geological Survey[4] on earthquake events to explore the effect of event occurrence on retweeting. Event saliency can influence both the propensity to retweet content as well as the attentiveness of users. Individuals may be more likely to search out content and thus be exposed faster. Additionally, emergency warnings and alerts present important information that had direct relevance to the behaviors of affected individuals. In each case, the consequence of an event may be shorter waiting times.

5 Results

To gain a sense of the explanatory power of the various features we discuss above, we compare the fit of the model using different combinations of features. To describe model fit we compute the *Bayesian Information Criterion* (BIC), a statistic which computes the loglikelihood as a measure of model fit while penalizing for the number of parameters used in the model [31]. In Table 2 we show the results of this analysis, which suggest that using each category of covariates play a role in the performance of the model. The best fit results from combining social, content and context features suggest each plays a role in waiting times.

Model	BIC
social + content + context	231275.92
social + content	231927.95
content + context	232292.08
social	237442.02
context	237132.26
content	233760.06

Table 2: Model selection via BIC. We compare the inclusion of each category of covariates. Lower values are better.

In Figure 4 we show the effect of the hour of the original post computed with respect to the timezone of the retweeter from the best performing model ($R^2 = .1281$). These results show a clear pattern of hourly seasonality. Waiting times are longer during night hours and become shorter during the day, suggesting optimal hours for fast retweeting. Results from the day of week seasonality term do not show as clear of a pattern; only Saturday shows longer waiting times than the reference category, Monday, with an estimate of 0.25 with standard error 0.039.

In Figure 3 we show selected estimates for social and content coefficients from the best performing model. Several features appear to be correlated with a decrease in the waiting time for a retweet. For example, $e^{-.31} = 0.733$ is the ratio of median waiting times between two populations whose log tweeter's status count differ by one unit. Likewise, for a 1% increase in the "hazard use propensity" estimate there is an expected 0.9% decrease in the median waiting time.

We find that inclusion of external URLs and @ mentions of other users (both third party and directed) are associated with longer waiting times. Hashtags, on the other hand are associated with shorter waiting times. Perhaps the inclusion of hashtags allows for increased exposure since individuals may search for content by hashtag. URLS and @ mentions present an interesting case; one explanation of these results might be that users are more likely to follow the external links and/or look at other users

[3]www.noaa.gov
[4]www.usgs.com

when evaluating content, delaying the overall process. The same might explain longer waiting times being associated with verified users. These features may also lead to heavier tailed distributions.

Social features are also associated with waiting times. The activity level of both the original poster and the retweeter are both correlated with shorter waiting times. High activity may indicate more attentive usage patterns. Interestingly, a higher number of followers and friends for the tweeter have positive effects (i.e. longer waiting times) while those of the retweeter have negative effects.

Finally, consider the relationship between the estimated hazard use proportion and waiting times. Recall, this estimate captures the propensity for keyword to be used in the context of hazard-related conversation. There is large variability in this proportion across keywords, even within the same hazard group. We find that the effect of hazard use is strongly associated with waiting times; higher propensities have smaller observed waiting times on average. These results suggest that hazard-related content demonstrates shorter waiting times when compared with control words.

We also considered event occurrence, which captures the event of a hazard in the 48 hour period prior to the posting of the tweet. Since only data on earthquake and tornado events is available, we fit this model to a subset of the full data. Results indicate that waiting times are significantly smaller after recent events on average. However, the magnitude of the effect is smaller than social and content features.

6 Discussion

In this research we compare waiting times between online messages and instances of their redistribution. Our results suggest that the distribution of waiting times is related to a variety of factors.

Content features such as hashtags are strongly correlated with smaller waiting times. The number of followers of the tweeter, on the other hand, is associated with longer waiting time. These results suggests that exposure through search may be an important factor. Tweets containing hashtags are easily found by users searching for content. This may be especially relevant in the context of extreme events where people are actively seeking specific information. Use of hashtags may be an easy means of increasing visibility of hazard-related content posted to Twitter.

Social features present an interesting case for further study. We might have predicted that a high number of followers for the tweeter would lead to shorter waiting times because initial exposure would be greater. However, results suggest just the opposite, pointing to an alternate mechanism. One plausible explanation is the classic theory of *diffusion of responsibility* [13, 20]. Typically used to explain the bystander effect, the theory attempts to explain phenomena in which the greater the number of people present, the less likely any single individual will take responsibility (e.g. in aiding an individual in distress). In the case of Twitter, messages broadcast to a larger number of individuals may decrease the perceived need among recipients to repost the content. An alternate explanation may be one of message superfluousness. If large numbers of individuals are initially exposed to the original content, retweeting might be redundant. Potential retweeters might assume that others have already been exposed to the content as well.

Comparisons between hazard-related and control-related content also have important implications. The strength with which specific keywords are associated with hazard-related conversation has one of the strongest associations with waiting times. Tweets containing keywords that are typically used to refer to hazard events have smaller waiting times on average. These results suggest that exposure to hazard-related content may in fact be faster than average content. These results have important consequences for disaster practitioners by demonstrating that Twitter could be a viable system for information dissemination. Findings are also important for research on information diffusion more generally.

References

[1] G.W. Allport and L. Postman. *The psychology of rumor.* Henry Holt, 1947.

[2] I. Altman. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* Brooks and Cole Publishing Company, Monterey, California, 1975.

[3] Susan Anthony. Anxiety and rumor. *The Journal of social psychology*, 89:91–98, 1973.

[4] E. Bakshy, J.M. Hofman, W.A. Mason, and D.J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM, 2011.

[5] Danah Boyd, Scott Golder, and Gilad Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*, pages 1–10, 2010.

[6] H. Taylor Buckner. A theory of rumor transmission. *Public Opinion Quarterly*, 29(1):54, 1965.

[7] Carter T. Butts, Jeannette Sutton, and Emma S. Spiro. Hazards, emergency response, and online informal communication project data, 2011.

[8] T Caplow. Rumors in war. *Social Forces*, (3):298–302, 1947.

[9] Carlos Castillo and Marcelo Mendoza. Information credibility on twitter. *Proceedings of the 20th*, pages 675–684, 2011.

[10] S. Counts and K. Fisher. Taking it all in? visual attention in microblog consumption. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[11] Onkar Dabeer and Prachi Mehendale. Timing Tweets to Increase Effectiveness of Information Campaigns. pages 105–112, 2011.

[12] E R Danzig, P W Thayer, and L R Glanter. The Effects of a Threatening Rumor on a Disaster-Stricken Community. Technical report, Washington, DC, 1958.

[13] J.M. Darley and B. Latane. Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology*, 8(4p1):377, 1968.

[14] B.H. Erickson, TA Nosanchuk, L. Mostacci, and C.F. Dalrymple. The flow of crisis information as a probe of work relations. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, pages 71–87, 1978.

[15] Alison R Fragale and Chip Heath. Evolving informational credentials: the (mis)attribution of believable facts to credible sources. *Personality and social psychology bulletin*, 30(2):225–36, February 2004.

[16] Mark S. Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360, 1973.

[17] Chip Heath, Chris Bell, and Emily Steinberg. Emotional Selection in Memes : The Case of Urban Legends. *Journal of Personality*, 81(6):1028–1041, 2001.

[18] J.N. Kapferer. A mass poisoning rumor in europe. *Public Opinion Quarterly*, 53(4):467, 1989.

[19] S. Kumar, G. Barbier, M.A. Abbasi, and H. Liu. Tweettracker: An analysis tool for humanitarian and disaster relief. In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[20] B. Latane and J.M. Darley. *The unresponsive bystander: Why doesn't he help?* Appleton-Century Crofts New York, 1970.

[21] N.H. Lee. *The search for an abortionist*. University of Chicago Press Chicago, 1969.

[22] Sofus A Macskassy and Matthew Michelson. Why Do People Retweet? Anti-Homophily Wins the Day! *International AAAI Conference on Weblogs and Social Media*, pages 209–216, 2011.

[23] S. Mehrotra, C. Butts, D. Kalashnikov, N. Venkatasubramanian, R. Rao, G. Chockalingam, R. Eguchi, B. Adams, and C. Huyck. Project rescue: Challenges in responding to the unexpected. In *In Proceedings of 16th Annual Symposium on Electronic Imaging Science and Technology*, 2004.

[24] O. Oh, M. Agrawal, and H.R. Rao. Information control and terrorism: Tracking the mumbai terrorist attack through twitter. *Information Systems Frontiers*, pages 1–11, 2011.

[25] Miles Osborne and Victor Lavrenko. RT to Win! Predicting Message Propagation in Twitter. *Artificial Intelligence*, pages 586–589, 2011.

[26] A Pal and S. Counts. Identifying topical authorities in microblogs. *Proceedings of the fourth ACM international*, 2011.

[27] A. Potgieter. Social media and the 2011 london riots. In *Annual Conference On WWW Applications*, 2011.

[28] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[29] R J Richardson, B H Erickson, and T A Nosanchuk. Community Size, Network Structure, and the Flow of Information. *Canadian Journal of Sociology*, 4:379–392, 1979.

[30] T.J. Scanlon. Post-disaster rumor chains: A case study. *Mass Emergencies*, 2(126):22–27, 1977.

[31] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[32] Tamo Shibutani. *Improvised News: A Sociological Study of Rumor.* Bobbs-Merrill, 1966.

[33] Kate Starbird and L. Palen. Pass It On?: Retweeting in Mass Emergency. In *Proceedings of the 7th International ISCRAM Conference, Seattle, WA*, number December 2004, pages 1–10, 2010.

[34] Kate Starbird, Leysia Palen, A.L. Hughes, and Sarah Vieweg. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 241–250. ACM, 2010.

[35] 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 *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 177–184. IEEE, 2010.

[36] J. Sutton, L. Palen, and I. Shklovski. Backchannels on the front lines: Emergent uses of social media in the 2007 southern california wildfires. *Crisis*, (May), 2008.

[37] J.N. Sutton. Twittering tennessee: Distributed networks and collaboration following a technological disaster. In *Proceedings of the 7th International ISCRAM Conference–Seattle*, volume 1, 2010.

[38] R H Turner. Rumor as Intensified Information Seeking: Earthquake Rumors in China and the United States. In R R Dynes and K J Tierney, editors, *Disasters, Collective Behavior, and Social Organization*, pages 244–256. University of Delaware Press, Newark, NJ, 1994.

[39] J S Victor. *Satanic Panic: The Creation of a Contemporary Legend*. Addison-Wesley, New York, 1993.

[40] Charles J Walker and Bruce Blaine. The Virulence of Dread Rumors: A Field Experiment. *Language & Communication*, 11(4):291–297, 1991.

[41] Jiang Yang and Scott Counts. Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. *Artificial Intelligence*, 2009.