How Do Multiple Topics in Terse Tweets Affect Retweeting?
Evidence from the 2013 Colorado Floods

Completed Research

Jaebong Son
California State University Chico
Jbson751212@gmail.com

Hyungkoo Lee
HEC Montreal
Hyung-koo.lee@hec.ca

Sung Jin
Metropolitan State University
Simon.Jin@metrostate.edu

Jiyoung Woo
Soonchunhyang University
jywoo@sch.ac.kr

Abstract

The rapid and wide dissemination of disaster information through tweeting and retweeting has made Twitter an important communication channel during disasters. However, due to its 140-character limit, a tweet could be considered uninformative during disasters. Based on Shannon and Weaver’s communication theory, we use the measure of entropy to quantify the extent to which a tweet is considered informative. We theorize that as a tweet’s entropy increases, its informativeness decreases, and importantly, so too does the probability of retweeting. To assess tweets’ entropy, we use topic modeling to discover topics in tweets. Using tweets collected during the 2013 Colorado floods, we empirically examine the relationship between tweets’ entropy and retweet frequency. We take this investigation one step further by examining the interaction effect of the number of URLs on that relationship. Our empirical results demonstrate the negative effect of entropy on retweet frequency. In addition, the effect of the number of URLs depends on entropy. Our findings suggest that tweets’ entropy is an important factor in explaining tweets’ retweeting mechanism during disasters and enhance the understanding of the relationship between short-length tweets and information dissemination. As a result, our study contributes to IS research on the role of Twitter in emergency communication.

Keywords
Terse Tweets, Entropy, Topic Modeling, Disaster Communication, Information Dissemination

Introduction

Characterized by a series of highly dynamic and urgent events (Sellnow and Seeger 2013; Starbird and Palen 2010), disasters are inherently associated with a lack of information (Mitroff 2004) that leads to high levels of situational uncertainty about impending life-threatening events. Under such circumstances, the public in affected areas experiences loss of control and stress (Spence et al. 2007) and thus, actively engages in seeking relevant information through mainstream and social media (Boyle et al. 2004; Procopio and Procopio 2007). While mainstream media plays important roles in providing critical information, it does not provide specific and timely information to local inhabitants in disaster stricken areas (Oh et al. 2013). Social media, on the other hand, is known to provide localized and up-to-date information that helps the public to be aware of their surroundings, to promote precautionary behaviors, and to avoid unexpected disaster-related events (Keller et al. 2006; Latonero and Shklovski 2011). Twitter, among the different social media platforms, has received great attention from both emergency officials and researchers (Chatfield and Brajawidagda 2013; Hughes et al. 2014) due to its ability to quickly broadcast critical information to the
mass public (Vieweg et al. 2010). Moreover, the Twitterverse—where a large scale of online users connect and communicate freely—becomes even more important for disseminating disaster information when mainstream media is disrupted (Winn 2011). For instance, during the 2011 Tohoku earthquake in Japan, people relied on the information disseminated on Twitter for rescue and evacuation tasks due to the disruptions of mainstream media (Chen and Sakamoto 2013).

With the recognition of Twitter’s capability for quick and widespread information dissemination, prior research has examined the mechanisms of information dissemination on Twitter during disasters. Twitter allows users (or twitterers) to post short messages of 140 characters or fewer (or tweets), and other twitterers can re-post such messages through retweeting, a key feature to share the original tweet with a large audience (Compston 2014). Retweeting allows voluntary participation of twitterers to forward interesting and informative tweets to their followers (Starbird and Palen 2010; Sutton et al. 2014b). The 140-character limit and retweeting mechanism are the two main factors that have made Twitter one of the most promising of communication media for disaster communication (Fraustino et al. 2012). On the other hand, Twitter also has a drawback due to its character limit in the sense that a single tweet may be restricted in delivering precise, accurate, and detailed disaster information (Sutton et al. 2015b). According to a survey study by Bean et al. (2016), short-length messages could be considered uninformative and confusing, leading to concerns of information quality and accuracy (Hu and Sundar 2009). This concern can become even more severe as twitterers include multiple topics in a single tweet for the following reasons: first, recipients will struggle with proper interpretation of the main topic of a tweet due to additional peripheral topics; second, multiple topics inevitably cause less information or fewer details per topic due to the character limit. Daft et al. (1987) argued that when the amount of information in a message decreases, this message’s uncertainty increases, hampering correct interpretation by recipients of the intended meaning of the message. Indeed, Shannon and Weaver (1949) communication theory stated that when a message includes multiple topics, its recipients have the freedom of choice in selecting the major topic, which leads to distortions and errors of this message’s main intention. They introduced the measure of entropy to quantify the extent to which a message is informative. Daft’s information uncertainty and Shannon and Weaver’s measure of entropy have important implications for disaster communication. That is, disaster-related information must not confuse recipients, due to the fact that disaster situations do not allow the public-at-risk the luxury of time to explore additional information (Oh et al. 2013; Runyan 2006). Therefore, we need to consider the degree of a tweet’s informativeness (or interpretability) when examining Twitter for disaster communication. In other words, Twitter studies must take into account tweets’ informativeness to better examine Twitter’s features relative to the dissemination of disaster information. In this research, we use Shannon and Weaver’s measure of entropy to quantify the extent of each tweet’s informativeness based on its topics and individual topic proportion, and we examine the effect of entropy on retweet frequency.

The paper is structured as follows: first, we discuss a few salient features of tweets for disaster communication; second, we provide our rationale for using Shannon and Weaver’s measure of entropy as an important independent variable of this study. We then establish research hypotheses. We follow with descriptions of our methodology, data, and empirical results, respectively.

Research Background

Tweets for Disaster Communication

The 140-character length tweets allow twitterers to quickly post up-to-date, first-hand information on disaster events and to effectively share such critical information with others in a timely manner (Hughes and Palen 2009; Murthy 2011; Vieweg et al. 2010). For instance, during the 2013 Boston Marathon bombing, critical information was effectively disseminated and amplified by terse tweets, minimizing further impact (Sutton et al. 2014a). In addition to sharing geo-location information during the 2009 Red River flood in North Dakota (Vieweg et al. 2010), Twitter was also used to exchange experiences and opinions during the 2011 Australian Queensland floods (Shaw et al. 2013). Furthermore, when a series of tsunamis destroyed the major communication infrastructure of Tohoku in Japan, Twitter was the only communication means that allowed the public in the disaster stricken areas to communicate with other people as well as to be aware of constantly moving threats (Acar and Muraki 2011). In fact, the U.S. Federal Emergency Management Agency (FEMA) had already attracted to these practical features of Twitter, so
much so that in 2011 the agency authorized management officials to use Wireless Emergency Alerts (WEA) to disseminate alerts and warnings to the public-at-risk (FEMA 2015). A WEA message is limited to 90 characters (Bean et al. 2016). In 2013, Twitter launched an alerting system in the U.S., Korea, and Japan to efficiently disseminate emergency information to the public (Protalinski 2013). Indeed, text-based, terse messages have been gaining momentum for disaster communication. Traditionally, however, the general guideline for disaster-related messages centers around 1,380 characters per message (Sutton et al. 2015b), which is much longer than the message length of WEA and Twitter. Without doubt, we know very little about how recipients interpret terse, disaster messages and then take appropriate actions accordingly. Bean et al. (2015) and Sutton et al. (2015a) raised a critical concern of short-length disaster messages “[T]erse communication can generate uncertainty, thereby promoting WEA and tweet recipients to mill for additional and confirming information (p. 9),” which in turn negatively affects the dissemination of information during disasters. In terms of communication on Twitter, tweets’ limited length contributes to the quick and rapid circulation of information. However, extremely truncated disaster messages could hinder tweeters from correctly grasping intended meanings and thus spark them up to seek supplemental information in order to better interpret intended meanings (Bean et al. 2015; Sutton et al. 2015b). In that regard, Twitter can be considered a double-edged sword for disaster communication.

To reflect the above-mentioned concern, we leverage Shannon and Weaver’s communication theory (1949). Particularly, the measure of entropy is designed to quantify how a message is noisy — the extent to which a message represents its main topic — which affects the extent of its recipients’ interpretation of a message’s meaning. The next section explains the details of entropy.

**Entropy and Its Implication for Tweets**

According to Shannon and Weaver (1949), entropy estimates a message’s noise, directly affecting its interpretability. There are two main components for calculating a message’s entropy: the number of topics and each topic’s proportion in a message. The following equation illustrates this measure, where $p_i$ is the proportion of the $i$th topic out of $n$ topics in a message $m$ (Shannon and Weaver 1949).

$$Entropy_m = - \sum_{i=1}^{n} p_i \ln p_i$$

For example, when a message includes only one topic with 100% (or 1.0) proportion, recipients will not be confused in selecting this topic as the main topic of the message. In other words, this message does not include any noise, so its entropy is 0. When a message has two topics with different proportions of 90% (or 0.9) and 10% (or 0.1) respectively, recipients will choose a topic with 90% proportion as primary, while possibly considering the other topic with 10% proportion as noise. Therefore, the entropy of this message is 0.325. However, when two topics with the same proportion are presented, recipients will be more confused about choosing the main topic because the two topics have the same proportion. In this case, the entropy is 0.693, which is much higher than the previous two cases. In sum, as the number of topics in a message increases, its entropy increases proportionately, confusing recipients regarding this message’s main topic. Table 1 summarizes entropy values by the number of topics and each topic’s proportion.

<table>
<thead>
<tr>
<th>Topic 1: 100%</th>
<th>Topic 1: 90%</th>
<th>Topic 2: 10%</th>
<th>Topic 1: 50%</th>
<th>Topic 2: 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>0</td>
<td>0.325</td>
<td>0.693</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1. Different Entropy Values by the Number of Topics in a Message**

The notion of entropy to measure a message’s noise fits well with terse tweets due primarily to the fact that Twitter, unlike other social media platforms, strictly limits the character length of tweets. Therefore, once $n$ characters are used to describe one topic in a tweet, other topics in this tweet must be explained by $140-n$ characters, reducing information per topic. For example, with only one topic in mind, tweeters can fully use 140 characters to describe that topic. When tweeters have two topics to convey, they will use $n$ characters for one topic and $140-n$ characters for the other, which possibly causes less specific information per topic in general compared to a single tweet with only one topic. From this perspective, multiple topics in a tweet cause the following two problems: first, as we discussed, tweeters can be confused in choosing a tweet’s main topic among its peripheral topics; second, tweeters can have difficulties in properly
interpreting a tweet’s main topic due to reduced information per topic. Summarily, as a tweet’s entropy increases, its interpretability decreases, and this phenomenon in turn provokes recipients’ additional information-seeking. Even though studies on the use of Twitter for disaster communication have focused on identifying factors that explain information dissemination, little is known about how tweets’ entropy affects information dissemination on Twitter. In the following section, we develop a set of hypotheses to examine the relationship between tweets’ entropy and their retweeting.

Research Hypotheses

As disaster events unfold, people become “information hungry” (Mileti and Sorensen 1990, p. 3.8). They immediately seek disaster-related information from available sources including television, radio, newspapers, and social media. Such information-seeking behavior is termed alternatively, sense-making (Oh et al. 2013; Sutton et al. 2014b) or milling (Mileti and Sorensen 1990; Oh et al. 2015). Desiring accurate and adequate information (Griffin et al. 2002), people receiving uninformative messages seek to verify whether or not they correctly interpret the intended meanings of these messages (Fraustino et al. 2012). Daft et al. (1987) and Zeng et al. (2010) demonstrated a significant negative correlation between uncertain information and information dissemination. Communication on Twitter is not an exception. For example, on recognizing a received tweet’s uninformativeness, twitterers become active in seeking further information for clarification by searching for other tweets, interacting with other twitterers, or turning to other information sources (Spiro et al. 2012). When they are convinced by the additional information that helps them to have the correct interpretation of this tweet, they re-post or retweet this tweet. Otherwise, twitterers may question this tweet, rather than retweeting. In this sense, entropy, as a measure for tweets’ informativeness, has a negative relationship with retweeting. Therefore, we propose the first hypothesis:

\[ H1. \text{ A tweet’s entropy is negatively associated with its retweet frequency.} \]

Twitter URLS are an interesting convention for overcoming tweets’ short message length (Hughes and Palen 2009; Purohit et al. 2013; Spiro et al. 2012). Twitterers leverage URLS when large amounts of information must be shared (Lachlan et al. 2014). Bruns and Stieglitz (2012) and Ma et al. (2013) reported that twitterers included URLS to make their tweets informative. In particular, Hughes and Palen (2009) argued that URLS played an important role in providing the affected public rich information during disasters. Suh et al. (2010), on the other hand, provided empirical evidence demonstrating the positive relationship between URLS and tweets’ retweet frequency. In general, embedding URLS in tweets is considered beneficial for twitterers to provide in-depth information to others. On the contrary, Sutton et al. (2015a) and Burnap et al. (2014) found that URLS have a negative impact on retweeting. Pervin et al. (2014) found both positive and negative effects of URLS according to different stages of disaster events. Such discrepancies can be explained if we consider tweets’ entropy. In other words, when a tweet’s entropy is zero, its recipients may or may not process information provided by URLS, primarily because they can successfully interpret this tweet’s meaning without URLS’ supplemental information. However, an increase in a tweet’s entropy requires that recipients seek more information to better interpret this tweet’s main meaning, in which case, URLS could play an important information source. Therefore, the effect of URLS on retweeting could be positive. From this perspective, we argue the interaction effect of URLS on the relationship between entropy and retweet frequency. The following hypothesis represents our interest:

\[ H2. \text{ The negative effect of entropy on retweet frequency depends on the number of URLS, such that this negative effect becomes weaker as the number of URLS increases.} \]
Data and Empirical Analysis

We used tweets about the 2013 Colorado floods to test our hypotheses. Project EPIC (Empowering the Public with Information in Crisis), hosted by the Department of Computer Science at the University of Colorado Boulder, collected tweets about the floods by systematically searching Twitter through the streaming API of Twitter (Dashti et al. 2014). The project produced a total of 106,479 original tweets and 124,251 retweets between September 11 and September 29, 2013. For data analysis, we chose tweets and their retweets posted during the most active period between September 12 and 23, accounting for 96.19% of the total original tweets (102,426) and for 91.85% of retweets (114,124).

Topic Modeling

By following Shannon and Weaver’s communication theory (1949), we leveraged entropy to quantify a tweet’s noise by discovering its topics and each topic’s proportion. Fortunately, Latent Dirichlet Allocation (LDA), a topic modeling algorithm (Blei 2012), is a good fit with our research in the sense that it produces topics and individual topic’s proportion per tweet. Therefore, we utilized the Machine Learning for Language Toolkit (MALLET), a Java library of the LDA-based topic modeling (McCallum 2002), to extract topics in tweets. MALLET has been widely used in studies related to enterprise blogosphere e-commerce (Singh et al. 2014), patent analytics (Hu et al. 2014), and crime prediction using tweets (Gerber 2014). The length of tweets is short, and therefore we expected one topic per tweet in general. Consistent with McCollister et al. (2015), we set the alpha parameter of MALLET to as low as 0.25 with 1,000 sampling iterations. Instead of using only uni-gram words as inputs for topic modeling, we used n-gram noun phrases (i.e., “colorado flood,” “flash flood warning,” and “pet disaster kit checklist”) to obtain more reliable topic models (Poulston et al. 2015). By using TweetNLP (Owoputi et al. 2013), we assigned a part-of-speech (POS) to each word in a tweet. Based on each word’s POS tag, we extracted multiple-word noun phrases. The maximum word count of the extracted n-grams was 6. A hashtag is an essential techno-linguistic feature of Twitter that reflects a tweet’s topic (Boyd et al. 2010; Cotelo et al. 2014; Ma et al. 2013; Oh et al. 2015). Hence, along with the identified n-gram noun phrases, hashtags embedded in tweets were used in modeling tweets’ topics. We also expanded MALLET’s basic stopwords based on TweetNLP’s POS tags. For concreteness, TweetNLP reserves a POS tag G to represent foreign words and “garbage,” and thus we used G-tagged words as stop words. To identify the optimal number of topics for the dataset, we produced 199 topic models by increasing the number of topics from 2 to 200, and then we calculated each topic model’s perplexity to measure its generalizability (Blei 2012). As a result, we found 57 topics as the optimal topic numbers of the dataset. It turned out that 67.4% of tweets had only 1 topic, 28.7% presented 2 topics, and approximately 3.5% included three topics.

Statistical Analysis

Several Twitter studies have consistently found that retweets are heavily concentrated within the 24 hours after posting of the original tweets. Achananuparp et al. (2012) and Van Liere (2010) reported that over 95% of the total retweets in their datasets were made within a day. Likewise, Kwak et al. (2010) noted that 75% of the retweets in their study were disseminated within 24 hours. Similarly, our data shows that 89.73% (95,549) of the total retweets (106,479) were made within 24 hours after posting. Therefore, the dependent variable of our research is the retweet frequency within the 24-hours after posting. With this count dependent variable which is nonnegative and skewed to the right, the Poisson or negative binomial regression analysis is considered to be better than other regression models (Cameron and Trivedi 2013; O’hara and Kotze 2010). Particularly, the negative binomial regression is more appropriate than the Poisson when over-dispersion occurs, indicating that the variance of a dependent variable is considerably larger than its mean (Hilbe 2011). As Table 2 shows, the variance of the dependent variable is larger than its mean, so we tested the dependent variable’s over-dispersion by following the procedures recommended by Cameron and Trivedi (2013). We confirmed the over-dispersion and the heteroscedasticity of variance (Breusch and Pagan 1979) in our data. As a result, we leveraged a robust negative binomial regression to evaluate our empirical models. Along with the factors that are known already to affect retweeting, we included the numbers of words and hashtags to control the length of tweets, such that we were able to statistically examine the effect of entropy on retweeting, regardless of the tweet length used. We summarize all the variables in Table 2.
How Do Multiple Topics in Terse Tweets Affect Retweeting?

Table 2. Variable Description

Table 3 shows the empirical results of the robust negative binomial regression of the dependent variable, the retweet frequency (\(\text{Retweets}_{24h}\)), on the independent variables. We generated four empirical models. Model 1 is the baseline model that includes only 6 control variables. Model 2 added Model 1 tweets’ entropy. Model 3 added Model 1 the number of URLs. The final model, Model 4, added Model 3 an interaction term between entropy and the number of URLs. As Aiken et al. (1991) recommended, entropy and the number of URLs were mean-centered from their means before being estimated. We utilized Model 4 to evaluate our hypotheses, because Model 4 includes all the variables of our interest (Wald Chi2=11211.41; df=9; \(p<0.000\)) and does not have multicollinearity issues (the average VIF of 1.56).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient (Robust Err.)</th>
<th>Coefficient (Robust Err.)</th>
<th>Coefficient (Robust Err.)</th>
<th>Coefficient (Robust Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy(_i)</td>
<td>-</td>
<td>-1.409*** (0.0438)</td>
<td>-1.405*** (0.0435)</td>
<td>-1.376*** (0.0447)</td>
</tr>
<tr>
<td>Supplemental Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URLs(_i)</td>
<td>-</td>
<td>-</td>
<td>0.0379 (0.0233)</td>
<td>0.0670** (0.0227)</td>
</tr>
<tr>
<td>Entropy(_i) × URLs(_i)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.462*** (0.0697)</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words(_i)</td>
<td>0.0569*** (0.00357)</td>
<td>0.0542*** (0.00367)</td>
<td>0.0563*** (0.00378)</td>
<td>0.0577*** (0.00372)</td>
</tr>
<tr>
<td>Hashtags(_i)</td>
<td>0.262*** (0.0100)</td>
<td>0.261*** (0.0100)</td>
<td>0.263*** (0.00999)</td>
<td>0.265*** (0.00901)</td>
</tr>
<tr>
<td>Ln(Followers(_i,t))</td>
<td>0.745*** (0.0101)</td>
<td>0.721*** (0.00946)</td>
<td>0.721*** (0.00947)</td>
<td>0.719*** (0.00943)</td>
</tr>
<tr>
<td>Ln(Followees(_i,t))</td>
<td>-0.0873*** (0.00819)</td>
<td>-0.0795*** (0.00787)</td>
<td>-0.0798*** (0.00786)</td>
<td>-0.0784*** (0.00778)</td>
</tr>
<tr>
<td>Ln(Likes(_i,t))</td>
<td>0.121*** (0.00639)</td>
<td>0.125*** (0.00600)</td>
<td>0.125*** (0.00613)</td>
<td>0.126*** (0.00612)</td>
</tr>
<tr>
<td>Ln(Status(_i,t))</td>
<td>-0.434*** (0.00932)</td>
<td>-0.438*** (0.00860)</td>
<td>-0.439*** (0.00856)</td>
<td>-0.438*** (0.00855)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.756*** (0.0171)</td>
<td>-0.826*** (0.0162)</td>
<td>-0.826*** (0.0161)</td>
<td>-0.820*** (0.0161)</td>
</tr>
</tbody>
</table>
How Do Multiple Topics in Terse Tweets Affect Retweeting?

Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Wald $\chi^2$</th>
<th>AIC (BIC)</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8106.49***</td>
<td>2.067 (-969890.9)</td>
<td>102,426</td>
</tr>
<tr>
<td>2</td>
<td>10790.34***</td>
<td>2.038 (-972824.7)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11035.56***</td>
<td>2.038 (-972820.6)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11211.41***</td>
<td>2.037 (-972907.8)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Results

Model 4 provides empirical evidence supporting the relationship between entropy and retweet frequency: as tweets’ entropy increased, their retweet frequency significantly decreased (Wald Chi$^2$=947.43; df=1; p<0.000). Therefore, Hypothesis 1 is supported. In addition, the effect of entropy is consistent across Models. We observe the significant interaction effect of the number of URLs on the relationship between entropy and retweet frequency (Wald Chi$^2$=43.93; df=1; p<0.000). More details are as follows: first, the main effect of the number of URLs was positive on retweet frequency, such that an additional URL increased retweet frequency by 6.93% on average. Second, this effect became significantly stronger as tweets’ entropy increased. In other words, the negative effect of entropy on retweet frequency weakened as the number of URLs increased. The meaning of this interaction is that when recipients need more information to better understand tweets with high entropy, information linked by URLs can be supplemental to these tweets. Therefore, Hypothesis 2 is statistically significant. The empirical results demonstrate well our major research interests that tweets’ entropy negatively affects the dissemination of information in tweets (i.e., retweeting), and that the effect of the number of URLs is conditional on entropy.

Discussion

As disasters impend, critical information must reach the public in disaster stricken areas in a timely manner (Bean et al. 2015; Mileti and Sorensen 1990; Mitroff 2004). As a means to quickly disseminate such critical information, short-message platforms for disaster communication have gained momentum (Bean et al. 2015; Sutton et al. 2015a). Some researchers have been concerned about the use of short message services (i.e., WEA and Twitter) for disseminating critical, disaster-related information due to restricted information intrinsic to short-length messages (Bean et al. 2016; Bean et al. 2015; Sutton et al. 2015a). However, prior to this study, no empirical research on this concern has been conducted. This study considers the negative aspect of short-length messages based on the foundation of Shannon and Weaver’s communication theory (1949). In other words, in this research, we introduced a new variable, entropy, for Twitter research, and provided empirical evidence for its affect based on the assumption that as a tweet’s entropy increases, its retweet frequency decreases due to additional information-seeking by the tweets’ recipients. The significant statistical evidence about entropy indicates that twitterers’ retweeting is negatively associated with tweets’ entropy. Another interpretation of this result is that the number of topics in a short-length tweet reduces its retweeting, because a tweet’s entropy is mostly determined by its number of topics. In addition, the effect of the number of URLs is conditional. That is, the effect of the number of URLs, as additional information source, becomes stronger as tweets’ entropy increases. We believe this phenomenon would be particularly true due to the 140-character limit of tweets and the highly uncertain and ambiguous nature of disaster situations.

The study offers practical implications for twitterers and emergency practitioners. First, due to tweets’ limited character length, twitterers and emergency practitioners should avoid including multiple topics in a single tweet. Instead, they can craft a series of tweets to disseminate multiple topics. We recommend one topic per tweet. Second, URLs, as a means of delivering rich information, are only useful when a tweet presents high entropy (or has multiple topics), and therefore, a caveat must be given to embedding URLs in a tweet. Finally, we suggest to other researchers planning to conduct studies on the use of Twitter for disaster communication to consider entropy in their empirical models. In so doing, we could have a more detailed understanding about Twitter features in association with information dissemination. Furthermore, we could enhance our knowledge-base that allows scholars and practitioners to effectively leverage Twitter to disseminate disaster information to others who desperately seek critical, highly effective information under great urgency.

Limitations and Future Research

This research suggests a new independent variable, entropy, and successfully provides empirical evidence for this variable. However, we are also aware of limitations of the study. First, during disasters, whether
How Do Multiple Topics in Terse Tweets Affect Retweeting?

tweets include situational information is an important factor affecting retweeting (Itakura and Sonehara 2013; Sutton et al. 2015b). Hence, for future research, we will take into consideration the contents of tweets (i.e., inclusion of situational and sentimental information) along with entropy to provide further insights into how Twitter is used for disaster communication. Second, the same entropy value can be produced by different numbers of topics. We will further investigate the relationship between the number of topics and each topic’s proportion in terms of entropy. Third, statistical models with more independent variables, such as whether tweets mention other twitterers, would be required to more precisely examine the effect of entropy. Fourth, our results were obtained from a dataset of tweets posted during the 2013 Colorado floods, thus limiting the generalizability of our findings. Therefore, empirical evidence from more datasets would be needed to strengthen the generalizability of the current results. Lastly, we plan to further investigate the role of entropy in different non-disaster contexts to have a better understanding of the relationship between entropy and information dissemination.

References


University of Illinois Press IL.


Twenty-third Americas Conference on Information Systems, Boston, 2017