WHAT HAVE WE MISSED WHEN EXAMINING TWITTER AS A COMMUNICATION MEDIUM DURING DISASTERS?

by

Jae Bong Son

B.B.A., Konkuk University, 2003

M.S., Korea University, 2006

M.S., University of Arizona, 2010

A thesis submitted to the

Faculty of the Leeds School of Business of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Operations and Information Management

2017

This dissertation entitled:

What Have We Missed When Examining Twitter as a Communication Medium during Disasters?

written by Jae Bong Son

has been approved for the Leeds School of Business

____

Jintae Lee

____

Laura Kornish

Date__________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Son, Jae Bong (Ph.D., Leeds School of Business)

What Have We Missed When Examining Twitter as a Communication Medium during Disasters?

Thesis directed by Associate Professors Jintae Lee and Laura Kornish

With the advancement of social networking and mobile technology, social media enables people to communicate with previously unreachable people at an unprecedented speed. Particularly, its importance becomes more salient during times of disaster in which localized and up-to-date information about unexpected, dynamically changing life-threatening events is highly necessary. From this perspective, Twitter has attracted the public at risk and online citizens who purposely relay information of local relevance at a faster rate than traditional media as it provides immediate, near-real time access to unique information. The follower-followee network, and short tweets up to 140 characters, are two major mechanisms that allow the public to rapidly exchange time-sensitive information at a large scale about disaster events.

However, traditional emergency messages have been longer, averaging 1,380 characters, and such a message length might be well aligned with the following criteria of disaster-related information: accurate, precise, specific, and clear. In fact, most of the previous research has ignored how short tweets would affect communication practices on Twitter during disaster events. Additionally, recent studies have kept arguing that more research should pay attention to possible influences of such short messages for disaster communication. Understanding how people interpret a brief tweet during disasters is important, as a short tweet may not always convey accurate, precise, specific, and clear messages. In my dissertation, “What Have We Missed When Examining Twitter as a Communication Medium during Disasters,” I closely investigate the shortness of tweets in association with retweeting as the short length of tweets can be viewed as a double-edged sword during times of disaster: one aspect allows for fast updates and the circulation of critical information among twitterers; on the other hand, a tweet may not convey all pertinent information about a disaster event.

DEDICATION

To my mother, father, and brother,

Seongja Yoo, Junho Son, and Jaejin Son

Who have made countless sacrifice for me and provided me with steady guidance and encouragement.

To my mother-in-law and father-in-law,

Yanggeun Ja and Kipyo Hong,

Who have given unconditional support, belief, and love.

To beloved my wife and son,

Minjeong Hong and Nathan Son,

Who are the reason of my life.

ACKNOWEDGEMENT

The completion of this dissertation would not have been possible without the support of the committee chair, Jintae Lee, and committee members, Kai R. Larsen, Laura Kornish, Karina Hauser, and Onook Oh. Their guidance and support have been invaluable in shaping and improving this dissertation. I would like to express my sincere appreciation to them. I am blessed to have Jintae as my advisor. He has trusted me throughout the whole process of my dissertation. Kai always welcomed me and made himself available, whenever I stopped by his office. Particularly, his insightful comments truly enriched my dissertation a lot. Like Kai, Onook's immense knowledge of research improved my dissertation. Thank you for your encouragement and advice!

My sincere thanks go to Hyungkoo Lee (HEC Montreal), Sung Jin (Metropolitan State University), and Jiyoung Woo (Soonchunhyang University) who have provided valuable comments, constructive criticism, and fresh perspectives, which contributed to enhancing my dissertation in many ways. I will be always grateful to you!

I would like to convey my heartfelt thanks to Tony Tong. He is warm-hearted and always encourages me in pursuit of accomplishing my dissertation. My next research must be with you, Tony. Thank you so much!

I would be always grateful to other doctoral students who have been excellent colleagues and friends. They are Yuchen Zhang (Tulane University), Kun Zhang, Aakash Saxena, James Endicott, and Unseok Cho. I feel blessed to meet you all in my Ph.D. program. Thank you for your encouragement and fellowship.

I am deeply indebted to Yongmoo Suh (Korea University), who advised my master thesis and greatly inspired me to pursue my career as a scholar. He introduced me to the field of data analytics 14 years ago! I have been always encouraged by his unwavering support for and trust in me. I will be always grateful to you.

Special thanks also go to the colleagues of mine at IBM Global Business Services. They are Sangho Lee, Yunjung Chang, Yunho Cheong, Kijeong Jeon, Jaeyong Lee, Suyeon Kim, and Jaeeun Chang. I am proud of becoming a member of such a great team and meet very brilliant, collaborative, and constructive professionals in consulting and data analytics. Thanks so much for all your help and encouragement! Particularly, I would like to convey honest thanks to Kijeong and Jaeyong. I am blessed to have both of you and to share a great experience. Thank you so much!

Last but not the least, I am sincerely grateful to my beloved wife Minjeong and son Nathan. Minjeong has made so much sacrifice so that I could pursue my career that I am really passionate about. I would not have completed my dissertation and accomplished my career goal without her unwavering love and sacrifices. I also thank Nathan, who always smiles at me. His great energy has positively affected me in so many ways. Minjeong and Nathan deserve credit for my achievement as much as I do. I love so much!

CONTENTS

TABLES

FIGURES

Chapter 1. Essay 1: Topic Diversity of Tweets and Its Effect on Retweeting during Disasters

Abstract

Information dearth is a central issue during disasters, such as floods, earthquakes, and hurricanes, when the rapid dissemination of up-to-date, localized information is considered critical. Twitter, a social media platform for sharing short messages of 140 characters or fewer, has emerged as a critical tool in disseminating timely information to a large audience at great speed through tweeting and retweeting. However, the character limit may not allow a tweet to clearly express its main message, especially when Twitter users (or twitterers) pack several topics into a single tweet. We theorize that message clarity in tweets is related to seeking additional information, which in turn affects the dissemination of information. That is, when twitterers receive tweets with insufficient message clarity, they may find additional information to have a better understanding of these tweets before retweeting, resulting in fewer retweets overall. In this study, we introduce a way to quantify message clarity in a tweet based on each tweet's number of topics and the entropy measure. Using tweets collected during the 2011 Queensland and the 2013 Colorado floods, we examine how a tweet's message clarity influences its retweet frequency, which hence serves as an indicator for information dissemination. Our findings confirm that a decrease in message clarity lowers retweet frequency, and this relationship is moderated by supplemental information, such as Twitter URLs. By enhancing the understanding of the relationship between message clarity and its influence on information dissemination, our study contributes to IS research on the role of Twitter, and of social media generally, in emergency communication.

Keywords: *Terse Tweets, Message Clarity, Topic Modeling, Entropy, Information Dissemination*

1.1 Introduction

Characterized by a series of uncertain, urgent, and non-routine events (Sellnow and Seeger 2013), disasters are inherently associated with lack of information among citizens in the stricken areas (Mitroff 2004), which leads in turn to high levels of situational uncertainty. Having experienced loss of control and stress (Spence et al. 2007), citizens tend to be motivated to seek disaster-related information in order to be aware of what they are facing and what kinds of disastrous events may yet be impending (Boyle et al. 2004; Procopio and Procopio 2007). Of course, mainstream media play an important role in resolving such situational uncertainty. However, they do not provide specific, timely information that helps local inhabitants at risk (Oh et al. 2013). Unlike mainstream media, social media play a critical role for local citizenry by providing direct observations of the situation at hand that could increase the awareness of impending events, promote precautionary behavior (Keller et al. 2006), and support crisis-event avoidance (Latonero and Shklovski 2011a). Twitter, relative to other social media channels, has received great attention from twitterers, emergency practitioners, communication researchers due to its ability to quickly broadcast critical information to the public (e.g., Chatfield and Brajawidagda 2013; Hughes et al. 2014).

Recognizing the aforementioned role of Twitter, prior research has examined the use of Twitter during disasters by attempting to identify the mechanisms of information dissemination. Twitter allows its users (twitterers) to share short messages of up to 140 characters in length, which others can share through retweeting. Additionally, Twitter provides an effective mechanism for twitterers to conveniently share and obtain information by subscribing to other twitterers' tweets and following them. Correspondingly, Twitter has become one of the most prominent communication channels for easily and rapidly disseminating critical information during disaster situations (Fraustino et al. 2012b). Twitter also has a drawback, however, as a result of the character limit of tweets. That is, the short length could restrict the amount of information for describing disaster events, possibly lowering tweets' message clarity. Message clarity may become even worse when twitterers try to pack several stories about disaster events into a single tweet. This tendency is due mostly to information insufficiency per story. In other words, topical diversity or the number of topics within a 140-character tweet may be negatively associated with message clarity in the sense that as a tweet conveys multiple topics, the amount of information per topic inevitably decreases. Information insufficiency per topic in turn lowers a tweet's message clarity as a whole. During disasters where accurate, precise, specific, and clearly stated information is required (Mileti and Sorensen 1990), the less message clarity a tweet represents, the higher confusion its recipients encounter. Accordingly, the reduced dissemination of such a tweet could prevent life-saving information from reaching recipients. Although the message clarity of a tweet should be taken into account especially for disaster-related research on Twitter, little is known about how such message clarity influences tweet propagation. In this study, we extract tweets' topics, quantify individual tweets' message clarity in terms of topical diversity, and examine the relationship between the message clarity and the dissemination of tweets. The research question we address is:

RQ. How does a tweet's message clarity influence its dissemination during times of disaster?

In the context of the use of Twitter for disaster communication, the primary contribution of our study is that as an exploratory work, we provide in-depth empirical evidence that shows how message clarity in terms of information sufficiency (or insufficiency) affects information propagation in disasters. The paper is structured as follows: first, we review distinctive features of Twitter; second, we introduce how we operationalize message clarity in a tweet based upon a topic modeling technique which groups documents based upon the similarity of topics (Blei 2012) and Shannon and Weaver's entropy measure (Shannon 1949); third, we develop a set of hypotheses to investigate the research question; fourth, we provide descriptions of data, statistical analyses, and empirical results; finally, we discuss the research findings, limitations, and implications for future research.

1.2 Literature Review

1.2.1 140-Character Limit of Tweets

"*just setting up my twttr*" was Twitter's first tweet, a simple message of 24 characters by co-creator Jack Dorsey in 2006 (Siese 2016). It signaled a new era of brevity in electronic media. Such brevity shapes the way the public communicates during times of disaster: first, the use of Twitter has been imperative as an information source to exchange alerts, warnings, and hazard messages (Sutton et al. 2014a) and to connect with other online citizens (Lachlan et al. 2014); second, short texts can be broadcast over virtually all communication platforms including the Web, mobile devices, and even cellular phones (Starbird and Palen 2010; Vieweg et al. 2010); and third, such brevity or terseness allows tweets to propagate at great speed and be available to a wide audience (Sutton et al. 2015b). Sutton et al. (2014a) found that at critical moments of the 2013 Boston Marathon bombing, terse tweets were effective in disseminating and amplifying messages aimed at warning about imminent threats as well as providing guidance for minimizing further damage. These practical features are a primary interest of the U.S. Federal Emergency Management Agency (FEMA). In 2011, FEMA authorized emergency management officials to leverage Wireless Emergency Alerts (WEA)—each of which is limited to 90 characters (Bean et al. 2016)—to alert the public about extreme weather, tsunami, tornado, flash flood warnings, and so on (FEMA 2015). After two years, based upon its message platform, Twitter launched an alerting system for official organizations in the U.S., Japan, and Korea to effectively propagate viable and accurate information about emergencies to online citizens (Protalinski 2013). In fact, text-based, terse messages are gaining momentum for disaster communication. Until recently, however, the guidance of crafting messages about emergencies has centered around longer messages of 1,380 characters (Sutton et al. 2015b). Undoubtedly, little is known about how citizens interpret pithy, brief tweets about non-routine, highly dynamic events and then take actions accordingly.

1.2.2 Topical Diversity and Message Clarity

When considering the length difference between the general emergency messages and tweets, we contend that disaster-related tweets' retweets could be affected by varying degrees of their message clarity and that such message clarity would be associated with the number of topics written into the tweet. That is, the greater number of topics a tweet bears, the less message clarity it may represent. In other words, as twitterers craft tweets with multiple topics, the amount of information per topic of each tweet inevitably decreases, and this negatively affects its message clarity as a whole. For example, when one tweet summarizes three topics and another tweet describes only one topic, the former is considered to have less information per topic than the latter. We argue that the three topics conveyed in the former tweet will be less clear than the one topic in the latter, primarily because the three topics have to be explicated within a range of 140 characters. For the latter tweet, 140 characters are available solely for its one topic. All in all, a single tweet aimed at carrying more topics may inevitably convey fewer specific details per topic (Bruns et al. 2012) and provide less sufficient information for the main topic (Mileti and Peek 2000) than another tweet conveying fewer topics. Because of tweet's brevity, the public may expect a single tweet not to hold diverse topics.

Shannon and Weaver already studied the clarity of a message, which he defined as noise—"a measure of one's freedom of choice in selecting a message"—and contended that if noise is present in a message, this message is assumed to contain some degree of distortions and errors, thus increasing the uncertainty of the message (Shannon 1949). They proposed the following equation to quantify noise in a message, where p_i is the proportion of the *i*th topic out of n topics of a message m.

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

According to Shannon and Weaver's information theory, a single message with two topics is harder to interpret than a message with only one topic. If two topics arise with equal proportion in a message, the recipients can interpret the message as being about either topic, meaning that it is noisy or unclear (entropy of greater than 0). Another possibility would be that if one topic exists with high proportion of almost 1, the proportion of the other topic should be 0.[1] Then, the entropy of the message becomes 0, which indicates that there is almost no chance of the message being interpreted as being about any second topic. This is to say that the message scarcely contains noise, and thus its clarity is highly assured. To take concrete instances, we consider the following three actual tweets about the warnings and alerts of flooding in the Boulder areas.

*Tweet 1: "If flooding is occurring or is expected, get to higher ground quickly. Remember *Turn Around, Don't Drown* #Boulderflood #coflood"*

Tweet 2: "I'm heading to <u>bed</u> now. Everyone in #boulder please be <u>safe</u>. Get to higher ground if possible. #boulderflood"

Tweet 3: "Man it's #<u>biblical</u> #cofloods #boulderfloods #southplatte lets <u>hope</u> the people made it to higher ground! http://t.co/5tRlSPOJaP"

Although all three tweets contained the message about an urgent situation that recommended the affected public seek higher ground for ensuring safety, Tweet 1 presented the urgent situation the most

[1] p_1 is the proportion of the first topic, and thus $1-p_1$ is for the second topic.

clearly of the three. According to the results of topic modeling, it turned out that Tweet 1 depicted urgency, with the topic proportion of 98.9% (*Urgency*—i.e., higher ground, see Table 1.1 for a list of topics), and therefore its entropy was 0.02 (very close to 0). However, Tweets 2 and 3 showed mixed topics. In Tweet 2, 65.53% of the message was about the urgent situation (*Urgency*) and 32.91% about relief (*Relief*—i.e., please be safe), and as such its entropy was 0.64, which is much higher than that of Tweet 1. Tweet 3 was packed with 3 topics: 49.03% about the urgency (*Urgency*), 37.46% about the relief (*Relief*—i.e., hope), and 12.36% about the current floods (*Current Flood*—i.e., #biblical). As expected, the entropy was highest for Tweet 3 at 0.9756.

[1] Table 1.1 Three Topics Corresponding to the Three Tweets[2]

Topic Labels	Keywords per Topic by Importance
Urgency (#46)	canyon, boulder, water, ground, higher, **higher ground**, wall, coming, boulder canyon, creek, **immediately**, **move**, boulder creek, gulch, emerson gulch, emerson, seek, debris, pearl, vehicles
Relief (#50)	**safe**, boulder, stay, rain, friends, prayers, thoughts, people, **hope**, affected, home, good, dry, family, love, raining, bad, crazy, victims, house, news, #longmontflood, praying, water, god, work, #prayforcolorado, live, staying, morning, coming, stop, heart, pray
Floods and Damage (#28)	damage, photos, aerial, images, flood damage, video, **biblical**, climate, line, trends, boulder, climate trends, views, **biblical** flood, show, waters, aerial views, lyons, shot, flood waters

We speculate that Tweet 3 could confuse the recipients more than Tweets 1 and 2 in the sense that Tweet 3 provides less information about the urgency than Tweet 1 and 2, but it presents the other topics –*Relief* and *Floods and Damage*. In other words, twitterers who received Tweets 1 and 3 will struggle more to understand the intent of the latter than that of former as Tweet 3 gives its recipients less information about its main topic – *Urgency*. Consequently, the twitterers will evaluate each tweet as a whole, finding that the message clarity of Tweet 1 is much higher than that of Tweet 3 in terms of the

[2] We produced topics by leveraging a Latent Dirichlet Allocation (LDA) method. Then, we showed three topics that are related to these example tweets. We labeled each topic based upon the interpretation on its list of keywords. Details are explained in the Data and Methods section.

emergency alerts. All in all, a single tweet trying to carry more topics may inevitably maintain lower consistency among topics in its message (Mileti and Peek 2000, p. 187), convey less specific (Bruns et al. 2012, p. 44), and provide less sufficient information for the main topic (Mileti and Peek 2000, p. 188) than another tweet conveying fewer topics. Having agreed on the function of the entropy measure, we believe that a tweet's message clarity could be measured in terms of its entropy or its topic quantity. In sum, we claim that because of their short length, individual tweets' message clarity can be quantifiable in terms of each tweet's topic quantity; additionally, a tweet's message clarity is negatively related to its topic quantity, and as a tweet's message clarity decreases, its credibility decreases as well.

1.2.3 Retweeting as Message Amplification

Communication is a purpose-driven process (Shannon 1949; Stephens and Barrett 2014). Emergency management officials and citizen journalists have an essential communicative goal of spreading disaster-related messages to as many people as possible in target areas within a short time span (Sutton et al. 2015b). As messages are disseminated widely, the number of people exposed to the messages increases as well (Sutton et al. 2014a). Accordingly, the levels of situational awareness among the intended target population expand and thereby increase individuals' own protective actions for safety. In this context, message amplification or retransmission is an efficient way to reach a wide target population. Twitter provides an effective network for message amplification by retransmission—retweeting. Retweeting is an act of re-posting an original tweet. Twitter's retweet functionality allows sharing of an original tweet with other twitterers (Compston 2014) when its information is considered to be interesting, useful, or imperative for others (Abdullah et al. 2014; Starbird and Palen 2010; Sutton et al. 2014b; Zubiaga et al. 2015). In a similar vein, the retweet function is also considered as a recommendation system (Boyd et al. 2010). That is, as prosumers of Twitter (Boyd et al. 2010), twitterers determine what information has to be emphasized, discussed, and diffused into their communities by shedding light on others' information sources and stories (Bruns 2008), along with generating their own information (Westerman et al. 2014). Retweeting allows twitterers to effectively pass along timely, critical information about approaching threats (Bruns and Stieglitz 2012) to citizens. We believe that one factor influencing the extent to which a tweet is retweeted during times of disaster is its message clarity. That is, as the message clarity of a tweet increases, its retweet frequency will increase as well.

1.3 Hypothesis Development

The affected public becomes "information hungry" as disaster events impend (Mileti and Sorensen 1990, p. 3.8). They immediately start to engage in seeking out information from sources such as television, terrestrial radio, newspapers, and social media. Among other information sources, Twitter is considered to have the potential for disseminating up-to-the-minute information (Lachlan et al. 2014) at critical moments of disaster events. Its unprecedented speed (Latonero and Shklovski 2011b; Sutton et al. 2015b) and ability to reach a large number of target audiences (Murthy 2012; Wilensky 2014) makes Twitter an imperative information source for disaster communication. The aforementioned advantages of Twitter are due to its "bite-sized, easily digestible doses of information" (Stephens and Barrett 2014). In this sense, Twitter's impact is due partly to the 140-character limit of tweets (Latonero and Shklovski 2011b; Murthy 2011; Sutton et al. 2015b; Zubiaga et al. 2015). While a tweet can effectively raise public awareness about impending threats in a timely manner, its terseness may make it an insufficient means for providing the detailed, accurate, and specific information that is required to adequately describe disaster events (Bean et al. 2016; Bean et al. 2015; Lachlan et al. 2014). In addition, the length limit could lower the information quality conveyed by tweets by encouraging twitterers to use shortened words, jargon, and even incorrect grammar (Cotelo et al. 2014; Murthy 2011) to compensate for such shortness. The problem may be severe as twitterers try to pack multiple topics into a 140-character length tweet. With multiple topics in a short message proposedly contributing to a lack of information per topic, and thus, to massage clarity of the tweet, twitterers become eager for additional and confirming information (Lindell and Perry 1987) to relieve the confusion and suspicion raised from insufficient message clarity (Bean et al. 2016; Bean et al. 2015).

Information-seeking behavior as an effort to obtain information has been termed sense-making (Sutton et al. 2014b) or milling (Mileti and Sorensen 1990). Chaiken and Eagly (1989) argued that a person desires accurate and adequate information (Griffin et al. 2002). For example, on recognizing the received tweets' insufficient message clarity, the recipients may wait for other related tweets to be received, explore twitterverse by hashtags or keywords, interact with others for clarification, or turn to more authoritative information sources (Spiro et al. 2012). In other words, twitterers who receive tweets with insufficient message clarity may spontaneously seek additional information in order to overcome such obscurity and to verify whether they properly understand the intended meaning of the tweets (Fraustino et al. 2012b), or they may misunderstand the intended message, negatively affecting their retweetability. Once they are convinced by the additional information that helps them to correctly grasp the main message of the tweets, they tend to promptly share the tweets with others. Without the correct understanding of received tweets, recipients may delay or abandon retweeting. Consequently, we hypothesize that as the message clarity of a tweet decreases, its retweet frequency decreases. Specifically, we contend that the relationship would be a curvilinear—the effect of message clarity in a tweet on retweet frequency is not constant, but conditional. Therefore, the hypotheses we are interested in are as follows:

H1a. The decrease in a tweet's message clarity negatively affects its retweet frequency during times of disaster.

H1b. The decrease in a tweet's message clarity negatively affects its retweet frequency, and this negative relationship decreases as message clarity in a tweet decreases.

Since we theorize that the issue of message clarity stems from information insufficiency (or sufficiency), we must examine whether additional information moderates the relationship between a tweet's message clarity and its retweet frequency. In the twitterverse, twitterers are able to obtain additional information based on hashtag or keyword search (Spiro et al. 2013), or from Twitter URLs linked to diverse information from external media (Bruns and Stieglitz 2012; Hughes and Palen 2009). In particular, Twitter URLs are an interesting convention for twitterers to overcome in the 140-character limitation of tweets (Hughes and Palen 2009; Purohit et al. 2013; Spiro et al. 2013) when a large amount of information must be disseminated (Lachlan et al. 2014). In fact, twitterers like to include Twitter URLs to make their tweets informative (Bruns and Stieglitz 2012; Ma et al. 2013). Hughes and Palen (2009) reported that roughly 50% of the tweets about a hurricane event included URLs, 10% higher than that of tweets about general events. They argued that embedding URLs into tweets becomes a key feature for disseminating rich information during disasters. Diverse information ranging from news articles and web sites, to multimedia sources such as video and audio (Kostkova et al. 2014), can be delivered through URLs. Because short tweets are inevitably related to an information dearth, Twitter URLs are the most fundamental and effective way to share in-depth information. In a similar vein, Suh et al. (2010b) empirically demonstrated the positive relationship between Twitter URLs and tweets' retweetability. Overall, studies that found a positive influence of Twitter URLs mainly argue that Twitter URLs are beneficial for conveying a large amount of interesting information. On the contrary, other studies revealed a negative association, claiming that URLs reduce the remaining space for twitterers to include other information that could be of more interest (Burnap et al. 2014), therefore impeding timely communication among twitterers in disaster situations. It is undoubtedly true during disasters that including as many Twitter URLs as possible in tweets provides an easy way to address the problem of information dearth, positively affecting the retweet frequency. Therefore, we believe that Twitter URLs can compensate for insufficient message clarity. The following hypothesis represents our interest:

H2. The decrease in a tweet's message clarity negatively affects its retweet frequency, and this negative relationship becomes weaker as the number of Twitter URLs increases.

As a means to express emotions and compensate for tweets' short length limit, twitterers use emoticons, which are constructed by typographical symbols and are used to convey feelings such as happy, sad, pleased, or agreeable (Rezabek and Cochenour 1998). Walther and D'Addario (2001) defined emoticons as pictorial representations that replicate facial expressions such as smiles :), winks ;), and frowns :(. Relative to face-to-face communication, Computer-Mediated Communication (CMC) lacks visual and nonverbal cues. Without such cues, communication could be seen as less social, less emotive, and less interpersonal (Rice and Love 1987). Along with the growing importance of mobile communication and the Internet, emoticons have become an important communication tool of CMC especially for text-based electronic media. Thus far, diverse research regarding emoticons has been conducted ranging from sentiment analysis (Agarwal et al. 2011; Gimpel et al. 2011; Pak and Paroubek 2010), to cross-cultural differences in the usage of nonverbal cues (Park et al. 2014; Park et al. 2013), to the general purposes of using Twitter (Westman and Freund 2010). It seems clear that CMC users apply emoticons to supplement, reiterate, or clarify the meaning of their texts as a means to assist conversation. At the same time, they interpret embedded emoticons in order to grasp the attitude and the subtle sentiment of other CMC users and to sense the message tones that textual elements alone do not provide (Lo 2008) (see Table 1.2). Rezabek and Cochenour (1998) emphasized the use of nonverbal cues for effective communication by stating, "Effective communication is not simply a matter of analyzing individual word denotations and connotations, it is a blend of many factors. Words, grammar and structure, context and experience, nonverbal signals, and other cues all contribute meaning in a message" (p. 202).

[2] Table 1.2 Top 10 Emoticons per Flood Incident

	2011 Queensland Floods			**2013 Colorado Floods**		
Ranking	Emoticon	Count	Meaning	Emoticon	Count	Meaning
1	:(	1279	A sad face	:(	384	A sad face
2	:)	1035	A smiling face	:)	164	A smiling face
3	<3	610	Love or heart	<3	97	Love or heart
4	:'(	152	A crying face	:/	76	"*This sucks*"
5	;)	131	Wink	:-(	42	Expression of sadness
6	:-(	117	Expression of sadness	;)	34	Wink
7	:/	114	"*This sucks*"	:-)	24	Same as :)
8	:-)	113	Same as :)	:D	15	A very happy smiley face
9	:D	96	A very happy smiley face	:'(	15	A crying face
10	:o	69	Surprised	:o	12	Surprised

However, we need to re-consider the usage of emoticons in terms of information value that they can supplement over and above verbal cues. For example, disaster situations are normally characterized by a lack of information, and the public in disaster stricken areas desperately seeks up-to-date, valid, and credible information in order to be properly aware of their surroundings. In this context, emoticons should be understood quite differently from Twitter URLs. That is, while URLs provide detailed, in-depth information over and above verbal cues of words and hashtags, emoticons deliver little information. Therefore, twitterers who received tweets with insufficient message clarity might judge that the emoticons embedded in the received tweets cannot address the problem of message clarity. As a result, we presume that during times of disaster, emoticons could be a factor that confuses twitterers who need to acquire additional information over and above verbal information. Thus, the last hypothesis is tendered with respect to the information value of emoticons in disasters.

H3. The decrease in a tweet's message clarity negatively affects its retweet frequency, and such a negative relationship is stronger as the number of emoticons increases.

1.4 Data and Methods

1.4.1 Two Flood-related Natural Disasters: 2013 Colorado and 2011 Queensland

While pouring 15 to 20 inches of rain on the Front Range area including Boulder, Colorado Springs, and Fort Collins, the 2013 Colorado floods caused a great deal of heartache and economic difficulty for Coloradans. A series of floods started on September 9, 2013 and lasted for seven days. Boulder County was hit the hardest with five days of rainfall exceeding its annual average of 20.7 inches. Fourteen counties in Colorado declared disaster emergencies with more than 11,000 residents evacuated. 1,750 residents, along with animals and 300 pets, were rescued by the U.S. Army and the Colorado National Guard (Connor et al. 2013). Immediately following the initial warnings by FEMA and the National Weather Services, people in several affected and remote areas started producing, sharing, and disseminating diverse flood-related information on Twitter.

[3] Table 1.3 Keywords and Hashtags used for Retrieving Tweets

2013 Colorado floods (Dashti et al. 2014b)		
Date	**Keywords**	**Hashtags**
September 11	boulderflood, cowx, nwsboulder	
September 12	coflood, cofloods, coflodding, cuboulder flood	#boulder, #cccf
September 15	Boulderfloods	
September 19	flood gas, flood infrastructure	#cofloodrelief
September 20		#coloradostrong
2011 Queensland floods		

(N/A)	Queensland, qldfloods, qldflood	#qldfloods, #thebigwet

During the 2013 Colorado floods, Project EPIC, hosted by the Department of Computer Science at the University of Colorado Boulder, collected tweets and their retweets about the flood events in near-real time. The research group was able to systematically retrieve relevant tweets and twitterers by incrementally adding keywords, hashtags, and twitterers (Dashti et al. 2014a). As a result, 102,426 original tweets and 122,276 retweets produced by 77,774 unique twitterers were collected between September 12 and September 25, 2013.

The research group also examined another Twitter flooding dataset. From December 2010 thru January 2011, with especially severe episodes between January 10 to 16, a series of floods hit much of the central and the southern parts of Australia, including Queensland (Shaw et al. 2013). The floods affected more than 200,000 residents living in 90 towns, caused A$2.38 billion of damage, and resulted in 38 casualties (Hanson March 08, 2012). Twitter was a crucial outlet for dissemination of emergency information as its users propagated the information and expanded its reach (Bruns and Stieglitz 2012). Twitter's GNIP[3] subsidiary provided the research group with data about the Queensland floods. Their data scientists followed Project EPIC's procedures in order to retrieve tweets, retweets, and twitterers' information on the 2011 Queensland floods. Between January 8, and January 21, 2011, 109,456 original tweets and 120,082 retweets produced by 33,565 unique twitterers were collected. Table 1.4 summarizes the descriptive statistics of the two Twitter datasets.

[4] Table 1.4 Descriptive Statistics of Two Flood Cases

Items \ Cases	2011 Queensland	2013 Colorado
Period of Data Collection	January 8^{th} ~ 21^{st}, 2011	September 12^{th} ~ 25^{th}, 2013
Total Tweets	109,456	102,426
Total Retweets	120,082	122,276
Unique Twitterers	33,565	77,774

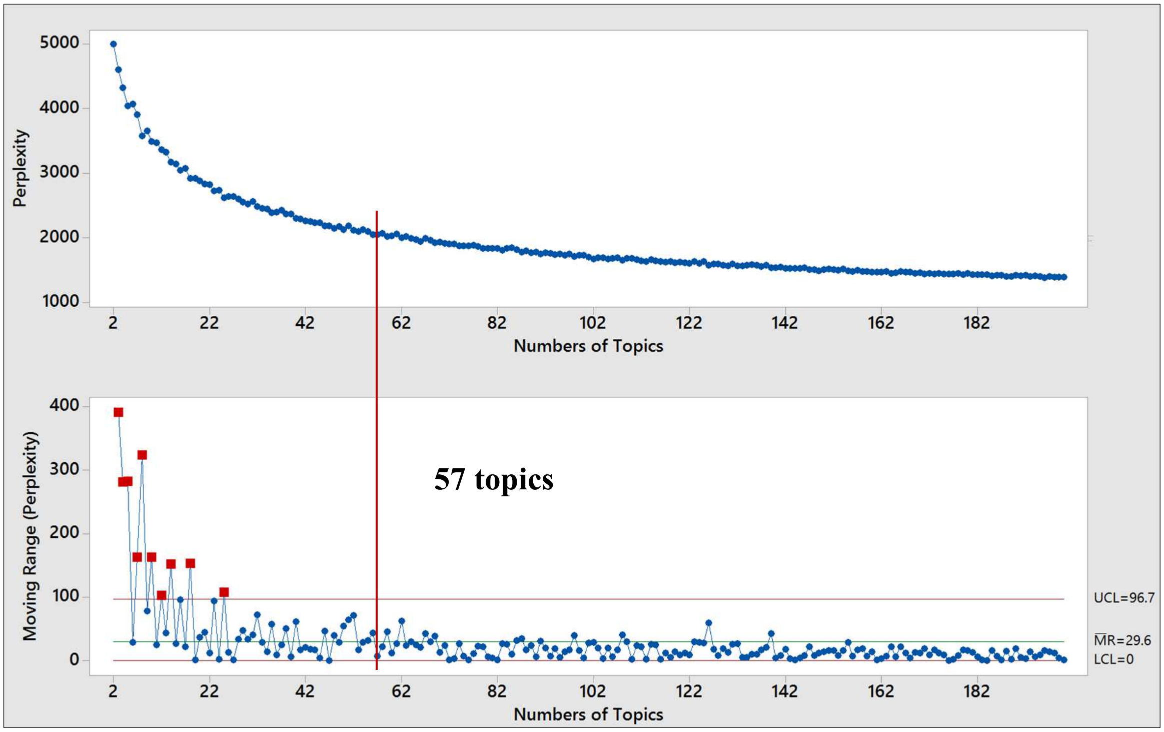
[3] GNIP (https://gnip.com) is a Twitter's subsidiary that provides an enterprise API platform.

1.5 Methods

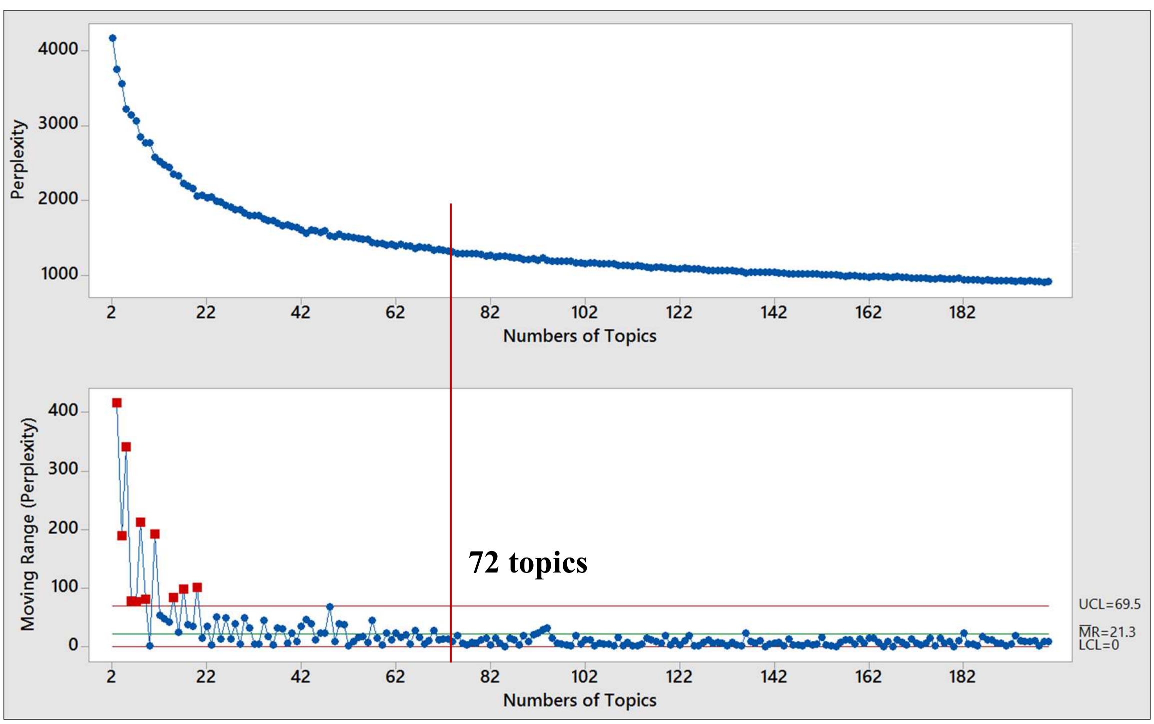
Along with statistical procedures, a series of analytical techniques were employed to fulfil the research goals of the study. We utilized natural language processing (NLP) techniques to analyze each tweet's unstructured message into its structured forms, which include words and their part-of-speech (POS) tags, URLs, hashtags, and mentions. Since the length of tweets is too short for modeling topics (Cataldi and Aufaure 2015; Wang et al. 2007), we extracted up to 6-gram noun phrases based on POS tags and used them as additional input to extract topics in tweets. That is, together with uni-gram words, bi- and tri-gram words—such as "flood victims," "colorado flood," and "higher ground," as well as "flood relief appeal—were used to extract topics. To achieve this analysis, the following steps were leveraged: first, we tagged tweets' message components by leveraging *TweetNLP*'s programming library (Owoputi et al. 2013); second, we extracted tweets' topics by utilizing the Machine Learning for LanguagE Toolkit (MALLET), a Java implementation of the Latent Dirichlet Allocation (LDA)-based topic modeling (McCallum 2002), to discover topics in tweets. As inputs for topic modeling, we included *n*-gram noun phrases and hashtags, which are an essential component to annotate individual tweets' conversation topics (Boyd et al. 2010; Bruns and Stieglitz 2012; Laniado and Mika 2010; Ma et al. 2013; Yang et al. 2012). However, we excluded embedded URLs comprised of random characters and numbers from the topic analysis (e.g., http://yfrog.com/hsi9sfj), because such URLs are devoid of the topic information needed to find topics. Topic modeling is considered a clustering method in the sense that documents are grouped together based upon the similarity of topics (Blei 2012). Accordingly, providing an optimal number of topics for the LDA will be critical to have topics that best represent target documents. To accomplish this goal, we generated topic models by increasing the number of topics from 2 to 200, calculated each topic model's goodness of fit, and evaluated the generalizability of each topic model in terms of its perplexity score (Blei et al. 2003), where M refers to the number of documents in the testing dataset, w_d refers to the words in document d, and N_d refers to the number of words in document d.

$$\text{Perplexity}\ (D_{test}) = \exp\left\{\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right\}$$

Each model's generalizability is inversely related to its perplexity score—the lower the score, the higher the generalizability. By sequentially ordering the perplexity scores by topic quantity, we applied the cumulative sum (CUSUM) procedure (Ellaway 1978) to each Twitter dataset to find an optimal topic quantity at which the changes in the perplexity score are negligible, indicating that additional topics would offer no significant benefits to generalizability. Figure 1.1 shows that as the quantity of topics increases, a series of perplexity scores and their moving ranges for the two flood incidents decrease. As a result, 72 and 57 topics were chosen as the optimal topic quantities for the tweets from the 2011 Queensland and the 2013 Colorado floods, repectively. All identified topics are listed in Appendix 1.A.



< 2011 Queensland floods >



< 2013 Colorado floods >

[1] Figure 1.1 The Optimal Topic Number by Perplexity

1.6 Empirical Analysis

1.6.1 Dependent Variable

The unit of analysis for our research is an individual, original tweet, and as such the dependent variable is the retweet frequency within a 24-hour time period. This means, we aggregated each original tweet's total number of retweets posted within 24 hours after its posting into a count variable. It has been reported that most regression models derived from the Gaussian probability distribution function (PDF) produce biased and inconsistent results when dealing with count dependent variables such as retweets (Cameron and Trivedi 2013). Even with transformed count variables, the Gaussian-based regressions performed poorly (O'hara and Kotze 2010). To address these concerns, we estimated our model using the Poisson, or a negative binomial regression, analysis. The negative binomial regression is more appropriate than the Poisson when the variance of a dependent variable is significantly larger than its mean, which is termed over-dispersion and indicates a violation of Poisson's distributional property (Hilbe 2011). As Table 1.5 shows that the variance of the dependent variable is larger than its mean, we tested the dependent variable's over-dispersion by following the procedures recommended by Cameron and Trivedi (2013). We confirmed over-dispersion in our datasets and observed the existence of the heteroscedasticity of variance (Breusch and Pagan 1979). As a result, robust negative binomial regression was employed to evaluate the empirical model.

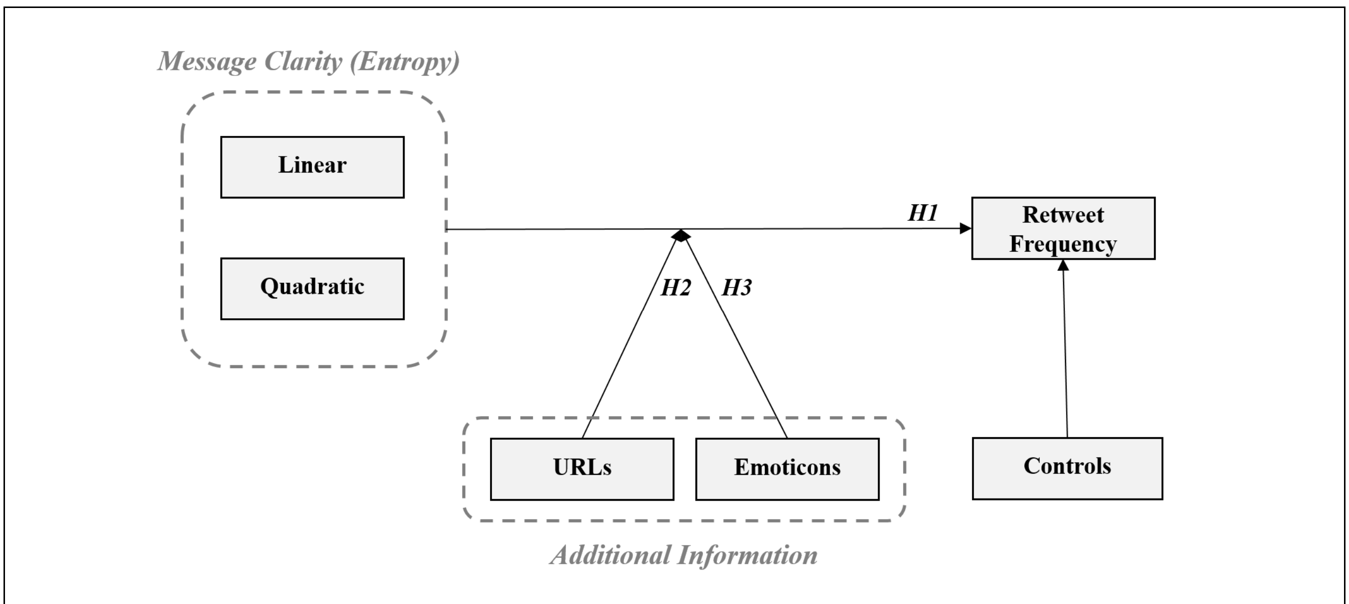
[5] Table 1.5 Variable Description

Variable Name	Explanation / Cases	2011 Queensland			2013 Colorado		
		Mean	S.D.	Range	Mean	S.D.	Range
Dependent Variable							
$Retweets\_24h_i$	Retweet frequency within 24 hours after the posting of tweet *i*	0.745	8.007	0-1684	1.146	6.51	0-741
Message Clarity (Entropy)							
$Entropy_i$	The entropy of tweet *i*	0.306	0.359	0-1.77	0.22	0.323	0-1.6
- Point Estimate of Message Clarity							
$Linear_i$	The linear relationship between the dependent variable and the entropy of tweet *i*						
$Quadratic_i$	The quadratic relationship between the dependent variable and the entropy of tweet *i*						
Additional Information							
$URLs_i$	The number of URLs in tweet *i*	0.462	0.553	0-5	0.667	0.535	0-4
$Emoticons_i$	The number of emoticons in tweet *i*	0.054	0.251	0-9	0.012	0.11	0-3
Message Clarity × URLs or Emoticons							
- Research Model 3							
- URLs							
$Linear_i$ x $URLs_i$	Moderation between *Linear x URLs* to examine information value of URLs over and above other information.						
$Quadratic_i$ x $URLs_i$	Moderation between *Quadratic x URLs* to examine information value of URLs over and above other information.						
- Emoticons							
$Linear_i$ x $Emoticons_i$	Moderation between *Linear x Emoticons* to examine information value of emoticons over and above other information.						
$Quadratic_i$ x $Emoticons_i$	Moderation between *Quadratic x Emoticons* to examine information value of emoticons over and above other information.						
Control Variables							
$Words_i$	The total number of words in tweet *i*	9.54	3.99	0-24	8.61	3.84	0-24
$Hashtags_i$	The total number of hashtags in tweet *i*	1.26	0.891	0-13	1.27	1.21	0-15
$Ln(Followers_{i,t})$	The log-transformed number of followers of tweet *i*'s author between his/her join date and the date of posting tweet *i*	5.465	1.799	0-15.1	6.106	2.306	0-16.4

	$Ln\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s Posting Date}}} Follower_{i,t}\right)$						
Ln($Followees_{i,t}$)	The log-transformed number of followees of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s Posting Date}}} Followee_{i,t}\right)$	5.37	1.61	0-12.1	5.834	1.948	0-12.7
Ln($Likes_{i,t}$)	The log-transformed number of likes of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s Posting Date}}} Like_{i,t}\right)$	1.783	1.945	0-9.32	3.304	2.602	0-13.6
Ln($Status_{i,t}$)	The log-transformed number of tweets of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\underbrace{\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s Posting Date}}} Tweet_{i,t}}_{Status}\right)$	7.443	1.983	0-12.7	8.140	2.234	0-14.0
$Mention\_YN_i$	Whether tweet *i* contains other twitterers' names – 1 for 'Yes' and -1 for 'No'						

1.6.2 Control Variables

Prior research has studied factors that affect the retweet frequency. Most research investigated the content features of tweets (e.g., the length of tweets, hashtags, and URLs) and twitterers' features (e.g., followers, followees, likes, and status). Sutton et al. (2014b, p. 779) performed empirical analyses on disaster-related tweets and found that hashtags, followers, or followees were positively related to the retweet frequency. By analyzing 74 million tweets, Suh et al. (2010b) reported the following findings: first, the number of followers and followees each had a positive relationship with the retweet frequency; second, the volume of past tweets (that is, status) and whether tweets include other twitterers' names (that is, mentions) negatively affected tweets' retweet frequency; lastly, the frequency of twitterers' likes did not have a significant effect on the retweet frequency. Therefore, we included the above mentioned features of Twitter as control variables.



[2] Figure 1.2 Research Model

By retaining the above-mentioned features as controls, we established the research model shown in Figure 1.2 that tests the effect of a tweet's message clarity, in terms of entropy, on its retweet frequency made within a period of 24 hours after its posting. We also included the number of *words* and *hashtags* to control tweets' length for better measuring the message clarity of tweets. For the control variables that were skewed to the right, we performed a log transformation on them for better normality (Judd et al. 2001). We based the log transformation on: 1) how many followers an author of tweet *i* has at the date of posting (*t*) – *Ln(Followers)*; 2) how many followees[4] an author of tweet *i* has at the date of posting (*t*) – *Ln(Followees)*; 3) how many likes an author of tweet *i* has at the date of posting (*t*) – *Ln(Likes)*, 4) how many tweets an author of tweet *i* has posted at the date of posting (*t*) – *Ln(Status)*. Then, the relationship between our dependent variable—the retweet frequency within a 24-hour time period—and the message clarity of tweets are estimated whether each relationship is significant over and above the control variables. Especially, we theorize that the relationship would not be constant as the message clarity decreases. Therefore, we add the 2 predictors of the linear (*Linear*) and the quadratic (*Quadratic*). Table 1.5 summarizes the descriptive statistics of the dependent, control, and predictor variables.

The extent of the dependence between the variables in the research model was investigated, and a corresponding correlation matrix was generated (see Appendix 1.B). Although a few relatively high correlations among control variables were identified, none of their variance inflation factors (VIF) exceeded 3.30, which is far below the acceptable VIF level of 5 (David A. Belsley 2005) (see Appendix 1.C), indicating that the empirical model did not have significant signs of a multicollinearity problem.

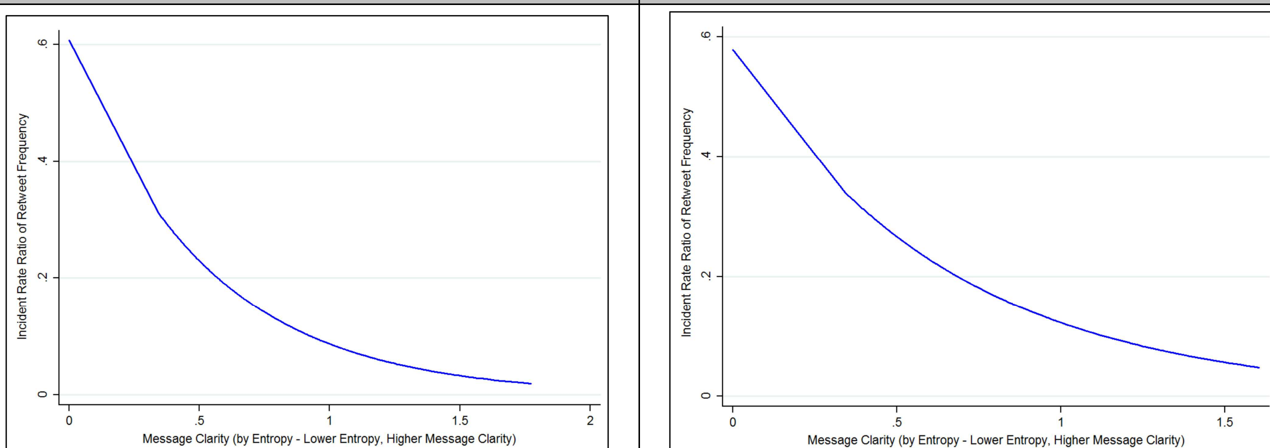
1.7 Results

Table 1.6 shows the result of the negative binomial regression of the dependent variable, *Retweets_24h*, on the independent variables of the two flood incidents. As we hypothesized, we found a significant relationship between the dependent variable and message clarity in both cases (Queensland:

[4] Followees refer to "friends" whom a twitterer follows, while followers are twitterers who follow a twitterer (Suh et al. 2010b).

Wald Chi^2=1574.96, df=2, p<0.000; Colorado: Wald Chi^2=1038.46, df=2, p<0.000), while the other variables were held constant at their means. In other words, as the message clarity decreased, the retweet frequency linearly decreased (Queensland: coefficient=-1.871, Wald Chi^2=1011.58.00, df=1, p=0.004; Colorado: coefficient=-1.287, Wald Chi^2=295.82, df=1, p<0.000). For Hypothesis 1b, we did not specify the effect's direction. However, we found the significant conditional effect of message clarity, which weakened the relationship found in Hypothesis 1a (Queensland: coefficient=-0.120, Wald Chi^2=12.70, df=1, p<0.000; Colorado: coefficient=-0.333, Wald Chi^2=32.02, df=1, p<0.000). In other words, as the relationship departed from linearity, such linearity became weaker as shown in Figure 1.3. Therefore, both Hypothesis 1a and 1b are supported.

2011 Queensland	2013 Colorado



[3] Figure 1.3 Relationship between Message Clarity and Retweet Frequency

We theorize that message clarity per tweet results from information insufficiency (or sufficiency), and Hypothesis 1 confirmed that the decrease in the message clarity of a tweet (or the decrease in information in a tweet) negatively affected its tweet frequency. One way to provide additional evidence for the effect of message clarity on the retweet frequency is to scrutinize the relationship between message clarity and the frequency of retweets when additional information is provided. As such, Hypothesis 2 and 3 were established to assess the conditional effect of message clarity on the retweet frequency by URLs and emoticons. For the hypotheses that include moderation terms, all continuous variables were deviated from their means to aid the interpretation of moderation effects and alleviate multicollinearity (Aiken et al. 1991). The results of our empirical model demonstrated that URLs increased the retweet frequency as tweets' message clarity decreased (Queensland: Wald Chi^2=26.19, df=2, p<0.000; Colorado: Wald Chi^2=43.40, df=2, p<0.000), clearly supporting Hypothesis 2. Interestingly, the effect of URLs on the retweet frequency was significant when the relationship between message clarity and the retweet frequency was linear (Queensland: coefficient=0.452, Wald Chi^2=22.46, df=1, p<0.000; Colorado: coefficient=0.444, Wald Chi^2=16.46, df=1, p<0.000); however, this moderation effect proved to be no longer true when such a linearity weakens (Queensland: coefficient=-0.240, Wald Chi^2=00.93, df=1, p=0.3339; Colorado: coefficient=0.0206, Wald Chi^2=0.01, df=1, p=0.9382). To interpret the moderation results of our negative binomial regression, we constructed Table 1.7, showing different standard errors (SEs) and confidence intervals (CIs) per URL (Hilbe 2011).

[6] Table 1.6 Statistical Results

Retweets_24h (DV) / Variables	2011 Queensland	2013 Colorado
	Coefficient (Robust Err.)	Coefficient (Robust Err.)
Information		
Message Clarity (Entropy)	(Wald Chi^2=1574.96 (2), p<0.000)	(Wald Chi^2=1037.46 (2), p<0.000)
$Linear_i$	-1.871*** (0.0577)	-1.287*** (0.0655)
$Quadratic_i$	-0.120 (0.137)	-0.333* (0.148)
Symbols	(Wald Chi^2=56.22 (2), p<0.000)	(Wald Chi^2=17.47 (2), p=0.0002)
$URLs_i$	0.236*** (0.0355)	0.0629 (0.0324)
$Emoticons_i$	-0.233** (0.0804)	-0.528*** (0.146)
Moderation		
Additional Info. – URLs	(Wald Chi^2=26.19 (2), p<0.000)	(Wald Chi^2=43.40 (2), p<0.000)
$Linear_i$ x $URLs_i$	0.452*** (0.0955)	0.444*** (0.110)

Quadratic$_i$ x URLs$_i$	-0.240 (0.249)	0.0206 (0.266)
Additional Info. – Emoticons	(Wald Chi^2=4.25 (2), p=0.1192)	(Wald Chi^2=0.06 (2), p=0.9689)
Linear$_i$ x Emoticons$_i$	-0.358 (0.354)	-0.109 (0.433)
Quadratic$_i$ x Emoticons$_i$	1.178 (0.619)	0.181 (1.025)
Control Variables	(Wald Chi^2=3276.7 (7), p<0.000)	(Wald Chi^2=10759 (7), p<0.000)
Words$_i$	0.0651*** (0.00403)	0.0577*** (0.00373)
Hashtags$_i$	0.242*** (0.0241)	0.264*** (0.00986)
Ln(Followers$_{i,t}$)	0.584*** (0.0142)	0.718*** (0.00934)
Ln(Followees$_{i,t}$)	-0.118*** (0.0122)	-0.0770*** (0.00785)
Ln(Likes$_{i,t}$)	0.0447* (0.0198)	0.127*** (0.00610)
Ln(Status$_{i,t}$)	-0.181*** (0.0174)	-0.438*** (0.00848)
Mention_YN$_i$	-0.211*** (0.0177)	-0.0260 (0.0146)
Constant	-1.008*** (0.0208)	-0.797*** (0.0209)
Model Summary		
Log-likelihood Ratio	19415.536***	31547.073***
Wald χ^2	6338.45***	12558.31***
McFadden's R^2	0.095	0.131
n	109456	102426

[1] All predictors are mean-centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

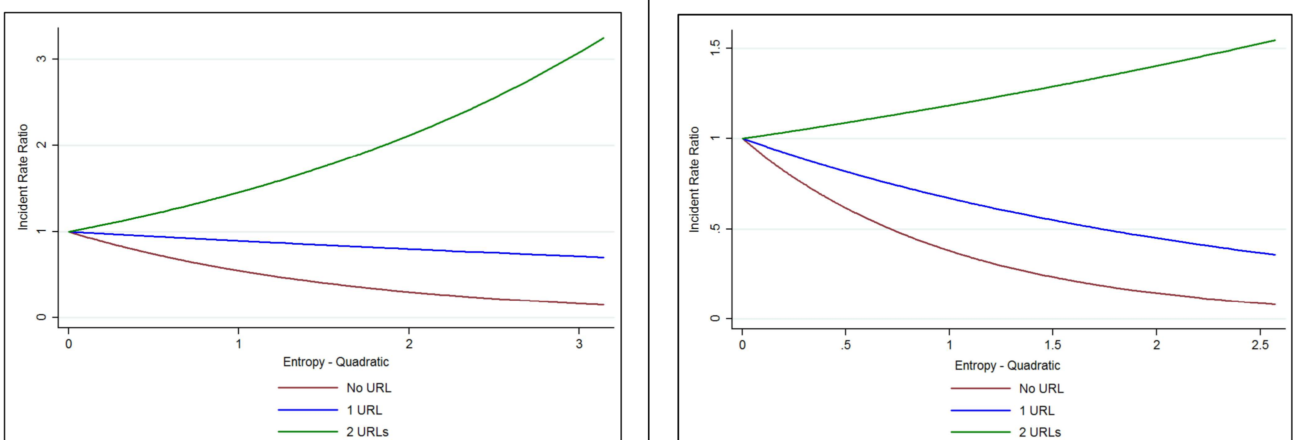
What we found interesting is that only one URL reliably moderated the relationship between message clarity and retweet frequency (Queensland: 95% CI[5]: 0.1169, 0.5009; Colorado: 95% CI: 0.2159, 0.8577) (see Figure 1.5).

[5] Confidence Interval

[7] Table 1.7 The Moderation IRRs of Twitter URLs

Cases / URLs	2011 Queensland			2013 Colorado		
	IRR	S.E.[6]	95% Confidence Intervals (CIs)	IRR	S.E.	95% Confidence Intervals (CIs)
0	0.1539 (-84.51%)	0.1784	(0.1085, 0.2184)	0.2761 (-72.39%)	0.1652	(0.1997, 0.3816)
1	0.2419 (-75.81%)	0.3713	(0.1168, 0.5009)	0.4304 (-56.06%)	0.3518	(0.2159, 0.8577)
2	0.3801 (-62%)	0.4986	(0.1431, 1.0103)	0.6710 (-32.90%)	0.4711	(0.2665, 1.6893)
3 or More	0.5975 (-40.25%)	0.6033	(0.1831, 1.9493)	1.0460 (+4.60%)	0.5673	(0.3440, 3.1800)

2011 Queensland	2013 Colorado



[4] Figure 1.5 Moderation Plots of Twitter URLs

Regarding Hypothesis 3, we did not find any significant moderation effect of emoticons on the relationship between message clarity and retweet frequency in either case. Therefore, Hypothesis 3 was rejected.

1.8 Discussion and Conclusion

[6] The IRR standard errors for the moderations are determined by the following variance equation (Hilbe 2011): $V_{URLs\ x\ Entropy} = V(Entropy) + URLs^2 * V(URLs) + 2 * URLs * Cov(Entropy, URLs)$, where V($a$) means the variance of a, Cov(a, b) represents the covariance of a and b.

During times of disaster, detailed and up-to-date information about disastrous events must reach the affected public in a timely manner (Bean et al. 2015; Mileti and Sorensen 1990; Mitroff 2004). While most Twitter studies have ignored the effect of tweets' limited length on information dissemination for disaster communication, this study proposed a new variable, *message clarity*, to reflect the length limit of tweets. Using two Twitter datasets, we provided statistical evidence that shows the significant relationship between message clarity in tweets and retweet frequency. The statistical results revealed the following two findings. First, as the message clarity of tweets decreases, their retweet frequency steeply decreases in a linear fashion, and then the effects weaken. Second, the inclusion of URLs in a tweet strengthened the aforementioned findings. That is, the additional information delivered by URLs compensated for insufficient message clarity. Importantly, however, only the inclusion of one URL was statistically effective in mitigating the negative effect of tweets' message clarity on the retweet frequency.

It is also necessary to discuss the insignificant moderation effect of emoticons with the relationship between the message clarity and the retweet frequency. In general, emoticons facilitate computer-mediated communication by increasing intimacy among communicators (Rezabek and Cochenour 1998). We assumed that emoticons would hamper the dissemination of tweets during disaster events in which information value is highly essential (Bean et al. 2016; Bean et al. 2015; Lachlan et al. 2014). Although the statistical results supported our fundamental assumption in the sense that the retweet frequency decreased as a function of emoticons, emoticons did not have a reliable moderation effect on the relationship between the message clarity and the retweet frequency. Our interpretation of this rejection is that emoticons confuse with twitterers to be vigilant to surroundings during disasters. Such rejection adds one important piece of evidence to the previous findings about message clarity: for disaster communication, verbal cues that directly describe one's surroundings are more important than non-verbal cues that cannot openly explicate situations. The relationships we investigated were consistent, appearing across two flood incidents. Table 1.8 summarizes the results of hypothesis testing.

[8] Table 1.8 Summary of Hypothesis Testing

Cases / Hypothesis	2011 Queensland	2013 Colorado	Results
H1	Supported	Supported	Fully Supported
H2	Supported	Supported	Fully Supported
H3	Not Supported	Not Supported	Not Supported (But the Same Sign)

This study makes empirical contributions to the literature on the use of Twitter in disaster communication. The short length of tweets is inherently associated with certain degrees of information insufficiency, and such insufficiency becomes severe as tweets convey more topics in general due to the fact that the 140 characters are split between topics. Even though some researchers have expressed concern about the use of short message services (i.e., Short Message Service [SMS], Wireless Emergency Alerts [WEA], and Twitter) for conveying disaster messages (Bean et al. 2016; Bean et al. 2015; Sutton et al. 2015a), to date, no empirical research on the use of Twitter during disasters has taken message clarity into account. Further, the topic of disaster message clarity has gained in prominence because of FEMA's 2011 decision authorizing emergency management officials to use WEA, with alerts length limited to 90 characters or less (Bean et al. 2016).

The study offers several implications for twitterers and emergency practitioners. First, due to the short length of tweets, twitterers should avoid packing diverse topics into a single tweet. One topic per tweet is recommended. Second, emergency practitioners should be thoughtful about embedding emoticons when crafting tweets during times of disaster. Although emoticons make tweets more friendly and emotive, they cannot provide necessary information that aimed at helping citizens at risk to be aware of their surroundings. Third, it turns out that Twitter URLs are an effective means of delivering additional information. However, given the statistical results, a caveat must be given to them regarding the inclusion of URLs: only 1 URL was seen to be reliably effective to replenish necessary information for tweets with low message clarity.

As with any other, the present study has several limitations that open opportunities for future research. First, as an exploratory study, the study lacks theoretical frameworks in testing specific empirical models. Therefore, future research should endeavor to confirm the current findings based on theoretical guidance. Second, message clarity per tweet was quantified based upon two algorithmic methods of the entropy model and the LDA technique. However, it will be useful to investigate the extent to which the message clarity of tweets conforms to how people actually interpret them. Third, the study relied on Twitter to answer the hypotheses. However, examining different media sources could enhance the current findings. Lastly, although this study reveals a "causal-mechanism" of Twitter URLs based upon the moderation analysis (Judd et al. 2001; Kenny 2015; Kraemer et al. 2002; Reis and Judd 2000), the causal-mechanism does not directly mean causality. Rather, it provides an important implication regarding the possible existence of casual mechanism. Therefore, a natural extension of the current study would be to provide empirical evidence about the casual effect of Twitter URLs on the relationship between message clarity and retweet frequency.

Overall, this study provides the new variable, *message clarity*, by grounding in Shannon's information theory. As a means to calculate each tweet's message clarity, we also introduced the computational framework consisting of computational linguistics techniques, such as POS tagging and topic modeling. As Twitter becomes more commonly used for disaster communication, our findings suggest that it is important for twitterers to enhance a tweet's message clarity by including only one topic for its wide dissemination and that the use of Twitter URLs compensates for the insufficient message clarity of tweets in association with retweeting.

Appendix 1.A

[9] Table 1.A.1 57 Topics and Keywords of the 2013 Colorado floods	
Topic #	**Keywords**
1	relief levels flood_levels give pic impression friends news add #twibbon create federer tennis victims online flood_relief flood_victims abc abc_news
2	need volunteer register volunteers clean cleanup #bnefloods volunteering brisbane food emergency accommodation #bakedrelief #bnecleanup needs needed
3	centre evacuation evacuation_centre showgrounds pets ipswich spread word ipswich_showgrounds rna evac rna_showgrounds centres lost found hills
4	change cross red climate red_cross #vicfloods climate_change clean rain australian towns weather relief services affected brace information brisbane
5	fill sandbags need free brisbane form council affected services nature disaster offer businesses local train_services stop contact mother_nature city
6	support map comparison map_comparison relief post affected blog rough event #vicfloods fundraiser peeps blog_post benefit devastation happening fundrasing
7	victims flood_victims stay released place ravaged advice friends airport police donation legal information free #vicfloods affected hotline recovery
8	volunteers helping proud disaster clean spirit hand #vicfloods army efforts relief together rescue australian amazing community #bnefloods workers
9	bligh anna_bligh anna premier brisbane queensland_premier low residents evacuate lying higher water ground inquiry ipswich urged #brisbane starting
10	crisis news flood_crisis bligh toll premier missing death latest anna_bligh anna dead disaster live death_toll online confirmed ahead buying
11	victims donate donating appeal remember sitting flood_victims donation link vic nsw left donations harvey amazing coast #auction vintage total
12	ipswich mayor looting ipswich_mayor paul pisasale city markers paul_pisasale find flood_markers brisbane higher mythbuster flood_mythbuster facing pi
13	water power brisbane residents safe supply ipswich #bnefloods shopping boil water_supply centre food victims drink advised cut flood_victims need
14	spirit aussie aussie_spirit amazing victims flood_victims home donate working flooded return family find cleaning heaps thanks_heaps strangers aussie
15	creek cars footage flash toowoomba washed video lockyer flash_flood lockyer_creek mil show evacuate gave film water mate oprah gympie god rises higher

16	victims flood_victims donate support donating every affected money #prayforaustralia raise need hope handset donations retweet visit generously coffee
17	#qld affected judgment judgment_day update insurance #bnefloods brisvenice brisvegas flood_update longer brisbane info hotline tourism #vicfloods bus
18	disaster size area zone declared texas disaster_zone times united flood_disaster france kingdom united_kingdom germany united_kingdoms kingdoms
19	power cut energex brisbane ipswich free affected homes image charge restore phones facing inundation families businesses mythbuster flood_myth
20	brisbane storage photos images free brisbane_floods live free_storage #bnefloods storage_king offering trucks #brisbane aerial affected pics amazing
21	cross red safe brisbane national registration system free cow roof #bnefloods clean water map inquiry place photos #brislantis damaged cross_national
22	high zoo swim crocs australia_zoo high_enough tying brisbane weather god biggest arrive bureau biggest_flood weather_bureau companies insurance_compa
23	media social social_media twitter #vicfloods health helping aid police hope australia_day need doctors join email stars disaster dept needed sunrise
24	brisbane river brisbane_river #bnefloods floating cbd farm drive streets list expected restaurant free park city affected coronation coronation_drive
25	man volunteers photo boatload kangaroos needed rescued #bnecleanup mayor kangaroo more_volunteers pic brilliant registration centres
26	volunteers auctions need awesome cahill qld_floods tim_cahill tim awesome_auctions cold beers ground high cbd mobile cold_beers handing high_ground
27	crisis flood_crisis list real media citizen reports citizen_reports died twitter related stories line info outlets lifeline twitter_list media_outlet
28	river brisbane broken brisbane_river banks end west library west_end wet sunny dry sunny_day wrap freezer gladwrap wet_photosbooks photosbooks
29	evacuation info centres financial brisbane app hit pledges evacuation_centres financial_help dogs cats owners recovery free staff information links b
30	river brisbane peak brisbane_river expected levels metres conference media ipswich #bnefloods media_conference flood_peak live level livestream tab
31	#bnefloods brisbane closed street bank ipswich bridge water pier eagle cbd shit open #brisbane south_bank river motorway holy holy_shit crap #fb road
32	stadium suncorp_stadium suncorp brisbane pool swimming picture field footy_field #bnefloods water fire bridge transformer emergency services silence

33	waters flood_waters children disaster helping barrier reef #auspol barrier_reef support office play water damage equipment replace stop homes pay
34	warning severe rain thunderstorm weather brisbane thunderstorm_warning flash coast hit #qld bay bom #tcanthony heavy river cyclone moreton
35	donate every appeal flood_appeal tweet cents aussie aussie_queensland #prayforaustralia retweet message #staystrong received qld_floods everyone
36	#vicfloods #nswfloods map need information road closures info flood_information road_closures contact crisis #tasfloods urgent list live flood_map
37	donate need queenslanders desperately police facebook updates page twitter phone flight qld_floods qld_police date change affected booking service
38	victims australian fundraiser items fan international win fan_fundraiser autographed auction autographed_items bed offer recent house affected spare
39	victims cahill auction experience flood_victims tim raise bid money tim_cahill #socceroos match ebay charity everton aid signed shirt cricket relief
40	affected survival animals offer email housing foster assistance email_floods foster_caretemporary caretemporary bill unnecessary lewis survival_value
41	abbott deep tony_abbott tony water #auspol dig flood_water donations bin wheelie indication wheelie_bin #nbn good_indication dollar political need
42	toll death death_toll valley lockyer found lockyer_valley missing rises bodies grantham police flood_death_toll dead flash news body man risen search
43	bligh anna_bligh anna premier conference gillard press julia crisis media julia_gillard press_conference leadership live pressure #abcnews leader qld
44	snake frog ride photo community hitches incredible escapes frog_escapes_flood incredible_photo looting bligh escape created riding anna red australia
45	appeal relief flood_relief_appeal aussies donate everyone thinking needs premier #aussies flood_appeal disaster donating relief_appeal donated
46	missing dead rice jordan jordan_rice confirmed #prayforaustralia hero brother died sad save queensland_floods lost boy saving homes rip god queensland
47	relief appeal flood_relief fund auction money raise proceeds donate donated flood_appeal raised signed bid relief_fund song funds raising sales donation
48	recovery tsunami inland biblical flood_recovery impact inland_tsunami crisis faces facing economic news hell official support force warns economic_impact
49	shark ipswich bull street goodna flooded brisbane bull_shark spotted flooded_street sharks affected update streets swimming bull_sharks main main_street

50	brisbane city council city_council latest live game news alert updates services support online #bnefloods notice media info collection drinking
51	rspca fairfield animals fairfield_rspca water qld_floods repost foster animal retweet shelter register origin raise money jerseys origin_jerseys
52	towns affected brisbane crisis news medical coal free water flood_crisis relief clean volunteer offering home car cities inundated recovery reds
53	safe affected everyone thoughts brisbane hope stay news #prayforaustralia friends prayers family home heart lost sad devastating hear watching rain
54	waters flood_waters city australian rockhampton braces brisbane fundraiser rise peak queensland_braces coastal rising river satellite bridge fundraising
55	end brisbane water west house home clean need #bnefloods mud west_end helping flooded hand hard river #vicfloods cleaning city power volunteers girl
56	flooded homes brisbane affected businesses need power stallion suburbs supply bay needed inundated water #bnefloods ipswich spare energex deception
57	jordan rice jordan_rice save swept younger rescuers brother life younger_brother blake own_life losing stop hero toowoomba aged waters #prayforaustralia
58	relief #vicfloods view volunteers cross hills needs bowen support concert red_cross neighbours bowen_hills service clean crisis brisbane continues
59	relief donate flood_relief donations needs appeal word spread information flood_relief_appeal everyone need #prayforaustralia qld_australia needed
60	heart health aussies praying safety prayers hearts breaks picture markets #bnefloods rocklea rocklea_markets brisbane disaster system team chopper fr
61	points velocity velocity_points brisbane closed donation allowing convert #bnefloods donate recovery donating donations road awesome page milton
62	appeal flood_appeal donate rspca animals give donations money qld_rspca raise need donated generously #vicfloods sales #prayforaustralia plead donati
63	damage insurance flood_damage need business food brisbane storm claims small milk pay water #bnecleanup supplies levy clean hit bread guide office
64	brisbane transport cross public red_cross centre public_transport volunteers red needs affected melbourne seekers asylum_seekers north needed based
65	brisbane cbd brisbane_cbd power closed evacuated transport coast myth public buster flood_myth_buster public_transport highway #bnefloods closing
66	found dogs dog disaster goodna island need floating toilet lost fraser block flood_disaster toilet_block fraser_island sharon pray god caltex sleep k

67	donate appeal fireworks day_fireworks cancel recovery donated money relief flood_appeal donation million free ride fund raising awareness twitter
68	levy video flood_levy tax #vicfloods friend gillard youtube relief #auspol pay victims images toowoomba view youtube_video nasa queensland_floods
69	water services fire lost normal goods kid home talent normal_kid stefanovic karl treatment affected karl_stefanovic plants supply summary room need s
70	coverage news abc maps radio brisbane live info local council information flood_maps online #abcnews site twitter updates channel #bnefloods city dig
71	town update residents dalby link area pool audio_link alert recovery pool_area emerald audio rockhampton road hit #police swimming power cut southern
72	dam wivenhoe brisbane water #bnefloods cbd lucia capacity view river street st_lucia brisbane_cbd albert farm full george new_farm southbank flooded

[10] Table 1.A.2 72 Topics and Keywords of the 2011 Queensland floods

Topic #	Keywords
1	toll, death, dead, rises, person, death_toll, flood_death_toll, evacuations, confirmed, people, deadly, presumed, woman, ordered, found, flood_toll, waters, missing
2	evacuation, center, head, jamestown, residents, notice, springs, eldorado, evac, creek, eldorado_springs, cty, evacuation_center, evacuation_notice, people, barn, ordered
3	towns, rescue, rain, rains, warnings, flood_warnings, diverse, closed, forecast, cats, flood_towns, colorado_towns, break, flood_rescue, stranded, brief_break, hamper, waters
4	schools, aurora, closed, creek, aurora_pd, creek_schools, aurora_schools, request, canyon, water, cherry, debris, valley, surge, foot, cars, other_debris, carrying, boulder
5	rescue, boulder, operation, water, flood_rescue_operation, area, report, continues, home, weather, leave, spill, chemical, historic, drive, fracking, rain, chemical_spill
6	record, breaking, guard, coast, led, worse, denver, concert, coast_guard, helicopters, relief, survivors, defense, coast_guard_helicopters, victims, benefit, state, coming
7	people, county, unaccounted, boulder, rescued, rescue, crews, sheriff, larimer, man, helicopters, save, officials, larimer_county, pets, boulder_county, racing, news, air
8	mountain, city, rocky, national, dam, commerce, arsenal, rocky_mountain_arsenal, evacuations, wildlife, refuge, failed, wildlife_refuge, impassable, roads, streets, east, dams
9	guard, national_guard, national, town, lyons, residents, jamestown, evacuations, moves, continue, boulder, evacuate, news, crest, downstream, colorado_town, students

10	creek, boulder, boulder_creek, move, broadway, sirens, sounding, #cuboulder, higher, canyon, cfs, east, ground, higher_ground, mesa, place, rising, table, flood_sirens, shelter
11	canyon, boulder, water, ground, higher, higher_ground, wall, coming, boulder_canyon, creek, immediately, move, boulder_creek, gulch, emerson_gulch, emerson, seek, debris, pearl
12	boulder, rain, evacuate, flash, more_rain, continue, ordered, live, county, officials, rescues, expected, braces, lyons, flash_flood, colorado_braces, damage, town, downtown
13	warning, flash, flash_flood_warning, boulder, flash_flood, county, issued, flood_warning, counties, effect, watch, skies, rain, warnings, evacuees, denver, springs
14	creek, boulder_creek, boulder, water, flow, wall, usgs, official, denver, term, experts, tsunami, experts_term, readings, creek_flow_readings, sensor, fourmile, usgs_sensor
15	platte, river, oil, south, spills, south_platte_river, gallons, tank, swollen, platte_river, spill, south_platte, damaged, morgan, reported, waters, water, oil_spill, Greeley
16	oil, gas, spills, zones, #fracking, wells, tracking, flood_zones, waters, sites, fracking, flood_waters, post, flooded, chemicals, water, gas_wells, leaks, denver, denver_post
17	gallons, locations, road, drenched, crude, dumps, spill, oil_spill_dumps, closures, waters, road_closures, flooded, boulder, water, many_locations, loved, shelter, affected
18	disaster, flood_disaster, media, blackout, media_blackout, #fracking, fracking, spills, happening, photos, update, toxic, worse, confirmed, shocking_photos, underwater, zone
19	waters, water, flood_waters, piano, house, play, home, sewage, wrecked, boulder, decided, man, contaminated, avoid, plays, sweep, moments, bike, creek, colorado_home, stay, video
20	vrain, water, river, creek, bridge, evac, roads, lyons, place, street, boulder, vrain_river, longmont, home, loveland, dry, center, big, #longmont, stay, hygiene, news, left
21	thompson, big, river, thompson_river, feet, county, ravaged, woman, pound, fatality, canyon, fifth_fatality, thompson_canyon, stage, record, loveland, central, thompson_flood
22	photo, car, havana, viewer, lyons, viewer_photo, swim, road, air, hwy, town, boulder, damage, hwy, news, water, dillon, pic, collapse, assessment, rescue, road_collapse, inside
23	longmont, #longmontflood, victims, water, lyons, view, rescues, equine, dam, storm, helicopter, vehicles, register, volunteers, image, urgent_call, woman, soldier, blog
24	long, water, city, safe, boulder, photo, rain, washed, picture, commerce, denver, commerce_city, rescue, stay, roads, house, areas, problems, live, couple, photos, send, yards, mile
25	images, unbelievable, unbelievable_images, boulder, map, google, tremendous, began, crisis, area, travel, water, notice, earth, evacuation, severe, google_earth, flash
26	game, football, school, state, path, bike, bike_path, postponed, high, fresno, field, pic, park, aurora, high_school, utah, utah_park, baseball, baseball_field, overland

27	front, range, front_range, boulder, coverage, open, space, water, emergency, relief, trucks, workers, rescue, hard, downtown, disaster, county, working, longmont, effort, parks
28	damage, photos, aerial, images, flood_damage, video, biblical, climate, line, trends, boulder, climate_trends, views, biblical_flood, show, waters, aerial_views, lyons, shot
29	campus, evacuation, damage, homes, water, mobile, school, mobile_homes, creek, high, epic, buildings, boulder, photo, shows, water_damage, city, shelters, closed, high_school
30	big, thompson, canyon, thompson_canyon, road, hwy, hwy, thousand, boulder, flooded, water, science, thompson_canyon_entr, entr, baseline, damage, photographers, cut, deep
31	road, closures, road_closures, map, list, county, updates, boulder, closure, updated, found, #copets, center, shelters, latest, shelter, evacuation, road_closure_map, roads, dog
32	park, hwy, hwy, closed, estes, estes_park, #cotraf, open, road, roads, highway, photos, disaster, #estespark, directions, news, fun, reporter, app, denver, evergreen
33	water, boil, residents, high, drinking, lyons, safe, treatment, drink, advisory, hand, boulder, district, city, vehicles, wastewater, left, bottled, town, levels, contaminated
34	recovery, information, response, volunteer, relief, resources, updates, communities, efforts, live, emergency, cleanup, blog, affected, local, boulder, long, flood_recovery
35	disaster, assistance, fema, boulder, emergency, county, recovery, center, counties, federal, disaster_assistance, declaration, map, affected, evacuation, register
36	damage, losses, billion, flood_damage, property_losses, relief, repairs, shutdown, property, million, government, flood_relief, highways, left, street, bridges, estimated
37	aid, unanimously, republicans, relief, sandy, sandy_aid, colorado_republicans, opposed, support, flood_relief, voted, house, house_republicans, flood_relief_unanimously
38	biden, recovery, hickenlooper, devastation, flood_devastation, damage, view, president, fema, efforts, joe, gov, response, vice_president, team, vice, joe_biden, news, rescue
39	victims, relief, word, free, spread, #cofloodrelief, storage, free_storage, flood_victims, fund, flood_relief, giving, donating, donated, flood_relief_fund, marijuana
40	relief, victims, flood_victims, #cofloodrelief, donate, efforts, flood_relief, support, fundraiser, benefit, affected, donations, relief_efforts, effort, raised, helping
41	people, unaccounted, oem, areas, boulder, rain, more_rain, awaits, number, center, flood_areas, boulder_oem, remain, home, shelter, stop, area, volunteers, listed, report
42	homes, unaccounted, people, destroyed, damaged, dead, evacuated, missing, shelters, search, homes_damaged, update, loved, safe, register, presumed, homes_destroyed, numbers
43	family, impacted, pray, fire, guard, epic, reach, flush, truck, zone, members, stranded, driving, food, video, housing, flood_zone, fire_truck, order, guard_members, residents

44	cross, red, victims, flood_victims, red_cross, give, texting, climate, change, shelter, climate_change, affected, shelters, people, volunteers, american, #cofloodrelief, safe
45	collins, fort, fort_collins, relief, south, view, north, support, efforts, friends, #foco, based, resorts, vail_resorts, denver, co_support, closed, relief_efforts, pass, season
46	canyon, boulder, residents, people, shelters, left, stayed, hand, water, boulder_canyon, springs, evacuated, overnight, creek, road, expected, support, providing, #redcross
47	safe, needed, share, #copets, pets, food, victims, volunteers, lost, animals, home, hay, register, #cofloodrelief, pet, loved, victim, longmont, check, disaster, donations, sign
48	pets, rescued, people, visit, best_way, evacuated, helicopter, victims, katrina, survivors, number, historic, #nationalguard, historic_flood, #copets, greatest_number, town
49	boulder, longmont, springs, closed, humane, open, manitou, society, humane_society, page, ave, #waldoflood, center, shelter, west, front_page, manitou_springs, animals, #hmrd
50	safe, boulder, stay, rain, friends, prayers, thoughts, people, hope, affected, home, good, dry, family, love, raining, bad, crazy, victims,
51	schools, aurora, closed, creek, aurora pd, creek schools, aurora schools, request, canyon, water, cherry, debris, valley, surge
52	rain, inches, totals, wild, instagrams, wild flood, rainfall, snow, boulder, received, map, record, past, annual, feet, rain totals
53	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery
54	live, victims, coverage, flood_victims, rocks, force, task, red, task_force, state, rain, red_rocks, news, rescues, continue, debris, good, water, honor, team, oil, photo, tribune
55	disaster, boulder, waters, flood_waters, people, allowed, fracking, tubing, boulder_pd, reminds, flood_disaster, cited, floodwaters, fracking_disaster, missing, sky, clears
56	county, weld, boulder, denver, post, weld_county, residents, denver_post, water, closed, boulder_county, evacuations, pipeline, road, oil_pipeline, roads, oil, blvd, rain
57	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery, more_rain, snarls, fundraiser, latest_updates, live

Appendix 1.B

Correlation Matrix 2011 Queensland											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Retweets_24h	1										
Entropy	-0.0518***	1									
URLs	0.00714*	-0.137***	1								
Emoticons	-0.00194	-0.00124	-0.0925***	1							
Words	-0.0252***	-0.109***	0.0599***	0.0362***	1						
Hashtags	0.0161***	0.123***	-0.316***	0.0723***	-0.0965***	1					
Ln(Followers)	0.0151***	0.00441	-0.0675***	-0.0136***	0.0339***	-0.173***	1				
Ln(Followees)	0.100***	-0.0640***	0.0557***	-0.00614*	0.0556***	-0.00775*	0.0665***	1			
Ln(Likes)	0.0268***	-0.0264***	-0.0465***	0.0120***	0.148***	-0.0235***	0.0855***	0.680***	1		
Ln(Status)	0.0202***	-0.0196***	-0.0289***	0.0629***	0.146***	-0.0449***	0.0914***	0.329***	0.362***	1	
Mention_YN	0.0368***	-0.0544***	0.0243***	0.0319***	0.0980***	-0.0207***	0.0491***	0.742***	0.488***	0.442***	1
2013 Colorado											
Retweets_24h	1										
Entropy	-0.0744***	1									
URLs	-0.00445	-0.119***	1								
Emoticons	-0.0108***	0.0170***	-0.0593***	1							
Words	0.00279	-0.0589***	-0.0284***	0.0179***	1						
Hashtags	0.0375***	0.0164***	-0.240***	0.0338***	-0.00794*	1					
Ln(Followers)	0.0315***	0.0498***	-0.106***	0.00880**	0.0822***	-0.207***	1				
Ln(Followees)	0.209***	-0.0947***	0.0818***	-0.0104***	0.140***	0.0846***	0.0517***	1			
Ln(Likes)	0.0733***	-0.0385***	0.0226***	0.0059	0.185***	0.0439***	0.0900***	0.737***	1		
Ln(Status)	0.0403***	0.0322***	-0.141***	0.0521***	0.187***	-0.0111***	0.144***	0.361***	0.476***	1	
Mention_YN	0.0541***	-0.129***	0.155***	-0.00939**	0.0517***	0.0948***	-0.0963***	0.670***	0.498***	0.347***	1

Appendix 1.C

[11] Table 1.C.1 Test of Multicollinearity – 2011 Queensland floods			
Linear	1.04	Ln(Followees)	2.03
URLs	1.17	Ln(Likes)	1.33
Emoticons	1.02	Ln(Status)	2.51
Words	1.18	Mention_YN	1.06
Hashtags	1.06		
Ln(Followers)	3.30	**Mean VIF**	1.57

[12] Table 1.C.2 Test of Multicollinearity – 2013 Colorado floods			
Linear	1.04	Ln(Followees)	2.49
URLs	1.19	Ln(Likes)	1.44
Emoticons	1.01	Ln(Status)	2.02
Words	1.16	Mention_YN	1.06
Hashtags	1.13		
Ln(Followers)	3.05	**Mean VIF**	1.56

Chapter 2. Essay 2: Complementary Effects Between Twitter's Heuristic and Systematic Information on Retweet Likelihood during Times of Disaster

Abstract

During disasters, Twitter has been used to rapidly disseminate disaster-related information to the affected public in a timely manner. However, there is a growing concern about the credibility of tweets, which sparks twitterers to seek confirming information and thus delays the dissemination of tweets. With a sense of urgency about disaster events, twitterers cannot afford to closely evaluate the credibility of tweets' content that requires high cognitive effort. Rather, heuristically processable twitterers' profile information can supplement the quick assessment of the credibility of the content. Using the notion of technological affordances, we interpret twitterers' profile information as source-credibility-cues. Based on the heuristic-systematic model (HSM) of information processing, we examine both twitterers' source-credibility-cues as heuristic information and tweets' content as systematic information in association with quick retweeting. From our analysis of tweets collected from the 2011 Queensland and 2013 Colorado floods, we demonstrate that twitterers' information, such as the number of followers, likes, tweets about the current events, and the length of affiliation, helped other twitterers quickly decide whether to retweet when they were in need of additional information. This study enhances our understanding of how the different types of information provided by Twitter can influence quick retweet decisions during disasters.

Keywords: *Twitter, Disaster Communication, Topic Modeling, Heuristic-Systematic Model of Information Processing, Technological Affordances, Source-Credibility-Cues*

2.1 Introduction

Creating emergency information and then spreading such information in a timely fashion to the at-risk public is of critical importance for the success of emergency management. Traditional and social media have played a pivotal role for communication during times of emergency (Oh et al. 2013; Sutton et al. 2008). Particularly, combined with the features of social networking and mobile technology, Twitter has attracted the public in disaster-stricken areas, emergency responders, and online citizens who purposely relay information, because it provides up-to-date information of local relevance (Fraustino et al. 2012b) at a faster rate than traditional media, and even other social media (Lachlan et al. 2014; Stephens and Barrett 2014). That said, as more and more information is available on social media, researchers are paying attention to the credibility of information in terms of content and its source (Bruns 2008; Hu and Sundar 2009; Lachlan et al. 2014; Sundar 2008; Westerman et al. 2014). The primary reason for this concern is that unlike traditional news media, social media in general significantly lacks the function of gatekeeping, the process by which topics, news, and issues should be verified for broadcast, publication, and dissemination (Westerman et al. 2014). Gatekeeping is the central role of traditional media in establishing fact-based, objective reporting as a means to prevent anything and everything from being published and to filter out false and even unnecessary information (Salcito 2009). While traditional media is maintained by diverse professional gatekeepers ranging from journalists to editors, to reliable organizations (Westerman et al. 2014), social media is managed mostly by its citizen members (Westerman et al. 2014). In this sense, it is highly probable that information disseminated throughout social media channels is not properly validated by professionals, and lacks either author identification or established reputation (Metzger 2007), a factor that confuses recipients with the trustworthiness of such information (Sutton et al. 2008).

As one of the most prominent social media for disaster communication, Twitter is no exception to the concern about information credibility. Several studies investigated the content credibility of tweets. Rumor research empirically demonstrated a negative correlation between the dissemination of the disaster-related content and its credibility (Castillo et al. 2011; Mendoza et al. 2010; Mendoza et al. 2013). More specifically, Acar and Muraki (2011) reported that the content-credibility of tweets was the major problem that rescuers encountered when deciding what to do with disaster-related tweets. Little research has been conducted, however, in the context of source-credibility during disasters, and whether or not the rapid and wide dissemination of *verified* information is necessary. Hu and Sundar (2009) pointed out that author information was correlated with source-credibility. Specifically, Sundar (2008) contended that as heuristic cues, the information about the author could influence recipients' credibility-assessment of information generated by him or her. In connection with the arguments previously mentioned, we claim that a twitterer's information—such as followers, followees (or friends), likes (one's tweets favorited or liked by other twitterers), status (one's total posted tweets), and so on—is associated with the credibility-assessment of his or her tweets. Therefore, investigating twitterers' information as credibility-cues will contribute to enhancing our understanding of information dissemination through Twitter during times of disaster.

Retweeting is an act of re-posting an original tweet, a key feature of Twitter to rapidly disseminate tweets to a large audience (Compston 2014). Retweeting allows voluntary and collective participation of twitterers to share interesting, useful, or imperative tweets with other twitterers (Abdullah et al. 2014; Starbird and Palen 2010; Sutton et al. 2014b; Zubiaga et al. 2015). Significantly, retweeting has been considered to be a gatekeeping process, in the sense that rather than crafting their own tweets, twitterers can collectively determine which information should be emphasized, further discussed, and diffused into their communities (Bruns 2008). Sutton (2010) observed that during emergencies, twitterers actively participated in fact-checking of received tweets before retweeting. In this regard, prior research viewed retweeting as a recommendation system for tweets, which harnesses the collective intelligence of loosely connected twitterers (Boyd et al. 2010; Starbird et al. 2010; Vieweg et al. 2010). Hence, compared to tweets with less retweet attention, tweets that receive more retweet attention are considered to have more important information (Abdullah et al. 2014; Shaw et al. 2013) and to maintain higher credibility (Shaw et al. 2013; Sutton 2010). By taking source-credibility, retweeting, and disaster situations into account, we address the following research question:

RQ: How does twitterers' information, as a source-credibility-cue, affect the retweeting of their tweets during times of disaster?

Due to human's limited capabilities for information processing (Lang 2000), recipients under the pressure of urgency should cope with information overload to reach a quick judgement (Metzger et al. 2010): the affected public tries to take protective actions for their safety before threats strike; citizens online should quickly determine whether to forward received information to others in need. We contend that such a situational restriction influences recipients to strategically process disaster-related information in an endeavor to minimize the information-processing effort and that, as Liu et al. (2012), Metzger et al. (2010), and (Sundar 2008) argued, the author's information, as heuristic information, can be used as a subsidiary means to ease processing effort. Interestingly, the heuristic-systematic model (HSM) of information processing theory specifies when people systematically interpret a message's content while processing its heuristic information (Chaiken 1980; Chaiken and Eagly 1989; Chaiken and Maheswaran 1994). The HSM provides a theoretical explanation for conditions related to when and why people engage in heuristic and systematic information-processing, which closely accords with our research purposes. Therefore, in order to examine the relationship specified in RQ, we should know when and why the public seeks heuristically processable, additional information while systematically processing disaster-related information.

The contribution of this study is twofold. First, in using the HSM, we demonstrate the interaction relationships of twitterers' information—as heuristic cues—with their tweets' content features—as systematic cues—relative to retweet likelihood. It is noteworthy that we utilize the notion of technological affordances to interpret twitterers' information as source-credibility. Second, drawing upon our empirical findings, we suggest a practical recommendation to emergency responders and online citizens for crafting effective warnings and alerts via short message services designed to help the public in disaster stricken areas.

The paper is structured as follows. First, we provide a literature review on Twitter and its use for disaster communication, credibility issues in social media in general, and the notion of technological affordances. We then present the theoretical model of this study and hypotheses to explore in order to empirically investigate the raised research questions. After that, descriptions of research methodology, data, and the results of hypothesis testing follow. We conclude with discussions of our findings and their implications for future research.

2.2 Literature Review

2.2.1 Twitter in Action for Disaster Communication

In 2006, Jack Dorsey, a co-founder of Twitter, posted the first tweet "*just setting up my twttr*" (Siese 2016). In tandem with the wide use of mobile phones, the advent of Twitter signaled the era of terse messages in our everyday mobile communication. In particular, short messages disseminated through Twitter have stood out in emergency situations where spreading emergency alerts and warnings to people in target areas is a most critical task (Sutton et al. 2015b). Along with its non-reciprocal, easily improvised following relationship and the 140-character limit of tweets, Twitter has become optimized not only for media-rich smartphones, but it is also compatible with most communication platforms on the Web, and even old-style cellular phones (Starbird and Palen 2010; Vieweg et al. 2010).

As a disaster communication medium, Twitter abounds with positive examples. Sutton et al. (2008) reported that, during the 2007 California wildfires, situational information broadcast by traditional media sources was not local and accurate enough, and even slowly updated. Therefore, the affected people turned to social media services, especially Twitter, to create, seek and share locally relevant information about the moving wildfires. With one accord, twitterers voluntarily traced a series of wildfires as they happened, and shared time-sensitive details about road closures, evacuation and shelter instructions, and the fire line shifts (Hughes and Palen 2009). In response to the Haiti earthquake in 2010, Twitter enabled the American Red Cross to raise emergency funds from 2.3 million participants in two days (Manjoo 2010). In 2011, when the Tohoku tsunami in Japan destroyed all communication and power infrastructure facilities, Twitter (installed in wireless mobile phones) was known to be the only communication means for the public living in affected areas to communicate, as well as to be aware of constantly moving threats (Acar and Muraki 2011). During the 2013, Boston Marathon bombings, twitterers improvised collaboration networks for disseminating information about warnings and imminent threats, guidance for minimizing further damage, and aid for recovering from the bombings (Sutton et al. 2014a). The aforementioned examples highlighted well for emergency communication the advantages of Twitter’s short message length and follower network. Twitter is not freed from the issue of information credibility, however, which primarily stems from the lack of a gatekeeping process. In the environment of traditional media, where the production cost of information is expensive and its distribution channels are scarce, central authorities (or gatekeepers) play a key role in verifying and confirming the credibility of information (Bruns 2008). In contrast, social media’s environment is to a great extent different from the traditional media environment in that: (1) diverse formats of information (i.e., text, picture, and audio/video) can be created at little or no cost; (2) abundant and diverse information can be instantly accessed and widely disseminated through other online channels; lastly (3) amateur citizen journalists are creators of such information and serve as gatekeepers as well. Although industry journalism has repeatedly denounced unskilled, amateurish citizen journalists (Bruns 2008), it is undeniable that when determining what events to report, what stories to cover, and how to effectively organize information, most citizen journalists lack expertise, compared to industry journalists and editors. When all things are taken together, the credibility of information on Twitter, as a prominent social media for disaster communication, is still largely unexplored.

2.2.2 Credibility-Assessment of Online Information

On social media, information credibility matters, and this issue is even more critical during emergencies as more and more online citizens desperately seek reliable information. Emergency alerts and warnings about approaching and current threats allow only short periods of time for the affected public to take protective actions (Sutton et al. 2014b). This is true also for online citizens who re-transmit critical information as a purposeful action to help others (Boyd et al. 2010). In this context, online citizens must constantly monitor information while quickly deciding whether or not to forward information based on its importance and credibility.

The credibility of information, as a complex and multifaceted variable, is generally evaluated by the extent to which a given medium, message, and source are believable. Therefore, the credibility-assessment is known as a time-consuming task that requires a high level of cognitive effort (Hu and Sundar 2009; Sundar 2008). However, several studies have claimed that source-credibility could provide a shortcut for the credibility-assessment of information: 1) Eastin (2001) reported empirical evidence that information about an author's expertise increased the perceived credibility of his or her information; 2) Metzger et al. (2010) found that during disasters, online information seekers actively engaged in processing source-related information to quickly assess the credibility of received information, while reducing cognitive effort and mitigating time pressure. These findings shed light on the influence of source-information as "credibility markers" to indicate the extent of information trustworthiness (Sundar 2008, p. 74) and as informational cues that are heuristically processable based on simple decision rules or learned knowledge (Todorov et al. 2002).

In disaster situations in which a high level of urgency restricts the public to have enough luxury to explore and process information (Runyan 2006), heuristic information, like source-information, has been highlighted as a quick decision-making tool that affects the credibility-assessment of information (Hu and Sundar 2009; Liu et al. 2012). Making decisions using heuristic information complements human's limited capacities for information processing (Lang 2000), and alleviates restricted cognitive resources deployable under the pressure of great urgency (Chaiken and Eagly 1989; Todorov et al. 2002). Simply put, heuristic information can decrease the amount of time needed for evaluating the credibility of information (Havard 2001). In this sense, source-credibility becomes a crucial factor for elaborating the intertwined relationship between Information and Communication Technology (ICT) and humans in pursuing their particular goals of disaster communication.

2.2.3 Technological Affordances of Twitter

Affordances are objects in the world that offer possibilities for action (Majchrzak et al. 2016) or that are designed to interact with people (Gaver 1991). Hutchby (2001) viewed affordances as "functional and relational aspects which frame, while not determining, the possibilities for agentic action in relation to an object." (p. 444) For instance, a rock may offer affordances for animals as a shelter from a storm or a concealment for protection from predators. A keyboard as an input device allows computer users to perform the action of typing characters and numbers. Similarly, a computer mouse provides possible actions for users to click buttons, scroll documents or web pages up and down, and drag certain objects on the computer screens. In this sense, an affordance is an action potential that influences goal-oriented actions by humans, but such potential can be perceived differently depending upon individuals' different levels of capabilities (Pozzi et al. 2014) or different contexts (Hutchby 2001). Also, an affordance is a relationship between an object and an actor, such that the actor leverages the features of the object to achieve particular goals. The implications of affordances can be equally applied to the relationship between a technology and its users, given that the users perceive its properties and features and then utilize the technology in different ways (Majchrzak et al. 2016). In other words, the meanings and usages of a technological artefact are socially shaped and reshaped through interactions with its users, while being framed by its users' practical usages (Hutchby 2001). In this perspective, technological affordances occur when both an IT artefact's capabilities and its users' purposes are equally realized (Pozzi et al. 2014), opening up a range of possibilities for investigating how IT artefacts are associated with some kinds of social effects.

As a technology and a communication medium, Twitter exists in socially interactive surroundings, and its meaning evolves from social relations among twitterers. Therefore, we view Twitter as an IT artefact comprised of a set of affordances that allow various interactions with twitterers, such that behavioral consequences (i.e., retweeting) have to be examined in line with how the properties or features of Twitter interact with twitterers in pursuit of their particular goals. Affordances are perceived differently from individual to individual as well as from circumstance to circumstance (Hutchby 2001). As such, Twitter's affordances are recognized in different ways during times of disaster. During disaster situations that are characterized by the challenges of information overload and a varying message quality

(Sundar 2008), twitterers are struggling with a limited capacity for information processing (Lang 2000) and are under great stress (Chaiken and Eagly 1989; Todorov et al. 2002). Therefore, we assume that twitterers would interpret Twitter-related affordances as a means to reach a fast judgement, such as the credibility of information (Sundar 2008; Westerman et al. 2014). Hence, source-information is our primary interest, and source-information is known to affect the credibility of information (Hu and Sundar 2009; Metzger et al. 2010; Sundar 2008; Westerman et al. 2014). The source-information of Twitter, *followers*, *followees*, *likes*, *status*, and *join date*, trigger the heuristics of bandwagon, authority, and social presence (Sundar 2008). In addition, one more related information is generated to reflect twitterers' *recency* pertinent to the current events—*tweets* about the current events. Overall, Twitter's source-information may help recipients to reduce cognitive effort in assessing the credibility of tweets' information during disasters.

2.3 Theoretical Background

2.3.1 A Theoretical Framework of Information Processing

"People rarely process information in perfect conditions. There are both environmental and cognitive constraints on information processing" (Todorov et al. 2002, p. 196). Chaiken's heuristic-systematic model (HSM) states that when processing information, people balance systematic and heuristic processing by considering both their motivation (environmental constraint) and cogntive capacity (cognitive constraint) (Chaiken 1980). That means, with sufficient motivation and available cognitive capacity, people likely employ *systematic* processing and, as such, examine a message's content carefully to form a judgement; on the other hand, people lacking motivation and cognitive capacity probably employ a *heuristic* processing strategy and, as such, heuristically process a message's superficial cues such as its length or author information to draw a conclusion (Todorov et al. 2002).

As an analytic orientation to information processing, systematic processing involves the comprehensive analysis of a message's content. Therefore, people in this systemic processing mode need higher cognitive effort and more cognitive resources than heuristic processing (Chaiken and Maheswaran 1994). Heuristic processing, by comparison, relies on heuristic cues that trigger simple decision rules, such as "Experts can be trusted" and "Consensus implies correctness" (Todorov et al. 2002, p. 197). Thus, people in this heuristic processing mode require less cognitive capacity than systematic processing in order to reach a message's conclusion (Chaiken and Maheswaran 1994).

The HSM postulates that heuristic and systematic processing can act together, and suggests the following three hypotheses of interplay (Todorov et al. 2002). First, the additive hypothesis assumes the independent effects of both heuristic and systematic processing on a message's conclusion due to consistent judgemental implications derived from both. Second, the attenuation hypothesis states that the judgemental implication of systematic processing can attenuate the implication of heuristic processing, because both processing modes are in opposition to one another. Lastly, the bias hypothesis states that the judgemental implication of systematic processing can be construed as agreement with that of heuristic processing, because a message's argument is unclear. For example, the same unclear message can be differently interpreted by recipients who believe that its source is reliable and by other recipients who do not believe (Todorov et al. 2002).

Among the three types of interplay between the two processing modes, the bias hypothesis better explains how heuristic and systematic processing interact with each other during disasters on Twitter. The reason for this is that as emphasized by Zeng et al. (2016), Bruns and Stieglitz (2012), and Heverin and Zach (2012), the public is highly motivated as 'public editors' of disaster-related information (Sutton 2010, p. 6). Therefore, individuals systematically process a received tweet's content features, such as words and hashtags, in order to understand its message. When noticing a tweet's unclear message, the public can take the following strategies: first, individuals try to judge this message in line with its twitterer (author) information; second, they try to judge this message by acquiring additional information through embedded Twitter URLs or other sources. Under time pressure, the first strategy should be more effective, such that they heuristically process twitterer's information as a way to minimize cognitive capacity and thus quickly reach a conclusion before retweeting. Therefore, to examine an individual twitterer's information as a credibility-cue, it is essential to consider tweets' message clarity for disaster communication.

2.3.2 Short-Length Tweets

Constrained by a 140-character limit, a tweet may not convey all pertinent information about a disaster event (Sutton et al. 2015a). Instead, atypically shortened messages (Stephens and Barrett 2014) carrying severely restricted details (Bean et al. 2015) might be broadcast through the Twittersphere by tweets. From this perspective, content is another durable determinant of public response (Bean et al. 2016; Bean et al. 2015; Sutton et al. 2015b). For example, Mileti and Sorensen (1990) suggest a guideline to craft clear warning messages to maximize their effects. First, warning messages should describe specific characteristics of a hazard, like "*a wall of water <u>20 feet high</u> <u>moving at 40 miles</u> per hour.*" Second, to maximize the public safety and minimize harm to life and property, actionable information should be unambiguously provided, like "*get to ground <u>higher than the top</u> of City Hall.*" Third, situation information, such as the location and time of the impending hazard, should be clearly included: "*the area of town that will flood will be between <u>Second</u> and <u>Fifth Street</u>, from <u>Elm Avenue</u> to <u>Magnolia Boulevard</u>*" or "*the tsunamis will not strike <u>before 10 p.m.</u> this evening, and you should be on <u>the northern side of U.S. Highway 72 9:45 p.m.</u> to be on the safe side.*" Lastly, to enhance the credibility of a warning message, the verifiable information source should be provided, like "*the National Weather*

Service, *the American Red Cross*, or *the head of civil defense*." Until recently, messages about emergencies have been viewed in light of quite a longer length of 1,380 characters (Sutton et al. 2015b). While Twitter has demonstrated the superior capabilities in its speed and scale of information dissemination, we should be mindful that tweets' 140 characters can restrict the chance that all the aforementioned criteria are incorporated within a single tweet (Sutton et al. 2015b). As a result, tweets could be considered to be uninformative and confusing (Bean et al. 2016), raising doubts about information credibility, quality, and/or accuracy (Hu and Sundar 2009).

2.3.3 Tweets with Multiple Topics for Warnings and Alerts

As we discussed above, tweets themselves are very short so as to satisfy the general requirements of warnings and alerts. Particularly, as twitterers craft a 140-character tweet with multiple topics, this tweet's information per topic unavoidably decreases, producing confusion for recipients (Bean et al. 2016). We coin the term "message clarity" to explain the relationship between tweets and the number of topics in them. That is, as a single tweet's number of topics increases, its message clarity decreases. For example, when Tweet 1 describes two topics and Tweet 2 has one topic, Tweet 1 could be considered to have less information per topic than Tweet 2. That is to say, the 140 characters of Tweet 1 can be used to describe only one topic, while Tweet 2's 140 characters are split between two topics. Accordingly, the message clarity of Tweet 1 is lower generally than that of Tweet 2.

Our argument is well aligned with Shannon's "noise" in communication theory—"a measure of one's freedom of choice in selecting a message" (Shannon 1949, p. 19). When noise is present in a message, the message is assumed to have certain degrees of uncertainty, causing distortions and errors (Shannon 1949). To observe and quantify noise in a message, they proposed the measure of entropy, consisting of p_i as the proportion of the ith topic out of n topics of a message m.

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

Entropy indicates that as a message's entropy increases, its noise increases as well. For example, when a message holds two topics with the same proportion of 0.5,[7] its entropy is 0.693.[8] In contrast, the entropy of another message with only one topic is 0.[9] Table 2.1 shows a few more cases that show the relationship between entropy values and different numbers of topics in a single message. We confirm that as a tweet's number of topics increases, its entropy value increases, implying that the clarity level of this tweet decreases accordingly. Therefore, we can safely argue that there is the negative relationship between the number of topics in a tweet message and its message clarity.[10]

[13] Table 2.1 Entropy by Different Numbers of Topics in a Single Message

Topic (Proportion) / **Message's Entropy**	• Topic 1 (1)	• Topic 1 (0.9) • Topic 2 (0.1)	• Topic 1 (0.5) • Topic 2 (0.5)	• Topic 1 (0.33) • Topic 2 (0.33) • Topic 3 (0.33)
Entropy	0	0.325	0.693	1.098

2.4 Research Setting and Hypotheses

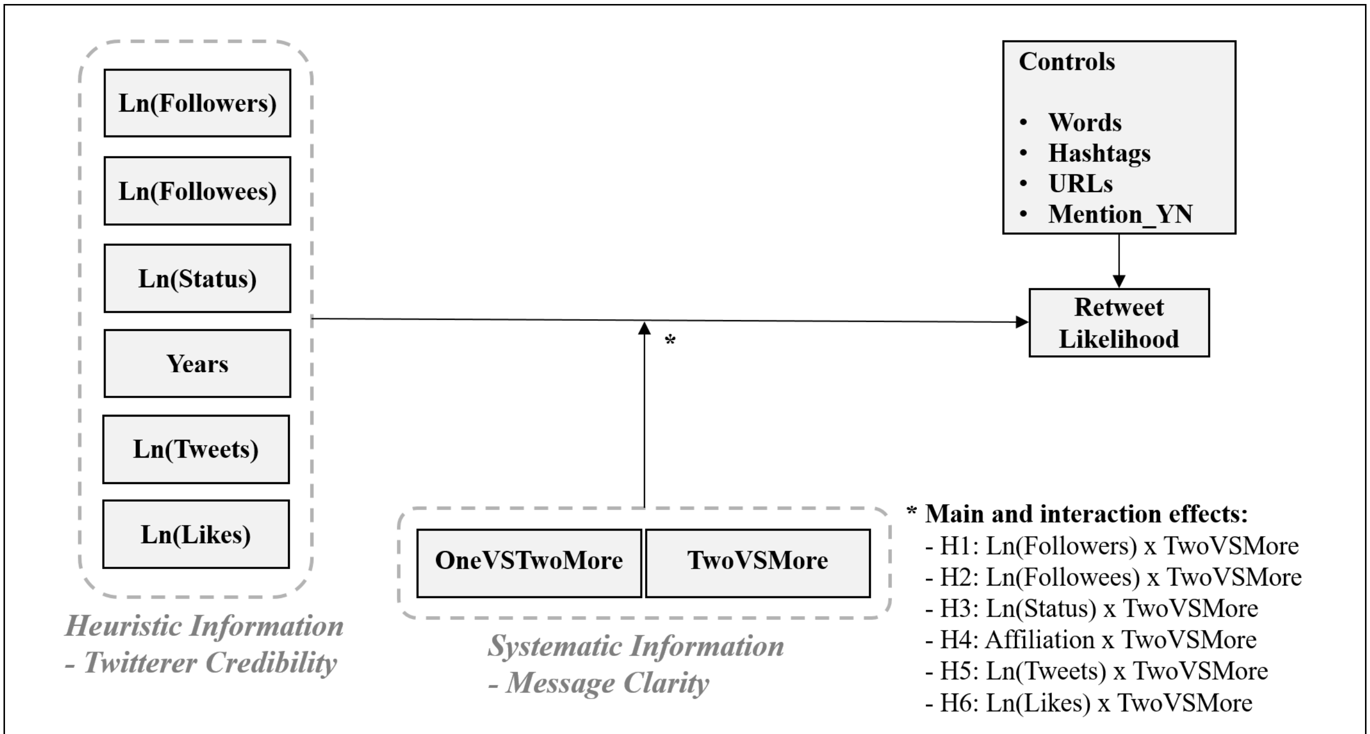
2.4.1 Research Setting

Heuristic-Systematic Model (HSM) as the Theoretical Research Model. Retweeting occurs after individual twitterers assess and validate information in tweets (Starbird and Palen 2010; Sutton et al. 2014b). Hence, retweeting is a decision-outcome of twitterers when processing tweets' content features and/or author information (Zhang and Watts 2008). Within the context of this study, we assume that the degree of a tweet's message clarity is determined by *systematic* information processing, while twitterers'

---

[7] p_1 is the proportion of the first topic, and thus $1-p_1$ is for the second topic.
[8] -(0.5*Log(0.5)+0.5*Log(0.5))=0.693
[9] -(1*Log(1))
[10] We empirically validated this relationship by using two Twitter datasets. See Appendix 2.A.

information is associated with *heuristic* information processing. Therefore, we utilize the HSM to establish relationships between heuristic and systematic information processing in order to observe how twitterers' information as a credibility-cue is associated with retweeting. In particular, we rely on the HSM's bias hypothesis to interrelate credibility-cues and message clarity. Figure 2.1 depicts all the relationships that are needed for the study.



[5] Figure 2.1 Research Model

Retweet Likelihood within the First 10 Minutes. As the outcome variable of the research model, tweets' retweet likelihood within the first 10 minutes after posting is of our primary interest, as the rate of response helps us to evaluate twitterers' quick retweet decisions during emergencies. There are several reasons for this. First, as we assumed, twitterers use heuristic cues to make a quick decision, rather than a deliberate decision. Therefore, we need to restrict a time interval between tweets and their retweets. Second, Twitter is a fast and responsive communication medium, such that information conveyed by tweets can easily become obsolete (Wilensky 2014). As a result, we need to define a short enough time interval to ensure that information by tweets is still valid. The same thing is true for additional information, such as twitterers' information. In addition, a quick judgment would not be necessary for already outdated information. Third, we suspect that varying time intervals could change explanatory factors' coefficients and even significant levels, resulting in undermining the confidence of the empirical results of the study. As a consequence, instead of using the retweet likelihood based upon a subjectively chosen time interval of an hour, a day, or a week, we have to objectively determine a time interval that could best represent both twitterers' quick judgement and tweets' retweetability. Through a series of statistical analyses, we found that the first 10 minutes after the original tweets' posting was optimal for evaluating their retweetability (See Appendix 2.B). Therefore, we leverage the first 10-minute interval between the original tweets and their first retweet to assess their retweetability.

Multiple Topics in Tweets for Additional Information Processing. When realizing insufficient information in disaster-related messages (Bruns et al. 2012; Mileti and Peek 2000), recipients seek additional information to confirm whether they correctly understand the argument of messages (Lindell and Perry 1987) and to relieve worry and fear about disasters (Bean et al. 2016; Bean et al. 2015; Oh et al. 2013). We contend that the same is true for Twitter. That is, multiple topics in tweets provoke additional information processing for the following reasons: 1) a tweet with 140 characters might not clearly describe multiple topics; 2) a tweet's main topic can be interfered with by other peripheral topics; or 3) all topics might be clearly stated, but recipients might be confused by which topic is the main argument. Any of the above cases requires that recipients obtain more information to have a better understanding of a received tweet. In a similar vein, recipients might consider that the message clarity of a tweet with two topics is less clear than another tweet with one topic. By the same logic, compared to a tweet with one topic, another tweet holding three topics may require twitterers to acquire far more clarifying and confirming information. Hence, twitterers would judge that a tweet with three topics has messages that are less clear than a tweet with one topic. Therefore, in order to find a condition that provokes twitterers to process additional information, we use the number of topics in tweets: a group of tweets with one topic (high message clarity), another group with two topics (intermediate message clarity), and the last group with more than two topics (low message clarity).

Between-Group Estimate. The processing of heuristic information (i.e., twitterers' profile information) comes into play as a means to reach a decision quickly. That is, the twitterer's heuristic information of a tweet can be supplemental to this tweet, in the sense that if he or she positively signifies an author's credibility, his or her tweets would receive benefits from this positive credibility in terms of retweeting. Therefore, a situation where the problem of message clarity exists is considered to test the effects of a twitterer's information with reference to retweet likelihood. Among the already defined three groups of tweets by message clarity from high, intermediate, to low, we conservatively choose two groups of tweets—tweets with intermediate message clarity versus tweets with low message clarity—to reliably observe the interaction of twitterers' information with tweets' message clarity. That is because when received tweets express only one topic, twitterers may or may not seek further information before retweeting. Accordingly, we are unsure whether or not twitterers reliably process additional information before retweeting. However, such information-seeking behavior would be more substantial when twitterers receive tweets with intermediate or low message clarity. From this perspective, we assume that the difference in retweet likelihood between the two groups (i.e., intermediate vs. low message clarity) is dependant upon the variability of twitterers' credibility-cues.

To perform the between-group estimate, we devise the orthogonal contrast codes (see Table 2.2) by following the guidelines of Judd et al. (2011). *OneVSTwoMore* aims to compare the mean difference between the retweet likelihood of tweets with one topic (high message clarity) and those with two (intermediate) or more topics (low). *TwoVSMore* compares the following two groups: tweets with two topics (intermediate) versus tweets with more than two topics (low). As we just discussed, *TwoVSMore* is used to examine the interaction effect of a twitterer's heuristic information on the relationship between his or her tweets' message clarity and the retweet likelihood.

[14] Table 2.2 Contrast Codes for Between-Group Estimate

Topic Quantity / Variables	One	Two	Three or more
OneVSTwoMore	1	-0.5	-0.5
TwoVSMore	0	1	-1

2.4.2 Hypotheses

Provoked by received tweets' unclear messages, additional information processing can hinder the dissemination of information in these tweets to the affected public and to online citizens to purposely relay critical information to others. The recipients of these tweets would strive to address such lack of clarity by searching for other tweets or obtaining relevant information from friends, neighbors, or other media sources. However, under the constraints of a limited amount of time for decision making, they would turn to information that can reduce their cognitive effort, such as twitterers' heuristic information. People generally process heuristic information by leveraging simple decision rubrics, cognitive heuristics, or learned knowledge structure (Liu et al. 2012). In this regard, Metzger et al. (2010) states, "A common strategy employed by Internet information seekers is to minimize cognitive effort and mitigate time pressures through the use of heuristics" (p. 426). Of greater interest here is that Twitter displays heuristic information such as twitterers' number of followers, followees, likes, and so on. Due to tweets' character limit, twitterers might exclude some information about themselves when crafting tweets (Bean et al. 2016), and thus, this heuristic information can be understood to represent the credibility of the tweets' source (i.e., author or twitterer). As a consequence, a twitterer's heuristic information is considered to be an effective way for recipients to lessen their cognitive effort when the credibility information is required for this twitterer's tweets (Todorov et al. 2002; Westerman et al. 2012; Westerman et al. 2014).

Reputation Heuristic. Relying on prior reputation or recognized names, the reputation heuristic implies that people favor reputable sources over unknown ones (Metzger et al. 2010). Therefore, people might believe that online information from reputable sources is more credible than that from lesser known ones. In this perspective, the reputation heuristic is viewed as a credibility-cue (Metzger and Flanagin 2013). A similar credibility-cue found on Twitter is that of a twitterer's number of followers. This number is mainly acquired through posting original tweets as a means to disseminate information or converse with other twitterers (Klotz et al. 2014). Therefore, a twitterer's number of followers indicates the extent to which he or she is popular (Hutto et al. 2013; Kwak et al. 2010), likeable (Liu et al. 2012), and influential (Christakou and Klimis 2013). Westerman et al. (2012) argued that twitterers' number of followers could be used for inferencing an author's credibility, because as the number of followers increases, gatekeeping processes could be strengthened. That is, twitterers have to be responsible for checking the veracity of information and for determining which information has to be released for their followers. In so doing, twitterers are able to keep their current social position (i.e., an opinion leader) on Twitter and to be influential to their followers. Therefore, we believe that a twitterer's number of followers adds his or her tweets positive credibility, contributing to an increase in the retweet likelihood. Moreover, that positivity becomes stronger as his or her tweets' message clarity decreases. The following hypothesis shows our interest:

H1: The positive effect of a twitterer's number of followers on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics).

When twitterer *A* follows twitterer *B*, we denote *A* as *B*'s follower, and *B* as *A*'s followee. The relationship of followees implies the opposite heuristic over followers, and as such, its notion would be opposite that of followers. Twitterers with a large number of followees could be seen as information consumers, not generators, because they are relying on other twitterers to acquire information. Consequently, their author credibility as information generator could weaken. We argue that a twitterer's number of followees adds to his or her tweets negative credibility, contributing to a decrease in his or her tweets' retweet likelihood, and that negativity becomes stronger as his or her tweets' message clarity decreases. The hypothesis with reference to the number of followees is:

H2: The negative effect of a twitterer's number of followees on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics)

Social Presence Heuristic. Most computer-mediated communication (CMC) aims to create systems that are similar to face-to-face communication and offer a richness and variety of interactions for communicators (Hollan and Stornetta 1992). Interpersonal interaction contributes to creating a feeling of being together. Such togetherness is considered to make communication media more effective as a functional alternative to traditional face-to-face interactions (Flaherty et al. 1998; Niinimäki et al. 2012). Because people perceive CMC to be more interpersonal and social, CMC can be used more for social interactions (Flaherty et al. 1998). Hence, it seems that the more social interaction cues a medium conveys, the closer the medium will be in fostering a sense of togetherness as face-to-face communication. In this viewpoint, Sundar (2008) defined such togetherness or social presence—"the user is communicating with a social entity rather than an inanimate object," (p. 84)—and believed that this social connectivity helps communication interaction between participants. Hence, as an author shows an active interaction online, communication participants consider him or her to be a good facilitator for communication, positively affecting his or her author credibility.

We consider that twitterers' total number of posted tweets since joining Twitter, called status, imply an interaction heuristic in the sense that as a twitterer posts tweets more frequently, other twitterers may feel a higher sense of his or her social presence. Therefore, we contend that a twitterer's total number of posted tweets adds to his or her tweets' positive credibility, contributing to a decrease in his or her tweets' retweet likelihood and that this positivity becomes stronger as his or her tweets' message clarity decreases. The following hypothesis represents our interest:

H3: The positive effect of a twitterer's total number of posted tweets on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics).

In a similar vein, twitterers could also get a sense of others' social presence by looking at their length of affiliation on Twitter. Sundar (2008) referred to this history of affiliation as loyalty. The longer the term of affiliation that a twitterer has, the higher the sense of being together that the twitterers would enjoy. In fact, Twitter provides twitterers' join date, and therefore twitterers may gauge one's loyalty on Twitter from one's join date. In this view, we assume that a twitterer's length of affiliation adds to his or her tweets' positive credibility, contributing to a decrease in his or her tweets' retweet likelihood and that this positivity becomes stronger as his or her tweets' message clarity decreases. The following hypothesis represents our interest:

H4: The positive effect of a twitterer's length of affiliation on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics).

Recency Heuristic. The recency of postings is one virtue of social media (Sutton et al. 2014b). As disaster events are largely unexpected and dynamic, disaster-related information becomes quickly obsolete and inaccurate (Wilensky 2014). In that regard, Sundar (2008) considered the timeliness of information an important cue of credibility. Metzger (2007) also argued that the recency of information is one factor for assessing the credibility of online information. In fact, Westerman et al. (2014) empirically showed that frequent updates indirectly increased the source credibility. Although Twitter does not automatically generate information cues with regard to twitterers' recency of postings, we include this heuristic cue into our research hypothesis by aggregating individual twitterers' tweets about the current disaster incident. Although this heuristic cue seems to be similar to the social presence heuristic in the sense that the tweeting frequency is of interest, it is different in terms of reflecting the recency of tweeting frequency. Taken together, we believe that a twitterer's tweeting frequency about the current events adds to his or her tweets' positive credibility, contributing to an increase in its retweet likelihood, and this positivity becomes stronger as his or her tweets' message clarity decreases. Therefore, the hypotheses we are interested in are as follows:

H5: The positive effect of a twitterer's tweeting frequency about the current events on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics).

Endorsement Heuristic. "People are inclined to perceive information and sources as credible if others do so also" (Metzger et al. 2010, p. 427). It is likely that individuals tend to consider something correct or truthful when many other people also believe it to be correct or truthful. Chaiken and Eagly (1989) termed such a phenomenon as a liking-agreement heuristic in which "people agree with people they like" or "people I like usually have correct opinions on issues" (p. 4). In a similar vein, Sundar (2008) introduced a bandwagon heuristic to influence one's credibility, by defining this heuristic as either the endorsement of a group or the reputation of a source. Thus, we argue that the credibility of a twitterer would be positively associated with the extent to which his or her tweets have been "liked" by other twitterers. As a matter of fact, Twitter provides a function, liking a tweet, for twitterers to express their appreciation, agreement, or acknowledgment for tweets (Twitter 2016; Warzel 2014). Liked tweets are summed up for each twitterer to represent his or her endorsement by others. Therefore, we consider that a twitterer's number of likes adds to his or her tweets' positive credibility and thus contributes to increasing his or her tweets' retweet likelihood. This positivity becomes stronger as his or her tweets' message clarity decreases. Thus, the last hypothesis is tendered with respect to the endorsement cue of likes.

H6: The positive effect of a twitterer's number of likes on his or her tweets' retweet likelihood would depend on the message clarity of tweets, such that the effect would become stronger for tweets with low message clarity (more than two topics) than for those with intermediate message clarity (two topics).

2.5 Data and Methods

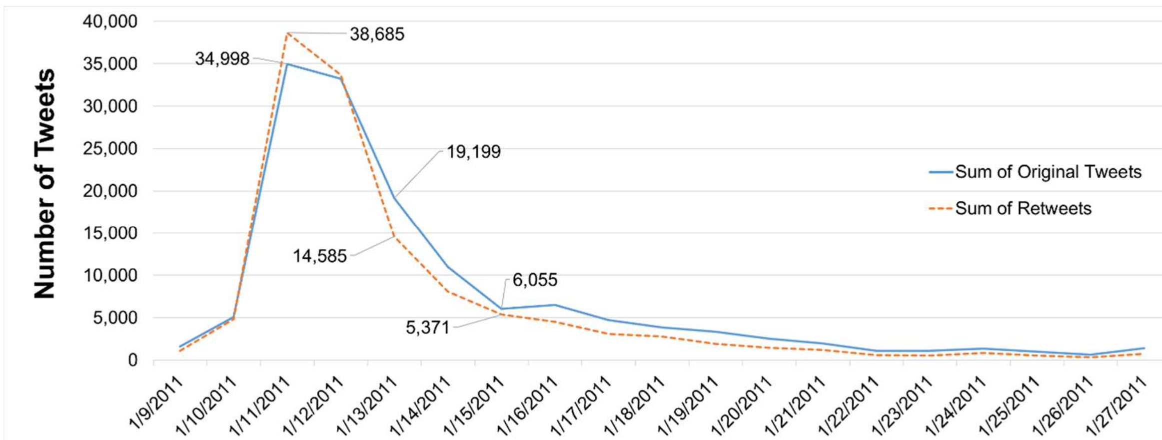
2.5.1 Data

2013 Colorado floods. In 2013, 15 to 20 inches of rain poured into the northern parts of Colorado, such as Boulder, Colorado Springs, and Fort Collins. Damage was substantial (Connor et al. 2013; Gochis et al. 2014):14 counties declared disaster emergency; 11,000 residents evacuated; 1,750 residents and more than 300 animals were rescued by the Colorado National Guard and U.S. Army; more than 20,000 homes were damaged; the State's infrastructure was severely affected. Most heartfelt is that fact that eight people were dead and five were missing (Gochis et al. 2014). Right after the initial warnings by FEMA, Coloradoans who lived in these affected areas and altruistic citizens online whose intention was to relay critical information to others started producing, sharing, and disseminating diverse flood-related information on Twitter.

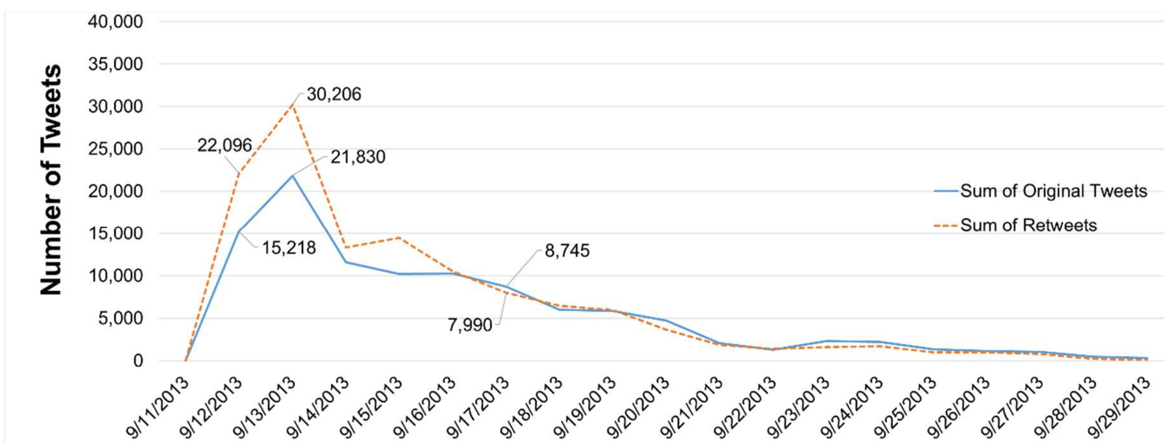
During the floods, Project EPIC, hosted by the Department of Computer Science at the University of Colorado Boulder, collected flood-related tweets and their retweets in near real-time by leveraging its data analytic infrastructure (Anderson and Schram 2011). By systematically identifying keywords, hashtags, and twitterers, EPIC's research group searched and stored relevant tweets and twitterers in their infrastructure (Dashti et al. 2014a) (see Table 2.3). Figure 2.2 shows the trends of the original tweets and their retweets from September 11 to September 29, 2013. For the empirical analysis of this study, we focus on tweets and their retweets generated during the two weeks of September 12 and September 25, which accounts for more than 95% of the total tweets (See Table 2.4).

[15] Table 2.3 Keywords and Hashtags		
2013 Colorado floods		
Date	**Keywords**	**Hashtags**
September 11th	boulderflood, cowx, nwsboulder	
September 12th	coflood, cofloods, coflodding, cuboulder flood	#boulder, #cccf

September 15th	Boulderfloods	
September 19th	flood gas, flood infrastructure	#cofloodrelief
September 20th		#coloradostrong
2011 Queensland floods		
(N/A)	Queensland, qldfloods, qldflood	#qldfloods, #thebigwet



< 2011 Queensland floods >



< 2013 Colorado floods >

[6] Figure 2.2 Trends of the Original Tweets and Their Retweets

2011 Queensland floods. In early 2011, a series of floods hit Australia's central and southern parts, including Queensland and Brisbane, and the floods were most intensified during January 10 and 16

(Shaw et al. 2013). Flood emergencies were declared for half of the Queensland territory, a size similar to that of France and Germany combined (Bruns et al. 2012). Queensland was substantially damaged: over 200,000 residents across 90 towns were affected; around 30,000 properties received damage; and 38 people were found to be dead (Davies 2013). From the beginning of this disaster, social media, such as Facebook and Twitter, were important means of communication. Particularly, Twitter was used by the public and online citizens to rapidly disseminate and amplify first-hand footage of emergency situations to others (Bruns et al. 2012).

GNIP[11], a subsidiary of Twitter, provided data about the Queensland floods. By following Project EPIC's data collection processes, the data scientists of GNIP identified keywords, hashtags, and twitterers to retrieve tweets and their related information. Figure 2.2 shows the trend of tweets collected in between January 1 and January 27, 2011. Similar to the 2013 Colorado floods, two weeks of tweet information was used for our statistical analysis (See Table 2.4).

[16] Table 2.4 Descriptive Statistics of Two Flood Incidents

Items \ Cases	2011 Queensland	2013 Colorado
Period of Data Collection	January 8 ~ 21, 2011	September 12 ~ 25, 2013
Total Tweets	109,456	102,426
Total Retweets	120,082	122,276
Unique Twitterers	33,565	77,774

2.5.2 Methods

In order to derive variables from tweets for statistical analysis, we leveraged natural language processing (NLP) techniques, such as part-of-speech tagging and topic modeling. For statistically parsing tweets, we used *TweetNLP*'s programming library (Owoputi et al. 2013) to extract each tweet's words and their part-of-speech tags, URLs, hashtags, and so on. For identifying topics in tweets, we

[11] GNIP (https://gnip.com) is a subsidiary of Twitter that provides an enterprise API platform.

utilized Machine Learning for LanguagE Toolkit (MALLET), a Latent Dirichlet Allocation (LDA) model (McCallum 2002) to automatically discover topics in a collection of documents. The LDA method defines a topic by words' distribution in documents, Therefore, multiple topics can occur in a single document in terms of different distributions of words per topic (Blei 2012). With a caveat in mind that tweets' lengths are too short to build topic models (Cataldi and Aufaure 2015; Wang et al. 2007), we add meaningful *n*-gram phrases to the original tweets as a means of compensating for the short length. For example, we extracted *n*-gram noun phrases based on part-of-speech tags as follows: "heavy rain" and "relief fundraiser" as bi-gram; "flood recovery efforts" and "road damage photos" as tri-gram; and "state emergency operation center" and "flood information resources list" as quad-gram. As a result, noun phrases up to 6-gram are used as inputs for topic modeling, along with hashtags that are known to summarize tweets' topic(s) (Boyd et al. 2010; Bruns and Stieglitz 2012; Laniado and Mika 2010; Ma et al. 2013; Yang et al. 2012). We do not include Twitter URLs that are comprised of randomly chosen characters and numbers (i.e., http://t.co/ntqdy1o7rw), however, because they have none of the topic information needed for topic modeling. Table 2.5 lists the most frequent, top 5 *n*-gram phrases.

[17] Table 2.5 Top 5 n-Gram Keywords

Rank	2011 Queensland floods	2013 Colorado floods
1	flood relief appeal	colorado flood
2	flood victims	flood victim
3	flood appeal	colorado relief
4	anna bligh	boulder creek
5	brisbane river	higher ground

Determining the Number of Topics. Topic-modeling groups similar documents together based upon each document's topic similarity, and thus, it is one type of clustering analysis that requires the expected number of clusters (or groups) as an input parameter (Blei 2012). Accordingly, determination of the number of topics for the LDA is needed in order to cluster tweets by their topics. To fulfill this task, we use the measure of perplexity to evaluate each topic model's generalizability (Blei et al. 2003), where M refers to the number of documents in the testing dataset, w_d refers to the words in document d, and N_d refers to the number of words in document d.

$$\text{Perplexity}\ (\mathrm{D_{test}}) = \exp\left\{\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right\}$$

Perplexity and generalizability are inversely associated, in that the lower the perplexity a topic model has, the higher the generalizability it maintains (Blei et al. 2003). In general, a topic-model with high generalizability is preferred. Hence, we designed the following procedures and proceeded thus to find a preferred topic model per Twitter dataset: 1) 199 topic-models were generated by the number of topics ranging from 2 to 200; 2) each model's perplexity was calculated; and lastly, 3) the cumulative sum (CUSUM) analysis (Ellaway 1978) was applied to the perplexity of the generated topic-models to find a favoured model whose perplexity significantly lowers and eventually becomes stable, signifying that additional topics do not substantially provide contribution to further topic-models' generalizability. Consequently, we found that the topic model with 72 topics for the 2011 Queensland floods and the model with 57 topics for the 2013 Colorado floods were preferred for our research. Appendix 2.C shows the relationship between the topic models having varying topic numbers and their perplexity values.

Variables. The dependent variable for this study is a binary response, which measures whether or not the original tweets are retweeted within the first 10 minutes after posting. We coded a positive response as 1 when a tweet was retweeted in this interval, while giving 0 as a negative response when a tweet was not retweeted. Accordingly, logistic regression was employed to examine the relationships between retweet likelihood and the exploratory variables of interest.

Except for the age of the Twitter accounts in years (the length of affiliation), the exploratory variables representing twitterers' heuristic information—the numbers of followers, followees, likes (the total number of one's liked or favorited tweets), status (the total number of one's tweets), and tweets about current events—were log-transformed for the following reasons: first, the log-transformation stabilizes data's variability and thus, could enhance statistical inference (Judd et al. 2011; Mosteller and Tukey 1977); second, the log-transformation can make skewed data conform to normality for a better model fit (Meaney et al. 2007). Tweets' numbers of URLs, words, and hashtags were used as control variables of the research model, such that the message clarity of tweets was estimated in association with retweet likelihood while accounting for individual tweets' total length[12]. Along with the above controls, we also included one additional control variable to indicate whether tweets included mentions. Spiro et al. (2012) found that the inclusion of Twitter mentions delayed retweet speed, which, in turn, negatively influenced a tweet's retweet likelihood compared to tweets that did not mention other twitterers. Similarly, Suh et al. (2010b) pointed out a negative effect of mentions on retweeting, but the effect was marginal. The dependent, control, and exploratory variables are summarized in Table 2.6 with their descriptive statistics. A correlation matrix can also be found among the variables in Table 2.7. The test of variance inflation factor (VIF) indicated that the proposed empirical model did not have significant signs of multicollinearity problem (Max of 3.37 and Mean of 1.61 for the Queensland floods; Max of 3.42 and Mean of 1.72 for the Colorado floods) (see Appendix 2.E). Due to the existence of the heteroscedasticity of variance, logistic regression was estimated using the Huber/White/sandwich estimator of variance (Huber 1967; White 1980). Figure 2.3 depicts a statistical expression for the empirical model of this study.

[12] As each tweet has a different character length, we control the total length of tweets to better estimate the effect of message clarity on retweeting.

[18] Table 2.6 Description of Variables							
Variables	**Explanation** / **Cases**	**2011 Queensland**			**2013 Colorado**		
		Mean	**S.D.**	**Range**	**Mean**	**S.D.**	**Range**
Dependent Variable							
Retweet_YN_10m$_i$	Whether or not tweet *i* is retweeted within the first 10 minutes after it is posted – ‘1’ for ‘Retweeted’ and ‘0’ for ‘Not retweeted’						
Systematic Information – ***Message Clarity (Entropy)***							
- Between-Group Estimate of Message Clarity							
OneVSTwoMore	A contrast code for group comparison between tweets with only one topic (high message clarity) and those with two or more topics (intermediate or low message clarity) – *High message clarity vs. Intermediate or Low message clarity*						
TwoVSMore	A contrast code for group comparison between tweets with two topics (intermediate message clarity) and those with three or more topics (low message clarity) – *Intermediate message clarity vs. Low or Low message clarity*						
Heuristic Information – ***Twitterer Credibility***							
Ln(Followers$_{i,t}$)	The log-transformed number of followers of tweet *i*'s author between his/her join date and the date of tweet *i's* posting $Ln\left(\sum_{\substack{Join\ date\ of\\ the\ author\ of\ tweet\ i\\ \le t \le\\ tweet\ i's\ posting\ date}} Follower_{i,t}\right)$	5.465	1.799	0~15.1	6.106	2.306	0~16.4
Ln(Followees$_{i,t}$)	The log-transformed number of followees of tweet *i*'s author between his/her join date and the date of tweet *i's* posting	5.37	1.61	0~12.1	5.834	1.948	0~12.7

	$Ln\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s posting date}}} Followee_{i,t}\right)$						
Ln($Status_{i,t}$)	The log-transformed number of tweets of tweet *i*'s author between his/her join date and the date of tweet *i's* posting $Ln\left(\underbrace{\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s posting date}}} Tweet_{i,t}}_{Status}\right)$	7.443	1.983	0~12.7	8.140	2.234	0~14.0
$Years_{i,t}$	The age of the Twitter accounts in years after each twitterer's account creation at time (t) of tweet *i's* posting $\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \leq t \leq \\ \text{tweet } i\text{'s posting date}}} Year_{i,t}\right)$	1.931	0.849	0~5	2.730	1.704	0~7
Ln($Tweets_{i,t}$)	The log-transformed number of tweets about current events of the twitterer of tweet *i* between the start date of an incident and the date of tweet *i's* posting	2.684	1.588	0.69 ~6.6	2.263	1.566	0.69 ~6.312

	$Ln\left(\sum_{\substack{Start\ of\ a\ flood\ incident \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Tweet_{i,t}\right)$						
Ln($Likes_{i,t}$)	The log-transformed number of likes of tweet *i*'s author between his/her join date and the date of tweet *i's* posting $Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Like_{i,t}\right)$	1.783	1.945	0~9.32	3.304	2.602	0~13.6
Dual Processing – *Message Clarity* × *URLs/Emoticons*							
Ln($Followers_{i,t}$) x TwoVSMore	Interaction effect between *Followers* x *TwoVSMore* to examine the degree to which the effect of *Followers* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						
Ln($Followees_{i,t}$) x TwoVSMore	Interaction effect between *Followees* x *TwoVSMore* to examine the degree to which the effect of *Followees* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						
Ln($Status_{i,t}$) x TwoVSMore	Interaction effect between *Status* x *TwoVSMore* to examine the degree to which the effect of *Status* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						
$Years_{i,t}$ x TwoVSMore	Interaction effect between *Years* x *TwoVSMore* to examine the degree to which the effect of *Years* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						
Ln($Tweets_{i,t}$) x TwoVSMore	Interaction effect between *Tweets* x *TwoVSMore* to examine the degree to which the effect of *Tweets* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						
Ln($Likes_{i,t}$) x TwoVSMore	Interaction effect between *Likes* x *TwoVSMore* to examine the degree to which the effect of *Likes* on the retweet likelihood would depend on the extent of the message clarity of tweets (Intermediate vs. Low)						

Control Variables							
$Words_i$	The total number of words in tweet *i*	9.54	3.99	0-24	8.61	3.84	0-24
$Hashtags_i$	The total number of hashtags in tweet *i*	1.26	0.891	0-13	1.27	1.21	0-15
$URLs_i$	The number of URLs in tweet *i*	0.462	0.553	0-5	0.667	0.535	0-4
$Mention\_YN_i$	Whether tweet *i* contains other twitterers' name – 1 for 'Yes' and -1 for 'No'						

[19] Table 2.7 Correlation Matrix of the Research Model

Queensland

	1	2	3	4	5	6	7	8	9	10	11
Retweet_YN_10m	1										
Entropy (Linear)	0.141***	1									
URLs	-0.0369***	-0.527***	1								
Emoticons	0.0196***	-0.118***	-0.0325***	1							
Words	0.0919***	-0.00966**	0.00600*	-0.173***	1						
Hashtags	0.0168***	0.130***	-0.0277***	-0.316***	-0.0675***	1					
Ln(Followers)	-0.0634***	0.105***	-0.0315***	-0.0965***	0.0339***	0.0599***	1				
Ln(Followees)	0.228***	0.0604***	-0.0206***	-0.00775*	0.0665***	0.0557***	0.0556***	1			
Ln(Likes)	0.107***	0.0261***	-0.0105***	-0.0235***	0.0855***	-0.0465***	0.148***	0.680***	1		
Ln(Status)	0.112***	0.0507***	-0.0127***	-0.0207***	0.0491***	0.0243***	0.0980***	0.742***	0.488***	1	
Mention_YN	0.0790***	0.000956	-0.00134	-0.0159***	0.0265***	-0.0149***	0.0232***	0.389***	0.323***	0.383***	1

Colorado

Retweet_YN_10m	1										
Entropy (Linear)	0.123***	1									
URLs	-0.0611***	-0.696***	1								
Emoticons	0.0655***	-0.0198***	-0.0668***	1							
Words	0.136***	-0.0551***	0.0343***	-0.207***	1						
Hashtags	-0.0562***	0.114***	-0.0511***	-0.240***	-0.106***	1					
Ln(Followers)	0.0555***	0.0532***	-0.0297***	-0.00794*	0.0822***	-0.0284***	1				
Ln(Followees)	0.332***	0.0897***	-0.0615***	0.0846***	0.0517***	0.0818***	0.140***	1			
Ln(Likes)	0.207***	0.0351***	-0.0285***	0.0439***	0.0900***	0.0226***	0.185***	0.737***	1		

Ln(Status)	0.101***	0.124***	-0.0792***	0.0948***	-0.0963***	0.155***	0.0517***	0.670***	0.498***	1	
Mention_YN	0.181***	-0.0156***	0.00506	0.0319***	0.0778***	-0.0212***	0.119***	0.516***	0.444***	0.345***	1

* p<0.05, ** p<0.01, *** p<0.001

$$
\begin{aligned}
Retweet\_YN\_10m_i = {} & \beta_0 + \underbrace{\beta_1 \cdot OneVSTwoMore_i + \beta_2 \cdot TwoVSMore_i}_{Systematic\ Information-Message\ Clarity} \\
& + \underbrace{\beta_3 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Follower_{i,t}\right) + \beta_4 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Followee_{i,t}\right) + \beta_5 \cdot Ln\left(\underbrace{\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Tweet_{i,t}}_{Status}\right)}_{Heuristic\ Information-Credibility} \\
& + \underbrace{\beta_6 \cdot \left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Year_{i,t}\right) + \beta_7 \cdot Ln\left(\sum_{\substack{Start\ of\ a\ flood\ incident \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Tweet_{i,t}\right) + \beta_8 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Like_{i,t}\right)}_{Heuristic\ Information-Credibility\ cont'd} \\
& + \underbrace{\beta_9 \cdot TwoVSMore_i \cdot \left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Follower_{i,t}\right) + \beta_{10} \cdot TwoVSMore_i \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Followee_{i,t}\right) + \beta_{11} \cdot TwoVSMore_i \cdot \left(\underbrace{\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Tweet_{i,t}}_{Status}\right)}_{Dual\ Processing} \\
& + \underbrace{\beta_{12} \cdot TwoVSMore_i \cdot \left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Year_{i,t}\right) + \beta_{13} \cdot TwoVSMore_i \cdot \left(\sum_{\substack{Start\ of\ a\ flood\ incident \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Tweet_{i,t}\right) + \beta_{14} \cdot TwoVSMore_i \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \leq t \leq \\ tweet\ i's\ posting\ date}} Like_{i,t}\right)}_{Dual\ Processing\ cont'd} \\
& + \underbrace{\underbrace{\beta_{15} \cdot Words_i + \beta_{16} \cdot Hashatgs_i}_{Tweet\ Length} + \beta_{17} \cdot URLs_i + \beta_{18} \cdot Mention\_YN_i}_{Control\ Variables} \\
& + \varepsilon_i
\end{aligned}
$$

[7] Figure 2.3 Statistical Expression of the Research Model

2.6 Results

Table 2.8 shows the results of the robust logistic regression of the two Twitter datasets—the 2011 Queensland and the 2013 Colorado floods. Strong evidence for the message clarity of tweets was found in both Twitter cases (Queensland – Wald $Chi^2=2078.03$, $df=2$, $p<0.000$; Colorado – Wald $Chi^2=1325.24$, $df=2$, $p<0.000$). The details are as follows: in the Queensland floods, while parsing out all other variables' effects in the empirical model, tweets with high message clarity (one topic) were 1.5 times more likely to be retweeted than tweets with intermediate or low message clarity (more than one topic) within the first 10 minutes after posting (coefficient=0.811, Wald $Chi^2=1755.24$, $p<0.0000$). Similarly, in the Colorado floods, tweets with high message clarity were 1.515 times more likely to be retweeted than those with intermediate or low message clarity (coefficient=0.831, Wald $Chi^2=1016.32$, $p<0.0000$). We also found in the Queensland Twitter data that tweets with intermediate message clarity (two topics) were 1.289 times more likely to be retweeted than tweets with low message clarity (more than two topics) (coefficient=0.509, Wald $Chi^2=330.17$, $p<0.0000$). Also, the difference in the odds ratio between tweets with intermediate message clarity and those with low message clarity was 1.384 in the Colorado case (coefficient=0.651, Wald $Chi^2=249.26$, $p<0.0000$). With the significant effect of tweets' message clarity on the retweet likelihood, we, therefore, assumed that a decrease in tweets' message clarity could motivate twitterers to search for verifying and confirming information. To examine such a relationship, a set of interaction hypotheses were used to test how the heuristic information of twitterers interacted with the message clarity of tweets in conjunction with the retweet probability.

[20] Table 2.8 Results of the Robust Logistic Regression

Variables \ Cases	2011 Queensland			2013 Colorado			Hypothesis
	Coefficient (Robust Err.)	Wald Chi-Square	Sig. Level	Coefficient (Robust Err.)	Wald Chi-Square	Sig. Level	
Heuristic Information – ***Twitterer Credibility***							
	Wald Chi-Square=4740.74***, df=6			*Wald Chi*-Square=11498.20***, df=6			
Ln(Followers$_{i,t}$)	0.501*** (0.00910)	3031.19	0.0000	0.574*** (0.00763)	5646.71	0.0000	(N/A)
Ln(Followees$_{i,t}$)	-0.0841*** (0.00705)	142.22	0.0000	-0.0661*** (0.00771)	73.63	0.0000	
Ln(Status$_{i,t}$)	-0.185*** (0.00781)	560.68	0.0000	-0.381*** (0.00704)	2926.69	0.0000	
Years$_{i,t}$	0.0798*** (0.0125)	40.70	0.0000	0.0545*** (0.00688)	62.65	0.0000	
Ln(Tweets$_{i,t}$)	0.116*** (0.00665)	305.03	0.0000	0.288*** (0.00639)	2032.29	0.0000	
Ln(Likes$_{i,t}$)	0.0187*** (0.00558)	11.20	0.0008	0.123*** (0.00456)	727.67	0.0000	
Systematic Information – ***Message Clarity***							
	Wald Chi-Square=2078.03***, df=2			*Wald Chi*-Square=1325.24***, df=2			
OneVSTwoMore$_i$	0.811*** (0.0193)	1755.24	0.000	0.831*** (0.0261)	1016.32	0.0000	
TwoVSMore$_i$	0.509*** (0.0280)	330.17	0.000	0.651*** (0.0412)	249.26	0.0000	
Dual Processing (or Bias) – ***Twitterer Credibility*** **and** ***Message Clarity***							**Hypotheses**
	Wald Chi-Square=38.94***, df=6			*Wald Chi*-Square=64.26***, df=6			
Ln(Followers$_{i,t}$) × TwoVSMore$_i$	-0.0363* (0.0172)	4.48	0.034	-0.0704*** (0.0166)	17.98	0.0000	H1: Supported
Ln(Followees$_{i,t}$) × TwoVSMore$_i$	-0.0109 (0.0129)	0.72	0.3973	-0.00315 (0.0163)	0.04	0.8467	H2: Rejected

$Ln(Status_{i,t})$ × $TwoVSMore_i$	0.0541*** (0.0141)	14.64	0.0001	0.0888*** (0.0148)	35.54	0.0000	H3: Rejected (Opposite Direction)
$Years_{i,t}$ × $TwoVSMore_i$	-0.0761*** (0.0221)	11.92	0.0006	-0.0310* (0.0137)	5.11	0.0238	H4: Supported
$Ln(Tweets_{i,t})$ × $TwoVSMore_i$	-0.0384** (0.0119)	10.42	0.0012	-0.0304* (0.0125)	5.88	0.0153	H5: Supported
$Ln(Likes_{i,t})$ × $TwoVSMore_i$	-0.00965 (0.00991)	0.95	0.3299	-0.0368*** (0.00929)	15.68	0.0001	H6: Partially Supported
Control Variables							
	Wald Chi-Square=1542.03***, df=4			*Wald Chi*-Square=1112.33***, df=4			
$Words_i$	0.0392*** (0.00239)	269.63	0.0000	0.0512*** (0.00260)	388.74	0.0000	(N/A)
$Hashtags_i$	0.256*** (0.00933)	752.58	0.0000	0.181*** (0.00745)	592.13	0.0000	
$URLs_i$	0.0136 (0.0170)	0.64	0.4239	-0.185*** (0.0181)	104.55	0.0000	
$Mention\_YN_i$	-0.248*** (0.00974)	646.93	0.0000	-0.0584*** (0.00951)	37.70	0.0000	
Constant	-2.395*** (0.0194)	-	0.0000	-2.432*** (0.0269)	-	0.0000	
Model Summary							
PCT. Predicted Correctly	85.02%			82.98%			(N/A)
Predicted Probability 1	10.81%			30.99%			
Predicted Probability 0	98.97%			95.97%			
Log Likelihood	11004.357***			22054.121***			
Wald Chi-Square	9124.25***			16551.18***			
McFadden's Pseudo R^2	0.115			0.215			
n	109,456			102,426			

[1] All predictors are mean centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

For the hypotheses that include higher order terms, we centered all continuous variables from their means to alleviate multicollinearity (Aiken et al. 1991) and to have better interpretations (Judd et al. 2011). On the whole, interesting empirical evidence supporting our hypotheses was found in both flood incidents (Queensland – Wald Chi^2=38.94, df=6, p<0.0000; Colorado – Wald Chi^2=64.26, df=6, p<0.0000). Of the six interaction hypotheses, four, including one partial support, were found statistically significant. Their interaction plots are shown in Figure 2.4. The details of Hypothesis 1 through 6 are as follows.

Hypothesis 1 is strongly supported by both Twitter datasets. In the Queensland floods, the positive effect of the number of followers (1.65 odds ratio[13]) on the retweet likelihood became significantly different for the different levels of tweets' message clarity. That is, the positive effect was significantly stronger for tweets with low message clarity than for those with intermediate message clarity (coefficient=-0.0363, Intermediate message clarity=1.62 odds ratio[14] vs. Low message clarity=1.68, Wald Chi^2=4.48, p<0.0000). That is to say, for every 1% increase in a twitterer's number of followers, the odds of his or her tweets being retweeted increased by a factor of 1.65, and those odds went up by a factor of 1.68 when his or her tweet's message clarity was low, but those odds decreased by a factor of 1.62 when his or her tweet's message clarity was intermediate. The difference between the odds of 1.68 and 1.62 is significantly different by the Wald Chi-Square test. In the Colorado floods, we also found that the relationship between the number of followers and the retweet likelihood was significantly interrelated with the different levels of the message clarity (the odds ratio of 1.775, coefficient=-0.0704, Intermediate message clarity=1.714 odds ratio vs. Low message clarity=1.84, Wald Chi^2=17.98, p<0.0000). Unlike the number of followers, the number of followees with regard to the retweet

[13] The odds ratio of the number of followers is $e^{0.501}$. See Table 2.8 for the specific number(s)
[14] The odds ratio of the interaction between the number of followers and intermediate message clarity is $e^{0.501-(0.0363/2)*1}$. Please see Table 2.8 for the specific number(s)

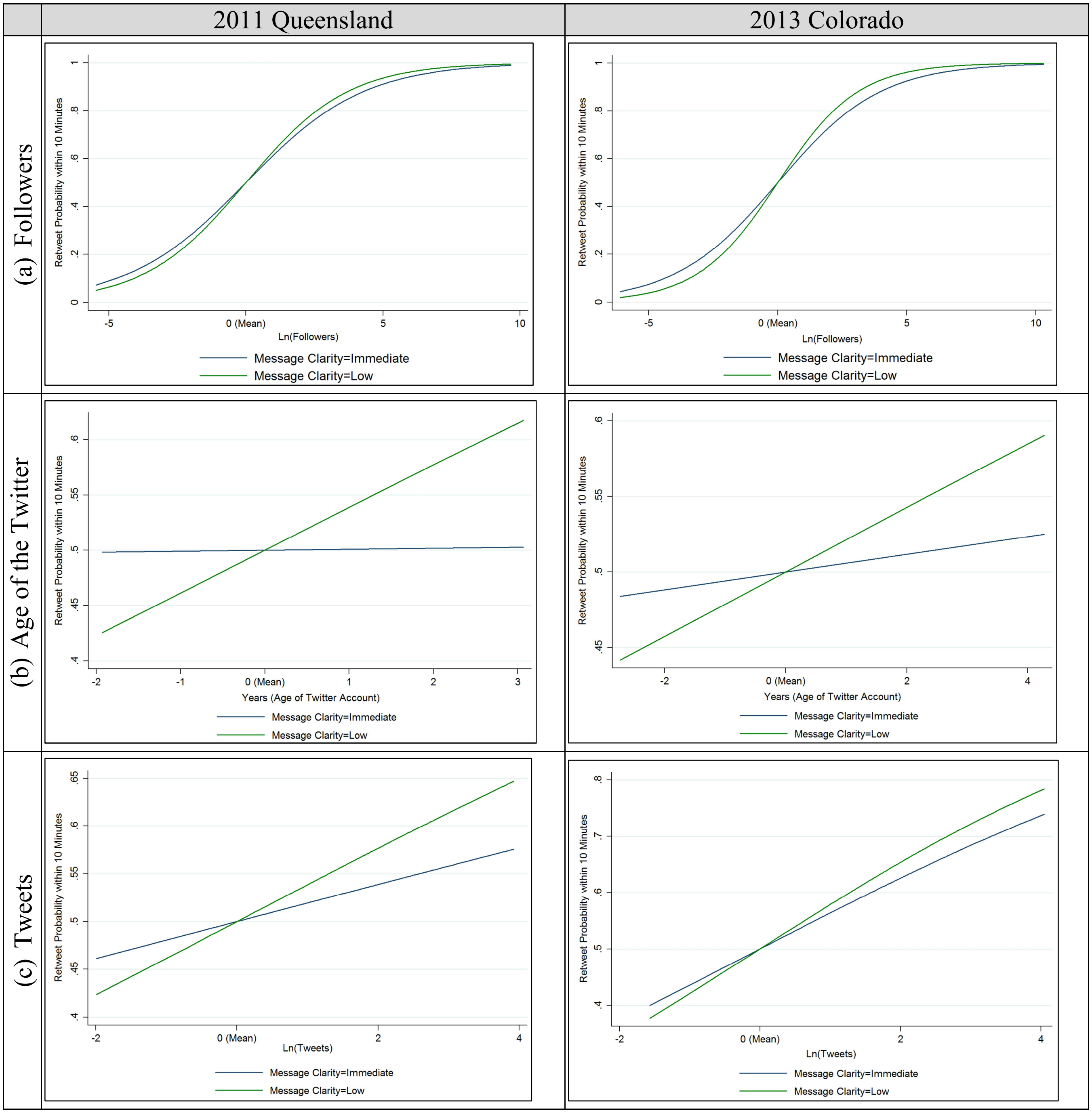
likelihood did not significantly interact with the different levels of the message clarity in both datasets. Thus, we reject Hypothesis 2.

Relating to the total number of posted tweets (or status) and the retweet likelihood, we found a statistically significant relationship in both datasets, but the direction of the relationship was opposite to what we expected. That means, the total number of posted tweets was negatively related to the retweet likelihood (0.831 odds ratio and 0.683 respectively for the Queensland and the Colorado floods), and that negative relationship considerably weakened when message clarity was intermediate as compared to when it was low (Queensland – coefficient=0.0541, Intermediate message clarity=0.854 odds ratio vs. Low message clarity=0.809, Wald Chi^2=14.64, p=0.0001; Colorado – coefficient=0.0888, Intermediate message clarity=0.714 odds ratio vs. Low message clarity=0.654, Wald Chi^2=17.98, p<0.0000). For this reason, we have to conclude that Hypothesis 3 is not supported, even though the opposite effect of the total number of posted tweets is statistically significant.

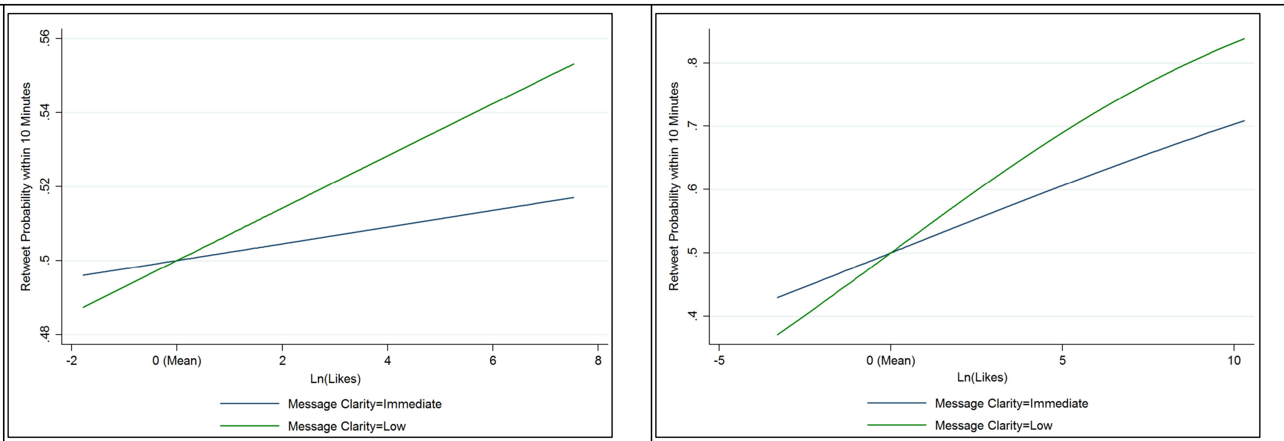
Hypothesis 4 turns out to be empirically significant. Statistically speaking, one additional year of affiliation increased the retweet odds ratio of 1.083 for the Queensland and the retweet odds of 1.056 for the Colorado floods, and each effect became significantly stronger for tweets with low message clarity than for those with intermediate message clarity in both datasets (Queensland – coefficient=-0.0761, Intermediate message clarity=1.043 odds ratio vs. Low message clarity=1.125, Wald Chi^2=11.92, p=0.0006; Colorado – coefficient=-0.0310, Intermediate message clarity=1.0397 odds ratio vs. Low message clarity=1.0725, Wald Chi^2=5.11, p=0.0238).

Our Twitter datasets consistently support Hypothesis 5. That is to say, the positive relationship between the number of tweets about current events and the retweet likelihood (1.123 odd ratio for the Queensland and 1.334 for the Colorado floods) became significantly stronger for tweets with low message clarity than for those with intermediate message clarity (Queensland – coefficient=-0.0384, Intermediate message clarity=1.1 odds ratio vs. Low message clarity=1.144, Wald Chi^2=10.42, p=0.0012; Colorado – coefficient=-0.0304, Intermediate message clarity=1.314 odds ratio vs. Low message clarity=1.354, Wald Chi^2=5.88, p=0.0153). For example, the tweets for the Queensland floods showed that as a twitterer's number of tweets about the current incident increased by 1%, the odds of his or her tweets being retweeted were raised by a factor of 1.123, and that odds went up by a factor of 1.144 when his or her tweets' message clarity was low, Moreover, those odds went down by a factor of 1.11 when his or her tweets' message clarity was intermediate.

Finally, we found partial support for Hypothesis 6, given that although no significant interaction effect of the number of likes was found in the tweets for the Queensland floods, the other Twitter dataset showed its significant interaction effect on the relationship between tweets' message clarity and the retweet likelihood. That is, we observed in the tweets about the 2013 Colorado floods that the positive effect of the number of likes (1.13 odds ratio) on the retweet likelihood was reliably higher for tweets with low message clarity than for those with intermediate message clarity (coefficient=-0.0368, Intermediate message clarity=1.11 odds ratio vs. Low message clarity=1.151, Wald Chi^2=15.68, p=0.001). Namely, an additional 1% of a twitterer's number of likes increased the odds of his or her tweets being retweeted by a factor of 1.13 on average, and those odds increased by a factor of 1.151 when his or her tweet's message clarity was low. Moreover, those odds decreased by a factor of 1.11 when his or her tweet's message clarity was intermediate.


2011 Queensland
2013 Colorado
(a) Followers
(b) Age of the Twitter
(c) Tweets
Retweet Probability within 10 Minutes
Ln(Followers)
Years (Age of Twitter Account)
Ln(Tweets)
0 (Mean)
Message Clarity=Immediate
Message Clarity=Low

(d) Likes[15]



[8] Figure 2.4 Interaction Plots Between Credibility-Cues and Message Clarity

2.7 Discussion

This study empirically demonstrated how twitterers' heuristic information influenced the retweet likelihood when tweets' messages are not clear enough. We examined retweet likelihood as a function of message clarity in tweets and in twitterers' cued information—followers, followees, likes, status, tweets, and the age of the Twitter accounts. In this context, our research hypotheses centered around the following question: how does twitterers' information as a source-credibility cue affect their tweets' retweeting during times of disaster? To better reflect communication practices in disasters where critical information must be consumed in a timely manner before being obsolete, we used the first 10 minutes after the original tweets' posting in order to evaluate the original tweets' retweetability rather than utilizing much longer intervals of a few hours, days, or weeks. By adapting the theoretical guidance of the HSM, we established the empirical model to investigate the interaction effect of twitterers' heuristic information as source-credibility cues on the relationship between the message clarity of tweets and the retweet likelihood.

[15] The effect of 'likes' was significant in the tweets for the 2013 Colorado floods, but not in the other Twitter dataset.

We found that as a tweet's message clarity decreased, its retweet likelihood lowered as well. The Queensland Twitter dataset showed that within the first 10 minutes after posting, the retweet likelihood of tweets with high message clarity (1 topic) was 12.03%,[16] while the retweet probability of tweets with intermediate message clarity (2 topics) was 10.52%,[17] and the retweet likelihood of tweets with low message clarity (more than 2 topics) was 6.6%.[18] Similarly, in the 2013 Colorado floods, the retweet probabilities of tweets with the different levels of message clarity were 11.74%, 10.84%, and 5.96% for high, intermediate, and low message clarity, respectively. We interpret these results in light of Twitter's double-edged aspects for communication: the 140-character limit allows online citizens to disseminate and receive urgent, time-sensitive information during emergencies when time is of the essence; however, such a limitation could negatively affect tweets' message clarity when twitterers try to write multiple topics together into a single tweet. We are confident about these findings for the reason that across varying time intervals, we were constantly able to observe these phenomena on the two Twitter datasets (see Table 2.9).

Based on the empirical findings of the message clarity of tweets, we tested the interaction effect of twitterers' information on the relationship between message clarity in tweets and retweet likelihood. In general, we expected that the effect of twitterers' heuristic information on the retweet likelihood varied by the different levels of the message clarity of tweets. Along with the length of affiliation, the number of followers, likes, and current incident-related tweets positively influenced the retweet likelihood of

---

16 The probability is calculated by $\widehat{Prob.} = {}^{odds}\!\big/_{1+odds}$.
Therefore, $\widehat{Prob.} = e^{\left(-2.395+\left(\frac{0.811}{2}\right)*1\right)} \Big/ 1+e^{\left(-2.395+\left(\frac{0.811}{2}\right)*1\right)}$. See Table 2.8 for the specific number(s)

17 $\widehat{Prob.} = e^{\left(-2.395+\left(\frac{0.509}{2}\right)*1\right)} \Big/ 1+e^{\left(-2.395+\left(\frac{0.509}{2}\right)*1\right)}$

18 $\widehat{Prob.} = e^{\left(-2.395+\left(\frac{0.509}{2}\right)*-1\right)} \Big/ 1+e^{\left(-2.395+\left(\frac{0.509}{2}\right)*-1\right)}$

tweets when tweets' message clarity was insufficient, and this positivity became stronger as tweets' message clarity lowered (intermediate vs. low message clarity). It is noteworthy that the number of likes was partially supported, albeit this effect's coefficient signs in both datasets were the same.

To strengthen the generalizability of this study's findings, we performed a series of additional analyses by varying time intervals. We presumed that the 10-minute interval between the posting of the original tweets and their first retweet would not be enough to represent either communication practices on Twitter during disasters or tweets' retweetability. Therefore, we defined three additional time intervals to cover a wide spectrum of time scales, with the intention of gauging the sensitivity and consistency of the empirical findings. The additional time intervals were Retweet_YN_5m, Retweet_YN_30m, and Retweet_YN_1h, which respectively correspond to the first 5 minutes, 30 minutes, and 1 hour, respectively, after posting to establish the twee–retweet relationship. Table 2.9 summarizes the results of these additional analyses. Except for the two inconsistent results, most of our interaction hypotheses were consistently supported. One inconsistency was found in the tweets for the Queensland floods in the sense that the interaction effect of the number of followers examined within the 5-minute interval was not significant. The other unreliable result was found in the Colorado dataset, given that the interaction effect of the age of the Twitter accounts was not reliable within the 5-minute interval. From the above cases, we speculate that the 5-minute interval is probably too short to consistently observe twitterers' retweet behaviors. In other words, such a short time interval would be inappropriate to reliably measure the tweet-retweet relationship.

[21] Table 2.9 Robustness Analysis

Retweet_YN$_i$ / Variable	- 2011 Queensland floods 5m Coefficient (Robust Err.)	10m Coefficient (Robust Err.)	30m Coefficient (Robust Err.)	1h Coefficient (Robust Err.)	- 2013 Colorado floods 5m Coefficient (Robust Err.)	10m Coefficient (Robust Err.)	30m Coefficient (Robust Err.)	1h Coefficient (Robust Err.)
Heuristic Information– *Twitterer Credibility*								
Ln(Followers$_{i,t}$)	0.496*** (0.00961)	0.501*** (0.00910)	0.521*** (0.00878)	0.527*** (0.00868)	0.569*** (0.00785)	0.574*** (0.00763)	0.599*** (0.00763)	0.612*** (0.00769)
Ln(Followees$_{i,t}$)	-0.0856*** (0.00737)	-0.0841*** (0.00705)	-0.0853*** (0.00686)	-0.0820*** (0.00681)	-0.0718*** (0.00805)	-0.0661*** (0.00771)	-0.0710*** (0.00756)	-0.0770*** (0.00754)
Ln(Status$_{i,t}$)	-0.177*** (0.00826)	-0.185*** (0.00781)	-0.197*** (0.00743)	-0.202*** (0.00732)	-0.366*** (0.00745)	-0.381*** (0.00704)	-0.399*** (0.00676)	-0.409*** (0.00668)
Years$_{i,t}$	0.0800*** (0.0134)	0.0798*** (0.0125)	0.0884*** (0.0118)	0.0869*** (0.0116)	0.0412*** (0.00739)	0.0545*** (0.00688)	0.0564*** (0.00652)	0.0541*** (0.00641)
Ln(Tweets$_{i,t}$)	0.122*** (0.00711)	0.116*** (0.00665)	0.107*** (0.00629)	0.102*** (0.00617)	0.285*** (0.00668)	0.288*** (0.00639)	0.277*** (0.00618)	0.272*** (0.00613)
Ln(Likes$_{i,t}$)	0.0151* (0.00598)	0.0187*** (0.00558)	0.0160** (0.00527)	0.0148** (0.00518)	0.114*** (0.00488)	0.123*** (0.00456)	0.129*** (0.00434)	0.133*** (0.00429)
Systematic Information– *Message Clarity*								
OneVSTwoMore$_i$	0.816*** (0.0213)	0.811*** (0.0193)	0.820*** (0.0179)	0.828*** (0.0174)	0.814*** (0.0287)	0.831*** (0.0261)	0.834*** (0.0239)	0.829*** (0.0232)
TwoVSMore$_i$	0.519*** (0.0313)	0.509*** (0.0280)	0.524*** (0.0257)	0.522*** (0.0249)	0.664*** (0.0457)	0.651*** (0.0412)	0.662*** (0.0374)	0.646*** (0.0362)
Dual Processing (or Bias) – *Twitterer Credibility* and *Message Clarity*								
Ln(Followers$_{i,t}$) × TwoVSMore$_i$	-0.0322 (0.0185)	-0.0363* (0.0172)	-0.0346* (0.0161)	-0.0312* (0.0158)	-0.0896*** (0.0176)	-0.0704*** (0.0166)	-0.0684*** (0.0159)	-0.0734*** (0.0159)
Ln(Followees$_{i,t}$) × TwoVSMore$_i$	-0.0118 (0.0139)	-0.0109 (0.0129)	-0.00858 (0.0122)	-0.00506 (0.0120)	0.00806 (0.0175)	-0.00315 (0.0163)	-0.00444 (0.0154)	-0.000532 (0.0153)
Ln(Status$_{i,t}$) × TwoVSMore$_i$	0.0426** (0.0152)	0.0541*** (0.0141)	0.0542*** (0.0131)	0.0514*** (0.0129)	0.0942*** (0.0159)	0.0888*** (0.0148)	0.0959*** (0.0139)	0.101*** (0.0136)
Years$_{i,t}$ × TwoVSMore$_i$	-0.0789** (0.0242)	-0.0761*** (0.0221)	-0.0877*** (0.0204)	-0.0841*** (0.0198)	-0.0143 (0.0149)	-0.0310* (0.0137)	-0.0400** (0.0127)	-0.0399** (0.0124)
Ln(Tweets$_{i,t}$) × TwoVSMore$_i$	-0.0358** (0.0130)	-0.0384** (0.0119)	-0.0397*** (0.0111)	-0.0368*** (0.0108)	-0.0341* (0.0134)	-0.0304* (0.0125)	-0.0297* (0.0119)	-0.0272* (0.0117)
Ln(Likes$_{i,t}$) × TwoVSMore$_i$	-0.00540 (0.0108)	-0.00965 (0.00991)	-0.00732 (0.00917)	-0.0116 (0.00899)	-0.0325** (0.0100)	-0.0368*** (0.00929)	-0.0438*** (0.00862)	-0.0401*** (0.00844)
Control Variables								

$Words_i$		0.0361*** (0.00254)	0.0392*** (0.00239)	0.0443*** (0.00226)	0.0465*** (0.00222)	0.0501*** (0.00278)	0.0512*** (0.00260)	0.0513*** (0.00246)	0.0515*** (0.00241)
$Hashtags_i$		0.262*** (0.00985)	0.256*** (0.00933)	0.262*** (0.00903)	0.262*** (0.00896)	0.162*** (0.00789)	0.181*** (0.00745)	0.213*** (0.00715)	0.222*** (0.00709)
$URLs_i$		-0.0404* (0.0184)	0.0136 (0.0170)	0.0698*** (0.0159)	0.0888*** (0.0155)	-0.257*** (0.0195)	-0.185*** (0.0181)	-0.112*** (0.0171)	-0.0798*** (0.0168)
$Mention\_YN_i$		-0.258*** (0.0106)	-0.248*** (0.00974)	-0.238*** (0.00911)	-0.234*** (0.00890)	-0.0693*** (0.0102)	-0.0584*** (0.00951)	-0.0480*** (0.00901)	-0.0431*** (0.00886)
Constant		-2.649*** (0.0216)	-2.395*** (0.0194)	-2.167*** (0.0178)	-2.079*** (0.0172)	-2.726*** (0.0298)	-2.432*** (0.0269)	-2.149*** (0.0245)	-2.039*** (0.0237)
Model Summary									
Percent Predicted Correctly		87.36%	85.02%	82.47%	81.41%	85.58%	82.98%	80.64%	79.79%
Predicted Probability	1	8.16%	10.81%	14.56%	15.99%	25.59%	30.99%	37.39%	39.59%
	0	99.33%	98.97%	98.26%	97.90%	97.28%	95.97%	94.23%	93.48%
Log Likelihood		9785.642***	11004.36***	12671.07***	13366.39***	18999.26***	22054.12***	24958.60***	25922.40***
Wald Chi-Square		8206.68***	9124.25***	10345.41***	10842.44***	14763.35***	16551.18***	18166.25***	18634.86***
McFadden's Pseudo R^2		0.115	0.115	0.120	0.122	0.208	0.215	0.222	0.223
n		109,456				102,426			

[1] All predictors are mean centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

2.8 Conclusion

Research on Twitter for disaster communication has made fewer efforts to explain twitterers' information in terms of information dissemination than tweets' content information. By using the notion of technological affordances, we construed twitterers' information as a heuristically accessible, source-credibility cue and empirically examined how such a source-credibility cue influenced recipients' quick decision-making. From this perspective, this study makes several contributions that enhance our understanding of Twitter as a communication medium.

Methodologically, it provides a way to reliably observe the effect of twitterers' information as a source-credibility cue on retweet likelihood. That is, to identify a condition that twitterers look for additional information, we attended to the 140-character limit of tweets, because such a length limit could provoke information-seeking behavior. In fact, tweets' 140 characters could positively or negatively influence the dissemination of tweets. On the one hand, the character limit allows twitterers to efficiently broadcast simple, up-to-date, and time-sensitive information; on the other hand, this limit can make it challenging for the twitters to clearly state the main topic or argument that they want to convey. Faced with these alternative views, we chose when the character limit was negatively manifested. That is, a decrease in tweets' message clarity could inspire recipients to search for additional information and thus, hinder the rapid dissemination of information in tweets. We believe that in this situation, twitterers' information could reduce help interpret tweets with less clear content.

Theoretically, the study relied on, and was guided by, the bias hypothesis of the HSM. Depending on the notion of technological affordances that introduces reputation, bandwagon, liking, and recency heuristics, we interpreted twitterers' heuristic information as a source-credibility cue. Then, the cued information was examined with the different levels of tweets' message clarity. In so doing, we provided empirical evidence about the bias hypothesis, which we expected to occur in disaster communication on social media. As a result, the empirical findings supported part of our argument that an author's credibility is an important heuristic cue that reduce cognitive processing efforts when a quick decision has to be formed.

Practically, this study highlighted one fundamental problem about deploying short messages for disaster alerts and warnings, and empirically examined how a twitterer's heuristic cues complemented the length-limit of his or her tweets. Based on the empirical results, we suggest the following guidelines for disaster communication. First, source or author information should be included in tweets as much as possible, such that recipients can reduce time to make their retweet judgment when facing insufficient message clarity in tweets. Second, during times of disaster, Twitter Company could include more specific information about twitterers as a means of better representing author credibility. Specifically, the company could remove the number of followees whose effect was not significant on retweeting. However, as the results indicate, the company should consider displaying twitterers' recent tweets about specific domains or topics. For example, when twitterers search tweets by certain keywords or hashtags, such as *Colorado flood* or *#coflood*, the company could show relevant tweets by displaying the number of recent tweets of each twitterer regarding the relevant tweets crafted about similar topics. Lastly, this research can be a humble reference for emergency management officials when writing alerts and warnings on the Wireless Emergency Alerts (WEAs) platform, which allows only 90 characters at maximum, or on Twitter.

Like many other studies, this study has limitations too. First, to answer the research question and a set of hypotheses, we leveraged two Twitter datasets about flooding. We believe, however, that testing the hypotheses in diverse types of natural disasters, such as hurricanes, earthquakes, or wildfires, will enhance the generalizability of this study's findings. In a similar vein, future research can leverage the methods of the study to investigate online citizen's information dissemination behavior on other social media platforms, such as Facebook, blogs, and discussion forums. Second, along with the method we used to define message clarity in tweets, taking into account people's direct responses to the distinguished message clarity will be another opportunity for future research. Furthermore, to observe the influence of twitterers' credibility-cues on retweeting, the use of another measure, rather than message clarity in tweets, should be plausible to firmly confirm the current findings.

Appendix 2.A

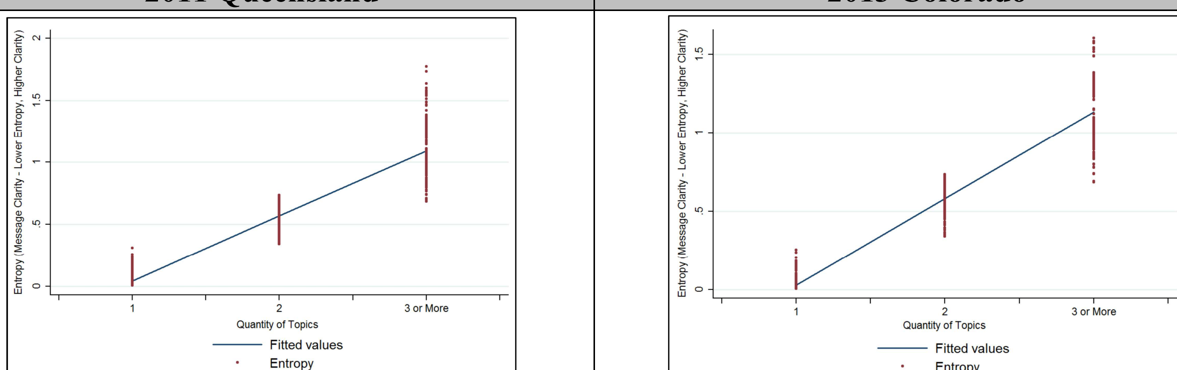
To empirically confirm the relationship between the number of topics and entropy, we performed one degree of freedom analysis using tweets corrected during the 2011 Queensland and 2013 Colorado floods. We used entropy as the dependent variable, while having the number of topics as the independent variable. The results indicated that in both cases a strong linear relationship existed (see Figure 2.A). As shown in Table 2.A, the number of topics explained 94.46% of the variance of the entropy in the Queensland case (R-squared=0.9446, Coefficient=0.5236, p<0.000) and 95.31% in the Colorado case (R-squared=0.9531, Coefficient=0.5516, p<0.000).

[22] Table 2.A Statistical Results between Entropy and the Number of Topics

Cases / Variables	2011 Queensland		2013 Colorado	
	Coefficient (Error)	Sig. Level	Coefficient (Error)	Sig. Level
$Topic\_Number_i$[1]	0.524*** (0.000528)	0.0000	0.552*** (0.000682)	0.0000
Constant	-0.479*** (0.000630)	0.0000	-0.523***(0.000755)	0.0000
Model				
R^2	0.9446		0.9531	
Adjusted R^2	0.945		0.953	
n	109456		102426	

[1] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

2011 Queensland	2013 Colorado



[9] Figure 2.A Group Comparisons of Tweets Conveying 1, 2, or More Topics

Appendix 2.B

10-Minute Time Interval to Establish Tweets and Their Retweets

In order to closely look into the first 60 minutes, we conducted statistical analyses by accumulating the retweet frequency of tweets within the first 24 hours after posting to present the retweetability of tweets. From the analyses, we found that the relationship between tweets and their first retweet established within the first 10 minutes after posting best represented tweets' retweetability among the relationships established within other time periods. Therefore, we use this 10-minute period to measure the retweet likelihood of tweets. The rest of this section is allocated to explain these statistical approaches that we devised.

[23] Table 2.B.1 Description of Variables

Variables	Explanation \ Cases	2011 Queensland			2013 Colorado		
		Mean	S.D.	Range	Mean	S.D.	Range
Dependent Variable							
$Retweets\_24h_i$	The total number of retweets of tweet *i* within the 24 hours after posting						
Explanatory Variables							
- Point Estimate							
$Retweet\_Minute_i$	The time interval in minutes between tweet *i* and its first retweet. If tweet *i*'s first retweet is not made within 60 minutes, its *Retweet_Minute* is coded 61.	49.4	22.8	0-61	47.3	24.1	0-61
$Retweet\_Minute^2_i$	A code for testing non-linearity of *Retweet_Minute* – Quadratic relationship						
$Retweet\_Minute^3_i$	A code for testing non-linearity of *Retweet_Minute* – Cubic relationship						
- Between Group Estimate							
50mVSOthers	(Not a meaningful contrast code)						

20mVS(50m, 30m, 40m, 10m)	(Not a meaningful contrast code)
50mVS(30m, 40m, 10m)	(Not a meaningful contrast code)
30mVS(40m, 10m)	(Not a meaningful contrast code)
10mVS40m	A contrast code to compare the retweet frequency between a group of tweets whose first retweet is made within the first 10 minutes after posting and a group of tweets whose first retweet is made between the 30 and 39 minutes after posting.

We identified an individual tweet's first retweet and calculated the time difference between this tweet and its first retweet. For instance, if a tweet is retweeted 15 minutes after its posting, this tweet's elapsed time is 15. However, if another tweet's first retweet is made after the first 60 minutes since its posting, '61' is given since we are only interested in a one-hour period. The control and exploratory variables are explained in Table 2.B.1.

[24] Table 2.B.2 Statistical Results between Entropy and Topic Quantity

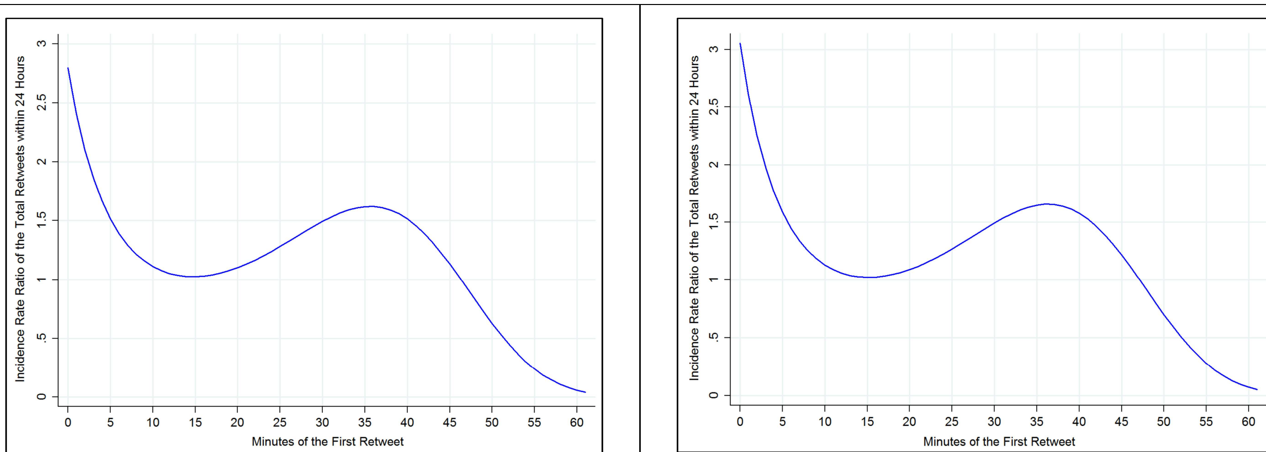
Cases / Variables	2011 Queensland		2013 Colorado	
	Coefficient (Robust Error)	Significance Level	Coefficient (Robust Error)	Significance Level
Explanatory	*Wald Chi*^{2}=7816.82, df=3, p<0.000		*Wald Chi*^{2}=12571.53, df=3, p<0.000	
$Retweet\_Minute_i$	-0.158*** (0.00920)	0.000	-0.168*** (0.00523)	0.000
$Retweet\_Minute^2_i$	0.00758*** (0.000388)	0.000	0.00784*** (0.000288)	0.000
$Retweet\_Minute^3_i$	-0.000100*** (0.00000416)	0.000	-0.000101*** (0.00000354)	0.000
Control	*Wald Chi*^{2}=882.74, df=8, p<0.000		*Wald Chi*^{2}=3142.94, df=8, p<0.000	
$Mention\_YN_i$	-0.0793*** (0.0180)	0.000	-0.0481*** (0.0124)	0.000
$Words_i$	0.0400*** (0.00412)	0.000	0.0230*** (0.00289)	0.000
$URLs_i$	0.124*** (0.0301)	0.000	0.170*** (0.0222)	0.000
$Hashtags_i$	0.00185 (0.0238)	0.938	0.0569*** (0.00845)	0.000
$Ln(Followers_{i,t})$	0.311*** (0.0135)	0.000	0.351*** (0.00747)	0.000

Ln($Followees_{i,t}$)	0.0323 (0.0168)	0.000	0.0494*** (0.00530)	0.000
Ln($Likes_{i,t}$)	-0.0935*** (0.0108)	0.055	-0.0576*** (0.00751)	0.000
Ln($Status_{i,t}$)	-0.130*** (0.0173)	0.000	-0.228*** (0.00829)	0.000
Constant	1.029*** (0.100)	0.000	1.117*** (0.0567)	0.000
Model Summary				
Log-likelihood Ratio	90102.500***		87104.669***	
Wald χ^2	12506.52***		24184.33***	
McFadden's R^2	0.339		0.363	
n	109,456		102,426	

[1] All predictors are mean centered in the regression.
[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.
[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

From a negative binomial analysis (see Table 2.B.2), we found that in both cases, the relationship between a tweet's first retweet and its retweet frequency within 24 hours was not linear, but rather curvilinear (Queensland – Wald Chi^2=7816.82, df=3, p<0.000; Colorado – Wald Chi^2=12571.53, df=3, p<0.000) (see Figure 2.4). That is, there was a steep decrease in the relationship between the first retweet made within 10 minutes (0~9 minutes) and the total retweet frequency. Then, the steep decrease plateaued (10~19 minutes). Next, the total retweet frequency gradually increased again (20~40 minutes) before drastically decreasing (40~59 minutes). Based upon this observation, ten-minute intervals seem to reflect well this non-linear pattern, and as such, could be a potential time unit to measure the retweet likelihood of tweets. Using a 10-minute interval, we created the following six intervals – 0_9m (an interval of 0 and 9 minutes), 10_19m, 20_29m, 30_39m, 40_49m, and 50_59m. Because we only considered the first retweet, each tweet belonged to only one period.

2011 Queensland floods	2013 Colorado floods



- *Retweet_Minute=-0.156****
- *Retweet_Minute^2=0.00758****
- *Retweet_Minute^3=-0.001****

- *Retweet_Minute=-0.168****
- *Retweet_Minute^2=0.00784****
- *Retweet_Minute^3=-0.00101****

[10] Figure 2.B Non-linear Relationship

By using the above six intervals, the subsequent analysis aims to identify the time interval that best represents the retweetability of tweets in terms of the total retweet frequency within the first 24 hours. Therefore, we performed another negative binomial regression to estimate between-group differences. To make an effective comparison of our multi-level categories, we ordered these six intervals by their relationship strength with the total retweet frequency and then devised completely orthogonal contrasts (Bruin 2016; Judd et al. 2011). Therefore, the intervals were ordered from low to high IRR (Incidence Rate Ratio) as follows: 50_59m, 10_19m, 40_49m, 20_29m, 30_39m, and 0_9m (see Table 2.B.3). In so doing, our statistical analysis to examine between-group differences became simple. That is, if we successfully demonstrate that the first retweet made within the first 10 minutes (0_9m) has a stronger relationship with total retweet frequency than the first retweet made between 30 and 39 minutes (30_39m), we can claim that the first retweet made within the first 10 minutes best describes tweets' retweetability.

[25] Table 2.B.3 Contrast Codes for Group Comparison

Intervals	50_59m	10_19m	40_49m	20_29m	30_39m	0_9m

Variables						
50mVSOthers	1	-1/5	-1/5	-1/5	-1/5	-1/5
20mVS(50m, 30m, 40m, 10m)	0	1	-1/4	-1/4	-1/4	-1/4
50mVS(30m, 40m, 10m)	0	0	1	-1/3	-1/3	-1/3
30mVS(40m, 10m)	0	0	0	1	-1/2	-1/2
10mVS40m	0	0	0	0	1	-1

The statistical results revealed that tweets whose first retweet was posted within the first 10 minutes (0_9m) after posting were retweeted significantly more than other tweets that were first retweeted between the 30 and 39 minute interval (30_39m) (Queensland – coefficient=-1.077, df=1, p<0.000; Colorado – coefficient=-0.717, df=1, p<0.000) (See Table 2.B.4). As a result, this time interval that we empirically found was utilized to establish the relationship between tweets and retweets for testing our hypotheses.

[26] Table 2.B.4 Statistical Results between Entropy and the number of Topics

Cases / **Variables**	**2011 Queensland**		**2013 Colorado**	
	Coefficient (Robust Error[1])	Sig. Level	Coefficient (Robust Error[1])	Sig. Level
Explanatory	*Wald Chi*^{2}=8236.6, df=3, p<0.000		*Wald Chi*^{2}=10013, df=3, p<0.000	
50mVSOthers	-9.314*** (0.909)	0.000	-10.01*** (0.557)	0.000
20mVS(50m, 30m, 40m, 10m)	0.567*** (0.122)	0.000	0.322*** (0.0764)	0.000
50mVS(30m, 40m, 10m)	-1.212*** (0.265)	0.000	-1.118*** (0.158)	0.000
30mVS(40m, 10m)	-0.166* (0.0681)	0.015	-0.280* (0.109)	0.010
10mVS40m	-1.077*** (0.0963)	0.000	-0.717*** (0.137)	0.000
Control	*Wald Chi*^{2}=1138.5, df=8, p<0.000		*Wald Chi*^{2}=3552.9, df=8, p<0.000	
Mention_YN$_i$	-0.103*** (0.0179)	0.000	-0.0502*** (0.0129)	0.000
Words$_i$	0.0446*** (0.00407)	0.000	0.0272*** (0.00305)	0.000
URLs$_i$	0.128*** (0.0298)	0.000	0.148*** (0.0241)	0.000

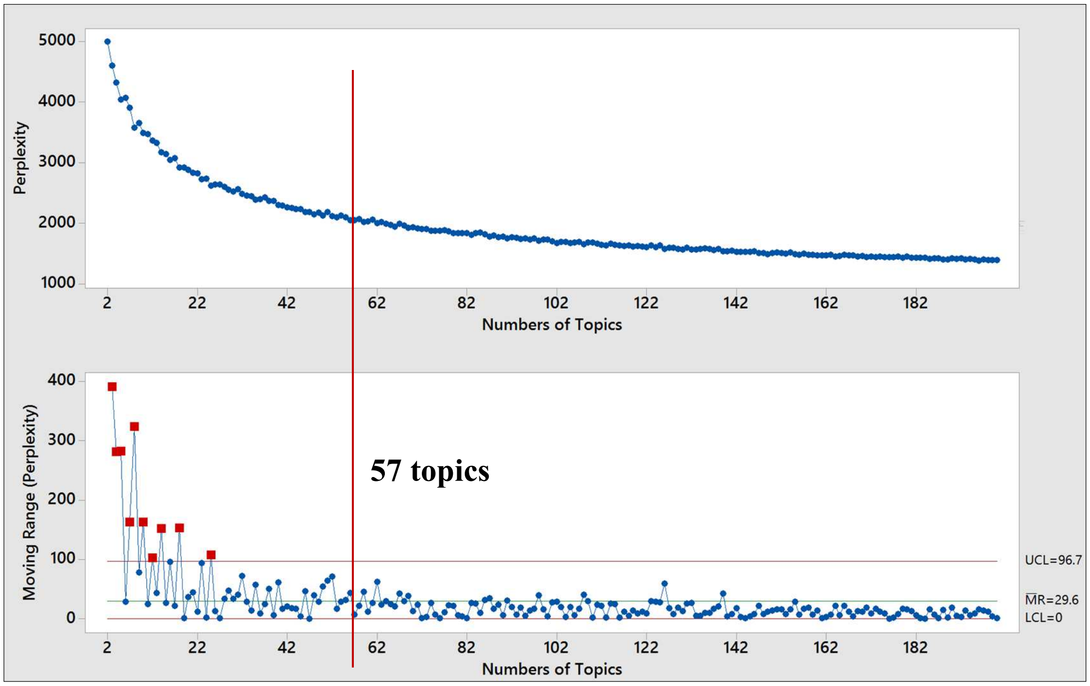
$Hashtags_i$	0.0415 (0.0226)	0.067	0.0855*** (0.00893)	0.000
$Ln(Followers_{i,t})$	0.360*** (0.0134)	0.000	0.411*** (0.00783)	0.000
$Ln(Followees_{i,t})$	0.0344 (0.0177)	0.000	0.0600*** (0.00576)	0.000
$Ln(Likes_{i,t})$	-0.0982*** (0.0105)	0.052	-0.0622*** (0.00742)	0.000
$Ln(Status_{i,t})$	-0.143*** (0.0170)	0.000	-0.258*** (0.00806)	0.000
Constant	-2.787*** (0.104)	0.000	-2.594*** (0.0644)	0.000
Model Summary				
Log-likelihood Ratio	69283.458***		71237.125***	
Wald χ^2	13923.93***		24092.02***	
McFadden's R^2	0.261		0.297	
n	109,456		102,426	

[1] All predictors are mean centered in the regression.

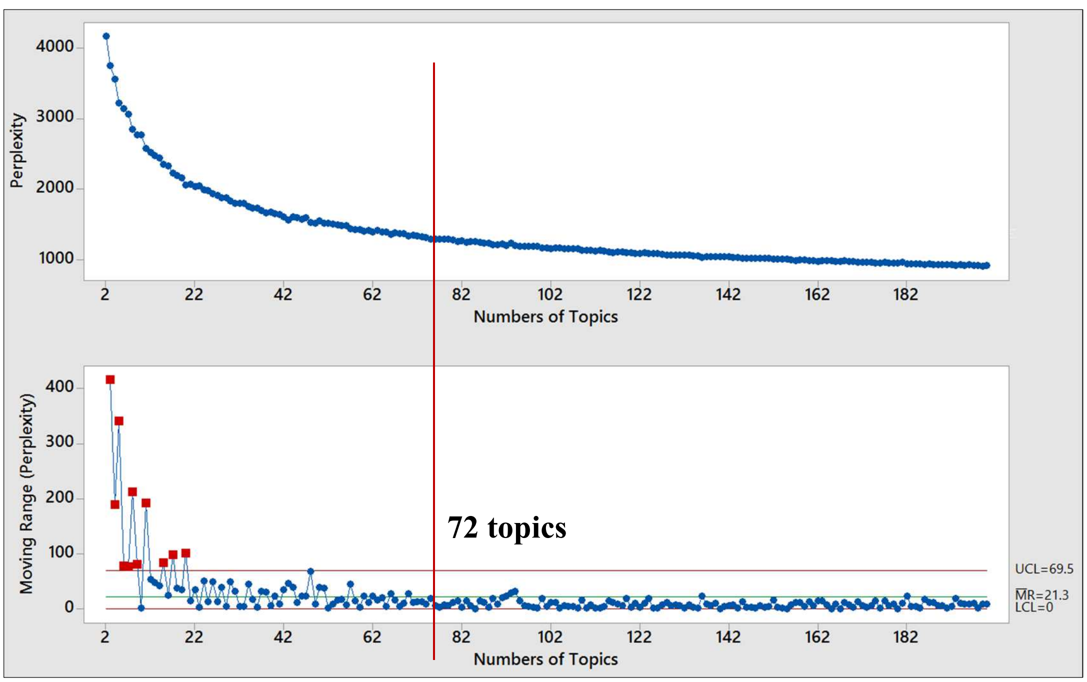
[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

Appendix 2.C



< 2011 Queensland floods >



< 2013 Colorado floods >

[11] Figure 2.C Perplexity Scores and Their Moving Ranges for the Two Twitter Datasets

Appendix 2.D

[27] Table 2.D.1 57 Topics and Keywords of the 2013 Colorado floods

Topic #	Keywords
1	relief levels flood_levels give pic impression friends news add #twibbon create federer tennis victims online flood_relief flood_victims abc abc_news
2	need volunteer register volunteers clean cleanup #bnefloods volunteering brisbane food emergency accommodation #bakedrelief #bnecleanup needs needed
3	centre evacuation evacuation_centre showgrounds pets ipswich spread word ipswich_showgrounds rna evac rna_showgrounds centres lost found hills
4	change cross red climate red_cross #vicfloods climate_change clean rain australian towns weather relief services affected brace information brisbane
5	fill sandbags need free brisbane form council affected services nature disaster offer businesses local train_services stop contact mother_nature city
6	support map comparison map_comparison relief post affected blog rough event #vicfloods fundraiser peeps blog_post benefit devastation happening fundrasing
7	victims flood_victims stay released place ravaged advice friends airport police donation legal information free #vicfloods affected hotline recovery
8	volunteers helping proud disaster clean spirit hand #vicfloods army efforts relief together rescue australian amazing community #bnefloods workers
9	bligh anna_bligh anna premier brisbane queensland_premier low residents evacuate lying higher water ground inquiry ipswich urged #brisbane starting
10	crisis news flood_crisis bligh toll premier missing death latest anna_bligh anna dead disaster live death_toll online confirmed ahead buying
11	victims donate donating appeal remember sitting flood_victims donation link vic nsw left donations harvey amazing coast #auction vintage total
12	ipswich mayor looting ipswich_mayor paul pisasale city markers paul_pisasale find flood_markers brisbane higher mythbuster flood_mythbuster facing pi
13	water power brisbane residents safe supply ipswich #bnefloods shopping boil water_supply centre food victims drink advised cut flood_victims need
14	spirit aussie aussie_spirit amazing victims flood_victims home donate working flooded return family find cleaning heaps thanks_heaps strangers aussie
15	creek cars footage flash toowoomba washed video lockyer flash_flood lockyer_creek mil show evacuate gave film water mate oprah gympie god rises higher

16	victims flood_victims donate support donating every affected money #prayforaustralia raise need hope handset donations retweet visit generously coffee
17	#qld affected judgment judgment_day update insurance #bnefloods brisvenice brisvegas flood_update longer brisbane info hotline tourism #vicfloods bus
18	disaster size area zone declared texas disaster_zone times united flood_disaster france kingdom united_kingdom germany united_kingdoms kingdoms
19	power cut energex brisbane ipswich free affected homes image charge restore phones facing inundation families businesses mythbuster flood_myth
20	brisbane storage photos images free brisbane_floods live free_storage #bnefloods storage_king offering trucks #brisbane aerial affected pics amazing
21	cross red safe brisbane national registration system free cow roof #bnefloods clean water map inquiry place photos #brislantis damaged cross_national
22	high zoo swim crocs australia_zoo high_enough tying brisbane weather god biggest arrive bureau biggest_flood weather_bureau companies insurance_compa
23	media social social_media twitter #vicfloods health helping aid police hope australia_day need doctors join email stars disaster dept needed sunrise
24	brisbane river brisbane_river #bnefloods floating cbd farm drive streets list expected restaurant free park city affected coronation coronation_drive
25	man volunteers photo boatload kangaroos needed rescued #bnecleanup mayor kangaroo more_volunteers pic brilliant registration centres
26	volunteers auctions need awesome cahill qld_floods tim_cahill tim awesome_auctions cold beers ground high cbd mobile cold_beers handing high_ground
27	crisis flood_crisis list real media citizen reports citizen_reports died twitter related stories line info outlets lifeline twitter_list media_outlet
28	river brisbane broken brisbane_river banks end west library west_end wet sunny dry sunny_day wrap freezer gladwrap wet_photosbooks photosbooks
29	evacuation info centres financial brisbane app hit pledges evacuation_centres financial_help dogs cats owners recovery free staff information links b
30	river brisbane peak brisbane_river expected levels metres conference media ipswich #bnefloods media_conference flood_peak live level livestream tab
31	#bnefloods brisbane closed street bank ipswich bridge water pier eagle cbd shit open #brisbane south_bank river motorway holy holy_shit crap #fb road
32	stadium suncorp_stadium suncorp brisbane pool swimming picture field footy_field #bnefloods water fire bridge transformer emergency services silence

33	waters flood_waters children disaster helping barrier reef #auspol barrier_reef support office play water damage equipment replace stop homes pay
34	warning severe rain thunderstorm weather brisbane thunderstorm_warning flash coast hit #qld bay bom #tcanthony heavy river cyclone moreton
35	donate every appeal flood_appeal tweet cents aussie aussie_queensland #prayforaustralia retweet message #staystrong received qld_floods everyone
36	#vicfloods #nswfloods map need information road closures info flood_information road_closures contact crisis #tasfloods urgent list live flood_map
37	donate need queenslanders desperately police facebook updates page twitter phone flight qld_floods qld_police date change affected booking service
38	victims australian fundraiser items fan international win fan_fundraiser autographed auction autographed_items bed offer recent house affected spare
39	victims cahill auction experience flood_victims tim raise bid money tim_cahill #socceroos match ebay charity everton aid signed shirt cricket relief
40	affected survival animals offer email housing foster assistance email_floods foster_caretemporary caretemporary bill unnecessary lewis survival_value
41	abbott deep tony_abbott tony water #auspol dig flood_water donations bin wheelie indication wheelie_bin #nbn good_indication dollar political need
42	toll death death_toll valley lockyer found lockyer_valley missing rises bodies grantham police flood_death_toll dead flash news body man risen search
43	bligh anna_bligh anna premier conference gillard press julia crisis media julia_gillard press_conference leadership live pressure #abcnews leader qld
44	snake frog ride photo community hitches incredible escapes frog_escapes_flood incredible_photo looting bligh escape created riding anna red australia
45	appeal relief flood_relief_appeal aussies donate everyone thinking needs premier #aussies flood_appeal disaster donating relief_appeal donated
46	missing dead rice jordan jordan_rice confirmed #prayforaustralia hero brother died sad save queensland_floods lost boy saving homes rip god queensland
47	relief appeal flood_relief fund auction money raise proceeds donate donated flood_appeal raised signed bid relief_fund song funds raising sales donation
48	recovery tsunami inland biblical flood_recovery impact inland_tsunami crisis faces facing economic news hell official support force warns economic_impact
49	shark ipswich bull street goodna flooded brisbane bull_shark spotted flooded_street sharks affected update streets swimming bull_sharks main main_street

50	brisbane city council city_council latest live game news alert updates services support online #bnefloods notice media info collection drinking
51	rspca fairfield animals fairfield_rspca water qld_floods repost foster animal retweet shelter register origin raise money jerseys origin_jerseys
52	towns affected brisbane crisis news medical coal free water flood_crisis relief clean volunteer offering home car cities inundated recovery reds
53	safe affected everyone thoughts brisbane hope stay news #prayforaustralia friends prayers family home heart lost sad devastating hear watching rain
54	waters flood_waters city australian rockhampton braces brisbane fundraiser rise peak queensland_braces coastal rising river satellite bridge fundraising
55	end brisbane water west house home clean need #bnefloods mud west_end helping flooded hand hard river #vicfloods cleaning city power volunteers girl
56	flooded homes brisbane affected businesses need power stallion suburbs supply bay needed inundated water #bnefloods ipswich spare energex deception
57	jordan rice jordan_rice save swept younger rescuers brother life younger_brother blake own_life losing stop hero toowoomba aged waters #prayforaustralia
58	relief #vicfloods view volunteers cross hills needs bowen support concert red_cross neighbours bowen_hills service clean crisis brisbane continues
59	relief donate flood_relief donations needs appeal word spread information flood_relief_appeal everyone need #prayforaustralia qld_australia needed
60	heart health aussies praying safety prayers hearts breaks picture markets #bnefloods rocklea rocklea_markets brisbane disaster system team chopper fr
61	points velocity velocity_points brisbane closed donation allowing convert #bnefloods donate recovery donating donations road awesome page milton
62	appeal flood_appeal donate rspca animals give donations money qld_rspca raise need donated generously #vicfloods sales #prayforaustralia plead donati
63	damage insurance flood_damage need business food brisbane storm claims small milk pay water #bnecleanup supplies levy clean hit bread guide office
64	brisbane transport cross public red_cross centre public_transport volunteers red needs affected melbourne seekers asylum_seekers north needed based
65	brisbane cbd brisbane_cbd power closed evacuated transport coast myth public buster flood_myth_buster public_transport highway #bnefloods closing
66	found dogs dog disaster goodna island need floating toilet lost fraser block flood_disaster toilet_block fraser_island sharon pray god caltex sleep k

67	donate appeal fireworks day_fireworks cancel recovery donated money relief flood_appeal donation million free ride fund raising awareness twitter
68	levy video flood_levy tax #vicfloods friend gillard youtube relief #auspol pay victims images toowoomba view youtube_video nasa queensland_floods
69	water services fire lost normal goods kid home talent normal_kid stefanovic karl treatment affected karl_stefanovic plants supply summary room need s
70	coverage news abc maps radio brisbane live info local council information flood_maps online #abcnews site twitter updates channel #bnefloods city dig
71	town update residents dalby link area pool audio_link alert recovery pool_area emerald audio rockhampton road hit #police swimming power cut southern
72	dam wivenhoe brisbane water #bnefloods cbd lucia capacity view river street st_lucia brisbane_cbd albert farm full george new_farm southbank flooded

[28] Table 2.D.2 72 Topics and Keywords of the 2011 Queensland floods

Topic #	Keywords
1	toll, death, dead, rises, person, death_toll, flood_death_toll, evacuations, confirmed, people, deadly, presumed, woman, ordered, found, flood_toll, waters, missing
2	evacuation, center, head, jamestown, residents, notice, springs, eldorado, evac, creek, eldorado_springs, cty, evacuation_center, evacuation_notice, people, barn, ordered
3	towns, rescue, rain, rains, warnings, flood_warnings, diverse, closed, forecast, cats, flood_towns, colorado_towns, break, flood_rescue, stranded, brief_break, hamper, waters
4	schools, aurora, closed, creek, aurora_pd, creek_schools, aurora_schools, request, canyon, water, cherry, debris, valley, surge, foot, cars, other_debris, carrying, boulder
5	rescue, boulder, operation, water, flood_rescue_operation, area, report, continues, home, weather, leave, spill, chemical, historic, drive, fracking, rain, chemical_spill
6	record, breaking, guard, coast, led, worse, denver, concert, coast_guard, helicopters, relief, survivors, defense, coast_guard_helicopters, victims, benefit, state, coming
7	people, county, unaccounted, boulder, rescued, rescue, crews, sheriff, larimer, man, helicopters, save, officials, larimer_county, pets, boulder_county, racing, news, air
8	mountain, city, rocky, national, dam, commerce, arsenal, rocky_mountain_arsenal, evacuations, wildlife, refuge, failed, wildlife_refuge, impassable, roads, streets, east, dams
9	guard, national_guard, national, town, lyons, residents, jamestown, evacuations, moves, continue, boulder, evacuate, news, crest, downstream, colorado_town, students
10	creek, boulder, boulder_creek, move, broadway, sirens, sounding, #cuboulder, higher, canyon, cfs, east, ground, higher_ground, mesa, place, rising, table, flood_sirens, shelter
11	canyon, boulder, water, ground, higher, higher_ground, wall, coming, boulder_canyon, creek, immediately, move, boulder_creek, gulch, emerson_gulch, emerson, seek, debris, pearl

12	boulder, rain, evacuate, flash, more_rain, continue, ordered, live, county, officials, rescues, expected, braces, lyons, flash_flood, colorado_braces, damage, town, downtown
13	warning, flash, flash_flood_warning, boulder, flash_flood, county, issued, flood_warning, counties, effect, watch, skies, rain, warnings, evacuees, denver, springs
14	creek, boulder_creek, boulder, water, flow, wall, usgs, official, denver, term, experts, tsunami, experts_term, readings, creek_flow_readings, sensor, fourmile, usgs_sensor
15	platte, river, oil, south, spills, south_platte_river, gallons, tank, swollen, platte_river, spill, south_platte, damaged, morgan, reported, waters, water, oil_spill, Greeley
16	oil, gas, spills, zones, #fracking, wells, tracking, flood_zones, waters, sites, fracking, flood_waters, post, flooded, chemicals, water, gas_wells, leaks, denver, denver_post
17	gallons, locations, road, drenched, crude, dumps, spill, oil_spill_dumps, closures, waters, road_closures, flooded, boulder, water, many_locations, loved, shelter, affected
18	disaster, flood_disaster, media, blackout, media_blackout, #fracking, fracking, spills, happening, photos, update, toxic, worse, confirmed, shocking_photos, underwater, zone
19	waters, water, flood_waters, piano, house, play, home, sewage, wrecked, boulder, decided, man, contaminated, avoid, plays, sweep, moments, bike, creek, colorado_home, stay, video
20	vrain, water, river, creek, bridge, evac, roads, lyons, place, street, boulder, vrain_river, longmont, home, loveland, dry, center, big, #longmont, stay, hygiene, news, left
21	thompson, big, river, thompson_river, feet, county, ravaged, woman, pound, fatality, canyon, fifth_fatality, thompson_canyon, stage, record, loveland, central, thompson_flood
22	photo, car, havana, viewer, lyons, viewer_photo, swim, road, air, hwy, town, boulder, damage, hwy, news, water, dillon, pic, collapse, assessment, rescue, road_collapse, inside
23	longmont, #longmontflood, victims, water, lyons, view, rescues, equine, dam, storm, helicopter, vehicles, register, volunteers, image, urgent_call, woman, soldier, blog
24	long, water, city, safe, boulder, photo, rain, washed, picture, commerce, denver, commerce_city, rescue, stay, roads, house, areas, problems, live, couple, photos, send, yards, mile
25	images, unbelievable, unbelievable_images, boulder, map, google, tremendous, began, crisis, area, travel, water, notice, earth, evacuation, severe, google_earth, flash
26	game, football, school, state, path, bike, bike_path, postponed, high, fresno, field, pic, park, aurora, high_school, utah, utah_park, baseball, baseball_field, overland
27	front, range, front_range, boulder, coverage, open, space, water, emergency, relief, trucks, workers, rescue, hard, downtown, disaster, county, working, longmont, effort, parks
28	damage, photos, aerial, images, flood_damage, video, biblical, climate, line, trends, boulder, climate_trends, views, biblical_flood, show, waters, aerial_views, lyons, shot
29	campus, evacuation, damage, homes, water, mobile, school, mobile_homes, creek, high, epic, buildings, boulder, photo, shows, water_damage, city, shelters, closed, high_school
30	big, thompson, canyon, thompson_canyon, road, hwy, hwy, thousand, boulder, flooded, water, science, thompson_canyon_entr, entr, baseline, damage, photographers, cut, deep
31	road, closures, road_closures, map, list, county, updates, boulder, closure, updated, found, #copets, center, shelters, latest, shelter, evacuation, road_closure_map, roads, dog
32	park, hwy, hwy, closed, estes, estes_park, #cotraf, open, road, roads, highway, photos, disaster, #estespark, directions, news, fun, reporter, app, denver, evergreen
33	water, boil, residents, high, drinking, lyons, safe, treatment, drink, advisory, hand, boulder, district, city, vehicles, wastewater, left, bottled, town, levels, contaminated

34	recovery, information, response, volunteer, relief, resources, updates, communities, efforts, live, emergency, cleanup, blog, affected, local, boulder, long, flood_recovery
35	disaster, assistance, fema, boulder, emergency, county, recovery, center, counties, federal, disaster_assistance, declaration, map, affected, evacuation, register
36	damage, losses, billion, flood_damage, property_losses, relief, repairs, shutdown, property, million, government, flood_relief, highways, left, street, bridges, estimated
37	aid, unanimously, republicans, relief, sandy, sandy_aid, colorado_republicans, opposed, support, flood_relief, voted, house, house_republicans, flood_relief_unanimously
38	biden, recovery, hickenlooper, devastation, flood_devastation, damage, view, president, fema, efforts, joe, gov, response, vice_president, team, vice, joe_biden, news, rescue
39	victims, relief, word, free, spread, #cofloodrelief, storage, free_storage, flood_victims, fund, flood_relief, giving, donating, donated, flood_relief_fund, marijuana
40	relief, victims, flood_victims, #cofloodrelief, donate, efforts, flood_relief, support, fundraiser, benefit, affected, donations, relief_efforts, effort, raised, helping
41	people, unaccounted, oem, areas, boulder, rain, more_rain, awaits, number, center, flood_areas, boulder_oem, remain, home, shelter, stop, area, volunteers, listed, report
42	homes, unaccounted, people, destroyed, damaged, dead, evacuated, missing, shelters, search, homes_damaged, update, loved, safe, register, presumed, homes_destroyed, numbers
43	family, impacted, pray, fire, guard, epic, reach, flush, truck, zone, members, stranded, driving, food, video, housing, flood_zone, fire_truck, order, guard_members, residents
44	cross, red, victims, flood_victims, red_cross, give, texting, climate, change, shelter, climate_change, affected, shelters, people, volunteers, american, #cofloodrelief, safe
45	collins, fort, fort_collins, relief, south, view, north, support, efforts, friends, #foco, based, resorts, vail_resorts, denver, co_support, closed, relief_efforts, pass, season
46	canyon, boulder, residents, people, shelters, left, stayed, hand, water, boulder_canyon, springs, evacuated, overnight, creek, road, expected, support, providing, #redcross
47	safe, needed, share, #copets, pets, food, victims, volunteers, lost, animals, home, hay, register, #cofloodrelief, pet, loved, victim, longmont, check, disaster, donations, sign
48	pets, rescued, people, visit, best_way, evacuated, helicopter, victims, katrina, survivors, number, historic, #nationalguard, historic_flood, #copets, greatest_number, town
49	boulder, longmont, springs, closed, humane, open, manitou, society, humane_society, page, ave, #waldoflood, center, shelter, west, front_page, manitou_springs, animals, #hmrd
50	safe, boulder, stay, rain, friends, prayers, thoughts, people, hope, affected, home, good, dry, family, love, raining, bad, crazy, victims,
51	schools, aurora, closed, creek, aurora pd, creek schools, aurora schools, request, canyon, water, cherry, debris, valley, surge
52	rain, inches, totals, wild, instagrams, wild flood, rainfall, snow, boulder, received, map, record, past, annual, feet, rain totals
53	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery
54	live, victims, coverage, flood_victims, rocks, force, task, red, task_force, state, rain, red_rocks, news, rescues, continue, debris, good, water, honor, team, oil, photo, tribune
55	disaster, boulder, waters, flood_waters, people, allowed, fracking, tubing, boulder_pd, reminds, flood_disaster, cited, floodwaters, fracking_disaster, missing, sky, clears

56	county, weld, boulder, denver, post, weld_county, residents, denver_post, water, closed, boulder_county, evacuations, pipeline, road, oil_pipeline, roads, oil, blvd, rain
57	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery, more_rain, snarls, fundraiser, latest_updates, live

Appendix 2.E

[29] Table 2.E.1 Test of Multicollinearity – 2011 Queensland floods			
OneVSTwoMore	1.45	Ln(Likes)	1.34
TwoVSMore	1.40	Words	1.19
Ln(Followers)	3.37	Hashtags	1.12
Ln(Followees)	2.06	URLs	1.27
Ln(Status)	2.66	Mention_YN	1.07
Year	1.25		
Ln(Tweet)	1.27	**Mean VIF**	1.61

[30] Table 2.E.2 Test of Multicollinearity – 2013 Colorado floods			
OneVSTwoMore	2.01	Ln(Likes)	1.46
TwoVSMore	1.97	Words	1.18
Ln(Followers)	3.42	Hashtags	1.20
Ln(Followees)	2.50	URLs	1.19
Ln(Status)	2.05	Mention_YN	1.07
Year	1.40		
Ln(Tweet)	1.25	**Mean VIF**	1.72

Chapter 3. Essay 3: Tweets' Initial Propagation: Why It Is Important and How to Measure It

Abstract

In disaster situations where dynamic, non-routine events appear and disappear in a short time span, timely information about emergency warnings and alerts determines whether people's lives will be saved or lost. Therefore, the propagation of information immediately after it is generated (or the initial propagation of information) is of primary interest to disaster researchers. With an awareness that either a tweet's average retweet interval or retweet frequency is not enough to explicate a tweet's initial propagation, we propose an index that quantifies the extent to which a tweet is propagated right after its posting. Using two Twitter datasets collected during the 2011 Queensland and the 2013 Colorado floods, we examine how well the proposed index reflects the initial propagation of tweets. To provide more solid empirical evidence of the index, we also examine two factors that are supposed to affect the initial propagation of tweets – a tweet's information sufficiency (or insufficiency) and Twitter URLs. Our findings demonstrate that the index was a better measure than either the average retweet time or retweet frequency in predicting a tweet's propagation, that information sufficiency in a tweet influenced its speed and scale of propagation, and that Twitter URLs were conditional depending on the degree of information sufficiency in a tweet.

Keywords: Tweet Propagation, Information Sufficiency, Disaster Communication

3.1 Introduction

Over recent years, Twitter has positioned itself as one of the most prominent microblogging platforms for communicating personal experiences and opinions in daily life (Aladwani 2015; Shi et al. 2014). Its ease of use (Murthy 2011, p. 781), immediacy of updating (Westerman et al. 2014, p. 174), and cross-platform accessibility (Vieweg et al. 2010, p. 1079) make it most appropriate for sharing information about fast-paced events. Online citizens learn about evolving events by receiving a tweet that a twitterer (i.e., user of Twitter) posts to those who are connected to that twitterer, or they can simply search the *Twitterverse* to acquire tweets of interest. In fact, Twitter is known as the first source to distribute information about breaking news and progressing events (Hu et al. 2012, p. 2751). Particularly, combined with mobile technology, Twitter plays important roles in disseminating disaster-related, time-sensitive information in a near real-time to citizens at risk (Hsu and Liao 2014). One such role is to request timely information of interest. A twitterer lets others know about certain urgent situations, and he/she can quickly receive relevant information in the matter of minutes (Covello et al. 2010, p. 145). Another role involves sharing emergency information during natural disasters. In the 2008 Sichuan earthquake (Li and Rao 2010), China experienced one of the largest earthquakes in their history. The magnitude was 7.9, and more than 67,000 people were reported to be killed or missing (pp. 1, 3). The overloaded cellular networks right after the disaster prevented them from obtaining time-critical information. Instead, the Chinese at risk from the disastrous event had to rely on messaging services such as Twitter to acquire information about the earthquake and to be aware of what was happening. As a matter of fact, the official report by the USGS (United States Geological Survey) was released 3 minutes after Twitter started disseminating information (p. 1). Similarly, in the 2011 Tohoku earthquake in Japan (Acar and Muraki 2011), a magnitude 9.0 earthquake caused catastrophic damages including the massive disruption of the phone and cell networks (p. 393). Right after the earthquake, Twitter turned out to be a critical communication medium that allowed people in the area to share information and to communicate with family and friends (Winn 2011). Lastly, emergency management officials leverage Twitter to spread disaster-related information immediately. During the 2013 Colorado flash floods, the Jefferson County Type II Incident Management Team (IMT) used various social media such as Blogs, Facebook, and Twitter to distribute the flood-related information to the public (St Denis et al. 2014, p. 741). The IMT stated that Twitter allowed the emergency messages to be quickly disseminated to a broad audience (p. 744). One testimonial illustrated that even without having press conferences, the messages crafted by the team were picked up and broadcasted by the local news agencies within minutes of posting. Along this line, in September 2013, Twitter launched *Twitter Alerts* in the U.S., Japan, and South Korea as a way to help online users to get timely information about emergencies and natural disasters (Pena 2013).

As shown in the above, timeliness is a quintessential aspect of disaster communication. Li and Rao (2010) stated that *"the initial hours following the disaster are the most important for emergency responders. Every single minute counts, since that is when lives will be saved and lost"* (p. 4). Due to the nature of highly dynamic, non-routine disaster events (Sellnow and Seeger 2013, p. 8), information loses its value far quicker than for any other type of event (Saunders and Pearlson 2009, p. 50). Accordingly, information can be considered to be valid only when it represents the up-to-the-minute state of constantly changing events. Likewise, Twitter has to be understood by its temporal aspects when it is used for disaster communications (Cotelo et al. 2014, p. 514). That is, under a situation where timely information has to be propagated as quickly and widely as possible to a threatened population, Twitter's temporal aspects could be represented by the extent to which a tweet is disseminated immediately after its posting. By taking this point of view, we argue that a shorter period of time for establishing the relationship between tweets and their retweets has to be deployed, because information conveyed by tweets during disasters would be highly volatile (Wilensky 2014, p. 705). In other words, Twitter features have to estimated only when tweets reliably represent evolving events. Therefore, we suppose that studying a tweet's initial propagation right after its posting should be of importance for enhancing our understanding of the use of Twitter features during disasters.

In order to consolidate our argument about the initial propagation of tweets, other factors are also taken into consideration. While the 140-character length limit contributes to rapid dissemination of tweets, such a short length could restrict the amount of information delivered to other twitterers. Particularly, as twitterers try to include multiple topics in a single tweet, its message clarity will decrease. That is to say, such a short message could not provide all the pertinent information for topics, leading to decreasing a tweet's overall meaning. Therefore, the recipients of the tweet may be in need of addressing such unclear meaning by seeking further information, affecting its initial retweet speed and scale (or its initial propagation). In addition, unlike words and hashtags, Twitter URLs are designed to deliver rich, external information that words and hashtags cannot convey (Hughes and Palen 2009, p. 9; Kostkova et al. 2014, p. 8:7; Sutton et al. 2014b, p. 6). However, twitterers pay specially attention to processing content such as lengthy news articles, complex maps, and/or time-consuming audio and video linked by embedded URLs. In that sense, Twitter URLs can be considered as a double-edged sword. That is, such URLs are important in carrying rich information to twitterers, while requiring them to make significant efforts to interpret linked information. Therefore, taking message clarity and Twitter

URLs into account should enrich our understanding on the initial propagation of tweets. As a result, we pose the overarching research questions as follows:

RQ1: How do we measure the initial propagation of tweets?

RQ2: How does the message clarity of tweets affect the initial propagation of tweets?

RQ3: How does additional information influence the relationship between the message clarity of tweets and the initial propagation of tweets?

The study makes the following contribution. For researchers, based upon exploring an issue about the initial propagation of emergency information, an index to estimate the propagation is suggested with empirical evidence. Therefore, it is fruitful to investigate factors that could affect the initial propagation of information in disasters. For communication participants about disasters on Twitter, this study suggests guidelines about the use of Twitter URLs as a means to provide in-depth, additional emergency information to the public at risk.

In the following sections, we develop the arguments that will inform the context of this study. First, we review the recent literature of Twitter and disaster communication. Then, we discuss the importance of emergency information's timeliness. Next, we propose a new index to measure such timeliness. Finally, we conclude by discussing the empirical findings and their implications for future research.

3.2 Research Background

3.2.1 Twitter

Twitter is a microblogging service that provides users the capability to post, exchange, and forward short messages of 140 characters or fewer (i.e., tweet). Currently, Twitter has over 284 million active users, with approximately 500 million tweets posted daily (Twitter 2015). Twitter provides several features that are similar to other social media platforms (i.e., Facebook, Linkedin, and Instagram) such as creating social connections between users and updating one's current status. Since its inception in 2006, Twitter has continuously added features to support user-driven linguistic conventions, which include hashtagging, mentioning other twitterers, and retweeting (Starbird and Palen 2010, p. 2). Hashtags provide a means of improving the user's understanding of a tweet's context, and makes tweets that are relevant to a topic discoverable by tagging or marking keywords within a message with the "#" symbol (Bruns and Stieglitz 2012, p. 164). Often, these keywords with hash symbols are appended at the end of tweets, or keywords in tweets are prefixed with the hash symbol (to maximize the limited length of characters). User designation, mentioning others, refers to adding the "@" symbol in front of the user name within a tweet, which allows the sender to post the tweet to a specific user. Retweeting or re-posting enables the propagation of information to a much broader audience, as the initial message is rebroadcast to the network of subscribers who retweeted the original message.

Factors that make Twitter such an viable communication medium over traditional media are the users' ability to post information in a near real-time basis and to easily redistribute information to a target audience through their follower networks (Liu et al. 2012). These features have made Twitter a dependable platform for disseminating time-sensitive information in a variety of different contexts (Bruns and Stieglitz 2012, p. 163; Fraustino et al. 2012a, p. 12). For this reason, Twitter has been gaining significant attention from the government, emergency organizations, and the general public as an effective tool for disseminating emergency information during natural disasters (Cheong and Cheong 2011, p. 2; Kongthon et al. 2012, p. 2227).

3.2.2 Timeliness of Retweeting during Disasters

Communication is a purpose-driven process (Shannon 1949, p. 5; Stephens and Barrett 2014, p. 3). Within a short time-span, emergency management officials and citizen journalists have an essential communicative goal to spread disaster-related messages to as many people as possible in affected areas (Sutton et al. 2015b, p. 5). As messages are disseminated widely, the intended-target population exposed the messages increases as well (Sutton et al. 2014a, p. 613). In this context, message amplification or retransmission is one of the efficient ways to reach a much wider target population, and Twitter provides an effective network for such amplification known as retweeting. Retweeting is an act of re-posting an original tweet. Twitter's retweet functionality allows sharing an original tweet with other twitterers (Compston 2014) when its information is considered to be interesting, useful, and/or imperative for others (Abdullah et al. 2014, p. 364; Starbird and Palen 2010, p. 3; Sutton et al. 2014b, p. 766; Zubiaga et al. 2015, p. 2). That is, retweeting is a means for twitterers to quickly share information that is deemed to be noteworthy to their followers (Liu et al. 2012, p. 445).

For disaster communication, such information sharing is an integral aspect for the public in disaster-stricken areas to increase situational awareness and guide them to be safe. Disaster alerts and warnings provide critical information about the nature and possible effects of approaching disaster events to people at risk who need to make sense of them (Mileti and Sorensen 1990, pp. 2-9). Therefore, they have to obtain relevant alerts and warnings about the threat before it strikes, in order to reduce the time taken for preparing protective actions accordingly (Mileti and Sorensen 1990, pp. 2-9, 3-11). In fact, Vieweg et al. (2010) revealed that tweets holding situational update information were more retweeted than other tweets (p. 1086). However, periods between warnings of disastrous events and their actual impacts may be very short, and thus affected people are required to make judgements based upon what they acquire in such short periods. Hermann (1963) also pointed out that an unexpected, highly dangerous situation allows the public at risk only a limited amount of time for a response (p. 64). That is, a terrifying sense of urgency resulting from such a highly uncertain incident requires people make snap judgement (Stein 2004, p. 1245). From this perspective, we have to emphasize two implications: first, most of disaster-related information has its value only before a specific event happens, otherwise it is no longer considered useful or relevant to the event regardless of its quality or accuracy; second, we speculate that the more timely information a tweet carries, the faster the tweet can be retweeted after its posting. Therefore, for examining Twitter as a means for disaster communication, the propagation of tweets immediately after the happening of an event will have more important implications than that made later on as the event precedes.

3.2.3 A Measure for the Initial Propagation of Tweets

Twitter allows the public to share first-hand observations about ongoing developments of emergency events by rapidly updating their status (or tweets), and for this reason, it seems to have a better fit to the information needs of the public at risk than other traditional and social media due to the following: first, the short message length and the follower network enable tweets to be propagated into twitterers' communities at an unprecedented speed and scale; second, retweeted tweets convey up-to-the-minute situational information to the public who needs to enhance situational awareness to make sense of surroundings; third, emergency information has to be most recent to reflect highly dynamic and unexpected nature of disaster events (Sellnow and Seeger 2013, p. 7), as affected people need the latest information in order to take appropriate protective actions as quickly as possible for avoiding or minimizing possible serious consequences. Otherwise, the information has to be considered obsolete or outdated. From the points of view above, we argue that the temporal aspect of disaster events and emergency information must be taken into account when estimating Twitter features in disaster communication. That is, studying Twitter features without considering such a temporal aspect could not capture their true effects. Here are two cases that we need to think about.

Case 1: Let's assume two tweets, *i* and *j*. Tweet *i* was retweeted at 2, 3, and 4 minutes after its posting. So, its total retweets and average retweet time in minute are 3 and 3 respectively. Tweet *j* was retweeted at 8, 9, and 10 minutes after its posting. Its total retweet count is 3, but the average retweet time is 9, 6 minutes later than that of Tweet *i*. In terms of the retweet total, both tweets looked the same, albeit Tweet *i* was retweeted much faster than the other tweet. The total retweet frequencies cannot capture such a difference between the two tweets. The average retweet time only represents the different speeds of retweeting between the two.

Case 2: There are three tweets, *i*, *j*, and *k*; each of which was retweeted 10, 20, and 30 times during the first 1 minute after posting. Although they have the same average retweet times, their retweet frequencies may significantly vary. If someone evaluates these tweets by the average retweet time, he/she will conclude no difference at all among the three tweets. However, the retweet total tells us the true difference among them. Both cases clearly demonstrate that considering either the average retweet time or the retweet total cannot fully reflect the relationship between tweets and retweets. This issue will be more severe as we investigate the use of Twitter during disasters when the rapid dissemination of timely information is considered critical.

Table 3.1 summarizes recent research studies about the use of Twitter in disasters. The studies investigated important phenomena for disaster communication. Zeng et al. (2016) and Oh et al. (2013) examined psychological factors and Twitter features to explain rumor-spreading during times of disaster by leveraging the median retweet time and the retweet likelihood for their respective dependent variables. Spiro et al. (2012) factored Twitter features into a waiting time measured by an average time difference between a tweet and its retweets. Sutton's series of studies looked deeply into how message content and style elements affected the retransmission of tweets (Sutton et al. 2015a; Sutton et al. 2015b; Sutton et al. 2014b). We also found a few empirical studies on Twitter and its use during disasters. Burnap et al. (2014) studied the information flow of tweets about a terrorist event. The authors summed up the number of seconds passed between a tweet and its first five retweets and used it as an independent variable for predicting the total retweet frequency (p. 2). Pervin et al. (2014)'s study is about factors that could affect a tweet's retweets in event-centric situations such as the 2013 Boston Marathon Bombing and the 2011 Great Eastern Japan Earthquake. Even though the total retweets of tweets were tallied per minute, the authors' purpose was about examining the impact of previous events on the current retweets (p. 6), which does not reflect the issues that we raised through Case 1 and 2. All in all, we could not find any empirical research that addresses the notion of obsolete information and as such takes retweet speed and size into account for analysis.

[31] Table 3.1 Studies on Twitter and Disaster Communication

Authors	Purpose	Case(s)	Dependent Variable	Tweet-Retweet Interval	Relevancy to Our Study
Zeng et al. (2016)	To explore whether rumors are transmitted faster than crisis information	Man-made Disaster • The 2014 Gunman Event in Sydney, Australia	Retweet time in median	Not considered	URL: +
Sutton et al. (2015b)	To examine content and style elements in association with the retransmission of tweets	Man-made Disaster • The 2013 Boston Marathon Bombing	Retweet count	Not considered	URL: –
Sutton et al. (2015a)	To study the effect of local network, content, and style on the retransmission of tweets	Disasters • Terrorist attack • Wildfire • Blizzard • Hurricane • Flood	Retweet count	Not considered	URL: –
Sutton et al. (2014b)	To examine the relationship between content, style, public attention and the retransmission of tweets	Natural Disaster • The 2012 Waldo Canyon Wildfire	Retweet count	Not considered	URL: –
Burnap et al. (2014)	To research information flows as the propagation of information over time	Man-made Disaster • The 2013 Terrorist Event in Woolwich, London	Retweet count	Considered as independent variables	URL: –
Pervin et al. (2014)	To investigate factors that affect the retweets of tweets in event-centric situations	Disasters • The 2013 Boston Marathon Bombing • The 2011 Great Eastern Japan Earthquake	Retweet count	Retweets per minute	URL: – and + (Boston) URL: + (Japan)
Oh et al. (2013)	To study factors that influence rumor spreading in disaster	Disasters • The 2008 Mumbai Terrorist Attack • The 2010 Toyota Recalls • The 2012 Seattle Café Shooting	Retweet likelihood	Not considered	None
Spiro et al. (2012)	To test Twitter features that influence waiting time between the original tweets and retweets	Natural Disaster • Earthquake/Mudslide • Tornado	Retweet time in average	Not considered	URL: +

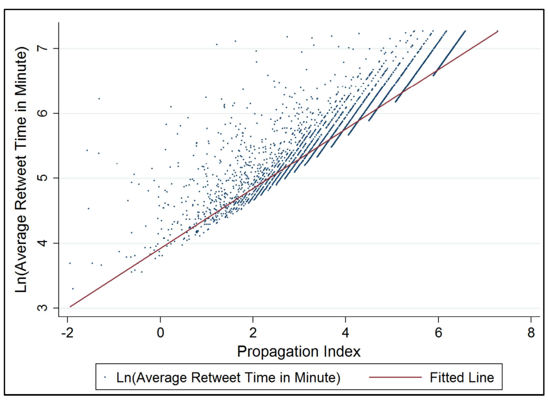
Therefore, we introduce the following measure, *Propagation Index (PI)*, that is designed to achieve our purpose:

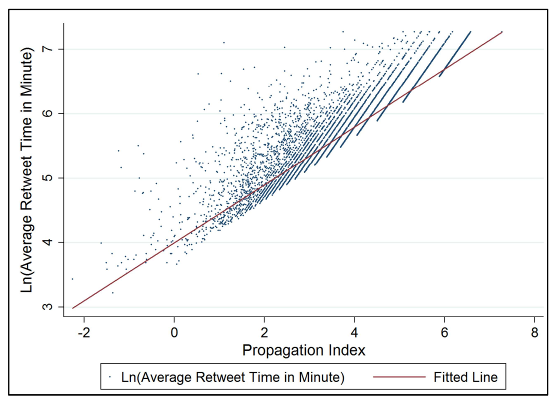
$$PI_{Tweet_i} = \frac{1}{N} \times \left(\frac{1}{N} \times \sum_{j=1}^{N} Time\_Diff(Tweet_i, Retweet_{i,j}) \right)$$

In this index, we use the following definition: $Tweet_i$ represents the original tweet; $Retweet_{i,j}$ is the jth retweet of $Tweet_i$; $Time\_Diff(Tweet_i, Retweet_{i,j})$ is the time lag, measured in minutes, between the original $Tweet_i$ and its jth $Retweet_{i,j}$; N is the total retweets of each original tweet.

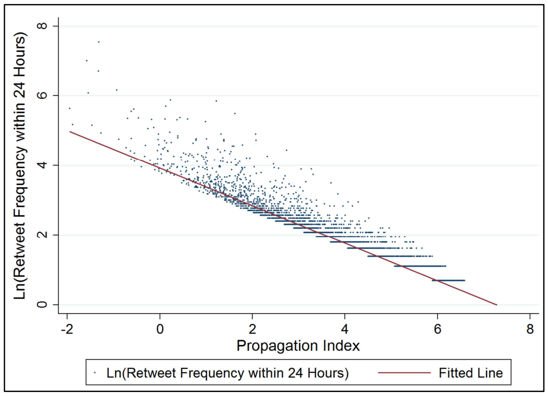
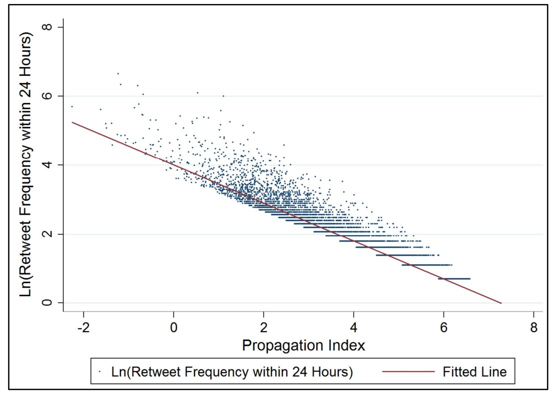
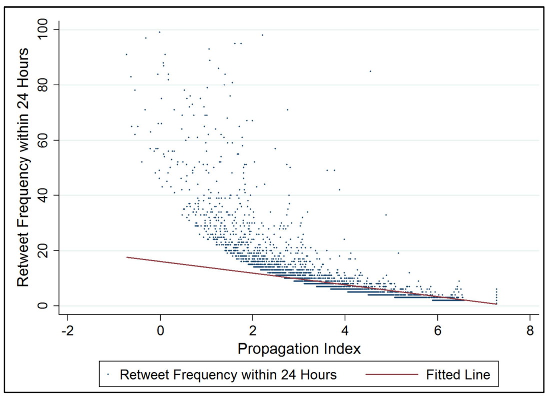
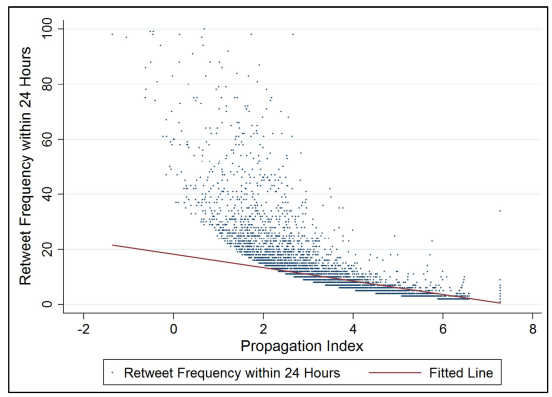
The above index (PI) resolves our concerns raised above: for the first case where the two tweets have the same quantity of retweets with a different average retweet time. The PI of *i* and *j* is 1 (i.e., 3/3) and 3 (i.e., 9/3) respectively. Although *i* and *j* have the same retweet frequency, the former were retweeted much earlier (or faster – 3 minutes on average) than the later (9 minutes on average) since each tweet's posting. Therefore, the index differentiates tweet *i* from *j* in terms of individuals' initial propagation. For the second case, the index of *i*, *j*, and *k* is 0.1 (i.e., 1/10), 0.05 (i.e., 1/20) and 0.033 (i.e., 1/30) respectively. That means, *k*'s initial propagation was the highest followed by *j* and then *i*. In fact, as the index decreases, the initial propagation of tweets increases. We conclude that unlike the retweet total and the retweet average time as measures, the PI successfully distinguishes tweets in the first and the second cases.

To provide more solid evidence on whether the PI well represents both aspects, we statistically examined it using two independent Twitter datasets. The first dataset contains tweets posted between January 8 and January 21 (two weeks) on the topic of the 2011 Queensland floods. The second dataset contains tweets disseminated from September 12 to September 25 (2 weeks) about the 2013 Colorado floods. In both datasets, tweets retweets were tallied only when retweets were posted within 24 hours of their original tweets. For the equation, we restricted both datasets to original tweets that were retweeted at least twice, and as such our sample contains 24,893 tweets for the Queensland flood incidents and 29,220 for the Colorado floods. We regressed the average retweet time and the retweet frequency (or count) on the PI respectively. A log transformation was applied on both the dependent variables – *the average retweet time* and *the retweet frequency* – and the PI in order to adjust highly skewed distributions of the above variables and thus to make the relationships more linear (Judd et al. 2011, p. 313; Lane 2016) (see Mode 1 and 2 of Figure 3.1). In addition to the log-transformed models, we also leveraged the negative binomial regression to estimate the effect of the log-transformed PI on the retweet frequency. We did this because the negative binomial distribution more accurately reflects counts data such as the retweet frequency (O'hara and Kotze 2010, p. 120) (see Model 3 of Figure 3.1).

<table>
<tr><th></th><th>2011 Queensland</th><th>2013 Colorado</th></tr>
<tr><td>Model 1. Ln(Average Retweet Time)</td><td>

R-squared=0.9330, $F_{1,\ 24891}$=45085.99, p<0.000</td><td>

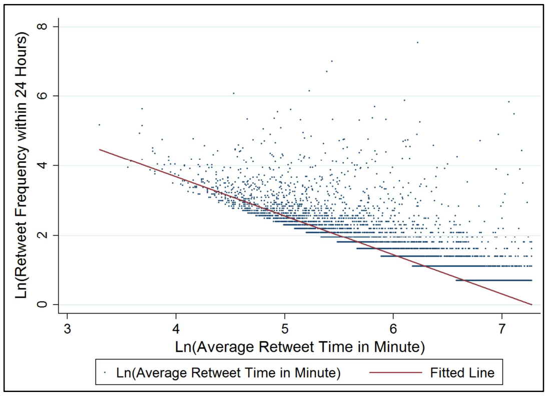
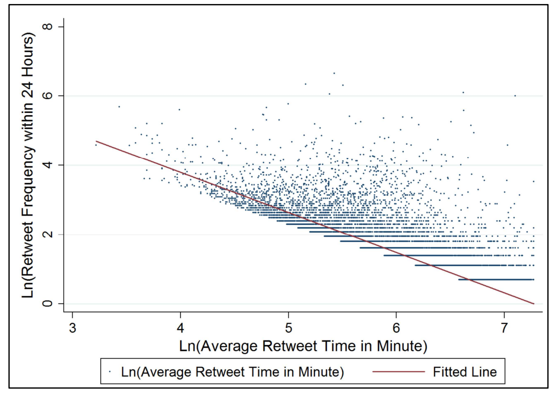
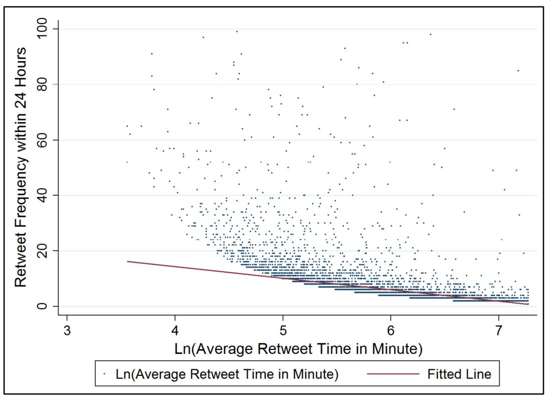
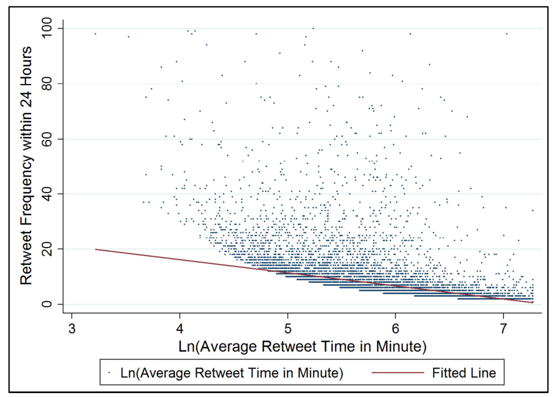
R-squared=0.9172, $F_{1,\ 29218}$=73660.70, p<0.000</td></tr>
</table>

Model 2. *Ln(Retweet Frequency)*	 R-squared=0.9253, $F_{1,\ 24891}$=53868.49, p<0.000	 R-squared=0.9100, $F_{1,\ 29218}$=83418.87, p<0.000
Model 3. *Retweet Frequency*[19]	 McFadden's Pseudo R-squared=0.331, Wald Chi^2=8362.18, p<0.000	 McFadden's Pseudo R-squared=0.324, Wald Chi^2=23529.49, p<0.000

[12] Figure 3.1 Relationships of PI on Retweet Time and Retweet Frequency

2011 Queensland	2013 Colorado

[19] For better visualizing the relationship between the retweet frequency and the PI, we only plotted tweets with less than 100 retweets. However, the models were tested based upon the full tweet datasets.

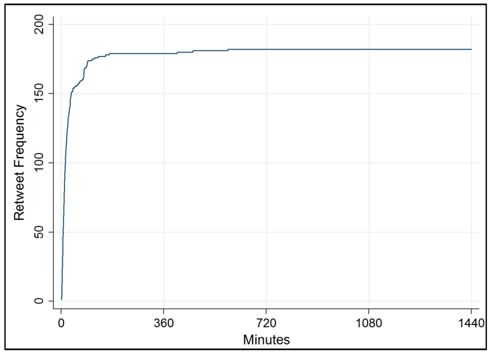
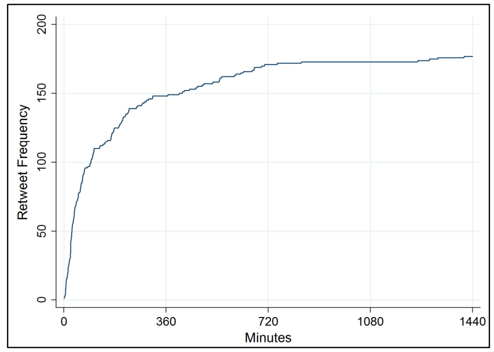
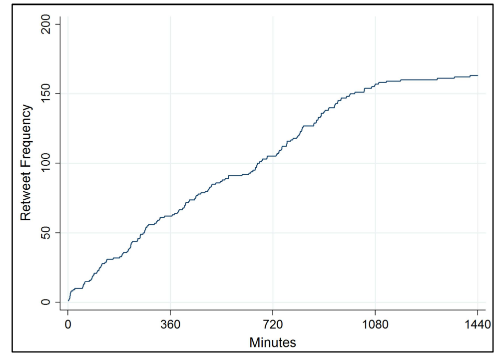
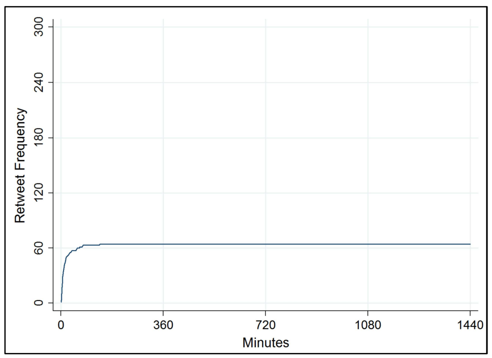
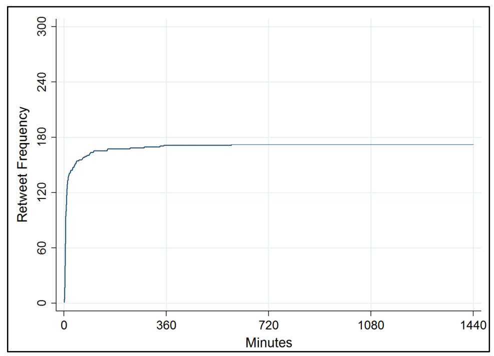
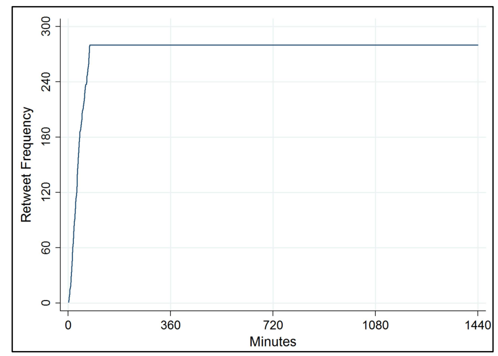
Model 1. *Ln(Retweet Frequency)*	 R-squared=0.7538, $F_{1,\,24891}$=30870.43, p<0.000	 R-squared=0.7038, $F_{1,\,29218}$=43252.37, p<0.000
Model 2. *Retweet Frequency*[20]	 McFadden's R-squared=0.159, Wald Chi^2=1319.65, p<0.000	 McFadden's R-squared=0.152, Wald Chi^2=2335.07, p<0.000

[13] Figure 3.2 Relationship between Retweet Time and Retweet Frequency

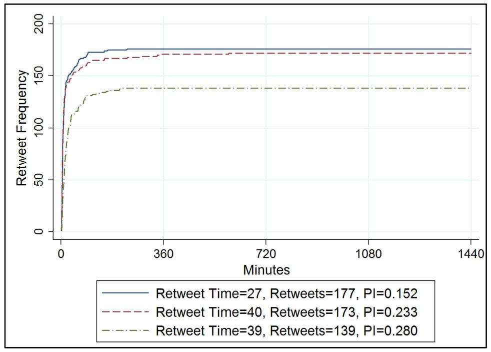
Model 1 of Figure 3.1 demonstrated that the PI well represents the average retweet time (Queensland – Coefficient=0.4335, R^2=0.933, $F_{1,\,24891}$=45085.99, p<0.000; Colorado – Coefficient=0.4234, R^2=0.9172, $F_{1,\,29218}$=73660.7, p<0.000). That is to say, as the PI increases, the average retweet time increases as well. Model 2 tested the relationship of the PI with the retweet frequency and showed that the PI is also highly related to the retweet frequency (Queensland – Coefficient=-0.5547, R^2=0.9253, $F_{1,\,24891}$=53868.49, p<0.000; Colorado – Coefficient=-0.5616,

[20] For better visualizing the relationship between the retweet frequency and the PI, we only plotted tweets with less than 100 retweets. However, the models were tested based upon the full tweet datasets.

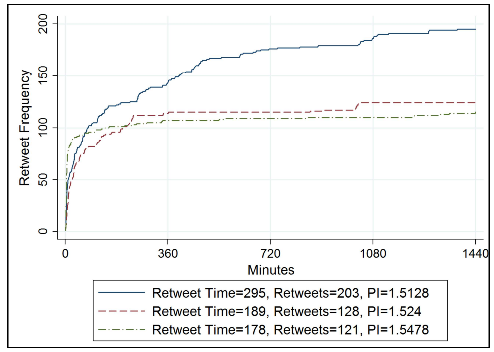
R^2=0.9172, $F_{1,2}$=73660.7, p<0.000). As the PI increases, the retweet frequency decreases. Interestingly, the devised index represents both the average retweet time and the total retweet frequency much better than the average retweet time explains the retweet frequency (Queensland – R^2=0.7538, $F_{1, 24891}$=30870.43, p<0.000; Colorado – R^2=0.7038, $F_{1,2}$=43252.37, p<0.000). As mentioned above, we also examined the relationship between the PI and the retweet frequency based upon a negative binomial regression, in which the estimation lent support to the previous results (see Model 3 and Model 2 of Figure 3.1) (Queensland – Pseudo R^2=0.331 vs R^2=0.159; Colorado – Pseudo R^2=0.324 vs R^2=0.152). Overall, a series of the statistical analyses demonstrated that the index of tweets' initial propagation (PI) well indicates both the average retweet time and the total retweet frequency at the same time. If so, we need to confirm that how well the PI does describe the initial propagation of tweets. Three examples were prepared (see Figure 3.3): for the first example, three tweets were compared, each of which had similar total retweets of 181, 176, 162 in order, but their average retweet times were 46, 176, and 162 minutes. Example 1 of Figure 3.3 depicts different patterns of the initial propagations of the three. The first tweet with its PI of 0.254 (181 retweets) was disseminated much quicker than the other two whose PIs were 1.64 and 4.129 correspondingly; for the second example, we intentionally chose three tweets that had the same average retweet time of 40 minutes, but the number of different retweets of 64, 172, and 280 in order to check whether the PI distinguishes a situation that tweets have the same average retweet time with the number of different total retweets. Example 2 of Figure 3.3 demonstrates that as the index decreases, the initial propagation soars up; the last example comprehensively shows the patterns of the initial propagation according to the PI scores ranging from 0.1, 1.5 to 3.0. In fact, Example 3 supports what we argued in Example 1 and 2. Consequently, it seems that the PI addresses the concerns raised by the early two cases and measures well the initial propagation of tweets. As a result, we make the following statement with confidence that the lower PI a tweet produces, the higher initial propagation it will show.

Example 1	 Avg. Retweet Time=46, Retweets=**181**, Propagation Index=**0.254** *(http://twitter.com/CUBoulderPolice/statuses/378369246272966656)*	 Avg. Retweet Time=290, Retweets=**176**, Propagation Index=**1.647** *(http://twitter.com/jeffjames3/statuses/378540653304508416)*	 Avg. Retweet Time=669, Retweets=**162**, Propagation Index=**4.129** *(http://twitter.com/JaymeMoye/statuses/378154652312686592)*
Example 2	 Avg. Retweet Time=**40**, Retweets=64, Propagation Index=**0.580** *(http://twitter.com/QPSmedia/statuses/25056946214150144)*	 Avg. Retweet Time=**40**, Retweets=172, Propagation Index=**0.232** *(http://twitter.com/QPSmedia/statuses/24674000789569536)*	 Avg. Retweet Time=**40**, Retweets=280, Propagation Index= **0.142** *(http://twitter.com/QLDFLOODRT/statuses/24772849948426240)*

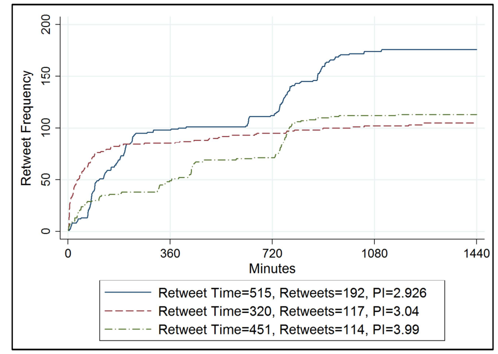
Example 3



Propagation Indexes around ***0.1~0.3***



Propagation Indexes around ***1.5***

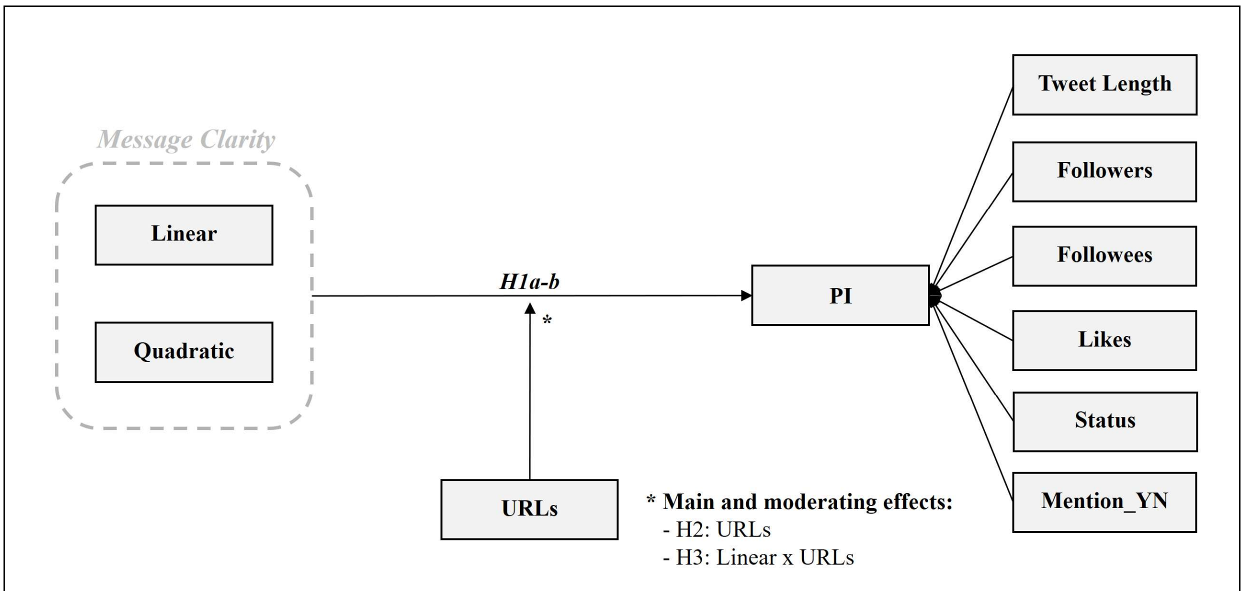


Propagation Indexes around **2.9~*4.0***

[14] Figure 3.3 Initial Tweet Propagation within 24 Hours

3.3 Hypothesis Development

When disasters strike, situational uncertainty increases. Accordingly, the public in disaster-stricken regions will be motivated to seek information for reducing such uncertainty (Spencer and Hiltz 2003, p. 657). Social media immediately provides near real-time (Fraustino et al. 2012b, p. 14) and first-hand information (Bruns et al. 2012, p. 8) to the public much faster than traditional news media and even disaster response agencies (Fraustino et al. 2012b, p. 16). Above all, Twitter is a simple but effective means to relay, redistribute, and share short messages of interest to one's followers (Suh et al. 2010a, p. 177). Its retweet function allows information to rapidly reach to a targeted population (Kwak et al. 2010, p. 599).



[15] Figure 3.4 Research Model

One factor that must be taken into account in association with such an unprecedented speed and scale in disseminating information is the short length of tweets. Due to the 140-character limit, tweets can be disseminated over virtually all communication platforms including the Web, mobile devices, and even cellular phones (Starbird and Palen 2010, p. 2; Vieweg et al. 2010, p. 2). During any period of heightened stress, tweets can be received anywhere in which electronic signals are present (Covello et al. 2010, p. 145). Immediately after receiving a tweet about alerts or warnings about disaster events, the recipients interpret them in order to understand its conveyed meaning(s), to develop an individual's risk perceptions, and to determine whether these tweets are believable (Bean et al. 2016, p. 4). Message quality of the tweet or the extent to which the tweet is written clearly, accurately, unambiguously, and consistently influences understanding and believing (Bean et al. 2016, p. 8). Its message quality will affect the recipients' decisions on whether to seek additional information or to take actions – either retweeting or proceeding with protective movements. In this perspective, message quality is critical for decision-making, and it becomes more important for people under imminent threats with a terrifying sense of urgency. In fact, the brevity of tweets for emergencies is inherently associated with insufficient information (Bean et al. 2016, p. 6). In other words, information insufficiency or information dearth can negatively affect the quality of tweets, which is *"the gap between what people know about a given risk (current knowledge) and what they say they need to know for their own purposes (the sufficiency threshold)"* (Kahlor et al. 2006, p. 171). Given such insufficiency, information-hungry people (Mileti and Sorensen 1990, pp. 3-8) will actively search for further information to address what they find deficiency in tweets. We argue that information insufficiency negatively affects the message clarity of tweets, and thus as a tweet is considered to show low message clarity, the recipients would aggressively engage in milling by seeking additional information to confirm whether they correctly understand its meaning (Bean et al. 2016, p. 2) and to affirm the suitability of their protective behaviors (Wood et al. 2012, p. 605). Milling is defined as the widespread search for collecting information and for confirming messages by people who are in disaster-stricken areas (Lindell and Perry 1987, p. 138; Schneider 2014,

p. 55). For example, on recognizing the lack of the message clarity in received tweets, the recipients may wait and search for other relevant tweets, interact with others for clarification, or turn to more credible news media or disaster response organizations (Spiro et al. 2012, p. 5). That is, the recipients will instantly seek additional information to minimize such obscurity and validate whether their understandings are correctly formed (Fraustino et al. 2012b, p. 13). With the trustworthiness of the tweets, they will quickly share what they received with others. Without the believability, they may deter or give up retweeting. As a result, milling activities could delay or even impede retweeting, resulting in slowing down the initial propagation of tweets. However, the negative relationship would become weakened as message clarity lowers further. That is to say, the effect of message clarity on the initial propagation of tweets would be not constant. Therefore, our first and second hypotheses are as follows:

H1a. A decrease in a tweet's message clarity negatively influences its initial propagation, such that as message clarity lowers, the propagation index increases.

H1b. A decrease in a tweet's message clarity negatively influences its initial propagation, and such a negative effect is less strong as message clarity in a tweet further decreases.

The length limitation is a big hurdle for twitterers who are in need of delivering large amounts of information. One convention to overcome such an obstacle is to embed Twitter URLs, which consist of randomly selected letters and numbers (e.g., http://yfrog.com/hsi9sfj), in tweets. In general, it is not possible for twitterers to directly extract meanings from embedded URLs before visiting sites linked by them. In other words, Twitter URLs serve as pointers to the public news articles and multimedia content such as YouTube, Flickr, and Google Maps that could be vital or valuable to other twitterers (Hughes and Palen 2009, p. 8). External resources introduced by Twitter URLs can enrich information contained in tweets (Ma et al. 2013, p. 1403), but at the same time the recipients have to make significant efforts to process such linked information. In fact, the inclusion of Twitter URLs seems to be a double-edged sword in the sense that rich information delivered by Twitter URLs increases the situational awareness of the affected public to help them make better decisions. On the other hand, easily digestible information within a very short time interval is also critical for snap judgement on rapidly changing events. However, embedded Twitter URLs require the public to spend extra time to encode associated external information. In light of the two different aspects of Twitter URLs – *information value* and *processing efforts*, we need to address how these features could be understood in association with retweeting. To the best of our knowledge, most Twitter research on disaster communication treated embedded URLs in terms of information value (Burnap et al. 2014, p. 206; Pervin et al. 2014, p. 8; Sutton et al. 2014b, p. 775), rather than processing effort. During highly uncertain and evolving disaster events, the extra time to read and comprehend in-depth information provided by URLs hampers a tweet's retweet speed and scale right after its posting. Therefore, we pose the third hypothesis about Twitter URLs by factoring processing effort into the initial propagation.

H2. Twitter URLs embedded in a tweet negatively affect its initial propagation.

Hypotheses 1 and 2 aim to look into the relationship of message clarity with the initial propagation of tweets. Both hypotheses shed light on the time required for additional milling and for processing external information linked by URLs, and their effects on how quickly and widely a tweet is disseminated since its posting – that is, its initial propagation. We believe that although time for processing and gathering further information might deter the initial propagation of tweets, rich information embedded in tweets could weaken the assumed negative influence of decreases in message clarity on their initial propagation. By using relevant hashtags or keywords, twitterers can search the twitterverse to acquire more information (Spiro et al. 2013, p. 7). Or, they can simply navigate external resources linked by Twitter URLs (Bruns and Stieglitz 2012, p. 178; Hughes and Palen 2009, p. 8). Twitter URLs are an interesting convention that enables twitterers to include rich, in-depth information (Hughes and Palen 2009, p. 8; Purohit et al. 2013, p. 2439; Spiro et al. 2013, p. 3). Also, it seems that twitterers like to include URLs to make their tweets instructive (Bruns and Stieglitz 2012, p. 178; Ma et al. 2013, p. 1400), especially for disaster communication. In this sense, Hughes and Palen (2009) reported one interesting finding that roughly 50% of the tweets about a hurricane event included Twitter URLs, 10% higher than that of the tweets about general events of interest. They considered Twitter URLs as an important means to overcome the information dearth of tweets in disasters. We claim that when a tweet presents low message clarity, its embedded URLs could complement information for the tweet, resulting in relieving the milling process of the recipients. That is, rather than searching the twitterverse or other media sources, twitterers who received tweets with low message clarity will process linked information by Twitter URLs, decreasing the strength of the relationship between message clarity and the initial propagation. Therefore, our last hypothesis states the moderating effect of Twitter URLs on the relationship.

H3. The negative relationship of message clarity with the initial propagation of tweets depends on Twitter URLs, such that the negativity becomes less strong as the number of Twitter URLs increases.

As a summary, Figure 3.4 depicts the relationships between the PI and the independent variables, as the all four hypotheses illustrated.

3.4 Data and Methods

3.4.1 Two Flood-related Natural Disasters

2013 Colorado floods: The 2013 Colorado floods caused a great deal of heartache and economic difficulties to Coloradans. A series of floods started on September 9, 2013 and lasted for 7 days while pouring 15 to 20 inches of rain in the Front Range area including Boulder, Colorado Springs, and Fort Collins. Boulder County was hit most with its five days' rainfall exceeding its annual average of 20.7 inches. Fourteen counties in Colorado declared disaster emergencies with more than 11,000 residents evacuated. The U.S. Army and the Colorado National Guard rescued 1,750 residents along with 300 animals and pets (Connor et al. 2013). Eight people were found to be dead and five were missing, according to the Federal Emergency Management Agency (FEMA) and the Colorado Office of Emergency Management (COEM) (Gochis et al. 2014). The economic damages were also substantial. Nearly 20,500 homes were damaged or destroyed and almost 50 state highway bridges were identified as needing repairs by the Colorado Department of Transportation (Gochis et al. 2014). Immediately following the initial warnings by FEMA and the National Weather Services, people around several affected areas and remote areas started producing, sharing, and disseminating diverse flood-related information on Twitter.

During the floods, Project EPIC, hosted by the Department of Computer Science at the University of Colorado Boulder, collected tweets and their retweets about the flood events in near-real time by leveraging its data analytic infrastructure (Anderson and Schram 2011). Based upon the infrastructure, the research group was able to systematically retrieve relevant tweets and twitterers by incrementally adding keywords, hashtags, and twitterers (Dashti et al. 2014a, p. 4). 102,426 original tweets and 122,276 retweets produced by 77,774 unique twitterers were collected between September 12 and September 25, 2013.

2011 Queensland: From late December 2010 to January 2011, a series of floods hit much of the central and southern parts of Australia including Queensland, which is Australia's second largest state. Queensland experienced the most intense flooding between January 10 and 16 (Shaw et al. 2013, p. 7). The floods affected more than 200,000 residents living across 90 towns and caused A$2.38 billion of damage, resulting in 38 casualties (Wikipedia 2016). Twitter was one of the crucial media outlets to disseminate and share emergency information, and its users (twitterers) played an important role to amplify the information (Bruns et al. 2012, p. 7) in a way that expands its reach. We were able to obtain Twitter data directly from one of the Twitter branches. By following the way to gather tweets about the 2013 Colorado floods, One Twitter branch followed the same procedures that Project EPIC utilized in order to identify and retrieve tweets, retweets, and twitterers' information on the 2011 Queensland floods. The retrieved tweets revealed that during January 8, 2011 to 21, 33,565 unique twitterers generated and forwarded 109,456 tweets and 120,082 retweets. Table 3.2 summarises the two Twitter datasets.

For the hypothesis testing, we selected the original tweets that were retweeted at least twice, such that the average retweet time can be calculated for the propagation index.

[32] Table 3.2. Descriptive Statistics of Two Flood Incidents

Items \ Cases	2011 Queensland	2013 Colorado
Period of Data Collection (2 Weeks)	January 8 ~ 21, 2011	September 12 ~ 25, 2013
Total Tweets	109,456	102,426
- Total Tweets w/ More Than 1 Retweet	24,893	29,110
Total Retweets	120,082	122,276
Unique Twitterers	33,565	77,774

3.4.2 Methods

Along with statistical procedures, a series of analytical techniques were employed to fulfil the research goals of the study. We utilized natural language processing techniques to analyze each tweet's unstructured message into its structured forms which include words and their part-of-speech tags, URLs, hashtags, and mentions. As the length of tweets are too short for modeling topics (Cataldi and Aufaure 2015, p. 576; Wang et al. 2007, p. 1), we extracted *n*-gram noun phrases based on part-of-speech tags and used them as additional input to model topics. That is, together with uni-gram words, bi- and tri-gram words, such as "flood victims," "colorado flood," and "higher ground,", and "flood relief appeal" were used to extract topics. Table 3.4 describes the top 5 n-gram noun phrases. To achieve this analysis, the following steps were leveraged: first, we tagged tweets' message components by leveraging *TweetNLP*'s programming library, which provides a tokenizer, a part-of-speech tagger, and a dependency parser for parsing tweets (Owoputi et al. 2013); second, we extracted tweets' topics by utilizing the Latent Dirichlet Allocation (LDA) model, which is a statistical method to summarize unstructured texts at a scale that might not be fulfilled by human commentators by automatically uncovering topics in a collection of text documents (Blei 2012, p. 78). As a result, The LDA model defines a topic by a distribution of words in documents, and as such multiple topics, i.e., different distributions of words, can exist (Blei 2012, p. 78). As outcomes, the LDA model produces topics in a set of documents and topic(s) per document (Blei 2012, p. 80). We employed a Machine Learning for LanguagE Toolkit (MALLET), a Java library for statistical natural language processing, to discover topics in tweets (McCallum 2002). Along with *n*-gram noun phrases, hashtags are an essential component to annotate individual tweets' conversation topic(s) (Boyd et al. 2010; Bruns and Stieglitz 2012; Laniado and Mika 2010; Ma et al. 2013; Yang et al. 2012). Therefore, we included hashtags when analyzing tweets' topics. However, we excluded embedded URLs comprised of random characters and numbers (e.g., http://yfrog.com/hsi9sfj), which are devoid of the topic information needed to find topics. Topic modeling is considered a clustering method in the sense that documents are grouped together based upon the similarity of topics (Blei 2012, p. 80). Accordingly, providing an optimal number of topics for the LDA will be critical to have topics that best represent target documents. To accomplish this goal, we generated topic models by increasing the number of topics from 2 to 200, calculated each topic model's goodness of fit, and evaluated the generalizability of each topic model in terms of its perplexity score (Blei et al. 2003, p. 1008), where *M* refers to the number of documents in the testing dataset, w_d refers to the words in document d, and N_d refers to the number of words in document d.

$$\text{Perplexity}\ (\mathrm{D_{test}}) = \exp\{\frac{\sum_{d=1}^{M} \log \mathrm{p}(\mathrm{w_d})}{\sum_{d=1}^{M} \mathrm{N_d}}\}$$

Each model's generalizability is inversely related to its perplexity score – the lower the score, the higher the generalizability. By sequentially ordering the perplexity scores by topic quantity, we applied the cumulative sum (CUSUM) procedure (Ellaway 1978) to each Twitter dataset to find an optimal topic quantity at which the changes in the perplexity score are negligible, indicating that additional topics would offer no significant benefits to generalizability. Appendix 3.B shows that as the quantity of topics increases, a series of perplexity scores and their moving ranges for the two flood incidents decrease. As a result, 72 and 57 topics were chosen as the optimal topic quantities for the tweets from the 2011 Queensland and the 2013 Colorado floods repectively (see Appendix 3.A).

3.4.3 Message Clarity

we contend that the retweeting of said tweets could be affected by varying degrees of their message clarity and that such clarity would be associated with the number of topics. That is, the greater number of topics a tweet bears, the less message clarity it may represent. In other words, as twitterers craft a tweet with multiple topics, information per topic of this tweet inevitably decreases, negatively affecting its message clarity as a whole. For example, when one tweet summarizes three topics and another tweet describes only one topic, the former is considered to have less information per topic than the latter. We argue that the three topics conveyed in the former tweet will be less clear than the one topic in the latter, primarily because the three topics have to be explicated within a range of 140 characters. For the latter tweet, 140 characters are available solely for its one topic. All in all, a single tweet aimed at carrying more topics may inevitably convey fewer specific details per topic (Bruns et al. 2012) and provide less sufficient information for the main topic (Mileti and Peek 2000) than another tweet conveying fewer topics. Because of tweet's brevity, the public may expect a single tweet not to hold diverse topics.

Shannon and Weaver already studied the clarity of a message, which he defined as noise—"a measure of one's freedom of choice in selecting a message"—and contended that if noise is present in a message, this message is assumed to contain some degree of distortions and errors, thus increasing the uncertainty of the message (Shannon 1949). They proposed the following equation to quantify noise in a message, where p_i is the proportion of the *i*th topic out of n topics of a message m.

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

According to Shannon and Weaver's information theory, a single message with two topics is harder to interpret than a message with only one topic. If two topics arise with equal proportion in a message, the recipients can interpret the message as being about either topic, meaning that it is noisy or unclear (entropy of greater than 0). Another possibility would be that if one topic exists with high proportion of almost 1, the proportion of the other topic should be 0.[21] Then, the entropy of the message becomes 0, which indicates that there is almost no chance of the message being interpreted as being about the second topic. This is to say that the message scarcely contains noise, and thus its clarity is highly assured. To take concrete instances, we consider the following three actual tweets about the warnings and alerts of flooding in the Boulder areas.

*Tweet 1: "If flooding is occurring or is expected, get to higher ground quickly. Remember *Turn Around, Don't Drown* #Boulderflood #coflood"*

Tweet 2: "I'm heading to <u>bed</u> now. Everyone in #boulder please be <u>safe</u>. Get to higher ground if possible. #boulderflood"

Tweet 3: "Man it's #<u>biblical</u> #cofloods #boulderfloods #southplatte lets <u>hope</u> the people made it to higher ground! http://t.co/5tRlSPOJaP"

[21] p_1 is the proportion of the first topic, and thus $1-p_1$ is for the second topic.

Although all three tweets contained the message about an urgent situation that recommended the affected public seek higher ground for ensuring safety, Tweet 1 presented the urgent situation the most clearly of the three. According to the results of topic modeling, it turned out that Tweet 1 depicted urgency, with the topic proportion of 98.9% (*Urgency*—i.e., higher ground, see Table 3.1 for a list of topics), and therefore its entropy was 0.02 (very close to 0). However, Tweets 2 and 3 showed mixed topics. In Tweet 2, 65.53% of the message was about the urgent situation (*Urgency*) and 32.91% about relief (*Relief*—i.e., please be safe), and as such its entropy was 0.64, which is much higher than that of Tweet 1. Tweet 3 was packed with 3 topics: 49.03% about the urgency (*Urgency*), 37.46% about the relief (*Relief*—i.e., hope), and 12.36% about the current floods (*Current Flood*—i.e., #biblical). As expected, the entropy was highest for Tweet 3 at 0.9756.

[33] Table 3.3 Three Topics Corresponding to the Three Tweets [22]

Topic Labels	Tokens per Topic by Importance
Urgency (#46)	canyon, boulder, water, ground, higher, **higher ground**, wall, coming, boulder canyon, creek, **immediately**, **move**, boulder creek, gulch, emerson gulch, emerson, seek, debris, pearl, vehicles
Relief (#50)	**safe**, boulder, stay, rain, friends, prayers, thoughts, people, **hope**, affected, home, good, dry, family, love, raining, bad, crazy, victims, house, news, #longmontflood, praying, water, god, work, #prayforcolorado, live, staying, morning, coming, stop, heart, pray
Floods and Damage (#28)	damage, photos, aerial, images, flood damage, video, **biblical**, climate, line, trends, boulder, climate trends, views, **biblical** flood, show, waters, aerial views, lyons, shot, flood waters

We speculate that Tweet 3 could confuse the recipients more than Tweets 1 and 2 in the sense that Tweet 3 provides less information about the urgency than Tweet 1 and 2, but it presents the other topics –*Relief* and *Floods and Damage*. In other words, twitterers who received Tweets 1 and 3 will struggle

[22] Three topics out of 57 were shown to explain the examples. Each topic was labeled based upon the interpretation on a list of tokens that are ordered by importance.

more to understand the intent of the latter than that of former as Tweet 3 gives its recipients less information about its main topic – *Urgency*. Consequently, the twitterers will evaluate each tweet as a whole, finding that the message clarity of Tweet 1 is much higher than that of Tweet 3 in terms of the emergency alerts. All in all, a single tweet trying to carry more topics may inevitably maintain lower consistency among topics in its message (Mileti and Peek 2000, p. 187), convey less specific (Bruns et al. 2012, p. 44), and provide less sufficient information for the main topic (Mileti and Peek 2000, p. 188) than another tweet conveying fewer topics. Having agreed on the function of the entropy measure, we believe that a tweet's message clarity could be measured in terms of its entropy or its topic quantity. In sum, we claim that because of their short length, individual tweets' message clarity can be quantifiable in terms of each tweet's topic quantity; additionally, a tweet's message clarity is negatively related to its topic quantity, and as a tweet's message clarity decreases, its credibility decreases as well.

3.4.4 Statistical Analysis

The dependent variable of the study is the propagation index (PI) that measures a tweet's propagation speed and size at the same time. Especially, to better represent normality between the dependent and the independent variables (Judd et al. 2011, p. 312; Keene 1995, p. 813), we performed a log transformation. The log-transformed PI of each original tweet was produced based upon its retweet frequency posted within 24 hours since its posting. As already shown in the earlier section, the PI well represents the extent to which a tweet is propagated in terms of its retweet speed and scale.

Previous research on Twitter under times of disaster has studied factors that affect retweeting in terms of the content features of tweets, such as the length of tweets, hashtags, and URLs and twitterers' features of followers, followees, likes, and status updates (total # of tweets posted by a twitterer). Zeng et al. (2016)'s rumor research revealed that whether a tweet included other twitterers' screen names (@mention) did not affect retweeting in terms of waiting times defined by time elapsed between a tweet and its retweets (p. 1975). Sutton et al. (2014b)'s empirical study on tweets about the 2012 Waldo Canyon wildfire reported that while followers and followees were positively related to the retweet rate, Twitter URLs decreased the rate (p. 779). Sutton et al. (2015a)'s subsequent empirical study reliably showed that @mention was negatively associated with the retweet total across five different disaster-related tweets and that three datasets out of the five supported the positive effect of followers on retweet counts (p. 14796). Similar to Zeng et al. (2016)'s study, Spiro et al. (2012) modeled elapsed time between the original tweets and their retweets. Their empirical results indicated that @mention, followers, and followees negatively affected retweet speed, but status updates positively influenced retweet speed (p. 7). The study of Burnap et al. (2014) on information flows during a terror attack found a significantly positive relationship between followers and retweet counts, but the effect of followees was not reliable (p. 9). In addition to the already known variables, the length of tweets and a binary indicator about whether a tweet is retweeted within 1 minute after its posting was included. Particularly, the tweet length controls the effect of message clarity on the PI for the following two cases: case 1 – a tweet represents one topic, while its character length is much fewer than 140; case 2 –another tweet conveys one topic by fully leveraging 140 characters. In order to take such cases into account, the total length of each tweet was leveraged in our empirical model. Like many other studies (Spiro et al. 2012, p. 6; Sutton et al. 2015a, p. 14796), this study also performed a log transformation on the variables that are related with twitterers, such as followers, followees, and status updates were skewed on the right. Table 3.4 summarizes the descriptive statistics of the dependent, control, and exploratory variables.

[34] Table 3.4 Variable Description							
Variable Name	**Cases** / **Explanation**	**2011 Queensland**			**2013 Colorado**		
		Mean	**S.D.**	**Range**	**Mean**	**S.D.**	**Range**
Dependent Variable							
$Ln(Propagation\_Index_i)$	The log-transformed propagation index of tweet *i* based upon its retweets made within 24 hours after its posting	5.24	1.08	-1.95-7.27	5.032	1.20	-2.25-7.27
Message Clarity (Entropy)							
$Linear_i$	The linear relationship between the dependent variable and the entropy of tweet *i*						
$Quadratic_i$	The quadratic relationship between the dependent variable and the entropy of tweet *i*						
Additional Information							
$URLs_i$	The number of URLs in tweet *i*	0.51	0.56	0-4	0.64	0.55	0-4
Message Clarity × URLs							
$Linear_i$ x $URLs_i$	Moderation between *Linear x URLs* to examine information value of URLs over and above other information.						
Control Variables							
$First\_Retweet\_1m\_YN_i$	Whether tweet *i* is retweeted within 1 minute after its posting – 1 for 'Yes' and -1 for 'No'						
$Tweet\ Length_i$	The character length of tweet *i* except embedded URLs	87.45	23.1	10-128	87.70	22.9	0-132
$Ln(Followers_{i,t})$	The log-transformed number of followers of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\sum_{\substack{\text{Join date of} \\ \text{the author of tweet } i \\ \le t \le \\ \text{Date of posting tweet } i}} Follower_{i,t}\right)$	6.29	1.88	0-15.1	7.42	2.16	0-16.4
$Ln(Followees_{i,t})$	The log-transformed number of followees of tweet *i*'s author between his/her join date and the date of posting tweet *i*	5.74	1.67	0-12.1	6.56	1.68	0-12.6

	$Ln\left(\sum_{\substack{Join\ date\ of\\ the\ author\ of\ tweet\ i\\ \le t \le\\ Date\ of\ posting\ tweet\ i}} Followee_{i,t}\right)$						
$\mathrm{Ln}(\mathrm{Likes}_{i,t})$	The log-transformed number of likes of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\sum_{\substack{Join\ date\ of\\ the\ author\ of\ tweet\ i\\ \le t \le\\ Date\ of\ posting\ tweet\ i}} Like_{i,t}\right)$	1.94	2.01	0-8.58	3.95	2.46	0-13.6
$\mathrm{Ln}(\mathrm{Status}_{i,t})$	The log-transformed number of tweets of tweet *i*'s author between his/her join date and the date of posting tweet *i* $Ln\left(\underbrace{\sum_{\substack{Join\ date\ of\\ the\ author\ of\ tweet\ i\\ \le t \le\\ Date\ of\ posting\ tweet\ i}} Tweet_{i,t}}_{Status}\right)$	7.84	1.89	0-12.4	8.47	1.96	0.69-14.1
$\mathrm{Mention\_YN}_i$	Whether tweet *i* contains other twitterers' screen name – 1 for 'Yes' and -1 for 'No'						

The regression analysis was performed to test our hypotheses by examining the relationships between the log-transformed PI and the predictors including control variables (see Figure 3.4). Despite relatively high correlations among the twitterer-related predictors such as followers, likes, friends, and status updates (see Table 3.5), in neither case did the VIF exceed 2.71 (Queensland – Mean VIF=1.41; Colorado – Mean VIF=1.51), which is well below the acceptable level of 5 (David A. Belsley 2005) (see Appendix 3.C), indicating that the proposed empirical model did not have significant signs of a multicollinearity problem. We confirmed the over-dispersion of our data and the existence of the heteroscedasticity of variance (Breusch and Pagan 1979). Therefore, a robust regression procedure was employed to estimate the empirical model. Before assessing the moderation hypotheses, all numerical variables were centered from their means in order to alleviate multicollinearity between the interaction term and its components, as recommended by Aiken and West (1991).

$$
\begin{aligned}
Ln(Propagation\_Index_i) = \beta_0 &+ \underbrace{\beta_1 \cdot Linear_i + \beta_2 \cdot Quadratic_i}_{Message\ Clarity} + \underbrace{\beta_3 \cdot URLs_i}_{Symbol} \\
&+ \underbrace{\beta_4 \cdot URLs_i \cdot Linear_i}_{Moderation} \\
&+ \underbrace{\beta_5 \cdot Tweet\_Length_i + \beta_6 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \le t \le \\ Date\ of\ posting\ tweet\ i}} Follower_{i,t}\right) + \beta_7 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \le t \le \\ Date\ of\ posting\ tweet\ i}} Followee_{i,t}\right)}_{Control} \\
&+ \underbrace{\beta_8 \cdot Ln\left(\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \le t \le \\ Date\ of\ posting\ tweet\ i}} Like_{i,t}\right) + \beta_9 \cdot Ln\left(\underbrace{\sum_{\substack{Join\ date\ of \\ the\ author\ of\ tweet\ i \\ \le t \le \\ Date\ of\ posting\ tweet\ i}} Tweet_{i,t}}_{Status}\right)}_{Control} \\
&+ \underbrace{\beta_{10} \cdot Mention\_YN_i + \beta_{11} \cdot First\_Retweet\_1m\_YN_i}_{Control} + \varepsilon_i
\end{aligned}
$$

[16] Figure 3.4 Statistical Expression

[35] Table 3.5 Correlation Matrix

Table 3.5.1. Correlation Matrix of the Research Model – Queensland

	1	2	3	4	5	6	7	8	9	10
PI	1									
Entropy (Linear)	0.170***	1								
URLs	-0.0087	-0.0582***	1							
First_Retweet_1m_YN	-0.189***	-0.0256***	-0.0452***	1						
Tweet_Length	-0.0223***	0.0401***	-0.422***	0.00176	1					
Ln(Followers)	-0.352***	-0.0799***	0.0830***	0.0880***	-0.0276***	1				
Ln(Followees)	-0.0768***	-0.0321***	-0.00425	0.0331***	0.0236***	0.586***	1			
Ln(Likes)	-0.0295***	-0.0216***	-0.00726	0.0229***	-0.00207	0.246***	0.316***	1		
Ln(Status)	-0.150***	-0.0544***	0.00691	0.0627***	-9.9E-05	0.672***	0.506***	0.443***	1	
Mention_YN	0.0878***	-0.0878***	0.0608***	-0.0348***	0.205***	0.00128	0.114***	0.0997***	0.106***	1

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 3.5.2. Correlation Matrix of the Research Model – Colorado

	1	2	3	4	5	6	7	8	9	10
PI	1									
Entropy (Linear)	0.169***	1								
URLs	0.0211***	-0.0474***	1							
First_Retweet_1m_YN	-0.229***	-0.0368***	-0.0647***	1						
Tweet_Length	-0.0724***	-0.0111	-0.404***	0.0327***	1					
Ln(Followers)	-0.399***	-0.102***	0.110***	0.0985***	0.00594	1				
Ln(Followees)	-0.109***	-0.0503***	0.0586***	0.0414***	-0.0105	0.595***	1			
Ln(Likes)	0.00211	-0.0112	-0.0215***	0.0139*	-0.00168	0.192***	0.396***	1		
Ln(Status)	-0.144***	-0.0705***	0.0792***	0.0540***	-0.0101	0.701***	0.631***	0.466***	1	
Mention_YN	0.0470***	-0.0631***	0.0232***	-0.0179**	0.233***	0.0731***	0.114***	0.103***	0.126***	1

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

3.5 Results

Table 3.6 shows the results of the robust regression analysis of the dependent variable, the propagation index, on the explanatory variables. For the 2011 Queensland and the 2013 Colorado flood incidents, the empirical models accounted for 20.9% ($F_{11, 24881}$=398.90, p<0.000) and 26.5% ($F_{11, 29208}$=780.93, p<0.000) of the variance respectively. Before evaluating the hypotheses, we examined the control variables. As a whole, seven control variables were statistically significant in predicting the PI (Queensland – $F_{7, 24881}$=524.82, p<0.000; Colorado – $F_{7, 29208}$=1081.61, p<0.000). It is noteworthy to point out a few interesting findings. During the Queensland and Colorado floods, tweets that were retweeted within one minute (*First_Retweet_1m_YN*) had a 10% and a 12.2% lower PI, respectively, than those that were retweeted after the first minute, while holding all other variables in the model at their means. That is, whether a tweet is retweeted within the first one minute reliably explains the extent to which how fast and wide the tweet is retweeted right after its posting. Partialling out the effects of the other variables in the model, we found empirical evidence for the negative relationship of the length of tweets (*Tweet_Length*) with the PI in both Twitter datasets (Queensland – $F_{1, 24881}$=102.38, p<0.000; Colorado – $F_{1, 29208}$=161.12, p<0.000), indicating that as the length of tweets increased, the PI significantly decreased. We interpret this result that as tweets deliver the more, directly interpretable information (except URLs), the faster and wider dissemination they could achieve due to the fact that readily digestible information reduces additional milling. Above all, followers decreased the PI significantly. That is, 1% increase in the number of followers on average decreased the PI of 25.0% in the Queensland floods and 28.33% in the Colorado ones (Queensland – $F_{1, 24881}$=2009.48, p<0.000; Colorado – $F_{1, 29208}$=4383.54, p<0.000). Practically, this result confirms that the size of one's followers greatly enables his or her tweets to be swiftly and broadly disseminated through the twitterverse.

Conversely, the size of followees had the opposite effect (Queensland – $F_{1,\ 24881}$=299.41, p<0.000; Colorado – $F_{1,\ 29208}$=239.26, p<0.000).

[36] Table 3.6 Results of the Logistic Regression		
Ln(Propagation Index) (DV) / Variables (IVs)	2011 Queensland	2013 Colorado
	Coefficient (Robust Error)	Coefficient (Robust Error)
Content Features		
- Message Clarity	$F_{2,\ 24881}$=448.26, p<0.000	$F_{2,\ 29208}$=393.66, p<0.000
$Linear_i$	0.594*** (0.0211)	0.675*** (0.0346)
$Quadratic_i$	-0.454*** (0.0764)	-0.413*** (0.0973)
- Symbol	$F_{1,\ 24881}$=1.48, p=0.2234	$F_{1,\ 29208}$=3.03, p=0.0819
$URLs_i$	-0.0141 (0.0116)	0.0211 (0.0121)
Clarity x Symbol		
- Tweet	$F_{1,\ 24881}$=6.53, p=0.0106	$F_{1,\ 29208}$=30.70, p<0.000
$URLs_i$ x $Linear_i$	-0.0823* (0.0322)	-0.205*** (0.0370)
Control Variables	$F_{7,\ 24881}$=524.82, p<0.000	$F_{7,\ 29208}$=1081.61, p<0.000
First_Retweet_1m_YN$_i$	-0.200*** (0.00835)	-0.260*** (0.00790)
Tweet_Length$_i$	-0.00291*** (0.000288)	-0.00364*** (0.000287)
Ln(Followers$_{i,t}$)	-0.287*** (0.00641)	-0.333*** (0.00503)
Ln(Followees$_{i,t}$)	0.105*** (0.00607)	0.0908*** (0.00587)
Ln(Likes$_{i,t}$)	-0.00610 (0.00358)	-0.0193*** (0.00286)
Ln(Status$_{i,t}$)	0.0675*** (0.00515)	0.135*** (0.00481)
Mention_YN$_i$	0.0906*** (0.00669)	0.0903*** (0.00645)
Constant	5.444*** (0.0118)	5.346*** (0.0128)
Model Summary		
R-squared	0.209***	0.265***
Adjusted R-squared	0.209***	0.264***

F Statistics	$F_{11,\ 24881}$=398.90	$F_{11,\ 29208}$=780.93
n	24893	29220

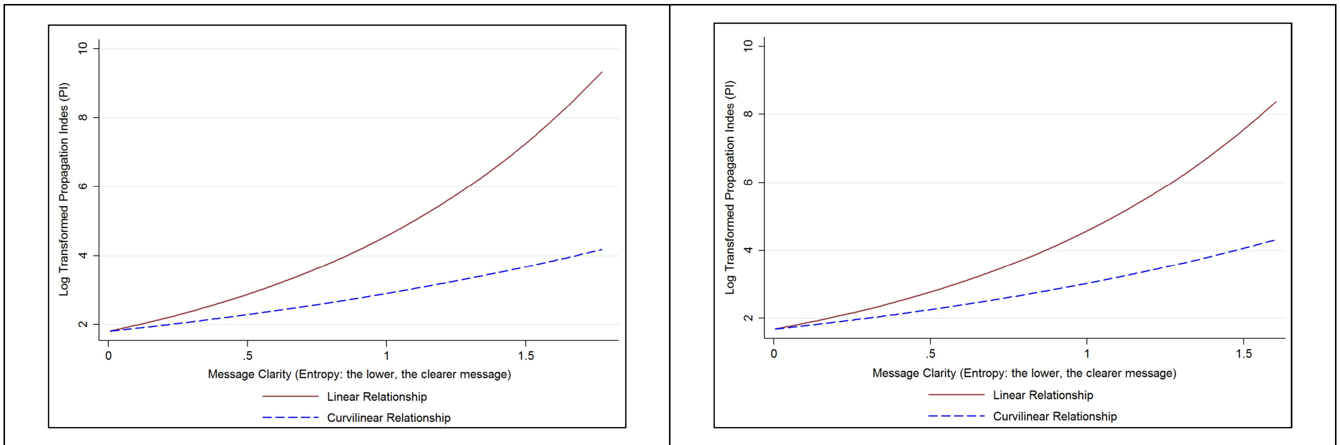
[1] All predictors are mean centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

In Hypotheses 1a and 1b, we speculated that as message clarity decreases, the initial propagation of tweets (H1a) would be hindered, and that relationship might weaken as message clarity further decreases (H1b). From Table 3.6, empirical evidence was found that the hypothesized relationships were to be statistically significant (Queensland – $F_{2,\ 24881}$=448.26, p<0.000; Colorado – $F_{2,\ 29208}$=393.66, p<0.000), after accounting for the effects of the other variables in the model. That is, as message clarity lowered, the propagation index (PI) increased (Queensland – $F_{1,\ 24881}$=789.98, p<0.000; Colorado – $F_{1,\ 29208}$=381.54, p<0.000). As expected, the strength of message clarity's effect weakened (Queensland – $F_{1,\ 24881}$=35.24, p<0.000; Colorado – $F_{1,\ 29208}$=18.03, p<0.000) while message clarity decreased even more. That is, as a tweet loses its message clarity, its initial propagation is impeded; however, the degree of such an impediment weakens. Therefore, both hypotheses are supported. This relationship is graphed in Figure 3.5.

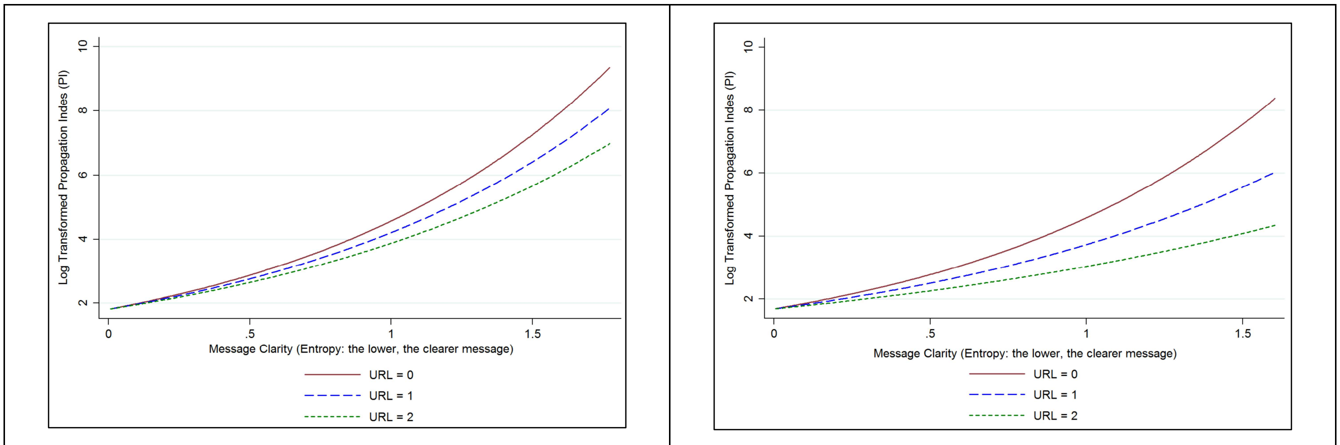
2011 Queensland	2013 Colorado



[17] Figure 3.5 Nonlinear Relationship between Message Clarity and Propagation Index (PI)

We could not find statistical evidence for Hypothesis 2, which states the negative effect of Twitter URLs on the initial propagation of tweets. Thus, this hypothesis is rejected. Even though there was no evidence for the main effect of Twitter URLs, the significant moderating effect of Twitter URLs on the relationship between message clarity and the PI was identified (Queensland – $F_{1,\ 24881}$=6.53, p=0.0106; Colorado – $F_{1,\ 29208}$=30.70, p<0.000). For the 2011 Queensland floods, one Twitter URL decreased 7.9% of the PI on average, when the relationship between message clarity and the PI was linear. Namely, while the initial propagation of a tweet is being hampering due to its lack of message clarity, an additional embedded URL in that tweet contributes to its propagation, which is improved by 7.9% on average with all other variables held constant at their means (see Figure 3.6). Likewise, we found a similar, moderating effect of Twitter URLs in the 2013 Colorado dataset in that its contribution to the initial propagation was 18.53%, which is much higher than that of 2011 Queensland dataset.

2011 Queensland	2013 Colorado



[18] Figure 3.6 Moderating Effect of Twitter URLs

3.6 Discussion

The study stems from the important, thought provoking two cases showed that neither average retweet time or retweet frequency provided satisfactory explanation of the initial tweet propagation, which is an important characteristic for disaster communication. From this perspective, we believed that a measure capable of differentiating the tweets illustrated in the two cases should be of importance, especially for disaster situations where rapid and wide dissemination of timely information is considered critical for safety (Li and Rao 2010, p. 4; Wilensky 2014, p. 705). Therefore, the primary purpose of this study was to explore a new measure that comprises both propagation *speed* and *scale* of tweets in great detail. Based upon the proposed measure and the elaborated research model, the study answered to the research questions we posed: how do we measure the initial propagation of tweets? how does message clarity affects the initial propagation of tweets? And finally, how do Twitter URLs influence the relationship between message clarity and the initial propagation of tweets? Regarding the first question, the empirical results shown in Figure 3.1 provided the high relevancy of the measure – the propagation index (PI) – with average retweet time and total retweet frequency simultaneously. Surprisingly, 93.3%

and 91.72% of the total variance of the logged average retweet time were predicted by the log-transformed PI, for the Queensland and Colorado incidents respectively. At the same time, the logged PI explained 33.1% of the variance of the retweet counts for the Queensland tweets and 32.4% for the Colorado ones. However, the log transformed average retweet time explained much smaller portions of the variance, which were 15.9% for the Queensland case and 15.2% for the Colorado one (see Figure 3.2). Above all, Figure 3.1 provides compelling evidence of the measure's capability in discerning different initial propagation patterns of tweets. The second question was designed to investigate the PI in more detail. As we expected, decrease in the message clarity of tweets slowed down the initial propagation of tweets, which is consistently shown in both datasets. This result provided a reasonable explication for the PI, because once realizing some degrees of message unclarity in received tweets, the recipients will seek more information to make sense out of the tweets, rather than just retweeting them. Based upon the research questions 1 and 2, the last research question added plausible evidence to our assumed mechanism of information with retweeting. That is, even though the main effect of Twitter URLs was not significant on predicting the dependent variable, the effect turned significant depending upon the levels of tweets' message clarity. In other words, additional information (i.e., Twitter URLs) helped twitterers cope with tweets revealing information dearth. Furthermore, as we defined message clarity in terms of the degree of information sufficiency (or insufficiency), this significant moderating effect empirically supports that among others information sufficiency in a tweet is one factor that causes its retweet speed and size. In that regard, it should be necessary to investigate why the effect of Twitter URLs was not significant on tweets' initial propagation. We considered Twitter URLs as a double-edged sword in the sense that while rich information linked by such URLs can provide supplemental information to tweets that are inherently associated with a certain degree of information dearth, it also requires twitterers spend extra time to process additional information. Therefore, the effect of Twitter URLs might be cancelled out when they were factored into the propagation index, which is designed to simultaneously measure tweets' retweet speed and scale. We performed post-hoc analysis to check the above possible explanations about Twitter URLs. As shown in Table 3.7, Twitter URLs increased the logged average retweet time. Although the effect was only significant in the 2013 Colorado floods, both directions were positive (Queensland – Coefficient=0.00723, $F_{1,\ 24881}$=1.81, p=0.1785; Colorado – Coefficient=0.048, $F_{1,\ 29208}$=76.28, p<0.000). This result partially supported our explanation about the processing aspect of Twitter URLs. In addition, the aspect of rich information was confirmed by the results of Table 3.8. It turned out that Twitter URLs increased the retweet frequency (Queensland – Coefficient=0.0506, Wald-Chi^2(1)=3.93, p=0.0474; Colorado – Coefficient=0.0883, Wald-Chi^2(1)=16.19, p=0.0001), and their effect became stronger as message clarity decreased (Queensland – Coefficient=0.166, Wald-Chi^2(1)=5.38, p=0.0204; Colorado – Coefficient=0.259, Wald-Chi^2(1)=16.65, p<0.0000). The results confirmed that during disasters information value positively contributes to retweeting.

[37] Table 3.7 Statistical Results of the Linear Regression

Ln(Avg. Retweet Time) (DV) / **Variables (IVs)**	**2011 Queensland**	**2013 Colorado**
	Coefficient (Robust Error)	Coefficient (Robust Error)
Content Features		
- Message Clarity	$F_{2,\ 24881}$=381.46, p<0.000	$F_{2,\ 29208}$=322.83, p<0.000
$Linear_i$	0.252*** (0.00972)	0.273*** (0.0159)
$Quadratic_i$	-0.193*** (0.0358)	-0.149** (0.0454)
- Symbol	$F_{1,\ 24881}$=1.81, p=0.1785	$F_{1,\ 29208}$=76.28, p<0.0000
$URLs_i$	0.00723 (0.00537)	0.0480*** (0.00550)

Clarity x Symbol		
- Tweet	$F_{1,\ 24881}$=7.87, p=0.0050	$F_{1,\ 29208}$=46.60, p<0.000
$URLs_i$ x $Linear_i$	-0.0419** (0.0149)	-0.116*** (0.0169)
Control Variables	$F_{7,\ 24881}$=563.06, p<0.000	$F_{7,\ 29208}$=982.86, p<0.000
$First\_Retweet\_1m\_YN_i$	-0.0998*** (0.00370)	-0.126*** (0.00352)
$Tweet\_Length_i$	-0.00106*** (0.000131)	-0.00141*** (0.000130)
$Ln(Followers_{i,t})$	-0.121*** (0.00274)	-0.128*** (0.00215)
$Ln(Followees_{i,t})$	0.0460*** (0.00273)	0.0327*** (0.00255)
$Ln(Likes_{i,t})$	-0.00103 (0.00159)	-0.00371** (0.00128)
$Ln(Status_{i,t})$	0.0204*** (0.00225)	0.0425*** (0.00215)
$Mention\_YN_i$	0.0425*** (0.00302)	0.0397*** (0.00291)
Constant	6.391*** (0.00544)	6.343*** (0.00583)
Model Summary		
R-squared	0.203***	0.236***
Adjusted R-squared	0.203***	0.235***
F Statistics	$F_{11,\ 24881}$=415.55	$F_{11,\ 29208}$=706.03
n	24893	29220

[1] All predictors are mean centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

[38] Table 3.8 Statistical Results of the Negative Binomial Regression		
Retweet Frequency (DV) / **Variables (IVs)**	**2011 Queensland**	**2013 Colorado**
	Coefficient (Robust Error)	Coefficient (Robust Error)
Content Features		
- Message Clarity	Wald $Chi^2(2)$=331.82, p<0.000	Wald $Chi^2(2)$=364.14, p<0.000
$Linear_i$	-1.023***	-0.861***

	(0.0593)	(0.0561)
$Quadratic_i$	0.863*** (0.125)	0.585*** (0.142)
- Symbol	Wald $Chi^2(1)=3.93$, p=0.0474	Wald $Chi^2(1)=16.19$, p=0.0001
$URLs_i$	0.0506* (0.0255)	0.0883*** (0.0220)
Clarity x Symbol		
- Tweet	Wald $Chi^2(1)=5.38$, p=0.0204	Wald $Chi^2(1)=16.65$, p<0.000
$URLs_i$ x $Linear_i$	0.199* (0.0859)	0.259*** (0.0635)
Control Variables	Wald $Chi^2(7)=1028.39$, p<0.000	Wald $Chi^2(7)=2592.10$, p<0.000
$First\_Retweet\_1m\_YN_i$	0.166*** (0.0223)	0.172*** (0.0126)
$Tweet\_Length_i$	0.00580*** (0.000959)	0.00418*** (0.000555)
$Ln(Followers_{i,t})$	0.302*** (0.0133)	0.354*** (0.00826)
$Ln(Followees_{i,t})$	-0.107*** (0.0125)	-0.0865*** (0.00875)
$Ln(Likes_{i,t})$	0.0460 (0.0258)	0.0346*** (0.00536)
$Ln(Status_{i,t})$	-0.0858*** (0.0208)	-0.171*** (0.00897)
$Mention\_YN_i$	-0.152*** (0.0227)	-0.0998*** (0.0127)
Constant	0.625*** (0.0230)	0.724*** (0.0216)
Model Summary		
Pseudo R-squared	0.076***	0.087***
Log Likelihood	8834.351	12729.737
n	24893	29220

[1] All predictors are mean centered in the regression.

[2] Results are estimated using robust regression with Huber-White sandwich estimators. Robust standard errors are in parentheses.

[3] Unstandardized regression coefficients are shown (* p<0.05, ** p<0.01, *** p<0.001).

As the summary of the study, Table 3.9 showed our research hypotheses and statistical results. Interestingly, the hypotheses were consistently supported or unsupported by both flood incidents.

[39] Table 3.9 Results of Hypothesis Testing

Hypothesis \ Dataset		Hypothesized Content	2011 Queensland	2013 Colorado	Consistency
H1	**H1a**	Message Clarity - Linear	Supported	Supported	Yes
	H1b	Message Clarity - Quadratic	Supported	Supported	Yes
H2		Twitter URLs	Not Supported	Not Supported	Yes
H3		Message Clarity x Twitter URLs	Supported	Supported	Yes

3.7 Conclusions

As a function of information sufficiency (or insufficiency), this study sheds light on information propagation in disasters. We viewed disaster situations where information is deficient, and thus considered the degree of a message's information sufficiency as one of the factors that influences the initial propagation of the message. Furthermore, we introduced the measure of PI for estimating initial propagation, which is a principal characteristic of emergency information during times of disaster. Based upon distinctive characteristics of Twitter such as a 140-character limit and the ability to link to external resources via URLs, we quantified each tweet's message clarity as a proxy for degrees of information sufficiency and considered Twitter URLs a supplementary means to deliver additional information. A series of empirical examinations verified the following: first, the proposed measure better predicts a tweet's initial propagation than its average retweet time or retweet frequency; second, as information sufficiency in a tweet decreases, its initial propagation steeply drops as well; last, when tweets suffer from information insufficiency, Twitter URLs reduce the negative effect of the problem of information dearth on the initial propagation of tweets.

In accordance with the implications of the empirical results, we make the following contributions. Disaster researchers, when examining factors that are supposed to influence disaster communication, should take into account degrees of information propagation immediately after information is generated. In addition, information sufficiency in a message affects its dissemination, and as such including variables reflecting the above notion in empirical models will enhance our understanding of how other message constituents promote or delay the sharing of a message in disasters. An acute understanding of message constituents, in terms of information value and leveraging them accordingly will help practitioners, such as emergency management officials or online journalists, reach a targeted audience via emergency messages in a timely manner. In particular, when using Twitter as a communication means in disasters, communication participants have to craft their tweets to express one clear topic as well as use Twitter URLs properly in order to supplement tweets with additional information.

As with any study, there are limitations. First, regarding the measure we proposed, conducting more research on it is a necessary step to accumulate empirical evidence. In addition, different types of disasters, including man-made devastating events, should improve the generalizability of the measure. In this sense, the two Twitter datasets alone would not be enough to establish the rigor of the measure. Second, to delve into the relationship between the PI (propagation index) and information sufficiency, other message constituents, except URLs, should be further examined in order to strengthen the argument regarding information sufficiency's role in association with emergency messages. Third, the generalizability of our empirical evidence is limited. Although we leveraged the two flood incidents to examine our study's hypothesis, other types of disasters such as earthquakes, wildfires, or man-made events can enhance the generalizability of our findings. Fourth, the unit of analysis for the study is an individual tweet. However, twitterers may post a series of tweets to deliver information about just one topic. Therefore, to minimize a limitation of the current study, future research can take a set of tweets by the same twitterer into account.

Appendix 3.A

[40] Table 3.A.1 57 Topics and Keywords of the 2013 Colorado floods	
Topic #	**Keywords**
1	relief levels flood_levels give pic impression friends news add #twibbon create federer tennis victims online flood_relief flood_victims abc abc_news
2	need volunteer register volunteers clean cleanup #bnefloods volunteering brisbane food emergency accommodation #bakedrelief #bnecleanup needs needed
3	centre evacuation evacuation_centre showgrounds pets ipswich spread word ipswich_showgrounds rna evac rna_showgrounds centres lost found hills
4	change cross red climate red_cross #vicfloods climate_change clean rain australian towns weather relief services affected brace information brisbane
5	fill sandbags need free brisbane form council affected services nature disaster offer businesses local train_services stop contact mother_nature city
6	support map comparison map_comparison relief post affected blog rough event #vicfloods fundraiser peeps blog_post benefit devastation happening fundrasing
7	victims flood_victims stay released place ravaged advice friends airport police donation legal information free #vicfloods affected hotline recovery
8	volunteers helping proud disaster clean spirit hand #vicfloods army efforts relief together rescue australian amazing community #bnefloods workers
9	bligh anna_bligh anna premier brisbane queensland_premier low residents evacuate lying higher water ground inquiry ipswich urged #brisbane starting
10	crisis news flood_crisis bligh toll premier missing death latest anna_bligh anna dead disaster live death_toll online confirmed ahead buying
11	victims donate donating appeal remember sitting flood_victims donation link vic nsw left donations harvey amazing coast #auction vintage total
12	ipswich mayor looting ipswich_mayor paul pisasale city markers paul_pisasale find flood_markers brisbane higher mythbuster flood_mythbuster facing pi
13	water power brisbane residents safe supply ipswich #bnefloods shopping boil water_supply centre food victims drink advised cut flood_victims need
14	spirit aussie aussie_spirit amazing victims flood_victims home donate working flooded return family find cleaning heaps thanks_heaps strangers aussie
15	creek cars footage flash toowoomba washed video lockyer flash_flood lockyer_creek mil show evacuate gave film water mate oprah gympie god rises higher

16	victims flood_victims donate support donating every affected money #prayforaustralia raise need hope handset donations retweet visit generously coffee
17	#qld affected judgment judgment_day update insurance #bnefloods brisvenice brisvegas flood_update longer brisbane info hotline tourism #vicfloods bus
18	disaster size area zone declared texas disaster_zone times united flood_disaster france kingdom united_kingdom germany united_kingdoms kingdoms
19	power cut energex brisbane ipswich free affected homes image charge restore phones facing inundation families businesses mythbuster flood_myth
20	brisbane storage photos images free brisbane_floods live free_storage #bnefloods storage_king offering trucks #brisbane aerial affected pics amazing
21	cross red safe brisbane national registration system free cow roof #bnefloods clean water map inquiry place photos #brislantis damaged cross_national
22	high zoo swim crocs australia_zoo high_enough tying brisbane weather god biggest arrive bureau biggest_flood weather_bureau companies insurance_compa
23	media social social_media twitter #vicfloods health helping aid police hope australia_day need doctors join email stars disaster dept needed sunrise
24	brisbane river brisbane_river #bnefloods floating cbd farm drive streets list expected restaurant free park city affected coronation coronation_drive
25	man volunteers photo boatload kangaroos needed rescued #bnecleanup mayor kangaroo more_volunteers pic brilliant registration centres
26	volunteers auctions need awesome cahill qld_floods tim_cahill tim awesome_auctions cold beers ground high cbd mobile cold_beers handing high_ground
27	crisis flood_crisis list real media citizen reports citizen_reports died twitter related stories line info outlets lifeline twitter_list media_outlet
28	river brisbane broken brisbane_river banks end west library west_end wet sunny dry sunny_day wrap freezer gladwrap wet_photosbooks photosbooks
29	evacuation info centres financial brisbane app hit pledges evacuation_centres financial_help dogs cats owners recovery free staff information links b
30	river brisbane peak brisbane_river expected levels metres conference media ipswich #bnefloods media_conference flood_peak live level livestream tab
31	#bnefloods brisbane closed street bank ipswich bridge water pier eagle cbd shit open #brisbane south_bank river motorway holy holy_shit crap #fb road
32	stadium suncorp_stadium suncorp brisbane pool swimming picture field footy_field #bnefloods water fire bridge transformer emergency services silence

33	waters flood_waters children disaster helping barrier reef #auspol barrier_reef support office play water damage equipment replace stop homes pay
34	warning severe rain thunderstorm weather brisbane thunderstorm_warning flash coast hit #qld bay bom #tcanthony heavy river cyclone moreton
35	donate every appeal flood_appeal tweet cents aussie aussie_queensland #prayforaustralia retweet message #staystrong received qld_floods everyone
36	#vicfloods #nswfloods map need information road closures info flood_information road_closures contact crisis #tasfloods urgent list live flood_map
37	donate need queenslanders desperately police facebook updates page twitter phone flight qld_floods qld_police date change affected booking service
38	victims australian fundraiser items fan international win fan_fundraiser autographed auction autographed_items bed offer recent house affected spare
39	victims cahill auction experience flood_victims tim raise bid money tim_cahill #socceroos match ebay charity everton aid signed shirt cricket relief
40	affected survival animals offer email housing foster assistance email_floods foster_caretemporary caretemporary bill unnecessary lewis survival_value
41	abbott deep tony_abbott tony water #auspol dig flood_water donations bin wheelie indication wheelie_bin #nbn good_indication dollar political need
42	toll death death_toll valley lockyer found lockyer_valley missing rises bodies grantham police flood_death_toll dead flash news body man risen search
43	bligh anna_bligh anna premier conference gillard press julia crisis media julia_gillard press_conference leadership live pressure #abcnews leader qld
44	snake frog ride photo community hitches incredible escapes frog_escapes_flood incredible_photo looting bligh escape created riding anna red australia
45	appeal relief flood_relief_appeal aussies donate everyone thinking needs premier #aussies flood_appeal disaster donating relief_appeal donated
46	missing dead rice jordan jordan_rice confirmed #prayforaustralia hero brother died sad save queensland_floods lost boy saving homes rip god queensland
47	relief appeal flood_relief fund auction money raise proceeds donate donated flood_appeal raised signed bid relief_fund song funds raising sales donation
48	recovery tsunami inland biblical flood_recovery impact inland_tsunami crisis faces facing economic news hell official support force warns economic_impact
49	shark ipswich bull street goodna flooded brisbane bull_shark spotted flooded_street sharks affected update streets swimming bull_sharks main main_street

50	brisbane city council city_council latest live game news alert updates services support online #bnefloods notice media info collection drinking
51	rspca fairfield animals fairfield_rspca water qld_floods repost foster animal retweet shelter register origin raise money jerseys origin_jerseys
52	towns affected brisbane crisis news medical coal free water flood_crisis relief clean volunteer offering home car cities inundated recovery reds
53	safe affected everyone thoughts brisbane hope stay news #prayforaustralia friends prayers family home heart lost sad devastating hear watching rain
54	waters flood_waters city australian rockhampton braces brisbane fundraiser rise peak queensland_braces coastal rising river satellite bridge fundraising
55	end brisbane water west house home clean need #bnefloods mud west_end helping flooded hand hard river #vicfloods cleaning city power volunteers girl
56	flooded homes brisbane affected businesses need power stallion suburbs supply bay needed inundated water #bnefloods ipswich spare energex deception
57	jordan rice jordan_rice save swept younger rescuers brother life younger_brother blake own_life losing stop hero toowoomba aged waters #prayforaustralia
58	relief #vicfloods view volunteers cross hills needs bowen support concert red_cross neighbours bowen_hills service clean crisis brisbane continues
59	relief donate flood_relief donations needs appeal word spread information flood_relief_appeal everyone need #prayforaustralia qld_australia needed
60	heart health aussies praying safety prayers hearts breaks picture markets #bnefloods rocklea rocklea_markets brisbane disaster system team chopper fr
61	points velocity velocity_points brisbane closed donation allowing convert #bnefloods donate recovery donating donations road awesome page milton
62	appeal flood_appeal donate rspca animals give donations money qld_rspca raise need donated generously #vicfloods sales #prayforaustralia plead donati
63	damage insurance flood_damage need business food brisbane storm claims small milk pay water #bnecleanup supplies levy clean hit bread guide office
64	brisbane transport cross public red_cross centre public_transport volunteers red needs affected melbourne seekers asylum_seekers north needed based
65	brisbane cbd brisbane_cbd power closed evacuated transport coast myth public buster flood_myth_buster public_transport highway #bnefloods closing
66	found dogs dog disaster goodna island need floating toilet lost fraser block flood_disaster toilet_block fraser_island sharon pray god caltex sleep k

67	donate appeal fireworks day_fireworks cancel recovery donated money relief flood_appeal donation million free ride fund raising awareness twitter
68	levy video flood_levy tax #vicfloods friend gillard youtube relief #auspol pay victims images toowoomba view youtube_video nasa queensland_floods
69	water services fire lost normal goods kid home talent normal_kid stefanovic karl treatment affected karl_stefanovic plants supply summary room need s
70	coverage news abc maps radio brisbane live info local council information flood_maps online #abcnews site twitter updates channel #bnefloods city dig
71	town update residents dalby link area pool audio_link alert recovery pool_area emerald audio rockhampton road hit #police swimming power cut southern
72	dam wivenhoe brisbane water #bnefloods cbd lucia capacity view river street st_lucia brisbane_cbd albert farm full george new_farm southbank flooded

[41] Table 3.A.2 72 Topics and Keywords of the 2011 Queensland floods

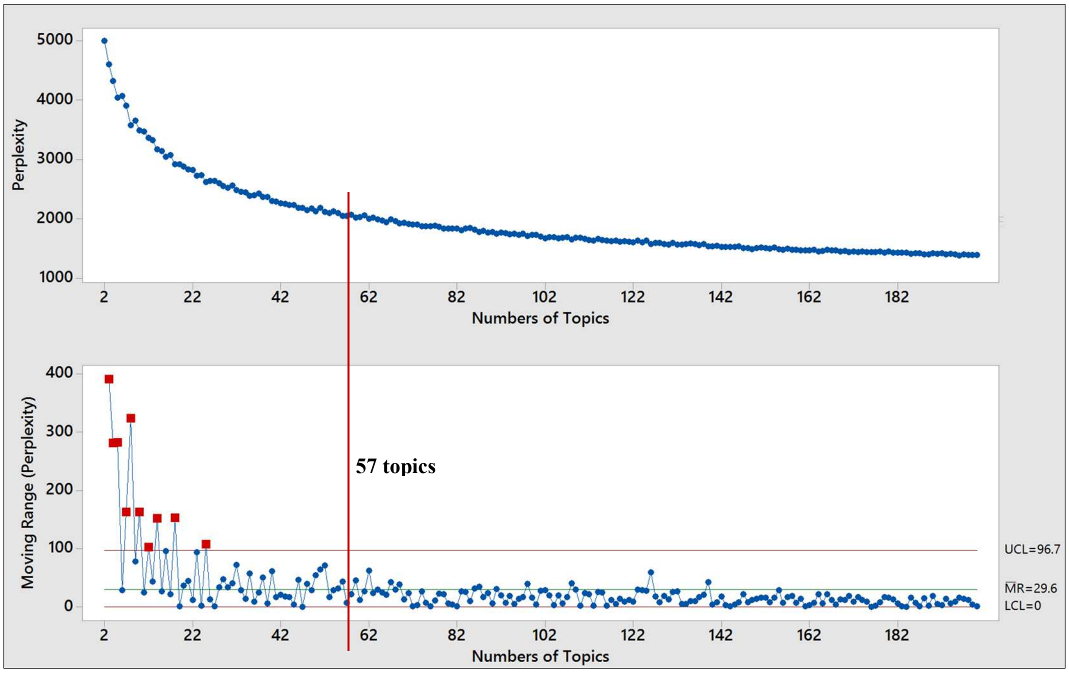
Topic #	Keywords
1	toll, death, dead, rises, person, death_toll, flood_death_toll, evacuations, confirmed, people, deadly, presumed, woman, ordered, found, flood_toll, waters, missing
2	evacuation, center, head, jamestown, residents, notice, springs, eldorado, evac, creek, eldorado_springs, cty, evacuation_center, evacuation_notice, people, barn, ordered
3	towns, rescue, rain, rains, warnings, flood_warnings, diverse, closed, forecast, cats, flood_towns, colorado_towns, break, flood_rescue, stranded, brief_break, hamper, waters
4	schools, aurora, closed, creek, aurora_pd, creek_schools, aurora_schools, request, canyon, water, cherry, debris, valley, surge, foot, cars, other_debris, carrying, boulder
5	rescue, boulder, operation, water, flood_rescue_operation, area, report, continues, home, weather, leave, spill, chemical, historic, drive, fracking, rain, chemical_spill
6	record, breaking, guard, coast, led, worse, denver, concert, coast_guard, helicopters, relief, survivors, defense, coast_guard_helicopters, victims, benefit, state, coming
7	people, county, unaccounted, boulder, rescued, rescue, crews, sheriff, larimer, man, helicopters, save, officials, larimer_county, pets, boulder_county, racing, news, air
8	mountain, city, rocky, national, dam, commerce, arsenal, rocky_mountain_arsenal, evacuations, wildlife, refuge, failed, wildlife_refuge, impassable, roads, streets, east, dams
9	guard, national_guard, national, town, lyons, residents, jamestown, evacuations, moves, continue, boulder, evacuate, news, crest, downstream, colorado_town, students

10	creek, boulder, boulder_creek, move, broadway, sirens, sounding, #cuboulder, higher, canyon, cfs, east, ground, higher_ground, mesa, place, rising, table, flood_sirens, shelter
11	canyon, boulder, water, ground, higher, higher_ground, wall, coming, boulder_canyon, creek, immediately, move, boulder_creek, gulch, emerson_gulch, emerson, seek, debris, pearl
12	boulder, rain, evacuate, flash, more_rain, continue, ordered, live, county, officials, rescues, expected, braces, lyons, flash_flood, colorado_braces, damage, town, downtown
13	warning, flash, flash_flood_warning, boulder, flash_flood, county, issued, flood_warning, counties, effect, watch, skies, rain, warnings, evacuees, denver, springs
14	creek, boulder_creek, boulder, water, flow, wall, usgs, official, denver, term, experts, tsunami, experts_term, readings, creek_flow_readings, sensor, fourmile, usgs_sensor
15	platte, river, oil, south, spills, south_platte_river, gallons, tank, swollen, platte_river, spill, south_platte, damaged, morgan, reported, waters, water, oil_spill, Greeley
16	oil, gas, spills, zones, #fracking, wells, tracking, flood_zones, waters, sites, fracking, flood_waters, post, flooded, chemicals, water, gas_wells, leaks, denver, denver_post
17	gallons, locations, road, drenched, crude, dumps, spill, oil_spill_dumps, closures, waters, road_closures, flooded, boulder, water, many_locations, loved, shelter, affected
18	disaster, flood_disaster, media, blackout, media_blackout, #fracking, fracking, spills, happening, photos, update, toxic, worse, confirmed, shocking_photos, underwater, zone
19	waters, water, flood_waters, piano, house, play, home, sewage, wrecked, boulder, decided, man, contaminated, avoid, plays, sweep, moments, bike, creek, colorado_home, stay, video
20	vrain, water, river, creek, bridge, evac, roads, lyons, place, street, boulder, vrain_river, longmont, home, loveland, dry, center, big, #longmont, stay, hygiene, news, left
21	thompson, big, river, thompson_river, feet, county, ravaged, woman, pound, fatality, canyon, fifth_fatality, thompson_canyon, stage, record, loveland, central, thompson_flood
22	photo, car, havana, viewer, lyons, viewer_photo, swim, road, air, hwy, town, boulder, damage, hwy, news, water, dillon, pic, collapse, assessment, rescue, road_collapse, inside
23	longmont, #longmontflood, victims, water, lyons, view, rescues, equine, dam, storm, helicopter, vehicles, register, volunteers, image, urgent_call, woman, soldier, blog
24	long, water, city, safe, boulder, photo, rain, washed, picture, commerce, denver, commerce_city, rescue, stay, roads, house, areas, problems, live, couple, photos, send, yards, mile
25	images, unbelievable, unbelievable_images, boulder, map, google, tremendous, began, crisis, area, travel, water, notice, earth, evacuation, severe, google_earth, flash
26	game, football, school, state, path, bike, bike_path, postponed, high, fresno, field, pic, park, aurora, high_school, utah, utah_park, baseball, baseball_field, overland

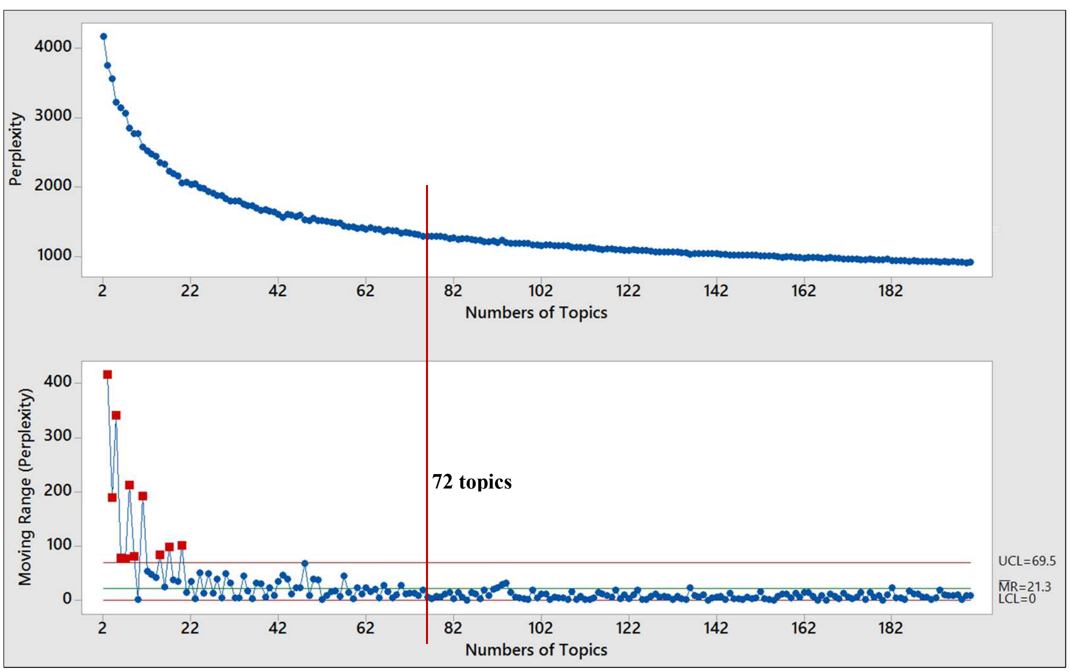
27	front, range, front_range, boulder, coverage, open, space, water, emergency, relief, trucks, workers, rescue, hard, downtown, disaster, county, working, longmont, effort, parks
28	damage, photos, aerial, images, flood_damage, video, biblical, climate, line, trends, boulder, climate_trends, views, biblical_flood, show, waters, aerial_views, lyons, shot
29	campus, evacuation, damage, homes, water, mobile, school, mobile_homes, creek, high, epic, buildings, boulder, photo, shows, water_damage, city, shelters, closed, high_school
30	big, thompson, canyon, thompson_canyon, road, hwy, hwy, thousand, boulder, flooded, water, science, thompson_canyon_entr, entr, baseline, damage, photographers, cut, deep
31	road, closures, road_closures, map, list, county, updates, boulder, closure, updated, found, #copets, center, shelters, latest, shelter, evacuation, road_closure_map, roads, dog
32	park, hwy, hwy, closed, estes, estes_park, #cotraf, open, road, roads, highway, photos, disaster, #estespark, directions, news, fun, reporter, app, denver, evergreen
33	water, boil, residents, high, drinking, lyons, safe, treatment, drink, advisory, hand, boulder, district, city, vehicles, wastewater, left, bottled, town, levels, contaminated
34	recovery, information, response, volunteer, relief, resources, updates, communities, efforts, live, emergency, cleanup, blog, affected, local, boulder, long, flood_recovery
35	disaster, assistance, fema, boulder, emergency, county, recovery, center, counties, federal, disaster_assistance, declaration, map, affected, evacuation, register
36	damage, losses, billion, flood_damage, property_losses, relief, repairs, shutdown, property, million, government, flood_relief, highways, left, street, bridges, estimated
37	aid, unanimously, republicans, relief, sandy, sandy_aid, colorado_republicans, opposed, support, flood_relief, voted, house, house_republicans, flood_relief_unanimously
38	biden, recovery, hickenlooper, devastation, flood_devastation, damage, view, president, fema, efforts, joe, gov, response, vice_president, team, vice, joe_biden, news, rescue
39	victims, relief, word, free, spread, #cofloodrelief, storage, free_storage, flood_victims, fund, flood_relief, giving, donating, donated, flood_relief_fund, marijuana
40	relief, victims, flood_victims, #cofloodrelief, donate, efforts, flood_relief, support, fundraiser, benefit, affected, donations, relief_efforts, effort, raised, helping
41	people, unaccounted, oem, areas, boulder, rain, more_rain, awaits, number, center, flood_areas, boulder_oem, remain, home, shelter, stop, area, volunteers, listed, report
42	homes, unaccounted, people, destroyed, damaged, dead, evacuated, missing, shelters, search, homes_damaged, update, loved, safe, register, presumed, homes_destroyed, numbers
43	family, impacted, pray, fire, guard, epic, reach, flush, truck, zone, members, stranded, driving, food, video, housing, flood_zone, fire_truck, order, guard_members, residents

44	cross, red, victims, flood_victims, red_cross, give, texting, climate, change, shelter, climate_change, affected, shelters, people, volunteers, american, #cofloodrelief, safe
45	collins, fort, fort_collins, relief, south, view, north, support, efforts, friends, #foco, based, resorts, vail_resorts, denver, co_support, closed, relief_efforts, pass, season
46	canyon, boulder, residents, people, shelters, left, stayed, hand, water, boulder_canyon, springs, evacuated, overnight, creek, road, expected, support, providing, #redcross
47	safe, needed, share, #copets, pets, food, victims, volunteers, lost, animals, home, hay, register, #cofloodrelief, pet, loved, victim, longmont, check, disaster, donations, sign
48	pets, rescued, people, visit, best_way, evacuated, helicopter, victims, katrina, survivors, number, historic, #nationalguard, historic_flood, #copets, greatest_number, town
49	boulder, longmont, springs, closed, humane, open, manitou, society, humane_society, page, ave, #waldoflood, center, shelter, west, front_page, manitou_springs, animals, #hmrd
50	safe, boulder, stay, rain, friends, prayers, thoughts, people, hope, affected, home, good, dry, family, love, raining, bad, crazy, victims,
51	schools, aurora, closed, creek, aurora pd, creek schools, aurora schools, request, canyon, water, cherry, debris, valley, surge
52	rain, inches, totals, wild, instagrams, wild flood, rainfall, snow, boulder, received, map, record, past, annual, feet, rain totals
53	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery
54	live, victims, coverage, flood_victims, rocks, force, task, red, task_force, state, rain, red_rocks, news, rescues, continue, debris, good, water, honor, team, oil, photo, tribune
55	disaster, boulder, waters, flood_waters, people, allowed, fracking, tubing, boulder_pd, reminds, flood_disaster, cited, floodwaters, fracking_disaster, missing, sky, clears
56	county, weld, boulder, denver, post, weld_county, residents, denver_post, water, closed, boulder_county, evacuations, pipeline, road, oil_pipeline, roads, oil, blvd, rain
57	rain, weather, snow, rescue, heat, efforts, fire, half, ass, blizzard, county, updates, people, blog, latest, await, recovery, more_rain, snarls, fundraiser, latest_updates, live

Appendix 3.B



< 2011 Queensland floods >



< 2013 Colorado floods >

[19] Figure 3.B The Optimal Topic Number by Perplexity

Appendix 3.C

[42] Table 3.C.1 Test of Multicollinearity – 2011 Queensland floods			
Entropy (Linear)	1.02	Ln(Followees)	1.65
URLs	1.28	Ln(Status)	2.20
First_Retweet_1m_YN	1.01	Ln(Likes)	1.29
Tweet_Length	1.31	Mention_YN	1.13
Ln(Followers)	2.28	**Mean VIF**	1.46

[43] Table 3.C.2 Test of Multicollinearity – 2011 Colorado floods			
Entropy (Linear)	1.02	Ln(Followees)	1.90
URLs	1.24	Ln(Status)	2.71
First_Retweet_1m_YN	1.02	Ln(Likes)	1.41
Tweet_Length	1.29	Mention_YN	1.10
Ln(Followers)	2.34	**Mean VIF**	1.56

References

Abdullah, N. A., Nishioka, D., Tanaka, Y., and Murayama, Y. 2014. "A Preliminary Study on User's Decision Making Towards Retweet Messages," *IFIP International Information Security Conference*: Springer, pp. 359-365.

Acar, A., and Muraki, Y. 2011. "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster," *International Journal of Web Based Communities* (7:3), pp. 392-402.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. 2011. "Sentiment Analysis of Twitter Data," *Proceedings of the workshop on languages in social media*: Association for Computational Linguistics, pp. 30-38.

Aiken, L. S., West, S. G., and Reno, R. R. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage.

Aladwani, A. M. 2015. "Facilitators, Characteristics, and Impacts of Twitter Use: Theoretical Analysis and Empirical Illustration," *International Journal of Information Management* (35:1), pp. 15-25.

Anderson, K. M., and Schram, A. 2011. "Design and Implementation of a Data Analytics Infrastructure in Support of Crisis Informatics Research (Nier Track)," *Proceedings of the 33rd International Conference on Software Engineering*, Honolulu, HI, pp. 844-847.

Bean, H., Liu, B. F., Madden, S., Sutton, J., Wood, M. M., and Mileti, D. S. 2016. "Disaster Warnings in Your Pocket: How Audiences Interpret Mobile Alerts for an Unfamiliar Hazard," *Journal of Contingencies and Crisis Management*).

Bean, H., Sutton, J., Liu, B. F., Madden, S., Wood, M. M., and Mileti, D. S. 2015. "The Study of Mobile Public Warning Messages: A Research Review and Agenda," *Review of Communication* (15:1), pp. 60-80.

Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), pp. 77-84.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3), pp. 993-1022.

Boyd, D., Golder, S., and Lotan, G. 2010. "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*: IEEE, pp. 1-10.

Boyle, M. P., Schmierbach, M., Armstrong, C. L., McLeod, D. M., Shah, D. V., and Pan, Z. 2004. "Information Seeking and Emotional Reactions to the September 11 Terrorist Attacks," *Journalism & Mass Communication Quarterly* (81:1), pp. 155-167.

Breusch, T. S., and Pagan, A. R. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica* (47:5), pp. 1287-1294.

Bruin, J. 2016. "Chapter 5: Additional Coding Systems for Categorical Variables in Regression Analysis " *Regression with SAS* Retrieved Semptember 7th, 2016, from http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter5/sasreg5.htm

Bruns, A. 2008. "3.1. The Active Audience: Transforming Journalism from Gatekeeping to Gatewatching."

Bruns, A., Burgess, J. E., Crawford, K., and Shaw, F. 2012. "# Qldfloods and@ Qpsmedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods,").

Bruns, A., and Stieglitz, S. 2012. "Quantitative Approaches to Comparing Communication Patterns on Twitter," *Journal of Technology in Human Services* (30:3-4), pp. 160-185.

Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., Knight, V., Procter, R., and Voss, A. 2014. "Tweeting the Terror: Modelling the Social Media Reaction to the Woolwich Terrorist Attack," *Social Network Analysis and Mining* (4:1), pp. 1-14.

Cameron, A. C., and Trivedi, P. K. 2013. *Regression Analysis of Count Data, 2nd Edition*. Cambridge University Press.

Castillo, C., Mendoza, M., and Poblete, B. 2011. "Information Credibility on Twitter," *Proceedings of the 20th international conference on World wide web*: ACM, pp. 675-684.

Cataldi, M., and Aufaure, M.-A. 2015. "The 10 Million Follower Fallacy: Audience Size Does Not Prove Domain-Influence on Twitter," *Knowledge and Information Systems* (44:3), pp. 559-580.

Chaiken, S. 1980. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of personality and social psychology* (39:5), p. 752.

Chaiken, S., and Eagly, A. H. 1989. "Heuristic and Systematic Information Processing within And," *Unintended thought* (212).

Chaiken, S., and Maheswaran, D. 1994. "Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment," *Journal of personality and social psychology* (66:3), p. 460.

Chatfield, A. T., and Brajawidagda, U. 2013. "Twitter Early Tsunami Warning System: A Case Study in Indonesia's Natural Disaster Management," *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pp. 2050-2060.

Cheong, F., and Cheong, C. 2011. "Social Media Data Mining: A Social Network Analysis of Tweets During the 2010-2011 Australian Floods," *Proceedings of the 15th Pacific Asia Conference on Information Systems*, Brisbane, Australia, pp. 1-16.

Christakou, E. R., and Klimis, G.-M. 2013. "Blogs and Social Media: The New Word of Mouth and Its Impact on the Reputation of Banks and on Their Profitability," in *Handbook of Social Media Management*. Springer, pp. 715-735.

Compston, S. 2014. "Identifying and Understanding Retweets." Retrieved August 24th, 2016, from http://support.gnip.com/articles/identifying-and-understanding-retweets.html

Connor, T., Alamaguer, M., and Silva, D. 2013. "'We Were Lucky to Get Out': Scores of People Unaccounted for in Colorado Flooding." Retrieved July 20, 2014, from http://www.nbcnews.com/news/other/we-were-lucky-get-out-scores-people-unaccounted-colorado-flooding-f8C11157645

Cotelo, J. M., Cruz, F. L., and Troyano, J. A. 2014. "Dynamic Topic□Related Tweet Retrieval," *Journal of the Association for Information Science and Technology* (65:3), pp. 513-523.

Covello, V., Becker, S., Palenchar, M., Renn, O., Sellke, P., Tzavellas, T., Morrell, P., Pfeifle, M., Tzavellas, A., and Bynum, R. 2010. "Effective Risk Communications for the Counter Improvised Explosive Devices Threat," *S4 Inc* (8).

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, S., and Anderson, S. 2014a. "Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods," *Proceedings of the 11th International ISCRAM Conference*, pp. 18-21.

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, S., and Anderson, S. 2014b. "Supporting Disaster Reconnaissance with Social Media Data: A Design-Oriented Case Study of the 2013 Colorado Floods," *Proceedings of the 11th International ISCRAM Conference-University Park.*

David A. Belsley, E. K., Roy E. Welsch. 2005. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

Davies, R. 2013. "Brisbane and Queensland Floods 2011." www. floodlist.com: FloodList - Reporting floods and flooding news.

Eastin, M. S. 2001. "Credibility Assessments of Online Health Information: The Effects of Source Expertise and Knowledge of Content," *Journal of Computer□Mediated Communication* (6:4), pp. 0-0.

Ellaway, P. 1978. "Cumulative Sum Technique and Its Application to the Analysis of Peristimulus Time Histograms," *Electroencephalography and clinical neurophysiology* (45:2), pp. 302-304.

FEMA. 2015. "Frequently Asked Questions: Wireless Emergency Alerts " Retrieved July 19th, 2016, from https://www.fema.gov/frequently-asked-questions-wireless-emergency-alerts

Flaherty, L. M., Pearce, K. J., and Rubin, R. B. 1998. "Internet and Face□to□Face Communication: Not Functional Alternatives," *Communication Quarterly* (46:3), pp. 250-268.

Fraustino, J. D., Liu, B., and Jin, Y. 2012a. *Social Media Use During Disasters: A Review of the Knowledge Base and Gaps*. National Consortium for the Study of Terrorism and Responses to Terrorism.

Fraustino, J. D., Liu, B., and Jin, Y. 2012b. "Social Media Use During Disasters: A Review of the Knowledge Base and Gaps,").

Gaver, W. W. 1991. "Technology Affordances," *Proceedings of the SIGCHI conference on Human factors in computing systems*: ACM, pp. 79-84.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. 2011. "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*: Association for Computational Linguistics, pp. 42-47.

Gochis, D., Schumacher, R., Friedrich, K., Doesken, N., Kelsch, M., Sun, J., Ikeda, K., Lindsey, D., Wood, A., Dolan, B., Matrosov, S., Newman, A., Mahoney, K., Rutledge, S., Johnson, R., Kucera, P., Kennedy, P., Sempere-Torres, D., Steiner, M., Roberts, R., Wilson, J., Yu, W., Chandrasekar, V., Rasmussen, R., Anderson, A., and Brown, B. 2014. "The Great Colorado Flood of September 2013," *Bulletin of the American Meteorological Society*), pp. 1-71.

Griffin, R. J., Neuwirth, K., Giese, J., and Dunwoody, S. 2002. "Linking the Heuristic-Systematic Model and Depth of Processing," *Communication Research* (29:6), pp. 705-732.

Hanson, D. C. J. March 08, 2012. "Floods: 10 of the Deadliest in Australian History," in: *Australian Geographic*.

Havard, T. M. 2001. "An Experimental Evaluation of the Effect of Data Presentation on Heuristic Bias in Commercial Valuation," *Journal of Property Research* (18:1), pp. 51-67.

Hermann, C. F. 1963. "Some Consequences of Crisis Which Limit the Viability of Organizations," *Administrative Science Quarterly* (8:1), pp. 61-82.

Heverin, T., and Zach, L. 2012. "Use of Microblogging for Collective Sense□Making During Violent Crises: A Study of Three Campus Shootings," *Journal of the American Society for Information Science and Technology* (63:1), pp. 34-47.

Hilbe, J. M. 2011. *Negative Binomial Regression*. Cambridge University Press.

Hollan, J., and Stornetta, S. 1992. "Beyond Being There," *Proceedings of the SIGCHI conference on Human factors in computing systems*: ACM, pp. 119-125.

Hsu, C.-L., and Liao, Y.-C. 2014. "Exploring the Linkages between Perceived Information Accessibility and Microblog Stickiness: The Moderating Role of a Sense of Community," *Information & Management* (51:7), pp. 833-844.

Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., and Ma, K.-L. 2012. "Breaking News on Twitter," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, pp. 2751-2754.

Hu, Y., and Sundar, S. S. 2009. "Effects of Online Health Sources on Credibility and Behavioral Intentions," *Communication Research*).

Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pp. 221-233.

Hughes, A. L., Denis, L. A. A. S., Palen, L., and Anderson, K. M. 2014. "Online Public Communications by Police & Fire Services During the 2012 Hurricane Sandy," in: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. Toronto, Ontario, Canada: ACM, pp. 1505-1514.

Hughes, A. L., and Palen, L. 2009. "Twitter Adoption and Use in Mass Convergence and Emergency Events," *International Journal of Emergency Management* (6:3-4), pp. 248-260.

Hutchby, I. 2001. "Technologies, Texts and Affordances," *Sociology* (35:2), pp. 441-456.

Hutto, C. J., Yardi, S., and Gilbert, E. 2013. "A Longitudinal Study of Follow Predictors on Twitter," *Proceedings of the sigchi conference on human factors in computing systems*: ACM, pp. 821-830.

Judd, C. M., Kenny, D. A., and McClelland, G. H. 2001. "Estimating and Testing Mediation and Moderation in within-Subject Designs," *Psychological methods* (6:2), p. 115.

Judd, C. M., McClelland, G. H., and Ryan, C. S. 2011. *Data Analysis: A Model Comparison Approach*. Routledge.

Kahlor, L., Dunwoody, S., Griffin, R. J., and Neuwirth, K. 2006. "Seeking and Processing Information About Impersonal Risk," *Science Communication* (28:2), pp. 163-194.

Keene, O. N. 1995. "The Log Transformation Is Special," *Statistics in medicine* (14:8), pp. 811-819.

Keller, C., Siegrist, M., and Gutscher, H. 2006. "The Role of the Affect and Availability Heuristics in Risk Communication," *Risk Analysis* (26:3), pp. 631-639.

Kenny, D. A. 2015. "Moderator Variables: Introduction." Retrieved November 29th, 2015

Klotz, C., Ross, A., Clark, E., and Martell, C. 2014. "Tweet!–and I Can Tell How Many Followers You Have," in *Recent Advances in Information and Communication Technology*. Springer, pp. 245-253.

Kongthon, A., Haruechaiyasak, C., Pailai, J., and Kongyoung, S. 2012. "The Role of Twitter During a Natural Disaster: Case Study of 2011 Thai Flood," *Proceedings of the 2012 Technology Management for Emerging Technologies*, Vancouver, British Columbia, Canada, pp. 2227-2232.

Kostkova, P., Szomszor, M., and St Louis, C. 2014. "# Swineflu: The Use of Twitter as an Early Warning and Risk Communication Tool in the 2009 Swine Flu Pandemic," *ACM Transactions on Management Information Systems (TMIS)* (5:2), p. 8.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., and Agras, W. S. 2002. "Mediators and Moderators of Treatment Effects in Randomized Clinical Trials," *Archives of general psychiatry* (59:10), pp. 877-883.

Kwak, H., Lee, C., Park, H., and Moon, S. 2010. "What Is Twitter, a Social Network or a News Media?," *Proceedings of the 19th international conference on World wide web*: ACM, pp. 591-600.

Lachlan, K. A., Spence, P. R., Lin, X., and Del Greco, M. 2014. "Screaming into the Wind: Examining the Volume and Content of Tweets Associated with Hurricane Sandy," *Communication Studies* (65:5), pp. 500-518.

Lane, D. M. 2016. "Online Statistics Education: An Interactive Multimedia Course of Study." *Transformations* Retrieved 18th September, 2016, from http://onlinestatbook.com/2/transformations/log.html

Lang, A. 2000. "The Limited Capacity Model of Mediated Message Processing," *Journal of communication* (50:1), pp. 46-70.

Laniado, D., and Mika, P. 2010. "Making Sense of Twitter," in *The Semantic Web – Iswc 2010,* P. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Pan, I. Horrocks and B. Glimm (eds.). Springer Berlin Heidelberg, pp. 470-485.

Latonero, M., and Shklovski, I. 2011a. "Emergency Management, Twitter, and Social Media Evangelism," *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* (3:4), pp. 1-16.

Latonero, M., and Shklovski, I. 2011b. "Emergency Management, Twitter, and Social Media Evangelism," *Latonero, M. & Shklovski, I.(2011). Emergency management, Twitter, & Social Media Evangelism. International Journal of Information Systems for Crisis Response and Management* (3:4), pp. 67-86.

Li, J., and Rao, H. R. 2010. "Twitter as a Rapid Response News Service: An Exploration in the Context of the 2008 China Earthquake," *The Electronic Journal of Information Systems in Developing Countries* (42).

Lindell, M. K., and Perry, R. W. 1987. "Warning Mechanisms in Emergency'response Systems '," *International Journal of Mass Emergencies and Disasters* (5:2), pp. 137-153.

Liu, Z., Liu, L., and Li, H. 2012. "Determinants of Information Retweeting in Microblogging," *Internet Research* (22:4), pp. 443-466.

Lo, S.-K. 2008. "The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication," *CyberPsychology & Behavior* (11:5), pp. 595-597.

Ma, Z., Sun, A., and Cong, G. 2013. "On Predicting the Popularity of Newly Emerging Hashtags in Twitter," *Journal of the American Society for Information Science and Technology* (64:7), pp. 1399-1410.

Majchrzak, A., Markus, M. L., and Wareham, J. 2016. "Designing for Digital Transformation: Lessons for Information Systems Research from the Study of Ict and Societal Challenges," *MIS Quarterly* (40:2), pp. 267-277.

Manjoo, F. 2010. "Online Giving, One Person at a Time." Retrieved 2016, 2016, from http://www.nytimes.com/2010/11/11/giving/11SOCIAL.html?_r=0

McCallum, A. K. 2002. "Mallet: A Machine Learning for Language Toolkit." Retrieved September 15, 2014, from http://mallet.cs.umass.edu

Meaney, P. M., Fang, Q., Rubæk, T., Demidenko, E., and Paulsen, K. D. 2007. "Log Transformation Benefits Parameter Estimation in Microwave Tomographic Imaging," *Medical physics* (34:6), pp. 2014-2023.
Mendoza, M., Poblete, B., and Castillo, C. 2010. "Twitter under Crisis: Can We Trust What We Rt?," *Proceedings of the first workshop on social media analytics*: ACM, pp. 71-79.
Mendoza, M., Poblete Labra, B., and Castillo Ocaranza, C. 2013. "Predicting Information Credibility in Time-Sensitive Social Media,").
Metzger, M. J. 2007. "Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research," *Journal of the American Society for Information Science and Technology* (58:13), pp. 2078-2091.
Metzger, M. J., and Flanagin, A. J. 2013. "Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics," *Journal of Pragmatics* (59), pp. 210-220.
Metzger, M. J., Flanagin, A. J., and Medders, R. B. 2010. "Social and Heuristic Approaches to Credibility Evaluation Online," *Journal of communication* (60:3), pp. 413-439.
Mileti, D. S., and Peek, L. 2000. "The Social Psychology of Public Response to Warnings of a Nuclear Power Plant Accident," *Journal of hazardous materials* (75:2), pp. 181-194.
Mileti, D. S., and Sorensen, J. H. 1990. "Communication of Emergency Public Warnings: A Social Science Perspective and State-of-the-Art Assessment," Oak Ridge National Lab., TN (USA).
Mitroff, I. I. 2004. *Crisis Leadership: Planning for the Unthinkable*. Hoboken, NJ: John Wiley & Sons.
Mosteller, F., and Tukey, J. W. 1977. "Data Analysis and Regression: A Second Course in Statistics," *Addison-Wesley Series in Behavioral Science: Quantitative Methods*).
Murthy, D. 2011. "Twitter: Microphone for the Masses?," *Media Culture and Society* (33:5), p. 779.
Murthy, D. 2012. "Towards a Sociological Understanding of Social Media: Theorizing Twitter," *Sociology* (46:6), pp. 1059-1073.
Niinimäki, T., Piri, A., Lassenius, C., and Paasivaara, M. 2012. "Reflecting the Choice and Usage of Communication Tools in Global Software Development Projects with Media Synchronicity Theory," *Journal of Software: Evolution and Process* (24:6), pp. 677-692.
O'hara, R. B., and Kotze, D. J. 2010. "Do Not Log□Transform Count Data," *Methods in Ecology and Evolution* (1:2), pp. 118-122.
Oh, O., Agrawal, M., and Rao, H. R. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises," *Mis Quarterly* (37:2), pp. 407-426.
Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A. 2013. "Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters," *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, GA, pp. 380-390.
Pak, A., and Paroubek, P. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *LREc*, pp. 1320-1326.
Park, J., Baek, Y. M., and Cha, M. 2014. "Cross□Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis," *Journal of Communication* (64:2), pp. 333-354.
Park, J., Barash, V., Fink, C., and Cha, M. 2013. "Emoticon Style: Interpreting Differences in Emoticons across Cultures," *ICWSM*.

Pena, G. 2013. "Twitter Alerts: Critical Information When You Need It Most." Retrieved September 14, 2016, from https://blog.twitter.com/2013/twitter-alerts-critical-information-when-you-need-it-most

Pervin, N., Takeda, H., and Toriumi, F. 2014. "Factors Affecting Retweetability: An Event-Centric Analysis on Twitter,").

Pozzi, G., Pigni, F., and Vitari, C. 2014. "Affordance Theory in the Is Discipline: A Review and Synthesis of the Literature,").

Procopio, C. H., and Procopio, S. T. 2007. "Do You Know What It Means to Miss New Orleans? Internet Communication, Geographic Community, and Social Capital in Crisis," *Journal of Applied Communication Research* (35:1), pp. 67-87.

Protalinski, E. 2013. "Twitter Launches Alerts Service in the Us, Japan, and Korea to Keep Users Informed During Emergencies." Retrieved July 20th, 2016, from http://thenextweb.com/twitter/2013/09/25/twitter-launches-alerts-service-in-the-us-japan-and-korea-to-keep-users-informed-during-emergencies/#gref

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., and Bhatt, S. 2013. "What Kind of #Conversation Is Twitter? Mining #Psycholinguistic Cues for Emergency Coordination," *Computers in Human Behavior* (29:6), pp. 2438-2447.

Reis, H. T., and Judd, C. M. 2000. *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press.

Rezabek, L., and Cochenour, J. 1998. "Visual Cues in Computer-Mediated Communication: Supplementing Text with Emoticons," *Journal of Visual Literacy* (18:2), pp. 201-215.

Rice, R. E., and Love, G. 1987. "Electronic Emotion Socioemotional Content in a Computer-Mediated Communication Network," *Communication research* (14:1), pp. 85-108.

Runyan, R. C. 2006. "Small Business in the Face of Crisis: Identifying Barriers to Recovery from a Natural Disaster1," *Journal of Contingencies and Crisis Management* (14:1), pp. 12-26.

Salcito, K. 2009. "Online Journalism Ethics: Gatekeeping," *Retrieved April* (14), p. 2010.

Saunders, C., and Pearlson, K. 2009. "Managing and Using Information Systems," *NY: Wiley*).

Schneider, S. K. 2014. *Dealing with Disaster: Public Management in Crisis Situations*. Routledge.

Sellnow, T. L., and Seeger, M. W. 2013. *Theorizing Crisis Communication*. John Wiley & Sons.

Shannon, C. E. 1949. "Weaver; W.(1949): The Mathematical Theory of Communication," *Urbana*).

Shaw, F., Burgess, J., Crawford, K., and Bruns, A. 2013. "Sharing News, Making Sense, Saying Thanks," *Australian Journal of Communication* (40:1), p. 23.

Shi, Z., Rui, H., and Whinston, A. B. 2014. "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter," *MIS Quarterly* (38:1), pp. 123-142.

Siese, A. 2016. "Twitter's First Tweet Ever Was This Simple, Effective Message from Jack Dorsey." Retrieved July 19th, 2016, from http://www.bustle.com/articles/149156-twitters-first-tweet-ever-was-this-simple-effective-message-from-jack-dorsey

Spence, P. R., Lachlan, K. A., and Burke, J. M. 2007. "Adjusting to Uncertainty: Coping Strategies among the Displaced after Hurricane Katrina," *Sociological Spectrum* (27:6), pp. 653-678.

Spencer, D. H., and Hiltz, S. R. 2003. "A Field Study of Use of Synchronous Chat in Online Courses," *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*: IEEE, p. 10 pp.

Spiro, E., Irvine, C., DuBois, C., and Butts, C. 2012. "Waiting for a Retweet: Modeling Waiting Times in Information Propagation," *2012 NIPS workshop of social networks and social media conference. http://snap. stanford. edu/social2012/papers/spiro-dubois-butts. pdf. Accessed*.

Spiro, E. S., DuBois, C. L., and Butts, C. T. 2013. "Waiting for a Retweet: Modeling Waiting Times in Information Propagation," in: *Neural Information Processing Systems Conference (NIPS) Workshop*. Lake Tahoe, Nevada.

St Denis, L. A., Palen, L., and Anderson, K. M. 2014. "Mastering Social Media: An Analysis of Jefferson County's Communications During the 2013 Colorado Floods," *11th International ISCRAM Conference*, pp. 737-746.

Starbird, K., and Palen, L. 2010. *Pass It On?: Retweeting in Mass Emergency*. International Community on Information Systems for Crisis Response and Management.

Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. 2010. "Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information," *Proceedings of the 2010 ACM conference on Computer supported cooperative work*: ACM, pp. 241-250.

Stein, M. 2004. "The Critical Period of Disasters: Insights from Sense-Making and Psychoanalytic Theory," *Human Relations* (57:10), pp. 1243-1261.

Stephens, K. K., and Barrett, A. K. 2014. "Communicating Briefly Technically," *International Journal of Business Communication*), p. 2329488414525463.

Suh, B., Hong, L., Pirolli, P., and Chi, E. H. 2010a. "Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," *Social computing (socialcom), 2010 ieee second international conference on*: IEEE, pp. 177-184.

Suh, B., Lichan, H., Pirolli, P., and Chi, E. H. 2010b. "Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pp. 177-184.

Sundar, S. S. 2008. "The Main Model: A Heuristic Approach to Understanding Technology Effects on Credibility," *Digital media, youth, and credibility* (73100).

Sutton, J., Gibson, C. B., Phillips, N. E., Spiro, E. S., League, C., Johnson, B., Fitzhugh, S. M., and Butts, C. T. 2015a. "A Cross-Hazard Analysis of Terse Message Retransmission on Twitter," *Proceedings of the National Academy of Sciences* (112:48), pp. 14793-14798.

Sutton, J., Gibson, C. B., Spiro, E. S., League, C., Fitzhugh, S. M., and Butts, C. T. 2015b. "What It Takes to Get Passed On: Message Content, Style, and Structure as Predictors of Retransmission in the Boston Marathon Bombing Response," *PLoS one* (10:8), p. e0134452.

Sutton, J., Palen, L., and Shklovski, I. 2008. "Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires," *Proceedings of the 5th International ISCRAM Conference*: Washington, DC, pp. 624-632.

Sutton, J., Spiro, E. S., Fitzhugh, S., Johnson, B., Gibson, B., and Butts, C. T. 2014a. "Terse Message Amplification in the Boston Bombing Response," *Proceedings of the 11th International ISCRAM Conference*: Pennsylvania State University University Park, PA, pp. 612-621.

Sutton, J., Spiro, E. S., Johnson, B., Fitzhugh, S., Gibson, B., and Butts, C. T. 2014b. "Warning Tweets: Serial Transmission of Messages During the Warning Phase of a Disaster Event," *Information, Communication & Society* (17:6), pp. 765-787.

Sutton, J. N. 2010. *Twittering Tennessee: Distributed Networks and Collaboration Following a Technological Disaster*. Citeseer.

Todorov, A., Chaiken, S., and Henderson, M. D. 2002. "The Heuristic-Systematic Model of Social Information Processing," *The persuasion handbook: Developments in theory and practice*), pp. 195-211.

Twitter. 2015. "About Twitter, Inc." Retrieved January 7, 2015, from https://about.twitter.com/company

Twitter. 2016. "Liking a Tweet." Retrieved August 2, 2016, from https://support.twitter.com/articles/20169874#

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. 2010. "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," *Proceedings of the SIGCHI conference on human factors in computing systems*: ACM, pp. 1079-1088.

Walther, J. B., and D'Addario, K. P. 2001. "The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication," *Social science computer review* (19:3), pp. 324-347.

Wang, X., McCallum, A., and Wei, X. 2007. "Topical N-Grams: Phrase and Topic Discovery, with an Application to Information Retrieval," *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*: IEEE, pp. 697-702.

Warzel, C. 2014. "Why We Favorite Tweets, According to Science." Retrieved August 2, 2016, from https://www.buzzfeed.com/charliewarzel/why-we-favorite-tweets-according-to-science?utm_term=.ewo73gyzqP#.erkxEvd3Mj

Westerman, D., Spence, P. R., and Van Der Heide, B. 2012. "A Social Network as Information: The Effect of System Generated Reports of Connectedness on Credibility on Twitter," *Computers in Human Behavior* (28:1), pp. 199-206.

Westerman, D., Spence, P. R., and Van Der Heide, B. 2014. "Social Media as Information Source: Recency of Updates and Credibility of Information," *Journal of Computer□Mediated Communication* (19:2), pp. 171-183.

Westman, S., and Freund, L. 2010. "Information Interaction in 140 Characters or Less: Genres on Twitter," *Proceedings of the third symposium on Information interaction in context*: ACM, pp. 323-328.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica: Journal of the Econometric Society*), pp. 817-838.

Wikipedia. 2016. "2007 California Wildfires." Retrieved July 17, 2016, 2016, from https://en.wikipedia.org/wiki/2007_California_wildfires

Wilensky, H. 2014. "Twitter as a Navigator for Stranded Commuters During the Great East Japan Earthquake," *Proc. of 11th International Conf. on Information Systems for Crisis Response and Management*, pp. 695-704.

Winn, P. 2011. "Japan Tsunami Disaster: As Japan Scrambles, Twitter Reigns." Retrieved July 10, 2014, from http://www.globalpost.com/dispatch/news/regions/asia-pacific/japan/110318/twitter-japan-tsunami

Wood, M. M., Mileti, D. S., Kano, M., Kelley, M. M., Regan, R., and Bourque, L. B. 2012. "Communicating Actionable Risk for Terrorism and Other Hazards⋆," *Risk analysis* (32:4), pp. 601-615.

Yang, L., Sun, T., Zhang, M., and Mei, Q. 2012. "We Know What @You #Tag: Does the Dual Role Affect Hashtag Adoption?," in: *Proceedings of the 21st international conference on World Wide Web*. Lyon, France: ACM, pp. 261-270.

Zeng, L., Starbird, K., and Spiro, E. S. 2016. "Rumors at the Speed of Light? Modeling the Rate of Rumor Transmission During Crisis," *2016 49th Hawaii International Conference on System Sciences (HICSS)*: IEEE, pp. 1969-1978.
Zhang, W., and Watts, S. A. 2008. "Capitalizing on Content: Information Adoption in Two Online Communities," *Journal of the Association for Information Systems* (9:2), p. 73.
Zubiaga, A., Spina, D., Martinez, R., and Fresno, V. 2015. "Real□Time Classification of Twitter Trends," *Journal of the Association for Information Science and Technology* (66:3), pp. 462-473.