Summer 8-1-2017

Evaluating Relevance and Reliability of Twitter Data for Risk Communication

Xiaohui Liu
University of Southern Mississippi

EVALUATING RELEVANCE AND RELIABILITY OF

TWITTER DATA FOR RISK COMMUNICATION

by

Xiaohui Liu

A Dissertation
Submitted to the Graduate School,
the College of Science and Technology,
and the Department of Geography and Geology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

August 2017

EVALUATING RELEVANCE AND RELIABILITY OF

TWITTER DATA FOR RISK COMMUNICATION

by Xiaohui Liu

August 2017

Approved by:

Dr. Bandana Kar, Committee Chair
Associate Professor, Geography and Geology

Dr. Chaoyang Zhang, Committee Member
Professor, Computing

Dr. David M. Cochran, Committee Member
Associate Professor, Geography and Geology

Dr. Carl A. Reese, Committee Member
Professor, Geography and Geology

Dr. George T. Raber, Committee Member
Associate Professor, Geography and Geology

Dr. Mark Puckett
Chair, Department of Geography and Geology

Dr. Karen S. Coats
Dean of the Graduate School

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

ABSTRACT

EVALUATING RELEVANCE AND RELIABILITY OF

TWITTER DATA FOR RISK COMMUNICATION

by Xiaohui Liu

August 2017

While Twitter has been touted to provide up-to-date information about hazard events, the relevance and reliability of tweets is yet to be tested. This research examined the relevance and reliability of risk information extracted from Twitter during the 2013 Colorado floods using five different approaches. The first approach examined the relationship between tweet volume and precipitation amount. The second approach explored the relationship between geo-tagged tweets and degree of damage. In the third approach, the spatiotemporal distribution of tweets was compared with flood extent. In the fourth approach, risk information from tweets were compared with survey responses obtained in a Department of Homeland Security report about risk communication to determine what people expect to be included in alerts vs. what is communicated via tweets. In the fifth approach, tweets containing top frequent keywords and hashtags were compared with official reports using cosine similarity method. For reliability assessment, contents of relevant tweets were manually compared with official data and images.

The findings indicated that relevant tweets provided information about the event, its impacts, and contained other risk information that public expects to receive via alert messages. Content analysis of images revealed that tweets were also reliable in disseminating information about damages and impacts. Given that the Crowdsourcing and Citizen Science Act (2016) authorizes agencies to use crowdsourcing to increase public response to emergency alerts, the methodology used in this study could be used by emergency management personnel (EMP). The findings could also be used by EMPs to identify relevant and reliable tweets. However, out of 1 million English tweets, 14% were relevant, 3% were reliable, and 0.44% were geo-tagged to Boulder. Although the geographically relevant tweets could have eliminated possible misinformation shared by "*outsiders*", very limited percentage of social media was useful, relevant, and reliable. Furthermore, social media analytics was time consuming and computationally intensive, which may not be feasible for EMP. Future research should focus on developing a matrix to assess data quality of crowdsourced data, automating implementation of analytics, and developing a citizen-science based approach to gather focused data about hazard events.

ACKNOWLEDGMENTS

Many thanks to my advisor, Dr. Bandana Kar, and my dissertation committee, Drs. David M. Cochran, Jr., George Raber, Chaoyang (Joe) Zhang, and Andy Reese, for advice and assistance in completing my doctoral study and dissertation successfully.

DEDICATION

I want to thank my parents, Zhizeng and Rongxia Liu, my soul mate, Haipeng Tang, my sister, Xiaomin and Xiaona Liu, and my extended family, especially, my mother-in-low, Guoxia, for her support throughout my doctoral study. The arrival of my daughter, Lucia Tang, in the final year of my Ph.D. allowed me to experience the joy of being a new mom, yet challenged my school work and dissertation. I am proud to have handled the demands of being a new mom and graduate within three years! I also want to thank my husband, Haipeng, for his continued support throughout my doctoral study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

DHS	Department of Homeland Security
EMA	Emergency Management Agency
EMP	Emergency Management Personnel
FEMA	Federal Emergency Management Agency
JSON	JavaScript Object Notation
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Services
OSM	Open Street Map
SVM	Support Vector Machine
USGS	United States Geological Survey
WEA	Wireless Emergency Alert

CHAPTER I - INTRODUCTION

1.1 Overview.

The purpose of this research was to examine the relevance and reliability of risk information extracted from Twitter in the context of emergency response, specifically, during risk communication following a hazard event. The main objectives of this study were to: (i) determine the relevance of risk information obtained from Twitter, and (ii) examine the reliability of Tweets in increasing situational awareness of public about a hazard event. This chapter introduces the research problem, research questions that were examined, and expected outcomes and intellectual merits of this research.

1.2 Problem Statement.

1.2.1 Natural Hazards.

Although the occurrence of natural hazards is not new, since 1970s, there has been a significant increase in severity and frequency of natural hazards, specifically, hydro-meteorological events (i.e. floods and tropical cyclones) (UN, 2013). Because of continued population growth and urbanization of coastal communities as well as rising sea level due to global warming and sinking or subsiding landmass, the natural and built environments of coastal communities are at higher risk to both coastal and riverine flooding events (Karl, 2009; Karl et al., 2009; National Research Council, 2010). These events tend to affect human lives and cause property damage more than any other types of sudden-onset hazard events. Several major floods that have occurred across the U.S. over the past decade including the widespread 2015/2016 winter floods which impacted the area surrounding Ohio River, Missouri River, Mississippi River, and spring floods across Texas, Louisiana Arkansas, and Mississippi (USGS, 2017) are examples of the significant adverse impacts of flooding events.

1.2.2 Risk Information.

Successful mitigation of natural hazard impacts on society and physical environments is closely linked to effective risk communication, which focuses on disseminating information about an impending disaster to citizens to help them take timely and preparatory actions to reduce hazard impacts (Covello, 1992). Due to the unpredictable nature of hazards, it is paramount to gather information regarding the nature, extent, and intensity of a hazard event, possible areas at risk, and possible impacts of the event. Additionally, other information, such as rescue support, safety information, resource location/allocation, and uncertainty of the hazard(s) are indispensable for emergency management. Thus, risk communication, a critical component of emergency management, could directly influence the emergency management activities of affected communities (Covello et al., 2012; Hughes et al., 2014; Lundgren and McMakin, 2013). Effective risk communication, however, depends on dissemination of timely, relevant, complete, and reliable information to enable EMPs, local public and other stakeholders undertake mitigation actions (Horita et al., 2013). Dissemination of tardy and incorrect information about a hazard event and its potential impacts could lead to slow response with serious consequences and failure of hazard mitigation (Erskine and Gregg, 2012).

1.2.3 Crowdsourcing in Emergency Response.

Crowdsourcing, a web 2.0 based phenomenon, is a relatively new concept. Yet it is widely adopted in a variety of fields including emergency management (Estellés-Arolas and González-Ladrón-De-Guevara, 2012). The three essential components of crowdsourcing are: “who” (individuals or organizations that form the “crowd”), “where” (internet as the venue), and “what” (information, service, or data are the content to be generated). Social media (i.e. Twitter, Facebook etc.) is a popular crowdsourcing venue that has been used for risk communication (Gao et al., 2011a). During the emergency response, social media serves as a hub for impacted people to post new information, obtain desired information, and share information obtained from other channels about a that otherwise may not be distributed to a broader audience (Preis et al., 2013).

Given that crowdsourcing allows the creation of online social networks, it has been extensively used during past hazard events by organizations, communities, and individuals to obtain and share information, coordinate disaster relief efforts, or seek assistance. For instance, in the aftermath of the Haiti earthquake (2010), crowdsourced information was generated through mapping sites including CrisisCamp Haiti, OpenStreetMap, Ushahidi, and GeoCommons (Zook et al., 2010). The information obtained via the web enabled first responders to coordinate search and rescue efforts on the ground (Heinzelman and Waters, 2010; Zook et al., 2010). First-hand information obtained on the ground were posted and shared via wikis and other collaborative workspaces, making those sites the main sources to share knowledge for involved U.S. government agencies (Yates and Paquette, 2011). Population movement information was also obtained by combining geographic positions of mobile phones before and after Haiti earthquake (Bengtsson et al., 2011). Likewise, during the 2014 Oso mudslide in Washington State, county emergency management officials used social media to update the public about the event, its impacts, and actions underway to mitigate impacts (Center for Digital Government, 2015). Even the United States Geological Survey (USGS) uses

Twitter data in their earthquake alert system (Bahir and Peled, 2015). The Crowdsourcing and Citizen Science Act (2016) and Social Media Improvement Act (2015) have also authorized emergency management agencies (EMA) to use social media to undertake emergency preparedness and response activities.

1.2.4 Data Relevance and Reliability.

Although crowdsourced data demonstrates promising prospects in promoting effective risk communication, quality of data and information obtained via crowdsourcing is a major concern that hinders its use and necessitates implementation of citizen science based approaches as a way to improve data quality. Some concerns stem from the following facts. First, crowdsourced data is generally unstructured, making it difficult to filter out credible and actionable information. Second, without complying with a standard for data generation, the quality of crowdsourced data varies. Although recent studies have focused on various aspects of crowdsourced data quality, such as positional accuracy, completeness, semantic accuracy (Arsanjani et al., 2013; Fan et al., 2014), credibility of tweets (Gupta and Kumaraguru, 2012), speed at which information is updated, relevance, reliability, and accessibility to information are still the aspects of data quality that are of significant concern by data users (Liu et al., 2016).

This research focuses on evaluating the relevance and reliability of crowdsourced data (i.e., tweets) using a case study of the 2013 Colorado flood. In this research context, relevance refers to data fitness, i.e. available and obtained data meet user needs (Grady and Lease, 2010; Vuurens and de Vries, 2012). Reliability means data are trustworthy, i.e., they are dependable in terms of content (Cai and Zhu, 2015; Mendoza et al., 2010). In the context of risk communication, relevant information may not be reliable because any information about a specific hazard event can be treated as relevant, such as mention of the time or location of the event. However, reliable information needs to be relevant before being considered reliable and trustworthy. For instance, a report of damage that occurred a few months prior in an impacted area should be categorized as misinformation in the context of risk communication and thus is not reliable. Based on this reasoning, this research examined the relevance and reliability of risk information obtained from tweets to answer the following research questions.

1.3 Research Objectives and Outcomes.

Despite social medias' popularity, the data available from these sites suffer from lack of veracity (IBM Big Data and Analytics Hub, 2016), and certainty that the data are useful in near real-time. This research examined the following objectives and research questions to assess the quality of content derived from tweets obtained for the 2013 Colorado floods.

1.3.1 Objectives.

1. Identify relevance of risk information obtained from tweets.
    a. What approaches can be used to evaluate relevance of the risk information?
    b. How different is risk information extracted from tweets to what public expects to be included in warning messages?
    c. How different is risk information extracted from tweets to those obtained from official warnings and damage assessment reports?
2. Identify reliability of the risk information obtained from tweets.

a. What approach can be used to assess reliability of contents extracted from tweets?
b. How reliable is the relevant risk information obtained from tweets?

1.3.2 Outcomes.

An important outcome of this research is a methodological framework to extract and evaluate relevance and reliability of risk information obtained from tweets that could facilitate risk communication, increase situational awareness, and public response to natural hazards. The extracted risk information could help emergency management agencies and first responders to coordinate relief efforts, and mitigate hazard impacts to lives and properties; meanwhile, the public could benefit from extracted risk information as it would increase their awareness and aid them in undertaking protective actions to minimize hazard impacts. The findings of this research could also be used in composing effective risk messages to increase public's response to alerts and warnings.

Other outcomes of this research include: (i) gaining knowledge about the extent to which crowdsourced data can be used for risk communication; (ii) combining geospatial data and official assessment reports to increase reliability of risk information for disaster preparedness and response; (iii) demonstrating an integrated use of spatiotemporal analysis and natural language processing techniques to extract relevant and reliable information from crowdsourced data for hazard events. In summary, this interdisciplinary research draws from geospatial and computer science to answer the research questions, and contributes to the broader literature of risk communication, and Geographic Information Science, specifically, to the research on data quality.

CHAPTER II - BACKGROUND

This chapter provides an overview of risk communication, and compares two types of risk communication: hierarchical and network-based. A comprehensive review of literature on crowdsourcing and its use during risk communications is presented, followed by a discussion of data quality of crowdsourced data including its relevance and reliability and prevalent methods of data quality assessment. Finally, a summary of the limitations of using crowdsourcing for risk communication is presented.

2.1 Risk Communication.

Risk communication, a principal element of emergency management, has varying definitions. However, risk communication is inherently defined as “the process of exchanging information among interested parties about the nature, magnitude, significance, or control of a risk” (Covello, 1992). Risk communication is paramount to governments, organizations, businesses, and individuals because it provides information about potential hazards, possible impacts or damages, and countermeasures, which directly influence the alert and warning message recipients’ decision-making process. Traditional risk communication follows a hierarchical, top-down, and centralized approach to deliver risk information to at-risk populations about potential adverse impacts of specific hazards (Gladwin et al., 2007). However, this approach often fails to motivate the public to respond positively to messages due to lack of trust in the message source or misunderstanding of the information provided by the message (Colley and Collier, 2009; Twyman et al., 2008). It is less likely that people would take actions to mitigate hazard impacts if they have limited trust in the message source or message content or do not know what actions to take. Public perception and cognition of risk is influenced by education, past experience, and socio-economic characteristics, which subsequently influence their decision to follow warnings (Kar and Cochran, 2015b).

Unlike traditional risk communication, network-based risk communication uses a bottom-up and collaborative approach that allows both impacted and interested populations to share unlimited information about a hazard, irrespective of its geographic location and time (Kar, 2015). Social networking sites (e.g. Facebook and Flickr), short-blog services (e.g. Twitter), and social mapping sites (e.g. Open Street Map and GeoCommons), are representatives of a network-based communication approach. During Hurricane Sandy, Flickr, a popular website for sharing photographs, was flooded with an extensive volume of pictures labeled with the terms “Hurricane”, “Sandy” or “Hurricane Sandy”, thereby serving as a valuable data source for emergency management (Preis et al., 2013). During the Haiti earthquake, Wikis, an online community forum, allowed users to collectively build textual and visual websites; this was the first time Wikis was used as a knowledge sharing site (Yates and Paquette, 2011). OpenStreetMap and GeoCommons were also used to produce and access spatial data and maps to assist geographically distributed volunteers Haitian earthquake relief efforts (Zook et al., 2010).

From a user’s perspective, effective risk communication requires dissemination of timely, complete, and accurate risk information to impacted populations by emergency management agencies before, during and after an emergency event. Risk information should reflect the current state of a hazard event (i.e., its location, possible risk, and potential impacts) as well as recommend mitigation actions to at-risk populations through continuous monitoring of the ongoing hazard event (Degrossi et al., 2014). For instance, topographic conditions and drainages systems are not the only environmental characteristics responsible for a flooding incident. Therefore, continuous monitoring of a flood incident should capture information about precipitation amount, flood extent, potential areas under flood, ongoing impacts, such as damage to roads, potential actions to undertaken for protection, and information about communities at risk. While information about the severity and extent of a flood event could be obtained from meteorological stations and through prediction of flood impacts, information about potential damages in at-risk communities could not be obtained from traditional information sources, i.e., National Weather Service, local EMA. Timely delivery of updated information about flood risk and possible consequences is crucial to ensure flood mitigation, which is only possible if local EMAs continuously analyze geospatial data sets (i.e. remote sensing images) in real-time or obtain crowdsourced data from public.

The ineffectiveness of hierarchical risk communication, as identified by other studies, is also caused by other factors, such as information sources (e.g., the credibility of risk experts or communicators), message design and style (e.g., the prior requirement to deliver 90-character Wireless Emergency Alert (WEA) messages), the delivery channel (e.g., radio, siren and cell phones), and the socio-economic characteristics of target audience (Covello, 1992; Kar et al., 2016; Mileti and Peek, 2000). By contrast, network-based risk communication has the potential to improve these aspects as it allows incorporation of risk information in any format (images, videos, and remote sensing imagery) from different sources (NOAA, Weather.com, USGS etc.), and dissemination of messages via diverse crowdsourcing sites (i.e., Twitter, Facebook, and other social media sites) to a broader audience regardless of their geographic locations (Palen et al., 2009;

Sheppard et al., 2012; Sutton et al., 2008). Because of its numerous advantages, network-based risk communication should be used to augment the traditional approach.

2.2 Crowdsourcing.

Crowdsourcing, coined in 2006 as a business model, was defined as “the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call” (Howe, 2006). It has evolved into a problem-solving model in the fields of software development, photography (Schenk and Guittard, 2009), and other, beyond business world (Brabham, 2008). This trend is partially due to the advent of the Web 2.0, which has changed the role of online users from passive web page readers or viewers to active contributors (Degrossi et al., 2014; Heipke, 2010; Rouse, 2010). Advancements in social computing have also contributed to the creation of online social networks such that the volume of user-generated information and data have increased tremendously. The end result of crowdsourcing is increase attention, interest, and use in different sectors including academia (Callison-Burch et al., 2015; Geiger et al., 2012; Hudson-Smith, 2014; Tripathi et al., 2014), industry (Callison-Burch et al., 2015; Geiger et al., 2012), business (Kern, 2014; Rouse, 2010), and government (Brabham, 2013).

Given its usability, not surprisingly, crowdsourcing has drawn much attention from practitioners, researchers, and agencies in the field of emergency management over the years (Chan, 2014; Chatfield and Brajawidagda, 2014; Frommberger and Schmid, 2013; Holderness and Turpin, 2015; Horita et al., 2013). Desktop or mobile applications designed to harness social media generated content about an impending hazard (i.e., flood, tornado, earthquake) have proven effective in disseminating information about the hazard, reporting damages, seeking help from stakeholders, and organizing relief efforts (Chatfield and Brajawidagda, 2014; Frommberger and Schmid, 2013; Holderness and Turpin, 2015). During the early aftermath of the 2010 Haiti earthquake, crowdsourced information was generated through mapping sites including CrisisCamp Haiti, OpenStreetMap, Ushahidi, and GeoCommons by geographically dispersed volunteers connected through the internet who coordinated with first responders on the ground (Heinzelman and Waters, 2010; Zook et al., 2010) by sharing and posting first-hand information about conditions on the ground via wikis and other collaborative workspaces, thereby making those platforms the main knowledge-sharing mechanism for involved U.S. government agencies (Yates and Paquette, 2011). Population movement information was also created by combining geographic positions of mobile phones that were used by affected people before and after the Haiti earthquake (Bengtsson et al., 2011).

Other than providing near-real time information during disaster response, crowdsourcing has been valuable in post-disaster assessment of damage caused by earthquakes (Barrington et al., 2012). Due to the lack of high-quality remote sensing images, non-authoritative data, such as social media, news, and mobile phone data, were used to assess damages to roads in New York by the flood following Hurricane Sandy at 2012 (Schnebele et al., 2014). During Hurricane Sandy, the Federal Emergency Management Agency (FEMA) recruited a team of volunteers from public and private institutions to analyze tweets and Instagram photos to identify communities requiring resources, and to process images as part of the OpenStreetMap-MapMill project to assess damages (Chan, 2012; Heaton, 2013).

With the advancements in geospatial technologies (i.e. GPS-enable devices) and location based services, novice citizens have become familiar with embedding geoinformation in their generated crowdsourced data. A great example is the Google Map Maker that allows communities to contribute local knowledge by editing and moderating features to improve the Google Map experience (Google, 2017). The geotagged crowdsourced data help pinpoint locations where people may be requiring help to access critical facilities (e.g., shelters, hospitals, churches, etc.) and provide information about transportation routes (highways, airports, country roads) (Ware, 2003). The popularity of mapping sites, i.e., OpenStreetMap and Ushahidi as well as the abundant map products that were generated during the Haiti earthquake are examples of crowdsourced data containing geospatial information about critical facilities and possible damages (Heinzelman and Waters, 2010; Soden and Palen, 2014; Zook et al., 2010).

Several factors have contributed to the widespread use of crowdsourcing in emergency management. First, the era of web 2.0 provides easily usable online platforms (i.e. Twitter, Facebook, etc.) for users to generate content and collaborate with others. This enables creation of a virtual community that enables public to create and share information, and helps form interest groups unconstrained by geographic locations and time. The role of online platforms is more critical during a hazard event when there is a great need for collaboration and generation of emergency-related information. Therefore, more and more people are turning to online platforms to gather risk information, share hazard related updates, and seek assistance before, during and after hazard events (Landwehr and Carley, 2014; St Denis et al., 2014; Zhao and Zhu, 2014).

Second, while geospatial data sets are extensively used in emergency management, these data often suffer from errors due to varying scales and resolutions, and are often unavailable in near real-time (Gao et al., 2011b; Pu and Kitsuregawa, 2014). Crowd-sourced data, however, fill the data gaps that exist with traditional geospatial and authoritative data sources, and are available in high temporal resolution to be useful during emergency management (Schnebele et al., 2015). As stated above, the geolocation enabled crowdsourced information serves as indispensable supplementary data to traditional geospatial data including remote sensing images, Census data, and other mapping data sets. The critical role of location information can never be exaggerated as it is valuable in targeting communities in need of resources and relief efforts, and in dispatching first responders for search and rescue operations.

Third, the characteristic of crowdsourcing that makes it so powerful is the ability to generate updated information that caters to the need for timely risk information during emergency responses. A very good example is the USGS Twitter earthquake detector that allows detection of aftershocks and dissemination of alerts within few minutes of receiving information from seismometers and Twitters (Earle et al., 2012). Any delay in risk information could lead to slow responses and may in turn cause serious consequences to human lives and property as was seen during Hurricane Katrina (Cole and Fellows, 2008). With the support of broadband and easy-to-use desktop or mobile applications, sending out first-hand texts or images is as quick as pressing a button, which far outweighs the information update speed of traditional media when journalists need to be sent to the spot (Sutton et al., 2008). Moreover, because the crowd serves as human sensors, they have the potential to capture first-hand information, which makes crowdsourcing unparalleled by any other type of traditional media.

Finally, crowdsourcing of risk information also promotes public participation in risk communication. Studies have indicated that public's growing desire to participate in policies pertaining to a variety of societal and environmental situations including risk communication influences the at-risk populations' response to messages (Covello and Sandman, 2001; Gurabardhi et al., 2005; McComas et al., 2009). For instance, in a U.S. watershed planning initiative, mail surveys were sent out to determine if the impacted population would like to participate in the watershed management efforts. Qualitative analysis of the 1% survey responses revealed that participants were most helpful in identifying and prioritizing issues, and that the participatory planning increased awareness of watershed conditions, strengthened inter-agency coordination, and assisted consensus building on resource management (Duram and Brown, 1999).

Public participation in risk communication refers to the involvement of citizens in assessing risk, disseminating risk information, and responding to risks (McComas et al., 2009). The expansion of social media and networking sites (e.g., Facebook, Twitter, Google + etc.) has become a major driving force for citizens to participate in risk communication (Gouveia et al., 2004; Krimsky, 2007; Laituri and Kodrich, 2008). Furthermore, because these sites allow near real-time delivery of warnings to a broad audience, EMAs, government officials, and first responders also use these technologies to disseminate and gather information about a hazard event (Laituri and Kodrich, 2008; Palen, 2008; Smith, 2010). For instance, during the 2010 Oso mudslide in Washington, the local county officials used social media to update news and inform the public;

meanwhile, local residents used social media as major information source (CDG, 2015). The 2015 Social Media Improvement Act and the 2016 Crowdsourcing and Citizen Science Act are motivating EMAs to work with public during emergency management activities. Therefore, using social media during risk communication might increase public responses to risk information and enable agencies to assist the public.

2.3 Crowdsourced Data Quality.

Since anyone can generate crowdsourced data, its data quality is a big concern from emergency management perspective. Specifically, information overloading is the first obstacle encountered by crowdsourced data users as massive amounts of user-generated content from diverse social media sites could be overwhelming for people to read and filter information, let alone validate. As of February 2017, Twitter alone has 313 million monthly active users (Twitter, 2017) and on average 500 million tweets (postings) are sent per day (InternetLiveStats, 2017). The number of active users and tweets skyrocket during large-scale crisis, as more people are turning to Twitter to read breaking news and keep abreast of event updates (Sutton et al., 2008). While Twitter is a news and social networking site where users post messages and interact with others, other social network/media sites with diverse focuses, including mapping sites (i.e., OpenStreetMap, Ushahidi, Bing Maps, Google Maps), video sharing sites (i.e., YouTube, YouKu, Yahoo), and photo sharing sites (i.e., Instagram, Flickr, Pinterest), are also continuously generating huge data sets. Without a strategic plan to select certain social network/media sites, it is difficult to decide which site(s) to use for obtaining data during a disaster response phase, which could eventually impact quality of data.

Besides information overloading, several reasons contribute to the difficulty in assessing quality of crowdsourced data. First, crowdsourced data often lack metadata (Meek et al., 2014). Without metadata, information about the author or creator, time and date of creation, location, device used to generate data, purpose, and standard used to create data cannot be confirmed (Meek et al., 2014). Therefore, the data tend to lack credibility (the extent to which the data can be relied upon to represent what it is supposed to represent) and authenticity (the guarantee that the data have not been manipulated) (Castillo et al., 2011; IGI-Global, 2016; Liu et al., 2016). Second, citizens who generate data tend to be from various backgrounds with varying perceptions and educational background, and possess distinct life experiences. Because there is no protocol in place to collect crowdsourced data, even during an emergency event, the data quality tends to be questionable (Kelling et al., 2015). Additionally, with the rise of crowdsourcing, robot-controlled social media accounts, commercial spam, collective attention spam, and hoaxes are all common phenomena occurring in social media (Lee et al., 2014; Starbird et al., 2014), which also impacts quality of crowdsourced data.

2.3.1 Crowdsourced Data Relevance and Reliability.

Data quality can be defined as a measure of fitness for specific purposes in a given context (SearchDataManagement, 2017). Accuracy, completeness, update speed, relevance, reliability, and accessibility are major components of data quality (Wang and Strong, 1996). Despite obvious differences, depending upon the purpose and context of data use, these components tend to overlap. In this research, relevance and reliability, the two components that have drawn significant attention from data users were assessed

(Gupta and Kumaraguru, 2012; Senaratne et al., 2017). A contextual definition of each component is provided in the following sections for the purpose of this research.

Relevance is “the condition of being connected or appropriate to what is being considered” (Cai and Zhu, 2015; Oxford, 2017a), or “if it has a logical, sensible relationship to the finding it supports” (Morgan and Waring, 2004). Although its connotations vary with context of usage, there are some shared common characteristics, such as the timeliness of relevant data and the closeness of the data to its context (Morgan and Waring, 2004). Reliability means “the quality or state of being trustworthy or consistently well” (Oxford, 2017b; United Nations Economic Commission for Europe, 2017), or “is a state that exists when data is sufficiently complete and error free to be convincing for its purpose and context” (Morgan and Waring, 2004). Other terms with similar meanings, such as credibility, trustworthiness, or correctness, are often used interchangeably (Hiltz et al., 2012). In this research, relevance refers to how well the information meets user needs in terms of what is represented and reliability corresponds to trustworthiness such that the data are dependable in terms of information provided by them to be used for emergency management activities.

2.4 Techniques for Analyzing Crowdsourced Data.

The methodologies used to analyze crowdsourced data for emergency response purpose could be classified into three categories: information extraction and content analysis, information classification, and social network analysis. Most of the research in information extraction and content analysis are computer algorithm-based applications relying on keyword searching in hazard response, with an emphasis on developing new methods or optimizing existing ones to extract useful or actionable information more efficiently and accurately (Imran et al., 2013). Some research has focused on automating information extraction so that self-contained, actionable, and useful information relevant to a hazard could be extracted (Atefeh and Khreich, 2015; Caragea et al., 2011).

Among the studies focusing on information extraction and analysis, social media content and sentiment analysis is significantly conducted to facilitate damage assessment (Cervone et al., 2016; Kryvasheyeu et al., 2016; Schnebele et al., 2014), to increase situational awareness (Vieweg et al., 2010), and to coordinate risk communication (St Denis et al., 2014). Kryvasheyeu et al. (2016) revealed that the spatiotemporal distribution of Hurricane Sandy related messages could help with real-time monitoring and assessment of the disaster itself (Kryvasheyeu et al., 2016). By searching hurricane related keywords ("sandy", "storm", "hurricane", etc.), the authors were able to plot the volume of messages with pre-defined keywords across cities with varying distance to the hurricane path. The result displayed that Twitter activity is related to the proximity of the region to the hurricane path. Additionally, the authors found that per-capita Twitter activity had a strong relationship with per-capita hurricane damage at county and zip-code level. Vieweg et al. (2010) compared Twitter posts generated during the Oklahoma Grassfires of April 2009 and the Red River Floods of March and April 2009 and found that geo-tagged tweets are more likely to contain situational information and thus are more likely to be retweeted (Vieweg et al., 2010). The authors proposed the development of an information extraction software with a microblog to enhance situational awareness during emergencies with regard to evacuation, sheltering, animal management, and damage (Vieweg et al., 2010).

The research focusing on information classification and event-detection are mostly based on keywords (Imran et al., 2013). Machine learning methods for extracting information nuggets from disaster-related tweets have been found to have accelerated disaster response and alleviate human and property loss (Imran et al., 2013). For instance, Sakaki et al. (2010) developed a support vector machine (SVM) based approach to classify real-time events about earthquakes by undertaking semantic analysis using keyword searches in tweets. The technique was later applied to an earthquake reporting system in Japan, which could detect earthquakes with high probability. In another study, Caragea et al., (2011) implemented SVM for text message classification and compared four types of feature representations for learning SVM classifiers. The experiment revealed that abstract features, which are generated by grouping "similar" features-based SVM classification outperformed other feature representation-based classifications. While these techniques extract information nuggets for some nugget types, such as source and casualties, with high accuracy, they fail to perform as well when extracting time, location, and caution/advice nuggets.

The studies researching social network communication in hazard response emphasize on examining communication patterns among affected communities. Following the trend of social network analysis, Cheong and Cheong (2011) analyzed the interaction of active Twitter users with local authorities during the 2010-2011 Australian floods to find the influential authorities/members during emergency communications. Social network analysis helped the authors identify active users in the online community, online communication patterns, and frequencies of communication, which are of significance to EMA personnel in guiding emergency communications and managing resources for response and relief efforts. In a similar study, Stephens and Poorthuis (2015) explored network density and network transitivity during crisis situations, and found that smaller networks are more socially clustered while large networks are physically dispersed. The authors also found that Twitter networks are more effective at transmitting information at local levels and within smaller networks than larger networks.

2.4.1 Data Quality Assessment Techniques.

Given the widespread use of social media data and presence of data quality issues, several studies have focused on assessing the quality of crowdsourced data and validate it. OpenStreetMap (OSM) is a worldwide crowdsourced spatial data layer that provides information about street networks. Forghani and Delavar (2014) compared street network data from OSM for Tehran, Iran with institutionally referenced geospatial databases and developed a new metric for quality assessment of OSM data. The metric used the following four measures: road length, minimum bounding geometry, directional distribution, and median center to compare the two datasets. The authors also evaluated crowdsourced data and reference data at a grid level using heuristic metrics such as Minimum Bounding Geometry area and directional distribution (Standard Deviational Ellipse), and found that the OSM data had a high quality. Eckle and de Albuquerque (2015) conducted a similar study to assess the quality of OSM by comparing maps created by remote mappers and expert mappers. Qualitative assessment of the OSM data sets suggested that a misinterpretation of roads, buildings, and other infrastructures could cause distortion on remotely produced maps. Brown et al. (2015) evaluated positional accuracy and data completeness of data collected from Google Map for conservation planning in comparison with empirically identified biological /conservation points. The results indicated that crowdsourced data may be "good enough" to complement biological data but cannot be regarded as the main data source for conservation planning.

Unlike approaches available to assess crowdsourced spatial data, several methods exist for evaluating quality of non-spatial crowdsourced data. These methods can be divided into the following categories: classification of information content or sources (Thomson et al., 2012), implementation of majority decision or control group evaluation (Hirth et al., 2013), and use of a reputation system designed for quality check (Alabri and Hunter, 2010). In a study, Thomson et al., (2012) examined the source credibility of tweets shared in relation to the Fukushima Daiichi nuclear power plant disaster in Japan (2011) by classifying users by location, language, and type (individual or institutions). The results indicated that institutional sources are more credible than individual sources, anonymous users tend to cite from less credible sources, Japanese-language tweets are more likely to reference credible third-party sources, and users proximal to the disaster post or share more credible tweets. Another study analyzed the credibility and relevance of tweets based on fourteen worldwide high impact events of 2011 (Gupta and Kumaraguru, 2012). This study adopted a supervised machine learning and relevance feedback approach to prioritize content features, such as the number of unique characters, swear words, emotions, and number of followers. The results revealed that the algorithm could automate credible information extraction from Twitter with high confidence. Other methods used to assess relevance include web page ranking to evaluate web links' relevance to queries (Page et al., 1999); supervised machine learning to prioritize relevance of tweets to queries (Duan et al., 2010); and crowdsourcing assessment based on Amazon Mechanical Turk (Alonso and Mizzaro, 2012).

Several methods exist for reliability assessment of crowdsourced data, which include data cleaning, automatic validation, authoritative data comparison, linked data analysis, and semantic harmonization (Meek et al., 2014). Mendoza et al., (2010) used the relationship between information propagation through Twitter network and the interrelationship among Tweeter users as a proxy to measure reliability of tweets. Human experts are also used in reliability testing, such as differentiation and justification of perceived “true incidents” from Twitter messages. Inspired by human assessment approach, Castillo et al., (2011) validated the reliability of tweets through human readers’ comparison of incidents identified in tweets with authoritative datasets.

Depending on the context of data usage, purposes, researchers’ or data users’ background, existing methods in assessing relevance and reliability vary. Using control data or authoritative data (Comber et al., 2013; Meek et al., 2014), experts’ knowledge (See et al., 2013), and crowdsourcing data (Goodchild and Li, 2012; Grady and Lease, 2010) are the prevalent approaches used to assess relevance and reliability. However, the studies focusing on classification and analysis of social media content mostly concentrate on developing algorithms and rarely incorporate other types of existing data related to hazards, such as meteorological and geospatial data (e.g., precipitation extent and volume in flood studies) (Cheong and Cheong, 2011; St. Denis et al., 2014) and (e.g., digital elevation models (DEM) in earthquake or landslide studies) (Caragea et al., 2011). Despite their effectiveness, due to the lack of geographic data/information about the hazards under study and other types of data sets set, these studies tend to be biased.

2.5 Summary

Risk communication necessitates dissemination of time-critical and rapidly changing information to help with decision-making processes (Beckman et al., 2007; Thompson et al., 2006). Soon after a hazard, information about the nature of the hazard, impacted area, mitigation measures to take, among other information need to be communicated to help with response and recovery efforts. Decisions that cannot be made in a timely fashion may adversely impact emergency management efforts. The collection, processing, and dissemination of risk information in the era of Web 2.0 could take advantage of fast updated and content-rich crowdsourced data.

Data quality of crowdsourced data is a big concern. Despite significant number of studies assessing crowdsourced data quality, very little research has been done to assess the quality of risk information extracted from social media (i.e. Twitter) to be useful during risk communication. As more and more EMAs are using social media for risk communication, lack of metadata about crowdsourced data could make these data unusable. Furthermore, as the information posted by EMAs on social media are perceived as credible and more likely to be shared (St. Denis et al., 2014; Starbird and Palen, 2010), it is crucial to assess relevance and reliability of these data.

Although authoritative data should be used as reference data for relevance assessment, proxy indicators for relevance could also be used if reference data are unavailable or could not be used for comparison (Senaratne et al., 2017). In this research, the relevance of tweets obtained during the 2013 Colorado flood were evaluated using 5 distinct approaches with help of geospatial data sets (i.e., precipitation, flood extent, and degree of damage) and authoritative data obtained from reports as proxy indicators of the information extracted from tweets and. The reliability of relevant tweets was evaluated by comparing the contents with human readers' judgement.

CHAPTER III - METHODOLOGY

This chapter provides an overview of the methodology implemented to answer the research questions. A discussion of the study site, the data sets, and analytics that were used to process and analyze the data sets, extract information as well as assess relevance and reliability of the information is also provided in this section.

3.1 Study Site.

The 2013 Colorado floods occurred in the Front Range, EL Paso County, Boulder County, and portions of the Denver metropolitan area. The devastating flash flooding was a result of historically severe precipitation that started on September 9 and continued until September 18, 2013. Figure 3.1 depicts the northern counties that were worst hit, severely hit, moderately hit, least hit areas based on extent of damages they experienced.

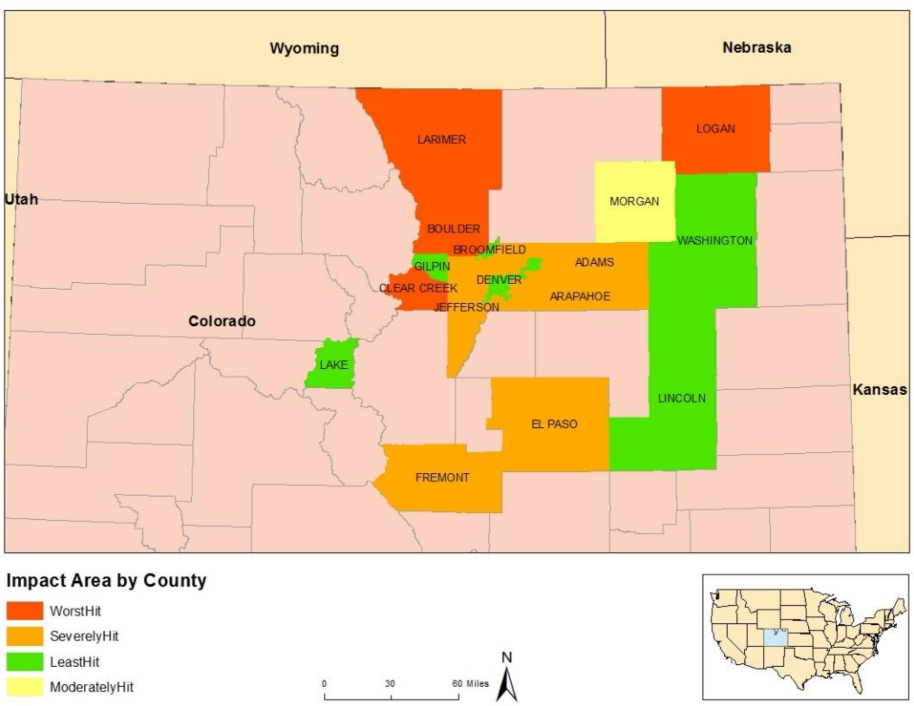

Figure 3.1 Study Site (FEMA, 2014)

On September 9, 2013, a cold front developed over Colorado that caused heavy rain during September 9th and September 15th. Figure 3.2 shows the hourly precipitation accumulations throughout the storm event at several different locations across the Front Range. It is evident from the figure that Boulder County was a worst hit area with 9.4 inches of precipitation on September 12th, which was comparable to the county's average annual precipitation. Except Boulder County other places had little to no accumulations until September 15th, and experienced a small amount of precipitation on September 15th.

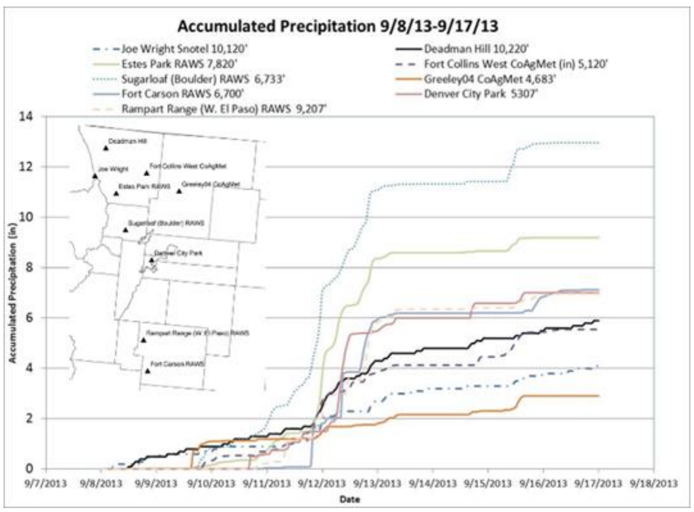

Figure 3.2 The 2013 Colorado Floods accumulated precipitation (CCC, 2013)

3.2 Data Sets and Processing.

Other than tweets, a variety of geospatial and survey data sets were obtained from several sources, which were used as auxiliary and reference data. A discussion of the data sets, and processing and analysis techniques used with each data set is discussed below.

3.2.1 Tweets of 2013 Colorado Floods.

3.2.1.1 Tweets.

Historical tweets were purchased from Twitter Inc. using the following keywords pertaining to location names (Colorado, Boulder, Front Range, El Paso County and

Boulder County, Denver metro), and hazard event and its impacts (flash flooding, flooding, rain 2013, emergency, impact, damaged bridges and roads, damaged houses, financial losses, evacuate, and evacuation). The tweets were purchased for a 10-day duration from September 9th to September 18th when majority of flooding occurred.

3.2.1.2 Tools & Preprocessing.

A total of 1,195,183 tweets were obtained in JSON (JavaScript Object Notation) format. Given the volume and unstructured format of the data, MongoDB was used to store, process, and analyze the data. MongoDB is an open-source cross-platform database for unstructured data including document-based data (e.g. JSON) that uses dynamic schemas. Robomongo, the client interface to visualize and interact with MongoDB was used to query and analyze tweets. Figure 3.3 demonstrates the steps implemented to process the data before implementing analytics.

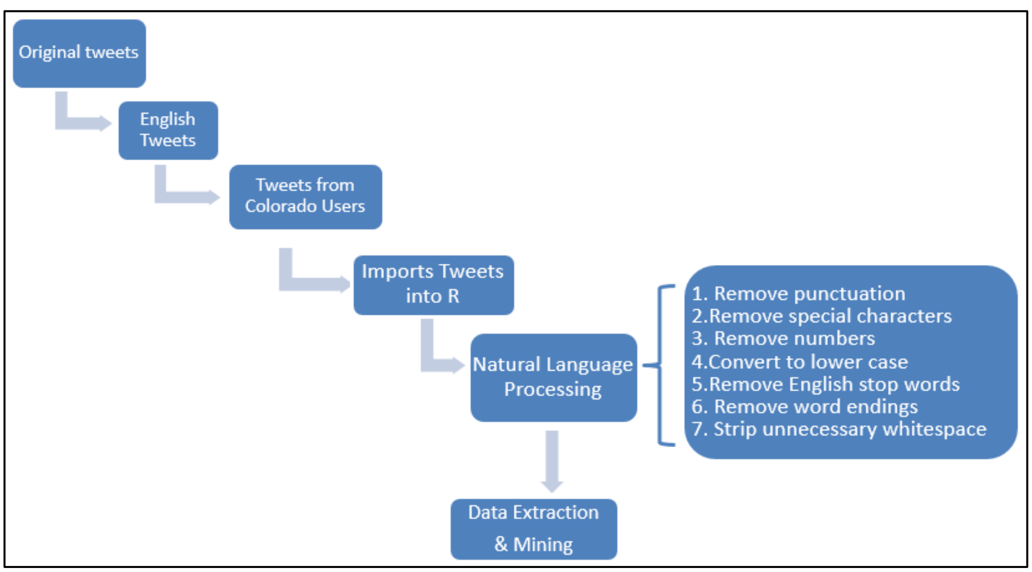

Figure 3.3 Flow Chart of Tweets Processing

Of the total tweets, 85% were in English, 1.38% had geo-location information available, and 0.44% of the tweets that were geo-tagged were found to be generated in Colorado. The 0.44% of the geo-located tweets were used in this study as these tweets were considered to be generated by those who experienced or witnessed the floods. The geographic relevance of the tweets eliminated the possibility of including misinformation or rumors generated by geographical "*outsiders*" that were not on the scene and enabled extraction of relevant risk information.

Table 3.1

Descriptive statistics of the Twitter dataset

Collection Name	**Number of Tweets**	**Percentage**
Total tweets	1,195,183	100%
Tweets in English	1,017,024	85%
Tweets with geo-location	16,551	1.38%
Tweets geo-located in Colorado	5,202	0.44%

After processing the tweets using steps identified in Figure 3.3, a list of top frequent words was created, which was used to mine flood-related tweets for further analysis. Figure 3.4 depicts the histogram of top frequent 30 words and associated word cloud. Evidently, the top frequent words were Colorado, Boulder, Denver, flooding, warning, September, etc., which are also relevant.

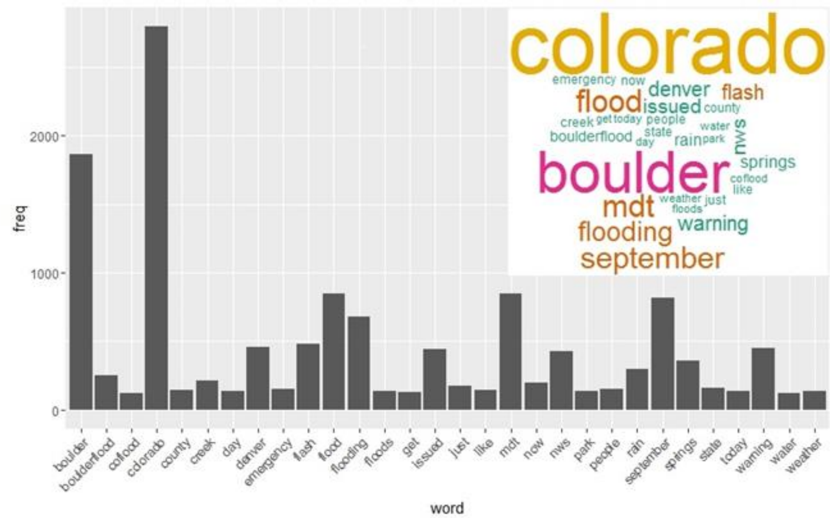

Figure 3.4 Top frequent words and corresponding word cloud

Apart from top frequent words, hashtags, words and phrases preceded by a pound sign (#), which represent messages on a specific topic were also used to extract flood-related tweets. Table 3.2 lists some of these hashtags, such as "Colorado", and "flooding" that coincide with top frequent words. High frequency hashtags, such as "NeverForget" and "GodBlessAmerica", are misleading, and therefore, were not used in tweet extraction to avoid extraction of irrelevant information.

Table 3.2

Top 10 Hashtags

1	Colorado
2	boulderflood
3	Coflood
4	cowx
5	NeverForget
6	flooding
7	GodBlessAmerica
8	news

Table 3.2 (Continued)

9	CORecall
10	Denver

3.2.2 Geospatial Data.

The devastating flooding event that started on September 9th, 2013 was a historically most severe flood since 1995 that Colorado had experienced. For this study, the hourly precipitation data was obtained in text format for all rainfall gauge stations located in the study site from the National Climate Data Center (CCC, 2013). These data were used to understand the relationship between temporal volume of tweets and precipitation volume, and to evaluate the temporal relevance of tweets to the flood event. Further discussion of this relationship and findings is presented in following chapters.

To understand the spatial distribution of tweets with respect to the flood impact areas, spatial data pertaining to flood extent was obtained from the City of Boulder (City of Boulder, 2014a). This data was generated by the City of Boulder using field surveys, Digital Globe Worldview satellite imagery, public input from Boulder crowdsourced online apps, and information from affected property owners. Street network data obtained from the City of Boulder was also used to evaluate relevance and reliability of tweets with regard to flood damages to roads and streets (City of Boulder, 2014b).

3.2.3 Survey Data.

Responses to a survey titled *Public Perceptions of Warning and Alert Messages* that was used in a Department of Homeland Security funded project to examine the Mississippi Gulf Coast residents' understanding of alerts and warnings was used as ancillary data for this study (Kar and Cochran, 2015b). The survey contained a question that investigated participants' opinions about the contents that should be included in an emergency alert message; the selectable choices were "*impact zone*", "*time frame*", "*recommended actions*", "*when to take action*", "*evacuation routes*", "*shelter location*", and "*who to contact for help*". Each of these choices reflects a critical aspect of risk information. The responses to this question represent public expectations of the content to be included in an alert message, which should not vary from place to place. Therefore, the results of this survey question were compared with information extracted from tweets to assess relevance of tweets in disseminating risk information.

3.2.4 NOAA Warning/alert Messages.

Warning/alert messages sent by NOAA-NWS during the 2013 Colorado flooding event were downloaded from NOAA Weather Forecast Office at Boulder in text format (US Department of Commerce, 2013). The messages document meteorological forecasts, observations, public watches, warnings, advisories, and other information as the flood unfolded. Therefore, these alert/warning messages were used as official reference information in evaluating reliability of tweets. However, instead of using individual messages, all messages were combined as a single text message for analytics purpose.

3.2.5 Official Warning and Damage Assessment Reports.

Different from NOAA warning/alert messages, which convey possible threats due to heavy rain and flood, assessment reports include information pertaining to post-event evaluation of the event and its impacts. There reports provide situational awareness about flooding and summarize damage to properties and infrastructures in the affected areas.

Additional official records, such as newspaper articles and town hall meeting briefings, that validated incidents and/or facts (i.e., damage to specific roads) were also used.

3.3 Analytics and Techniques

This section discusses the steps that were implemented to extract relevant risk information from tweets, and assess their reliability in disseminating valid risk information. Five different approaches were used to evaluate relevance of tweets based on extracted risk information, and binary change detection approach was used to determine reliability of tweets (Figure 3.5).

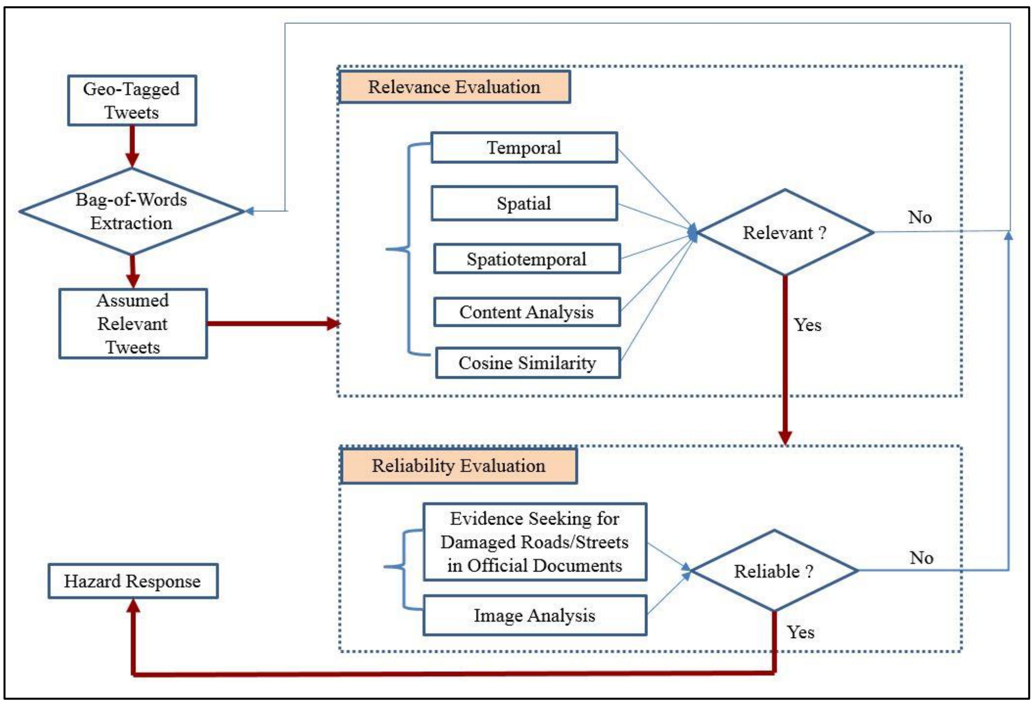

Figure 3.5 Research Workflow

3.3.1 Extraction of Relevant Risk Information: Bag-of-words Model.

To filter tweets pertinent to the 2013 Colorado flooding event, a bag-of-words model was used, which is widely used in natural language processing and information retrieval (Filliat, 2007; Tirilly et al., 2008; Wallach, 2006). A bag-of-words contains topic-specific search terms to measure the relevance of a document to the search terms.

The bag-of-words in this study included Colorado, Boulder, Front Range, El Paso County and Boulder County, Denver metro, flash flooding, flooding, rain 2013, Coflood, cowx, CoRecall, news, emergency, impact, damaged bridges and roads, damaged houses, financial losses, evacuate, and evacuation. These words were derived from two sources: top frequency words and hashtags corresponding to impacted locations, event impacts, and emergency management activities. The top frequency words served as indicator of popular topics that were covered by tweets; top hashtags were used to identify messages on specific topics. Using the bag-of-words, tweets were extracted from the list of geo-tagged tweets present in Colorado (0.44% of the original data set), which were considered to be relevant, and were used in subsequent analysis.

3.3.2 Survey Responses to Warning/alert Message Content.

The analysis of the survey responses (discussed above) indicated that citizens hold various viewpoints towards the contents of warning/alert messages. Despite their varying preferences, all participants indicated that the following risk information should be included in warning/alert messages: "*nature of disaster*", "*impact zone*", "*time frame*", "*recommended actions*", "*when to take action*", "*evacuation routes*", "*shelter location*", and "*who to contact for help*". The findings of descriptive statistics conducted on the survey responses were compared with the risk information extracted from tweets to find if difference exists between what people expect and what was conveyed via social media.

3.3.3 Evaluation of Relevance.

Five approaches using five official reference data sets were used to evaluate relevance of tweets and extract risk information from relevant tweets. A brief discussion of each approach is presented in the following section.

1. *Temporal*: Exploring the relationship between temporal distribution of tweet volume and precipitation amount could be used to determine temporally relevant tweets. Precipitation was used as the reference data as it was the primary cause of severe flooding in the study site that lasted almost a week. Continuous heavy rainfall not only caused flooding, but induced heavy traffic on Twitter about the flooding event. Therefore, comparing the trend of tweet volume and precipitation amount was one way to evaluate relevance of tweets based on the topic.
2. *Spatial*: If tweets are relevant to the floods, their spatial distribution should not be random. Rather, they should demonstrate a correlation with the degree of damage experienced across the study site. Therefore, the relationship of spatial distribution of tweets and degree of damage was examined statistically.
3. *Spatiotemporal*: Because of the most intensive flood and damage, the 2013 Colorado floods is also called the Boulder flood. Hence, tweets that were generated by residents of Boulder could be more representative and more reflective of the spatial and temporal distribution of the flood and its associated impacts. The spatial distribution of tweets over a six-day period with respect to the flood extent was mapped to identify relevance tweets.
4. *Content analysis*: The seven choices from the survey correspond to seven aspects of risk information that were used to classify risk information extracted from tweets. No matter which category a tweet belongs to, it could be considered

relevant. By comparing the percentage of risk information extracted from survey responses and tweets, the difference between what people expected in warning/alert messages and what was conveyed in tweets could also be detected.

5. *Cosine similarity comparison*: Cosine similarity comparison is a vector space model mostly used for comparing document relevance or similarity. In this research, contents of tweets were compared with official warnings or damage reports using cosine similarity approach. The approach calculates the cosine angle between two non-zero vectors (or two documents) and the similarity score represents the degree of relevance (0 means no relevance and 1 means very relevant or the same). (Indurkhya and Damerau, 2010).

   Documents are represented as vectors, and each vector holds a place for every term in the document collection. Binary approach is used for converting a document into a vector, namely, a value of 1 or 0 was used to represent a present or an absent term. The binary approach was chosen instead of using term frequency for conversion because tweets consist of similar or repetitive compressed messages that could significantly increase term frequency in tweets, thereby making the tweets and official warning or damage reports incomparable. Given two documents, d1 and d2:

$$\text{Similarity} = \cos(\theta) = \cos(d1, d2) = \frac{d1 \cdot d2}{\|d1\|\|d2\|}$$

   where • indicates vector dot product, ‖ d ‖ is the length of the vector d.

   Due to the unstructured nature of tweets and their 140-character limit, they could not be directly compared with official reports. Furthermore, despite the short length of each tweet, the large volume of tweets is likely to contain repetitive words. Therefore, instead of direct comparison of documents, top 50 frequent words and top 10 non-redundant hashtags were compared with top 50 frequent words from both NOAA warning/alert messages and official damage assessment reports. The rationale for using top frequent words and top hashtags was that top frequent words from tweets represent information from social media, and top frequent words from combined warning and damage assessment report represent risk information from the authority. If a certain degree of similarity exists between a tweet and the report, then the tweet is relevant to the event.

6. *Relevance score:* MongoDB has a built-in function ($meta) that returns a matching score based on the provided terms to match with. Top 10 frequent words and top 10 hashtags were used as the terms to compute a relevance score for each of the 5202 geo-tagged tweets in Colorado.

3.3.4 Evaluation of Reliability.

In the context of risk communication, relevant information may not be reliable such as mention of the time or location of the event. However, reliable information must be relevant before being considered trustworthy. Once relevant tweets were extracted, the tweets were manually analyzed to identify names of damaged roads and streets, and the posted time of each tweet. The identified roads and streets were used as keywords to search for related information in official damage assessment reports and news reports. If a discussion of the roads/streets or the immediate neighborhoods were found in the official reports, the damage was reported to have happened around the same time as the posted tweets, and similar flood situations were described, then the tweets were considered reliable. Images of the impacted roads and streets that were identified from the tweets were also compared with those obtained from the reports/newspaper archives to assess reliability of tweets.

CHAPTER IV – RESULTS

4.1 Evaluation of Relevance.

4.1.1 Temporal Trend of Tweets Volume vs. Precipitation Amount.

To investigate the relationship between tweet volume and precipitation over time, the daily volume of tweets and total precipitation across all rain gauge stations in Colorado from September 11th - September 15th were obtained. Both data sets were normalized to have the values range between 0 and 1 for comparison. Table 4.1 lists daily distribution of tweet volume and precipitation amount. Figure 4.1 plots the two data sets, and the correlation between them. It is apparent from Figure 4.1 that tweet volume increased with increase in precipitation amount, and experienced a significant increase on September 12th. Both tweet volume and precipitation dropped continuously after September 12th until both increased by a small amount on September 14th. A Pearson correlation between tweet volume and precipitation concentration resulted in a correlation coefficient of 0.778 ($p = 0.05$), which indicated the presence of a very strong relationship between the two variables. Therefore, it could be concluded that relevant tweets were produced on days when Boulder experienced significant rainfall and flooding.

Table 4.1

Normalized Tweets Volume and Precipitation

Date	**Tweets**	**Precipitation**
September 11th	0.03	0.19
September 12th	0.41	0.35
September 13th	0.33	0.20
September 14th	0.11	0.11
September 15th	0.12	0.14

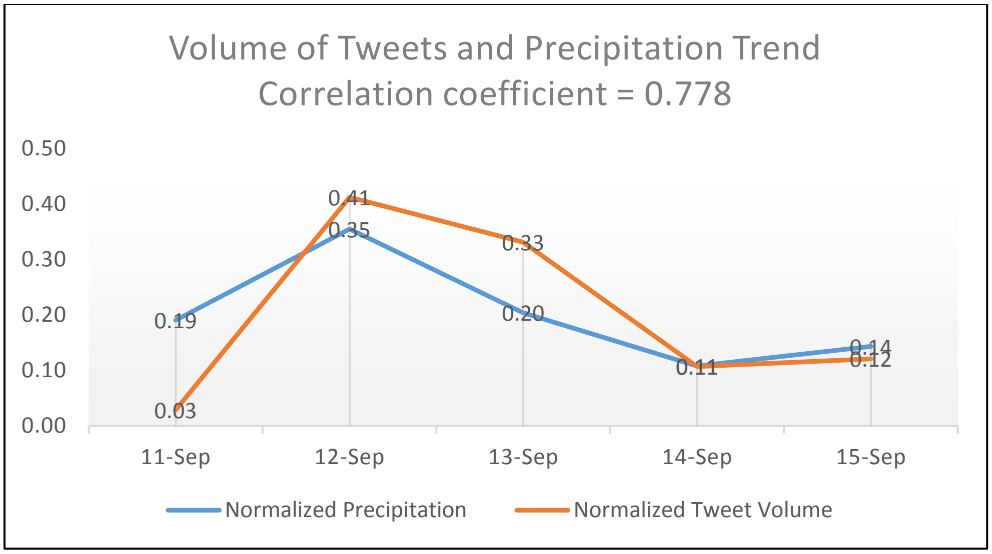

Figure 4.1 Correlation between volume of tweets and precipitation (CCC, 2013)

4.1.2 Spatial Distribution of Tweets vs. the Degree of Damage.

To understand the extent to which volume and spatial distribution of tweets can reflect the spatial distribution of degree of damage, the tweets were aggregated by city, and overlaid with the impacted area map (Figure 4.2). A visual interpretation of the map clearly reveals that the tweets were not randomly distributed, but rather were concentrated in counties/cities that experienced severe damages, some of which have a large population. Table 4.2 lists the cities plotted in Figure 4.2 along with volume of tweets generated in each city, total population of each city, and the degree of damage experienced by each city ("Colorado City Rank," 2016). It is evident from Table 4.2 that tweet volume is dependent on population of a city (high population density means high tweet volume) and is also influenced by the degree of damage. Denver, Colorado Springs, and Fort Collins are among the top four cities by population and by Tweet volume; with far larger populations than other cities, their tweet volume correspond to their population. However, the higher tweet volume for the following five cities - Boulder, Longmont, Broomfield, Centennial, and Loveland is a result of flood damage. Therefore, it could be concluded that tweet volume is impacted by both population and damage extent.

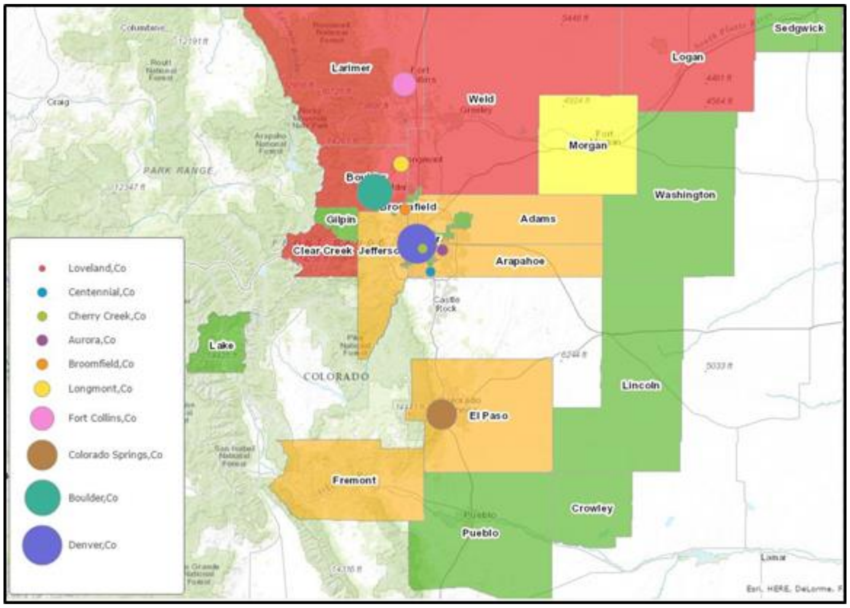

Figure 4.2 Tweets across damaged counties

Table 4.2

Rankings of City by Tweet Volume vs. Population

City	By Tweet Volume	Tweet Volume	Degree of Damage	By Population	Population
Denver	1	16053	Severe	1	682,545
Boulder	2	12665	Worst	11	456,568
Colorado Springs	3	1776	Severe	2	359,407
Fort Collins	4	1311	Worst	4	161,175
Longmont	5	540	Worst	13	109,741
Broomfield	6	362	Severe	15	107,349
Aurora*	7	315	Severe	3	92,088
Centennial	9	231	Severe	9	75,182
Loveland	10	164	Worst	14	65,065

Note: Cherry Creek is a neighborhood in Denver, so it was removed from the city list.

To determine the extent to which tweet volume is influenced by city population and degree of damage, tweet volume and population were converted to a scale of 0 to 1 and plotted in a stacked line chart (Figure 4.3). Tweet volume (blue line) decreases as population decreases with an exception of a trough and a crest. Being the worst damaged city, Boulder ranks the second in tweet volume but eleventh in population and therefore creates a trough in population curve (Freedman, 2013). Aurora is the only exception among the cities that ranked much higher in population than in tweet volume, which might be due to less severe damage in Aurora. Given the dependency of tweet volume on both population and degree of damage, it could be concluded that the extracted tweets are relevant to the flood event.

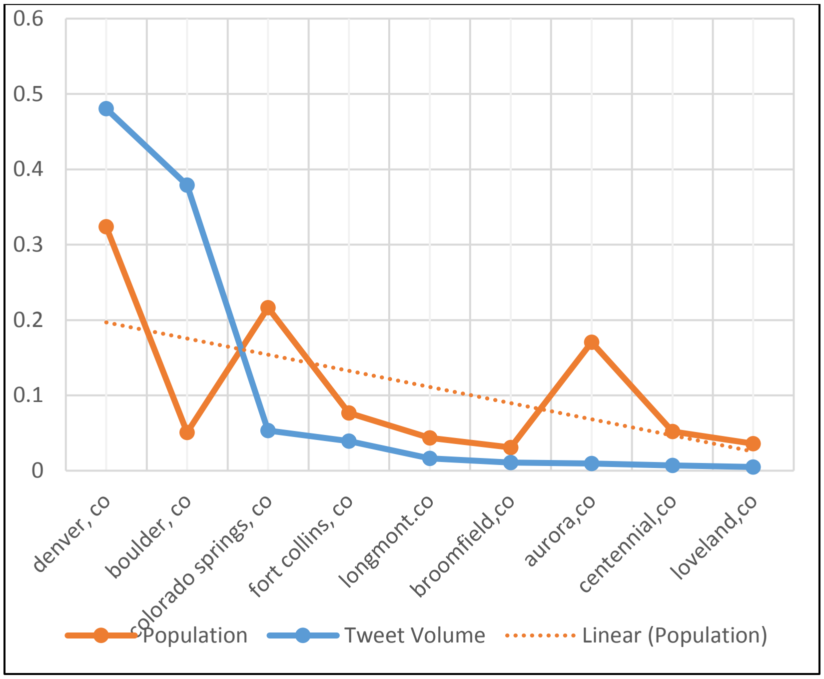

Figure 4.3 Tweet volume vs. city population

4.1.3 Spatiotemporal Analysis of Tweets.

Because of the intensity and severity of floods in Boulder, the 2013 Colorado floods is known as the Boulder flood. Therefore, tweets geo-tagged to Boulder were examined to reflect the spatiotemporal distribution of flood area. Figure 4.4 depicts the spatial distribution of geo-tagged tweets within Boulder city limit over six days (September 10^{th} – September 15^{th}).

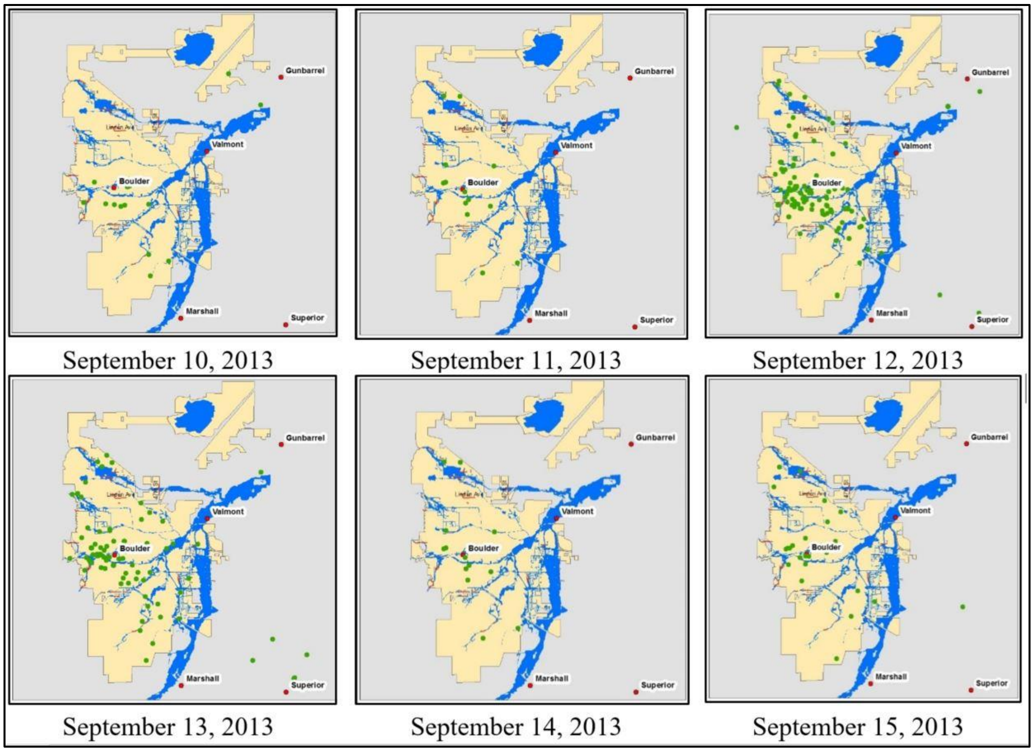

Figure 4.4 Spatiotemporal distribution of tweets across Boulder

The time-series tweet distribution (a green dot indicates one or more tweets if several tweets are from the same location, the blue represents the flood extent across Boulder) in Figure 4.4 indicate that the spatial proliferation of tweets occurred on September 12th and September 13th when the heaviest precipitation and subsequent flooding occurred. By contrast, tweets generated on other days are fewer and are sparsely distributed. It is evident from the spatial distribution of tweets with respect to flood extent that geo-tagged tweets are concentrated along the flooded river/creek channels rather than spread across the city. Even though the tweets were extracted based on their geographic location, the above-mentioned findings prove that the tweets are relevant to the flood event. Imposing geographical constraints to extract relevant information from social media was found to be useful and could narrow the massive data to a small portion for effective use for emergency management purpose.

4.1.4 Content Analysis.

The risk communication that was conducted along the Mississippi Gulf Coast revealed that the most critical contents that should be included in an emergency alert message are nature of disaster, followed by "*impact zone*", "*time frame*", "*recommended actions*", "*when to take action*", "*evacuation routes*", "*shelter location*", and "*who to contact for help*". A content analysis of relevant tweets was conducted using a list of keywords (Table 4.3) to extract risk information belonging to the above-mentioned categories. The results of the analysis (Figure 4.6) were then compared with survey responses (Figure 4.5), which indicated that no matter which category a tweet belongs to, the tweet could be considered relevant.

Table 4.3

Keyword for Each Category of Content Analysis

Category	**Keywords**
Contact for help	Help, need, assistance
Damage, loss, and road closure	Flooded, road, basement
Shelter location	Shelter, church, place, center
Recommended action	Action, evacuate, alert, siren, warning, stay safe, stay dry, stay inside, higher ground
Impact zone	Boulder
Nature of disaster	Flood, flooding

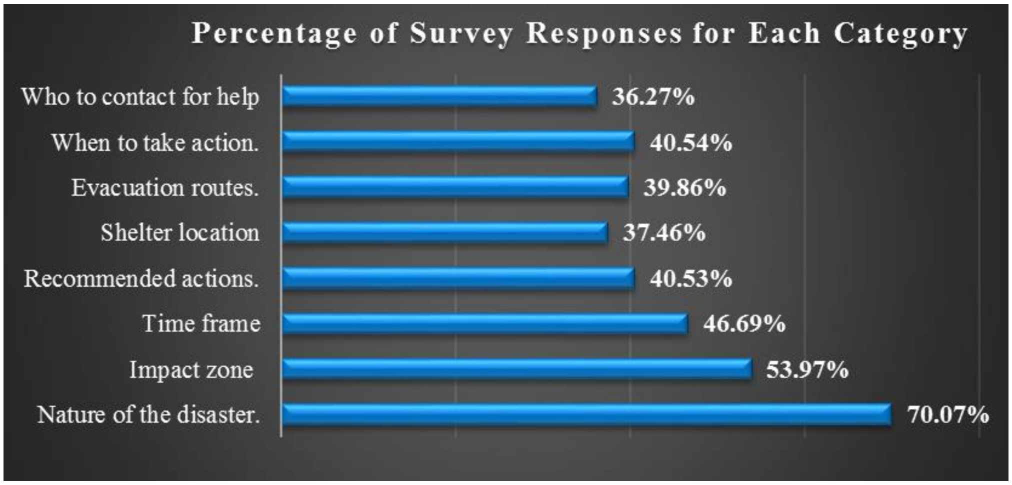

Figure 4.5 Percentage of survey responses for each category

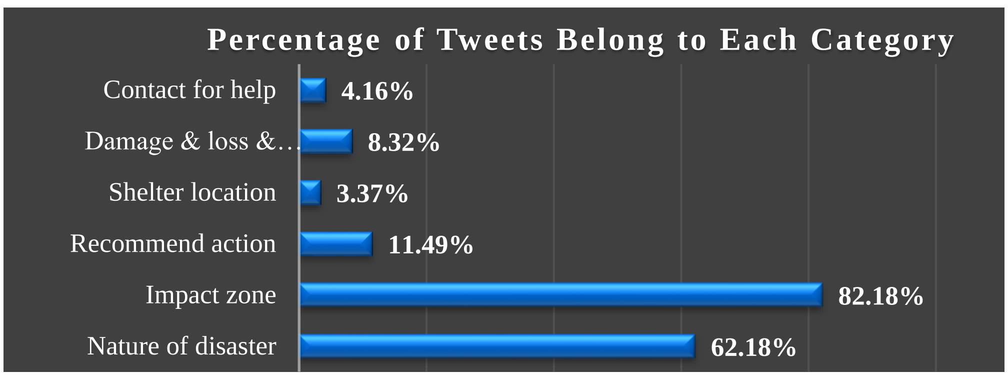

Figure 4.6 Percentage of risk information from tweets

The survey responses (Figure 4.5) revealed that 70.07% of respondents expect information about *nature of the disaster* in alert/warning messages, 53.97% respondents require information about *impact zone*. Almost 40 – 45 % participants indicated their preference to have information about *time frame*, *recommended actions*, and *when to take action* in messages, and only 36 – 40% participants expect information about *evacuation routes, shelter location*, and *who to contact for help* in warning messages. The content analysis of tweets resulted in 62.18% and 82.18% tweets discussing *nature of the disaster* and *impact zone*, respectively (Figure 4.6) followed by *recommended action* (11.49%), *damage & loss* (8.32%), *contact for help* (4.16%), and *shelter location* (3.37%).

Evidently, the percent of survey responses belonging to each category of risk information differ from the percentage of tweets in each category. Because the keywords used to extract tweets in each category influenced the results, a better selection of keywords is needed to extract the content to eliminate bias. Nonetheless, both the survey and tweet analysis revealed that the public requires information about *nature of the disaster* and *impact zone* rather than other aspects of risk, which is reasonable because citizens' need for information regarding *contact for help* or *shelter location* is dependent on their socioeconomic condition, degree of preparations, and/or past experiences with similar situations. Likewise, *recommended action* or *damage & loss* is meaningless unless *the nature of the disaster* and *impact zone* are known.

4.1.5 Cosine Similarity Comparison.

Cosine similarity comparison generates a similarity score representing the degree of relevance between documents. One of the documents used in the analysis contained top 50 frequent words from tweets along with top 10 non-redundant hashtags (see Appendix B), and the other document contained top 50 frequent words from NOAA warning/alert messages and official damage assessment reports (see Appendix B). Several methods can be used to create document vectors, such as raw term frequency and binary weights. Raw term frequency approach includes the frequency of occurrence for the term in each document in the vector, and binary weights approach considers the presence (1) or absence (0) of a term in the vector.

While tweets are a collection of keywords-centric documents, NOAA alert messages and official damage assessment reports are story-based documents with contextual content. This inherent distinction made it difficult to create vectors for comparison using raw frequency approach. Therefore, the presence or absence of a term in respective documents was used in a binary weights approach. Before converting each document to a vector (Table 4.4), the following steps were implemented: (i) eliminate special character (i.e. “â€”), (ii) remove meaningless character combinations (i.e. “wfos”, “awips”), (iii) combine words of different forms (i.e. “colorado”, “colo”,“coc”), (iv) eliminate adjective or auxiliary words that are general in meaning and have no relationship to flood (i.e. “great”, “may”, and “love”).

Table 4.4

Document Vectors of Tweets and Official Reports

Terms	Tweets	Official Reports
Boulder	1	1
Center	1	1
closed	1	0
coflood	1	0
colorado	1	1
county	1	1
creek	1	1
denver	1	1
emergency	1	0
flash	1	1
flood	1	1
flooding	1	1
flows	0	1
forecast	0	1
front	0	1
heavy	0	1
help	1	0
hydrologic	0	1
issued	1	0
news	1	0
noaa	0	1
nws	1	1
park	1	1
precipitation	0	1
rain	1	1
recommendation	0	1
river	1	1
road	1	0
safe	1	0
september	1	1
springs	1	1
stream	1	1
warning	1	0
water	1	1
weather	1	1

Cosine similarity scores range from 0 (meaning dissimilar or not relevant) to 1 (meaning very similar or the same), and in-between values indicates intermediate similarity or relevance. The cosine similarity score in this research is:

Similarity = cos (45.5°) = cos (d_{tweets}, $d_{official\ reports}$) = 0.7.

A relevance score of 0.7 indicates that the two term lists are inclined to be similar. Thus, conclusions can be made that the extracted tweets are relevant to official warning messages and damage assessment reports in terms of content.

4.1.6 Relevance Score

A relevance score was generated for each of the 5202 geo-tagged tweets in Colorado. Table 4.5 lists eight randomly selected tweets in descending order of their respective relevance score. Take the first, sixth, and eighth tweet for example with high relevance score, these tweets contain more flood relevant risk information than lower scored ones. The relevance score varied between 1.3 (lowest score) and 3.79 (highest score) for the 5,202 tweets. The lowest scored tweet, "*Denver is a mess ;(Flooding!*" (767th tweet in the database) rarely had any relevant information regarding the flood event. About 14% of the tweets based on their score were found to be relevant.

Table 4.5

Relevance Score of Geo-Tagged Tweets in Colorado

	Tweet	**Relevance Score**
1	Boulder Flash Flood: Four Mile Creek being flooded in Boulder, Colorado after several days of rain http://t.co/R86BI2kXec #iReport	3.79
2	80720: Flash Flood Warning issued September 11 at 3:23PM MDT until September 11 at 6:15PM MDT by NWS Boulder http://t.co/qi9DvK1pP7	3.44
3	Flooding on the Boulder Creek #boulderflood @ Boulder Creek http://t.co/Brdi9YM2MO	2.91
4	Shout out to Tweeps in Denver and Boulder with flooding. Stay safe! 3 dead so far due to flash floods.	2.45
5	Evacuations for all along boulder creek north to at least spruce. Or go south if on that side. Do not cross boulder creek #boulderflood	2.25
6	Flash flooding along Fourmile Creek.	1.83
7	Colorado flooding: How you can help: With deadly flooding inundating communities across Colorado, many are asking... http://t.co/7Nxds82RXq	1.69
8	Boulder's still gorgeous even after a storm @ University of Colorado Boulder http://t.co/r3ymLfkFG5	1.42

4.2 Evaluation of Reliability.

4.2.1 Evaluation of Text Content.

The findings of five different approaches (discussed above) indicated that the extracted tweets were relevant to the 2013 Colorado flood. Manually, from the relevant tweets, the name of damaged roads/streets, time of impact, and type of impact, were extracted (Table 4.6 and Appendix C).

Table 4.6

Example of Identified Roads/Streets

	Roads/streets	Posted Time	Associated Risk Information
1	West of Broadway	09/12 03:02	Boulder Creek is about to spill its bank.
2	Broadway & Arapahoe Avenue	09/12 05:30	Water at Boulder Creek has come up 2.5 feet in
3	8th Street & Marine Street	09/12 05:52	Gregory canyon drainage overtopping the underground culvert, flowing onto 8th St. near Marine.
4	28th Street & Colorado Avenue	09/12 06:09	Knee deep water at 28th St & Colorado Ave.
5	15th Street	09/12 08:39	River taking back Boulder neighborhood
6	Highway 36 underpass	09/12 22:23	It's raining! It's pouring!
7	8th Street between University of Colorado and Marine	09/13 03:22	…basically, a raging torrent.
8	30th Street & Foothills	09/13 00:49	Colorado Avenue is closed between 30th and Foothill.
9	30th Street	09/13 01:08	Water is coming up through drains on 30th and Colorado Ave…this could get ugly.
10	Highway 36	09/13 01:30	Barely make it out of Boulder. Couldn't get to hwy 36.
11	Highway 36	09/13 02:33	Highway 36 is flooded, not way out.
12	Highway 36 & Foothills	09/13 05:32	Over 3 feet of water flooding.

The extracted information of roads and streets that were damaged by the flood were used as keywords to search for related information in official damage assessment reports and news articles. If information about the same roads/streets or the immediate neighborhoods were found in the reports, and similar flood situations were described, it was concluded that the impacts occurred roughly around the same time as the posted tweet and that the tweets were reliable. Figure 4.7 displays the roads/streets identified in Table 4.6, which are marked on the map with a serial number. A detailed description of reliability assessment of the tweets listed in Table 4.6 is presented below.

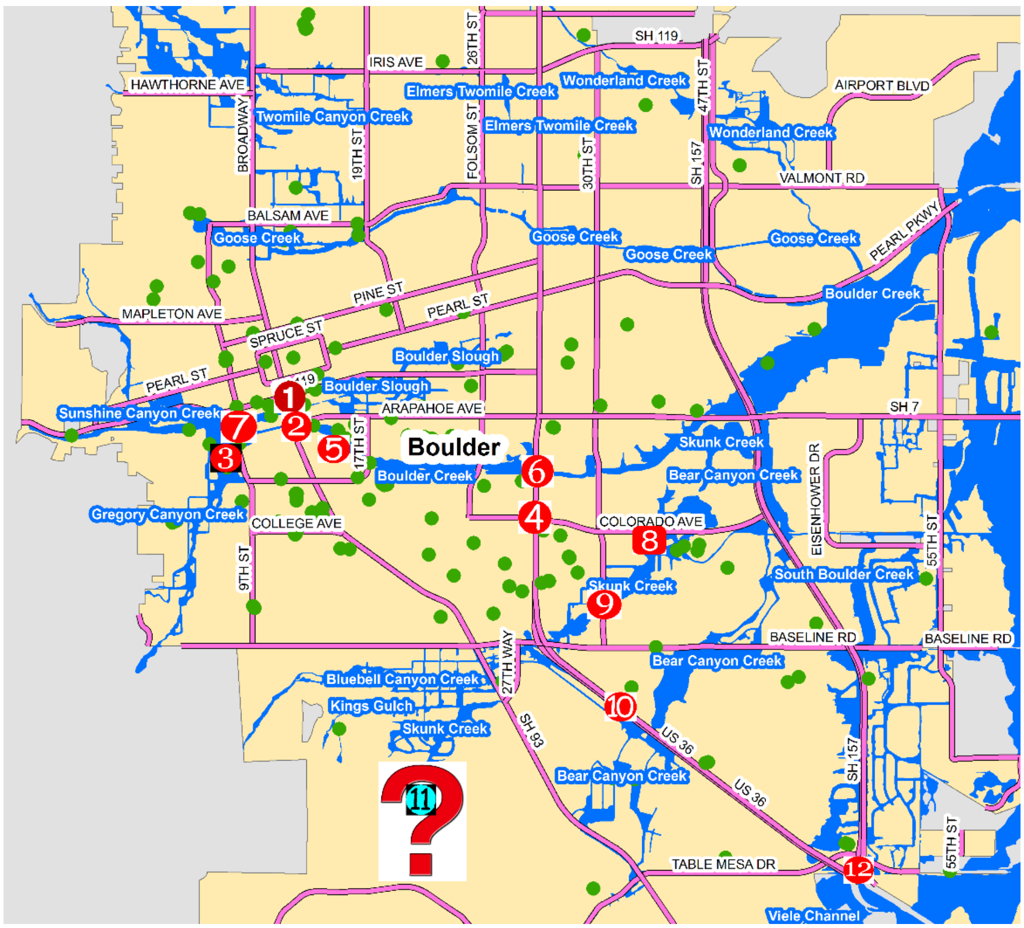

Figure 4.7 Example of identified roads/streets

1. Using keywords “west of *Broadway*”, a related NOAA warning/alert message was found: *Hourly rainfall intensity at the Sugarloaf RAWS station 6 mi. west of Boulder compared with gage height on Boulder Creek at Boulder* (**west of Broadway**). *The first flood peak closely followed the* ***heavy rainfall before***

midnight on 9/11-12*, when 3.5" fell in 6 hours. (Data: rainfall: RAWS via WRCC; and streamflow: Colorado DWR; plotted by Jeff Lukas, WWA).*

The above message specifically mentioned the gauge height on Boulder Creek at west of *Broadway* following the flood peak that resulted from heavy rainfall before midnight on September 11th, which corresponds to the tweet and explains why "Boulder Creek is about to spill its bank at west of *Broadway*" at 3:02 am on September 12th. With this warning message as proof, the first tweet in Table 4.6 could be considered reliable in terms of its location, time, and content.

2. When searching for "*Broadway*" and "*Arapahoe Avenue*", no direct evidence was found, which could be because *Arapahoe Avenue* being a county road is generally not included in official warning or damage assessment reports where a larger scale is used. However, as seen in Figure 4.7, the Boulder Creek flooded the crossing of Broadway and *Arapahoe Avenue* (marked 2), which probably made it possible for the observer to detect increased water level of 2.5 feet within 10 minutes. Additionally, the tweet was posted at 5:30 am, which falls exactly within the period when Boulder Creek was officially identified to have experienced a rapid rise of water level (see Figure 3.2).
3. The crossing of *8th Street* and *Marine Street* (number 3) was impacted by flooding of Gregory Canyon Creek, which corresponds to the tweet content that the drainage of Gregory Canyon overflooded 8th street. Based on time, the previous tweet identified a rapid rising of water level on Boulder Creek at 5:30am, and then this tweet 20 minutes later reported inundation of roads due to flooding of

Gregory Canyon Creek which is adjacent to Boulder Creek. This shows that the risk information in the third tweet is reliable based on content and time.

4. The intersection of *28th Street* and Colorado Avenue (number 4 in Figure 4.7) is in the flooded area between Boulder Creek and Skunk Creek, and the tweet was posted at the height of flooding when the creeks rose rapidly above flood stage simultaneously. Due to continuous rainfall that flooded most tributaries, rushing water inundated most roads in Boulder City. An estimation of road damage was found in an official damage assessment report by NOAA (2014): *Authorities estimate the flooding damaged or destroyed almost 485 miles of roads and 50 bridges in the impacted counties.* The content of this tweet indicates flooded roads with "knee deep water" and it was posted right after continuous heavy rainfall. Therefore, it can be considered a reliable tweet.
5. The fifth tweet was posted in a similar context as the fourth tweet, and the user appears to have witnessed the neighborhood streets were all flooded. Considering this tweet was reliable, *15th street* could be marked inundated so that these roads could be avoided for evacuation.
6. State *Highway 36* was mentioned several times in tweets, with the earliest mention being on September 12th when excessive rainfall continued to intensify the flooding situation. The 6th, 10th, 11th and 12th tweets also referred to the condition of *Highway 36*, such as nearby raining and pouring, flooded situation with over 3 feet of water, and its subsequent closure. Evidence of this situation was also found in an official damage assessment report (NOAA, 2014): *Based on FEMA information, the flooding destroyed more than 350 homes with over 19,000*

homes and commercial buildings damaged, many of which were impossible to reach except on foot. Flooding resulted in a total of 485 miles of damaged roadway, destroyed 30 state highway bridges, and severely damaged another 20 bridges. During the height of the flooding, authorities were forced to close ***36 state highways****. Some highways could not be repaired for weeks or even months.*

Therefore, these tweets were considered to be reliable. The sixth, tenth, and twelfth tweets were geo-located along *Highway 36*, but the eleventh tweet was posted beyond the city limits of Boulder. Being posted in a place that is far from the site, it is hard to prove the reliability without referring to other tweets that also mentioned *Highway 36*. In this case, the eleventh tweet is reliable based on content though the posted location did not correspond to the impact location. From this point of view, keywords that were verified to be related to important incidents, such as *Highway 36*, could be used to extract tweets that are beyond the spatial limit of the study area or even do not possess any geo-location information. This approach would yield a large volume of relevant tweets.

7. The seventh tweet posted that a portion of *eighth Street* between University of Colorado Boulder (CU Boulder) and Marine Street (number 7 in Figure 4.7) experienced severe rainfall at 3:22 am on 9/13. Geographically speaking, the site is present near the juncture of Boulder Creek, Sunshine Canyon Creek, and Gregory Canyon Creek. Thus, the street was highly likely to be flooded at that time. A piece of news by Huffing Post, "*around 80 buildings on campus were damaged in some form, CU Boulder police tweeted, and raw sewage was flowing from a pipe in one area.*" (Kingkade, 2013) confirmed this tweet. A campus

damage assessment report (Department of Higher Education, 2013) also reported that "*80 of 300 structures on the Boulder campus sustained some damage. The damage is described as "widespread" but not severe*." These two news articles confirmed the reliability of the tweet.

8. The eighth and ninth tweets were geo-located along the flooded Skunk Creek (Figure 4.7). While 30th Street was still getting flooded, the adjacent *Colorado Avenue* was already closed. Both streets are located in the Foothills area, which was reported to have been seriously impacted by flood in a damage assessment report summary: "*Foothills around Boulder also saw severe flooding and debris flows*" (Western Water Assessment, 2014).

4.2.2 Evaluation of Image.

Risk information conveyed by images available from tweets is more emotionally appealing and influencing when used to motivate public response. Often, it is through images that people develop a deep impression of how destructive natural hazards could be (Vis et al., 2013). However, according to a study on tweet content categorization, around 4% of tweets are spams at all times (Kelly, 2009), which doubtlessly include images. Research has also shown that fake images tend to be propagated via web during crises (Gupta et al., 2013). Despite abundant research on filtering out spam or phishing tweets (Benevenuto et al., 2010; Song et al., 2011; Wang, 2010a, 2010b), studies focusing on diffusion of fake images are sparse (Gupta et al., 2013). Techniques used to eliminate spam tweets include URL shortening services, domain and popular blacklists detection, and machine learning. However, none of the above-mentioned approaches were implemented in this study; rather, a manual content analysis of images was conducted. The images were considered reliable if they: (i) correspond to facts mentioned in tweets, (ii) mutually prove each other, or (iii) gain support from other sources.

From the relevant tweets, 42 images were randomly selected of which 33 images reflect the facts/incidents that have been validated in the previous section. A common characteristic shared by the images was that they all corresponded to a specific location. Figure 4.8, 4.9, 4.10, and 4.11 display the images extracted from tweets.

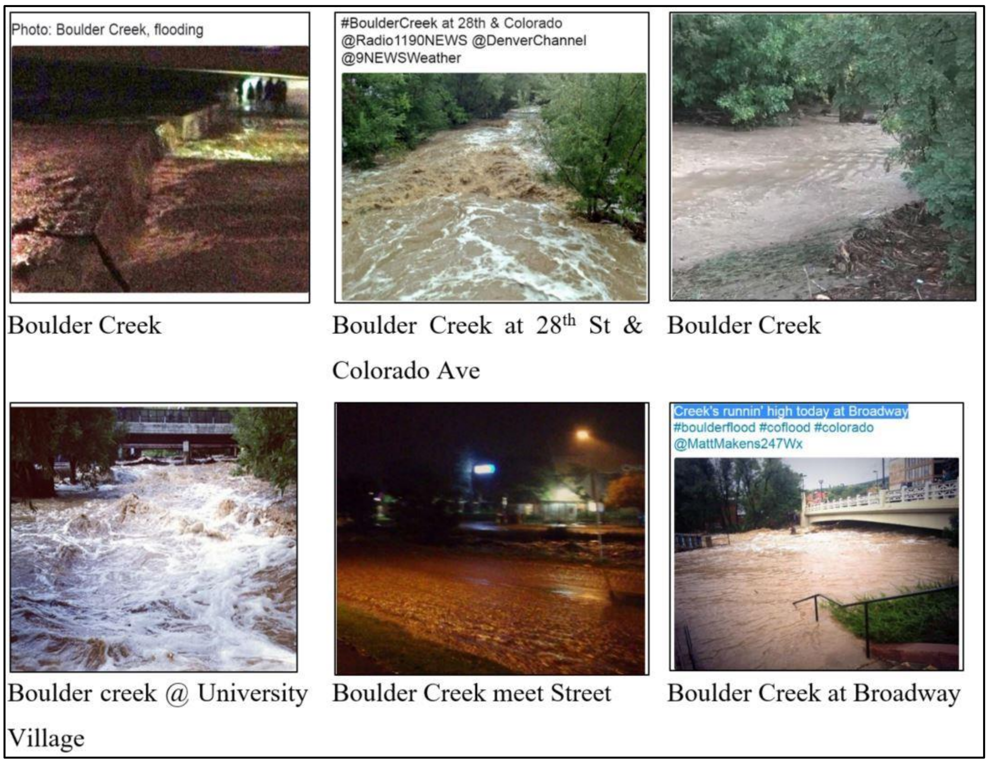

Figure 4.8 Images of Boulder Creek

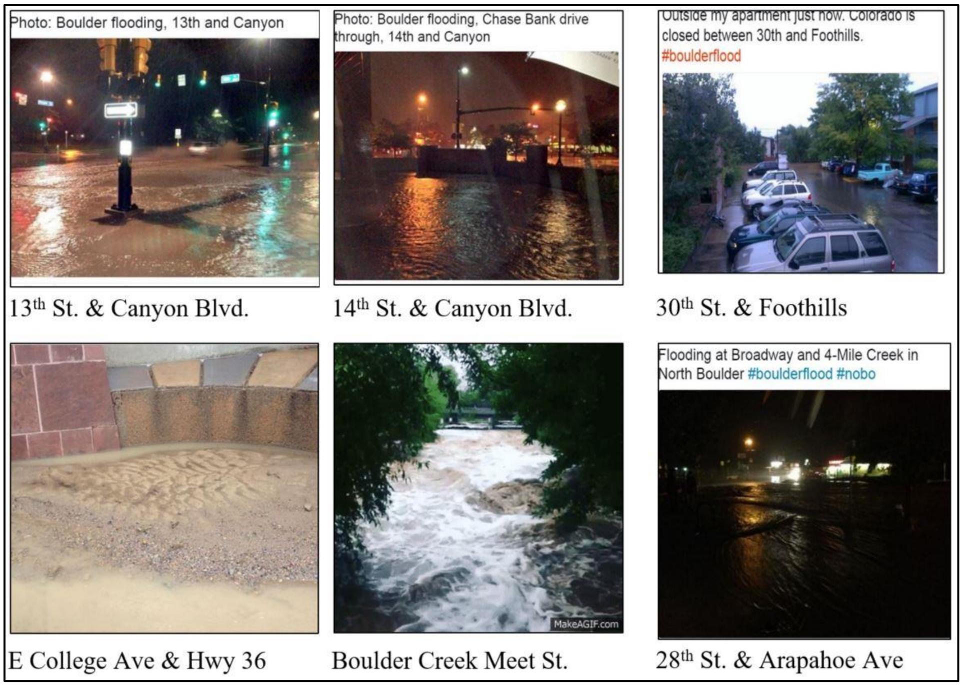

Figure 4.9 Images of flooded streets

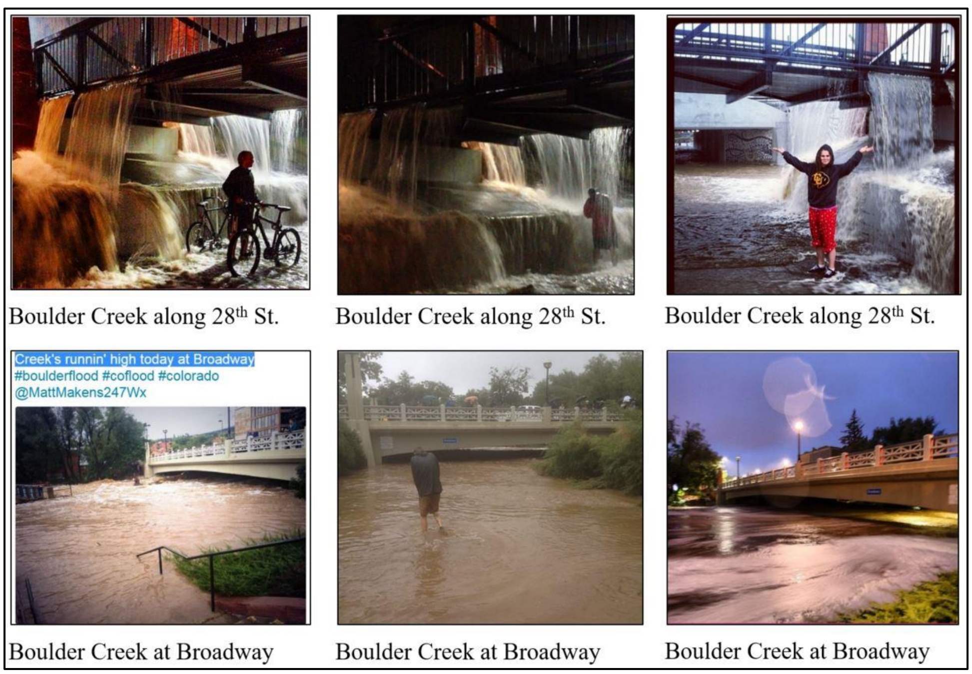

Figure 4.10 Images mutually prove each other

The images shown in Figure 4.10 were taken at the same location by different people, at different time, and from different angles. The flood water falling from the bridge created the "beautiful" waterfall and, thus, attracted people to take pictures to document the severity and rarity of the flood. The bottom three images in Figure 10 recorded the increased water level at Boulder Creek under Broadway Bridge, which clearly displays the temporal change of flood severity. This finding is of critical importance for crowdsourcing-based risk communication as massive images could mutually verify each other despite lack of external information.

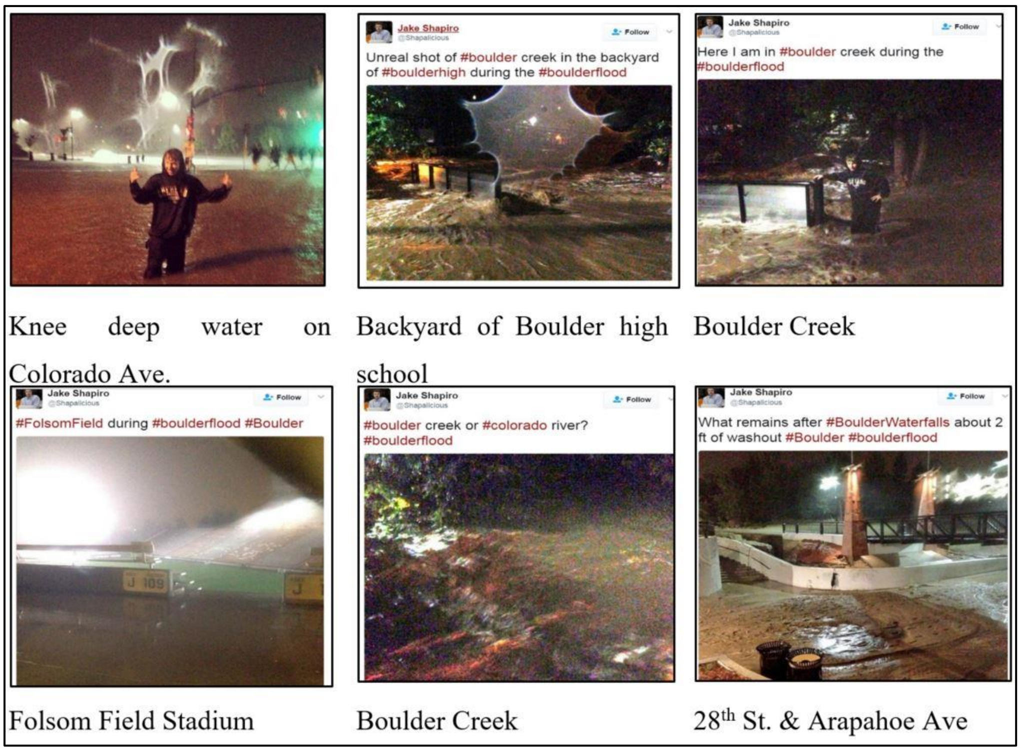

Figure 4.11 Images took by a local news reporter

Images in Figure 4.11 were taken by a local news reporter, Mr. Jake Shapiro, who posted tweets about flood situations in several locations along with pictures. The locations that were mentioned by the reporter were: *Colorado Avenue*, the backyard of *Boulder High School*, *Folsom Field Stadium*, and 28th *Street* & *Arapahoe Ave*. The text and images posted by the reporter could be regarded as reliable.

CHAPTER V DISCUSION AND CONCLUSION

Despite its popularity for providing up-to-date information pertaining to disasters, and although several studies have evaluated different aspects of data quality of crowdsourced data, little research has been conducted to examine relevance and reliability of risk information available from crowdsourced data. This research focused on analyzing geo-tagged tweets to ensure that the contents of tweets were generated by those who experienced or witnessed the 2013 Colorado flood rather than by "*outsiders*" who were not on the scene. This chapter discusses the findings of various methods used to assess relevance and reliability, and their significance to risk communication. The chapter also identifies limitations of the study and future research directions, and finally, summarizes the contributions of this study.

5.1 Relevance of Tweets to Risk Communication.

Five distinct approaches - temporal, spatial, spatiotemporal, content analysis, and cosine similarity comparison were used to find relevant tweets that conveyed risk information to help public undertake preparatory actions to mitigate flood impact. Each approach helped extract relevant tweets based on their spatial and temporal distribution in relation to intensity and severity of flood, flood impact areas, and flood induced damages to road networks. Despite their effectiveness in extracting relevant tweets, a combination of these approach should be implemented to ensure extraction of all relevant tweets. Nevertheless, the implementation of any of these approaches necessitated extraction of geo-tagged tweets that were concentrated in Boulder (the study site) to eliminate introduction of rumors and misinformation. Although the spatial and temporal approaches helped extract relevant tweets that were spatially clustered on days with the heaviest precipitation in areas with high population density and high degree of damage, tweet volume was found to be dependent largely on population density and damage extent. Therefore, spatial distribution of tweet volume could not be used as an indicator of severely damaged areas that need to be targeted for emergency management efforts as pointed out by Kryvasheyeu et. al. (2016).

The content analysis of tweets revealed that what people expect to be included in an alert message is different from what is conveyed by individuals in social media. While the survey responses indicated public's need to gain information about *nature of the disaster*, *impact zone*, *time frame*, *recommended actions*, *when to take action*, *evacuation routes*, *shelter location*, and *who to contact for hel*p in order of rank via alert messages, majority of the tweets tend to provide information about *nature of the disaster* and *impact zone*. Although information about these two components of a disaster is crucial for individuals to take appropriate actions to reduce disaster impacts, from a risk communication perspective, it may not be prudent to use tweets to disseminate information about other components identified by the survey response.

The cosine similarity comparison approach compared selected terms in tweets with official warning/assessment reports. The approach resulted in a relevance score of 0.7, which indicates that the two documents are relevant. However, just because the tweets are relevant to the reports does not mean that the tweets provide an in-depth information about the event or its impacts like the reports. So, the relevant tweets must be used with caution, and be complemented with other ancillary data sets and reports. Different from cosine similarity comparison approach, the relevance score of 5202 geo-tagged tweets in Colorado was derived using the MongoDB built-in function ($meta).

Top frequent words and hashtags are the terms based on which the scores were generated, which is different from the reference data in cosine similarity comparison.

5.2 Reliability of Tweets to Risk Communication.

To assess reliability, two main approaches were used. In the first approach, experts with knowledge of the area and flood event verified and validated the risk information obtained from tweets by comparing them with authoritative data, specifically, reports obtained from NOAA and FEMA. In this approach, the focus was roads and street networks that were damaged or flooded by the event on different dates and times. In the second approach, images obtained from tweets were compared with images and text-based contents derived from official reports and news articles. In some of the incidents, multiple images were posted about the damage, which enabled mutual validation of the information, and identification of reliable tweets. In other cases, the content analysis of images either corresponded or supplemented the text-based content derived from reports. The 720 (14%) tweets that were used in reliability assessment were relevant ones with a relevance score higher than 1.3 (see Table 4.7). It is apparent that relevant tweets tend to be reliable, but only 3% of the relevant tweets contained names of flooded or damaged roads/streets/rivers/creeks, and thus are validated to be reliable.

5.3 Research Outcomes.

5.3.1 Implications for Risk Communication.

An important outcome of this research is an integrated methodological framework to extract and evaluate relevant and reliable risk information that could facilitate risk communication, increase situational awareness, and public response to natural hazards. Different from traditional risk communication approach, which is a top-down and centralized model, crowdsourcing based risk communication offers bottom-up, centralized, and collaborative mechanism that enable more communication possibility among EMAs and the public. Based on literature review of existing research (Alabri and Hunter, 2010; Forghani and Delavar, 2014; Meek et al., 2014), this research reassured the important role of crowdsourcing-based risk communication and revisited the issues with crowdsourced data quality. Information overloading, lack of metadata, and uneven data quality are some of the key issues with crowdsourced data. Therefore, relevance and reliability evaluations is vital to alleviate or eliminate the above-mentioned problems.

The extracted risk information could help emergency organizations and responders to coordinate relief efforts and mitigate hazard impacts to lives and properties. Given the time-consuming and expensive nature of the implemented approaches, the findings may not be referable and the methodology may not be replicated in an emergency setting. However, automating the methodology could provide critical risk information in a timely manner to be useful for EMA activities.

The public could benefit from the extracted risk information by increasing their situational awareness, take preparatory actions, and minimize hazard impacts. As social media has become popular during risk communication (Ding and Zhang, 2010; Veil et al., 2011; Wendling et al., 2013), it is vital for public to realize that social media sites may contain useful or actionable information and the specific types of risk information. The methodology implemented in this study could help the public avoid blind usage of social media, such that the public could wisely choose the social media sites that suit them and make the most of social media risk communication. Nevertheless, it is also important to understand the proportion of useful or actionable information on social media sites, and learn how to distinguish relevant and reliable information from massive social data. This research demonstrated ways to choose reference data according to corresponding situations and the possible ways to determine relevant and reliable crowdsourced data. The research also demonstrated the extent to which crowdsourced data could be used for risk communication and to increase situation awareness of public following a disaster.

5.3.2 Implications for GIScience.

Despite being interdisciplinary, this study is positioned in the field of GIScience and the broader geography. Use of various geospatial datasets and the spatiotemporal approaches/thinking implemented in this research and the use of Volunteer Geographic Information derived from crowdsourced data makes the study a geographical study. Combing distinct subfields of geography, including hazard geography (research subject) and computational geography (research methodology), this research enriches the content of geographical research and promotes the intersection between different sub-disciplines.

The role of GIScience in emergency response is not new. Geospatial concepts and technologies have been applied throughout each stage of emergency response, such as preparedness and rescue (Cutter, 2012). However, this research demonstrated an integrated use of spatiotemporal and temporal approaches to assess data quality of crowdsourced data using intrinsic approaches, which is an area of ongoing research and of concern among crowdsourced data users (Alonso and Mizzaro, 2012). This research also enabled combining geospatial data (precipitation, flood extent, degree of flood damage, etc.) and non-geospatial data (survey data, official warnings/alerts, and official assessment reports) for data quality assessment and risk communication.

The geographical methods and techniques implemented in this study could be used to assess relevance and reliability of data that could be used during emergency management. Despite the time-consuming nature of the application, the findings are valuable to complement other emergencies management resources. Therefore, this research is of significance to GIScience discipline, specifically, data quality of data sets as well as to risk communication that relies on timely and updated data to save lives. The research also strengthens the need to integrate GIScience and crowdsourcing to increase situational awareness and enable public response to natural hazards in web 2.0.

5.4 Limitations and Future Research.

Although the geo-location approach implemented to extract tweets for this study probably eliminated misinformation or rumors (Lee et al., 2011), it may not be effective in case of other hazards occurring in other countries. For instance, an analysis of 2015 Nepal earthquake resulted in less than 0.44% of geo-tagged tweets, which could be because Nepal has a very limited Twitter population compared to developed countries, i.e. U.S. (Kulshrestha et al., 2012). Out of one million tweets, only 0.44% were geo-tagged to the study site and were used in this study. The low percent of geo-tagged tweets could decrease in case of extreme weather events that cause power failure and communication disruption as well as due to rising concern about privacy that prohibits citizens from sharing their personal information including location.

There are other issues that could hinder the use of tweets by EMPs during real-time emergency management. First, the data used in this research was purchased from Gnip Inc. (used to be a subsidiary company of Twitter Inc.) for $1,250.00. Purchasing tweets is not a wise way to spend money during emergencies when disaster response and recovery efforts call for a great portion of disaster relief fund. Furthermore, purchasing tweets ensures post-processing rather than real-time analytics for emergency response purpose, which is the need of EMPs to reduce impact. Although real-time tweets could be obtained using Twitter API for free, this automatic download tends to return only 1% of the total tweet volume that match the search terms (Twitter Inc., 2017). It is probable that the low percentage of real-time tweets obtained using API would eliminate a significant number of tweets that might be relevant. Second, data collection, data cleaning, and implementation of the methodologies are time consuming, computationally intensive, and require skilled professionals. Therefore, using crowdsourced data for emergency management activities would be inefficient for local EMPs unless automated tools and algorithms are developed to enable tweet use in real-time. Third, given that the EMAs' main responsibilities during an emergency is to coordinate response and recovery efforts, they may not be interested in knowing what people expect to be communicated in an alert message. However, knowing the kind of risk information the public expects will help EMAs send out the specific information via social media and also use social media to increase situation awareness in a timely manner.

The methodology used in this research is dependent on the hazard event and available reference data sets. For instance, precipitation data was collected to understand flood impacts and extent, and extract tweets pertaining to this event and location. This methodology demonstrated a way of integrating available data and approaches to evaluate relevance and reliability of tweets in disseminating risk information, which is gaining interest and focus in emergency management. However, this methodology has a number of limitations. First, despite having local knowledge, manual identification of reliable tweets is not only time consuming, it could introduce potential errors. Second, the total number of tweets that could be manually interpreted is limiting in comparison to the potentially reliable tweets. Third, biases could be introduced by human readers due to varying cognition and judgement toward the same issue. Thus, use of manual check of reliability prevented from real-time reliability evaluation.

Despite the limitations associated with human annotation, it is commonly used as the main or supplementary method for evaluating reliability (Castillo et al., 2011; Grady and Lease, 2010; Gupta and Kumaraguru, 2012). In a study on automating assessment of the reliability of Twitter messages (Castillo et al., 2011), people were paid to determine the authenticity ("true" or "not true") of each pre-identified emerging topic on Twitter. Another study assessing crowdsourced data quality employed human experts and authoritative data, and found that involvement of experts played a critical role in the control of data quality (See et al., 2013). Human annotators were asked to rate the credibility (which is used interchangeably with reliability in the context) of information based on given choices, including *Definitely Credible*, *Seems Credible*, *Definitely Not credible*, and *I Can't Decide* (Gupta and Kumaraguru, 2012). Future research should therefore use deep learning to develop a metabase of keywords for reliability assessment that would eliminate human involvement.

The steps implemented in this research were progressive, i.e., each step is based on the implementation of a previous step and could not be reversed. The joint use of geo-tagged tweets and bag-of-words extracted a significant number of tweets with high relevance and reliability. However, it was assumed that a large number of local public generated the tweets, which in reality may not be the case as seen from Nepal

Earthquake. If that happens, then the tweet volume would be small enough to be useful for emergency management activities. Furthermore, the methodology developed in this research may not be useful with a small percent of tweets. Therefore, it might be efficient and effective to use a citizen science based portal that would allow impacted population to share information specific to a hazard event (Kar, 2015) or use the new social media site “next door” that eliminates participation of “*outsiders*” beyond certain zip codes.

Out of 100% of tweets, 14% were relevant and 3% reliable. Given these results, despite the rich content of tweets, the time and money spent on obtaining tweets and other data sets, and implementing the methodology is not justifiable from EMA perspective. Therefore, future research should focus on developing a matrix to assess data quality of tweets, automating implementation of techniques, and implementing machine learning approaches to assess reliability.

APPENDIX A - Code

A.1 MongoDB Code

Extract English tweets:

```
db.Twitter.aggregate({$match:{ twitter_lang:"en"}},{$out:”en_tweets”})
```

→1017024

Extract tweets with coordinates and save into geo collection:

```
db.en_tweets.find().forEach(function(doc){
if(doc.geo)
db.geo.save(doc);})
```

→16551

Extract tweets posted or shared from users in Colorado among geo-tagged tweets:

```
db.en_twitter.aggregate({$match:
{$or:[{"actor.location.displayName":/colorado/},
{"actor.location.displayName":'co'},
{"actor.location.displayName":/boulder/},
{"actor.location.displayName":/front range/},
{"actor.location.displayName":/el paso/},
{"actor.location.displayName":/denver metro/},
{"actor.location.displayName":/boulder,co/},
{"actor.location.displayName":/denver,co/},
{"actor.location.displayName":/boulder,colorado/},
{"actor.location.displayName":/denver,colorado/},
{"actor.location.displayName":/colorado,us/},
{"actor.location.displayName":/denver metro/},
```

```
{"actor.location.displayName":/denver/},
{"actor.location.displayName":/deadman hill/},
{"actor.location.displayName":/joe wright/},
{"actor.location.displayName":/fort collins/},
{"actor.location.displayName":/sugarloaf/},
{"actor.location.displayName":/fort carson/},
{"actor.location.displayName":/adams county/},
{"actor.location.displayName":/arapahoe county/},
{"actor.location.displayName":/broomfield/},
{"actor.location.displayName":/fremont county/},
{"actor.location.displayName":/jefferson county/},
{"actor.location.displayName":/fremont county/},
{"actor.location.displayName":/larimer/},
{"actor.location.displayName":/logan county/},
{"actor.location.displayName":/morgan county/},
{"actor.location.displayName":/pueblo county/},
{"actor.location.displayName":/weld county/},
{"actor.location.displayName":/clear creek/}] }
{$or:[{"actor.location.displayName":{$ne:/texas/}},{"actor.location.displayName
":{$ne:/utah/}},{},]},{$out:"en_geo_co"})
```

A.2 R Code

```
cname =
setwd("C:/Users/XiaohuiLiu/Dropbox/TwitterData/CompareStudy/CO_R")
```

```
dir(cname)

docs <- Corpus(DirSource(cname))

library(tm)

docs <- Corpus(DirSource(cname))

summary(docs)

docs <- tm_map(docs, removePunctuation)

for(j in seq(docs))

{

docs[[j]] <- gsub("/", " ", docs[[j]])

docs[[j]] <- gsub("@", " ", docs[[j]])

docs[[j]] <- gsub("\\|", " ", docs[[j]])

}

docs <- tm_map(docs, removeNumbers)

docs <- tm_map(docs, tolower)

docs <- tm_map(docs, removeWords, stopwords("english"))

library(SnowballC)

docs <- tm_map(docs, stemDocument)

docs <- tm_map(docs, stripWhitespace)

docs <- tm_map(docs, PlainTextDocument)
```

APPENDIX B – Top Frequent Words & Hashtags

1. Top 50 frequent words list from tweets.

	Words	Frequency
1	colorado	2798
2	boulder	1865
3	flood	851
4	mdt	848
5	september	815
6	flooding	681
7	flash	478
8	denver	459
9	warning	448
10	issued	438
11	nws	429
12	springs	356
13	rain	298
14	boulderflood	248
15	creek	210
16	state	156
17	emergency	151
18	people	150
19	like	144
20	county	141
21	park	136
22	floods	135
23	today	135
24	weather	135
25	day	132
26	water	122
27	coflood	119
28	colo	112
29	new	112
30	will	109
31	good	108
32	time	101
33	closed	93
34	one	93
35	love	92
36	safe	91
37	home	88
38	stay	85
39	center	84

40	news	84
41	know	83
42	come	79
43	great	77
44	road	76
45	city	75
46	night	75
47	see	75
48	front	74
49	school	74
50	help	73

2. Top 10 hashtags from tweets.

	Hashtag
1	Colorado
2	boulderflood
3	Coflood
4	cowx
5	NeverForget
6	flooding
7	GodBlessAmerica
8	news
9	CORecall
10	Denver

3. Top 50 frequent words from NOAA warning/alert messages.

	Word	**Frequency**
1	flood	5198
2	flooding	2706
3	rain	2442
4	boulder	2328
5	feet	2238
6	heavy	2136
7	sep	1840
8	inch	1598
9	near	1550
10	counti	1432
11	weather	1320
12	denver	1284
13	county	1276

14	service	1188
15	national	1180
16	south	1036
17	coc	1008
18	colorado	1000
19	water	902
20	kbou	836
21	larimer	804
22	counties	802
23	act	776
24	weld	764
25	rainfall	760
26	area	750
27	northeast	738
28	wgus	692
29	latlon	676
30	precautionarypreparedness	610
31	jefferson	604
32	stream	594
33	central	586
34	continue	586
35	areas	566
36	park	532
37	locations	490
38	afternoon	486
39	north	486
40	may	484
41	roads	446
42	stat	414
43	adams	390
44	arapahoe	386
45	flows	386
46	across	382
47	minor	370
48	west	370
49	douglas	358
50	number	350

4. Top 50 frequent words from official damage assessment reports.

	Word	**Frequency**
1	boulder	722
2	flood	589

3	september	530
4	river	513
5	wfo	495
6	colorado	470
7	flooding	442
8	nws	435
9	forecast	417
10	flash	393
11	rainfall	391
12	event	384
13	â€“	371
14	new	345
15	creek	337
16	weather	330
17	south	275
18	hpwwacoloradoedu	249
19	platte	242
20	hydrologic	235
21	data	232
22	front	220
23	â€â€	213
24	heavy	208
25	central	207
26	precipitation	207
27	figure	200
28	partners	198
29	service	198
30	range	196
31	wfos	186
32	time	182
33	area	172
34	system	169
35	pueblo	168
36	recommendation	162
37	local	158
38	training	153
39	county	151
40	finding	150
41	basin	148
42	events	148
43	denver	144
44	services	141
45	forecasts	135
46	information	132

47	rain	130
48	â€•	129
49	noaa	127
50	awips	126

APPENDIX C – Examples of Identified Road/Streets

	Roads/streets	**Posted**	**Associated Risk Information**
1	West of Broadway	09/12 03:02	Boulder Creek is about to spill its bank.
2	Broadway & Arapahoe Avenue	09/12 05:30	Water at Boulder Creek has come up 2.5 feet in 10 mins.
3	8th Street & Marine Street	09/12 05:52	Gregory canyon drainage overtopping the underground culvert, flowing onto 8th St. near Marine.
4	28th Street & Colorado Avenue	09/12 06:09	Knee deep water at 28th St & Colorado Ave.
5	15th Street	09/12 08:39	River taking back Boulder
6	Highway 36 underpass	09/12 22:23	It's raining! It's pouring!
7	8th Street between University of Colorado and Marine	09/13 03:22	…basically, a raging torrent.
8	30th Street & Foothills	09/13 00:49	Colorado Avenue is closed between 30th and Foothill.
9	30th Street	09/13 01:08	Water is coming up through drains on 30th and Colorado Ave…this could get ugly.
10	Highway 36	09/13 01:30	Barely make it out of Boulder. Couldn't get to hwy 36.
11	Highway 36	09/13 02:33	Highway 36 is flooded, not way out.
12	Highway 36 & Foothills	09/13 05:32	Over 3 feet of water flooding.

REFERENCES

Agarwal, P., 2005. Ontological considerations in GIScience. Int. J. Geogr. Inf. Sci. 19, 501–536. doi:10.1080/13658810500032321

Alabri, A., Hunter, J., 2010. Enhancing the quality and trust of citizen science data, in: E-Science (E-Science), 2010 IEEE Sixth International Conference on. IEEE, pp. 81–88.

Alonso, O., Mizzaro, S., 2012. Using crowdsourcing for TREC relevance assessment. Inf. Process. Manag. 48, 1053–1066. doi:10.1016/j.ipm.2012.01.004

Arsanjani, J.J., Barron, C., Bakillah, M., Helbich, M., 2013. Assessing the quality of OpenStreetMap contributors together with their contributions, in: Proceedings of the AGILE. pp. 14–17.

Atefeh, F., Khreich, W., 2015. A Survey of Techniques for Event Detection in Twitter. Comput. Intell. 31, 132–164. doi:10.1111/coin.12017

Bahir, E., Peled, A., 2015. Real-Time Major Events Monitoring and Alert System through Social Networks. J. Contingencies Crisis Manag. 23, 210–220.

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., Lin, A.Y.-M., Huyck, C., 2012. Crowdsourcing earthquake damage assessment using remote sensing imagery. Ann. Geophys. 54. doi:10.4401/ag-5324

Beckman, P., Nadella, S., Trebon, N., Beschastnikh, I., 2007. SPRUCE: A system for supporting urgent high-performance computing, in: Grid-Based Problem Solving Environments. Springer, pp. 295–311.

Benevenuto, F., Magno, G., Rodrigues, T., Almeida, V., 2010. Detecting spammers on twitter, in: Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS). p. 12.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., Schreeb, J. von, 2011. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLOS Med. 8, e1001083. doi:10.1371/journal.pmed.1001083

Brabham, D.C., 2013. Using crowdsourcing in government. IBM Cent. Bus. Gov. 1–42.

Brabham, D.C., 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence 14, 75–90. doi:10.1177/1354856507084420

Brown, G., Weber, D., de Bie, K., 2015. Is PPGIS good enough? An empirical evaluation of the quality of PPGIS crowd-sourced spatial data for conservation planning. Land Use Policy 43, 228–238.

Cai, L., Zhu, Y., 2015. The challenges of data quality and data quality assessment in the big data era. Data Sci. J. 14.

Callison-Burch, C., Ungar, L.H., Pavlick, E., 2015. Crowdsourcing for NLP., in: HLT-NAACL. pp. 2–3.

Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H.-W., Mitra, P., Wu, D., Tapia, A.H., Giles, L., Jansen, B.J., others, 2011. Classifying text messages for the haiti earthquake, in: Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management (ISCRAM2011). Citeseer.

Castillo, C., Mendoza, M., Poblete, B., 2011. Information credibility on twitter, in: Proceedings of the 20th International Conference on World Wide Web. ACM, pp. 675–684.

CCC, 2013. Colorado Climate Center.

CDG, (Center for Digital Government), 2015. Social Media: Emergency Communications' Best Ally. [WWW Document]. URL http://825d0007e19cfb8330f5-793aa0e2839afbbc4a0b9a46376ed589.r13.cf1.rackcdn.com/CDG14_BRIEF_ArchiveSocial_V.pdf (accessed 11.14.16).

Center for Digital Government, 2015. Social Media: Emergency Communications' Best Ally 2015. [WWW Document]. URL http:// archivesocial.com/social-media-resources-for-government

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N., 2016. Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. Int. J. Remote Sens. 37, 100–124.

Chan, C.J., 2014. Crowdsourcing Disaster Response. ResearchGate.

Chatfield, A.T., Brajawidagda, U., 2014. Crowdsourcing hazardous weather reports from citizens via twittersphere under the short warning lead times of EF5 intensity tornado conditions, in: System Sciences (HICSS), 2014 47th Hawaii International Conference on. IEEE, pp. 2231–2241.

Cheong, F., Cheong, C., 2011. Social media data mining: A social network analysis of tweets during the 2010-2011 Australian floods.

City of Boulder, 2014a. The 2013 Boulder Flood Extent Map.

City of Boulder, 2014b. Main Roads and Streets map.

Cole, T.W., Fellows, K.L., 2008. Risk communication failure: A case study of New Orleans and Hurricane Katrina. South. Commun. J. 73, 211–228.

Colley, K.L., Collier, A., 2009. An overlooked social media tool? Making a case for wikis. Public Relat. Strateg. 34–35.

Colorado City Rank [WWW Document], 2016. URL http://worldpopulationreview.com/states/colorado-population/cities/

Comber, A., See, L., Fritz, S., Van der Velde, M., Perger, C., Foody, G., 2013. Using control data to determine the reliability of volunteered geographic information about land cover. Int. J. Appl. Earth Obs. Geoinformation 23, 37–48.

Covello, V., Sandman, P.M., 2001. Risk communication: evolution and revolution. Solut. Environ. Peril 164–178.

Covello, V.T., 1992. Risk communication: An emerging area of health communication research. Ann. Int. Commun. Assoc. 15, 359–373.

Covello, V.T., McCallum, D.B., Pavlova, M.T., 2012. Effective Risk Communication: The Role and Responsibility of Government and Nongovernment Organizations. Springer Science & Business Media.

Cutter, S.L., 2012. Hazards Vulnerability and Environmental Justice. Routledge.

Degrossi, L.C., de Albuquerque, J.P., Fava, M.C., Mendiondo, E.M., 2014. Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil., in: SEKE. pp. 570–575.

Department of Higher Education, 2013. 2013 Colorado flood: Campus impacts.

Dictionary.com, 2017. Reliability | Define Reliability at Dictionary.com [WWW Document]. URL http://www.dictionary.com/browse/reliability (accessed 2.21.17).

Ding, H., Zhang, J., 2010. Social media and participatory risk communication during the H1N1 flu epidemic: A comparative study of the United States and China. China Media Res. 6, 80–91.

Duan, Y., Jiang, L., Qin, T., Zhou, M., Shum, H.-Y., 2010. An empirical study on learning to rank of tweets, in: Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, pp. 295–303.

Duram, L., Brown, K., 1999. Insights and Applications Assessing Public Participation in U.S. Watershed Planning Initiatives. Soc. Nat. Resour. 12, 455–467. doi:10.1080/089419299279533

Earle, P.S., Bowden, D.C., Guy, M., 2012. Twitter earthquake detection: earthquake monitoring in a social world. Ann. Geophys. 54.

Eckle, M., de Albuquerque, J.P., 2015. Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. Retrieved Sept. 24, 5–1.

Erskine, M., Gregg, D., 2012. Utilizing Volunteered Geographic Information to Develop a Real-Time Disaster Mapping Tool: A Prototype and Research Framework (PDF Download Available), in: ResearchGate.

Estellés-Arolas, E., González-Ladrón-De-Guevara, F., 2012. Towards an integrated crowdsourcing definition. J. Inf. Sci. 38, 189–200.

Fan, H., Zipf, A., Fu, Q., Neis, P., 2014. Quality assessment for building footprints data on OpenStreetMap. Int. J. Geogr. Inf. Sci. 28, 700–719.

FEMA, 2014. 2013 Colorado Floods Situational Awareness.

Filliat, D., 2007. A visual bag of words method for interactive qualitative localization and mapping, in: Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE, pp. 3921–3926.

Forghani, M., Delavar, M.R., 2014. A quality study of the OpenStreetMap dataset for Tehran. ISPRS Int. J. Geo-Inf. 3, 750–763.

Freedman, 2013. Flood-Ravaged Boulder, Colo., Sets Annual Rainfall Recor [WWW Document]. Clim. Cent.

Frommberger, L., Schmid, F., 2013. Mobile4D: crowdsourced disaster alerting and reporting, in: Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2. ACM, pp. 29–32.

Gao, H., Barbier, G., Goolsby, R., 2011a. Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell. Syst. 26, 10–14.

Gao, H., Wang, X., Barbier, G., Liu, H., 2011b. Promoting coordination for disaster relief–from crowdsourcing to coordination, in: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. Springer, pp. 197–204.

Geiger, D., Rosemann, M., Fielt, E., Schader, M., 2012. Crowdsourcing information systems-definition, typology, and design.

Gladwin, H., Lazo, J.K., Morrow, B.H., Peacock, W.G., Willoughby, H.E., 2007. Social science research needs for the hurricane forecast and warning system. Nat. Hazards Rev. 8, 87–95.

Goodchild, M.F., 2010. Twenty years of progress: GIScience in 2010. J. Spat. Inf. Sci. 2010, 3–20. doi:10.5311/JOSIS.2010.1.32

Goodchild, M.F., Li, L., 2012. Assuring the quality of volunteered geographic information. Spat. Stat. 1, 110–120.

Google, 2017. Countries editable in Google Map Maker (as of February, 2017) [WWW Document]. URL http://www.google.com/mapmaker/mapfiles/s/launched.html

Gouveia, C., Fonseca, A., Câmara, A., Ferreira, F., 2004. Promoting the use of environmental data collected by concerned citizens through information and communication technologies. J. Environ. Manage. 71, 135–154.

Grady, C., Lease, M., 2010. Crowdsourcing Document Relevance Assessment with Mechanical Turk, in: Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT '10. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 172–179.

Gupta, A., Kumaraguru, P., 2012. Credibility ranking of tweets during high impact events, in: Proceedings of the 1st Workshop on Privacy and Security in Online Social Media. ACM, p. 2.

Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A., 2013. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy, in: Proceedings of the 22nd International Conference on World Wide Web. ACM, pp. 729–736.

Gurabardhi, Z., Gutteling, J.M., Kuttschreuter, M., 2005. An empirical analysis of communication flow, strategy and stakeholders' participation in the risk communication literature 1988–2000. J. Risk Res. 8, 499–511.

Heinzelman, J., Waters, C., 2010. Crowdsourcing crisis information in disaster-affected Haiti. US Institute of Peace Washington, DC.

Heipke, C., 2010. Crowdsourcing geospatial data. ISPRS J. Photogramm. Remote Sens. 65, 550–557.

Hiltz, S.R., González, J.J., Van de Walle, B., 2012. Assessing and improving the trustworthiness of social media for emergency management: A literature review, in: Proceedings of the Norwegian Information Security Conference (NISK), Bodo, Norway. pp. 19–21.

Hirth, M., Ho\s sfeld, T., Tran-Gia, P., 2013. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. Math. Comput. Model. 57, 2918–2932.

Holderness, T., Turpin, E., 2015. From Social Media to GeoSocial Intelligence: Crowdsourcing Civic Co-management for Flood Response in Jakarta, Indonesia, in: Social Media for Government Services. Springer, pp. 115–133.

Horita, F.E.A., Degrossi, L.C., de Assis, L.F.G., Zipf, A., de Albuquerque, J.P., 2013. The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: a systematic literature review.

Howe, J., 2006. Crowdsourcing: Tracking the Rise of the Amateur [WWW Document]. Crowdsourcing. URL http://www.crowdsourcing.com/cs/ (accessed 5.10.17).

Hudson-Smith, A., 2014. NCRM podcast: The potential of crowdsourcing for research and funding in academia.

Hughes, A.L., St Denis, L.A., Palen, L., 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. ACM, pp. 1505–1514.

IBM Big Data and Analytics Hub, 2016. The Four V's of Big Data [WWW Document]. IBM Big Data Anal. Hub. URL http://www.ibmbigdatahub.com/infographic/four-vs-big-data (accessed 5.10.17).

IGI-Global, 2016. What is Data Authenticity.

Imran, M., Elbassuoni, S.M., Castillo, C., Diaz, F., Meier, P., 2013. Extracting information nuggets from disaster-related messages in social media. Proc ISCRAM Baden-Baden Ger.

Indurkhya, N., Damerau, F.J., 2010. Handbook of natural language processing. Chapman and Hall/CRC.

InternetLiveStats, 2017. Twitter Usage Statistics - Internet Live Stats [WWW Document]. URL http://www.internetlivestats.com/twitter-statistics/ (accessed 2.14.17).

Kar, B., 2015. Citizen science in risk communication in the era of ICT. Concurr. Comput. Pract. Exp.

Kar, B., Cochran, D., 2015a. Department of Homeland Security (DHS): Public Perceptions of Warning and Alert Messages. Department of Homeland Security.

Kar, B., Cochran, D., 2015b. Exploring public perceptions of alert and warning messages. DHS - S&T.

Kar, B., Cochran, D., Liu, X., Callais, N., Zale, J., 2016. Integrated Approach Geo-Target At-Risk Communities Deploy Effective Crisis Communication Approaches | Homeland Security.

Karl, T.R., 2009. Global climate change impacts in the United States. Cambridge University Press.

Kelling, S., Fink, D., La Sorte, F.A., Johnston, A., Bruns, N.E., Hochachka, W.M., 2015. Taking a “Big Data”approach to data quality in a citizen science project. Ambio 44, 601–611.

Kelly, R., 2009. Twitter study reveals interesting results about usage. PearAnalytics August 12th.

Kern, R., 2014. Introduction, in: Dynamic Quality Management for Cloud Labor Services. Springer, pp. 3–7.

Kingkade, T., 2013. University Of Colorado Students Post Scenes From Flooded Boulder Campus (PHOTOS/VIDEO). Huffington Post.

Krimsky, S., 2007. Risk communication in the internet age: The rise of disorganized skepticism. Environ. Hazards 7, 157–164.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrian, M., 2016. Rapid assessment of disaster damage using social media activity. Sci. Adv. 2, e1500779.

Kulshrestha, J., Kooti, F., Nikravesh, A., Gummadi, K.P., 2012. Geographic dissection of the twitter network, in: Sixth International AAAI Conference on Weblogs and Social Media.

Laituri, M., Kodrich, K., 2008. On line disaster response community: People as sensors of high magnitude disasters using internet GIS. Sensors 8, 3037–3055.

Landwehr, P.M., Carley, K.M., 2014. Social media in disaster relief, in: Data Mining and Knowledge Discovery for Big Data. Springer, pp. 225–257.

Lee, K., Caverlee, J., Pu, C., 2014. Social spam, campaigns, misinformation and crowdturfing., in: WWW (Companion Volume). pp. 199–200.

Lee, R., Wakamiya, S., Sumiya, K., 2011. Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web 14, 321–349.

Liu, J., Li, J., Li, W., Wu, J., 2016. Rethinking big data: A review on the data quality and usage issues. ISPRS J. Photogramm. Remote Sens. 115, 134–142.

Lundgren, R.E., McMakin, A.H., 2013. Risk communication: A handbook for communicating environmental, safety, and health risks. John Wiley & Sons.

McComas, K.A., Arvai, J., Besley, J.C., 2009. Linking public participation and decision making through risk communication. Handb. Risk Crisis Commun. 364–385.

Meek, S., Jackson, M.J., Leibovici, D.G., 2014. A flexible framework for assessing the quality of crowdsourced data.

Mendoza, M., Poblete, B., Castillo, C., 2010. Twitter Under Crisis: Can we trust what we RT?, in: Proceedings of the First Workshop on Social Media Analytics. ACM, pp. 71–79.

Mileti, D.S., Peek, L., 2000. The social psychology of public response to warnings of a nuclear power plant accident. J. Hazard. Mater. 75, 181–194.

Morgan, S.L., Waring, C.G., 2004. Guidance on Testing Data Reliability. Retrieved Jan 23, 2005.

National Research Council, 2010. Adapting to the Impacts of Future Climate Change.

NOAA, 2014. The record Front Range and eastern Colorado Floods of September 11-17, 2013.

Oxford, 2017a. relevant - definition of relevant in English | Oxford Dictionaries [WWW Document]. Oxf. Dictionaries Engl. URL https://en.oxforddictionaries.com/definition/relevant (accessed 5.13.17).

Oxford, 2017b. reliability - definition of reliability in English | Oxford Dictionaries [WWW Document]. Oxf. Dictionaries Engl. URL https://en.oxforddictionaries.com/definition/reliability (accessed 5.13.17).

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

Palen, L., 2008. Online social media in crisis events. Educ. Q. 31, 76–78.

Palen, L., Vieweg, S., Liu, S.B., Hughes, A.L., 2009. Crisis in a networked world features of computer-mediated communication in the April 16, 2007, Virginia Tech Event. Soc. Sci. Comput. Rev. 27, 467–480.

Preis, T., Moat, H.S., Bishop, S., Treleaven, P., Stanley, H.E., 2013. Quantifying the digital traces of Hurricane Sandy on Flickr. Sci Rep 3, 3141.

Pu, C., Kitsuregawa, M., 2014. Big data and disaster management: A Report from the JST/NSF Joint Workshop [WWW Document]. URL https://grait-dm.gatech.edu/wp-content/uploads/2014/03/BigDataAndDisaster-v34.pdf

Rouse, A.C., 2010. A preliminary taxonomy of crowdsourcing. ACIS 2010 Proc. 76, 1–10.

Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World Wide Web. ACM, pp. 851–860.

Schenk, E., Guittard, C., 2009. Crowdsourcing: What can be Outsourced to the Crowd, and Why, in: Workshop on Open Source Innovation, Strasbourg, France.

Schnebele, E., Cervone, G., Waters, N., 2014. Road assessment after flood events using non-authoritative data. Nat. Hazards Earth Syst. Sci. 14, 1007–1015.

Schnebele, E., Oxendine, C., Cervone, G., Ferreira, C.M., Waters, N., 2015. Using non-authoritative sources during emergencies in urban areas, in: Computational Approaches for Urban Environments. Springer, pp. 337–361.

SearchDataManagement, 2017. What is data quality? - Definition from WhatIs.com [WWW Document]. SearchDataManagement. URL http://searchdatamanagement.techtarget.com/definition/data-quality (accessed 2.18.17).

See, L., Fritz, S., Leeuw, J. de, 2013. The rise of collaborative mapping: trends and future directions. Multidisciplinary Digital Publishing Institute.

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M., 2017. A review of volunteered geographic information quality assessment methods. Int. J. Geogr. Inf. Sci. 31, 139–167.

Sheppard, B., Janoske, M., Liu, B., 2012. Understanding Risk Communication Theory: A Guide for Emergency Managers and Communicators.

Smith, B.G., 2010. Socially distributing public relations: Twitter, Haiti, and interactivity in social media. Public Relat. Rev. 36, 329–335.

Soden, R., Palen, L., 2014. From crowdsourced mapping to community mapping: The post-earthquake work of OpenStreetMap Haiti, in: COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France). Springer, pp. 311–326.

Song, J., Lee, S., Kim, J., 2011. Spam filtering in twitter using sender-receiver relationship, in: International Workshop on Recent Advances in Intrusion Detection. Springer, pp. 301–317.

St Denis, L.A., Palen, L., Anderson, K.M., 2014. Mastering social media: An analysis of Jefferson county's communications during the 2013 Colorado floods, in: 11th International ISCRAM Conference. pp. 737–746.

Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R.M., 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. IConference 2014 Proc.

Starbird, K., Palen, L., 2010. Pass it on?: Retweeting in mass emergency. International Community on Information Systems for Crisis Response and Management.

Stephens, M., Poorthuis, A., 2015. Follow thy neighbor: Connecting the social and the spatial networks on Twitter. Comput. Environ. Urban Syst. 53, 87–95.

Sutton, J.N., Palen, L., Shklovski, I., 2008. Backchannels on the front lines: Emergency uses of social media in the 2007 Southern California Wildfires. University of Colorado.

Thompson, S., Altay, N., Green III, W.G., Lapetina, J., 2006. Improving disaster response efforts with decision support systems. Int. J. Emerg. Manag. 3, 250–263.

Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., Isochi, R., Wang, Z., 2012. Trusting tweets: The Fukushima disaster and information source credibility on Twitter, in: Proceedings of the 9th International ISCRAM Conference. pp. 1–10.

Tirilly, P., Claveau, V., Gros, P., 2008. Language modeling for bag-of-visual words image categorization, in: Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval. ACM, pp. 249–258.

Tripathi, A., Tahmasbi, N., Khazanchi, D., Najjar, L., 2014. Crowdsourcing typology: a review of is research and organizations. Proc. Midwest Assoc. Inf. Syst. MWAIS.

Twitter, 2017. Company | About [WWW Document]. Twitter About. URL https://about.twitter.com/company (accessed 2.14.17).

Twitter Inc., 2017. API Rate Limits — Twitter Developers [WWW Document]. URL https://dev.twitter.com/rest/public/rate-limiting (accessed 6.11.17).

Twyman, M., Harvey, N., Harries, C., 2008. Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. Judgm. Decis. Mak. 3, 111.

UN, U., 2013. Humanitarian and disaster relief assistance. Acessado Em 2, 2013.

United Nations Economic Commission for Europe, 2017. A Suggested Framework for the Quality of Big Data - Google Search [WWW Document]. URL https://www.google.com/search?q=A+Suggested+Framework+for+the+Quality+of+Big+Data&oq=A+Suggested+Framework+for+the+Quality+of+Big+Data&aqs=chrome..69i57j69i64l3j69i60.343j0j4&sourceid=chrome&ie=UTF-8 (accessed 5.13.17).

US Department of Commerce, N., 2013. Denver/Boulder, CO [WWW Document]. URL http://www.weather.gov/bou/ (accessed 11.7.16).

USGS, 2017. USGS 2016 Flood Events [WWW Document]. URL https://water.usgs.gov/floods/events/2016/ (accessed 2.22.17).

Veil, S.R., Buehner, T., Palenchar, M.J., 2011. A Work-In-Process Literature Review: Incorporating Social Media in Risk and Crisis Communication. J. Contingencies Crisis Manag. 19, 110–122.

Vieweg, S., Hughes, A.L., Starbird, K., Palen, L., 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 1079–1088.

Vis, F., Faulkner, S., Parry, K., Manyukhina, Y., Evans, L., 2013. Twitpic-ing the riots: analysing images shared on Twitter during the 2011 UK Riots.

Vuurens, J.B., de Vries, A.P., 2012. Obtaining high-quality relevance judgments using crowdsourcing. IEEE Internet Comput. 16, 20–27.

Wallach, H.M., 2006. Topic modeling: beyond bag-of-words, in: Proceedings of the 23rd International Conference on Machine Learning. ACM, pp. 977–984.

Wang, A.H., 2010a. Don’t follow me: Spam detection in twitter, in: Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on. IEEE, pp. 1–10.

Wang, A.H., 2010b. Detecting spam bots in online social networking sites: a machine learning approach, in: IFIP Annual Conference on Data and Applications Security and Privacy. Springer, pp. 335–342.

Wang, R.Y., Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. J. Manag. Inf. Syst. 12, 5–33.

Ware, J.L., 2003. Geospatial data fusion: Training GIS for disaster relief operations. Natl. Geospatial Intell. Sch. Esri User Conferance July 7–13.

Wendling, C., Radisch, J., Jacobzone, S., 2013. The use of social media in risk and crisis communication.

Western Water Assessment, 2014. WWA | Front Range Floods [WWW Document]. URL http://wwa.colorado.edu/resources/front-range-floods/ (accessed 3.24.17).

Yates, D., Paquette, S., 2011. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. Int. J. Inf. Manag. 31, 6–13.

Zhao, Y., Zhu, Q., 2014. Evaluation on crowdsourcing research: Current status and future direction. Inf. Syst. Front. 16, 417–434.

Zook, M., Graham, M., Shelton, T., Gorman, S., 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Med. Health Policy 2, 7–33.