

Assessing Relevance of Tweets for Risk Communication

Xiaohui Liu^a, Bandana Kar^b, Chaoyang Zhang^c and David M. Cochran^d

^aResilience Research Centre, Dalhousie University, Halifax, NS, Canada

^bUrban Dynamics Institute. Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

^cSchool of Computing, University of Southern Mississippi, Hattiesburg, MS, USA

^dDepartment of Geography and Geology, University of Southern Mississippi, Hattiesburg, MS, USA

Abstract

Although Twitter is used for emergency management activities, the relevance of tweets for risk communication is still open to debate. In this study, six different analytical techniques from computational and geospatial sciences were implemented to assess relevance of risk information extracted from tweets obtained during the 2013 Colorado flood event.

Primarily, tweets containing information about the flooding events and its impacts were analysed. Examination of the relationships between tweet volume and its content with precipitation amount, damage extent, and official reports revealed that relevant tweets provided information about the event and its impacts rather than any other risk information that public expects to receive via alert messages. However, only 14% of the geo-tagged tweets and only 0.06% of the total fire hose tweets were found to be relevant to the event. By providing insight into the quality of social media data and its usefulness to emergency management activities, this study contributes to the literature on quality of big data. Future research in this area would focus on assessing reliability of relevant tweets for disaster related situational awareness.

Keywords: Content Relevance, Twitter, Spatiotemporal Data Mining, Emergency Management, Risk Communication

1. Introduction

Since 1970s, there has been a world-wide increase in severity and frequency of hydro-meteorological events (i.e. floods and tropical cyclones) (UN 2013). Due to continued population growth and rapid urbanization, the natural and built environments of coastal communities are at elevated risk to both coastal and riverine flooding events (Karl 2009; Karl et al. 2009; National Research Council 2010). Hurricanes Harvey, Irma and Maria that impacted the U.S. Atlantic Coast in 2017 provide an insight into the severity of climate related extreme events.

Although the occurrence of these events could not be prevented, their impacts on society and physical environments could be reduced through emergency management and mitigation activities. Risk communication is a crucial component of emergency management, which focuses on disseminating information about an impending disaster to citizens to help them take timely preparatory actions to reduce impacts (V. T. Covello, McCallum, and Pavlova 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 first responders, local public and other stakeholders undertake mitigation actions (Horita et al. 2013).

The ubiquity of computing and IoT devices has enabled individuals and organizations to use social media and social network sites to obtain and share information, coordinate disaster relief efforts, seek assistance during disasters (Estellés-Arolas and González-Ladrón-De-Guevara 2012). In fact, Twitter is considered crucial for risk communication in addition to conventional technologies, such as mass media, TV, radio, etc. (Houston et al. 2015; Ng and Lean 2012; Takahashi, Tandoc, and Carmichael 2015). In 2013, Twitter launched the Twitter Alerts service to help agencies share information during crisis or emergency events (Twitter, 2013).

Although, Twitter allows impacted and interested parties to share near real-time risk information, tweets tend to suffer from data quality issues, due to their unstructured data format, lack of a data and metadata standard, lack of positional accuracy, completeness and semantic accuracy, among others (Meek, Jackson, and Leibovici 2014). This project focused on evaluating the relevance of social media (i.e., twitter) for risk communication during the 2013 Colorado flood. The main questions that were examined in this project were: (i) what techniques and reference data could be used to evaluate the relevance of tweets' content for risk communication? and (ii) what percentage of risk information extracted from tweets is relevant? Relevance in this research refers to *data fitness, i.e. availability and suitability of data and information for specific purpose* (Grady and Lease 2010; Vuurens and de Vries 2012).

2. Literature Review

Risk communication is defined as “the process of exchanging information among interested parties about the nature, magnitude, significance, or control of a risk” (Vincent T. Covello 1992). Traditional risk communication follows a hierarchical, top-down and centralized approach to deliver risk information to at-risk populations about a hazard event and its potential adverse impacts (Gladwin et al. 2007). This approach, however, often fails to motivate public to respond positively to alerts and warnings due to lack of specific instructions, misunderstanding of the information provided by messages, and language barrier (Colley and Collier 2009; Kar and Cochran 2015). By contrast, network-based risk communication technologies, such as 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), use a bottom-up and collaborative

approach that allows both impacted and interested populations to share unlimited information about a hazard event, irrespective of their geographic location and time (Kar 2015). These sites have been used extensively during past disasters, such as Hurricanes Sandy (2012) and Harvey (2017), and Haiti earthquake (2010) (Preis et al. 2013; Yates and Paquette 2011; Boulianne, Minaker, and Haney 2018).

Despite popularity of network-based risk communication approach, several limitations prohibit emergency management personnel (EMP) from using the data obtained from social media and social network sites. First, there is no transparent procedure in place to allow use of social media data for decision making purposes (Department of Homeland Security Press Office 2004; Hughes and Palen 2012). Second, the dynamic nature of social media, and its abbreviated and informal styled content makes it difficult to craft and disseminate alert messages for public consumption (Crowe 2010; Hughes and Palen 2012). Third, citizen generated data often lacks metadata (i.e. creator, location, and device information), thereby making it less trustworthy (Meek, Jackson, and Leibovici 2014). Last, but not least, citizens who generate data tend to have varying perceptions and educational backgrounds, and possess distinct life experiences, which often leads to generation of dubious quality data (Porter, Verdery, and Gaddis 2016).

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. This research examines *relevance of data*, which is defined as “the condition of being connected or appropriate to what is being considered” (Cai and Zhu 2015; Oxford 2017), or “if it has a logical, sensible relationship to the finding it supports” (Morgan and

Waring 2004). Although connotation of relevance varies with context of usage, there are some shared common characteristics, such as the timeliness of relevant data and closeness of data to its context, which were used to assess quality of tweets in this research.

The different analytical techniques and approaches used to assess relevance include – algorithm development using authoritative data (Comber et al. 2013; Meek, Jackson, and Leibovici 2014), experts’ knowledge (See, Fritz, and Leeuw 2013), semantic analysis (Gabrilovich and Markovitch 2007; Hofmann 1999), contextual factors (i.e. seasonal effect) (Baltrunas et al. 2012), weighted criteria (Gough 2007), and internet-based human labour (Alonso and Mizzaro 2012; Goodchild and Li 2012; Grady and Lease 2010). For example, Page et al. (1999) used web page ranking to evaluate web links’ relevance to queries; Duan et al. (2010) applied supervised machine learning to prioritize relevance of tweets to queries; Alonso and Mizzaro (2012) used crowdsourcing, such as Amazon Mechanical Turk to assess relevance; and Chae et al. (2012) used Latent Dirichlet Allocation (LDA) to extract relevant information by identifying major topics and by probability ranking. These models and techniques tend to not incorporate any other type of data, such as geospatial and meteorological data to assess relevance of crowdsourced and social media data (Cheong and Cheong 2011; Caragea et al. 2011; St Denis, Palen, and Anderson 2014).

The lack of inclusion of other data sets/information pertaining to hazards and specific emergency management activities tends to make these studies less useful. In this research, the relevance of tweets was evaluated using six geospatial and computation techniques (i.e., spatial, spatiotemporal, and temporal data parameterization approaches, content analysis, cosine similarity comparison, and relevance score assessment). Several geospatial data sets (i.e.,

precipitation, flood extent, and degree of damage) and authoritative data (i.e., damage assessment reports) were used as reference data for relevance assessment.

3. Methodology

3.1. Study site:

In 2013, the Front Range, EL Paso County, Boulder County, and portions of the Denver metropolitan area in Colorado were flooded by severe flooding. The devastating flash flood that resulted from historically severe precipitation during September 9th through September 18th caused considerable damage to a number of counties. Figure 1 depicts the counties that experienced worst, severe, moderate, and least damage due to flooding (Federal Emergency Management Agency 2014).

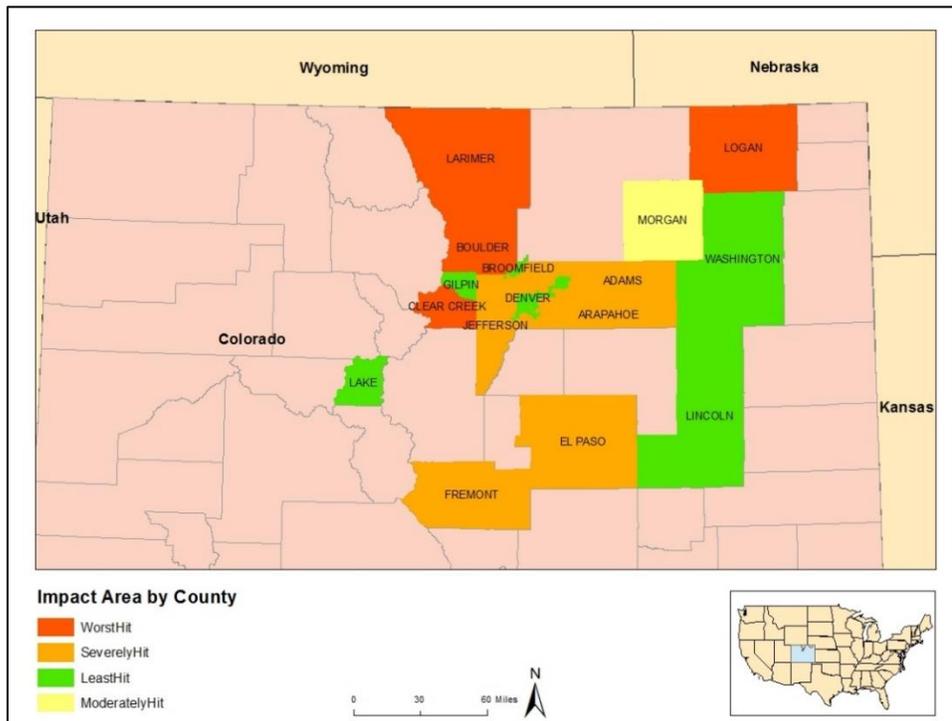


Figure 1 Study site (Counties Impacted by 2013 Flood in Colorado)

Figure 2 shows the hourly precipitation accumulation during the storm event across Front Range. Boulder County, one of the worst hit areas, experienced 9.4 inches of precipitation on Sep. 12th, which was comparable to the county’s average annual precipitation. Other areas in the Front Range experienced a small amount of precipitation and accumulation on September 15th.

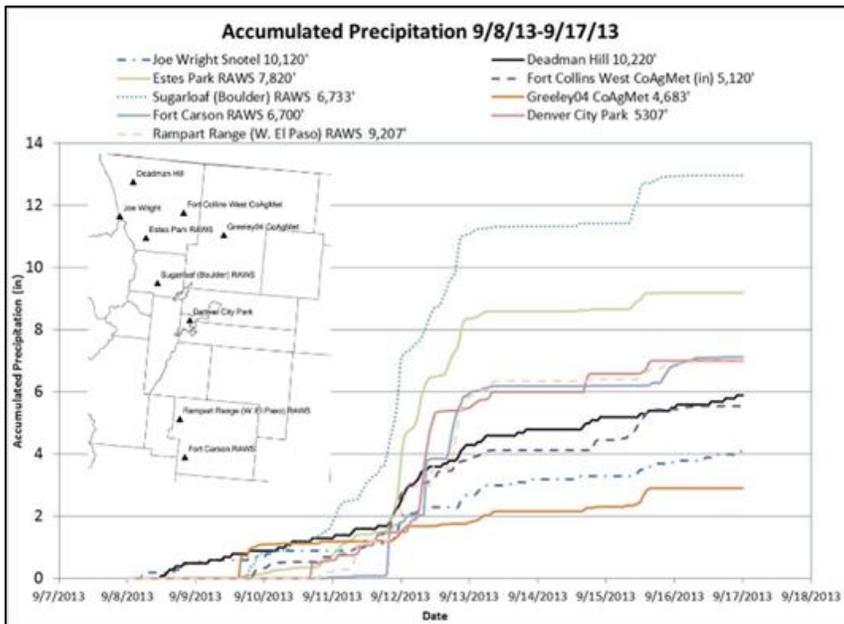


Figure 2 Accumulated precipitation resulting from the 2013 Colorado Floods (CCC, 2013)

3.2. Data sets and processing:

For this study, in addition to tweets, geospatial and survey data sets were obtained from various sources to assess relevance of tweets in terms of their content about the flooding event. A discussion of data sets, data processing and analytical techniques used to answer the research questions is presented in the following sections.

(1) *Tweets of 2013 Colorado Floods*: Historical tweets were purchased from Twitter Inc. using a set of keywords pertaining to locations (Colorado, Boulder, Front Range, El Paso County and

Boulder County, Denver metro), 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. The firehose data contained a total of 1,195,183 tweets in JSON (JavaScript Object Notation) format, which were stored in MongoDB - an open-source cross-platform database for unstructured data that uses dynamic schemas. Figure 3 depicts the steps that were implemented to process the tweets for analytics.

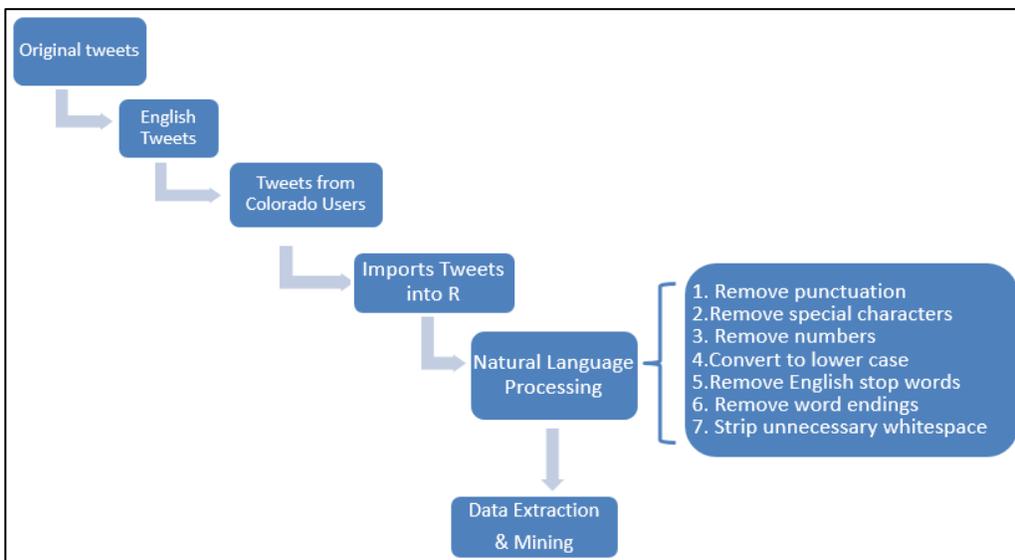


Figure 3 Flow chart of tweets processing

Among the total tweets, 85% were in English which were used in this study. From the English tweets, 1.38% had geo-location information, and 0.44% of the geo-tagged tweets were generated within Colorado. The 0.44% of the geo-located tweets were used in this study as it was assumed that these tweets were generated by those who experienced or witnessed the floods. This assumption assigned geographic relevance to the tweets and eliminated the possibility of including tweets containing misinformation or rumours, and that were generated by geographical “outsiders” who did not experience or witness the flood.

Table 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%

From the geo-located tweets, a list of top frequent words and hashtags (words and phrases prefixed by #) corresponding to the flood event was extracted, which was used to mine flood-related tweets. The top frequent words (Figure 4) were: *Colorado, Boulder, Denver, flooding, warning, September*. The top 10 hashtags with high frequency were: *Colorado, boulderflood,*

Coflood, cowx, NeverForget, flooding, GodBlessAmerica, news, COREcall, Denver. High frequency hashtags, such as “NeverForget” and “GodBlessAmerica”, that were misleading were removed from the list of keywords and hashtags before creating the corpus of tweets containing information relevant to the flooding event for analysis.

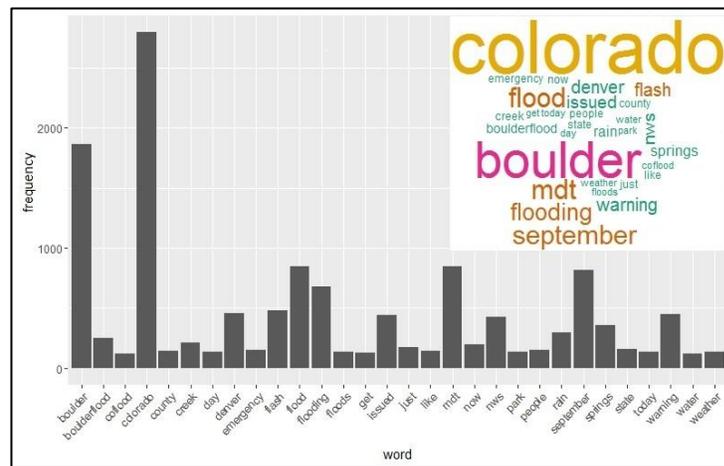


Figure 4 Top frequent words and corresponding word cloud

(2) *Geospatial Data*: The flooding event of September 9th, 2013 was a historically severe flood since the 1995 flood that occurred in Colorado. For this event, 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). The precipitation data was used to explore the relationship between temporal volume of tweets and precipitation amount, and to evaluate the temporal relevance of tweets to the flood event. To understand the spatial distribution of tweet volume with respect to flood impacted areas, flood extent data was obtained from the City of Boulder (City of Boulder 2014c). This data was generated based on field surveys, Digital Globe Worldview satellite imagery, crowdsourcing, and input from affected property owners. Street network data was also obtained from the City of Boulder to evaluate relevance of tweets with regard to flood damages to roads and streets (City of Boulder 2014a).

(3) *Survey Data*: A survey data set containing information about residents' choice of contents that should be included in an emergency alert message was obtained from the University of Southern Mississippi (Kar and Cochran 2015). This data was collected as part of a Department of Homeland Security (DHS) funded project that examined the risk communication requirements of the Mississippi Gulf Coast residents, specifically, public expectations of contents to be included in an alert message. The information extracted from tweets was compared with survey data to assess relevance of tweets in disseminating risk information as needed by at-risk population.

(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 (NOAA 2013). The messages contained meteorological forecasts, observations, warnings, advisories, and other information as the flood unfolded. These alert/warning messages were combined as a single text message before using them as reference to evaluate relevance of tweets.

(5) *Official Warning and Damage Assessment Reports*: The official damage assessment reports convey possible threats due to heavy rain and flood, and contain information pertaining to post-event impacts (City of Boulder 2014c, 2014b; Lukas 2014; NOAA 2014; Wright Water Engineers, Inc. 2014). These reports also provide situational awareness about flooding and summarize damage to properties and infrastructures. In addition to these reports, newspaper articles that provided other facts (i.e. local damage, individual responses to the floods) were obtained to assess tweet relevance.

3.3. Analytics and techniques

Six different computational and geospatial techniques were used to evaluate relevance of tweets based on contents, specifically, risk information (Figure 5). The following section discusses the steps implemented to extract relevant risk information from tweets.

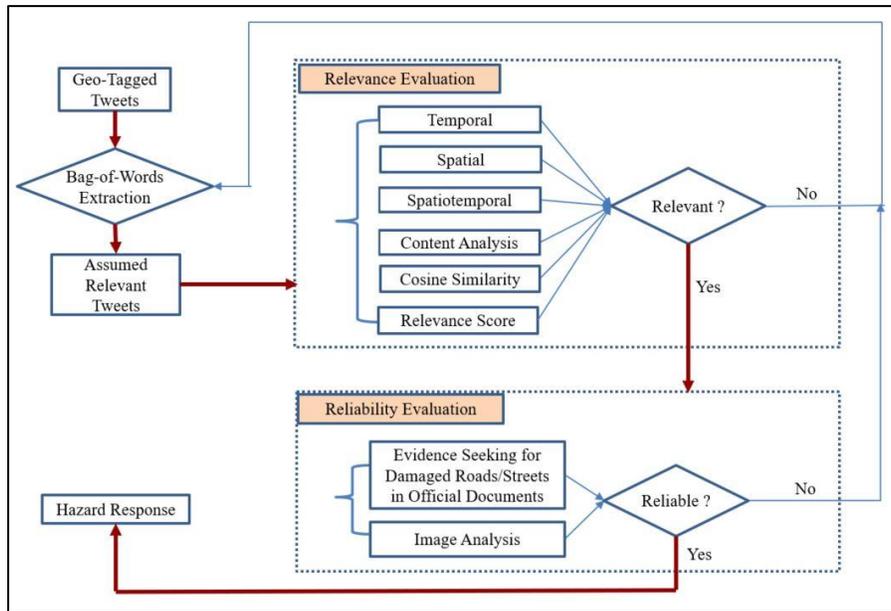


Figure 5 Flow-diagram of analytical steps

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, Claveau, and Gros 2008; Wallach 2006). The bag-of-words contains topic-specific search terms to measure the relevance of a document to the search terms. In this study, the bag-of-words contained top frequency words and high frequency hashtags (discussed above) corresponding to impacted locations, event impacts, and emergency management activities. While the top frequency words served as indicators of popular topics that were covered by tweets, the high frequency hashtags helped identify tweets related to specific topics. Using the bag-of-words model, tweets containing information about the flood event were extracted from the 0.44% of geo-tagged tweets.

Survey Responses to Warning/alert Message Content: The analysis of survey responses (discussed above) revealed that citizens prefer inclusion of following information 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”. Each choice reflects a critical component of risk information. The findings of statistical analyses of survey responses were compared with risk information extracted from tweets to determine if difference exists between what people expect and what was conveyed via tweets.

3.3.2 *Evaluating relevance of tweets*

- (1) *Temporal*: Continuous heavy rainfall caused flooding in the Front Range in Colorado in September 2013, which also contributed to an increase in tweet volume during September

9th – 18th. Therefore, the relationship between temporal distribution of tweet volume and precipitation amount was determined to evaluate temporal relevance of tweets.

- (2) *Spatial*: If tweets are relevant to the flooding event, then it could be assumed that their spatial distribution should not be random. Therefore, the spatial distribution of tweets should correlate to the degree of damage experienced across the study site, which was examined statistically to assess spatial relevance of tweets.
- (3) *Spatiotemporal*: Because of intense flooding and extensive damage in Boulder County, the 2013 Colorado floods is called the Boulder flood. Therefore, it could be assumed that the tweets were representative of the spatial and temporal distribution of flood and its associated impacts in the study site. 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*: A classification of risk information, which contains seven categories, was adopted from a survey that was created for a DHS funded project. Referred to this classification, content information belongs to the seven categories was extracted respectively. No matter which category a tweet belongs to, it could be considered relevant. The number of tweets belongs to each category were then divided by the number of geo-tagged tweets to yield the percentage of each category. On the other hand, the number of survey responses to each category were divided by the total survey responses to generate the percentage of each category for the survey. The percentage information derived from both datasets were then compared. 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 was determined.

(5) *Cosine similarity comparison*: Cosine similarity comparison is a vector space model used for comparing document relevance or similarity (Erk and Padó 2008; Turney and Pantel 2010). In cosine similarity, documents are represented as vectors, and each vector holds a place for every term in the document collection. The approach calculates the cosine angle between two non-zero vectors (two documents) and the similarity score represents the degree of relevance (0 - no relevance and 1 - relevant) (Indurkha and Damerau 2010). Given two documents, $d1$ and $d2$, cosine similarity could be computed using equation 1:

$$\text{Similarity} = \cos(\theta) = \cos(d1, d2) = \frac{d1 \cdot d2}{\|d1\| \|d2\|} \quad \text{equation (1)}$$

where “ \cdot ” indicates vector dot product, $\|d\|$ is the length of the vector d .

In this project, contents of tweets were compared with official warnings and damage reports using cosine similarity. Each document vector (tweets and official reports) was assigned a value of “1” or “0” to represent a presence or absence of a 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. Although term frequency-inverse document frequency could have been used to assess similarity, due to the unstructured nature of tweets and their 140-character limit, they could not be directly compared with official reports. Hence, instead of direct comparison of documents, top 50 frequent words and top 10 non-redundant hashtags extracted from tweets 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 trending topic(s) from social media, and top frequent words from official

reports represent risk information from authorities. If a certain degree of similarity existed between a tweet and the report, then the tweet was considered relevant.

- (6) *Relevance score*: MongoDB's built-in function (*\$text*) returns a score representing the relevance of a document to a given text search query. Particularly, when one or more search terms are provided, MongoDB tokenizes and stems the text content of each tweet by creating an index for each tweet using the *\$text* operator, and then determines relevance of search terms to each tweet using the *{ \$meta: "textScore" }* expression and yields a score for each tweet. The top 10 frequent words and top 10 hashtags were used as search terms to compute a relevance score for each of the 5202 geo-tagged tweets in Colorado. Three differences exist between relevance score computed by MongoDB and cosine similarity comparison approach: (i) the former used predefined terms, while the latter compared top keywords from tweets and official reference documents; (ii) the former used default MongoDB function (*\$text*), and the latter used cosine similarity function; and (iii) the former yielded a relevance score for each tweet, whereas the latter returned a single similarity score measuring relevance of the documents being compared.

4. Results and Discussion

4.1. Temporal trend of tweet volume vs. precipitation amount

The daily volume of tweets and total precipitation across all rain gauge stations in Colorado from September 11th - September 15th were compared to investigate the relationship between tweet volume and precipitation amount over time. Both data sets were normalized to have the values range between "0" and "1" for comparison. Figure 6 plots the daily distribution of tweets and

precipitation amount. Figure 7 depicts the hourly distribution of these data sets.

It is apparent from Figure 6 that tweet volume increased with increase in precipitation amount and experienced a significant increase on September 12th. Both tweet volume and precipitation dropped after September 12th. Interestingly, tweet volume increased from 9:00 am until 6:00 pm when it reached the daily peak. The daily distribution of tweet volume and precipitation roughly have the same pattern for every day except for September 13th, when tweet volume was high although precipitation was low. Although the Pearson correlation coefficient between hourly tweet volume and precipitation amount was 0.544 ($p = 0.05$), the Pearson correlation coefficient between daily tweet volume and precipitation amount was 0.778 ($p = 0.05$). This indicates that large volume of tweets was generated on the days when precipitation was also heavy except for September 13th. This could be because the flooding was so destructive that it continued to be a significant topic of discussion on social media despite the one-day halt in precipitation on September 13th. It could also be concluded that relevant tweets were produced on days when Boulder experienced significant rainfall and flooding.

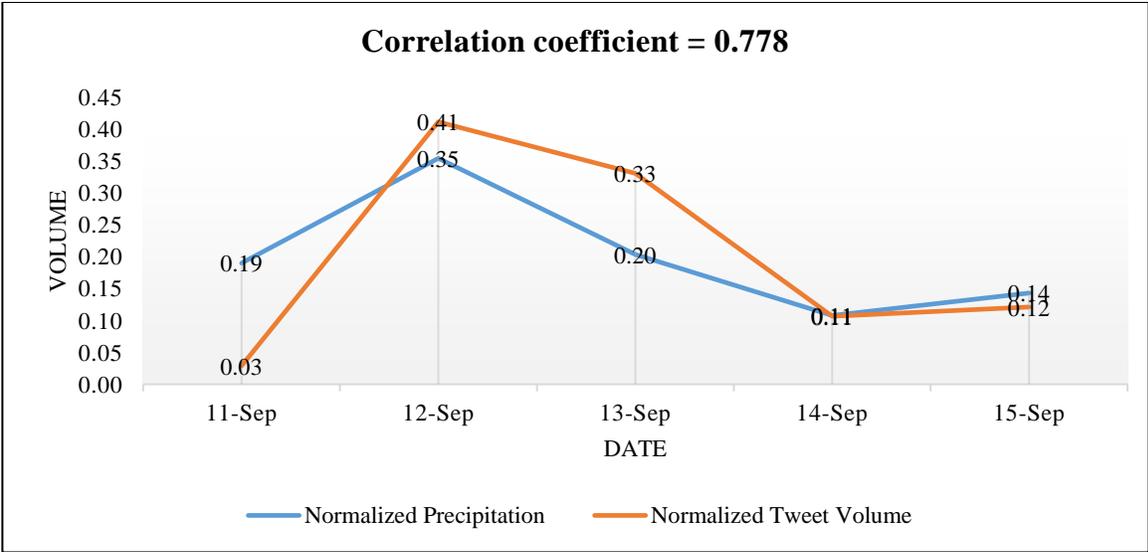


Figure 6 Correlation between daily volume of tweets and precipitation

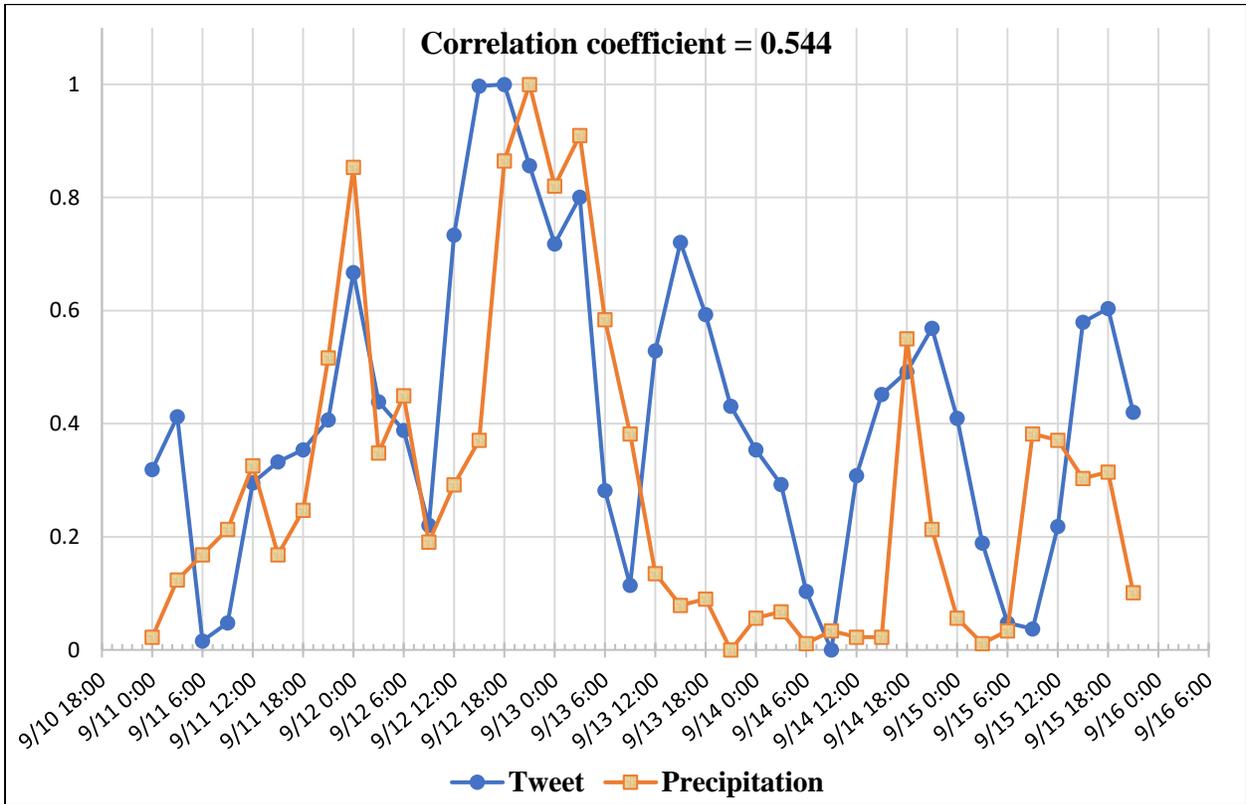


Figure 7 Correlation between hourly volume of tweets and precipitation

4.2. Spatial distribution of tweets vs. degree of damage

To understand the relationship between spatial distribution of tweet volume and spatial distribution of damage, the tweets were aggregated by city, and overlaid with the impacted area map (Figure 8). Instead of creating buffers surrounding flood extent boundary, administrative boundaries were used due to the following two reasons: (i) lack of data about flood extent in counties other than Boulder, and (ii) the degree of damage data used in this analysis was only available at administrative boundary level.

A visual interpretation of the map clearly reveals that the tweets were concentrated in counties/cities that experienced severe damages. A cluster analysis (Average Nearest Neighbour) of the tweets based on spatial distribution of degree of damage yielded a z-score of -113.78 (p-value = 0.000). This indicates that there is less than 1% likelihood that the clustered pattern could be the result of random chance. Table 2 lists the cities plotted in Figure 7 along with the 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 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 the top four cities by population and by Tweet volume; 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.

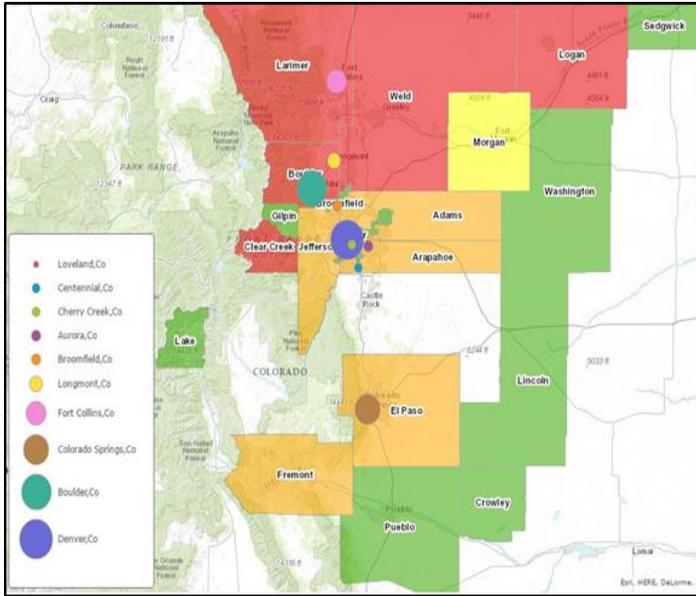


Figure 8 Tweet volume across damaged counties

Table 2 Rankings of City by Tweet Volume vs. Population

City	Rank By Tweet Volume	Tweet Volume	Degree of Damage	Rank By Population (“Colorado City Rank” 2016)	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

Tweet volume, city population and degree of damage were normalized to a scale of “0” to “1” and plotted in a stacked line chart (correlation coefficient = 0.54, p=0.05) (Figure 9). It is evident from Figure 8 that tweet volume (blue line) decreased with decreasing population except for one trough and one crest.

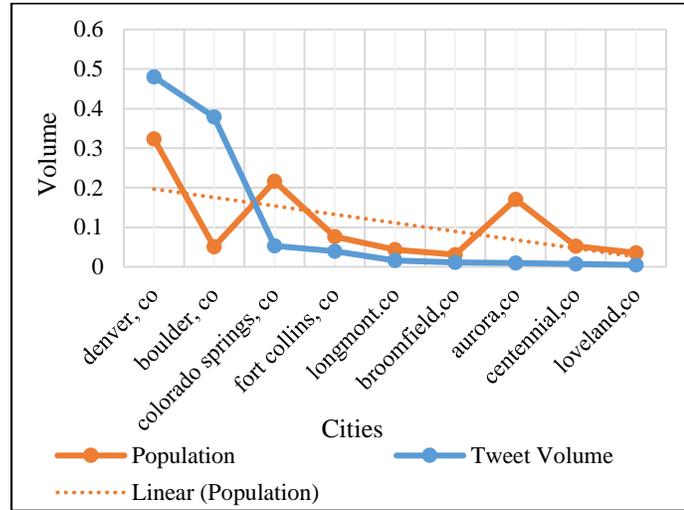


Figure 9 Tweet volume vs. city population

Being the worst damaged city, Boulder ranked second in tweet volume but eleventh in population and therefore created a trough in population curve. Although Aurora ranked higher in population, due to less severity in damage, the tweet volume for Aurora was also low.

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, and tweet volume is impacted by both population and damage extent. To quantify the impact of population and degree of damage on tweet volume, a regression analysis was implemented on the normalized values, which yielded a relation presented by equation 2:

$$Tweet\ Volume = 0.02 + 0.3 * Population + 0.1 * Degree\ of\ Damage \quad \text{equation (2)}$$

With a R square of 0.59, the equation explains 59% changes in tweet volume that are caused by population and degree of damage. Based on coefficient values, it is evident that both population and degree of damage contribute positively to increase in tweet volume. Also, population count appears to have a bigger impact on increase in tweet volume.

4.3. Spatiotemporal analysis of tweets

The geo-tagged tweets for Boulder (the county that experienced the worst damage) were examined to understand their spatiotemporal distribution in flood impacted areas of Boulder.

Figure 10 depicts the spatial distribution of geo-tagged tweets within Boulder city limit over six days (September 10th – September 15th).

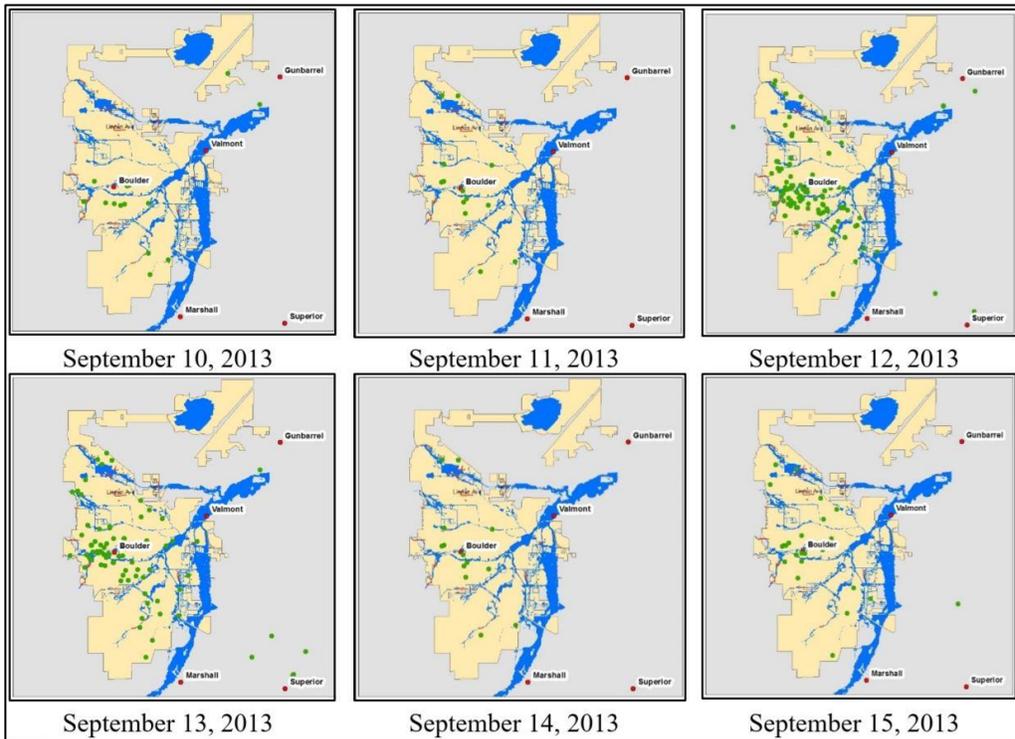


Figure 10 Spatiotemporal distribution of tweets across Boulder

The time-series distribution of tweets (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 10 indicates 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 sparsely distributed. Furthermore, the geo-tagged tweets are concentrated along the flooded river/creek channels rather than spread across the city.

The maximum threshold distance (between tweets and flood extent, in Kilometres) within which the tweets are presented on each day was derived to depict the spatial distribution of the tweets relative to the flooded area. Figure 11 depicts the relationship between the threshold distance and percentage daily distribution of tweets within the distance threshold. A Pearson correlation coefficient of -0.76 indicates a strong negative relationship between the two measures. Although the tweet volume increased daily with increasing precipitation, their spatial distribution is confined to the areas where flooding was severe. It could be concluded that relevant tweets were generated by population that experienced flooding and subsequent damage.

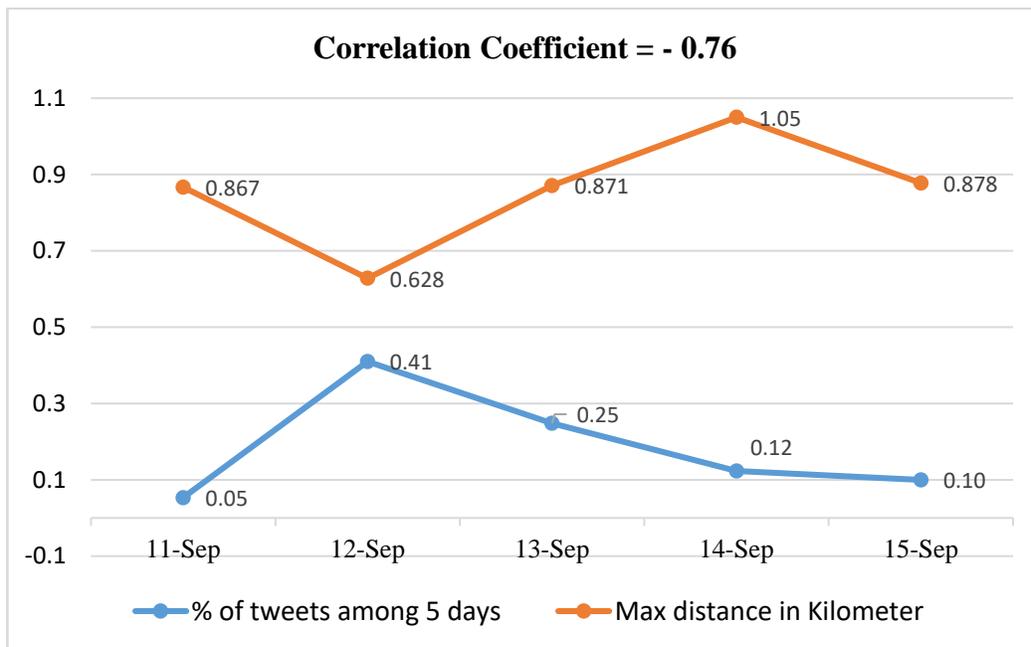


Figure 11 Relationship between daily tweets and distance to flooded area

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 as it reduced the massive data to a small portion to be effectively used for emergency management purpose.

4.4. Content analysis

A content analysis of geo-tagged tweets was conducted using a list of keywords (Table 3) to extract risk information belonging to the categories identified by the DHS study about the information that need to be included in an alert message (discussed above). The results of content analysis were compared with survey responses (Figure 12). The survey responses 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. By contrast, 62.18% and 82.18% tweets contained information about *nature of the disaster* and *impact zone*, respectively followed by *recommended action* (11.49%), *damage & loss* (8.32%), *contact for help* (4.16%), and *shelter location* (3.37%).

Table 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, centre
Recommended action	Action, evacuate, alert, siren, warning, stay safe, stay dry, stay inside, higher ground
Impact zone	Boulder
Nature of disaster	Flood, flooding

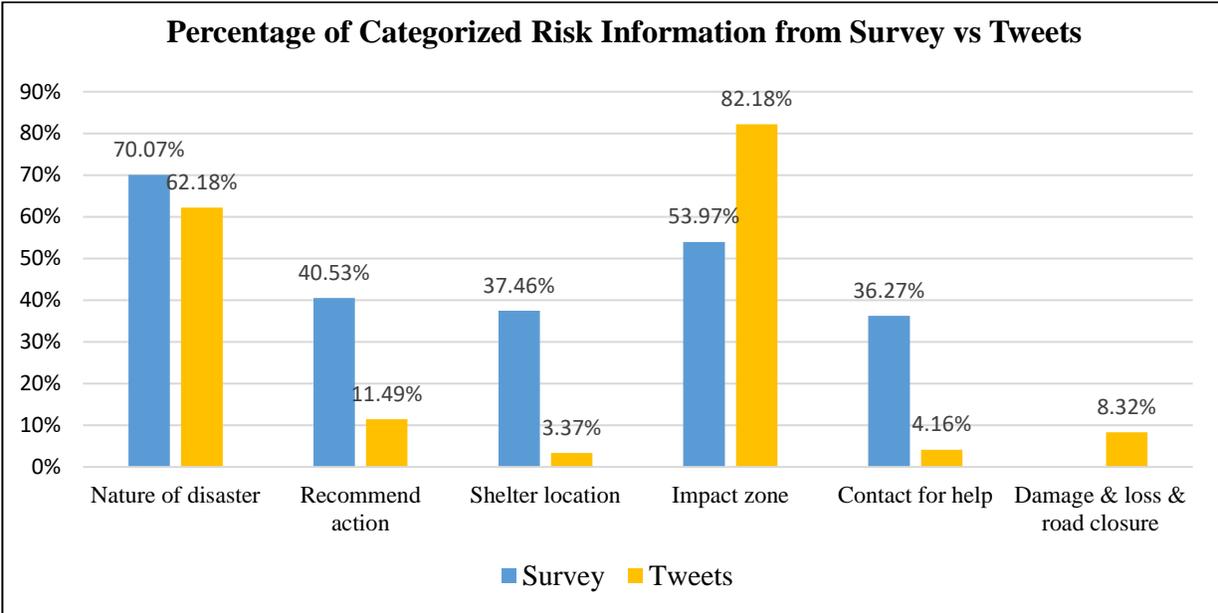


Figure 12 Percentage of categorized risk information from survey vs tweets

Although the kind of risk information people expect in an alert message differs from what is shared by individuals via tweets, both the survey and tweet analysis revealed that the public requires and shares information about *nature of the disaster* and *impact zone* rather than other aspects of risk. This 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 (Morrow 1999). Likewise, *recommended action* or *damage & loss* is meaningless unless *the nature of the disaster* and *impact zone* are known. Nonetheless, 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.

4.5. Cosine similarity comparison

Cosine similarity comparison generated 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, and the other document contained top 50 frequent words from NOAA warning/alert messages and official damage assessment reports.

Several methods can be used to create document vectors, such as raw term frequency and binary weights. While tweets are a collection of keyword-centric documents, NOAA alert messages and official damage assessment reports are story-based documents with contextual content. This 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. Furthermore, the following steps were implemented to convert each document to a vector (Table 4): (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 auxiliary words that had no relationship to flood (i.e. “great”, “may”, and “love”).

Table 4 Document Vectors of Tweets and Official Reports

Terms	Tweets	Official	Terms	Tweets	Official
Boulder	1	1	noaa	0	1
Center	1	1	nws	1	1
Closed	1	0	park	1	1
coflood	1	0	precipitation	0	1
colorado	1	1	rain	1	1
County	1	1	recommendation	0	1
Creek	1	1	river	1	1
denver	1	1	road	1	0
emergency	1	0	safe	1	0
Flash	1	1	september	1	1
Flood	1	1	springs	1	1

Flows	0	1	warning	1	0
forecast	0	1	water	1	1
Front	0	1	weather	1	1
Heavy	0	1	hydrologic	0	1
Help	1	0	issued	1	0
News	1	0	flooding	1	1

Cosine similarity scores ranged from “0” (meaning dissimilar or not relevant) to “1” (meaning very similar or the same). The cosine similarity score for this study was: $Similarity = \cos(45.5^\circ) = \cos(d_{tweets}, d_{official\ reports}) = 0.7$, which 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.

4.6.Relevance score

A relevance score was generated for each of the 5202 geo-tagged tweets in Colorado. Table 5 lists five randomly selected tweets in descending order of their respective relevance score. The relevance scores varied between 0.5 (lowest score) and 3.79 (highest score) for the 5,202 tweets. Figure 12 shows the score distribution. The tweets with high relevance score (i.e. the first and second tweet) contained flood relevant risk information than lower scored ones. The lowest scored tweets, such as the fifth tweet does not contain any information relevant to the flood event. After careful screening of the tweet content, a threshold value of 1.3 was manually selected to separate relevant tweets from irrelevant ones, which results in roughly 14% relevant tweets based on their score among the geo-tagged tweets in Colorado.

Table 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
	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	Boulder's still gorgeous even after a storm @ University of Colorado Boulder http://t.co/r3ymLfkFG5	1.42

To better assess relevance, the tweets were categorized into four groups based on their relevance score, *low relevance* (1.3 ~ 1.5), *moderately low* (1.5 ~2), *moderately high* (2 ~2.5), and *high relevance* (2.5 ~ 3.79). The distribution of the relevance score can be found in Figure 13, the approximate normal distribution indicates a reasonable categorization of the relevance levels. A The cut-off point for each group was chosen around the inflexion of relevance score trend line. Based on their relevance score, tweets were classified into the risk information categories (Table 3). From Figure 14, it is obvious that relevant tweets tend to provide information about *nature of the disaster* and *impact zone* as opposed to other tweets. Also, the percentage distribution of tweets in each category is similar to the trend observed by the entire volume of tweets based on their content. It could be concluded that relevance score is an important measure to identify relevant tweets for emergency management purpose.

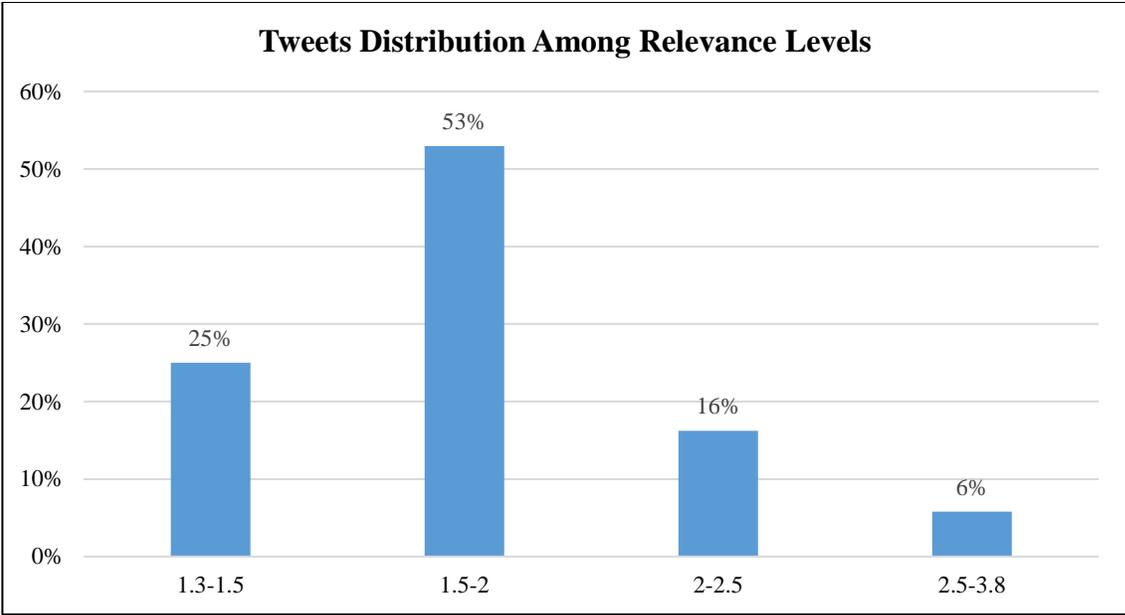


Figure 13 Tweet distribution among relevance levels

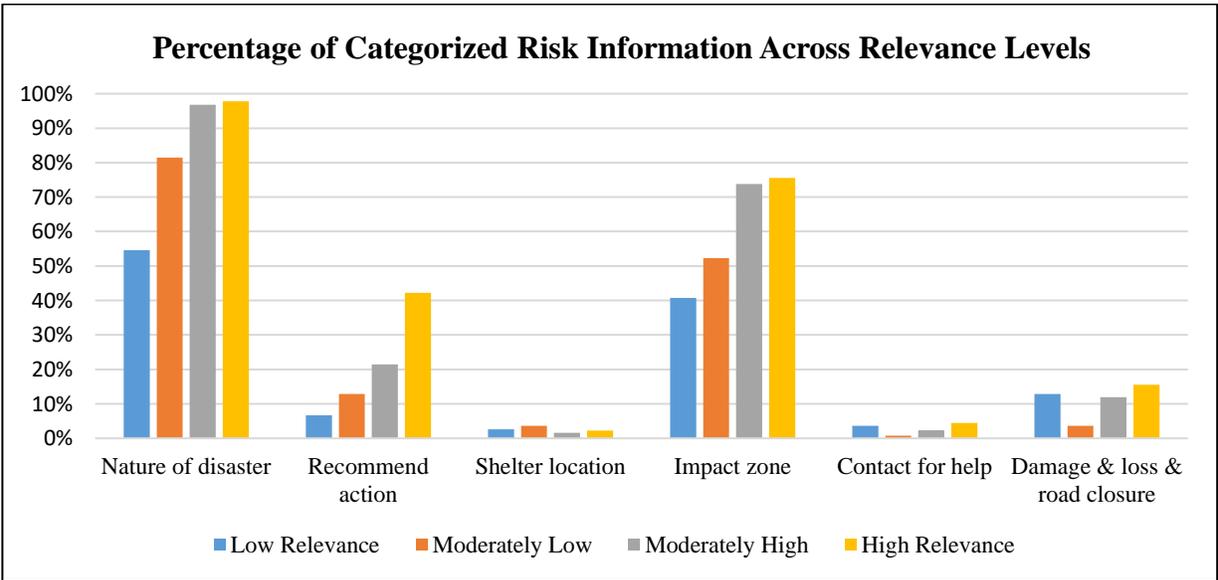


Figure 14 Percentage of categorized risk information across relevance levels

5. Conclusion and future work

Despite its popularity for providing up-to-date information pertaining to disasters, and although several studies have evaluated data quality of tweets, research examining relevance of tweets

based on risk information availability is at its infancy. In this research, an attempt was made to evaluate relevance of geo-tagged tweets to ensure that the tweets were generated by those who experienced or witnessed the 2013 Colorado flood rather than by “*outsiders*”.

Each technique used to find relevant tweets 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. Although the spatial and temporal techniques helped extract relevant tweets that were spatially clustered on days when heaviest precipitation occurred 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 the only indicator of severely damaged areas to be targeted for recovery and response efforts as pointed out by (Kryvasheyev 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. 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 DHS study.

The cosine similarity score (0.7) indicated that the tweets contained information similar to those provided by NOAA and official reports. However, just because the tweets are relevant to the reports does not mean that the tweets contained detailed information about the event or its impacts. So, relevant tweets must be used with caution and complemented with ancillary data.

6. Limitations and future research.

Although the geo-location approach implemented to extract tweets for this study probably eliminated misinformation or rumours (Lee, Wakamiya, and Sumiya 2011), it may not be effective in case of other hazards occurring in other countries. 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 was seen during Hurricane Maria (2017) that left the entire island of Puerto Rico without power and subsequently, without any means to communicate using social media. The rising concern about privacy also prohibits citizens from sharing their personal information including location. There are other limitations (discussed below) that also hinder the use of tweets by EMP during emergency management.

- (1) The data used in this research were purchased, which is not always possible in real-time. From an emergency management perspective, purchasing tweet is not feasible during emergencies when disaster response and recovery efforts require large amount of fund. Purchasing tweets ensures post-processing rather than real-time analytics for emergency response purposes. Although real-time tweets could be obtained freely using Twitter API, this automatic download tends to return only 1% of the total tweet volume that match the search terms (Twitter Inc. 2017). Though probable, it is unclear to what extent the low percentage of real-time tweets obtained using API would eliminate relevant and irrelevant tweets that were otherwise obtained by purchasing from Twitter Inc.
- (2) Data collection, data cleaning, and analytics are time consuming, computationally intensive, and require skilled professionals. Therefore, using social media data for

emergency management activities would be inefficient for local EMP unless automated tools and algorithms enable tweet use in real-time.

- (3) Given that the Emergency Management Agencies' (EMA) responsibilities during an emergency involve coordination of 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, thereby increasing situational awareness.
- (4) 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. However, it was assumed that a large number of local public generated the tweets, which in reality may not be the case. If that happens, then the tweet volume would be small enough to be useful for emergency management activities. Furthermore, the methodology implemented herein may not be useful with a small percent of tweets. Therefore, it might be efficient and effective to use crowdsourcing or a citizen science based approach to impacted population to share information specific to a hazard event (Kar 2015).
- (5) The bag-of-words model aided with extracting relevant tweets based on predominant keywords, which did not account for the underlying context of those words. Although this model is appropriate to extract specific risk information, it may not be useful for categorizing tweets based on their content for emergency management efforts.
- (6) Out of the total geo-tagged tweets (0.44% of the total tweets) used in this study, only 14% were relevant to the flood event. 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. Moreover, relevant tweets must be assessed for reliability before using them for emergency management efforts.

Acknowledgement

Dr. Bandana Kar has participated in this project in her own independent capacity and not on behalf of UT-Battelle, LLC, or its affiliates or successors. The views and conclusions expressed in this article are those of the authors and do not reflect the policies or opinions of the funding agency, Oak Ridge National Laboratory, UT-Battelle, the Department of Energy, or the US Government.

Disclosure Statement

No potential conflict of interest was reported by the authors

Funding

This research was funded partially by the National Science Foundation [grant no CMMI-1335187], and the 2014, 2015, and 2016 Arthell Kelley Scholarships from the Department of Geography and Geology at The University of Southern Mississippi

References

- Alonso, Omar, and Stefano Mizzaro. 2012. "Using Crowdsourcing for TREC Relevance Assessment." *Information Processing & Management* 48 (6): 1053–66. <https://doi.org/10.1016/j.ipm.2012.01.004>.
- Baltrunas, Linas, Bernd Ludwig, Stefan Peer, and Francesco Ricci. 2012. "Context Relevance Assessment and Exploitation in Mobile Recommender Systems." *Personal and Ubiquitous Computing* 16 (5): 507–26. <https://doi.org/10.1007/s00779-011-0417-x>.

- Boulianne, Shelley, Joanne Minaker, and Timothy J. Haney. 2018. "Does Compassion Go Viral? Social Media, Caring, and the Fort McMurray Wildfire." *Information, Communication & Society* 21 (5): 697–711. <https://doi.org/10.1080/1369118X.2018.1428651>.
- Cai, Li, and Yangyong Zhu. 2015. "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era." *Data Science Journal* 14. <http://datascience.codata.org/articles/10.5334/dsj-2015-002/>.
- Caragea, Cornelia, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, et al. 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. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.370.6804&rep=rep1&type=pdf>
- CCC. 2013. "Colorado Climate Center." <http://coflood2013.colostate.edu/>.
- Cheong, France, and Christopher Cheong. 2011. "Social Media Data Mining: A Social Network Analysis of Tweets during the 2010-2011 Australian Floods." https://works.bepress.com/christopher_cheong/2/.
- City of Boulder. 2014a. "Main Roads and Streets Map." <https://bouldercolorado.gov/open-data>.
- . 2014b. "Summary Report of Private Property and Resident Flood Impact Survey and Analysis September 2013 Flood Disaster."
- . 2014c. "The 2013 Boulder Flood Extent Map."
- Colley, Kay L., and Amy Collier. 2009. "An Overlooked Social Media Tool? Making a Case for Wikis." *Public Relations Strategist*, 34–35.
- "Colorado City Rank." 2016. <http://worldpopulationreview.com/states/colorado-population/cities/>.
- Comber, Alexis, Linda See, Steffen Fritz, Marijn Van der Velde, Christoph Perger, and Giles Foody. 2013. "Using Control Data to Determine the Reliability of Volunteered Geographic Information about Land Cover." *International Journal of Applied Earth Observation and Geoinformation* 23: 37–48.
- Covello, V. T., David B. McCallum, and Maria T. Pavlova. 2012. *Effective Risk Communication: The Role and Responsibility of Government and Nongovernment Organizations*. Springer Science & Business Media.
- Covello, Vincent T. 1992. "Risk Communication: An Emerging Area of Health Communication Research." *Annals of the International Communication Association* 15 (1): 359–373.
- Crowe, Adam. 2010. "The Elephant in the JIC: The Fundamental Flaw of Emergency Public Information within the NIMS Framework." *Journal of Homeland Security and Emergency Management* 7 (1).
- Department of Homeland Security Press Office. 2004. "Department Of Homeland Security Secretary Tom Ridge Approves National Incident Management System (NIMS)." 2004. <http://www.fireengineering.com/articles/2004/03/department-of-homeland-security-secretary-tom-ridge-approves-national-incident-management-system-nims.html>.
- Erk, Katrin, and Sebastian Padó. 2008. "A Structured Vector Space Model for Word Meaning in Context." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 897–906. Association for Computational Linguistics.
- Estellés-Arolas, Enriquet, and Fernando González-Ladrón-De-Guevara. 2012. "Towards an Integrated Crowdsourcing Definition." *Journal of Information Science* 38 (2): 189–200.

- Filliat, David. 2007. "A Visual Bag of Words Method for Interactive Qualitative Localization and Mapping." In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 3921–3926. IEEE.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4209698.
- Gabrilovich, Evgeniy, and Shaul Markovitch. 2007. "Computing Semantic Relatedness Using Wikipedia-Based Explicit Semantic Analysis." In *IJCAI*, 7:1606–1611.
- Gladwin, Hugh, Jeffrey K. Lazo, Betty Hearn Morrow, Walter Gillis Peacock, and Hugh E. Willoughby. 2007. "Social Science Research Needs for the Hurricane Forecast and Warning System." *Natural Hazards Review* 8 (3): 87–95.
- Goodchild, Michael F., and Linna Li. 2012. "Assuring the Quality of Volunteered Geographic Information." *Spatial Statistics* 1: 110–120.
- Gough, David. 2007. "Weight of Evidence: A Framework for the Appraisal of the Quality and Relevance of Evidence." *Research Papers in Education* 22 (2): 213–28.
<https://doi.org/10.1080/02671520701296189>.
- Grady, Catherine, and Matthew Lease. 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*, 172–179. CSLDAMT '10. Stroudsburg, PA, USA: Association for Computational Linguistics.
<http://dl.acm.org/citation.cfm?id=1866696.1866723>.
- Hofmann, Thomas. 1999. "Probabilistic Latent Semantic Analysis." In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 289–296. Morgan Kaufmann Publishers Inc.
- Horita, Flávio Eduardo Aoki, Lívia Castro Degrossi, Luiz Fernando Gomes de Assis, Alexander Zipf, and João Porto de Albuquerque. 2013. "The Use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: A Systematic Literature Review." <http://aisel.aisnet.org/amcis2013/eGovernment/GeneralPresentations/4/>.
- Houston, J. Brian, Joshua Hawthorne, Mildred F. Perreault, Eun Hae Park, Marlo Goldstein Hode, Michael R. Halliwell, Sarah E. Turner McGowen, et al. 2015. "Social Media and Disasters: A Functional Framework for Social Media Use in Disaster Planning, Response, and Research." *Disasters* 39 (1): 1–22. <https://doi.org/10.1111/disa.12092>.
- Hughes, Amanda L., and Leysia Palen. 2012. "The Evolving Role of the Public Information Officer: An Examination of Social Media in Emergency Management." *Journal of Homeland Security and Emergency Management* 9 (1).
- Hughes, Amanda L., Lise AA St Denis, Leysia Palen, and Kenneth M. Anderson. 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*, 1505–1514. ACM. <http://dl.acm.org/citation.cfm?id=2557227>.
- Indurkha, Nitin, and Fred J. Damerou. 2010. *Handbook of Natural Language Processing*. Chapman and Hall/CRC. <http://www.crcnetbase.com/doi/pdf/10.1201/9781420085938-f>.
- Kar, Bandana. 2015. "Citizen Science in Risk Communication in the Era of ICT." *Concurrency and Computation: Practice and Experience*.
<http://onlinelibrary.wiley.com/doi/10.1002/cpe.3705/full>.
- Kar, Bandana, and David Cochran. 2015. "Department of Homeland Security (DHS): Public Perceptions of Warning and Alert Messages." Department of Homeland Security.

- Karl, Thomas R. 2009. *Global Climate Change Impacts in the United States*. Cambridge University Press. <https://books.google.com/books?hl=en&lr=&id=UCg7inA-HksC&oi=fnd&pg=PA13&dq=Global+Climate+Change+Impacts+in+the+United+States+&ots=uYgbGdQQ-K&sig=jMrd-tKMmMXnoKS6RHdQIjDbsuw>.
- Karl, Thomas R., Jerry M. Melillo, Thomas C. Peterson, and others. 2009. "Global Climate Change Impacts in the United States." *Global Climate Change Impacts In The United States*. <https://www.ars.usda.gov/research/publications/publication/?seqNo115=244618>.
- Kryvasheyeu, Yury, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. 2016. "Rapid Assessment of Disaster Damage Using Social Media Activity." *Science Advances* 2 (3): e1500779.
- Lee, Ryong, Shoko Wakamiya, and Kazutoshi Sumiya. 2011. "Discovery of Unusual Regional Social Activities Using Geo-Tagged Microblogs." *World Wide Web* 14 (4): 321–349.
- Lukas, Jeff. 2014. "The Front Range Floods of September 2013- An Overview of the Weather, Hydrologic Impacts, and Climate Context." CIRES Western Water Assessment.
- Lundgren, Regina E., and Andrea H. McMakin. 2013. *Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=hImIVRZsPw0C&oi=fnd&pg=PA10&dq=Risk+communication+response+activities&ots=LLD9Vhuhhg&sig=lmUwGG_eEsAHJBOTr5C_7o-FVkm.
- Meek, Sam, Mike J. Jackson, and Didier G. Leibovici. 2014. "A Flexible Framework for Assessing the Quality of Crowdsourced Data." <http://repositori.uji.es/xmlui/handle/10234/98927>.
- Morgan, S. L., and C. G. Waring. 2004. "Guidance on Testing Data Reliability." Retrieved Jan 23: 2005.
- Morrow, Betty Hearn. 1999. "Identifying and Mapping Community Vulnerability." *Disasters* 23 (1): 1–18.
- National Research Council. 2010. *Adapting to the Impacts of Future Climate Change*. <https://www.nap.edu/catalog/12783/adapting-to-the-impacts-of-climate-change>.
- Ng, Kwan-Hoong, and Mei-Li Lean. 2012. "The Fukushima Nuclear Crisis Reemphasizes the Need for Improved Risk Communication and Better Use of Social Media." *Health Physics* 103 (3): 307–310. <https://doi.org/10.1097/HP.0b013e318257cfcb>.
- NOAA. 2013. "Denver/Boulder, CO." 2013. <http://www.weather.gov/bou/>.
- . 2014. "The Record Front Range and Eastern Colorado Floods of September 11-17, 2013."
- Oxford. 2017. "Reliability - Definition of Reliability in English | Oxford Dictionaries." Oxford Dictionaries | English. 2017. <https://en.oxforddictionaries.com/definition/reliability>.
- Porter, Nathaniel, Ashton M. Verdery, and S. Michael Gaddis. 2016. "Enhancing Big Data in the Social Sciences with Crowdsourcing: Data Augmentation Practices, Techniques, and Opportunities."
- Preis, Tobias, Helen Susannah Moat, Steven Bishop, Philip Treleaven, and H. Eugene Stanley. 2013. "Quantifying the Digital Traces of Hurricane Sandy on Flickr." *Sci. Rep* 3: 3141.
- SearchDataManagement. 2017. "What Is Data Quality? - Definition from WhatIs.com." SearchDataManagement. 2017. <http://searchdatamanagement.techtarget.com/definition/data-quality>.

- See, Linda, Steffen Fritz, and Jan de Leeuw. 2013. *The Rise of Collaborative Mapping: Trends and Future Directions*. Multidisciplinary Digital Publishing Institute.
<http://www.mdpi.com/2220-9964/2/4/955/htm>.
- St Denis, Lise Ann, Leysia Palen, and Kenneth M. Anderson. 2014. “Mastering Social Media: An Analysis of Jefferson County’s Communications during the 2013 Colorado Floods.” In *11th International ISCRAM Conference*, 737–746.
<http://www.iscram.org/legacy/ISCRAM2014/papers/p93.pdf>.
- Takahashi, Bruno, Edson C. Tandoc, and Christine Carmichael. 2015. “Communicating on Twitter during a Disaster: An Analysis of Tweets during Typhoon Haiyan in the Philippines.” *Computers in Human Behavior* 50 (Supplement C): 392–98.
<https://doi.org/10.1016/j.chb.2015.04.020>.
- Tirilly, Pierre, Vincent Claveau, and Patrick Gros. 2008. “Language Modeling for Bag-of-Visual Words Image Categorization.” In *Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval*, 249–258. ACM.
<http://dl.acm.org/citation.cfm?id=1386388>.
- Turney, Peter D., and Patrick Pantel. 2010. “From Frequency to Meaning: Vector Space Models of Semantics.” *Journal of Artificial Intelligence Research* 37: 141–188.
- Twitter Inc. 2017. “API Rate Limits — Twitter Developers.” 2017.
<https://dev.twitter.com/rest/public/rate-limiting>.
- UN, UN. 2013. “Humanitarian and Disaster Relief Assistance.” *Acessado Em* 2 (12): 2013.
- Vuurens, Jeroen BP, and Arjen P. de Vries. 2012. “Obtaining High-Quality Relevance Judgments Using Crowdsourcing.” *IEEE Internet Computing* 16 (5): 20–27.
- Wallach, Hanna M. 2006. “Topic Modeling: Beyond Bag-of-Words.” In *Proceedings of the 23rd International Conference on Machine Learning*, 977–984. ACM.
<http://dl.acm.org/citation.cfm?id=1143967>.
- Wang, Richard Y., and Diane M. Strong. 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers.” *Journal of Management Information Systems* 12 (4): 5–33.
- Wright Water Engineers, Inc. 2014. “Rainfall—Runoff Analysis for September 2013 Flood in the City of Boulder, Colorado.”
- Yates, Dave, and Scott Paquette. 2011. “Emergency Knowledge Management and Social Media Technologies: A Case Study of the 2010 Haitian Earthquake.” *International Journal of Information Management* 31 (1): 6–13.