Understanding Citizens' and Local Governments' Digital Communications during Natural Disasters: The Case of Snowstorms

Lingzi Hong
iSchool
University of Maryland
lzhong@umd.edu

Cheng Fu
Department of Geographical Sciences
University of Maryland
cfu@umd.edu

Paul Torrens
Deparment of Computer Science and Engineering
New York University
torrens@nyu.edu

Vanessa Frias-Martinez
iSchool
University of Maryland
vfrias@umd.edu

ABSTRACT

A growing number of citizens and local governments have embraced the use of Twitter to communicate during natural disasters. Studies have shown that online communications during disasters can be explained using crisis communication taxonomies. However, such taxonomies are broad and general, and offer little insight into the detailed content of the communications. In this paper, we propose a semi-automatic framework to extract and compare, in retrospect, the digital communication footprints of citizens and governments during disasters. These footprints, which characterize the topics discussed during a disaster at different spatio-temporal scales, are computed in an unsupervised manner using topic models, and manually labelled to identify specific issues affecting the population. The end objective is to offer detailed information about issues affecting citizens during natural disasters and to compare these against local governments' communications. We evaluate the framework using Twitter communications from 18 snowstorms (including two blizzards) on the US east coast.

CCS CONCEPTS

• **Information systems** → **Data analytics**; **Social networks**; *Geographic information systems*; • **Computing methodologies** → **Information extraction**; **Topic modeling**;

KEYWORDS

crisis communication, disaster analytics, topic models,spatio-temporal analysis

1 INTRODUCTION

Natural disasters such as snowstorms, floods or tornados affect millions of citizens every year [4, 36]. The use of social media during such events has been extensively studied in the literature, with a heavy focus on Twitter. Researchers have analyzed the use of Twitter as an information spreading tool during disasters [31]; as a platform to generate situational awareness for humanitarian response [34]; as a way for governments to connect and engage with citizens [37]; or to extract and classify *information nuggets* in an automatic manner [15].

A growing number of citizens and local government agencies recognize its importance and embrace the use of social media platforms such as Twitter during natural disasters [37]. Most of the work in understanding citizens' and governments' digital communications during natural disasters has focused on using taxonomies with pre-defined sets of crisis communication categories [26, 30]. For example, two of the pre-defined categories in crisis communications are direct experience and media sharing which reflect a common human behavior during natural disasters, that of sharing what is seen. Such categorization is highly useful to understand behaviors from a sociological perspective, but it does not provide insights into the types of issues being discussed.

In this paper, we present an approach to extract and compare, in retrospect, the communication footprints of citizens and local governments during natural disasters. Our objective is twofold: first, we are interested in gaining more insight into the specific types of issues citizens and local governments discuss through social media. We propose to do this by freely exploring the topics discussed online, rather than using a pre-determined taxonomy of categories. For these topics to be meaningful, we also need to be able to frame them within various spatio-temporal scales that characterize different time periods and regions affected by the disaster. Second, we are interested in comparing the official communications from local governments with the citizens' digital communications. Ultimately, although not the focus of this paper, such comparisons could be used to analyze whether local governments' social media strategies are covering the information that citizens seek online during disasters.

With that purpose in mind, we present a framework to extract, with limited human supervision, a set of insightful topics shared online by citizens or local governments at fine-grained spatio-temporal scales during natural disasters. Broadly speaking, the framework uses geolocated tweets collected during several occurrences of a given type of natural disaster to identify, in retrospect, the issues discussed online at two spatial scales: counties and roads; and during

WebSci'17, June 25-28, 2017, Troy, NY, USA.
 ACM ISBN 978-1-4503-4896-6/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3091478.3091502

three periods: before, during and after the disaster took place. Since the collection of geolocated tweets can include tweets potentially not related to the disaster, the framework uses hashtagged tweets, collected during the same natural disaster, as a gold standard of what disaster-related tweets look like, and ranks the geolocated tweets based on their similarity with respect to the gold standard. The similarity is computed using a language-based technique known as *ranking by cross-entropy* [3] that ranks each tweet by its similarity to the n-gram probabilistic distribution of the gold standard. The final subset of disaster-related tweets is selected from the ranked pool as the set of tweets with the largest topic coherence and interpretability when topics are extracted using Single-topic Latent Dirichlet Allocation (ST-LDA)[5, 13], a fully unsupervised topic modeling technique that creates probabilistic associations of words to topics and assigns a unique topic per tweet.

The resulting topics are manually labelled by annotators and the framework computes the *digital communication footprints* that characterize the types and distribution of topics for local governments and citizens, using only tweets with high topic membership probabilities. Finally, the framework analyzes the footprints observed in citizens' and local governments' digital communications and compares similarities and differences across space and time.

The proposed framework advances the state of the art in several fronts: (1) it extracts and compares, in a semi-automatic manner, digital communication footprints that characterize the types of topics local governments and citizens discuss online at different spatio-temporal scales. The objective is to extract and compare, in retrospect, detailed topics rather than general, pre-defined categories that albeit useful from a sociological point of view, do not provide insights into the specific types of issues discussed online [30]; (2) it requires a small amount of manual work carried out by annotators labeling the topics extracted by ST-LDA in a fully unsupervised manner; compared to previous automatic approaches of tweet classification during disasters that require massive labelings of individual tweets to train the classifiers [26]; and (3) it presents a novel approach to select tweets relevant to a natural disaster using a *ranking by cross-entropy* technique combined with topic models that eliminates the need to manually pre-define a set of keywords to identify tweets related to disasters [25]. We will also show that this technique provides a ranking that results in more cohesive and informative topics than using keyword-based rankings.

2 RELATED WORK

Social media, with a heavy focus on Twitter, has been extensively used in the past to retrieve information being exchanged during natural disasters [1, 31, 34]. Some approaches focus on understanding communication patterns via the analysis of tweeting behaviors [7, 11]; on studying the types of information being exchanged [25, 32]; or on understanding whether Twitter can be used as a proxy for damage assessment during disasters [18]. A complementary line of research has also analyzed official and media responses to natural disasters, and how these might differ with respect to citizens' responses across various types of disasters [6, 26]. The approach presented in this paper, is based on the fact that all the previous findings show that Twitter, with all of its limitations, can be used as a tool to understand communications during natural disasters, both in terms of behaviors and content being exchanged. Unlike previous comparative analyses that focus on the use of pre-defined taxonomies to characterize communications during disasters, our approach presents a framework that extracts detailed information using topic modeling techniques that allow to freely explore, in retrospect, the topics discussed online at various spatio-temporal scales [17].

From an analytical point of view, most of the work in this area has either focused on collecting tweets via hashtags and analyzing communication patterns and content during disasters [8, 29, 33]; on collecting geolocated tweets and/or tweets with disaster-related keywords and developing automatic classifiers to extract the tweets that are talking about the disaster [2, 15, 16, 25] or on a combination of both [19]. In this paper, we present a solution that uses both hashtagged and geolocated tweets, but that unlike previous approaches does not require the labeling of massive amounts of tweets to train automatic tweet classifiers. We achieve this by using unsupervised topic models and manual topic labeling to both extract disaster-related tweets and to analyze the types of topics discussed online during disasters.

3 THE FRAMEWORK

In this section we present the framework to extract digital communication footprints from the twitter activity of local governments and citizens during natural disasters, and to analyze their similarities and differences. The framework has four sequential components (see Figure 1): (1) data preparation, to clean and prepare the tweet datasets from local governments and citizens; (2) tweet ranking and selection, to determine the geolocated tweets that could be potentially related to the natural disaster; (3) topic identification, to extract and annotate the topics communicated online by local governments and by citizens; and (4) computation of digital communication footprints, to characterize and compare the distribution of topics discussed online by local governments and citizens at two spatial scales: counties and roads; and during three time periods: before, during and after the natural disaster.

In the next sections, we describe each component in detail and evaluate the framework for a specific type of natural disaster: snowstorms. The underlying hazards of snowstorms are varied, including vehicle accidents due to icy conditions on roads or power outages, and disrupt the lives of citizens even when their cities are, in principle, well prepared to cope with them [28]. We evaluate the framework on the snowstorms that took place across 24 counties in the state of Maryland for two winters from November, 2014 till April, 2016. This period was very harsh with 18 snowstorms including two blizzards. We define as snowstorms periods of one or more consecutive days when the snowfall was disruptive *i.e.,* when the snow depth provided by the National Oceanic and Atmospheric Administration was of at least $2in$ [24]. Although the evaluation focuses on snowstorms, the framework can be used to extract and compare, in retrospect, digital communication footprints from local governments and citizens for any type of acute event were both local agencies and the population participate and as long as it has one or multiple hashtags associated to it.

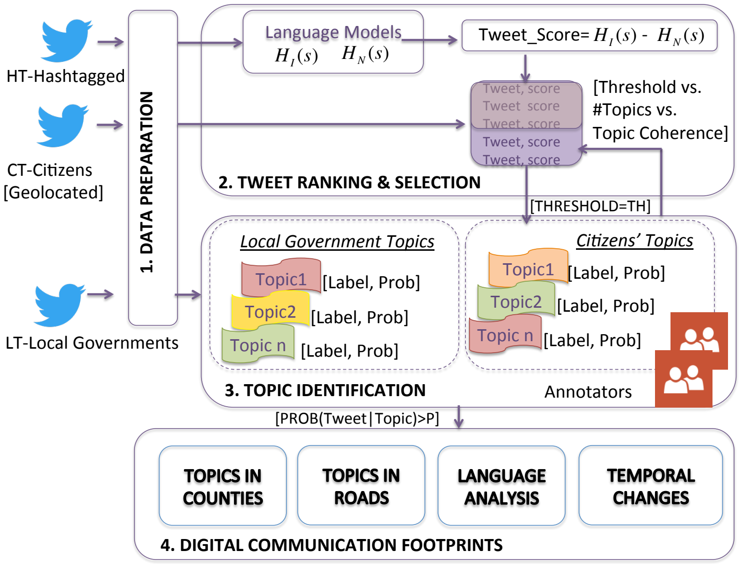

Figure 1: Overview of the proposed Framework.

4 DATA PREPARATION

The framework uses three data sources: geolocated tweets from citizens (CT), hashtagged tweets (HT), and tweets from local government agencies (LT). All these data sources are collected using Twitter's streaming API. The framework uses the geolocated tweets (CT) to extract the digital communication footprints of the citizens affected by the snowstorms at two spatial scales: counties and roads; and three time periods: before, during and after the natural disaster. CT was collected during the 2014-2016 winter seasons for the 24 counties in the state of Maryland using a bounding box over the whole state. Since the bounding box covered some areas outside of the limits of the state, we used the state's shapefile [9] to filter out tweets outside of its geographical boundary. We collected a total of $723,033$ geolocated tweets. However, since we are only interested in tweets published during disruptive snowstorms, we only kept tweets from CT that were posted on days when there was snow and its depth was of, at least, $2in$. To analyze the digital communications before and after the snowstorms took place, we also kept CT tweets published two days before and two days after each snowstorm. Finally, since we are interested in modeling the online voice of *average* citizens during the snowstorms, it is important to eliminate from CT all accounts potentially representing organizations or spammers. For that purpose, we retrieved the number of followers and followees for each twitter account in CT and filtered out accounts whose: (1) number of followers and followees were two standard deviations larger than the average values [38]; and (2) whose ratio of followees over followers was smaller than two standard deviations from the average value of the dataset [39]. The final size of CT after applying the described filters was of $169,987$ tweets.

The framework also requires all the hashtagged tweets (HT) associated to the area affected by the natural disaster under study. This second dataset is used in the Tweet Ranking and Selection component of the framework as a gold standard of the language employed and the topics discussed during the natural disaster. To collect HT during the snowstorms, we manually curated a list of 12 hashtags that were used during the 2014-2016 snowstorms in Maryland (MD) and collected all the tweets that had at least one of the hashtags in the text of the tweet. Some examples of the collected hashtags include #mdblizzard or #mdsnow. We did not collect more generic hashtags (*e.g., #blizzard*) since they could represent topics and language from other regions. The final HT dataset contained a total of $96,423$ hashtagged tweets.

The framework also requires a dataset with the tweets generated by local government agencies (LT) in the area where the natural disaster takes place. To carry out this data collection for the snowstorms, we computed a curated list of 83 Twitter handles representing local government agencies from the 24 counties in the state of Maryland; and collected all their tweets for the 2014-2016 winter seasons. Similarly to CT, we only kept LT tweets that were published either two days before, during, or two days after a snowstorm with a depth of at least $2in$. The final size of LT was $18,688$ tweets.

Local government tweets are typically not geolocated. However, we need to assign a location to be able to compare the digital communication footprints of local governments and citizens at two spatial scales: counties and roads. As a result, an official tweet will be assigned a location based on the location of the agency itself. For example, if the local agency is the Allegany County Department of Emergency Services (@Allegany911), the location assigned to the tweets generated by this account will be Allegany county. On the other hand, the tweets created by local transportation agencies are associated to roads. For example, tweets by Montgomery County's Highway Service (@MontCo_Highways) would be used to compute the digital footprints on roads.

Finally, we clean the text of all the collected tweets as follows: we eliminate stop words, emoticons, and tweets with less than 4 words; we maintain retweets because they are indicative of individuals agreeing with a message; we replace specific numbers and road names with the generic words *number* and *roadname* since these are frequent in our tweets but hard to model due to the different types of addresses, numbers and roads involved; we replace links with the word *link* to keep track of media being shared; and we lemmatize all words using NLTK's WordNet [21]. As opposed to stemming, lemmatization (which reduces words to lemmas, or dictionary forms) has been frequently used with topic models providing good results [35]. The final size of the CT dataset after this cleaning was $163,019$ tweets since there were $\approx$ 6000 tweets with less than four words. HT and LT kept the same number of tweets. It is important to clarify that, by exclusively using geolocated tweets from citizens to compare against local governments' tweets, we acknowledge that we are only able to capture the voice of those who have geolocation enabled on their Twitter accounts. Although that is a limitation, it is the unique way to extract local information that is meaningful and actionable at the granular spatial scales we are interested in studying. Other approaches with larger, non-geolocated, datasets could provide potentially additional information; however, it would come at the cost of loosing the local scale, which is the focus of this research.

5 TWEET RANKING AND SELECTION

CT contains tweets that were published by citizens during snowstorms. However, it is highly probable that not all the collected tweets are talking about the snowstorms. This framework component focuses on: (1) the ranking of the CT tweets with respect to their similarity to snowstorm-related tweets; and (2) the computation of the threshold to select the final subset of ranked CT tweets that potentially talk about snowstorms (see Tweet Ranking in Figure 1). This component is exclusively executed on CT since local governments' tweets during a disaster (LT) typically focus on disaster-related topics [14].

The framework performs the ranking process using the hashtagged tweets (HT) as a gold standard of what snowstorm-related tweets might look like. By comparing the CT tweets against the ground truth, the framework can assign a similarity score to each geolocated tweet and rank them. Specifically, the framework uses a method called *ranking by cross-entropy* [3]. This method uses all the hashtagged tweets (HT) to compute an *in-domain* language model that characterizes the frequency of different n-grams in tweets that talk about snowstorm-related topics. HT tweets are considered to be *in-domain* as opposed to more generic tweets that can talk or not about snowstorms, which the method will refer to as *general-domain* tweets. The method also computes the language model for the *general-domain* CT tweets, that can talk about any topic, calculating the frequency distribution of n-grams across all the CT tweets. These two language models are used to compute, for each CT tweet, the difference between the tweet's perplexity using the in-domain model and the tweet's perplexity using the general-domain model. The idea behind this scoring is that tweets that are highly similar to in-domain (snowstorm) tweets and highly dissimilar to the general-domain tweets will be ranked highly. Formally, given a string s with empirical n-gram distribution p in a language model q, we calculate its perplexity as: $P(s) = 2^{H(s)}$ where $H(s)$ is the cross-entropy of the string s computed as $H(s) = -\sum_x p(x) \log q(x)$. This process results in a list of ranked CT tweets based on their perplexity score, from smaller (most similar) to larger (most different) values. The framework uses the SRI Language Modeling Toolkit with one to four n-grams to compute the language models [22]. To account for the sparsity and the fact that many words are not present in our datasets, we use the interpolated modified Kneser-Ney discount when building the language models.

Next, the framework performs the computation of the threshold that determines the percentage of tweets from the ranked pool to be selected as snowstorm-related tweets. Since the final goal of the framework is to extract the topics discussed online during snowstorms, we use the quality of the topics obtained as a measure of how good the selection of in-domain, snowstorm-related tweets is. The underlying assumption is that if the CT tweets selected are snowstorm-related tweets, then the quality of the topics extracted by the topic modeling technique will be high; while if the selected CT tweets are mostly noisy *i.e.,* not related to snowstorms, then the quality of the topics would be low. To determine the best threshold value, the framework uses the PMI score (point-wise mutual information) as a measure of the quality and interpretability of the topics extracted using ST-LDA over the CT tweets [5, 13]. This measure has been showed to best model topic coherence and interpretability

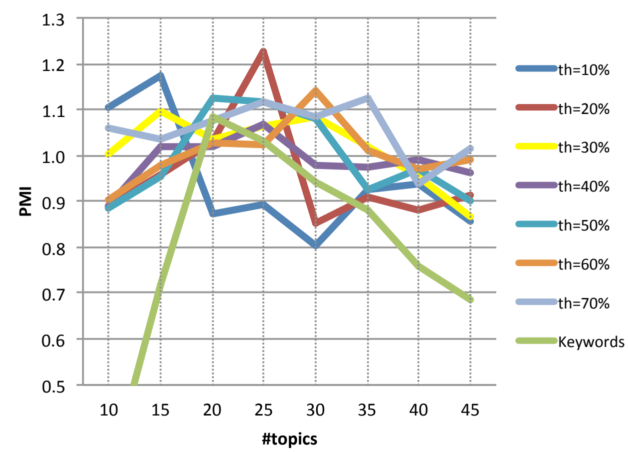

Figure 2: PMI values for various threshold (*th*) and number of topics (*K*) combinations; and for the keywords approach.

when compared to other topic scoring methods [23]. Specifically, the method computes the PMI score for various combinations of threshold values (th) and numbers of topics (K) and selects the final th and K based on the largest PMI achieved. Formally, the framework computes the PMI for each topic and averages across all topics. The PMI for a topic is computed as $PMI(x, y) = \log \frac{p(y,x)}{p(y)}$ where x and y are word n-grams from the top t most frequent words for that topic, and the PMI measures the probability of co-occurrence of those two words. Details about topic identification with ST-LDA are described in the next component of the framework.

Figure 2 shows the results of the ranking process and threshold-based tweet selection using the CT and HT datasets containing snowstorm tweets across the 24 counties in the state of Maryland. Each data point represents the PMI for a given combination of threshold and number of topics, and is computed as the average PMI across all the topics when the top 10 1-gram words are considered. We show results for $K = 10$ to $K = 45$ different topics and threshold values from 10% to 70% (other values perform worse and are thus not shown). We observe that the best values for our dataset were 25 topics and a threshold of 20% *i.e.,* selecting the top 20% of the ranked tweets as being snowstorm related and extracting 25 topics provides the most coherent and interpretable set of topics discussed online during snowstorms.

Computing the two language models for the ranking process is computationally expensive. One could argue that a keyword-based approach would be a much more efficient way to select the subset of CT tweets related to snowstorms. To assess this hypothesis, we compare the performance of both approaches as follows. We collected snowstorm-related keywords using two sources: (1) words provided by five individuals who were asked to create a list of words they would use on Twitter to talk about their experiences during the 2014-2016 snowstorms (filtering out repeated words); and (2) a list of 380 disaster-related words from the dataset published in [25]. The total set of keywords collected was of 419. We used this set of keywords to filter out the CT tweets that did not contain any of these words in their text; and computed the PMI for the same sets of topics (K values) reported in Figure 2. As the green line in the Figure shows, the keyword-based PMI values are worse than the proposed ranking-selection approach for most of the thresholds explored. The best result for the keyword-based approach (with $K = 20$) gives a $PMI = 1.08$, which is much worse than the best result for the proposed ranking approach with a $PMI = 1.23$ (with $K = 25$ and $th = 20\%$). This result shows that our ranking and selection approach helps in identifying a subset of disaster-related tweets that result in more coherent and interpretable topics than when using tweets exclusively selected via keywords. Given that the keyword list we used is quite comprehensive, we argue that this finding could be related to the fact that Twitter language is very informal and words are prone to polysemy (*e.g.,* Jonas: the name of a snowstorm and a music band); as well as to slang (*e.g.,* crash: fall asleep or accident). Using such informal language as keywords, we could be labeling as snowstorm-related tweets, tweets that are not in reality.

6 TOPIC IDENTIFICATION

This component focuses on identifying the topics discussed online by local governments and citizens. The framework runs two steps: (1) topic extraction and (2) topic labeling. The topic extraction is executed using Single-topic Latent Dirichlet Allocation (ST-LDA), a variant of the topic modeling technique LDA. Similarly to LDA, it considers each tweet as a document and clusters them automatically into latent topics based on word co-occurrence. However, acknowledging the fact that the length of a tweet does not allow for a lot of issues to be discussed, it exclusively assigns a unique topic per tweet [13]. ST-LDA also computes, for each tweet, the topic membership probability, which can be used as a measure of how representative of a topic a given tweet is. Two separate topic models are run, one for the local governments' tweets (LT) and another one for the citizens' tweets (CT). The reason for this is that the number of tweets by local governments is typically much smaller than the volume of citizen tweets, and as a result, local government's topics could be hindered by the citizens' ones if a unique ST-LDA is used [10]. The models are run for various number of topics K, and the final value is selected based on the coherence and interpretability of the topics, which is measured by the PMI over the top ten most frequent words in each topic.

The topic extraction for the citizens' tweets is incorporated in the Tweet Ranking and Selection component described in the previous section. Recall that the topic extraction is used to select the snowstorm-related tweets from the CT pool, and that the most coherent set of topics was obtained for $K = 25$. On the other hand, the topics discussed by local governments are extracted applying ST-LDA over all the tweets in LT; which, using the best PMI value, resulted in $K = 45$ deemed as the most coherent and interpretable set of topics.

Once local governments' and citizens' topics are extracted, these need to be labeled to identify the types of issues discussed online during disasters. For that purpose, we recruited four annotators who were given access to the top ten words per topic and, for completeness, to the full set of words and their topic membership probabilities, prior lemmatization. Each topic had on average 96 words. Each annotator was asked to assign a label (of maximum five words) to each of the 25 citizen topics and 45 local government topics. If they believed a pair of citizen and local government topics were similar, they were required to provide the same labels. It is important to highlight that annotators were trained to understand the task, but were not offered any guidance in terms of labels or topics they should identify. From that perspective, the labeling was completely unsupervised and bottom-up, and did not include any bias as of what types of topics should be found in the online communications. However, that freedom implied that annotators could be using different words to refer to the same topic *e.g.,* "complaints school delays" versus "venting school closed". To bridge those differences, we designed a protocol by which once all labeling was done, annotators met and discussed their labels in person so as to evaluate whether they were referring to the same topics or not, and achieve a final agreement over the names of the labels assigned (categories). Due to the nature of this protocol, we did not use crowdsourced platforms for the labeling, since we required annotators to be able to meet in person. Finally, we computed the inter-annotator agreement using Fleiss' Kappa since we had multiple annotators [12]. The test gave a strong inter-rater reliability of $\kappa = 0.78$.

Table 1 shows the topics identified by the annotators. For discussion purposes, we have grouped some of the topics into themes that bring together issues that were highly related. For example, the topics of "traffic accidents" and "driving conditions" were grouped together in the "Traffic" theme that includes two topics: accidents and driving conditions. Topics that were not grouped have the same theme and topic name in the Table. Overall, there were seven general themes identified namely, economic and social factors, government offices and schools, traffic, public transportation, preparedness, response and weather information. The table shows whether the topic was exclusive to citizens $[CT]$, to local governments $[LT]$ or shared by both $[CT, LT]$; and it also displays, for each theme and topic, two top ranked tweets based on their topic membership probability. Next, we describe each theme in detail. The discussion about how each topic is used by local governments or citizens is covered in the next section.

The *economic and social factors* topic is mostly represented by words reflecting thoughts and opinions around the impact of snowstorms on citizens' daily routines using terms related to work, money and access to food stores and supplies. This topic, which is exclusive to citizens, had top ranked tweets expressing opinions about the impact of snowstorms on the local economy *e.g.,* companies forcing their employees to go to work under harsh weather conditions; about citizens having a hard time finding food supplies in stores; or about citizens spending the snowstorms working from home, or eating and drinking with family and friends.

The second general theme identified was *government offices and schools*. This theme covered two topics: (1) information about delays and closures, and (2) citizen complaints about official decisions with respect to offices' or schools' schedule changes. The words in the first topic were related to updates and schedules including times and places such as elementary schools, colleges or government offices. On the other hand, the second topic was represented by complaint, decision and choice words, verb negations as well as by local government agency names. Some of the top tweets in this topic show citizen tweets asking their local governments to reconsider

Table 1: Themes and topics for citizens' tweets [CT], local governments' tweets [LT] or both [CT,LT] and sampled top ranked tweets.

Themes	Topics	Top Tweets
Economic and Social Factors	Economic and Social Factors [CT]	a. Sears has no love for their employees making me work in this bullshit weather b. [...] If you are on verge of suicide, go grocery shopping at Walmart before [...] the snow
Gov.Offices and Schools	Delays and Closures [CT,LT]	a. UPDATE: Anne Arundel County Government Offices Will Be closed Today b. Due to the inclement weather, all BCPSchools [...] will be closed Friday [...]
	Complaints [CT]	a. Mcps is the type of county that would still give school even with this weather on Tuesday b. Howard it's too cold, kids can't go outside in this, 2 hour delay please
Public Transportation	Delays and Closures [CT,LT]	a. MARC Train service is closed Thursday snow is expected [...][link] b. RTA bus service remains suspended & will not operate 1/25 [...] for service updates visit [link]
	Private Options [CT]	a. If anyone wants to make $1 billion, be an Uber driver in Baltimore [...] 4x surge pricing b. How did I get an Uber_Maryland today without Surge pricing? [...] Thank you Uber
Traffic	Accidents [CT,LT]	a. Accident in #BaltimoreCity on I83 NB between Northern Pkwy and Ruxton Rd b. Disabled snow plow in #GlenBurnie on Baltimore Annapolis Blvd near Marley Neck Rd
	Driving Conditions [CT,LT]	a. Snow on roadway on I-695 Outer Loop at X16 b. Icy conditions in #Bethesda on The Beltway Inner Loop at I-270 Spur and before Connecticut Ave
Preparedness	Preparedness [LT]	a. Not sure what you should have in your emergency kit? Free checklist from APHA's b. Protect your pipes against freezing by letting a thin stream of water flow through faucet
Response	Snow Removal [CT,LT]	a. 98% of County roads plowed at least once. Constant work continues to the 2,600 miles [...] b. Thank you MCPS for not plowing your parking lots
	Emergency Services [LT]	a. MCFRS Emergency Communications on 'CONDITION RED' due to heavy call volume [...] b. For an emergency during the #WinterStorm call 911. For non-emergencies call 410-887-2222 [...]
Weather Info	Weather Info [CT,LT]	a. Snow totals for AACO are up. Now in the 8-10 inch range b. Old Village and Mechanicsville Road (picture was taken around 8:30 a.m.) this morning

delay or closure decisions given the harsh weather conditions (see Table 1).

The third theme identified was *public transportation*. This theme grouped two topics, delay and closure information and private transportation options. The first topic focuses on words explaining changes in schedules or closures in trains, metro, light rail or buses. Table 1 shows an example from Baltimore's Department of Transportation announcing the closure of several services due to snow. The second topic in this theme, exclusive to citizens, was represented by words related to private transportation alternatives and its implications; including words such as Uber, Lyft, service, or price surges during snowstorms.

Another theme identified was *traffic*. Our annotators labeled two distinct topics in this domain: road accidents and driving conditions. The topic of road accidents included words related to cars, accidents and road numbers and names; while the driving conditions topic gathered words mostly around speed, ice, snow, and road names and numbers. Examples of some of the top ranked tweets are showed in the table. The topic of *preparedness*, exclusive to local governments, mostly reflected advisory and safety words about how to cope with harsh weather conditions. Some top tweets offered information about emergency kits or about how to protect home pipes under freezing conditions.

The theme of *response* included two topics: snow removal and emergency services; the latter exclusive to local governments. The first topic was represented by snow plowing related words, announcement and update words, locations and times. Table 1 shows a tweet from Baltimore County explaining that most of its roads have been plowed. As for emergency services, most words were related to emergency-specific phone numbers and their status. Finally, the last topic identified, *weather information*, had words mostly related to weather conditions including snow depth, temperatures and links to pictures showing those conditions during the snowstorms.

7 DIGITAL COMMUNICATION FOOTPRINTS

Once topics have been identified, the framework focuses on the computation of the digital communication footprints during disasters for local governments and citizens at various spatio-temporal scales. Specifically, the framework performs four retrospective analyses: (1) comparison of theme and topic usage by local governments and citizens in counties, (2) comparison of theme and topic usage by local governments and citizens in roads, (3) language used in citizens' communications, and (4) changes in local governments' and citizens' themes and topics before, during and after snowstorms.

For the spatial analysis, the footprints are computed by counting the number of tweets per topic identified for each of the 24 counties in Maryland and for the roads in the state. The objective is to understand the topic distribution for a given spatial scale; and to compare across local governments and citizens. To carry out the analysis, the framework assigns each geolocated tweet in CT to the county whose geographical boundaries contain the geolocation of the tweet; and to the road whose boundaries, with a $2m$ buffer, include the location of the tweet. TIGER/Line shapefiles are used to extract the exact boundary information [9] for the 24 counties and for five types of roads: interstates, freeways, principal arterials, minor arterials and collector streets. As explained earlier, local government tweets in LT are not geolocated and as a result are assigned to the county of

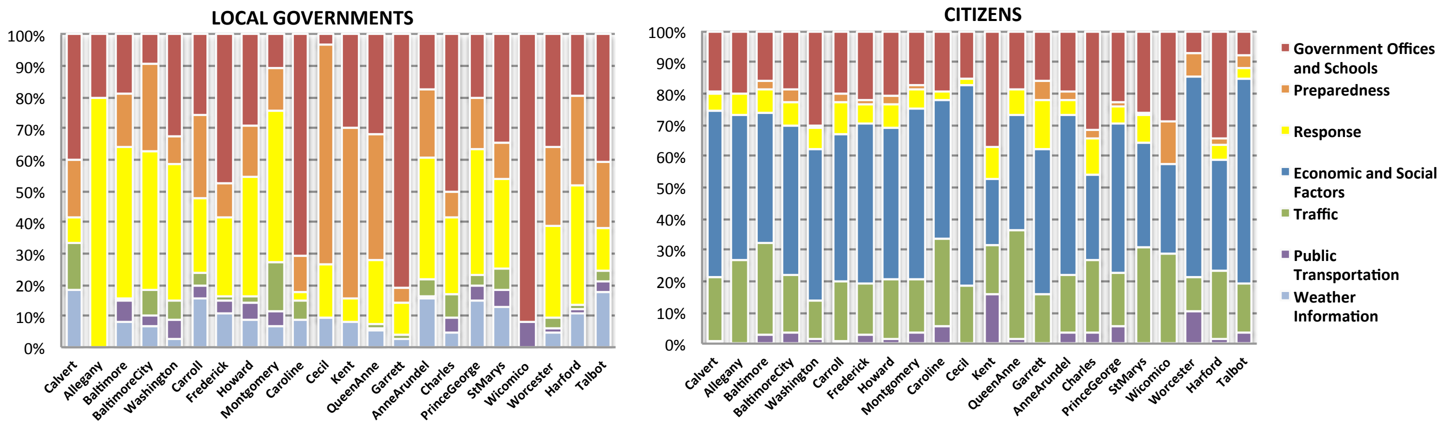

Figure 3: Digital Communication Footprints of Local Governments and Citizens for each county in the state of Maryland.

the agency represented by the twitter handle; or to roads (without a specific type) if the twitter account is from a local transportation agency. We also present a language analysis to further understand the thought processes behind the topics discussed online by citizens. Finally, we discuss the results of the temporal analysis, focused on exploring differences in topic usage before, during and after disasters take place.

Rather than using all the tweets in CT and LT to compute the digital communication footprints, the framework only uses tweets whose topic membership probability is sufficiently large to guarantee representativity for that topic. This approach eliminates from the footprints tweets that have been assigned to a topic with very little certainty. Specifically, the framework requires the topic membership probability assigned by ST-LDA to each tweet to be of at least $p > .75$ to be counted as a communication with that topic label. Such p value guarantees a maximum error of $e = 1 - p \leq .25$ in tweet to topic associations. That error value is a lower bound, since the average topic membership probability for tweets whose $p > .75$ in CT and LT is 0.96 ($e \leq 0.04$), which is comparable to error rates achieved when identifying topics using classifiers trained with labeled tweets [25]. After applying the probabilistic filter, the total number of tweets in counties was of $148{,}103$. To guarantee analytical significance, we eliminated from the analysis counties with less than 50 tweets throughout all the snowstorms, which excluded two counties: Somerset and Worcester. As for roads, we collected 7904 tweets from both CT and LT. CT tweets covered 31% of all road segments in the state of Maryland, almost evenly distributed across all five types of roads. Next, we describe each analysis in detail.

7.1 Themes and Topics in Counties

Figure 3 shows the digital communication footprints of local governments and citizens for all the counties in Maryland. The counties on the x-axis are ordered by the median amount of snow throughout the 18 snowstorms in our dataset. Each bar in the histogram represents the digital communication footprint for a county, with the percentage of the overall communication (y-axis) associated to each theme.

An important first observation is that the types of themes discussed by citizens and local governments are largely different; and that the amount of snow does not appear to play a role in the theme choice. Local governments tend to focus their online communications on three themes: Government Offices and Schools, Response and Preparedness, representing an average of 30.1%, 26.6% and 16.9% of their total communications, respectively. Hierarchical clustering of the local government county footprints grouped together urban counties such as Montgomery or Prince George, characterized by higher response communication volumes than their rural counterparts (using the 2013 urban-rural classification scheme by the NCHS/CDC); which might be indicative of urban counties being more informative about their decisions to cope with snowstorms. On the other hand, citizens mostly use Twitter to talk about Economic and Social factors (47.6%), Government Offices and Schools (19.8%), and Traffic (18.8%), with no clear clusters differentiating urban and rural citizen communications. These findings suggest that while local governments mostly use Twitter to prepare the community for the snowstorms and to inform them about how they are responding; citizens use Twitter as both a social network: discussing and sharing thoughts about the economic and social impact of the snowstorms, and as a news platform: sharing news about closures, delays or driving conditions [20].

Next, we look in depth into the use of topics within the themes shared by local governments and citizens in their communications. Figure 4 represents the difference between official and citizen footprints for the topics in the Government Offices and Schools theme. The differences are computed subtracting from the local governments' percentage of tweets per topic, the percentage of citizen tweets for that same topic. Positive values imply that local governments dominate the topic while negative values represent citizens talking more about that topic. Situations were the topic is exclusive to either local governments or citizens are clarified in the discussion. Figure 4 shows that while for most of the counties the topic of delays and closures is dominated by local government agencies, possibly making official announcements (with average 23% more communications than citizens); the topic of complaints is exclusive

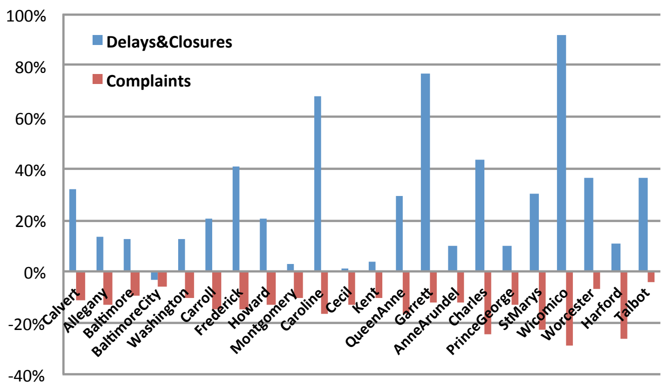

Figure 4: Government Offices and Schools: topic differences between local government and citizen communications.

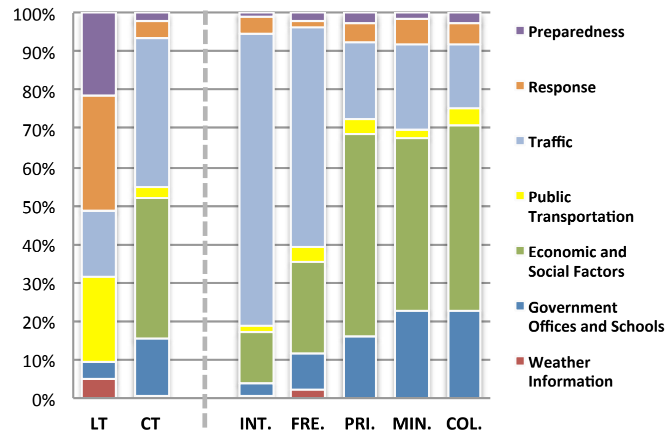

Figure 5: Distribution of themes and topics on roads. Two bars on the left show the average footprint across all types of roads; five bars to the right show footprints per type of road: interstate, freeway, principal and minor arterials and collector

to citizens demanding their local governments different scheduling decisions (with average communication volumes of 12%).

Similar analyses showed that for the Traffic theme, accident reporting online is done almost equally by citizens and local governments for most of the counties. However, citizens dominate the conversation with respect to reporting driving conditions, with an average of 16% more communications than their local governments. The topic of delays and closures in the Public Transportation theme has local governments talking slightly more about it than citizens, possibly making official announcements (4% higher volumes on average); while citizens were the only ones to talk about the private options topic, with a conversation around Lyft and Uber when public transportation was not an option. In the Response theme, the topic of emergency services is exclusive to local governments who share information about how to use these types of services during snowstorms; while local governments also lead communications about snow removal in their communities (7% larger). Finally, the weather information topic was common for both local governments and citizens, with only 1% more official communications.

7.2 Themes and Topics in Roads

Figure 5 shows the digital communication footprints for local governments and citizens on roads. The first two bars on the left, represent the average footprint across all types of roads for local government (*LT*) and citizen (*CT*) communications. We observe that, similarly to the county communication footprints, local government communications focus on preparedness and response, possibly offering pre-snowstorm advice and updates about snow removal strategies. However, we also observe that the official discourse on roads puts a lot of weight on communications about traffic and public transportation, with tweets spreading details about accidents or driving conditions; and offering information about delays and closures of public transportation services. On the other hand, citizen communications on roads still focus on economic and social factors (35%), and government offices and schools (14%). However, citizen traffic-related communications have almost doubled with respect to their county communications, from 18% to 35%. Looking in depth into the citizens' communication footprints for each type of road (Figure 5, five bars to the right), we observe that the distribution of topics changes from larger roads (interstates, *INT.* or freeways *FRE.*) to smaller roads (principal and minor arterials *PRI.,MIN.* or collectors *COL.*). Interestingly, while citizens mostly focus on sharing accident information and driving conditions with others when they are on large interstates or freeways (65% average communications on the traffic theme), the discourse changes towards economic and social factors (48%) when we move to smaller arterial and collector roads. This reflects that citizens' communications during snowstorms change depending on the physical spaces they occupy. Driving in large roads during snowstorms encourages a utilitarian approach with drivers sharing useful information (Twitter as a news platform) while smaller roads bring them back to opinion and thought sharing (Twitter as a social network).

7.3 Language Analysis

Understanding the types of topics discussed online by citizens is highly relevant for local governments wishing to serve their populations during snowstorms. However, since topics simply provide lists of words, it would be important for local governments to gain more insights into the potential emotions and thought processes behind the citizen communications shared online. To carry out this analysis, the framework uses the psycholinguistic lexicon LIWC, which has been extensively validated in determining mental states for different types of text sources [27]). Specifically, the framework puts in a document all the citizens' tweets for each theme and topic identified (across counties and roads) and analyzes the following LIWC dimensions per topic: positive and negative emotion, anxiety and anger from *affect words*; insight, cause, discrepancies and tentativeness from *cognitive processes*; seeing and hearing from *perpetual processes*; and work, leisure, home and money from *personal concerns*.

Figure 6 shows a radar chart with the LIWC dimension values for the most popular citizen themes and topics. Each line in the chart represents the distribution of the values for a given topic. The line for economic and social factors shows that there is a heavy use of

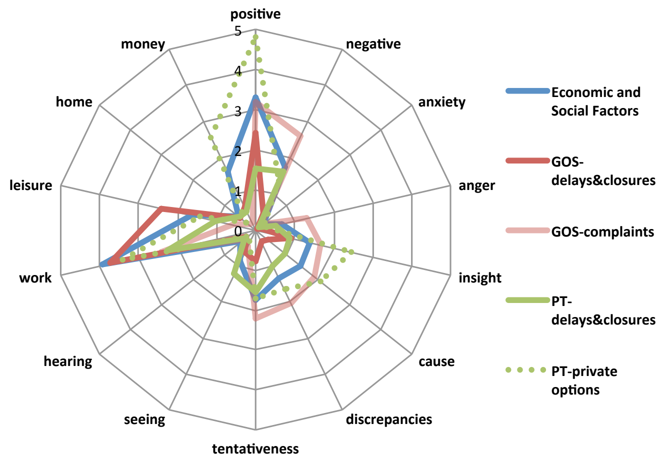

Figure 6: LIWC dimension values for citizen topics.

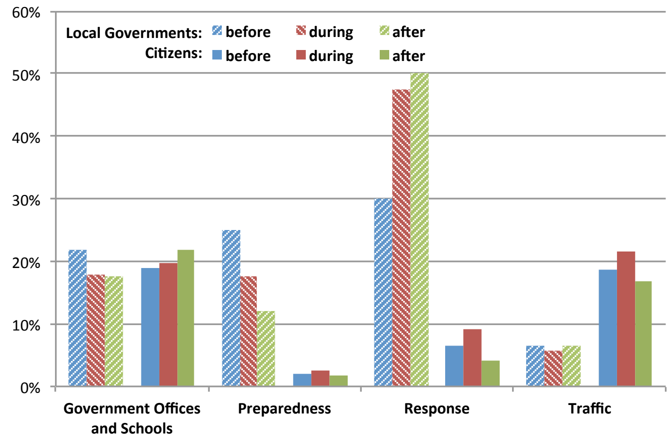

Figure 7: Percentage of tweets per topic and time period for local governments (texture bars) and citizens (solid bars).

work, money and leisure-related words (with values of 3.9, 1.8 and 1.6, respectively), with a mix of positive and negative emotions, and a heavy use of words related to tentativeness, discrepancy, causal effects and insights. This reflects very well the nature of the topic, with individuals sharing strong positive and negative opinions about the impact of the snowstorms on their daily routines *e.g.,* staying at home with the family versus going to work under harsh weather conditions.

The line for government offices and schools (*GOS*) shows that communications about delays and closures (*GOS-delays&closures*) are mostly represented by work and leisure related words mixed with positive emotions (3.6, 2.4 and 2.5, respectively), implying that, in general, the population accepts the closures as something positive and leisurely *e.g., staying at home*. On the other hand, the topic about complaints with respect to closures (*GOS-complaints*) shows a heavy use of angry, tentative, causal and discrepancy words probably reflecting anger, confusion and uncertainty about how to plan a day when schools and offices are closed. Public transportation topics (*PT-delays&closures* and *PT-private options*) also show a peak in the use of work related words, probably reflecting people's concerns about reaching work on time; additionally, a very high peak (4.9 and 2.5) for positive emotion and insight words is observed when talking about private transportation alternatives, which might be indicative of the fact that people are very happy to have other transportation options when public transportation is not fully working. Finally, the traffic topics (not shown in graph for clarity), reflected a heavy use of work-related and negative emotion words (3.5 and 3.2) when citizens were reporting driving conditions; while causal and negative words were mostly present when reporting accidents (2.1 and 2.3).

7.4 Temporal Changes

In this component, the framework evaluates whether the communication patterns of local governments and citizens change over time as the snowstorm evolves. Specifically, it computes the percentage of tweets per topic communicated by local governments and citizens in three distinct time periods: two days before a snowstorm, snowstorm days and two days after a snowstorm. Figure 7 shows the percentage of local governments' tweets (texture bars) and citizens' tweets (solid bars) for the topics that showed a distinct change across time periods.

We can observe that local governments' tweets talking about delays and closures of offices and schools are slightly larger before the snowstorms than during or after (from 21% to 18% of the total communications), probably related to the fact that official closure announcements are typically made before the snowstorm fully develops. The opposite trend is observed for citizens' communications about government offices and schools which increase during and after the snowstorms (from 18% to 22%), probably reflecting continued discussions about closure decisions. Preparedness communications from local governments show a large, decreasing trend in volumes (from 22% to 11%), which might indicate that local governments focus on spreading information about how to prepare for the snowstorm well in advance; while no significant change is observed in citizens' preparedness communications. On the other hand, local governments' response communications heavily increase (from 29% to 49%) during and after the snowstorms take place, indicating that local governments share updates about snow removal and information on emergency services mostly once the snowstorm has reached their geographical boundaries and several days later; while citizen tweets focus on communicating about response topics mostly during snowstorms (9%). Finally, citizen tweets showed that traffic information was mostly posted during snowstorms, reflecting the use of Twitter by citizens as a (real-time) news sharing platform; while local governments' traffic communications did not show any significant change.

8 CONCLUSIONS

We have presented a semi-automatic framework to extract and compare, in retrospect, the digital communication footprints of citizens and local governments during snowstorms at various spatio-temporal scales. The framework uses a fully unsupervised ST-LDA model to select disaster-related tweets and to extract the topics that are discussed online during snowstorms; together with human annotators to give labels and meaning to the topics identified. We acknowledge that by exclusively using geolocated tweets from citizens to compare against local governments' tweets, we are only able to capture the voice of those who have geolocation enabled on their Twitter accounts. Although that is a limitation, it is the unique way to extract local information that is meaningful and actionable at the granular spatio-temporal scales we are interested in studying.

Our analysis of tweets from 18 snowstorms shows that the digital communication behaviors of citizens and local governments are vastly different. While citizens use Twitter heavily to share thoughts and opinions about the implications of the snowstorms; and to a lesser degree, to exchange information about delays, closures and traffic conditions, local governments heavily focus on preparedness and response.

Although the evaluation focuses on snowstorms, the framework can be used to extract and compare, in retrospect, digital communication footprints from local governments and citizens for any type of acute event were both local agencies and the population participate online via Twitter. In the future, we plan to use the proposed framework to analyze whether local governments' social media strategies are covering the information that citizens seek online during disasters; and to provide information to help local governments cater their social media messages to citizens' needs.

REFERENCES

[1] Adam Acar and Yuya Muraki. 2011. Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities* 7, 3 (2011), 392–402.

[2] Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. 2014. Tweedr: Mining twitter to inform disaster response. *International Community on Information Systems for Crisis Response and Management* (2014).

[3] A. Axelrod, X. He, and J. Gao. 2010. Domain Adaptation via Pseudo In-domain Data Selection. *Annual Meeting of the Association for Computational Linguistics ACL* (2010).

[4] R.W. Belter and M.P. Shannon. 1993. Impact of natural disasters on children and families. In *Children and disasters*. Springer, 85–103.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine earning research* 3, Jan (2003), 993–1022.

[6] A. Bruns, J.E. Burgess, K. Crawford, and F. Shaw. 2012. # qldfloods and@ QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods. http://www.cci.edu.au/floodsreport.pdf[Accessed March-2-2017]. (2012).

[7] Axel Bruns and Jean E Burgess. 2012. Local and global responses to disaster:# eqnz and the Christchurch earthquake. In *Disaster and emergency management conference*, Vol. 2012. 86–103.

[8] Axel Bruns and Stefan Stieglitz. 2012. Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services* 30, 3-4 (2012), 160–185.

[9] CensusBureau. 2016. USA County Shapefiles. https://www.census.gov/cgi-bin/geo/shapefiles/index.php. (2016).

[10] Enhong Chen, Yanggang Lin, Hui Xiong, Qiming Luo, and Haiping Ma. 2011. Exploiting probabilistic topic models to improve text categorization under class imbalance. *Information Processing & Management* 47, 2 (2011), 202–214.

[11] F. Cheong and C. Cheong. 2011. Social media data mining: A social network analysis of tweets during the 2010-2011 Australian floods. *Pacific Asia Conference on Information Systems* (2011).

[12] J. L. Fleiss. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* ((1971)), 378.

[13] L. Hong, W. Yang, P. Resnik, and V. Frias-Martinez. 2016. Uncovering Topic Dynamics of Social Media and News: the Case of Ferguson. *International Conference on Social Informatics* (2016).

[14] Amanda L Hughes, Lise AA St Denis, Leysia Palen, and Kenneth M Anderson. 2014. Online public communications by police & fire services during the 2012 Hurricane Sandy. In *CHI Conference on Human Factors in Computing Systems*.

[15] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Practical Extraction of Disaster-relevant Information from Social Media. In *World Wide Web Conference*.

[16] Muhammad Imran, Shady Mamoon Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Extracting information nuggets from disaster-related messages in social media. *International Community on Information Systems for Crisis Response and Management* (2013).

[17] K. Kireyev, L. Palen, and K. Anderson. 2009. Applications of topics models to analysis of disaster-related twitter data. In *NIPS Conference on Neural Information Processing Systems*.

[18] Yury Kryvasheyeu, 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* 2, 3 (2016).

[19] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief.. In *International Conference on Web and Social Media ICWSM*.

[20] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *World Wide Web Conference*.

[21] Edward Loper and Steven Bird. 2002. NLTK: The natural language toolkit. In *Annual Meeting of the Association for Computational Linguistics ACL Workshop*. Association for Computational Linguistics, 63–70.

[22] R. Moore and W. Lewis. 2010. Intelligent Selection of Language Model Training Data. *Annual Meeting of the Association for Computational Linguistics ACL* (2010).

[23] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic Evaluation of Topic Coherence *(Human Language Technology Conference)*.

[24] NOAA. 2017. http://www.nohrsc.noaa.gov/. (2017). [Accessed March-2-2017].

[25] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises.. In *International Conference on Web and Social Media ICWSM*.

[26] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Computed-Supported Cooperative Work and Social Computing CSCW*.

[27] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC. *Lawrence Erlbaum Ass.* 71 (2001).

[28] John F Jr. Rooney. 1967. The urban snow hazard in the United States: An appraisal of disruption. *Geographical Review* (1967), 538–559.

[29] H.M. Saleem, Y. Xu, and D. Ruths. 2014. Effects of disaster characteristics on Twitter event signature. *Procedia Engineering* 78 (2014), 165–172.

[30] F. Shaw, J. Burgess, K. Crawford, and A. Bruns. 2013. Sharing News, Making Sense, Saying Thanks: Patterns of Talk on Twitter during the Queensland Floods. *Australian Journal of Communication* (2013).

[31] Kate Starbird and Leysia Palen. 2010. *Pass it on?: Retweeting in mass emergency*. International Community on Information Systems for Crisis Response and Management.

[32] Kate Starbird and Jeannie Stamberger. Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In *International Community on Information Systems for Crisis Response and Management*.

[33] T. Terpstra, A de Vries, R Stronkman, and GL Paradies. 2012. *Towards a realtime Twitter analysis during crises for operational crisis management*. Simon Fraser University.

[34] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *CHI Conference on Human Factors in Computing Systems*. 1079–1088.

[35] Christian Wartena and Rogier Brussee. 2008. Topic detection by clustering keywords. In *DB and Expert Systems Applications*.

[36] Andrew Worthington and Abbas Valadkhani. 2004. Measuring the impact of natural disasters on capital markets: an empirical application using intervention analysis. *Applied Economics* 36, 19 (2004), 2177–2186.

[37] C. Wukich and I. A. Mergel. 2014. Closing the Citizen-Government communication gap: Content,audience, and network analysis of government tweets. *Socal Science Research Network Working Paper* (2014).

[38] Sarita Yardi, Daniel Romero, Grant Schoenebeck, and others. 2009. Detecting spam in a twitter network. *First Monday* 15, 1 (2009).

[39] Louis Yu, Sitaram Asur, and Bernardo A Huberman. 2011. What Trends in Chinese Social Media. In *CoRRabs/1107.3522*.