Computers, Environment and Urban Systems 83 (2020) 101514

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters

Chao Fan[a,*], Miguel Esparza[a], Jennifer Dargin[a], Fangsheng Wu[b], Bora Oztekin[a], Ali Mostafavi[a,*]

[a] Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840, USA
[b] Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77840, USA

ARTICLE INFO

Keywords:
Social media
Big data fairness
Bias analysis
Disaster informatics

ABSTRACT

The objective of this study is to examine and quantify the relationships among sociodemographic factors, damage claims, and social media attention on areas during natural disasters. Social media has become an important communication channel for people to share and seek situational information to learn of risks, to cope with community disruptions, and to support disaster response. Recent studies in disaster informatics have recognized the presence of bias in the representation of social media activity in areas affected by disasters. To explore related factors for such bias, existing studies have used geo-tagged tweets to assess the extent of social media activity in disaster-affected areas to evaluate whether vulnerable populations remain silent on social media. However, less than 1% of all tweets are actually geo-tagged; therefore, attempts to understand the representativeness of geotagged tweets to the general population have shown that certain populations are over- or underrepresented. To address this limitation, this study examined the attention given to locations based on social media content. The study conducted a content-based analysis to filter tweets related to 84 super-neighborhoods in Houston during Hurricane Harvey and 57 cities in North Carolina during Hurricane Florence. By examining the relationships among sociodemographic factors, the number of damage claims, and the volume of tweets, the results showed that social media attention concentrates in populous areas, independent of education, language, unemployment, and median income. The relationship between population and social media attention is characterized by a sub-linear power law, indicating a large variation among the sparsely populated areas. Using a machine-learning model to label the topics of the tweets, the results showed that social media users pay more attention to rescue- and donation-related information; nevertheless, the topic variation is consistent across areas with different levels of attention. These findings contribute to a better understanding of the spatial concentration of social media attention regarding posting and spreading situational information in disasters. The findings could inform emergency managers and public officials to effectively use social media data for equitable resource allocation and action prioritization.

1. Introduction

Equitable response and recovery, such as equitable allocation of relief resources and prioritization of actions, are important for reducing causalities and property losses in areas unevenly impacted by a disaster. One key approach to equitable disaster response is to gain awareness of the situations and disparities of the impacts of disasters among affected areas. Enhanced equitable situational awareness, and thus response, is critical to adapting to uneven disaster impacts (Seppänen & Virrantaus, 2015). With the increased use of digital devices and social platforms, social media offers the possibility of improved disaster communication. Unlike traditional media, such as radio and news articles, social media enables augmented information capacity and rapid interactivity, by allowing the general population to efficiently post and share situational information (Zhang, Fan, Yao, Hu, & Mostafavi, 2019). This phenomenon makes social media data attractive to large-scale laboratories for disaster research (Ogie, Clarke, Forehead, & Perez, 2019), including detecting disruptive events (Fan & Mostafavi, 2019; Fan, Mostafavi, Gupta, & Zhang, 2018), characterizing information propagation (Cvetojevic & Hochmair, 2018; Sutton et al., 2015; Yang et al., 2019), and measuring human sentiments (Ragini, Anand, & Bhaskar, 2018) during disasters. Twitter, in particular, has a geotagging feature that associates tweets with accurate longitude and latitude, enabling pinpointing of locations and time to show where the tweets were generated. Due to this benefit, a vast number of studies have analyzed the geotagged tweets for various objectives, such as estimating disaster

* Corresponding authors.
E-mail addresses: chfan@tamu.edu (C. Fan), amostafavi@civil.tamu.edu (A. Mostafavi).

https://doi.org/10.1016/j.compenvurbsys.2020.101514
Received 29 February 2020; Received in revised form 16 May 2020; Accepted 4 June 2020
Available online 11 June 2020
0198-9715/

scales (Kumar, Hu, & Liu, 2014), mapping disaster situations (Li, Wang, Emrich, & Guo, 2018), and quantifying human movement (Wang, 2015). These studies show the potential of social media to enhance situational awareness and disaster management (Samuels, Taylor, & Mohammadi, 2020).

Despite the opportunities brought by social media, the analysis of social media data may lead to issues of uneven representativeness in a disaster situation. Existing studies show that social media users pay varying levels of attention to different areas in disasters (Malik, Lamba, Nakos, & Pfeffer, 2015). There are some factors such as population size, minority percentage, or median income, that might be related to the variation in the number of social media posts in different affected areas. For example, Xiao, Huang, and Wu (2015) shows that areas where the percentage of people using social media and the percentage of households with communication services are quite low usually gain less tweets related to them, compared to the areas with high percentage of people using social media (Xiao et al., 2015). Hence, studies on the variation of social media attention in terms of the social and geographical disparities has emerged (Zou et al., 2018). For example, Madianou (2015) revealed that low-income participants have fewer social media opportunities (Madianou, 2015). Because of that, less situational information related to this group of people and the areas where they live are reported on social media. Other users and disaster managers would think there is less need of disaster relief and recovery. Hence, using social media data for disaster management and resource allocation may deepen social inequalities in disaster-affected areas. This leads to a deepening of social inequalities on social media and disaster recovery. In addition, Xiao et al. (2015) examined the spatial heterogeneity in the generation of geotagged tweets and found that socioeconomic factors are an important predictor of the number of tweets in Census tracts. Kryvasheyeu et al. (2016) found relationships between the proximity to a hurricane's path and the quantity of hurricane-related geotagged tweets (Kryvasheyeu et al., 2016). The findings of these studies indicate the influence of sociodemographic and damage factors on social media attention through the use of geotagged tweets.

Most existing studies have examined geo-tagged tweets for analyzing biases and disparities. As documented in existing studies, however, fewer than 0.42% of all tweets are associated with accurate geospatial information (latitudes and longitudes) (Cheng, Caverlee, & Lee, 2010), and only about 1% to 1.5% of the tweets are geo-coded with cities or neighborhoods (Morstatter, Pfeffer, Liu, & Carley, 2013). The low representation is due to the fact that the majority of the users disable the geotagging function to protect their privacy (Malik et al., 2015). Hence, a mass of disaster-related tweets is proactively ignored in existing studies. In addition, situation information provided in the texts of the tweets may not match the geotags where users generated the tweets (Fan, Wu, & Mostafavi, 2020). Moreover, trending topics for the information that people care about on social media also have spatial patterns, which correspond to the disaster footprint (Resch, Usländer, & Havas, 2018). These shortfalls in existing research could lead to limitations in understanding the relationship between geographical heterogeneity and social media attention, which may exacerbate the inequality of response actions and resource allocation in disasters.

The issues related to geotagged tweets and activity-only analysis, beg content-based analysis to examine the disparities of social media attention on areas varying demographics. To this end, this study endeavors to examine the influence of sociodemographic factors and disaster-related damage on social media attention during disasters by identifying the location-relevant tweets from their content, rather than from geotags. In this study, we measure the social media attention by counting the number of tweets in which the name of an area (i.e., a super neighborhood or a city) is mentioned. Three research questions guide this study:

- To what extent social-demographic and damage factors do influence levels of social media attention in different areas during disasters?
- What is the relationship between social media attention and the social-demographic and damage factors in different areas? To what extent do these relationships vary across different scales (neighborhood vs. city scale)?
- What information topics do users pay more attention to on social media? To what extent the topics differ across different areas with different levels of social media attention.

To address these questions, we filtered relevant tweets based on the mention of disaster-affected super-neighborhoods and cities in the contents of the tweets, quantified the relationship between social media attention and the sociodemographic factors of areas using regression models, and examined the variation of topics with different levels of user attention. The study examined tweets from Houston during Hurricane Harvey and from North Carolina during Hurricane Florence. The super-neighborhoods in Houston during Hurricane Harvey and cities in North Carolina during Hurricane Florence were both included in this study to uncover the cross-hazard similarity and cross-scale variability of social media attention patterns. Assessing these patterns regarding geospatial disparities through the content-based analysis inform about the biases in social media attention leading to inequalities that may cause unfair disaster treatment.

2. Related work

A number of prior studies have examined the geospatial patterns of social media activities, investigated social media use among different groups of people, and developed techniques for classifying tweets with humanitarian topics (Huang, Cervone, & Zhang, 2017; Stock, 2018). This section discusses these prior research works to highlight the need for a content-based location analysis in examining social media biases and disparities in disasters.

2.1. Geotagged tweets for disaster mapping

The geotags of tweets enable capturing the specific locations where the tweets were created (Sloan & Morgan, 2015). The geospatial patterns of the volume of tweets, to some degree, can imply population activities or disaster situations. Hence, the geospatial patterns of social media data using geo-tagged tweets during recent disasters have been extensively investigated in recent disasters, such as hurricanes (MacEachren et al., 2010); and flooding, fires, and haze (Kibanov, Stumme, Amin, & Lee, 2017). One avenue of studies focused on mapping and assessing the damages in disaster-affected areas using geotagged tweets. For example, Middleton et al. proposed a social media disaster-mapping platform to geo-parse real-time data streams and to assess disaster impacts (Middleton, Middleton, & Modafferi, 2014). To estimate the scale of disasters, predictive and approximate methods including kernel density estimation (Fan, Jiang, & Mostafavi, 2020) and network prediction model (Rahimi, Cohn, & Baldwin, 2015) were adopted for analyzing the geographical distribution of geotagged tweets. As researchers realize the limitations of geotagged tweets (Malik et al., 2015), recent studies have tried to uncover the underlying factors that induce biases in geotagged tweets. The primary factors examined in existing research are demographic and socioeconomic factors (Hecht & Stephens, 2014). Prior studies showed a positive association between locations whose populations have high-level academic degrees and the number of geotagged tweets (Jiang, Li, & Ye, 2019). Although existing studies have shown the influence of social media activities, the geotagged tweets account for only about 1% of the total volume of tweets and are not representative of the entirety of social media activities in disasters. In addition, it is also a possibility that the geotags of the tweets do not match the content of the tweets. People might post a geotagged tweet in a location different from the location described in the content of tweets. Hence, to better understand biases and disparities, it is necessary to analyze the social media attention by parsing the content of the tweets.

2.2. Effects of sociodemographic factors on social media use

Sociodemographic factors, such as education, minority status, and median income usually affect not only preparedness and response to disasters, but also the accessibility to and use patterns of social media in the affected areas. Specifically, as shown in the results of large-scale surveys, education can reveal people's ability to understand information about emergency plans or warning information to avoid dangerous situations (Cutter, Boruff, & Shirley, 2003). In a disaster, the language gap creates cultural barriers in a community that can exacerbate the disaster response for people who do not speak the native language well (Frigerio & De Amicis, 2016). A large percentage of racial or ethnic minorities or poverty within a population contributes to social vulnerability due to a diminished access to resources and internet during and after disasters (Cutter et al., 2003). In addition, existing studies have revealed that racial/ethnic and health status–related disparities are factors internet access, but do not significantly affect social media use patterns (Chou, Hunt, Beckjord, Moser, & Hesse, 2009). People with different demographic and psychosocial backgrounds also have different perceptions of using social media (Keating, Hendy, & Can, 2016). In particular, Neubaum, Rösner, Rosenthal-Von Der Pütten, and Krämer (2014) found that the motivations of people using social media in disasters is to share the emotions and to parse empathic concerns (Neubaum et al., 2014). Given this finding, population size tends to strongly influence social media attention within an area. These existing studies identified multiple potential factors that might influence the social media attention, which provide empirical evidence for us to select variables in this study. Despite progress made in capturing the relationships between social media use and sociodemographic factors, the relationship between the social media attention in an actual disaster event and its related factors remains to be clarified. This limited understanding is mainly due to the shortcomings of geo-tagged tweets, as discussed earlier.

2.3. Machine learning techniques for topic analysis

Finally, the topics of information delivered by tweets have also been studied in the context of disasters. An understanding of the bias in social media attention vis-à-vis the topics of information, however, is absent in existing studies. Topic analysis is usually achieved by adopting the techniques for batch labeling of social media data. To achieve batch labeling, researchers have developed multiple advanced techniques for classifying tweets with humanitarian categories (Imran, Castillo, Diaz, & Vieweg, 2014). The machine learning approaches for disaster applications started with unsupervised learning, such as Latent Dirichlet allocation topic modeling for detecting the trending topics in disasters (Fan et al., 2018; Hidayatullah, Aditya, Karimah, & Gardini, 2019). While these approaches can capture the topics of massive tweets in near real-time, the outputs are not stable due to excessive noise present in tweets. With the development of supervised learning, multiple advanced learning approaches such as Naïve Bayes classifier (Hutto & Gilbert, 2014) have been proposed for identifying the tweet topics. These approaches usually trained the models on labeled data and implemented on the datasets from similar hazards or crises. For example, Li et al. proposed a domain adaptation approach integrating with the Naïve Bayes classifier to label social media data in emerging target disasters (Li, Caragea, Caragea, & Herndon, 2018). Caragea et al. presented an approach based on Convolutional Neural Networks to identify informative messages in social media streams in disasters (Caragea, Silvescu, & Tapia, 2016). A recent study which fine-tuned a BERT-based (Bidirectional Encoder Representations from Transformers) classifier further enhanced the performance of the deep learning models for categorizing tweets (Fan, Wu, & Mostafavi, 2020). The advancement of machine learning approaches provides unique opportunities for analyzing the distribution of social media attention regarding the topics of the information.

In summary, recent work not only offers the techniques and empirical evidence for analyzing social media data in disasters, but also points out the necessity of examining inequalities in social media content regarding the volume and topics of social media posts. Important knowledge is still missing, however, regarding the spatial biases inherent in social media content in disasters. For example, certain areas receive more attention on social media during disasters than others; however, empirical studies that show that tweet concentrations in one area are necessarily associated with a sociodemographic factor. Without this information, it is likely that certain population groups are not receiving needed assistance during disasters. To this end, this paper discusses a conceptual model and applies statistical tests to analyze the relationships among social media attention, sociodemographic factors, damage claims, and evaluation of information topics.

3. Conceptual model and hypotheses development

In order to examine spatial biases existing in social media content, this study considers various sociodemographic factors associated with different sub-populations residing in the selected spatial areas. Multiple sociodemographic factors characterize societal attributes of disaster areas. As suggested by existing studies discussed in section 2, sociodemographic factors such as population size, percentage of people with high-school degrees, percentage of people whose native language is English, percentage of unemployment, percentage of minorities, and median income characterize societal attributes of disaster areas (Jiang et al., 2019). This study focuses mainly on these factors and their relationships with social media attention. Despite the potential influence of these factors, according to the psychological evidence (Neubaum et al., 2014), social media attention and activities are motivated by acquiring empathic concerns from people in the same situation. Due to the localized impacts of disasters, people living in vicinity of each other tend to be in the same situation. Hence, social media attention might be driven by the population of a specific affected area. As such, we hypothesize that:

H1. Population size and volume of damage claims in affected areas have a strong, positive relationship with social media attention during disasters, and social media attention is independent of other sociodemographic factors.

To test this hypothesis, a number of considerations should be examined. In particular, the variation in both population size (and damage claims) and the number of tweets could be extremely large between different areas. The variation might be further amplified by expanding the scale of the areas (from super-neighborhood scale to city scale). Thus, linear models in linear space might not be able to accurately capture the relationship. Shrinking the scale using logarithmic representation of the data space, responds to the skewness toward large variations in the dataset. Meanwhile, the percentage changes or multiplicative factors can also be identified from the data in logarithmic space. Then, linear models can be applied to the logarithmic space to quantify the relationship. This would lead to an examination of a power-law relationships between two variables. Therefore, we hypothesize that:

H2a. Social media attention and population (and damage claims) follow a power-law relationship of the form $y \approx x^{\beta}$, where x is the population of a city or super-neighborhood, y is the number of tweets, and β is the scaling exponent.

H2b. The power-law relationship is consistent at both city and super-neighborhood scales.

Social media platforms have no limitations with respect to the topics of the situational information posted by users. Based on the experiences

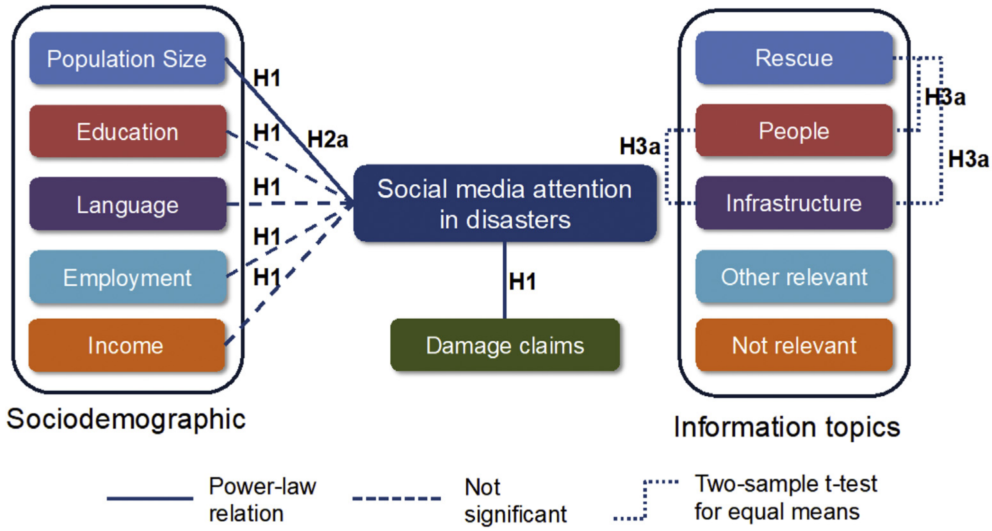

Fig. 1. Conceptual model for examining social media attention in disasters.

of previous disasters, situational information included infrastructure and utility damages, affected individuals, rescue, and volunteer efforts. This information could inform first responders, relief organizations, and residents outside the affected areas. Prior studies have shown that, by analyzing tweet content, social media serves as an important tool for communicating situational information (Fan, Wu, & Mostafavi, 2020). For example, relief organizations share the information about their resources and locations, and people at risk post their needs to connect to the relief organizations. In this study, we aim to understand if users pay different attention to different topics and the extent to which the topics differ across areas with different levels of social media attention. According to the findings of prior studies related to tweet topics, we hypothesize that:

H3a. Social media users pay most attention to rescue, volunteering, and donation-related information on Twitter.

H3b. Among areas of differing levels of social media attention, the topics of greatest interest show little variation.

Based on the aforementioned constructs, we propose a conceptual model for examining the relationships among multiple variables and developing hypotheses (Fig. 1).

4. Materials and methods

4.1. Study context and data collection

To test the hypotheses and answer the research questions, we utilized data related to super-neighborhoods in Houston during Hurricane Harvey and cities in North Carolina during Hurricane Florence. Hurricane Harvey, a category 4 storm, made landfall August 27, 2017, in Houston. The torrential rainfall that occurred during Harvey and the slow movement of the storm system caused intense flooding in Harris County, destroying 50,000 homes (Pulcinella, Winguth, Allen, & Dasa Gangadhar, 2019). Since the city of Houston sustained the majority of damages during Harvey, we determined to use Houston super-neighborhoods as a testcase (see Figure A3). Hurricane Florence formed in August 2018, and the early warning was sent about September 6. On September 14, Florence struck the southeastern coast of North Carolina and caused 53 fatalities and $16 to $40 billion in damage (Paul, Ghebreyesus, & Sharif, 2019). As Hurricane Florence damage affected multiple cities in North Carolina, we selected several cities in this case study (see Figure A3).

The super-neighborhoods are defined by City of Houston as a geographically designated area where residents, civic organizations, institutions and businesses work together to identify, plan, and set priorities to address the needs and concerns of their community (City of Houston, 2017). The majority of the damages caused by Harvey was in Houston, which is the fourth largest city in the United States. The damages and social-demographic characteristics of the super-neighborhood vary. Hence, to better capture the relation among damages, social-demographic factors, and social media attention, we examined super-neighborhoods as the spatial unit of analysis for the case study of Hurricane Harvey in Houston. In addition, Hurricane Florence struck the southeastern coast of the United States, including multiple cities in North Carolina. The affected areas include a number of small cities (but not large metropolitan areas like Houston). These small cities are analogous to the super-neighborhoods in Houston. That is, the social-demographic characteristics and damages are usually measured at city level by the official government. But, the scale of the cities is different from the scale of super-neighborhoods. Hence, to examine the consistency of the findings at different scales, we examined cities as spatial units of analysis in the case study of the Hurricane Florence (see Appendix A3).

To examine the social media attention on different areas, we collected 2 million tweets for both disaster events. The time period for collecting tweets for Hurricane Harvey spans from August 26 to September 4, 2017, which encompasses the response and recovery phases of Hurricane Harvey. Since Hurricane Florence was a longer disaster event, data was collected from September 6 until September 2, 2018. We collected the tweets using Twitter PowerTrack API (Application Programming Interface) with two filters. The Twitter PowerTrack API provides a complete historical dataset of the tweets posted on Twitter platform. To collect the tweets posted by the users affected by the disasters, the first filter was defined to identify the tweets posted by the users whose profiles show the locality of the areas of interest (e.g., Houston or the studied cities in North Carolina), and the tweets that have geotags in our predefined bounding boxes (e.g., the boundary of Houston neighborhoods or cities in North Carolina) (see Appendix A3). The second filter was the keywords that were used to identify the tweets specifically related to a super-neighborhood in Houston or a city in North Carolina. The keywords include the names and abbreviations of the areas. By applying these two filters, we can retrieve the tweets for our case studies.

4.1.1. Social media attention disparity indicators

Seven attributes were used to indicate attention disparities on social media during the disaster events. In this study, Census data from

Table 1
Summary statistics for sociodemographic data and FEMA damage claims.

Variables	Hurricane Harvey					Hurricane Florence				
	Obs.	Mean	SD	Min	Max	Obs.	Mean	SD	Min	Max
Population size	84	25,955	20,863	2031	119,598	57	6824	17,007	112	106,476
Education (%)	84	23.3	9.5	2	40	57	89.3	7.3	67.7	100
Unemployment (%)	84	7.9	4.0	2	23	57	5.4	3.6	0	19.6
Language (%)	84	47.2	21.7	10	89	57	5.6	4.9	0	25.9
Minority (%)	84	76.3	22.6	23	99	57	21.9	17.9	1.0	61.0
Median Income ($)	84	57,029	27,137	25,489	152,092	57	64,940	37,338	2453	160,311
Damage claims	84	626	3102	7	28,579	57	1141	1916	6	10,821

Note: "Obs." means number of observations.

American Factfinder (American FactFinder, 2020) was collected for the following social characteristics: population size, percentage of the educated population, percentage of unemployed persons, percentage of people whose native language is not English, proportion of minority population, and the median income. These social groups were observed in 84 super-neighborhoods for Hurricane Harvey in Houston, and 57 cities in North Carolina affected by Hurricane Florence.

To assess the physical damage in the super neighborhoods and cities, this study relied on the number of Federal Emergency Management Agency (FEMA) claims in each area of study. The damage claims were filed by people whose properties were damaged or services disrupted. This data was collected using HydroShare, which enables storage, management, sharing, publication, and annotation of data associated with hydrological studies (Horsburgh et al., 2016). FEMA claims data provide an effective way to measure the physical damage of a disaster because claims are used to compensate the public's property damage.

After removing areas without enough geotagged tweets, we had 84 super-neighborhoods and 57 cities in this study. The descriptive statistics were computed and are summarized in Table 1. Each attribute provides unique information about the social composition of the areas of study, as summarized in Table 1.

4.1.2. Social media attention

In this study, we define social media attention based mainly on the total number of tweets with content related to a particular spatial area. The name of spatial areas—super-neighborhoods names in the context of Houston and city names in the context of Florence—were used to determine the total number of tweets based on their content for each spatial area. We manually identified the names, abbreviations, main buildings, key roads, and relevant keywords for super-neighborhoods in Houston so that we could filter the relevant tweets for each super-neighborhood. We first downloaded the demographic data of super-neighborhoods, and the geo-coordinates of the main buildings and roads from the official website of the City of Houston (City of Houston, 2017). Then, using the geographical boundaries of the super-neighborhoods and the geo-coordinates of the main buildings and roads such as galleria mall and state highway 6, we associated the main buildings and roads to the super-neighborhoods. The names and abbreviations of the super-neighborhoods and the names of the main buildings and roads that belong to a specific super-neighborhood are selected as keywords to go through the content of the tweets, and associate the tweets with relevant super-neighborhoods. Cities in Florence are much larger in scale than super-neighborhoods and include a vast number of buildings and roads. Hence, only the names of the cities were employed for filtering relevant tweets. The descriptive statistics of the Twitter data for super-neighborhoods and cities are summarized in Table 2. As shown in Table 2, apparently, retweets account for the largest proportion of the social media attention with largest variations and maximum values in both super-neighborhood and city scales, and replies and quotes account for a very small proportion in the total amount of tweets.

4.2. Variable correlation analysis

Pearson's correlation coefficients were calculated for each pair of variables (Fig. 2) to confirm independence among the selected variables in the analysis. The results show that the number of damage claims is strongly and positively correlated with population size, while other factors, such as median income, unemployment, minority status, language, education, are all independent of the population size and damage claims. For example, areas with a high proportion of minority residents can be either populous or sparsely populated. This finding can be observed at both super-neighborhood and city scales. The relationships among education, unemployment, language, minority and median income are rather different at different scales. The results indicate that, to some degree, these variables are correlated with each other. This signifies that we cannot include these variables in a single regression model due to their relationship. This observation indicates that a pairwise comparison between social media attention and social variables is needed.

4.3. Machine learning for topic classification

Advances in machine learning techniques enable automatic labeling of tweet topics, which allows examination of the type of information social media users pay most attention to. Existing studies have developed machine-learning models, in particular, for classifying disaster-related tweets. We adopted a recently published work (Fan, Wu, & Mostafavi, 2020) in which an advanced BERT-based model is trained and tested for labeling hurricane-specific tweets with humanitarian categories. The BERT-based model is designed to train bidirectional representations that embed the words, position of the words, and the segments of a tweets. To implement the model, we first employed the definitions of the humanitarian categories proposed by Alam, Ofli, and Imran (2018). Table 3 shows example tweets during Hurricane Harvey and Florence for four humanitarian topics: (1) information from infrastructure damage, (2) affected individuals, (3) rescue efforts, and (4) other relevant information in related to the disaster events but may not deliver specific information related to the three main categories. There are tweets that mentioned relevant place names but do not contain information related to these specific categories, although these tweets still indicate the attention of social media users to specific areas. Our adopted machine learning method is able to identify these tweets. Based on the output of the method, we found that about 27% of the tweets detected by our machine learning method are non-relevant to humanitarian categories. In this study, we did not include these non-relevant tweets in the topic analysis.

Analysis of disparities in social media attention focused mainly on the first three main categories (Table 3). After training the model by defining these humanitarian topics, the adopted BERT-based machine-learning model was applied to the filtered tweets for super-neighborhoods in Houston and impacted cities in North Carolina. The model generated measures of the closeness of the tweets to each humanitarian

Table 2
Summary statistics for Twitter data.

Variables	Hurricane Harvey					Hurricane Florence				
	Obs.	Mean	SD	Min	Max	Obs.	Mean	SD	Min	Max
Original tweets	84	99.5	155.4	0	671	57	376.5	1194.8	0	8354
Retweets	84	374.7	674.5	0	3165	57	2968.8	7860.8	0	37,659
Replies	84	14.3	23.0	0	136	57	82.5	254.2	0	1508
Quotes	84	18.5	36.9	0	214	57	50.0	150.3	0	968
Total tweets	84	477	815	0	3721	57	3395	9148	0	48,198

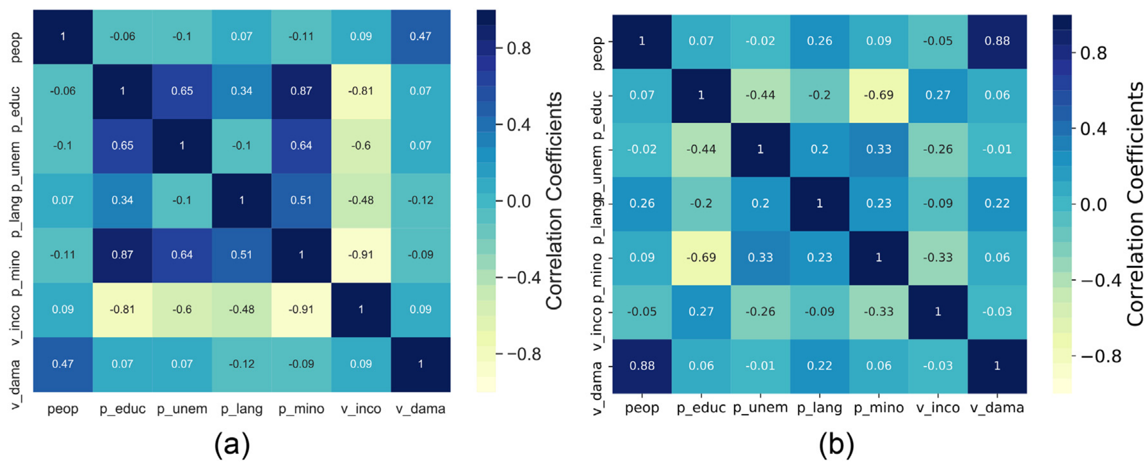

Fig. 2. Correlation analysis for sociodemographic and damage variables. (a) super-neighborhoods in Houston. (b) cities in North Carolina. "peop" represents population; "p_educ" represents the proportion of people who are at least high school gradudates; "p_unem" represents the proportion of unemployment population; "p_lange" represents the proprotion of people who speak languages other than English; "p_mino" represents the proportion of minority population; "v_inco" presents the median income of households; and "v_dama" represents the number of damage claims that FEMA received.

category, represented by a probability. The output of the model was the label for each tweet based on the highest probability of the category. Accordingly, we aggregated the number of tweets in each topic for each super-neighborhood and city.

5. Results

5.1. Influence of sociodemographic factors

As mentioned previously, variation in social media attention can be significantly large among different geographic areas (super-neighborhoods or cities). To make the comparison straightforward, we intuitively divided the super-neighborhoods and cities each into areas of high and low social media attention, with 300 tweets set as the threshold. This threshold level, 300 tweets for an area, was set to balance the size of the two groups of areas in both cases such that the results were comparable and the effect of group size was negligible. Then we conducted a two-sample *t*-test to assess the difference of means in the two groups for each hurricane event. This step will provide evidence about the presence of discrepancies in social attributes (e.g., population size, minority ratio, and median income) across different areas with different levels of social media attention. The results can support the quantitative analysis of the relationships between social media attention and social attributes.

Fig. 3 shows the differences of population (Fig. 3a and c) and damage claims (Fig. 3b and d) in groups of areas with high or low social media attention. In the case of Hurricane Harvey, the population mean in areas with high social media attention was about 35,000, while the population mean in areas with low attention was only 2000. Since damage levels were correlated to the population size, the mean of damage claims in areas with high social media attention is about 3000, compared with areas with low social media attention, which showed an average of 800 damage claims. The population size within a standard deviation in the group of areas with high attention was always greater than that of areas with low attention. We observed these patterns at the city scale during Hurricane Florence as well. As shown in Fig. A1 and A2 in Appendix A, the super-neighborhoods and cities with low social media attention concentrate in the bottom left corner, where both population size and damage claims are low. In contrast, areas with high social media attention distribute in the top-right corner, which

Table 3
Example Tweets for humanitarian topics in Hurricane Harvey and Florence.

Topics	Example tweets
Infrastructure and utility damages	*"Streets in downtown are filling with a lot of water – pls don't try to drive right now #Houston #Harvey* *"I-40 is flooded at Burgaw and at Castle Hayne. "We don't have a land access to Wilmington."*
Affected and injured individuals	*"I dover three hours from the North Carolina beaches but saw people from the coast in stores today buying emergency supplies"* *"my sister is at the house. No word on her apt downtown but we think it probably flooded"*
Rescue, volunteering, or donation effort	"*Night shift volunteers also sought at downtown Brown convention center shelter. #Harvey2017* *"Beulah Baptist Church in Calabash is taking donations for storm victims. They are in serious need of dog and cat food."*
Other relevant information	"find open restaurants and details on flooded areas at Downtown Houston" *"Hurricane #Florence rain totals so far. Oriental 21.6 Surf City: 16.6 New Bern: 14.26 Swansboro: 14.25 Calabash: 12…"*

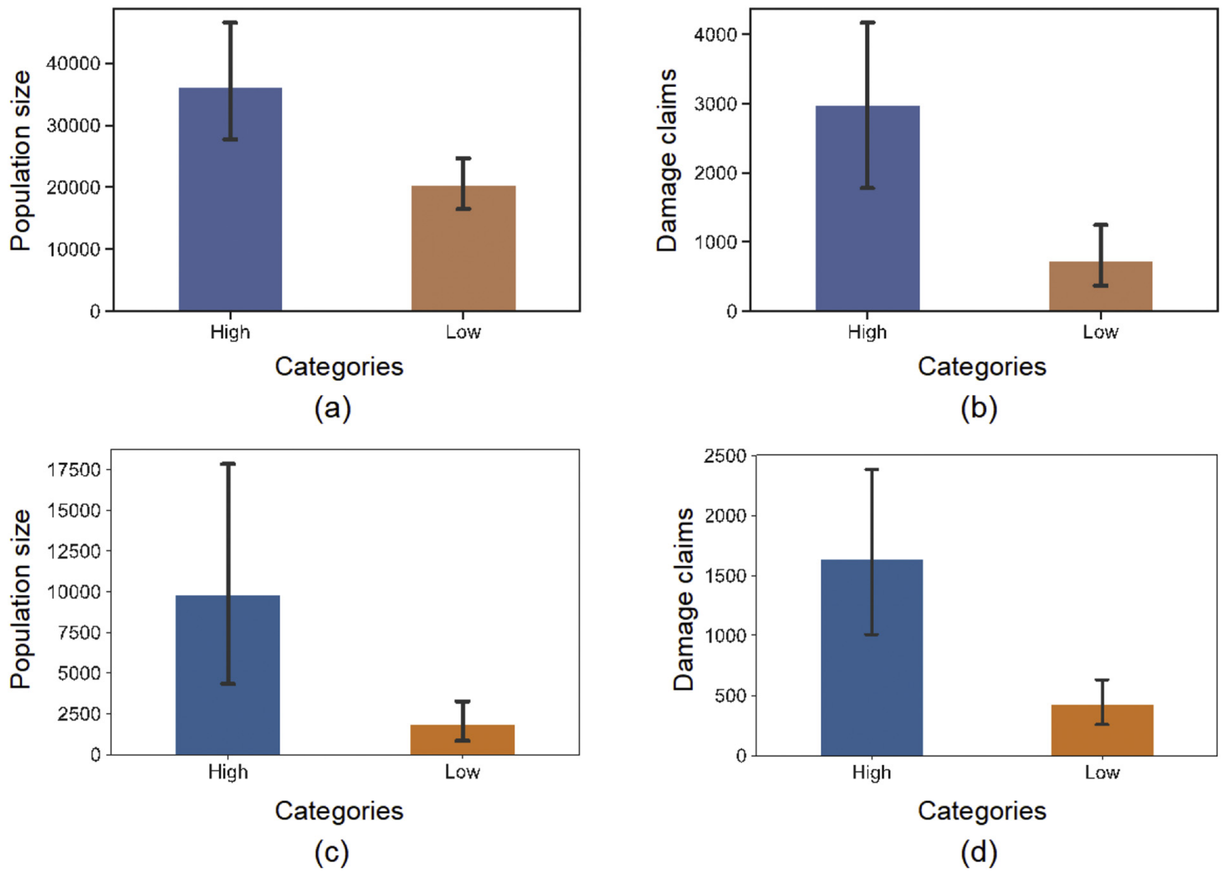

Fig. 3. The population and number of damage claims in areas with different level of social media attention. (a) population size of super-neighborhoods in Hurricane Harvey ($p < .01$); (b) damage claims of super-neighborhoods in Hurricane Harvey ($p < .01$); (c) population size of cities in Hurricane Florence ($p < .05$); and (d) damage claims of cities in Hurricane Florence ($p < .01$).

indicates that both damage claims and population size contribute to the level of social media attention (Fig. A1 and A2 in Appendix A). Hence, we can conclude that areas with high social media attention have much larger populations and a greater number of damage claims than those of areas with low social media attention. Through the test of significance, this finding is significant and consistent at both super-neighborhood and city scales. Since the number of damage claims is correlated to the population of the areas (Fig. 2.), we can conclude that the social media attention is concentrated in populous areas.

The focus of social media attention might also be affected by the composition of the population. As shown in Table 4 below, however, the means of the two groups in the areas are close for all variables. In addition, the population size is not related to any sociodemographic factors (Fig. 2). That means, for example, the areas with a higher percentage of minorities areas can be either populous or sparsely populated. It is possible that minority areas with high population size can still receive high social media attention (Fig. A1 and A2 in Appendix A). This result implies that the composition of the population in terms of any other sociodemographic factors (i.e., education, unemployment, language, minority, or median income) in the two groups is not significantly different. The effects of the sociodemographic composition of the population on the social media attention is negligible. This pattern can be identified from both super-neighborhoods and cities. Hence, social media attention is independent of these sociodemographic factors and is influenced only by population size and damage claims. This result supports hypothesis **H1.**

5.2. Analyzing the power-law relationship

In the next step, the study focused on analysis of the relationship between the population size and social media attention. Fig. 4 shows the concentration of social media attention by showing the scaling laws

Table 4
Differences and significance of sociodemographic variables in areas with different level of attention.

Cases	Variables	Mean of high attention areas	Mean of low attention areas	*P-value*
Super-neighborhoods in Hurricane Harvey	Education (%)	21.8	24.8	Not significant
	Unemployment (%)	7.5	8.1	Not significant
	Language (%)	44.4	48.7	Not significant
	Minority (%)	71.2	80.0	Not significant
	Median income ($)	63,459	52,255	Not significant
Cities in Hurricane Florence	Education (%)	88.6	90.0	Not significant
	Unemployment (%)	5.3	5.6	Not significant
	Language (%)	6.0	4.9	Not significant
	Minority (%)	23.3	19.3	Not significant
	Median income ($)	71,966	55,978	Not significant

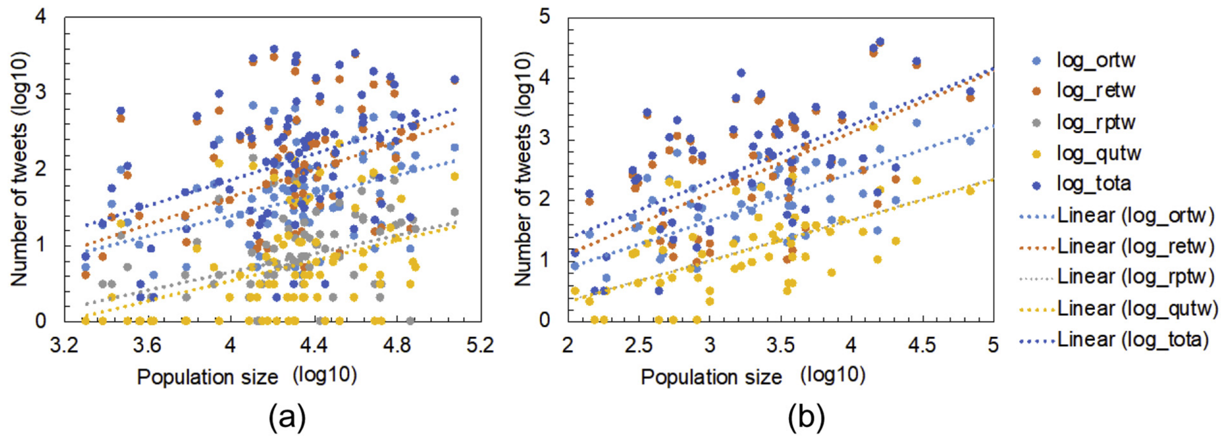

Fig. 4. Relations between number of tweets and population for super-neighborhoods in Hurricane Harvey (a) and cities in Hurricane Florence (b). Here, "log_ortw" represents the logarithmic number of original tweets; "log_retw" represents the logarithmic number of retweets; "log_rptw" represents the logarithmic number of replies; "log_qutw" represents the logarithmic number of quotes; and "log_tota" represents the number of all types of tweets.

Table 5
The coefficients in the relations between population and different types of tweets.

Variables	Coefficient (β)	
	Super-neighborhoods	Cities
Original tweets	0.67***	0.78***
Retweets	0.91***	1.00***
Replies	0.60***	0.67***
Quotes	0.66***	0.65***
Total tweets	0.86***	0.93***

Note: * $P < .05$, ** $P < .01$, *** $P < .001$.

followed by different types of tweets for areas in Hurricane Harvey (Fig. 4a) and Florence (Fig. 4b). Scaling laws in super-neighborhoods follow power-law relationships of the form $y \approx x^{\beta}$, where x is the population of a city or super-neighborhood, y is the number of tweets, and β is the scaling exponent (Table 5). This result indicates that social media attention is concentrated in highly populated areas. Such variation of social media attention is the consequence of the nonlinearity of user behaviors in online social networks (Balland et al., 2020). Specifically, in the case of Hurricane Harvey (Fig. 4a), the number of original tweets related to an area grows with the $\beta = 0.67$ power of the population size of that area. For retweets, the number of retweets whose content related to a super-neighborhood scales sub-linearly with population size with an exponent of $\beta = 0.91$. Similarly, the number of replies grows as the $\beta = 0.60$ power of the population size, the number of quotes scales as the $\beta = 0.66$ power of the population size, and the total number of tweets scales as the $\beta = 0.86$ power of the population size. These results support the hypothesis **H2a**.

We repeated this exercise by studying the scaling laws related to different types of tweets for cities and found similar, but a bit greater coefficients to those for super-neighborhoods (Table 5). Thus, the variation of social media attention (in relation with population size) at the city scale is slightly greater than that at the super-neighborhood scale. Despite these small differences, generally, the coefficients for the relationships in city and super-neighborhood scales are close to each other. Hence, the quantitative relationships and corresponding findings could be scalable for different disasters and areas. This result supports hypothesis **H2b**. In addition, almost all coefficients are smaller than 1, which indicates a sub-linear relationship between population size and social media attention in the logarithmic space. This implies that that the bias in social media attention vis-a-vis the size of the cities is weak (Barabási, 2013). The attention variation is greater among less populated areas than that among populous areas.

5.3. Variation of attention on topics

Next, we applied the BERT-based machine-learning model to label the tweets with humanitarian topics and investigated whether the concentration of social media attention shows variations in information topics and areas. Fig. 5a shows the mean percentages of the tweets in

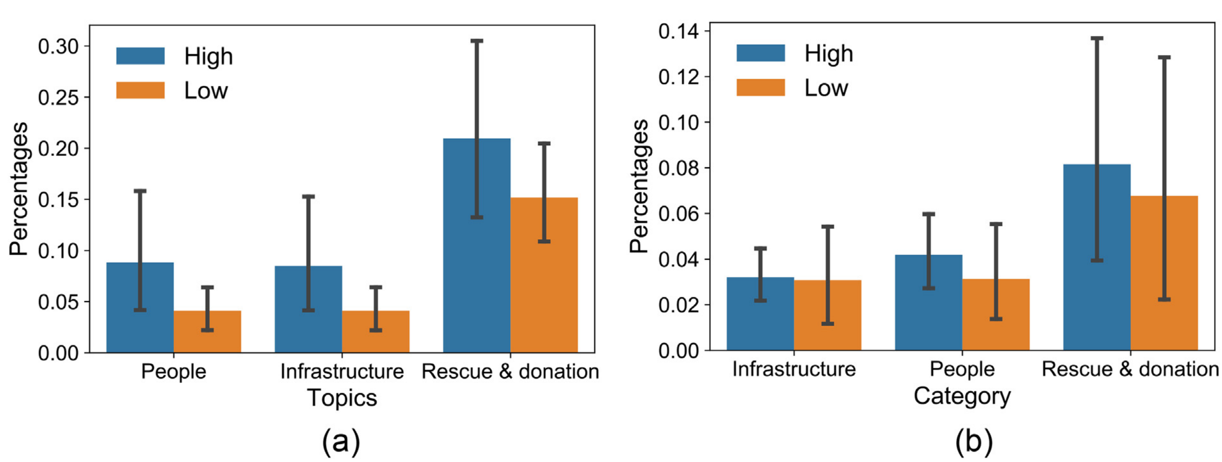

Fig. 5. The percentages of the tweets of the three categories of topics for areas with different levels of social media attention: (a) super-neighborhoods in Hurricane Harvey; (b) cities in Hurricane Florence.

Table 6
The differences and significance of social media attention for different types of tweets.

Variables	*P-value*	
	Super-neighborhoods	Cities
Infrastructure v. people	Not significant	Not significant
Infrastructure v. rescue & donation	***	*
People v. rescue & donation	***	Not significant

Note: * $P < .05$, ** $P < .01$, *** $P < .001$.

each topic for groups of areas with different levels of attention for super-neighborhoods. The figure shows that, in general, high attention areas have proportionally slightly more tweets with specific topics than those of low attention areas. This difference, however, is not statistically significant. So, the average percentages of tweets for each topic are consistent in areas with different levels of attention. The same pattern is observed at the city scale for Hurricane Florence (Fig. 5b). By conducting pairwise comparisons for mean percentages of tweets of different topics, we find that the tweets related to rescue and donation efforts are predominant (Table 6). The pattern is less significant at the city scale than that at the super-neighborhood scale during Hurricane Harvey. But, some of the findings are still similar at both scales. For example, compared to the difference between infrastructure and rescue information, users paid more attention to rescue- and donation-related information. The information related to affected individuals and infrastructure damage accounts for similarly low percentages of the total amount of tweets and does not show significant difference in the proportions of the tweets. Consequently, social media posts pay more attention to posting and spreading rescue and donation information in for disaster-affected areas. This finding supports hypotheses **H3a** and **H3b**.

6. Discussion

This study examined the influence of sociodemographic factors, population size, and damage claims on the attention of social media to areas in times of disasters. The results of the study show that the population size of an area explains the variations in the degree to which social media attention concentrates. We show this relationship to be present for tweets of several topics and to be independent of other sociodemographic factors. The findings show that social media attention is biased toward more populous areas during disasters. A sub-linear scaling law was adopted to quantify the relationship between attention and population density. The sub-linearity suggests that social media attention may vary in sparsely populated areas, while the attention is concentrated and constant among populous areas. With regard to topics of tweets, more attention was paid to rescue and donation efforts. The variation in topic attention is consistent in both super-neighborhood and city scales; hence, the identified relationships can be generalized to different areas and scales.

6.1. Theoretical contributions

The quantitative findings in this study contribute to a more precise understanding of spatial biases in social media information in disasters. First, different from prior studies using geotagged tweets, this study is the first to analyze social media attention by investigating the content of the tweets. The variation of social media attention across different areas confirms the findings from existing studies that have used geotagged tweets; however, the patterns are different. From the content-based analysis, the only significant factor is the population size. The quantity of damage claims is also related to the quantity of tweets, but it is also correlated to the population size (Fig. 2).

This study contributes to the literature concerning the effectiveness of social media for needs and disruption assessment. For example, studies leveraging social media to assess disruptions and needs in disasters should take population variation into consideration. Existing techniques and tools, such as event detection and topic modeling for understanding disruptions and needs, usually miss the representativeness issue of social media data. This issue would cause biases in these techniques (Du, Yang, Zou, & Hu, 2019). The evidence of the bias of social media attention present in this study is able to point out the source of the bias (i.e., population size of an area). By mitigating the spatial concentration of the data, existing machine-learning models will be more effective at capturing situations from social media in disasters.

6.2. Implications for practice and policy

The idea that social media attention is concentrated more in populous areas poses a range of questions for emergency managers, responders, and city planners and officials. This finding suggests that emergency management agencies and relief organizations need to rethink their response strategies if they rely social media information to make response and recovery decisions. Highly populated areas have more resources to consume and process compared to their less-populated counterparts. This can be observed from the results for the variation of social media attention in information topics. The results in section 5 show that the proportions of tweets with different topics remain constant across areas with different levels of social media attention. Since the tweets are concentrated in populous areas, however, the variation of the number of tweets for each topic is amplified by the variation in population size. This finding signifies that the rescue- and donation-related information is concentrated in populous areas. This signifies possible potential inequities in the allocation of relief resources and the prioritization of response actions due to the unequal attention paid by the public to different areas. For example, public information being posted and spread on social media attracts the attention of the relief organizations; however, little would be known about the situation in sparsely populated areas. Although the damages and needs in populous areas might be more severe than those of sparsely populated areas, the variation of social media attention may exacerbate the inequality of the disaster treatment in terms of the efficiency of response and the sufficiency of supplies. Hence, if the social media attention and the population size cannot be dissociated, the spatial inequality observed among populous and sparsely populated areas are likely to increase in real-word disaster management. Policymakers and response managers must recognize such inequality both within cities and between cities.

6.3. Limitations and future research

The limitations of this study should be taken into consideration in future studies. The descriptive nature of our analysis focused only on examining the associations between social and damage factors with respect to social media attention, but does not provide a clear indication of the causal factors leading to increases in the spatial concentration of social media attention in disasters. The topics in the texts of the tweets might provide cues for understanding the causes of this concentration pattern. For example, rescue and donation efforts might be more concentrated in populous areas, which could lead to a concentration of social media attention. The concentration of other relevant tweets in populous areas, however, cannot be clearly explained. Hence, future studies can look into the causal effects between societal characteristics and social media attention in disaster-affected areas. Secondly, while this study obtains the quantitative findings in both super-neighborhood and city scales for two national-level representative disasters, the next question to investigate is how much these patterns vary from these cases to other contexts and scales. Future studies can adopt this analysis method in other disaster events to evaluate the universality of these patterns regarding social media attention from our study. Also, with advances in name entity recognition, the content of tweets could be associated with finer-scale spatial areas in order to examine the attention disparity in more detail.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant Number CMMI-1846069 (CAREER), the National Academies' Gulf Research Program Early-Career Research Fellowship and the Amazon Web Services (AWS) Machine Learning Award. The authors also would like to acknowledge the funding support from the National Academies' Gulf Research Program Early-Career Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation and Amazon Web Services.

Appendix A

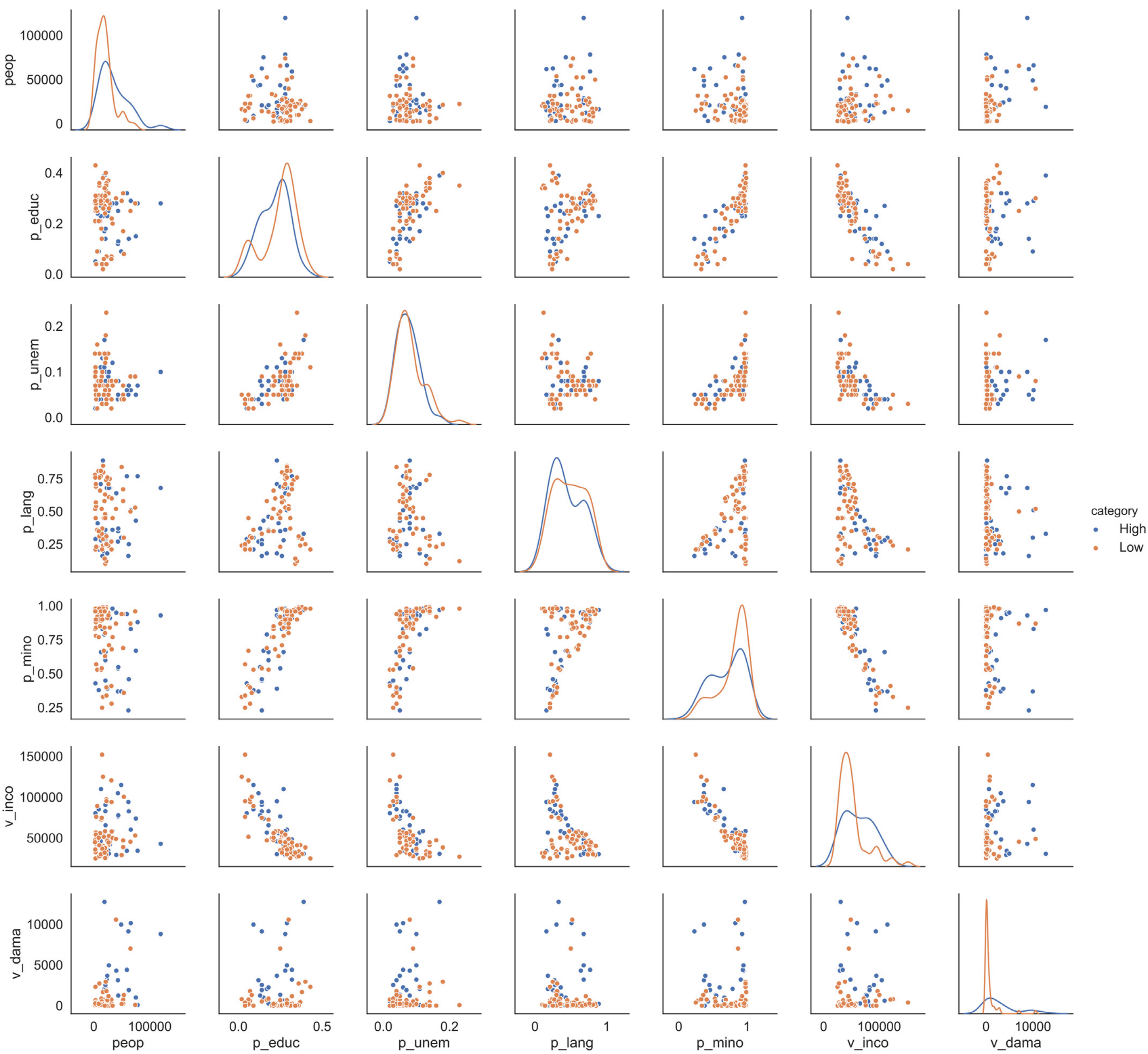

Fig. A1. Distribution of the sociodemographic and damage variables for the super-neighborhoods in the case of Hurricane Harvey. "peop" represents the population size; "p_educ" represents the proportion of people who are at least high school gradudates; "p_unem" represents the proportion of unemployment population; "p_lange" represents the proprotion of people who speak languages other than English; "p_mino" represents the proportion of minority population; "v_inco" prepresents the median income of households; and "v_dama" represents the number of damage claims that FEMA received. The blue dots represent the areas with high social media attention; and orange dots represent the areas with low social media attention

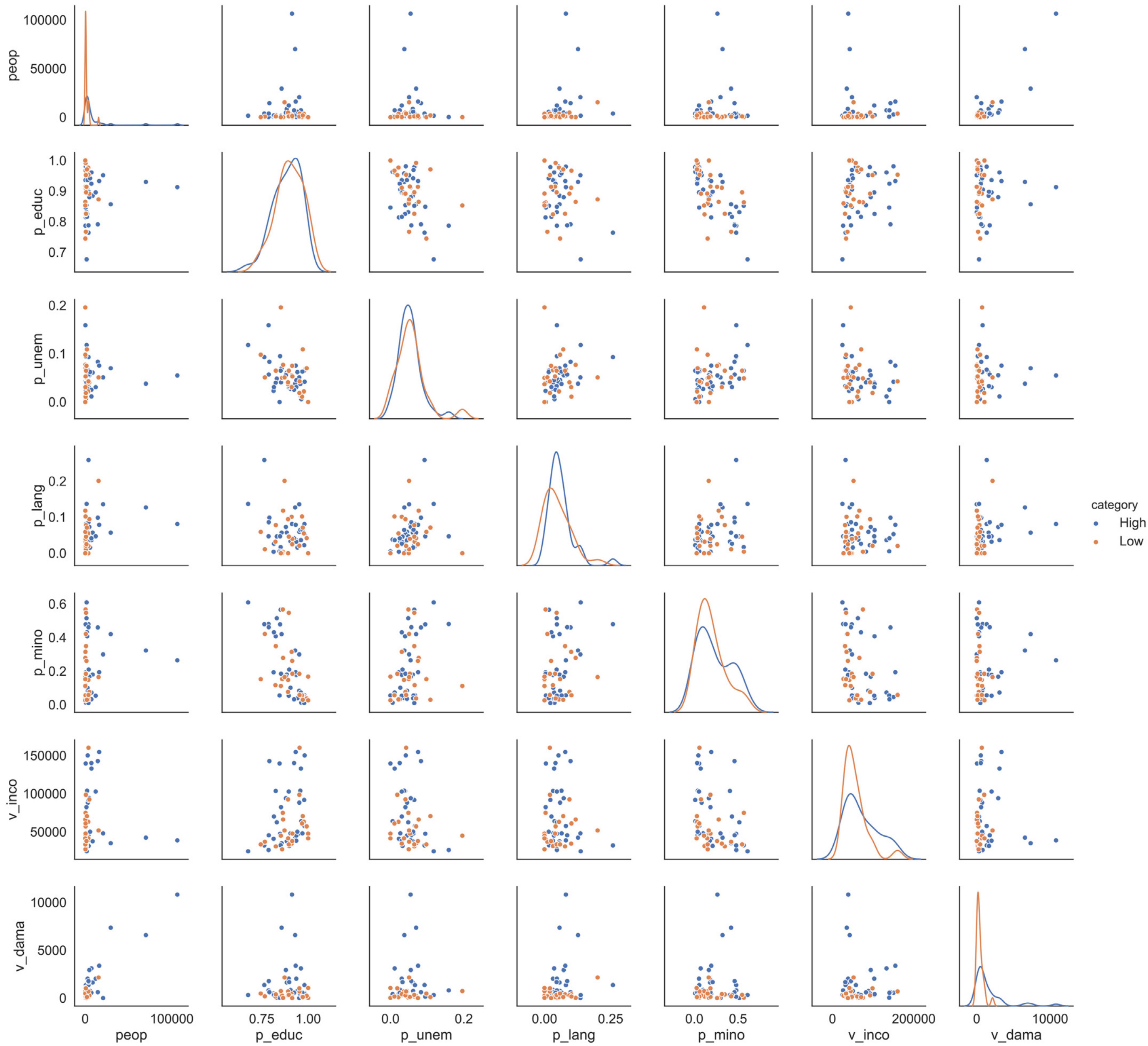

Fig. A2. Distribution of the sociodemographic and damage variables for the cities in the case of Hurricane Florence. “peop” represents the population size; “p_educ” represents the proportion of people who are at least high school gradudates; “p_unem” represents the proportion of unemployment population; “p_lange” represents the proprotion of people who speak languages other than English; “p_mino” represents the proportion of minority population; “v_inco” prepresents the median income of households; and “v_dama” represents the number of damage claims that FEMA received. The blue dots represent the areas with high social media attention; and orange dots represent the areas with low social media attention. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

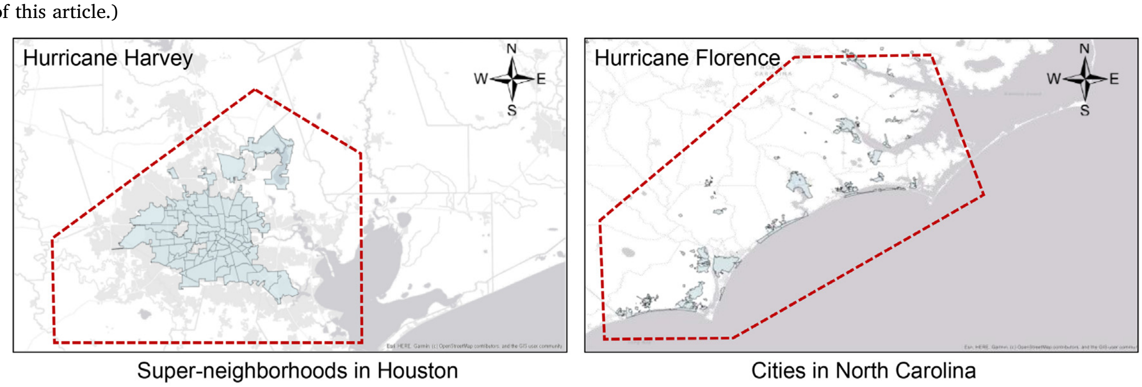

Fig. A3. Maps of studied super-neighborhoods in Houston and cities in North Carolina. Red bounding boxes represents the disaster-affected areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Alam, F., Ofli, F., & Imran, M. (2018). CrisisMMD: Multimodal Twitter datasets from natural disasters. Retrieved from http://arxiv.org/abs/1805.00713.

American FactFinder (2020). American FactFinder. Retrieved February 19, 2020, from https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

Balland, P. A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0803-3.

Barabási, A.-L. (2013). Network science. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*. https://doi.org/10.1098/rsta.2012.0375.

Caragea, C., Silvescu, A., & Tapia, A. H. (2016). *Identifying informative messages in disaster events using convolutional neural networks.* In Proceedings of the International ISCRAM Conference.

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet : A content-based approach to geo-locating twitter users. *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 759–768). . https://doi.org/10.1145/1871437.1871535.

Chou, W. Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social media use in the United States: Implications for health communication. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.1249.

City of Houston (2017). *Super neighborhoods demographic data.* Retrieved December 5, 2019 http://www.houstontx.gov/planning/Demographics/super_neighborhoods_2.html.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*. https://doi.org/10.1111/1540-6237.8402002.

Cvetojevic, S., & Hochmair, H. H. (2018). Analyzing the spread of tweets in response to Paris attacks. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.03.010.

Du, M., Yang, F., Zou, N., & Hu, X. (2019). Fairness in deep learning: A computational perspective. Retrieved from http://arxiv.org/abs/1908.08843.

Fan, C., Wu, F., & Mostafavi, A. (2020). A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. *IEEE Access, 8*, 10478–10490. https://doi.org/10.1109/ACCESS.2020.2965550.

Fan, C., Jiang, Y., & Mostafavi, A. (2020). Social sensing in disaster city digital twin: an integrated textual-visual-geo framework for situational awareness during built environment disruptions. *Journal of Management in Engineering,* 1–12.

Fan, C., & Mostafavi, A. (2019). A graph-based method for social sensing of infrastructure disruptions in disasters. *Computer-Aided Civil and Infrastructure Engineering, 34*(12), 1055–1070. https://doi.org/10.1111/mice.12457.

Fan, C., Mostafavi, A., Gupta, A., & Zhang, C. (2018). A system analytics framework for detecting infrastructure-related topics in disasters using social sensing. In I. F. C. Smith, & B. Domer (Eds.). *Advanced computing strategies for engineering* (pp. 74–91). Cham: Springer International Publishing.

Frigerio, I., & De Amicis, M. (2016). Mapping social vulnerability to natural hazards in Italy: A suitable tool for risk mitigation strategies. *Environmental Science and Policy*. https://doi.org/10.1016/j.envsci.2016.06.001.

Hecht, B., & Stephens, M. (2014). *A tale of cities: Urban biases in volunteered geographic information. In Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.

Hidayatullah, A. F., Aditya, S. K., Karimah, & Gardini, S. T. (2019). *Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA). In IOP Conference Series: Materials Science and Engineering.* https://doi.org/10.1088/1757-899X/482/1/012033.

Horsburgh, J. S., Morsy, M. M., Castronova, A. M., Goodall, J. L., Gan, T., Yi, H., ... Tarboton, D. G. (2016). HydroShare: Sharing diverse environmental data types and models as social objects with application to the hydrology domain. *Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12363.

Huang, Q., Cervone, G., & Zhang, G. (2017). A cloud-enabled automatic disaster analysis system of multi-sourced data streams: An example synthesizing social media, remote sensing and Wikipedia data. *Computers, Environment and Urban Systems, 66*, 23–37. https://doi.org/10.1016/j.compenvurbsys.2017.06.004.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Eighth international AAAI conference on weblogs and social media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109.

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2014). *Processing social media messages in mass emergency: A survey*. https://doi.org/10.1145/2771588.

Jiang, Y., Li, Z., & Ye, X. (2019). Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. *Cartography and Geographic Information Science, 46*(3), 228–242. https://doi.org/10.1080/15230406.2018.1434834.

Keating, R. T., Hendy, H. M., & Can, S. H. (2016). Demographic and psychosocial variables associated with good and bad perceptions of social media use. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.12.002.

Kibanov, M., Stumme, G., Amin, I., & Lee, J. G. (2017). Mining social media to inform peatland fire and haze disaster management. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-017-0446-1.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances,* (March), 1–12. https://doi.org/10.1126/sciadv.1500779.

Kumar, S., Hu, X., & Liu, H. (2014). *A behavior analytics approach to identifying tweets from crisis regions. Proceedings of the 25th ACM Conference on Hypertext and Social Media - HT '14*. 255–260. https://doi.org/10.1145/2631775.2631814.

Li, H., Caragea, D., Caragea, C., & Herndon, N. (2018). Disaster response aided by tweet classification with a domain adaptation approach. *Journal of Contingencies & Crisis Management*. https://doi.org/10.1111/1468-5973.12194.

Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: A case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science, 45*(2), 97–110. https://doi.org/10.1080/15230406.2016.1271356.

MacEachren, A. M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., & Mitra, P. (2010). Geo-Twitter analytics: Applications in crisis management. *Proceedings of the 25th international cartographic conference*.

Madianou, M. (2015). Digital inequality and second-order disasters: Social media in the Typhoon Haiyan Recovery. *Social Media and Society*. https://doi.org/10.1177/2056305115603386.

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). *Population Bias in geotagged tweets. ICWSM-15 Workshop: Standards and Practices in Large-Scale Social Media Research.* 18–27. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10662.

Middleton, S. E., Middleton, L., & Modafferi, S. (2014). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems, 29*(2), 9–17. https://doi.org/10.1109/MIS.2013.126.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). *Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose. In Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*.

Neubaum, G., Rösner, L., Rosenthal-Von Der Pütten, A. M., & Krämer, N. C. (2014). Psychosocial functions of social media usage in a disaster situation: A multi-methodological approach. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2014.01.021.

Ogie, R. I., Clarke, R. J., Forehead, H., & Perez, P. (2019). Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.09.002.

Paul, S., Ghebreyesus, D., & Sharif, H. O. (2019). *Brief communication: Analysis of the fatalities and socio-economic impacts caused by hurricane florence. 2005*, 1–12. https://doi.org/10.3390/geosciences9020058.

Pulcinella, J. A., Winguth, A. M. E., Allen, D. J., & Dasa Gangadhar, N. (2019). Analysis of flood vulnerability and transit availability with a changing climate in Harris County, Texas. *Transportation Research Record*. https://doi.org/10.1177/0361198119839346.

Ragini, J. R., Anand, P. M. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management, 42*, 13–24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004.

Rahimi, A., Cohn, T., & Baldwin, T. (2015). Twitter user geolocation using a unified text and network prediction model. *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*https://doi.org/10.3115/v1/p15-2104.

Resch, B., Usländer, F., & Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science*. https://doi.org/10.1080/15230406.2017.1356242.

Samuels, R., Taylor, J., & Mohammadi, N. (2020). Silence of the Tweets: incorporating social media activity drop-offs into crisis detection. *Natural Hazards*. https://doi.org/10.1007/s11069-020-04044-2 In press.

Seppänen, H., & Virrantaus, K. (2015). Shared situational awareness and information quality in disaster management. *Safety Science, 77*, 112–122. https://doi.org/10.1016/j.ssci.2015.03.018.

Sloan, L., & Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS One, 10*(11), 1–15. https://doi.org/10.1371/journal.pone.0142209.

Stock, K. (2018). Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.05.007.

Sutton, J., Gibson, C. B., Phillips, N. E., Spiro, E. S., League, C., Johnson, B., ... Butts, C. T. (2015). A cross-hazard analysis of terse message retransmission on twitter. *Proceedings of the National Academy of Sciences, 112*(48), 14793–14798. https://doi.org/10.1073/pnas.1508916112.

Wang, Q. (2015). Human mobility perturbation and resilience in natural disasters. Retrieved from https://vtechworks.lib.vt.edu/handle/10919/51955.

Xiao, Y., Huang, Q., & Wu, K. (2015). Understanding social media data for disaster management. *Natural Hazards, 79*(3), 1663–1679. https://doi.org/10.1007/s11069-015-1918-0.

Yang, Y., Zhang, C., Fan, C., Yao, W., Huang, R., & Mostafavi, A. (2019). Exploring the emergence of influential users on social media during natural disasters. *International Journal of Disaster Risk Reduction, 38*, 101204. https://doi.org/10.1016/J.IJDRR.2019.101204.

Zhang, C., Fan, C., Yao, W., Hu, X., & Mostafavi, A. (2019). Social media for intelligent public information and warning in disasters: An interdisciplinary review. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2019.04.004.

Zou, L., Lam, N. S. N., Shams, S., Cai, H., Meyer, M. A., Yang, S., ... Reams, M. A. (2018). Social and geographical disparities in Twitter use during Hurricane Harvey. *International Journal of Digital Earth, 0*(0), 1–19. https://doi.org/10.1080/17538947.2018.1545878.