Disaster Medicine and Public Health Preparedness

www.cambridge.org/dmp

Original Research

Cite this article: Muniz-Rodriguez K, Schwind JS, Yin J, Liang H, Chowell G, Fung IC-H. Exploring social media network connections to assist during public health emergency response: A retrospective case-study of Hurricane Matthew and Twitter users in Georgia, USA. *Disaster Med Public Health Prep*. **17**(e315), 1–9. doi: https://doi.org/10.1017/dmp.2022.285.

Keywords:
Twitter; social media; content analysis; natural disasters; location

Corresponding author:
Kamalich Muniz-Rodriguez,
Email: km11200@georgiasouthern.edu.

SDMPH
SOCIETY FOR DISASTER MEDICINE & PUBLIC HEALTH

Exploring Social Media Network Connections to Assist During Public Health Emergency Response: A Retrospective Case-Study of Hurricane Matthew and Twitter Users in Georgia, USA

Kamalich Muniz-Rodriguez DrPH[1,2], Jessica S. Schwind PhD[1], Jingjing Yin PhD[1], Hai Liang PhD[3], Gerardo Chowell PhD[4] and Isaac Chun-Hai Fung PhD[1]

[1]Department of Biostatistics, Epidemiology, and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA, USA; [2]Ponce Research Institute, Ponce Medical School Foundation, Ponce, Puerto Rico; [3]School of Journalism and Communication, Chinese University of Hong Kong, Hong Kong and [4]Department of Population Health Sciences, Georgia State University, Atlanta, Georgia, USA

Abstract

Objective: To assist communities who suffered from hurricane-inflicted damages, emergency responders may monitor social media messages. We present a case-study using the event of Hurricane Matthew to analyze the results of an imputation method for the location of Twitter users who follow school and school districts in Georgia, USA.

Methods: Tweets related to Hurricane Matthew were analyzed by content analysis with latent Dirichlet allocation models and sentiment analysis to identify needs and sentiment changes over time. A hurdle regression model was applied to study the association between retweet frequency and content analysis topics.

Results: Users residing in counties affected by Hurricane Matthew posted tweets related to preparedness ($n = 171$; 16%), awareness ($n = 407$; 38%), call-for-action or help ($n = 206$; 19%), and evacuations ($n = 93$; 9%), with mostly a negative sentiment during the preparedness and response phase. Tweets posted in the hurricane path during the preparedness and response phase were less likely to be retweeted than those outside the path (adjusted odds ratio: 0.95; 95% confidence interval: 0.75, 1.19).

Conclusions: Social media data can be used to detect and evaluate damages of communities affected by natural disasters and identify users' needs in at-risk areas before the event takes place to aid during the preparedness phases.

In their response to natural disasters, emergency management agencies must have access to real-time information to respond to the situation. One potential tool is social media data analysis. In recent years, the usefulness of social media for public health surveillance and their use during natural disasters has been proposed.[1–5] Social media offers emergency management agencies a tool to communicate emergency information, warnings, and updates in their profiles using short messages, photos, and videos.[1,6] Content analysis and sentiment analysis can help classify information extracted from social media messages into different categories and help identify those in need of assistance and the geographical areas affected by an event.[5,7,8]

The possible roles of social media data analysis during natural disasters have been studied before.[1,5] Researchers used social media data analysis to study the content of shared posts during emergencies, identify user's locations, develop mapping applications as a visual aid for emergency responders, and communicate emergency warnings.[3,6,9–13] However, several limitations were identified in the analyses, including a low number of geolocated tweets, large datasets with a reduced number of natural disaster-related posts, and tweets being posted from areas not affected by the disaster.[5] Given such limitation, an imputation method was developed to impute Twitter user geolocations, using the social network connections of Twitter users and the accounts they follow.[14]

We applied such an imputation method to analyze the social media behavior of users who followed schools and school districts in Georgia during Hurricane Matthew. Based on the identified hashtags and information from the National Hurricane Center, Hurricane Matthew was selected as the case study to validate the imputation method used to impute the Twitter users' locations.[15–17] Hurricane Matthew was a category 5 storm that affected the Caribbean islands, Georgia, and North and South Carolina from September 28 to October 9, 2016. The southwest and coastal regions of Georgia were heavily affected, recording winds from a category 2 hurricane.

This retrospective case study, using a secondary dataset, showcased how the imputation method mentioned above can be applied to impute Twitter users' locations and its potential to facilitate the communication efforts of emergency responders if applied in real-time. This study aims: (1) to describe the topics and sentiment of Twitter users who follow schools' and school districts' accounts in Georgia before and during Hurricane Matthew; and (2) to evaluate the association between retweet frequency and topics posted by Twitter users during Hurricane Matthew.

Methods

Data Collection

The analysis uses secondary data from the social media platform Twitter, as described in Ahweyevu et al.[18] Ahweyevu and collaborators downloaded publicly available public school and school districts data for the state of Georgia from the National Center for Education Statistics (NCES) (nces.edu.gov) and identified Twitter profiles for the schools and school districts.[18] For details on the data collection process, refer to Ahweyevu et al.[18]

Missing Data Imputation

We developed a method to impute the information of location for Twitter users who do not share their self-reported locations in their profiles.[14] A location at the Metropolitan Statistical Area (MSA) level, in Georgia, USA, was assigned to Twitter users who did not share any location or a real location. An MSA is defined as a region with a minimum of 1 community with at least 50,000 people.[19] There are 14 MSAs in Georgia. The public schools' and public school districts' Twitter accounts in these MSAs were identified.[18] The imputation method used the follower- "followee" relation as a proxy to impute a location to users.

The total sample size was 27,598 followers from 53 school or district accounts.[14] The analysis presented in this article used the sample of Twitter users and their imputed locations to explore their social media behavior during Hurricane Matthew.

Selection of Hurricane-Related Tweets

The hurricane-related tweets from users in the imputed sample were extracted by the keywords of "hurricane" and "hurricanes". From a total of 26,274 hurricane-related tweets extracted, 3,753 tweets were posted during Hurricane Matthew, from September 28 to October 9, 2016. Three datasets were created to analyze the tweet content shared by the users. Dataset 1 comprised only those tweets considered original content posted by the users (1,679 tweets). Dataset 2 included tweets identified as retweets in the sample (2,033 tweets), and dataset 3 contained replies to tweets in the sample (41 tweets). Given its very small size, dataset 3 was excluded from further analysis.

Content Analysis

Content analysis was done to describe the topics mentioned by Twitter users who followed schools' and school districts' accounts in Georgia before and during Hurricane Matthew. The steps were repeated for original content tweets and retweets to assess the differences in content per type of Twitter post and for counties in the actual hurricane path. We implemented a probabilistic topic model known as the latent Dirichlet allocation (LDA) model, which is a Bayesian mixture model[20] to determine the importance of a term in the analyzed text corpus.[21] The LDA model was trained using 90% of the dataset in this project, and the model was tested using the remaining 10% percent of data.[22,23] Before model fitting, the number of topics (k) was determined by running model simulations with $k = 5$ to $k = 100$ in the increment of 5 units,[24] with 30 iterations, using the training datasets to assess the value of k. The optimal number of topics for dataset 1 (original tweets) and that for dataset 2 (retweets) were both 30 topics.

Sentiment Analysis

Sentiment analysis was applied to describe the sentiment of Twitter users who followed schools' and school districts' accounts in Georgia before and during Hurricane Matthew. A lexicon-approach method was implemented to calculate the average sentiment of words in the tweets.[25] Two different lexicon libraries, Afinn and Bing,[25] were compared in their evaluations in a preliminary analysis and the Afinn lexicon was found to be the more preferred library and thus the following analysis used the sentiment scores based on Afinn. Next, general descriptive frequencies were studied for original tweets and retweets. Finally, the overall changes in sentiment scores were plotted over time.

Hurdle Regression Model to Evaluate the Association Between Retweet Frequency and Tweet Topics

We fitted hurdle regression models to evaluate the association between retweet frequency and topics posted by Twitter users during Hurricane Matthew. The response variable, the number of retweets a tweet received, was analyzed in association with the independent variable topic categories obtained from content analysis and US Census demographic data as covariates.[26] The hurdle model was divided into 2 components. The first part was a zero-mass component model that determined the chance of having a zero number of retweets. The second part of the model was a truncated Poisson model that considered only the positive retweet counts to determine the likelihood ratio of having higher number of retweets.[27,28] The level of significance was specified as 0.05 a priori.

Results

Descriptive Statistics

Hurricane-related tweets were identified through their hashtags ($n = 168,184$). "Hurricanemaria" ($n = 16,346$; 0.10%), "Hurricaneharvey" ($n = 12,728$; 0.08%), and "Hurricanematthew" ($n = 11,508$; 0.07%) were identified as the 3 most common hurricane-related hashtags in the tweets collected from followers of schools and school districts in Georgia. Observing tweet frequency and time of posting, our analysis focused on major hurricanes in the Atlantic region and those that directly affected the state of Georgia (Supplementary Materials, Figure S1).

Description of the Topics and Sentiment of Tweets From Users Who Followed Schools' and School Districts' Accounts in Georgia Before and During Hurricane Matthew

The topics identified by the LDA model in each dataset were manually categorized into 10 different categories (Table 1). The top 3 categories of tweets were "awareness," "preparedness," and "call for help or action" for original tweets and retweets datasets (Supplementary Materials, Table S6). Users in the Hinesville MSA, 1 of the MSAs in the hurricane path, posted the highest

Table 1. Number (%) of tweets by content analysis category for MSAs in or out of Hurricane Matthew's path posted by followers of schools and school districts in Georgia, USA, during Hurricane Matthew

		Tweet in the hurricane path	
		Yes	No
Content analysis categories	Definition	No. of tweets (%)	No. of tweets (%)
Awareness	Topics related to hurricanes information	407 (38%)	180 (37%)
Call for help or action	Topics related to asking for help or action from individuals or government	206 (19%)	101 (21%)
Preparedness	Topics related to preparation before the event	171 (16%)	81 (16%)
Evacuation or Migration	Topics with words related to moving from the area or evacuations	93 (9%)	57 (12%)
Damage	Topics related to any structural damage	73 (7%)	34 (7%)
Warnings	Topics related to emergency warnings	34 (3%)	0
Miscellaneous	Topics that do not fit in any other designated category	32 (3%)	0
Emotions or Religious	Topics related to any type of emotion or religion	30 (3%)	19 (4%)
Shelter	Topics related to shelter needs	24 (2%)	20 (4%)

number of original tweets related to preparing for the weather event (Supplementary Materials, Figure S4).

When focusing on the emergency cycle phases, it was found that most original tweets were posted during the preparedness phase of the emergency response cycle and were mainly associated with content categories "preparedness," "awareness," and "call for action or help." Original tweets posting frequency decreased during the response phase, but high numbers of "awareness" tweets and "call for help or action" tweets were found. Compared with prior phases, the response phase saw the least number of tweets captured in the dataset, however, the "awareness" category as the most identified one (Figure 1). When focusing on the retweets during Hurricane Matthew, it was observed that all categories had a higher number of tweets during the emergency cycle's preparedness phase than other phases, with "awareness," "call for help or action," and "preparedness" as the 3 most common categories (Figure 2).

Analysis of tweet count by MSA during Hurricane Matthew reflected a spike in tweet frequency was observed near the end of the preparedness phase of the emergency response cycle for all MSAs. Original tweet signal decreased as the response phase started, with the lowest number of original tweets detected during the recovery phase for all MSAs. Savannah and Hinesville MSAs had the highest number of original tweets during the recovery phase (Table 1).

The sentiment changes throughout all phases of the emergency response cycle presented a decrease in sentiment value, accompanied by a decline in the number of Twitter posts related to Hurricane Matthew. On September 28, both original tweets and retweets reflected a positive sentiment score. On this day, the National Hurricane Center declared the development of the weather event as a tropical storm Matthew.[29] Overall, among both original tweets and retweets, an increase in negative sentiment through the preparedness phase was observed with a change to an increase in positive sentiment during the response phase. As the day of landfall in Georgia approached, negative sentiment values increased. The days after hurricane landfall, overall sentiment started to show more positive values for original tweets and retweets (Supplementary Materials, Figure S5; Figure S6).

Hurdle Regression Model to Evaluate the Association Between Retweet Frequency and Content Categories Posted by Twitter Users During Hurricane Matthew

A multivariable hurdle regression model was adjusted for confounding variables to evaluate the association between retweet frequency and Twitter content categories (Table 2). The logistic model component presents the adjusted odds ratio (aOR) of a tweet being retweeted; the truncated Poisson model component presents the adjusted risk ratio (aRR) of retweet count if retweeted. As seen in Table 3, compared with tweets in the preparedness category, tweets in the hurricane damage category were less likely to be retweeted (aOR: 0.84; 95% confidence interval [CI], 0.63, 1.12); however, if retweeted, they were retweeted 53% more (aRR: 1.53; 95% CI: 1.52, 1.53). Likewise, tweets in the awareness category were less likely to be retweeted (aOR: 0.83; 95% CI, 0.69, 1); however, if retweeted, they were retweeted 74% more (aRR: 1.74; 95% CI, 1,74, 1.74). Similarly, tweets "calling for help" were 30% less likely to be retweeted (aOR: 0.7; 95% CI: 0.57, 0.85); if retweeted, the retweet count was estimated to increase by 1.62 (95% CI: 1.61, 1.62) compared with tweets in the preparedness category. Location is important when studying Twitter behavior. If the user who posted the tweet was in Hurricane Matthew's path, their tweet's probability of being retweeted was reduced by 5% (aOR: 0.95; 95% CI: 0.75, 1.19), and if it was retweeted, its retweet count was reduced by 89% (aRR: 0.11; 95% CI: 0.11, 0.11) (Table 3).

When we stratified our data by the phase of the emergency cycle, our results demonstrated that the timing of the tweet (in terms of the phase of the emergency cycle) was an important factor to consider in social media analysis for emergency response. If a tweet was posted during the preparedness phase and was published in the path of the hurricane, it was 1.16 (95% CI: 0.85, 1.58) times as likely to be retweeted, and if the post was retweeted, being posted from the hurricane path reduced the retweet count by 91% (aRR 0.09; 95% CI: 0.09, 0.09). During the preparedness phase of the emergency cycle, the retweet count of tweets in the "damage" category, if retweeted, was 73% more than tweets in the preparedness category (aRR, 1.73; 95% CI, 1.72, 1.73); the retweet count for tweets in the "call for help or action" category was estimated to increase by 1.40 (95% CI: 1.39, 1.40) compared with tweets in the preparedness category when retweeted. Also compared with the retweet count of tweets retweeted in the preparedness category, if retweeted, tweets posted in the warning category was 1.89 (95% CI: 1.88, 1.89) times in their retweet count, those in the shelter category was 1.82 (95% CI: 1.81, 1.82) times in their retweet count, and those in the emotion or religious categories was 1.58 (95% CI: 1.58, 1.59) times in their retweet count (Table 4). When analyzing the same model with tweets only posted during the response phase of the emergency cycle, it was observed that those users who

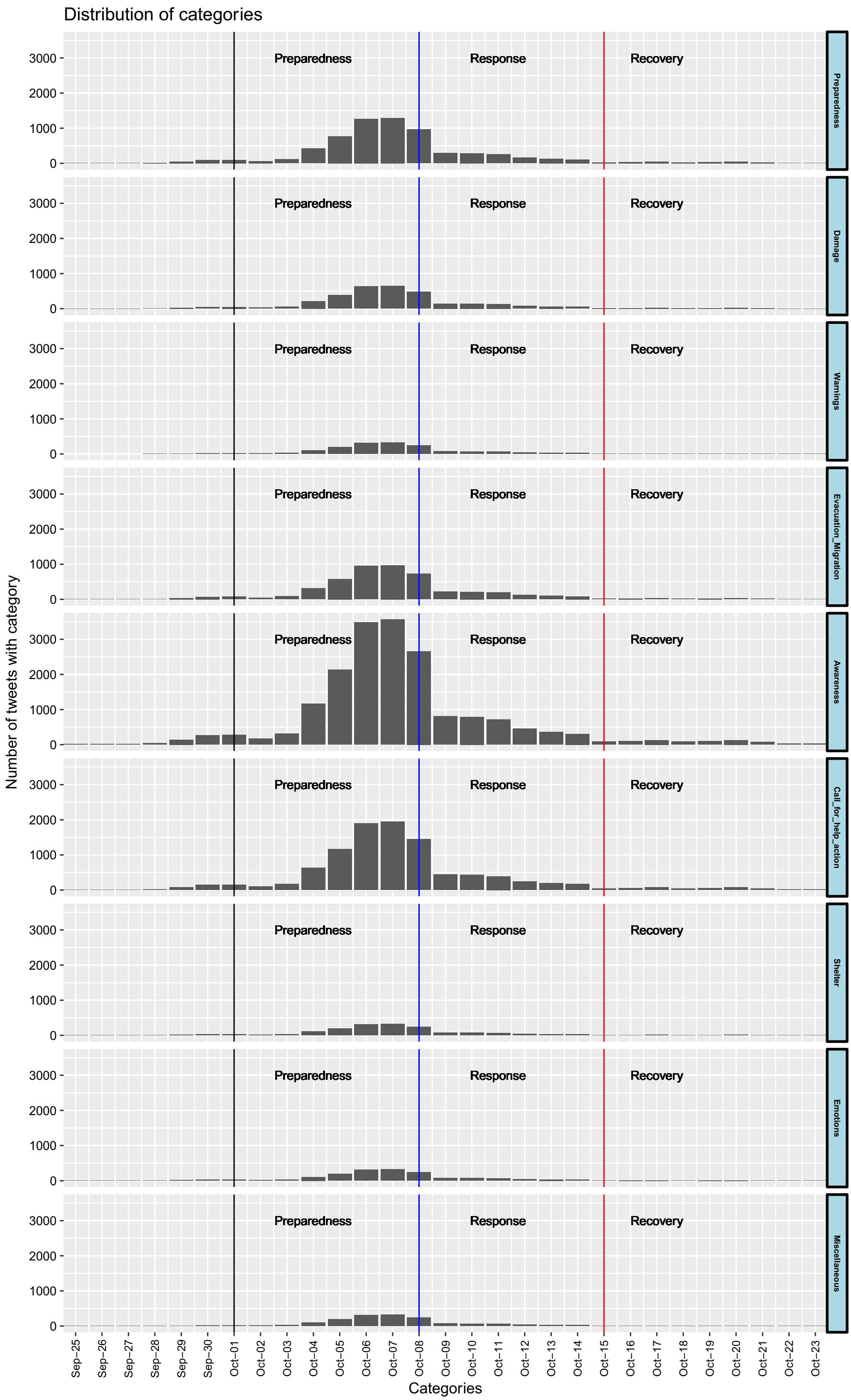

Figure 1. Distribution of number of original tweets by category and emergency management cycle phase during Hurricane Matthew posted by followers of schools and school districts in Georgia, USA. The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]

https://doi.org/10.1017/dmp.2022.285 Published online by Cambridge University Press

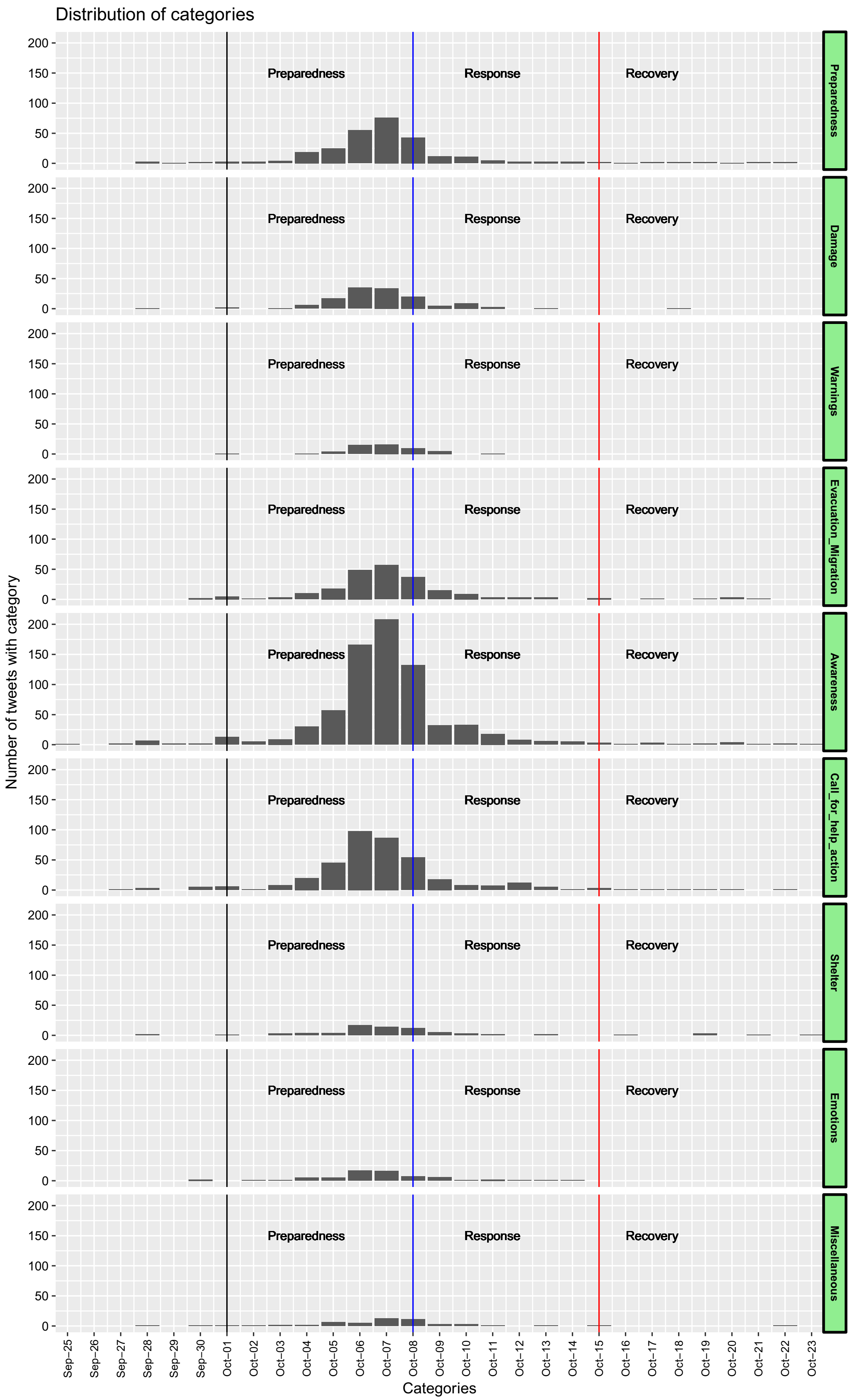

Figure 2. Distribution of number of retweets by category and emergency management cycle phase during Hurricane Matthew posted by followers of schools and school districts in Georgia, USA. The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]

https://doi.org/10.1017/dmp.2022.285 Published online by Cambridge University Press

Table 2. Content analysis categories by emergency response cycle phase for tweets during Hurricane Matthew in Georgia, USA analyzed in the hurdle regression model (logistic model for the probability of being retweeted and Poisson model for the positive retweet count)

	Preparedness + response + recovery phases			Preparedness phase			Response phase		
Content analysis categories	% RT	RT total	RT median	% RT	RT total	RT median	% RT	RT total	RT median
Preparedness	14.8%	732	14.8	14.4%	455	0.35	15.9%	252	0.37
Damage	5.9%	294	5.9	5.6%	176	0.23	6.7%	107	0.25
Warnings	2.5%	126	2.5	2.4%	75	0.15	3.2%	51	0.18
Evacuations or migrations	9.6%	473	9.6	9.6%	305	0.3	9.4%	150	0.29
Awareness	37.5%	1856	37.5	37.9%	1199	0.49	36.9%	587	0.48
Call for help or action	20.8%	1028	20.8	21.4%	676	0.41	19.3%	307	0.39
Shelter	3.1%	152	3.1	3.0%	94	0.17	2.8%	44	0.16
Emotions	3.1%	155	3.1	3.1%	98	0.17	3.1%	49	0.17
Miscellaneous	2.6%	128	2.6	2.7%	84	0.16	2.6%	42	0.16

Note: The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]
Abbreviation: RT, retweet.

Table 3. Association between content analysis categories and retweet count of tweets tweeted in the preparedness, response, and recovery phases of Hurricane Matthew in Georgia, USA, as given by the hurdle regression model (logistic model for the probability of being retweeted and Poisson model for the positive retweet count)

	Zero hurdle model coefficients (binomial with logit link)			Count model coefficients (truncated Poisson with log link)		
Coefficients	aOR	aOR 95% CI	*P*-Value	aRR	aRR 95% CI	*P*-Value
Intercept	11.01	(7.57, 16.00)	<0.001	7690.36	(7663.53, 7717.28)	<0.001
In the hurricane path	0.95	(0.75, 1.19)	0.66	0.11	(0.11, 0.11)	<0.001
Content analysis categories						
Preparedness	REF					
Damage	0.84	(0.63, 1.12)	0.23	1.53	(1.52, 1.53)	<0.001
Awareness	0.83	(0.69, 1)	0.05	1.74	(1.74, 1.74)	<0.001
Call for help or action	0.7	(0.57, 0.85)	<0.001	1.62	(1.61, 1.62)	<0.001
Warnings	0.91	(0.61, 1.36)	0.65	1.02	(1.02, 1.03)	<0.001
Evacuation	1.1	(0.82, 1.36)	0.68	1.37	(1.37, 1.38)	<0.001
Shelter	1.28	(0.86, 1.90)	0.22	1.71	(1.70, 1.71)	<0.001
Emotions or religious	1.27	(0.86, 1.88)	0.23	1.56	(1.55, 1.56)	<0.001
Miscellaneous	0.85	(0.57, 1.26)	0.42	0.36	(0.36, 0.37)	<0.001
Percent of poverty level	0.94	(0.91, 0.96)	<0.001	0.95	(0.94, 0.95)	<0.001
Percentage of not owning a car in the house	0.93	(0.87, 0.97)	0.01	1.09	(1.08, 1.09)	<0.001
Percentage of mobile homes	1.01	(0.99, 1.02)	0.42	0.98	(0.97, 0.97)	<0.001

Note: The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]
Abbreviations: aOR, adjusted odds ratio; aRR, adjusted relative risk; CI, confidence interval; REF, reference category.

resided in counties in the path of the hurricane were 9% less likely (aOR: 0.91; 95% CI: 0.63, 1.32) of being retweeted, and if retweeted, the retweet count was lowered by 74% (aRR: 0.26; 95% CI: 0.26, 0.26) (Table 5).

Discussion

This case-study incorporates the results from a new imputation method of Twitter users' locations[14] into a retrospective analysis of Hurricane Matthew-related Twitter corpus. The analysis identified higher tweet frequency in the preparedness phase and a decline in tweets after the response phase. Also, the results showed that tweets posted by those in the actual path of the hurricane and those in low-income areas were less likely to be retweeted, presenting a challenge if help is needed in these areas. Our results highlight the strengths and limitations of Twitter data analysis for public health emergency response.

The literature suggests that less than 1% of Twitter users share their exact geolocations with geographical coordinates and that users with privacy settings share their location when they feel safe.[4,32] The lack of geolocated data presents a challenge for public health agencies interested in harvesting social media information for emergency response purposes. Our analysis uses the locations of schools and school districts with Twitter accounts as a proxy for user location, imputing the location of 67.0% of the sample.[14] Public health agencies can use this newly available information

https://doi.org/10.1017/dmp.2022.285 Published online by Cambridge University Press

Table 4. Association between content analysis categories and retweet count of tweets tweeted in the preparedness phase of Hurricane Matthew in Georgia, USA, as given by the hurdle regression model (logistic model for the probability of being retweeted and Poisson model for the positive retweet count)

	Zero hurdle model coefficients (binomial with logit link)			Count model coefficients (truncated Poisson with log link)		
Coefficients	aOR	aOR 95% CI	*P*-Value	aRR	aRR 95% CI	*P*-Value
Intercept	9.39	(5.87, 15.01)	<0.001	10463	(10421.00, 10506.00)	<0.001
In the hurricane path	1.16	(0.85, 1.58)	0.343	0.09	(0.09, 0.09)	<0.001
Content analysis categories						
Preparedness	REF					
Damage	0.92	(0.63, 1.35)	0.684	1.73	(1.72, 1.73)	<0.001
Awareness	0.93	(0.55, 1.58)	0.792	0.98	(0.97, 0.98)	<0.001
Call for help or action	1.05	(0.76, 1.44)	0.780	1.40	(1.39, 1.40)	<0.001
Warnings	0.83	(0.65, 1.04)	0.109	1.89	(1.88, 1.89)	<0.001
Evacuation	0.74	(0.57, 0.95)	0.019	1.63	(1.63, 1.64)	<0.001
Shelter	1.49	(0.88, 2.52)	0.133	1.82	(1.81, 1.82)	<0.001
Emotions or religious	1.11	(0.68, 1.81)	0.669	1.58	(1.58, 1.59)	<0.001
Miscellaneous	0.51	(0.32, 0.82)	0.006	0.23	(0.23, 0.23)	<0.001
Percent of poverty level	0.96	(0.93, 0.99)	0.047	0.89	(0.89, 0.89)	<0.001
Percentage of not owning a car in the house	0.89	(0.83, 0.95)	<0.001	1.14	(1.14, 1.14)	<0.001
Percentage of mobile homes	1.01	(0.98, 1.03)	0.603	1.01	(1.01, 1.01)	<0.001

Note: The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]
Abbreviations: aOR, adjusted odds ratio; aRR, adjusted relative risk; CI, confidence interval; REF, reference category.

Table 5. Association between content analysis categories and retweet count of tweets tweeted in the response phase of Hurricane Matthew in Georgia, USA, as given by the hurdle regression model (logistic model for the probability of being retweeted and Poisson model for the positive retweet count)

	Zero hurdle model coefficients (binomial with logit link)			Count model coefficients (truncated Poisson with log link)		
Coefficients	aOR	aOR 95% CI	*P*-Value	aRR	aRR 95% CI	*P*-Value
Intercept	12.15	(5.96, 24.77)	<0.001	933.88	(924.67, 943.18)	<0.001
In the hurricane path	0.91	(0.63, 1.32)	0.63	0.26	(0.26, 0.26)	<0.001
Content analysis categories						
Preparedness	REF					
Damage	0.78	(0.49, 1.25)	0.31	0.8	(0.80, 0.81)	<0.001
Awareness	0.97	(0.51, 1.83)	0.92	1.26	(1.26, 1.28)	<0.001
Call for help or action	1.23	(0.79, 1.92)	0.35	1.27	(0.97, 0.98)	<0.001
Warnings	0.98	(0.72, 1.34)	0.91	0.97	(0.99, 1.01)	<0.001
Evacuation	0.76	(0.53, 1.08)	0.12	1.27	(1.27, 1.28)	<0.001
Shelter	0.97	(0.49, 1.90)	0.92	1	(0.99, 1.01)	0.97
Emotions or religious	1.46	(0.73, 2.92)	0.28	1.75	(1.173, 1.76)	<0.001
Miscellaneous	2.81	(1.19, 6.65)	0.02	0.64	(0.63, 0.65)	<0.001
Percent of poverty level	0.93	(0.89, 0.98)	0.01	1.13	(1.13, 1.13)	<0.001
Percentage of not owning a car in the house	0.91	(0.83, 1.01)	0.08	0.95	(0.95, 0.96)	<0.001
Percentage of mobile homes	1.00	(0.97, 1.03)	0.93	0.89	(0.89, 0.90)	<0.001

Note: The timeframe for each response phase was determined based on the reviewed literature, the emergency cycle phases, and the official FEMA incident period for Hurricane Matthew in Georgia (October 4, 2016, to October 15, 2016).[5,30,31]
Abbreviations: aOR, adjusted odds ratio; aRR, adjusted relative risk; CI, confidence interval; REF, reference category.

to understand the needs, worries, and awareness of individuals residing in the MSA included in our analysis.

This study analyzed Twitter data and observed its possible uses as a tool by emergency response agencies during the preparedness and response phases. The "awareness" category was identified as the most frequent category in both original (37.64%) and retweeted 37.0)%) content associated with Hurricane Matthew. The majority of tweets in the "awareness" category were related to weather information pertinent to Hurricane Matthew. The identification of the "awareness" category as the most common content category in the sample was consistent with findings of social media data analysis during flooding and earthquake events.[5,9,33–35] Other common

https://doi.org/10.1017/dmp.2022.285 Published online by Cambridge University Press

content categories were "preparedness" and "call for help or action." A higher number of retweets from the "damage" category were detected during the response phase than the preparedness phase. An increase in negative sentiment as the hurricane approached the state was observed in the results. A similar pattern was observed during Hurricane Sandy.[36] A change to more positive sentiment, expressing hope through religious language, was detected after landfall.

The analysis identified a higher number of original tweets and retweets pertinent to Hurricane Matthew during the preparedness and response phases than the other cycle stages, with tweets peaking days before the hurricane landfall. Similar to the results found by other social media researchers, a low number of tweets were posted after landfall and during the recovery phase in our sample.[37,38] It is understood that the low number of tweets found during the recovery and mitigation phases establishes that Twitter does not present as a viable tool to study for long-term follow-up of areas affected by natural disasters. Previous research found that most social media communication from emergency management agencies is 1-sided, meaning the agency does not interact with their followers.[5,13] The increased number of tweets observed during the preparedness phase of the emergency can represent an increased awareness of the event, and public health professionals can take this opportunity to perform communication campaigns to help alleviate the information gap.

Retweeted content can help information go viral, and their role in social media communication strategies has been studied. For example, Liang et al. found that on Twitter, Ebola-related information primarily reached a user's followers (the "broadcast model"). To make a tweet retweeted beyond the immediate group of followers, having individuals who have many followers (such as celebrities) to retweet a public health agency's tweet may be a key. This suggests that the identities of Twitter users and their followers can influence the reach of a tweet.[39] This study did not find that celebrities were the most retweeted accounts in our sample; instead, individual personal accounts were more frequently retweeted, contrary to other studies.[13,39,40] Higher Twitter activity levels were observed in geographical areas (MSAs) outside of the hurricane path, contrary to other studies.[33,41] Twitter users outside the hurricane path and those in the hurricane path posted tweets related to "awareness" and "call for action or help," which can be driven by the news cycle and proximity of the storm.[36] Users in the hurricane path are less likely to be retweeted than those outside the hurricane path. Therefore, the development of a content analysis guide for training is highly recommended. For example, it may include a step-by-step checklist to complete the analysis, what questions can be answered, and specialists that can assist in the analysis if necessary.

The regression modeling results suggested no evidence to support the hypothesis that higher levels of hurricane-related Twitter activity are associated with the actual hurricane path. During the emergency response phase, the results demonstrated that original tweets that were retweeted from low-income areas had an increased retweet count as the poverty percentage in the area increased (albeit statistically insignificant). This can help emergency responders quickly identify those that could have been heavily affected by the event.

Public Health Implications

This case-study demonstrates that retrospective Twitter data analysis can provide emergency response agencies with insights into the needs of social media users who might be affected by natural disasters. However, it is important to recognize that the analysis is time-consuming. It is difficult to make all the data identification, data cleaning and processing, and content and sentiment analyses in real-time. Therefore, to apply this type of analysis in practice, it is recommended to conduct data verification before the start of the Atlantic hurricane season or during the planning phase of emergency management agencies to avoid delays in the emergency communication response.

Strengths and Limitations

There are several limitations to this study. The results are not generalizable to the general population of the state of Georgia. The findings only apply to Twitter followers of schools and school districts in our sample. Also, the user locations analyzed in our study were based on the locations of the schools or school districts they followed. We were not able to verify the veracity of the locations at the time of the analysis. Results were based on post frequency; network analyses for information dissemination were not conducted.

Public health researchers previously employed the dataset used in this case study to detect unplanned school closures, establishing the social media platform's usefulness to detect a higher number of school closures than the current systems.[18,42] The analysis presented in this research project gave an existing dataset a new purpose, demonstrating how we can repurpose public health datasets from 1 field into a completely new area.

Conclusions

In times where social media is a core component of public health interventions, emergency response should not be the exception. Despite not being able to pinpoint a location if the social media user does not share coordinates, our results showed that our imputation method could help impute users' geolocations and, thereby, through Twitter data analysis, help provide an overview of the situation in areas affected by natural disasters. It can help understand the needs of social media users in at-risk areas before the event takes place. Future research to further test the imputation method should focus on official emergency response agencies' pages and their followers.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/dmp.2022.285.

Acknowledgment. This manuscript is part of K.M.R.'s doctoral dissertation project, titled "Social Media Data Analysis, a Tool for Public Health Emergency Management During Natural Disasters," Fall 2020. K.M.R. thanks the Department of Biostatistics, Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, for her graduate assistantship. The authors thank Bryan Omar Sepulveda Bahamundi for his work and assistance in the content analysis portion for this project.

Funding. No external funding reported.

Conflict of interest. No conflict of interest declared.

IRB statement. This research project was approved by the Georgia Southern University Institutional Review Board with a B2 exemption category for the use of Twitter data with project number H15083.

References

1. **Finch KC, Snook KR, Duke CH, *et al.*** Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. *Nat Hazards (Dordr).* August 01 2016;83(1): 729-760. doi: 10.1007/s11069-016-2327-8
2. **Kabir AI, Karim R, Newaz S, *et al.*** The power of social media analytics: text analytics based on sentiment analysis and word clouds on R. *Informatica Economica.* 2018;22(1):25-38. doi: 10.12948/issn14531305/22.1.2018.03
3. **Sherchan W, Pervin S, Butler CJ, *et al.*** Harnessing Twitter and Instagram for disaster management. *IBM J Res Dev.* 2017. doi: 10.1147/JRD.2017.2729238
4. **Fu K-w, White J, Chan Y-y, *et al.*** Enabling the disabled: media use and communication needs of people with disabilities during and after the Sichuan earthquake in China. *Int J Emerg Manag.* 2010/01/01 2010; 7(1):75-87. doi: 10.1504/IJEM.2010.032046
5. **Muniz-Rodriguez K, Ofori SK, Bayliss LC, *et al.*** Social media use in emergency response to natural disasters: a systematic review with a public health perspective. *Disaster Med Public Health Prep.* 2020;14(1):139-149. doi: 10.1017/dmp.2020.3
6. **Kim J, Hastak M.** Social network analysis: characteristics of online social networks after a disaster. *Int J Inform Manag.* 2018;38(1):86-96. doi: 10.1016/j.ijinfomgt.2017.08.003
7. **Adams J, Raeside R, Khan HTA.** *Research Methods for Business and Social Science Students.* 2nd edition. Sage Publications Pvt. Ltd; 2014.
8. **Kiatpanont R, Tanlamai U, Chongstitvatana P.** Extraction of actionable information from crowdsourced disaster data. *J Emerg Manag.* 2016 2016;14(6):377-390. doi: 10.5055/jem.2016.0302
9. **Andrews S, Gibson H, Domdouzis K, *et al.*** Creating corroborated crisis reports from social media data through formal concept analysis. *J Intell Inf Syst.* 2016;47(2):287-312. doi: 10.1007/s10844-016-0404-9
10. **Brandt HM, Turner-McGrievy G, Friedman DB, *et al.*** Examining the role of Twitter in response and recovery during and after historic flooding in South Carolina. *J Public Health Manag Pract.* 2019;25(5):E6-E12. doi: 10.1097/phh.0000000000000841
11. **Cervone G, Sava E, Huang Q, *et al.*** Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *Int J Remote Sens.* 2016;37(1):100-124. doi: 10.1080/01431161.2015.1117684
12. **Kaufhold M-A, Reuter C.** The self-organization of digital volunteers across social media: the case of the 2013 European floods in Germany. *J Homel Secur Emerg Manag.* 2016;13(1):137-166. doi: 10.1515/jhsem-2015-0063
13. **Tang Z, Zhang L, Xu F, *et al.*** Examining the role of social media in California's drought risk management in 2014. *Nat Hazards.* Oct 2015; 79(1):171-193. doi: 10.1007/s11069-015-1835-2
14. **Muniz-Rodriguez K.** *Social Media Data Analysis, a Tool for Public Health Emergency Management During Natural Disasters.* Electronic Theses and Dissertations. Georgia Southern University; 2020. https://digitalcommons.georgiasouthern.edu/etd/2175
15. **National Hurricane Center**. 2016 Atlantic Hurricane Season. National Oceanic and Atmospheric Administration. Accessed August 3, 2020. https://www.nhc.noaa.gov/data/tcr/index.php?season=2016&basin=atl
16. **National Hurricane Center**. 2017 Atlantic Hurricane Season. National Oceanic and Atmospheric Administration. Accessed August 3, 2020. https://www.nhc.noaa.gov/data/tcr/index.php?season=2017&basin=atl
17. **National Hurricane Center**. Glossary of NHC Terms. National Oceanic and Atmospheric Administration. Accessed October 11, 2020. https://www.nhc.noaa.gov/aboutgloss.shtml
18. **Ahweyevu JO, Chukwudebe NP, Buchanan BM, *et al.*** Using Twitter to track unplanned school closures: Georgia public schools, 2015-17. *Disaster Med Public Health Prep.* 2021;15(5):568-572. doi: 10.1017/dmp.2020.65
19. **United States Census Bureau**. About. Updated October 15, 2018. Accessed October 25, 2019. https://www.census.gov/programs-surveys/metro-micro/about.html
20. **Grün B, Hornik K.** topicmodels: An R package for fitting topic models. Cran.R-Project. Accessed July 15, 2020. https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf
21. **Blei DM.** Probabilistic topic models. *Commun ACM.* 2012;55(4):77-84. doi: 10.1145/2133806.2133826
22. **Ghatak A.** *Machine Learning with R.* Springer; 2017:224.
23. **Kumar A.** *Mastering Text Mining with R.* 1st ed. Packt Publishing Ltd; 2016.
24. **Adnan MM, Yin J, Jackson AM, *et al.*** World Pneumonia Day 2011-2016: Twitter contents and retweets. *Int Health.* 2019;11(4):297-305. doi: 10.1093/inthealth/ihy087
25. **Silge J, Robinson D.** *Text Mining with R: A Tidy Approach.* O'Reilly Media; 2019. https://www.tidytextmining.com
26. **United States Census Bureau**. Country profile: Georgia. Accessed August 20, 2020. https://data.census.gov/cedsci/table?q=United%20States&g=0400000US13_310M200US10500,12060,15260,25980_310M300US12020,17980,19140,23580,25980,31420,40660,42340,46660,47580&y=2016&tid=ACSST1Y2016.S0101&hidePreview=false
27. **Love TE.** Data science for biological, medical and health research: notes for 432. Updated May 1, 2018. Accessed September 5, 2020. https://thomaselove.github.io/432-notes/index.html
28. **Rodriguez G.** Mean and variance in models for count data. Princeton University. Accessed September 13, 2020. https://data.princeton.edu/wws509/notes/countmoments
29. **National Hurricane Center**. Hurricane Matthew (AL142016). Updated April 7, 2017. Accessed September 1, 2020, 2020. https://www.nhc.noaa.gov/data/tcr/AL142016_Matthew.pdf
30. **Federal Emergency Management Agency**. The four phases of emergency management Accessed December 11, 2018. https://training.fema.gov/emiweb/downloads/is10_unit3.doc
31. **Federal Emergency Management Agency**. Georgia Hurricane Matthew (DR-4284-GA). FEMA. Updated March 20, 2020. Accessed August 31, 2020. https://www.fema.gov/disaster/4284
32. **Liang H, Shen F**, Fu K-w. Privacy protection and self-disclosure across societies: a study of global Twitter users. *New Media Society.* 2016; 19(9):1476-1497. doi: 10.1177/1461444816642210
33. **Grasso V, Crisci A.** Codified hashtags for weather warning on Twitter: an Italian case study. *PLoS Curr.* 2016. doi: 10.1371/currents.dis.967e71514ecb92402eca3bdc9b789529
34. **Yuan F, Liu R.** Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. *Int J Disaster Risk Reduct.* 2018;28:758-767. doi: 10.1016/j.ijdrr.2018.02.003
35. **Kryvasheyeu Y, Chen H, Obradovich N, *et al.*** Rapid assessment of disaster damage using social media activity. *Sci Adv.* 2016;2(3):e1500779. doi: 10.1126/sciadv.1500779
36. **Zou L, Lam NSN, Cai H, *et al.*** Mining Twitter data for improved understanding of disaster resilience. *Ann Am Assoc Geographers.* 2018;108(5): 1422-1441. doi: 10.1080/24694452.2017.1421897
37. **David CC, Ong JC, Legara EF.** Tweeting Supertyphoon Haiyan: evolving functions of Twitter during and after a disaster event. *PLoS One.* 2016;11(3): e0150190. doi: 10.1371/journal.pone.0150190
38. **Kim J, Bae J, Hastak M.** Emergency information diffusion on online social media during storm Cindy in US. *Int J Inf Manag.* 2018;40:153-165. doi: 10.1016/j.ijinfomgt.2018.02.003
39. **Liang H, Fung IC, Tse ZTH, *et al.*** How did Ebola information spread on twitter: broadcasting or viral spreading? *BMC Public Health.* 2019/04/25 2019;19(1):438. doi: 10.1186/s12889-019-6747-8
40. **Comunello F, Parisi L, Lauciani V, *et al.*** Tweeting after an earthquake: user localization and communication patterns during the 2012 Emilia seismic sequence. *Ann Geophys.* 2016;59(5). doi: 10.4401/ag-6945
41. **Zahra K, Ostermann FO, Purves RS.** Geographic variability of Twitter usage characteristics during disaster events. *Geo Spat Inf Sci.* 2017;20(3):231-240. doi: 10.1080/10095020.2017.1371903
42. **Jackson AM, Mullican LA, Tse ZTH, *et al.*** Unplanned closure of public schools in Michigan, 2015-2016: cross-sectional study on rurality and digital data harvesting. *J Sch Health.* 2020;90(7):511-519.