IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 29, 2019, accepted April 23, 2019, date of publication May 6, 2019, date of current version May 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914797

Twitter and Disasters: A Social Resilience Fingerprint

BENJAMIN A. RACHUNOK[1], JACKSON B. BENNETT[1,2], AND ROSHANAK NATEGHI[1,2,3]
[1]School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA
[2]Ecological Science and Engineering Department, Purdue University, West Lafayette, IN 47907, USA
[3]Environmental and Ecological Engineering Department, Purdue University, West Lafayette, IN 47907, USA

Corresponding author: Benjamin A. Rachunok (brachuno@purdue.edu)

This work was supported by the U.S. National Science Foundation under Grant 1826161 and Grant 1728209.

ABSTRACT Understanding the resilience of a community facing a crisis event is critical to improving its adaptive capacity. Community resilience has been conceptualized as a function of the resilience of components of a community such as ecological, infrastructure, economic, and social systems, etc. In this paper, we introduce the concept of a "resilience fingerprint" and propose a multi-dimensional method for analyzing components of community resilience by leveraging existing definitions of community resilience with data from the social network Twitter. Twitter data from 14 events are analyzed and their resulting resilience fingerprints computed. We compare the fingerprints between events and show that major disasters such as hurricanes and earthquakes have a unique resilience fingerprint which is consistent between different events of the same type. Specifically, hurricanes have a distinct fingerprint which differentiates them from other major events. We analyze the components underlying the similarity among hurricanes and find that ecological, infrastructure and economic components of community resilience are the primary drivers of the difference between the community resilience of hurricanes and other major events.

INDEX TERMS Data analysis, human computer interaction, resilience, Twitter.

I. INTRODUCTION

There is a temporal trend toward more frequent and more unexpectedly intense natural disasters [1], [2]. To prepare for uncertain future disasters, it is fundamental to question what constitutes a *resilient community* so as to build a body of knowledge useful in enhancing communities' adaptive capacity in the face of the next generation of unforeseen disasters. Resilience is a concept with multiple definitions, all of which stem from understanding how elements of a community protect against, respond to, and recover from a disruption [3]–[10]. At their core, these definitions establish how an exogenous disruption bears on the dynamic interactions and responses inside a community whether through ecological, infrastructure, social, or economic mechanisms. However, previous analyses do not directly incorporate the experience of individuals during disasters when measuring the totality of a community's resilience. Instead, (community) resilience analyses examine the impact of a disaster or disruption on individuals as manifested through an existing social, physical, economic, or ecological systems [4], [7], [11], [12]. Recent work has hypothesized that *online social networks* (OSNs) can fill this gap in the study of resilience by incorporating the direct measurement of individuals in a community throughout the response to a major disruption [13], [14].

In this work, we formulate measurements of the resilience of a community by augmenting existing conceptualizations of community resilience with data from online social networks, namely the microblogging platform Twitter. In 2017, 80% of the US population is estimated to have a social media account; of those Twitter is among the most popular with 62 million monthly active users in the US in 2018 [15], [16]. Twitter is a platform for disseminating and consuming content at an unprecedented scale, providing a direct conduit into the response of individuals to major events. Interactions on Twitter are based on short messages of 280 characters. These messages (called *tweets*) are broadcast to a user's *followers*. Particularly during major events, the follower–followee relationships leads to emergent social properties at a macro-scale which are driven by a bottom-up self-organization of information [17], thus providing unique access to information deemed important by the community.

The associate editor coordinating the review of this manuscript and approving it for publication was Yuedong Xu.

Consequently when the resilience of a community is tested by a major event, the self-organization of Twitter discourse indicates that topics which are relevant to the resilience of a community are detectable. In this paper, we leverage this bottom-up information to develop a multi-dimensional *social resilience fingerprint* which analytically captures the interactions within pillars of community resilience during a disruption.

We introduce the *resilience fingerprint* as a multi-dimensional concept for understanding community resilience. A resilience fingerprint is the unique combination of components of community resilience in response to a major event or disruption. We use the analogy of a fingerprint to emphasize the identifiability of components critical to community resilience. In this way, we move away from evaluating resilience in one dimension and instead propose a relative-mapping of the interrelated aspects of resilience to one another. Rather than asking *how resilient is a community* we ask *what constitutes a resilient community*. We subsequently describe methods for measuring the resilience fingerprint of communities impacted by major events through analysis of the social media discourse surrounding the event thus establishing a *social resilience fingerprint*.

A social resilience fingerprint is an analytical method for understanding the interactions between components of community resilience as observed through social media. This is calculated first by defining community resilience as a set of resilience *components* suitable for measurement by social media, then categorizing the macro-scale Twitter response of a community before, during, and after a major event by its impact on the individual components. The relative measurements of each resilience component –along with the interaction between components– form the basis of the social resilience fingerprint.

The remainder of this paper is as follows: Section II provides background on community resilience and describes our categorization of community resilience in the context of online social network analysis; Section III describe the data used in this analysis, as well as the methods used to turn large corpora of tweets into a social resilience fingerprint. Finally, Section IV applies the techniques presented to 14 events with a significant Twitter response, the results of which are presented in Section V.

II. BACKGROUND

Externally, communities are the ''totality of social system interactions within a defined geographic space such as a neighborhood, census tract, city, or county'' [5], and can be characterized by internal dynamics which comprise combinations of individuals and groups with multiple –potentially competing– interests and associations [18], [19]. The broad scope of communities leads to a vast number of approaches and methods for the study of their resilience. In this section we discuss how conceptualizing resilience as a multidimensional fingerprint fits within context of existing studies of resilience and online social networks.

A. COMMUNITY RESILIENCE

In order to understand how multiple dimensions of disaster resilience can be studied through social media, we establish a definition of community resilience based on previous constructions and in alignment with evaluation through online social networks. Community resilience has been formalized as a comparative assessment of the resilience of community *components* or *categories* [5], [6], [20]. Category-based definitions of community resilience share substantial overlap. One such definition is given by the Multidisciplinary Center for Earthquake Engineering Research, which categorize community resilience with the acronym `PEOPLES`: **P**opulations, **E**nvironment and ecosystem, **O**rganized government, **P**hysical infrastructure, **L**ifesyle and community, **E**conomic development, and **S**ocial-cultural capital [6]. A similar definition proposes a framework which distinguishes categories of resilience by how they are measured [5]. They include ecological resilience, social resilience, economic resilience, institutional resilience, infrastructure resilience, and community competence [5]. We leverage a multi-dimensional categorization of community resilience, defined as a set of components which are derived from previous definitions of community resilience so as to theoretically ground our analysis [5]. We define the categories of community resilience in an OSN context as the Ecological, Economic, Institutional, Social, Infrastructure, and Quality of life categories. These categorizations are not mutually exclusive, but are collectively exhaustive. Table 1 lists high level descriptions of the components of a social resilience fingerprint and the topics they encompass through Twitter.

TABLE 1. Resilience components, their description, and community elements from that category.

Component	Description	Example elements
Ecology	Related to natural systems and features of the environment and ecosystem	Coasts, marshes, streams, beaches, wetland
Economy	Financial, economic, and business aspects within a community	Currency, business operation, labor
Institutions	Government and service-based institutions providing community function and care	Police, hospital, FEMA, government officials
Social	Non-institutional support systems within a community	Humanatarian aid, volunteerism, neighbors
Infrastructure	Physical infrastructure systems and their dependencies	Pipelines, power systems, cell communication
Quality of life	The health and wellbeing of the community	Health, hospital, mental well-being

B. TWITTER

Since its inception in 2006, Twitter has been a common source of academic inquiry particularly relating to its use

TABLE 2. Size, date, and type of twitter events analyzed. Ordered by starting date of event.

Event	Dates	Total corpus IDs[a]	Final tweet count[b]	Retention[c]
Hurricane Sandy	10/22/2012 to 11/02/2012	6,554,744	3,252,011	49.61%
Ebola Outbreak	08/18/2014 to 01/19/2015	5,085,767	993,905	19.54%
California Earthquake	08/24/2014 to 08/30/2014	254,529	50,414	19.81%
Nepalese Earthquake	04/25/2015 to 05/19/2015	4,223,983	509,299	12.06%
Brexit	05/05/2016 to 08/24/2016	23,733,133	3,884,599	16.36%
Charlottesville Riots	08/14/2017 to 10/23/2017	3,015,437	207,098	6.87%
Eclipse	08/17/2017 to 08/23/2017	13,548,321	1,211,729	8.94%
Hurricane Harvey	08/25/2017 to 10/23/2017	18,352,142	1,062,127	5.78%
Hurricane Irma	09/01/2017 to 10/23/2017	17,244,139	976,294	5.66%
Hurricane Maria	09/20/2017 to 10/03/2017	1,096,335	87,160	7.95%
Las Vegas shooting	09/29/2017 to 10/07/2017	14,108,104	866,758	6.14%
Ireland 8th Amendment	04/13/2018 to 06/04/2018	2,279,396	195,050	8.56%
Aretha Franklin's death	08/08/2018 to 08/18/2018	2,832,128	252,433	8.91%
Hurricane Florence	09/05/2018 to 10/03/2018	4,971,575	488,106	9.82%
Total		117,299,733	14,036,983	11.97%

[a] Original number of tweet ids published in corpus.

[b] Total number of tweets used in this analysis after deleted tweets are accoutned for and retweets removed.

[c] Tweet retention is calculated as the final tweet count divided by the total corpus ids for each event.

during disasters and major events. Since Twitter is a platform for sharing and consuming media, early work in the evaluation of tweet content established relationships between public Twitter posts and internal sentiment, situational awareness during disaster, and psychological trauma [17], [21].

Twitter has also been studied as a form of sensing network which can augment more traditional analyses performed during a disaster such as the study of vulnerability or resilience [13], [14]. Understanding how online social networks can be used to derive meaningful insight has been defined as *social media analytics* [13], [22]. Work in this area is typically broken down into multiple dimensions based on how social media is used for analysis (e.g., tweet location, tweet content etc.) [22]. What follows is a review of literature relating to understanding disasters and communities through social media.

Social media analytics has been previously used in many disaster-related contexts to gather information about the spatial distribution of disasters in an attempt to correlate measurable elements of a disaster with measurable elements of social media. Tweets related to a topic of a disaster were shown to be more likely to occur near disaster-related areas during a flood of the Elbe river [23], [24]. There is also significant evidence to suggest GIS and remote-sensing applications can be significantly improved by augmentation with social media data [25]. The primary benefit of this augmentation is that social media provides a ground-up network of sensors which can allow for hyper-local and rapid updating of geographic systems [26].

Temporal associations between tweets and disasters have also been investigated. A study of Hurricane Sandy found the time for an individual to learn about a disaster through social media was proportional to an individual's distance from the impact [27]. In a different context, the role of individuals in a disaster is found to be temporally-dependent [28]. During times of disasters, individuals are observed to transition toward an information-sharing role on Twitter, broadcasting and exchanging information [29].

Another thrust of social media analytics is an analysis of tweet content, in which a semantic understanding of a tweet is used to make assessments of the tweet author [22]. Related to disasters, the 'mood' of tweets was tracked through multiple disasters affecting North America as a proxy for how individuals recover psychologically from disasters [30]. Other analyses use the content of social media networks to understand the patterns of information diffusion in disaster [31].

III. DATA AND METHODS

The accessibility of tweets issued prior to 7 days in the past as well as Twitter's terms of service make acquiring corpora of tweets a non-trivial task. In this section we first briefly discuss the process of tweet acquisition, and follow with the methods used to analyze the Twitter corpora.

A. TWEET ACQUISITION

Our tweet datasets were retrieved from various archival sources [32]–[42]. High-level descriptions of the corpora are listed in Table 2, with more details presented in Appendix Table 4. Over 14 million tweets were analyzed spanning 14 major events. The major events include 5 hurricanes, 2 events of public violence, 2 political referendums, 2 earthquakes, 1 public health crisis, 1 death of a celebrity, and 1 solar eclipse. Events were chosen based on the scale of the social-media response to the event, but little other restriction was placed on inclusion in our study. This results in a corpus of tweets which spans multiple years, sizes, event types, and archival methods.

As of early 2019, Twitter limits access to the entire body of published tweets via a paid subscription service. Additionally, Twitter's Terms of Service prohibit the reproduction or distribution of datasets of *whole* tweets and instead only allow for the distribution of lists of numerical serial numbers corresponding to each tweet, called tweet *IDs* [43]. Hence, the medium of tweet compilation and sharing is the tweet ID, which can be used to re-construct the original tweets. IDs are simply serial numbers corresponding to each tweet and provide no actionable information, therefore, the process of *hydrating* tweets must be carried out to convert tweet IDs into a full tweet as it would be seen on the platform. Hydrating repeatedly calls the Twitter API with a specified tweet ID and returns the associated tweet content as well as additional meta-data such as the author, the date of publication, whether it is a retweet of someone else, etc. As this is a process of retroactively accessing data, there may be a loss of data. Tweets may not be available due to deletion of the previous tweet, tweet-author's user account, or change in privacy settings of a user account. Recent work has shown that despite this data loss, remaining samples of tweets are still representative of the data published in real time [43]. Based on previous findings which indicate that Twitter messages sent during consequential events are more focused on information-broadcasting and information sharing [28], we additionally remove retweets (*ie* a user re-broadcasting the tweet originally authored by someone else) from our data to focus on originally produced content.

B. DATA PROCESSING

After hydrating and removal of retweets, the text data of each tweet are processed to remove abnormalities. First, URLs, and non-ASCII characters are removed using customized regular expressions [44] [1]. English and Spanish stop words are then removed. Stop words are non-informative, frequently-used words which do not contribute to a semantic understanding of text [45]. In this case stop words are defined using the popular `stopwords` R package [46]–[48]. Each event's tweets are then processed to remove words occurring less than 10 times through all tweets related to an event. Additionally -if the dataset was complied based on keyword filtering- the words used for filtering were removed from the corpora, as they would otherwise be included in all tweets by construction. Finally, the remaining words are *stemmed* to remove word endings using the Porter stemming algorithm, implemented in R [49]–[51]. Stemming removes word endings to avoid differentiating between words of similar meaning used in different tenses, conjugations etc. For example ecolog**ical** and ecolog**y** would both stem to the same root: ecolog. Word stemming has been previously shown to greatly improve text processing and analysis [45].

[1] This is increasingly important in recent datsets as the use of emojis becomes more prevalent

C. SOCIAL RESILIENCE FINGERPRING

At the core of the methodology proposed in this paper is understanding how individual *components* of community resilience can be measured and understood through the lens of social media. Formally we have a set of all events $\mathbb{E}$ comprised of n individual events E such that $E_1, E_2, \ldots, E_n \in \mathbb{E}$. For a given event E^*, we have a set of tweets, $t_1^{E^*}, t_2^{E^*}, \ldots, t_m^{E^*} \in E^*$, where m is the total number of tweets compiled for each event after hydration and processing. Each tweet is subsequently comprised of a series of *features*, f, which are the individual words in each tweet such that for a given tweet $t^*, f_1^{t^*}, f_2^{t^*}, \ldots, f_l^{t^*} \in t^*$. As each tweet can contain multiple copies of the same word, we also have a set of all features F^E for a given event E.

Additionally, we manually coded a set of words for each category in order to map the set of features to our pre-defined resilience categories. Thus, each category of resilience contains a set of words which indicate associated discourse. For example, $C_{\text{infrastructure}} = \{$ *power, water, cell, outage, road, …* $\}$. The words were manually selected by two groups individually, then consensus was established between the two sets. As such, the sets of words used in each category are not mutually exclusive. The full listing of words coded for each category is listed in Appendix Table 6.

To parse the features for a given event into categories we construct a *category co-association* matrix, A for each event. A is a symmetric 6 by 6 matrix with each row and column corresponding to a resilience category. A_{ij} is then the co-association of category i with category j. The co-association of a given category is based on the co-occurrence of words from categories. As such, for resilience categories i and j,

$$A_{ij} = \sum_{\substack{r \in C_i \\ s \in C_j}} \sum_{t} \sum_{r \in t} \mathrm{occ}(r, s) \tag{1}$$

where

$$\mathrm{occ}(r, s) = \begin{cases} 1 & \text{if word } r \text{ occurs with } s \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

After fixing a word r from one category, and word s from another, the `occ` function is an indicator function taking a value of 1 each time word r occurs with word s in a given tweet. This is summed over all occurrences of r in a given tweet (innermost summation of (1)), then subsequently summed over every tweet. This is done for all combinations of words in category i and category j. Thus A_{ij} is the total times a word from category i occurs in the same tweet as category j. The matrix of values A –one for each event– form the social resilience fingerprint. Off diagonal values of A represent the frequency of resilience categories appearing together in Twitter discourse. The diagonals of A are less intuitive, representing the relative frequency of words from the same category appearing in a tweet. This is a modified version of a *co-occurrence* matrix, used for term clustering in natural language processing [52]–[54]. This extension uses apriori

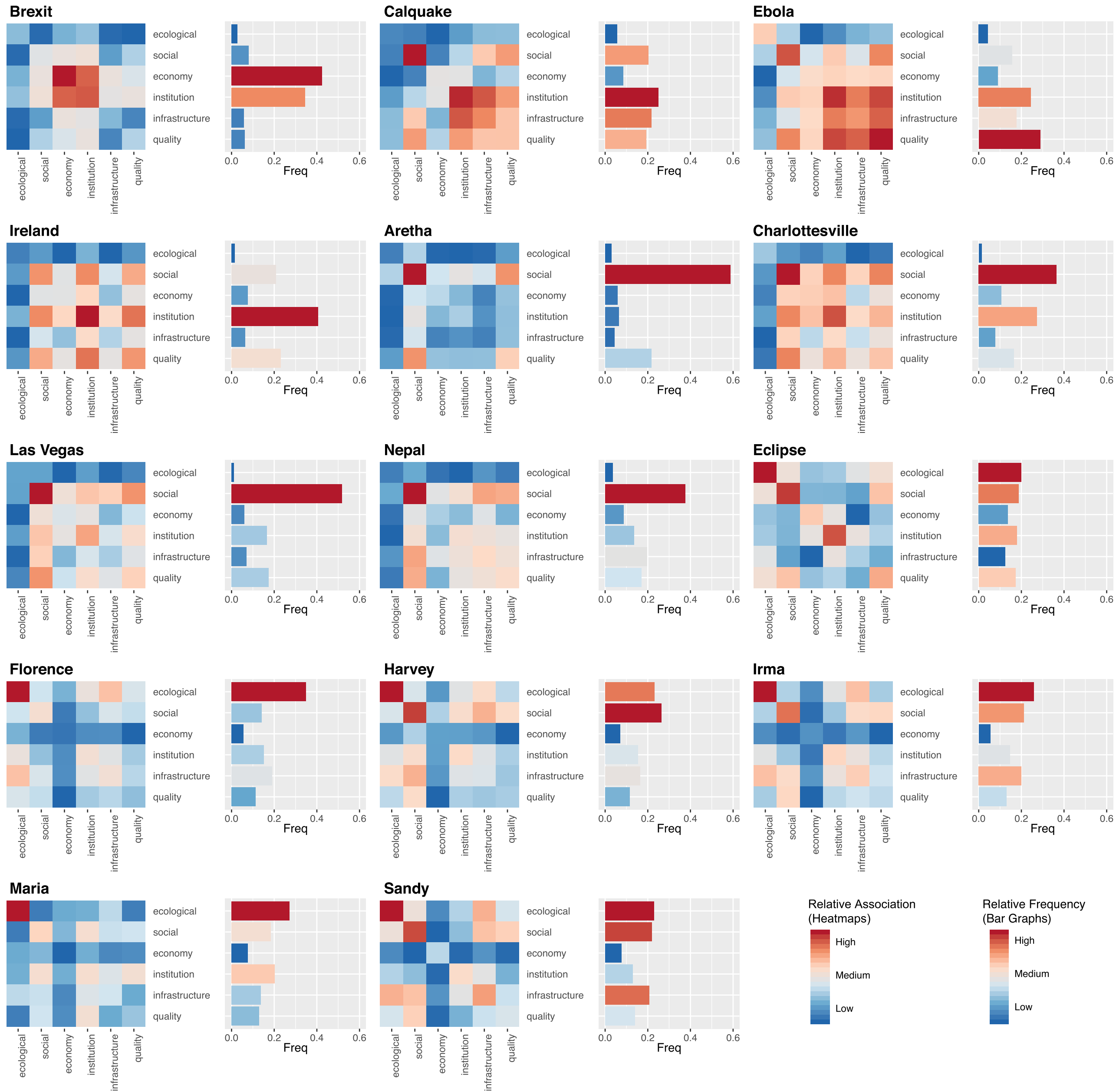


FIGURE 1. **Visualizing social resilience fingerprints. Each heatmap represents the association between one category with another. The color red indicates the most association, and the color blue represents the least association. Diagonal values in the heatmap are indicative of how self-associative a category is. Bar-graphs show the relative occurrence of each category. Note the color scheme in this plot is based on log-normalization of *A*, as opposed to Sinkhorn-Knopp, to aid in visualization.**

categorizations –grounded in the theoretical definitions of community resilience– to find associations within topics to determine the relative association of categories of resilience. In the following section, we apply this fingerprinting methodology to multiple major events and discuss the feasibility of extracting category-based insights using this method. For the 14 events listed in Table 2, the categorical binning described in (1) and (2) are used to establish the social resilience fingerprint. As the total number of tweets gathered for each event vary substantially, the A matrices are scaled. This allows for a more balanced comparison between events as it removes information regarding the total number of tweets from the fingerprint so that any comparison made between events is based solely on the pattern of interactions among the components of the resilience fingerprint. Sinkhorn-Knopp matrix regularization is used on A matrices [55]; this preserves the structure of the fingerprint while allowing the relations between categories to be compared across events.

IV. RESULTS

Visual representations of the social resilience fingerprints are shown in Figure 1. Each heatmap and associated bargraph show the relative association of each category and the frequency of each category respectively. The respective heatmaps are visual examples of the matrices A, representing the co-association of discourse related to components of community resilience. Because of the self-organization of tweets in response to major events, we hypothesize that stronger textual association of categories indicate a stronger underlying relationship between the categories in the community and by extension in the resilience of the community. This is in line with previous findings which found that event-related keywords were indicative of a major event's impact on an individual [27].

A. EVENT SIMILARITY

From the wide range of the events studied, we hypothesize that the Twitter discourse in reaction to similar events will itself be similar, as measurable through the resilience fingerprint. To evaluate this, we measure the component-wise Spearman distance between scaled A matrices for all events. The result is a numerical measure of similarity among the structure of the resilience fingerprints in which a smaller distance represents a more-similar pattern of Twitter discourse between two events. In Figure 2, the resulting pair-wise distances are visualized in a heatmap after hierarchical clustering is performed on the rows and columns –a technique called VAT or a Visual assessment of Cluster Tendency [56]–[58]. Each element in Figure 2 represents the distance between the row and column event.

The VAT methodology is formulated to allow for visual identification of trends in data [56]. A VAT cluster appears visually as a square block along the lower-left to upper-right diagonal of the heatmap. In Figure 2, there are clear clusters corresponding to Hurricanes Florence, Irma, Sandy, Harvey, Maria as well as the 2018 Eclipse. Additionally, a case could be made for the clustering of the Nepalese earthquake and the California earthquake. Finally, the upper right of Figure 2 provides evidence of clustering of the Las Vegas shootings, Charlottesville riots, Ireland's 8th constitutional amendment, and the death of Aretha Franklin. As we calculated the distance between the events by summing component-wise distances between two fingerprints –each scaled from their original counts– these clusters are representative of similarity in the pattern of associations between components of resilience. From this, we see a similarity in the social resilience fingerprints of alike events, providing evidence that our proposed methodology has discriminating power.

Based on this distance measure, we subsequently analyzed each event's closest match using an alternative distance measure, namely, the Pearsons's correlation coefficient. This is also computed between each pair of fingerprints. The closest-correlated event to each event are listed in Table 3, along with the associated correlation. The results paint a similar picture to the VAT comparison. Natural disasters, such as hurricanes and earthquakes, pair closely with one another, as do acts of violence like the Las Vegas shooting and Charlottesville riots.

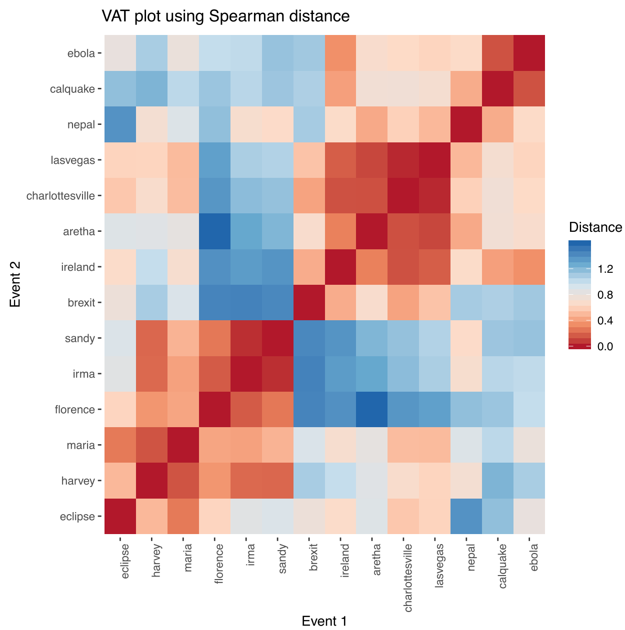


FIGURE 2. VAT assessment. Each element represents the distance between the event in the row and the event in the column, red indicating closer, blue farther away.

TABLE 3. Closest events. The pearson correlation is calculated between all pairs of events, with the closest match listed. The correlations above 80% have been highlighted in bold.

Event	Best Match	Correlation
aretha	lasvegas	0.77
brexit	charlottesville	0.50
calquake	**nepal**	**0.85**
charlottesville	**lasvegas**	**0.86**
ebola	irma	0.56
eclipse	charlottesville	0.51
florence	**irma**	**0.94**
harvey	**irma**	**0.89**
ireland	calquake	0.78
irma	**florence**	**0.94**
lasvegas	**charlottesville**	**0.86**
maria	**florence**	**0.90**
nepal	**calquake**	**0.85**
sandy	**irma**	**0.88**

We perform another similarity measurement by comparing the clusters generated via the k-means clustering algorithm. We select 3 clusters as the marginal within-cluster error does not improve greatly with additional number of clusters. Figure 3 shows the three clusters plotted on the axes of the first two principle components of the data. The cluster containing the hurricanes and eclipse differs most greatly in the direction of the first principle component, while the remaining two clusters differ based on the second component. The first principle component is driven by differences in ecological categories of resilience while the second is a difference in social, economic, and institutional resilience.

From the results of these clustering methods, we hypothesize that the similarity between the fingerprints of similar

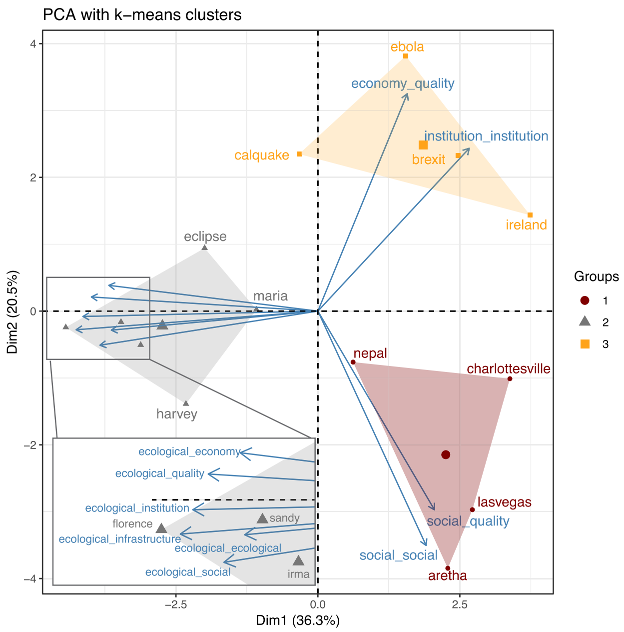


FIGURE 3. PCA loading plot. The top 10 most contributing elements of each fingerprint are projected onto the first two principle components. Components and events along the negative x axis are shown in the bottom left inset for clarity. *k*-means clusters ($k = 3$) are additionally overlain on the PCA plot, with clusters denoted by shaded areas and the shape of the icon.

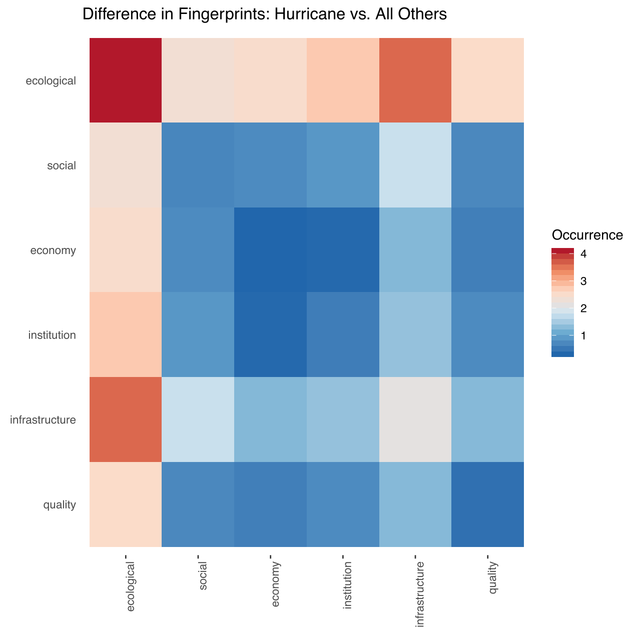


FIGURE 4. Category-based difference between the average hurricane fingerprint and all average non-hurricane fingerprint. Each element represents the difference between the hurricane and non-hurricane association of categories. Blue indicates stronger association in non-hurricane events, while red indicates stronger association in hurricane events. Coloration based on log-normalization of *A* matrices rather than Sinkhorn-Knopp for visual clarity.

events indicates that much of the emergent properties of the resilience of a community is driven by the specific disaster. Through this hypothesis, we propose that the elements of community resilience common to each type of event are distinct enough to affect the Twitter discourse of the individual communities to an extent that it is measurable at a macro scale.

B. CRITICAL COMPONENTS OF COMMUNITY RESILIENCE

To further understand the importance of the categories of community resilience, we now ask which elements of community resilience drive the similarity among events by looking at the loading of each variable –corresponding to an i, j element of the fingerprint across all events– as projected onto the first two principle components. The variable loading for the 10 most contributing variables are plotted in Figure 3 along with the events. Figure 3 additionally includes the k-means clusters described in Section IV-A.

From the variable loadings, we can see that –as expected– the vectors with similar directionality have overlapping categories. Along the x-axis of Figure 3 are the associations of the ecological category with all others, indicating they are strong contributors to the similarity of the hurricane-events. Likewise, institutional and economic category dominate the first quadrant. Finally, social components tend in the direction of the cluster associated with the Charlottesville riots and Las Vegas shootings. In Figure 3, a small angle between vector loading indicate high correlation between variables. From this we can generally infer a positive correlation within the ecological and social categories as well as between the institution and economy categories.

From Figure 3 we can also interpret that the first principle component is driven by changes in the ecological categories indicating this may be primary drivers behind the clustering of the hurricanes, and consequently a significant component of community resilience.

C. POSTERIOR ANALYSIS

To further investigate the components most influential in the social resilience fingerprint, we look at the explicit difference between events of different types. The most apparent cluster of events are Hurricanes Florence, Irma, Sandy, Harvey, and Maria. As such, we compute the element-wise mean fingerprint of the hurricane-events and non-hurricane-events. The element-wise difference –calculated as the hurricane mean minus the non-hurricane mean– is visualized in a heatmap in Figure 4. For each pair of categories, the color of the cell value indicates whether those categories have a stronger association among the hurricane fingerprints (colored red), or the non-hurricane fingerprints (colored blue).

Figure 4 confirms the results of the PCA analysis and indicates the ecological and infrastructure categories of resilience are much stronger in the hurricane fingerprints than in the non-hurricane fingerprints. The interaction of infrastructure and ecological categories are the strongest for the hurricane category among the non-diagonal elements. At the same time,

TABLE 4. Tweet corpora summary. Description of events used along with the quantity of tweets and their acquisition methods and respective sources.

Event	Event Description	Event Dates	Total Tweet IDs	Resulting Tweets	Tweet acquisition method	Reference
Aretha Franklin's death	The death of singer Aretha Franklin	08/08/2018-08/18/2018	2,832,128	252,433	Keyword filtering	[32]
Brexit	The referendum to remove the UK from the European Union	05/05/2016-08/24/2016	23,733,133	3,884,599	Keyword filtering	[33]
California earthquake	Magnitude 6.0 earthquake striking south of Napa, CA	08/24/2014-08/30/2014	254,529	50,414	Keyword filtering	[34]
Charlottesville	White supremacist rally which resulted in significant counter-protesting and violence in Charlottesville, VA	08/14/2017-10/23/2017	3,015,437	207,098	Keyword filtering	[35]
Ebola Outbreak	Ebola epidemic in Guinea, Liberia, Sierra Leone and other parts of West Africa	08/18/2014-01/19/2015	5,085,767	993,905	Keyword filtering	[34]
Eclipse	2017 Solar eclipse passing over much of the United States	08/17/2017-08/23/2017	13,548,321	1,211,729	Keyword filtering	[36]
Hurricane Florence	Major Atlantic Hurricane impacting the Eastern United States	09/05/2018-10/03/2018	4,971,575	488,106	Keyword filtering	[37]
Hurricane Harvey	Major Atlantic Hurricane impacting the Gulf Coast	08/25/2017-10/23/2017	18,352,142	1,062,127	Keyword filtering	[38]
Ireland 8th	Referendum to remove the Eight Amendment from the Irish Constitution, governing the legality of abortion	04/13/2018-06/04/2018	2,279,396	195,050	Keyword filtering	[39]
Hurricane Irma	Major Atlantic hurricane impacting the Caribbean and Florida Keys	09/01/2017-10/23/2017	17,244,139	976,294	Keyword filtering	[38]
Las Vegas shooting	Lone-gunman attack on a music festival in Las Vegas, NV	09/01/2017-10/23/2017	14,108,104	866,758	Keyword filtering	[40]
Hurricane Maria	Major Atlantic hurricane severely impacting Puerto Rico	09/20/2017-10/03/2017	1,096,335	87,160	Keyword filtering	[41]
Nepal	Magnitude 7.8 earthquake in the Gorkha District of Nepal	04/25/2015-05/19/2015	4,223,983	509,299	Keyword filtering	[34]
Hurricane Sandy	Major Atlantic storm impacting much of the Caribbean and East Coast of the US	10/22/2012-11/02/2012	6,554,744	3,252,011	Bounding box surrounding CT, DE, MA, MD, NJ, NY, NC, OH, PA, RI, SC, VA, WV.	[42]

the economic-institutional relationship is most strong among the non-hurricane events.

V. DISCUSSION

After clustering the social resilience fingerprints for all events and analyzing what drives their similarity, we identify two major trends: first is the strong distinction between hurricane and non-hurricane events with respect to fingerprint similarity, and second is the importance of ecological and infrastructure resilience in making that distinction.

We see a strong association, not just of one hurricane with another, but among *all* hurricanes for which we could collect data. The hurricane-related tweet corpora were collected in a variety of ways and span distinct spatial and temporal scales. Despite these differences, the similarity in the fingerprints indicate generalizable patterns in community resilience in the face of hurricane impacts. Moreover, it provides a strong evidence supporting the fingerprinting methodology. It also suggests that Twitter is a persistent source of data about individual responses to a disaster within a community, establishing Twitter as a valuable tool for measuring disaster resilience across communities.

Additionally, general similarity among specific non-hurricane events indicates emergent themes in the Twitter responses manifesting as similar social resilience fingerprints of related events, and thus similarities in the underlying resilience. The relative similarity of the California and Nepalese earthquakes, as well as the public violence in Charlottesville and Las Vegas, both indicate that other types of major events may also have fundamental, emergent themes decodable through Twitter discourse. We conjecture that similarity in the social resilience fingerprints of related events is indicative of fundamental similarity in the resilience of the communities facing such events. That is, there are

TABLE 5. Tweet acquisition. Keywords, keyword phrases, and hashtags used to create the tweet datasets.

Event	Keywords
Aretha Franklin's death	aretha_franklin, queen_of_soul
Brexit	brexit
California earthquake	napa_earthquake, sonoma_earthquake, bay_area_earthquake, california_earthquake, ca_earthquake, sfearthquake, san_francisco_earthquake, napaearthquake, sfquake, napaquake, napa_quake, sonoma_quake, bay_area_quake, california_quake, ca_quake, san_francisco_quake
Charlottesville	charlottesville, standwithcharlottesville, defendCville, heatherheyer, unitycville
Ebola Outbreak	ebola, ebola_virus
Eclipse	solareclipse2017, solareclipse, eclipse2017, eclipseday, eclipse
Florence	florence, hurricaneflorence, florencehurricane, hurricane_florence", florencenc, hurricaneflorence2018, hurricaneflorence
Harvey	hurricane_harvey, hurricaneharvey, harvey, hurricane
Ireland 8th	8thref, hometovote, jointherebellion, trustwomen, repealthe8th, together4yes, togetherforyes, voteyes, time4choice, knowyourrepealers, mybodymychoice, savethe8th, loveboth, lovebothvoteno, votenotoabortion, standupforlife, lifecanvass, protectthe8th, 8thamendment, whoneedsyouryes, men4yes, register4yes, roadtorepeal, repealfacts, healthcarenotairfare, repeal, trustwomen, itstime, whyimvotingyes, deaftogetherforyes, doctorsforyes, repeal4betterbirth, togetherforno, men4no, whoneedsyourno, rallyforlife, votenotoabortion, bemyyes, academicsforyes, hometovoteno, hometocanvass, abortionreferendum, savita, repealshield, farmersforyes, lawyersforchoice, lawyersforyes, studentsforchoice, archivingthe8th, repealedthe8th, wemadehistory, nowforni, wetrustwomen,
Irma	irma, hurricane_irma, irmastrong
Las Vegas shooting	vegas
Maria	hurricane_maria, hurricanemaria, tropical_storm_maria, maria_storm
Nepal	basantapur, patan, anamnagar, bhaktapur, durbar_square, nuwakot, dharahara_tower, gorkha, lamjung, khudi, kathmandu, sankhu, sunsari, solu_district, okhaldhunga, nepal, nepal_earthquake, ktmearthquake, indiawithnepal, nepalquake, nepalquakerelief, nepalearthquake, kathmanduquake, kathmanduquakerelief, kathmanduearthauqake, quakenepal, earthquakenepal, quakekathmandu, earthquakekathmandu, prayfornepal
Sandy	NA

emergent similarities between the way different communities respond to the same event across all types of events. However, we recognize the limitation of drawing conclusions from the similarity of only two events studied in this paper and intend to expand upon this analysis to test our conjecture.

The second major trend in the analysis of the social resilience fingerprints is the influence of individual components of resilience in the separation of one event from another. Ecology, infrastructure, and economic categories drive much of the separation between the emergent clusters in the data. Economic resilience is intuitively intertwined with all other categories in our definition [3], [6], [7], and is seen in the Principle Component Analysis to contribute greatly to the distinction between clusters of non-hurrican events.

The significance of infrastructure resilience in differentiating between hurricane and non-hurricane events –as seen in Figure 4– is likely due to the significance of infrastructure damage in communities affected by hurricanes. Ecological resilience and its close ties to sustainability, have been previously shown to be strong drivers of community resilience at all levels [3], [5]. We see the distinction in Figure 3, manifesting as the ecological loadings in the direction of the first principle component –indicating that ecology explains the largest degree of variance among the fingerprints. This reveals that the resilience fingerprint method is not limited by what has hampered the previous attempts in quantifying community resilience –namely the difficulty in acquiring data related to specific ecosystems. Due to the difficulties in finding relevant measurement indicators, ecological resilience has previously been excluded from resilience assessments [60].

The resilience fingerprints of three events were not revealed as expected: The Irish constitutional amendment, Brexit, and the Ebola outbreak. The authors hypothesized that the Irish constitutional amendment and Brexit would be similar events due to their close physical proximity and the general political nature of the event; a trend which did not emerge from our analysis. One explanation for the difference are in the specificity of search terms used for the generation of the Irish amendment tweet dataset. The Irish amendment tweet dataset used 52 terms to filter by, the most most filter terms used by almost a factor of 2 (See Appendix Table 5 for terms); the Brexit dataset was built on only one search term: `brexit`. The terms used to filter the Irish referendum dataset are also more specific than the others, leading to a corpus of tweet text which may be overly specific to the Irish political system and the issues of the referendum, lacking substantive information about the community's response in favor of the individuals. Tweets related to the Ebola virus additionally showed little relation with other events. In this case, we hypothesize that the location of the event relative to major Twitter-adoptive societies may affect the ability of fingerprinting to detect a signal. International Twitter use is

TABLE 6. Categories of resilience and associated keywords. Keywords are manually coded based on conceptual definitions of resilience categories.

Ecological	Social	Economy	Institution	Infrastructure	Quality
ecological	social	economy	institution	power	community
ecology	love	nation	nation	nation	love
erosion	peace	market	hospital	emergency	life
wetlands	prayer	business	vote	flight	home
biology	family	bank	poll	airplane	hospice
coast	life	trade	country	safe	hospital
marsh	bless	stock	police	water	protest
dune	spirit	politic	mayor	power	health
fish	protest	money	president	relief	school
bird	rally	dollar	governor	city	doctor
river	monument	credit	senator	coal	nurse
climate_change	god	job	flag	evacuate	medic
rainfall	church	jobs	doctor	airport	safe
nature	donate	work	nurse	cell	found
floodwater	aid	money	govern	water	aid
beach	network	wealth	school	outage	humanatarian
sun	church	property	medic	road	life
stream	faith	pay	fema	bus	health
flood	friend	employer	shelter	car	depression
storm	friends	employee	school	infrastructure	
wind	family		potus		
rain	pray		red_cross		
water	neighborhood		church		
weather	town		evacuate		
beach			homeland		
tropic			responders		
climate			fema		
			ems		
			police		
			fire		
			government		
			alderman		
			county		
			officials		

lower than that of the US [16]. As such we hypothesize that someone tweeting about Hurricane Florence was more likely witnessing community impacts due the storm than someone tweeting about the ebola outbreak.

VI. CONCLUSION AND FUTURE RESEARCH

In this paper, we present the *resilience fingerprint* as a concept for understanding community resilience as the relationship of individual components. We then calculate a *social resilience fingerprint* by leveraging social media analytics guided by the community resilience theory. We find evidence that resilience fingerprinting can highlight the different community responses to a variety of major events and identify the components of community resilience which most contribute to the overall response. We leverage a category-based definition of community resilience to classify the macro-scale response on Twitter to a disaster into elements of community resilience.

In summary, the resilience fingerprint provides a concept for the multi-dimensional analysis of the emergent responses of communities to major events. The rapid spread of information via social media makes social resilience fingerprinting a vital complement to existing resilience analyses, capable of categorizing the community response to a disaster.

In this work, the categories were manually coded, as guided by the literature in community resilience. However, an ongoing extension of this work is to use automated topic detection to both determine what individual words best comprise a resilience category, and to determine the emergent resilience categories in an unsupervised way. Additionally, we aim to extend the classification of tweets beyond word-association based on recent developments in the classification of tweets related to disasters [59], [61].

This work does not include retweets in the data analyzed with the resilience fingerprint. A sensitivity analysis is ongoing as to assess the impact of retweets on event similarity.

Finally, we are expanding the fingerprinting methods to allow for the creation of a resilience fingerprint in real time. This will provide a dynamic look at the interactions among communities as they respond to major disasters and events.

APPENDIX

See Tables 4–6.

REFERENCES

[1] *U.S. Billion-Dollar Weather and Climate Disasters*, NOAA National Centers for Environmental Information, 2019. [Online]. Available: https://www.ncdc.noaa.gov/billions/

[2] A. B. Smith and J. L. Matthews, "Quantifying uncertainty and variable sensitivity within the us billion-dollar weather and climate disaster cost estimates," *Natural Hazards*, vol. 77, no. 3, pp. 1829–1851, 2015.

[3] W. N. Adger, T. P. Hughes, C. Folke, S. R. Carpenter, and J. Rockström, "Social-ecological resilience to coastal disasters," *Science*, vol. 309, no. 5737, pp. 1036–1039, 2005. [Online]. Available: http://science.sciencemag.org/content/309/5737/1036

[4] C. S. Holling, "Resilience and stability of ecological systems," *Annu. Rev. Ecol. Syst.*, vol. 4, no. 1, pp. 1–23, 1973. doi: 10.1146/annurev.es.04.110173.000245.

[5] S. L. Cutter *et al.*, "A place-based model for understanding community resilience to natural disasters," *Global Environ. Change*, vol. 18, no. 4, pp. 598–606, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959378008000666

[6] G. P. Cimellaro, C. Renschler, A. M. Reinhorn, and L. Arend, "PEOPLES: A framework for evaluating resilience," *J. Struct. Eng.*, vol. 142, no. 10, 2016, Art. no. 04016063. doi: 10.1061/(ASCE)ST.1943-541X.0001514.

[7] A. Rose and S.-Y. Liao, "Modeling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions," *J. Regional Sci.*, vol. 45, no. 1, pp. 75–112, 2005. doi: 10.1111/j.0022-4146.2005.00365.x.

[8] M. Bruneau *et al.*, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthq. Spectra*, vol. 19, no. 4, pp. 733–752, 2003. doi: 10.1193/1.1623497.

[9] R. Nateghi, "Multi-dimensional infrastructure resilience modeling: An application to hurricane-prone electric power distribution systems," *IEEE Access*, vol. 6, pp. 13478–13489, 2018.

[10] B. Rachunok and R. Nateghi. (2019). "The sensitivity of electric power infrastructure resilience to the spatial distribution of disaster impacts." [Online]. Available: https://arxiv.org/abs/1902.02879

[11] C. N. Kaiser-Bunbury *et al.*, "Ecosystem restoration strengthens pollination network resilience and function," *Nature*, vol. 542, no. 7640, pp. 223–227, 2017.

[12] M. Ouyang, "Review on modeling and simulation of interdependent critical infrastructure systems," *Rel. Eng. Syst. Safety*, vol. 121, pp. 43–60, Jan. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832013002056

[13] K. Barker *et al.*, "Defining resilience analytics for interdependent cyber-physical-social networks," *Sustain. Resilient Infrastruct.*, vol. 2, no. 2, pp. 59–67, Mar. 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/23789689.2017.1294859

[14] P. Meier, "Human Computation for Disaster Response," in *Handbook of Human Computation*, P. Michelucci, Ed. New York, NY, USA: Springer, 2013. pp. 95–104. doi: 10.1007/978-1-4614-8806-4_11.

[15] (2018). *U.S. Population With a Social Media Profile*. Accessed: Dec. 2018. [Online]. Available: https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/

[16] *Top U.S. Mobile Social APPS by Users 2018 | Statistic*. Accessed: 12/2018. [Online]. Available: https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/

[17] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on the red: What hazards threat reveals about the social life of microblogged information," in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*. New York, NY, USA: ACM, 2010, pp. 241–250. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1718918.1718965

[18] A. Agrawal and C. C. Gibson, "Enchantment and disenchantment: The role of community in natural resource conservation," *World Develop.*, vol. 27, no. 4, pp. 629–649, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305750X98001612

[19] F. Berkes and H. Ross, "Community resilience: Toward an integrated approach," *Soc. Natural Resour.*, vol. 26, no. 1, pp. 5–20, 2013. doi: 10.1080/08941920.2012.736605.

[20] S. W. Gilbert, *Disaster Resilience: A Guide to the Literature*. Scotts Valley, CA, USA: CreateSpace, 2016.

[21] K. Glasgow, C. Fink, and J. Boyd-Graber, "'Our Grief is unspeakable': Automatically measuring the community impact of a tragedy," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–10.

[22] Z. Wang and X. Ye, "Social media analytics for natural disaster management," *Int. J. Geographical Inf. Sci.*, vol. 32, no. 1, pp. 49–72, 2018. doi: 10.1080/13658816.2017.1367003.

[23] J. P. de Albuquerque, B. Herfort, A. Brenning, and A. Zipf, "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management," *Int. J. Geograph. Inf. Sci.*, vol. 29, no. 4, pp. 667–689, 2015. doi: 10.1080/13658816.2014.996567.

[24] B. Herfort, J. P. de Albuquerque, S.-J. Schelhorn, and A. Zipf, "Exploring the geographical relations between social media and flood phenomena to improve situational awareness," in *Connecting a Digital Europe Through Location and Place* (Lecture Notes in Geoinformation and Cartography), J. Huerta, S. Schade, and C. Granell, Eds. Springer, pp. 55–71, 2014. doi: 10.1007/978-3-319-03611-3_4.

[25] K. Boulos *et al.*, "Crowdsourcing, citizen sensing and sensor Web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples," *Int. J. Health Geograph.*, vol. 10, no. 1, p. 67, 2011. doi: 10.1186/1476-072X-10-67.

[26] S. Roche, E. Propeck-Zimmermann, and B. Mericskay, "GeoWeb and crisis management: Issues and perspectives of volunteered geographic information," *GeoJournal*, vol. 78, no. 1, pp. 21–40, 2013. doi: 10.1007/s10708-011-9423-9.

[27] Y. Kryvasheyeu *et al.*, "Rapid assessment of disaster damage using social media activity," *Sci. Adv.*, vol. 2, no. 3, 2016, Art. no. e1500779. doi: 10.1126/sciadv.1500779.

[28] A. L. Hughes and L. Palen, "Twitter adoption and use in mass convergence and emergency events," *Int. J. Emergency Manage.*, vol. 6, nos. 3–4, p. 248, 2009. [Online]. Available: http://www.inderscience.com/link.php?id=31564

[29] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, " Microblogging during two natural hazards events: What twitter may contribute to situational awareness," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2010, pp. 1079–1088.

[30] A. López-Cuevas, J. Ramírez-Márquez, G. Sanchez-Ante, and K. Barker, "A community perspective on resilience analytics: A visual analysis of community mood," *Risk Anal.*, vol. 37, no. 8, pp. 1566–1579, 2017. doi: 10.1111/risa.12788.

[31] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *Int. J. Inf. Manage.*, vol. 38, no. 1, pp. 86–96, 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S026840121730525X

[32] E. Summers. *Aretha-Tweets*. Internet Archive. Accessed: Dec. 2018. [Online]. Available: https://archive.org/details/aretha-tweets

[33] D. Milajevs. (Jan. 2017). *#Brexit Tweets Collected From the 5th of May to the 24th August 2016*. Accessed: Dec. 2018. [Online]. Available: https://doi.org/10.5281/zenodo.263584

[34] M. Imran, P. Mitra, and C. Castillo, "Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages," in *Proc. 10th Lang. Resour. Eval. Conf. (LREC)*. Paris, France: European Language Resources Association (ELRA), May 2016, pp. 1638–1643. [Online]. Available: http://crisisnlp.qcri.org/lrec2016/lrec2016.html

[35] J. Littman. (2018). *Charlottesville Tweet Ids*. Accessed: Dec. 2018. [Online]. Available: https://doi.org/10.7910/DVN/DVLJTO

[36] E. Summers. *Eclipse Tweet Ids*. Internet Archive. Accessed: Dec. 2018. [Online]. Available: https://archive.org/details/eclipse-tweets.csv

[37] M. Philips. *Hurricanhe Florence Twitter Dataset*. Accessed: Dec. 1, 2018. [Online]. Available: digital.library.unt.edu/ark:/67531/metadc1259406/

[38] J. Littman. (2017). *Hurricanes Harvey and IRMA Tweet Ids*. [Online]. Available: https://doi.org/10.7910/DVN/QRKIBW

[39] J. Littman. (2018). *Ireland 8th Tweet Ids*. [Online]. Available: https://doi.org/10.7910/DVN/PYCLPE

[40] University of Nevada. (2017). *Web archive on the Cctober 1, 2017 shooting in Las Vegas, Nevada*. Accessed: Dec. 1, 2018. [Online]. Available: https://n2t.net/ark:/62930/f1207c

[41] F. Alam, F. Ofli, M. Imran, and M. Aupetit. (2018). "A twitter tale of three hurricanes: Harvey, IRMA, and maria." [Online]. Available: https://arxiv.org/abs/1805.05144

[42] H. Wang, E. Hovy, and M. Dredze, "The hurricane sandy twitter corpus," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–4.

[43] A. Zubiaga, "A longitudinal assessment of the persistence of twitter datasets," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 8, pp. 974–984, 2018. doi: 10.1002/asi.24026.

[44] T. W. Rinker. (2017). qdapRegex: Regular Expression Removal/Extraction/Replacement Tools, University at Buffalo/SUNY, Buffalo, NY, USA. [Online]. Available: http://github.com/trinker/qdapRegex

[45] C. Silva and B. Ribeiro, "The importance of stop word removal on recall values in text categorization," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 3, Jul. 2003, pp. 1661–1666.

[46] K. Benoit, D. Muhr, and K. Watanabe, *StopWords: Multilingual Stopword Lists*, document R Package Version 0.9.0, 2017. [Online]. Available: https://CRAN.R-project.org/package=stopwords

[47] I. Feinerer and K. Hornik, *TM: Text Mining Package*, document R Package Version 0.7-5, 2018. [Online]. Available: https://CRAN.R-project.org/package=tm

[48] D. Meyer, K. Hornik, and I. Feinerer, "Text mining infrastructure in R," *J. Stat. Softw.*, vol. 25, no. 5, pp. 1–54, Mar. 2008. [Online]. Available: http://www.jstatsoft.org/v25/i05/

[49] K. Benoit *et al.*, "quanteda: An r package for the quantitative analysis of textual data," *J. Open Source Softw.*, vol. 3, no. 30, p. 774, 2018. [Online]. Available: https://quanteda.io

[50] M. F. Porter. (2001). *Snowball: A Language for Stemming Algorithms*. [Online]. Available: http://snowball.tartarus.org/texts/introduction.html

[51] R. C. Team. (2018). R: A Language Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[52] S. Momtazi, S. Khudanpur, and D. Klakow, "A comparative study of word co-occurrence for term clustering in language model-based sentence retrieval," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 325–328.

[53] D. Jurafsky and J. H. Martin, *Vector Semantics*. Speech and Language Processing. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/6.pdf

[54] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," in *Proc. 27th Annu. Meeting Assoc. Comput. Linguistics*, 1989, pp. 76–83. [Online]. Available: http://portal.acm.org/citation.cfm?doid=981623.981633

[55] P. A. Knight, "The Sinkhorn–Knopp algorithm: Convergence and applications," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 1, pp. 261–275, 2008.

[56] J. C. Bezdek and R. J. Hathaway, "VAT: A tool for visual assessment of (cluster) tendency," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 3, 2002, pp. 2225–2230.

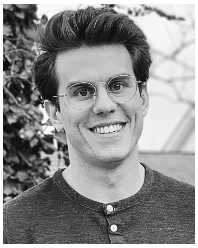
[57] A. Kassambara and F. Mundt, *factoextra: Extract Visualize Results Multivariate Data Analyses*, document R Package Version 1.0.5, 2017. [Online]. Available: https://CRAN.R-project.org/package=factoextra

[58] R. C. Team, *R: A Language Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018.

[59] C. Zhang, C. Liu, X. Zhang, and G. Almpandis, "An up-to-date comparison of state-of-the-art classification algorithms," *Expert Syst. Appl.*, vol. 82, pp. 128–150, Oct. 2017.

[60] S. L. Cutter, C. G. Burton, and C. T. Emrich, "Disaster resilience indicators for benchmarking baseline conditions," *J. Homeland Secur. Emergency Manage.*, vol. 7, no. 1, pp. 1–25, 2010. [Online]. Available: https://www.degruyter.com/view/j/jhsem.2010.7.1/jhsem.2010.7.1.1732/jhsem.2010.7.1.1732.xml?format=INT&intcmp=trendmd

[61] K. Stowe, J. Anderson, M. Palmer, L. Palen, and K. Anderson, "Improving classification of twitter behavior during hurricane events," in *Proc. 6th Int. Workshop Natural Lang. Process. Social Media*. Assoc. Comput. Linguistics, 2018, pp. 67–75. [Online]. Available: http://www.aclweb.org/anthology/W18-3512



BENJAMIN A. RACHUNOK received the B.S. degree in industrial and systems engineering from North Carolina State University. He is currently pursuing the Ph.D. degree with the School of Industrial Engineering, Purdue University. He is also an R&D Intern with Sandia National Labs. His research interest includes developing data scientific methods for the holistic measurements of community resilience in anticipation and response to major natural hazards. He was a recipient of the Lee Chaden Fellowship and the Estus H. and Vashti L. Magoon Excellence in Teaching Award.

JACKSON B. BENNETT received the B.S. degree in ecological and environmental engineering from Purdue University, in 2018, where he is currently pursuing the M.S. degree with the Laboratory for Advancing Sustainable Critical Infrastructure (LASCI), School of Industrial Engineering. He is also conducting the Interdisciplinary Ecological Sciences and Engineering Graduate Program in Purdue University. His research interests include the intersection of data mining, and the environment with a focus on sustainable and resilient systems.

ROSHANAK NATEGHI received the M.Eng. degree in mechanical engineering from Imperial College London, in 2006, and the M.S.E. and Ph.D. degrees in environmental engineering from Johns Hopkins University. She was an NSF Science Engineering and Education for Sustainability (SEES) Fellow; jointly appointed between Johns Hopkins University and Resources for the Future. She is currently a Faculty Member with Purdue University, jointly appointed between the Industrial Engineering and Environmental Department and Ecological Engineering Department. She is also the Director of the Laboratory for Advancing Sustainable Critical Infrastructure (LASCI). Her research interests include advancing the theory, and practice of community-centric resilience through developing advanced analytics to facilitate better anticipation and mitigation of both rapid and slow-onset hazards.