



Tipping the scales: how geographical scale affects the interpretation of social media behavior in crisis research

Rachel Samuels¹ · Jiajia Xie² · Neda Mohammadi¹ · John E. Taylor¹

Received: 20 October 2020 / Accepted: 22 December 2021 / Published online: 5 February 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Our relationship with technology is constantly evolving, and how we use technology in disasters has evolved even faster. Understanding how to utilize human interactions with technology and the limitations of those interactions will be a crucial building block to contextualizing crisis data. The impact of geographic scale on behavioral change analyses is an unexplored facet of our ability to identify relative severities of crisis situations, magnitudes of localized crises, and total durations of disaster impacts. Within this paper, we aggregate Twitter and hurricane damage data across a wide range of geographic scales and assess the impact of increasing scale on both the recognition of extreme behaviors and the correlation between activity and damage. The power-law relationships identified between many of these variables indicate a direct, definable scalar dependence of social media aggregation analyses, and these relationships can be used to inform more intelligent, equitable, and actionable social media usage in emergency response.

Keywords Crisis response · Crisis informatics · Social media · Vulnerability · Hurricanes

1 Background

As the supply of data from humans-as-sensors continues to increase, understanding individual data streams in the context of our multi-layered and multi-networked society is becoming more difficult. Social media is increasingly looked to as a potential source of additional information in the notoriously information-scarce environment of crises (Reuter and Kaufhold 2018) for increasing hyper-local situational awareness and improving localized crisis response. The crisis informatics field has continued to flourish and expand alongside the seemingly ever-increasing quantities of available social media data and methods of analyzing that data. As a result, the applications for social media on situational awareness during crises has expanded to include event detection (Sakaki et al. 2010), resource availability and need (Choe et al. 2017), and mobility monitoring (Wang and Taylor 2016a). Analytical methods range from sieving individual posts for information (Ashktorab et al. 2014), to

✉ John E. Taylor
jet@gatech.edu

¹ Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

² Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

analyzing geographic changes in sentiment and behavior for informing gestalt-level decisions (Jongman et al. 2015; Kryvasheyev et al. 2016).

The big data revolution has clearly opened a vast area of possibilities for crisis response (Qadir et al. 2016). One of the greatest strengths of the field is the diversity it contains and the range of techniques available for the processing of this ever-increasing and changing pool of data. Applications developed in the field are being used by international aid organizations (Imran et al. 2014), and strides are being made for local response implementations of social media analysis as well (Tapia and Moore 2014). That said, the range of diversity of applications and methods can also impede the process of building a solid foundation. Researchers both external and internal to crisis informatics have noted criticism of social media applications' limits with respect to data bias, social inequality, and lack of confirmed validity (Imran et al. 2015; Jiang 2018).

More data is available; however, big data are not complete data. There has been a consistent call for us to critically interrogate the assumptions and capabilities of big data in the context of our political and urban usage (Boyd and Crawford 2012). As the reach and amount of available data increase, holes in that data become both less obvious due to the existing volume and yet more harmful due to the increasing prevalence of that data's use (Morstatter and Liu 2017). Social media, especially when used for crisis response, is not exempt from this call. This is especially true in the case of crisis response, where information availability alone can tip the scales of resource distribution. To ensure more equitable and intelligent use of social media data in crisis response, researchers need to understand the social, spatial, and sociospatial limitations of that data. One critical piece of that understanding is understanding the non-conventional geographic scale at which social media data are capable of identifying disasters, and how much information is gained or lost by varying that arbitrary geographic scale.

Geographic scale (referring to the spatial extent of a geographic area under study—i.e., smaller scale refers to a smaller geographic area; not to be mistaken for scale in cartography, which is a representative fraction) is less important in analyses sifting through individual posts, but it becomes more relevant when determining the likelihood of each of those posts appearing in a specific place and time. Spatiotemporally aggregated data can be key to identifying an expected baseline level of activity (Toepke 2018); identifying the proportions of a population represented by that Tweet (Mislove et al. 2011); monitoring human mobility during and after crises (Wang and Taylor 2016a); identifying drop-offs in activity alongside spikes in activity (Samuels et al. 2020); detecting urban emergencies through geographic and semantic clustering (Wang and Taylor 2019); and ultimately providing a social lens through which the social benefits or ramifications can be, at a minimum, glimpsed (Shelton et al. 2015).

Aggregate behaviors in real-time can be used to analyze relative disaster severity and magnitude (Zhang et al. 2016, Chen et al. 2020). As is often echoed in the literature (Wang and Taylor 2017, Wang and Taylor 2016a), disasters are not disasters because of high wind speeds or unprecedented amounts of flooding; disasters are disasters because of how they interrupt and, sometimes, forever change a society's functions. Disasters are inherently social and, ultimately, defined by societal vulnerabilities. The impacts and vulnerabilities associated with a disaster need to be defined in multiple dimensions, with an emphasis on social norm disruptions (Guan and Chen 2014). Spatiotemporal aggregation, then, can give us the pre- and post-impact phase reference points called for by Guan and Chen (2014).

However, we currently lack an understanding of the scale at which to generate these reference points. Previous research, particularly focusing on social disruptions, has identified the existence of a scalar dependence on the correlation between hurricane damage

and Twitter activity fluctuations (Shelton et al. 2014), but did not further investigate it. Research into the spatial biases of social media during disasters has identified a power law relationship between the number of people in cities and the number of Tweets generated in those cities during a disaster. However, the authors note the need for information on how these patterns vary at different spatial scales (beyond the city and super-neighborhood scales) (Fan et al. 2020). As power law relationships have been often identified in social networks, understanding the prevalence of scale-independent behaviors within the context of disasters is necessary to understand the intrinsic biases of social media information. Additionally, subsequent calls for further research highlight how understanding the scalar dependencies of social media data will improve our reference for the data's place in geographic, temporal, and social space (Jessop et al. 2008), thus improving our total understanding of the social significance of Twitter activity trends.

A massive disaster such as a hurricane or an earthquake is composed of hundreds of small ones: flooded neighborhoods, downed overhead power transmission lines, and trees thrown through roofs by gale-force winds (Wurman and Kosiba 2018). These disasters happen to more than individuals, but less than the whole of society. Disasters of varying magnitudes can happen to small neighborhoods, along vast swathes of a river, or through power outages across a city. If a 911 call can recognize a disaster happening to an individual, at what scale can social media recognize and triage emergencies affecting more than individuals? Is it limited to disasters occurring to thousands of people, or can it also identify disasters affecting smaller groups? The more we understand the capacity of humans-as-sensors to identify the location, relative severity, and magnitude of the localized disaster, the more useful social media will be to emergency response (Raue et al. 2013). Understanding how scale impacts the recognition of behavior will also enable us to reduce the obfuscation of any minority behaviors occurring at small scales that are drowned out by those happening at larger ones (Chen et al. 2013).

Within this paper, we analyze and present the scalar dependencies of aggregate social media analyses. We focus primarily on the ability of social media to identify localized disasters, i.e., to distinguish groups or areas that are being impacted by the disaster more extremely than the broad geographic region. The connection between social media activity and the presence of extreme danger or disaster has been noted in several pieces of literature (Guan and Chen 2014; Kryvasheyeu et al. 2015), and we are specifically testing the scalar dependencies of that connection. In order to do this, we test at different scales (1) the power of social media (Twitter) activity to identify distinct clusters of similar behavior, (2) how much behavior is identified as non-normal or extreme during a disaster, and (3) how the strength of the correlation between Twitter activity and hurricane damage varies. Understanding the shape of the relationships between these three factors and changing scale will improve our understanding of how to maximize the benefits of decreasing scale (more specificity of place) while minimizing the costs (less confidence in correlations). These assessments are codified in the following hypotheses:

H1 The distribution of changes in social media behavior and the identification of behavioral clusters are statistically distinct at smaller scales.

In order to assess this hypothesis, we qualitatively investigated the statistical distribution of Twitter representation (Tweets per person) across the Houston Metropolitan Area using each of the spatial nets (see Fig. 1 and Sect. 2.3). We secondarily quantitatively investigated the scale at which the distributions of Twitter representation cannot be statistically

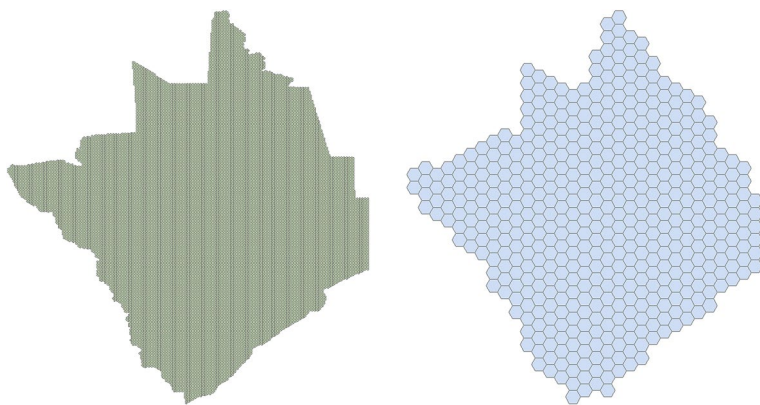


Fig. 1 (Left) Spatial net consisting of 1 km² hexagons across Houston. (Right) Spatial net consisting of 80 km² hexagons across Houston

distinguished between increasing spatial aggregation scales for both steady and perturbed state days.

Provided that some aspects of Twitter behavior are scale dependent, we should be able to use the dependency to identify how scale affects crisis-relevant analyses. Most of social media crisis analyses operate on the assumption that humans affected by crisis will change their behavior, and increasing amounts of evidence show that hurricanes produce localized crises at a small scale (Lieberman-Cribbin et al. 2017; Wurman and Kosiba 2018). In order to be useful to crisis managers who need specific and localized information, we should understand the smallest scale at which in-crisis behavior changes are identifiable and the potential information value trade-offs of decreasing or increasing the scale of analysis. This leads to our second hypothesis:

H2 The identification of crisis-induced, extremely high or extremely low amounts of non-normal Twitter activity is scale-dependent.

For this second hypothesis, we quantitatively assess the percentages of areas that have deviated from their steady state norms during Hurricane Harvey. The assumption of crisis informatics is that disasters break the ability of society to function normally (Guan and Chen 2014). How social media identifies and codifies those breaks in normal functioning is important to understand for trying to identify small-scale crises in a sea of larger crises. We also need to understand if the aforementioned connection between social media activity and hurricane damage is more or less consistent for identifying those small-scale crises. This leads us to our third hypothesis:

H3 The strength of the correlation between Twitter activity and infrastructural damage is scale-dependent.

For this, we test previously identified correlations between Twitter activity deviations and records of hurricane damage at increasing spatial scales. Knowing how scale affects correlation strength, and therefore how it affects how confident we can be that activity is indicative of a disaster, is important for communicating with emergency managers.

These quantitative comparisons of the representative capabilities of aggregation techniques will also inform future tools and algorithms that seek a real-time metric for human need expressed through social media. To address these hypotheses, we chose to focus our efforts on the city of Houston, Texas circa Hurricane Harvey. With a population greater than 2.3 million people (U.S. Census Bureau 2016), the city of Houston had a large hurricane-affected population that, based on our analysis of Gulf-based city Tweeting behavior, also had a substantial number of affected Twitter-users.

2 Methods

2.1 Twitter data

All of the geolocated Twitter data for the greater metropolitan area of Houston for seven weeks prior to and one week following Hurricane Harvey's landfall were streamed by the authors through the Twitter API (Wang and Taylor 2016b, Roesslein 2020). Hurricane Harvey made its first landfall in Houston on August 25th in the evening; the hurricane then pivoted and returned on August 27th to deposit torrential, record-breaking rains. For our analysis, we defined our perturbed state—the period of time during which non-normal behavior would be expected—as one day prior to the first landfall through the week following landfall (August 24th–September 1st). To identify non-normal behavior, we needed to select a steady state to act as our baseline for “normal” behavior. We defined this steady state as the period from July 15th to August 16th, following prior research describing the time period length necessary to generate a sufficiently stable analysis (Toepke 2018a); a longer period would increase the influence of both seasonality and population flux. The steady state behavior has a left-leaning lognormal distribution, matching prior findings (Zaman et al. 2014). We also allowed for a transitional state, during which the hurricane would broadly impact Twitter posting behavior through anticipation of harm but not through actual hurricane damage or events. This state is defined as the period from the day Harvey was identified as a tropical storm through the day before our perturbed state begins (August 17th–August 23rd).

With respect to important dates, it should be noted that Houston experienced the most infrastructural damage and flooding on August 27th and not when Hurricane Harvey first made landfall. As such, many of the following analyses focus on behaviors identified on August 27th. The sets of steady state and perturbed state Tweets were temporally aggregated by day, transformed into individual points through ArcGIS, and plotted using their latitude and longitude attribute information in ArcGIS. The Zip Code Tabulation Areas (ZCTAs) and 2010 census tract shapefiles were downloaded from the Harris County GIS data portal (Harris County 2019).

2.2 Population data

The census data and census tracts are not at a sufficiently fine resolution to enable understanding of the nuances of neighborhood-scale behavior during a crisis. The tracts further from the city center can be as large as 1500 km², so we need to find a method of increasing the resolution of the population data. The geographic information science field has historically utilized National Land Cover Database (NLCD) data to increase the granularity of census data with substantial accuracy (Reibel and Agrawal 2007). The NLCD contains

Table 1 Multiple linear regression results for the NLCD land class types and the census data

NLCD Class	Constrained optimum model coefficient
Open area	0.071
Low intensity	NA
Medium intensity	1.79
High intensity	NA
Rest	NA

All Coefficients are subject to a non-negative constraint (Intercept = 2973.72)

a raster file with 30 meter (m) by 30 m cells that have been classified, through satellite imagery, as one of 16 classes. The classification includes four classes of developed land: open space, low intensity, medium intensity, and high intensity. These classifications are determined using the amount of water-permeable and water-impermeable land; thus, the “open space” designation does not necessarily describe areas with no people, but rather areas with a relatively smaller (<20%) amount of concrete-covered land, like suburbs. The raster cells from the 2011 dataset that were classified as “developed” were extracted and clipped to the greater metropolitan Houston Area. Using ArcGIS’ Raster to Point function, each of the raster cells were transformed into points located at the center of each cell and spatially joined by count into the census tracts for Houston. Using the counts of each type of NLCD class and the population record for each census tract, a multiple linear regression analysis was performed to determine the expected contribution of each type of land type to the tract’s population (Bian and Wilmot 2015). The model is presented as:

$$Pop_t = \beta_0 + \beta_1 OpenSpace_t + \beta_2 LowInt_t + \beta_3 MedInt_t + \beta_4 HighInt_t + \beta_5 Rest_t + \epsilon \quad (1)$$

For estimating the parameters, Pop_t is the population for a given census tract t ; $OpenSpace_t$ is the number of NLCD cells described as “Open Space (Developed)” within the census tract; $LowInt_t$ is the number of cells described as “Low Intensity (Developed)”; $MedInt_t$ is the number of cells described as “Medium Intensity (Developed)”; $HighInt_t$ is the number described as “High Intensity (Developed)”, and $Rest_t$ is the accumulated number of cells with other labels. Note that, since population cannot be negative, we added non-negative constraints on β_1 , β_2 , β_3 , β_4 , and β_5 . The results of the regression analysis are presented in Table 1. The model has an adjusted R-squared of 0.65. This model is used to estimate population at smaller geographical scales such as spatial nets defined in the next section. The five coefficients explained above are aggregated to assigned weighted population to each cell.

2.3 Spatial nets

A common method of aggregation, particularly when social factors are considered, is to use census tracts or ZCTAs (Grubesic and Matisziw 2006). Although this gives arguably the most accurate nighttime population of the aggregated areas, their boundaries can be of widely varying shapes and sizes, and their jurisdictions are strongly influenced by the boundaries of socially biased fragmentations. These social factors can contribute to data

bias, and the size and shape differences can contribute to the Modifiable Area Unit Problem (MAUP) (Jelinski and Wu 1996; Grubestic and Matisziw 2006). As such, we chose to create our own spatial nets of varying scales and to include the ZCTA boundaries for comparison.

With the social media data temporally aggregated and mapped and the population data spatially disaggregated, we designed twelve spatial nets within which to aggregate the population data and the Twitter stream for each day. These spatial nets are composed of a series of interlocking shapefiles that cover the greater Houston area. One of these nets consists of the Houston ZCTA zones; the others were composed of uniformly shaped, tiled hexagons. As many of the problems identified with using manmade boundaries for spatial aggregation are related to their varying sizes and shapes, we wanted to develop a spatial net design that could be deployed across a large area, regardless of country of interest, and could be scaled according to the intended research design. This has the advantage over ZCTAs, which vary in size and shape. The ZCTAs within Houston alone, for instance, range in size from 0.16 square kilometers (km^2) to 677.20 km^2 .

We generated the spatial nets of equally sized and shaped hexagonal polygons through ArcGIS' Generate Tessellation function. Hexagons are better suited for tiling large geospatial areas because they reduce edge effects that can be exacerbated by intersecting rectangles and are more scalable on a curved surface like the globe (Carr et al. 1992; Polisciuc et al. 2016). The eleven hexagonal nets consist of hexagons that have square areas of, respectively, 0.25 km^2 , 0.5 km^2 , 0.75 km^2 , 1 km^2 , 5 km^2 , 10 km^2 , 15 km^2 , 20 km^2 , 35 km^2 , 50 km^2 , and 80 km^2 . These areas were chosen following the guidelines listed in the Spatial Association of Scalable Hexagons described by Potter et al., which suggests choosing sizes that closely mimic the behavior that is being studied or the sampling size, or the spatial dependence of the data (Potter et al. 2016). As such, these sizes mimic the range of sizes of Houston census tracts and the spatial spread of the Twitter data. With respect to the upper limit of 80 km^2 , we determined from the stated interests of emergency management personnel that information on areas larger than that scale provided very little actionable or useful information in terms of aid distribution or the presence of local disasters. A comparison between the 1 km^2 and 80 km^2 nets is shown in Fig. 1 for a scalar reference.

We summed the number of Twitter posts for each day and the population values of each of the NLCD points within each polygon of each net. Following the methods listed in Kryvasheyev et al. (2016), polygons that did not contain at least one Tweet per day during either the steady state or perturbed state were removed. For validation purposes with respect to the hurricane damage, we additionally plotted the Federal Emergency Management Agency building level damage assessments (FEMA damage assessments) collected in the days following Hurricane Harvey. These assessments are geolocated and the buildings they reference are classified as "Affected", "Minorly Affected", "Majorly Affected", and "Destroyed". We converted this ordinal scale of damage into a numerical ordinal scale, for which "Affected" is classified as a "1" and "Destroyed" is classified as a "4". For each polygon of each net, we additionally calculated the total number of damage assessments performed and the maximum and average assigned damage value of the polygons.

2.4 Analytical methods

H1 The distribution of changes in social media behavior and the identification of behavioral clusters is statistically distinct at smaller scales.

In order to identify the scalar interval of aggregation at which the distribution of Tweets per person changes significantly, we created empirical cumulative density functions (CDFs) of the Twitter activity per person for each net on a daily basis. These intra-net CDFs consisted of all of the Twitter representation values identified on a single day in all of the polygons of the net. We produced thirteen CDFs for five days of the steady state and for each day of the perturbed state. We then used the Kolmogorov–Smirnov II test on each pair of CDFs to test the likelihood that the CDFs were produced from the same parent distribution (Massey 1951).

H2 The identification of crisis-induced, extremely high or extremely low amounts of Twitter activity is scale-dependent.

In order to understand the distribution of perturbed state Twitter posting counts that were either much higher or much lower than the “normal” behavior observed in the steady state, we used cumulative distribution functions (CDFs) to compare perturbed state Twitter activity to the steady state Twitter activity. We used the steady state post counts for each of the spatial nets to generate a series of CDFs. These CDFs represented the distribution of Twitter activity for a single area across each day of the steady state period. For instance, for a given area **A** within the 10 km² net, we created a CDF from all of area **A**’s steady state Twitter activity counts by day. We then took the perturbed state Twitter activity on a given day, such as August 27th, and used the generated steady state CDF of activity to determine what percentage of steady state days had produced less than the number of Twitter posts produced on August 27th in area **A**. For instance, a result of 0.90 would indicate that the perturbed state activity on August 27th was higher than the activity produced on 90% of the days in the steady state, and a result of 0.10 would indicate that the perturbed state activity was only higher than 10% of days in the steady state.

We used each CDF to assess, for each area and each day of the perturbed state, the cumulative likelihood of observing a certain number of Tweets in that area on that day. We categorized this likelihood as being normal, non-normal, or extreme. Although, as stated, we are most interested in extreme values, we included an analysis of non-normal to provide a reference for the impact of how the threshold of “extreme” amounts of activity is defined. Using empirical rule values ($\bar{v} \pm C\sigma$, where $C = 1, 2, 3$, $\bar{v}$ is the population mean, and σ is the population standard deviation), non-normal social media behavior was defined as being less than 16% of steady state values or greater than 84% of steady state values. Extreme social media behavior was defined as being less than 5% of steady state values or greater than 95% of steady state values. To identify the effect of scale on observing extreme (and non-normal) values and so understand the prevalence and significance of activity bursts, clustering, or drop-offs, we took the distribution of the likelihood of observing the perturbed state Twitter activity and analyzed the distributions of those probabilities across nets and days of the perturbed state.

H3 The strength of the correlation between Twitter activity and infrastructural damage is scale-dependent.

Finally, to assess how scalar aggregation affects the previously identified correlation between non-normal Twitter activity and hurricane damage, we applied the statistical test for Kendall’s rank coefficient to the Twitter activity per person within each area and the

average FEMA building level assessment designation recorded within each area (Kendall 1938). These values were determined on a daily basis and compared across spatial nets.

3 Results

H1 The distribution of changes in social media behavior and the identification of behavioral clusters is statistically distinct at smaller scales.

The comparisons of the Kolmogorov–Smirnov II test results are depicted in Fig. 2. The null hypothesis for the test is that the two populations are drawn from the same population. With larger p-values (distribution pairs with p-values greater than 0.05 are displayed as gray squares), the less certain we are that we can reject the null hypothesis.

The top two figures show the results of two days of the steady state, August 8th and August 12th (Fig. 2a–b), and the bottom two figures show the results of two days of the

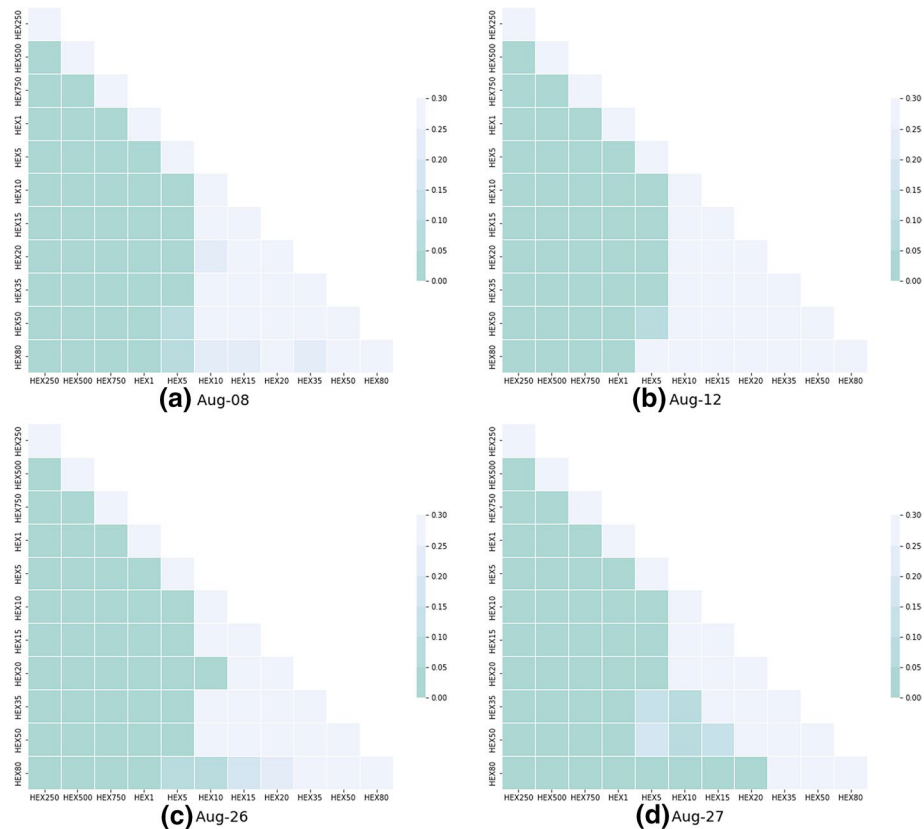


Fig. 2 a–d Comparison of the ρ values, showing that the spatial nets have distinct daily distributions using the Kolmogorov–Smirnov II test. The census tracts net is represented in the furthest left column and bottom row. The top two figures were created from steady state values (August 8th and August 12th), and the bottom two figures were created from perturbed state values (August 26th and August 27th). Teal color represents statistical significant ($p < 0.05$)

perturbed state, August 26th and August 27th (Fig. 2c-d). It should be noted that the CDF and statistical test results for the census tracts are on the outermost edge of the figures, and those results, interestingly, are not equivalent to either the very small nets or the very large ones. Across each set of tests, the larger nets' distributions are more similar to each other, and nets that are more similar to each other in size also have more similar distributions. Apart from those trends, both of the steady state graphs show the significant decrease in the certainty that the spatial nets' CDFs are from distinct distributions between 5km² and 10km². This pattern of 5-10km² spatial trends breaks during the perturbed state, in which small-scale areas begin to behave more similarly to each other. This opened window in the smaller spatial scales exists from August 26th through September 1st, although it diminishes in size beyond August 29th. It is largest on the day of maximum damage from rainfall, August 27th.

H2 The identification of crisis-induced, extremely high or extremely low amounts of Twitter activity is scale-dependent.

In order to understand the distribution of perturbed state values that were either much higher or much lower than the “normal” social media behavior observed in the steady state, we used CDFs to compare perturbed state values to the steady state values. For instance, for a given area A, within the 10 km² net, we estimated a CDF from all of that area's steady state values. We then took the perturbed state Twitter activity on a given day and used the CDF to determine what percentage of steady state values were lower than that perturbed state value. For instance, a value of 0.90 would indicate that the perturbed state value was higher than 90% of the steady state values. We then determined the distribution of those probabilities and compared them across scales to identify the effect on scale on the prevalence of extreme or non-normal values.

We quantitatively assessed the percentage of values on the day of maximum rainfall and damage, August 27th, which exhibited non-normal or extreme social media behavior. Non-normal behavior was defined as being one standard deviation from the mean, i.e., less than 16% of steady state values or greater than 84% of steady state values. Extreme social media behavior was defined as being less than 5% of steady state values or greater than 95% of steady state values. We additionally included the percentage of areas that exhibited normal Twitter activity, within the central 68% of observed Twitter activity, on both Figures 3 and 4 for comparison.

The percentage of areas identified with each of those kinds of social media interactions compared to the size of the spatial nets are compared in Fig. 3 and Fig. 4. To define the relationships between each set of variables, we used the method of maximum likelihood to estimate a scaling exponent for a power law relationship. That is we estimate a discrete power law probability defined as follows (Gillespie 2014):

$$f(x) = \frac{x^{-\alpha}}{\xi(\alpha, x_{\min})} \quad (2)$$

$$\xi(\alpha, x_{\min}) = \sum_{n=1}^{\infty} (n + x_{\min})$$

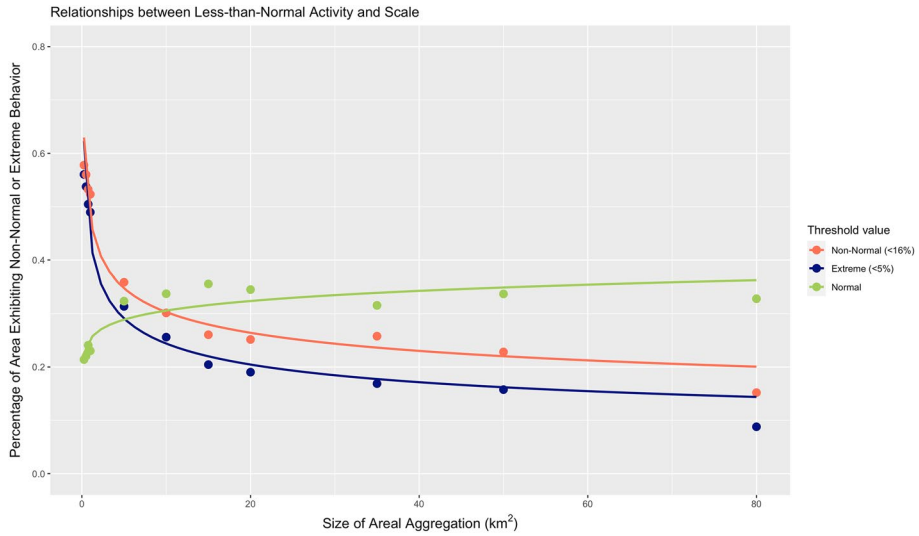


Fig. 3 Comparison of the percentages of areas exhibiting non-normal or extremely low Twitter activity behavior between spatial nets, fit to discrete power law distributions

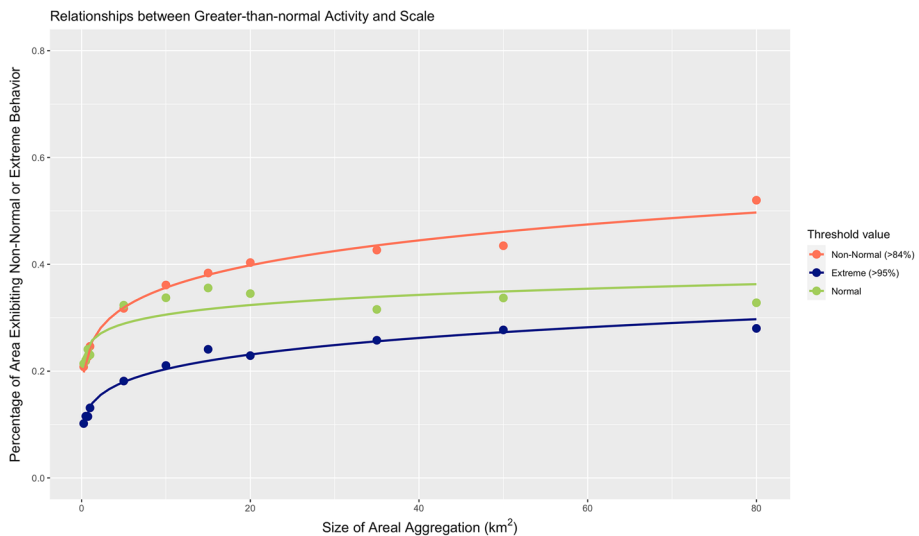


Fig. 4 Comparison of the percentages of areas exhibiting non-normal or extremely high Twitter activity behavior between spatial nets, fit to discrete power law distributions

Here x represents the geographic scale, and $f(x)$ denotes the predicted variable (i.e., the percent of areas with extremely high or low numbers of Tweets, the correlation coefficient between Tweets and FEMA damage ratings, or the percentage of geographic area excluded from analysis) (Stumpf and Porter 2012). The minimum x value, x_{min} , (geographic scale) was determined by minimizing the Kolmogorov–Smirnov test statistic, and the goodness

of fit of the estimated parameters for these relationships were assessed using the Kolmogorov–Smirnov test (Clauset et al. 2009). Although there are too few data points to more positively confirm the presence of a power law relationship, the Kolmogorov–Smirnov test statistics show that the data distributions for each of the tested relationships could be derived from a power law distribution. Vuong’s test was used to compare the relative distance between the sample distributions, the estimated power law distributions, and log-normal distributions estimated from the same data (Vuong 1989). The estimated discrete power law parameter, the minimum x values, the Kolmogorov–Smirnov test statistics, the Kolmogorov–Smirnov test results, and the Vuong test results are presented in Table 2.

Of most importance for applications, the relationship between increasing geographic scale and the identification of extremely low social media interaction is negative, while the relationship between scale and the identification of extremely high social media interaction is positive. The inclusion of non-normal activity analysis was provided as a reference for the impact of the threshold at which the amount of activity could be interpreted as “extreme”. We see the impact of increasing the boundary for the classification from outside the central 68% to outside of the central 95% in the identification of non-normal and extreme perturbed state activity for both sets of behavior trends. Decreasing the threshold at which an observed behavior is classified as noteworthy obviously increases the number of noteworthy observations; however, this effect is slightly larger for the extremely high values, and the effect is more profound at higher scales (greater than 40 km²).

The difference between the percentage of areas identified as exhibiting extreme at the smaller scales is much larger (approximately 80% for scales less than 0.5 km²) than the percentage identified at the larger scales (approximately 53%). The geographic coverage of those areas, however, is quite similar due to the increased removal of areas without Twitter activity at the smaller scale. By excluding from analysis all areas that did not have a single Tweet across the steady state period or a single Tweet across the perturbed state, this resulted in a substantial amount of geographic coverage reduction in the smaller scales. We identified a logarithmic increase in excluded area with decreasing scale. For example, 56% of Houston was excluded in the 0.25 km² net, 48% was excluded from the 1 km² net, and only 19% was excluded from the 20 km² net. This relationship is represented in Fig. 5.

H3 The strength of the correlation between Twitter activity and infrastructural damage is scale-dependent.

Table 2 Results of the maximum likelihood estimation of sample distribution fit to power laws

Variable	α	$x_{\min}$	K-S test	Vuong’s test
Extremely high activity	1.93	10000m ²	0.23*	Lognormal More Likely
Non-normally high activity	1.91	10000m ²	0.23*	Lognormal More Likely
Non-normally low activity	1.43	250m ²	0.17*	Power Law More Likely
Extremely low activity	1.47	250m ²	0.16*	Power Law More Likely
% of area removed	1.49	250m ²	0.16*	Power Law More Likely

* $p < 0.05$ for the Kolmogorov–Smirnov test statistic’s null hypothesis (that the sample distribution was drawn from a power law distribution) could not be rejected

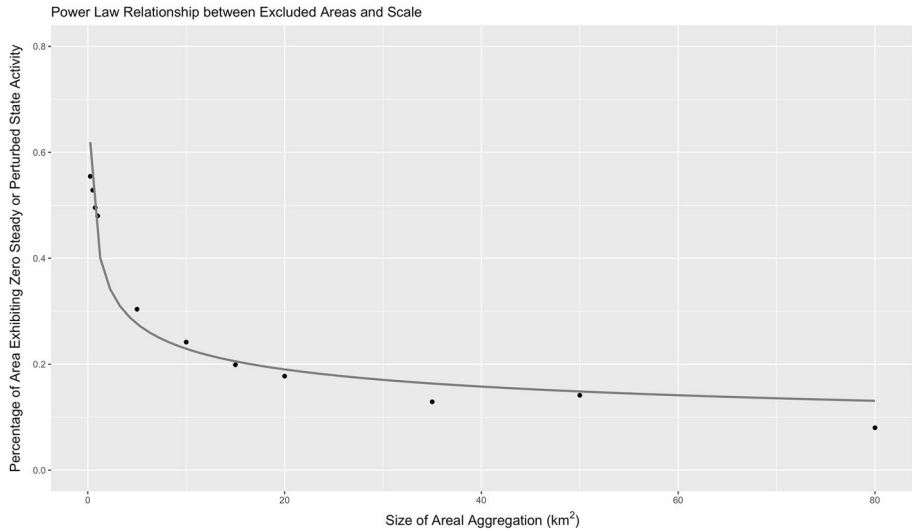


Fig. 5 The relationship between the percentage of the total study area (the greater metropolitan area of Houston) excluded from the analysis on account of not having sufficient Twitter activity, as defined in Sect. 2.3, and the geographic scale at which the data was aggregated

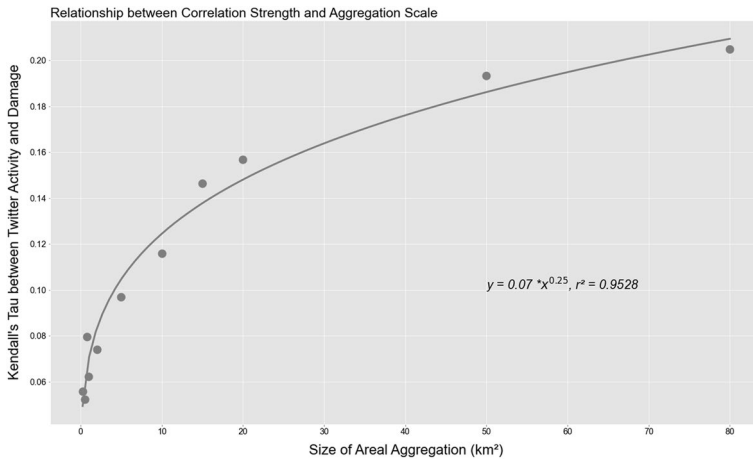


Fig. 6 Kendall Rank Correlation Coefficient strength between Twitter Activity per person and the average FEMA building damage assessment rating within each area

The relationship between scalar aggregation and the strength of the correlation between Twitter activity on August 27th and hurricane damage is presented in Fig. 6. This data also shows a direct proportionality between the scale of aggregation and the strength of human social media behavior signals in the analysis. To quantify this, we also fit the data to a continuous power law relation, commonly defined as: $f(x) = \beta x^\gamma$. It should be noted that each value is significant except for that of the 35 km² net, which

was excluded from the model fitting due to its lack of statistical significance ($p = 0.11$). The correlation strength increases sharply until the 10 km² scalar aggregation with no apparent sacrifice of statistical significance.

4 Discussion

First and foremost, the analyses presented herein show that the stories data tell differ when read at different scales. With any scalar analysis, we would expect more variation at smaller scales across time and between areas. There is an expected tradeoff between certainty at large scales and specificity at small scales. The crisis community has recognized this; however, the effect of scale on the ability of social media to (H1) identify distinct clusters of geographic social media interaction changes, (H2) identify non-normal or extreme social media behaviors, and (H3) provide statistical confidence that social media behavior changes indicate danger had not been explored. Within our three analyses, we have identified the precise relationship between scale and social media signal behavior in the hopes of making the crisis informatics community more aware of how scale can influence multiple facets of the findings of social media analytics.

In addressing H1, the steady distribution of Twitter activity observed within spatial nets of an area greater than 10 km² is not statistically distinct from the distribution observed in another spatial net greater than 10 km². The inverse is also true; nets smaller than 5 km² are not as likely to be from the same distribution as each other. This confirms the existence of social and place-related social norms occurring at scales smaller than 5 km²; i.e., in the broader scheme of human activity, most locations and events occupy a geographic space smaller than 5 km², so smaller scale analyses incorporate spatial segments that contain different locations and events. With analyses that compute a spatial relationship across those segments, the produced answer may not accurately reflect any of the spatial constituents. At the 10 km² scale, the highs and lows of activity are averaged across more data, minimizing the impact of the grouped extremes. Even at the 0.25 km² scale, the analysis identifies different distributions of activity and bursting than those from the 0.5 km² scale.

This clear cut-off of statistical difference disappears during a crisis state, however. People within 0.5–5 km² begin behaving in more similar ways, and, during the day of maximum rainfall and damage, the difference between the distributions at most of the smaller scales decreases substantially. This confirms research showing hurricanes impacting cities differently at small scales due to small vortices and flooding susceptibility. It is likely that this sudden homogenizing of social media behavior at the 0.5–5 km² is indicative of the average effect of distinct, human-impacting hurricane phenomena. The similarity of social media behaviors at these scales during a disaster may also be indicative of reduced population mobility or infrastructure limitations.

With respect to our second hypothesis, the percentage of extreme social media behavior in a crisis state and the strength of the correlation between extreme behavior and hurricane damage are both definably dependent on geographic scale. We identify and define the effect of geographic scale on the identification of extreme social media behaviors, and we show clearly that the effect is different for different extreme behaviors. When we assessed the distribution of the likelihood of seeing each value in the perturbed state using the empirical cumulative distribution function generated in the steady state, we found that the smallest scales are likely to define the majority of active areas as

having either very low or very high activity. Few perturbed state values lie close to the average of the steady state. If the purpose of social media analysis is to identify areas with higher or lower severity, this indiscriminate binning of most values as “extreme” would not be ideal. The consistency across the perturbed state does, however, note a reliably consistent categorization of specific areas into extremes. This is the inverse of the social media behavior observed at the larger spatial scales, which has a wide variety of probabilities of the occurrence of values, and yet varies drastically from day to day of the perturbed state.

The effect of geographic scale on variability in data distributions is reinforced by the relationships between aggregation scale and each representation of Twitter behavior assessed above. In our data, we have identified six possible power law distributions, although three of them are more likely to truly be drawn from a power law distribution than the others. Power law distributions have attracted a large amount of attention in almost every field, ranging from microbiology to economics. They have been suggested as being present in nearly every natural system, although the statistical confirmation of the reality of such a claim has been questioned (Stumpf and Porter 2012). In the realm of social behavior and social networks, power law relationships have been shown to develop due to growth and preferential attachment (Barabási and Albert 1999). The prevalence of those two features in human networks has contributed to the identification of many social power law relationships, such as the famous “rich-get-richer” phenomenon (Newman 2005). In seeking a theoretical reason for our observations, we propose that the flux of people into urban centers could easily account for growth, and preferential attachment could be due to how social media-using demographics more commonly flock to urban centers with similar types of people (Shelton et al. 2015).

Recent work in social media has also identified additional power law relationships between the population size of city centers and the number of different types of Tweets generated within (Fan et al. 2020). Those results, which stem from the number of people in specific cities, and our results, which stem from the number of people within geographic scales, seem to indicate that growth and preferential attachment are present in the relative spatial clustering of those who want and are *able* to use social media more during a disaster. Because of these qualities, despite continuous growth in urban populations and social media users, the network of users specifically using social media during a disaster is organizing itself into a scale-free clustered network. This phenomenon may affect the relative ability of areas with greater numbers of social media users to receive more resources during a disaster purely due to the concentration of their voices.

In our data, we show that the number of identified extremely low events decreases exponentially with increasing scale, and the number of identified extremely high events increases exponentially with increasing scale. This increased identification of non-normal social media interaction at increased scales suggests the need to apply more stringent thresholds for activity marked as abnormally or extremely high at larger spatial scales. That said, the many decreased values identified at the smaller scale call for more stringent methods of investigation into these areas that are suddenly silent. In terms of sudden silence, previous research has identified that drop-offs in Twitter activity are also correlated with damage and theorized that those drop-offs are caused by social vulnerabilities more than social media behavioral choices (Samuels and Taylor 2020). Increased scales minimize the potential for a social media analysis to identify these drop-offs as extreme events, a factor that needs to be considered and addressed in social media applications.

In applying these to future analyses, we also show the influence of the MAUP on the use of ZCTAs in crisis informatics. ZCTAs vary widely in size and shape; the ones within

Houston, for instance, range in size from 0.16 km² to 677.20 km². This variance in size and socially constrained boundaries has been substantially critiqued in the field of critical GIS (Jelinski and Wu 1996; Saib et al. 2014). Across all analyses, the distributions appear closest to the values for the 5 km² net. As the average size of the census tracts for the area is approximately 7 km², this suggests that the potential spatial biases of census tracts in terms of Twitter representation may be more directly related to the tracts' geographic size and less their socially constructed boundaries. The variances in the sizes of the ZCTAs is an additional variable that, as we have shown, has a significant effect on the analytical results of an analysis.

As for the third hypothesis, concerning the relationship between these extreme values and hurricane damage, there have been multiple remarks in the literature regarding an expected relationship. Shelton et al. (2014) identified a discrepancy in correlation values at varying scales, noting the apparent necessity of including scale as a factor in any analysis comparing Twitter activity and hurricane damage. County-scale and state-scale correlations have been found to be moderately strong (Guan and Chen 2014; Shelton et al. 2014; Kryvasheyev et al. 2016); however, each author notes the influence of scale on their analyses. Guan and Chen hypothesize that “moving upward on the scale is likely associated with a larger amount of disruptions at a higher level of severity”, which would lead to a stronger, more significant “disaster” signal. Within our analysis, we are able to show the likelihood of a power law relationship between increasing analytical scale and the strength of the correlation between damage and Twitter activity. This relationship shows that we can be more confident in social media activity behaviors indicating a local hazard when we look at larger spatial scales.

Ultimately, the certainty involved in whether extreme social media behaviors function as good indicators of hurricane damage is dependent on geographic scale, showing once again that the tradeoff for geographic specificity is certainty of the identification of hurricane damage, and analyses performed at the census tract or county level need to incorporate the analysis' scale into their explanations of their findings.

5 Limitations and future work

5.1 NLCD data

Although one of the first uses of NLCD data was to estimate populations distributions on a fine scale, and despite its storied history in population estimation, it is still a fallible metric. In our case, two out of the five developed areas are considered “irrelevant”. Additionally, census data is limited to the “nighttime” population, i.e., where people sleep (home). Based on the Tweet text, we find that most people stayed home during the worst of the hurricane, and thus this nighttime population may double as hurricane-time population. We were incapable of incorporating evacuation dynamics for the hurricane, but evacuation orders were not widely issued for Houston. Evacuation likely had less impact on population dynamics for Houston than for areas more often affected by hurricanes.

5.2 Hurricane-specific Tweeting

Many studies in the field filter for Tweets that are directly related to the hurricane through text analysis. This limits the application of steady state versus perturbed state analysis, as

no one was Tweeting about Hurricane Harvey before it formed in the Gulf. We additionally wanted to incorporate areas Tweeting in a “business-as-usual” fashion during the hurricane. As such, using hurricane/disaster-specific Tweets was not possible, and our analysis undoubtedly incorporated some Twitter bots (Yang et al. 2019). We manually filtered some of the bots based on keywords (i.e., “jobs”, “FloodWatch”) determined through manual application of the OSoMe tool BotOrNot (Indiana University 2018). Furthermore, as the bots are likely unaffected by the hurricane, their activity changes would be normalized to zero, thus minimizing their effect on our analyses. Further research is necessary into the influence of Twitter bots on the Twitter distributions and analyses specific to disasters.

5.3 Power law relationships

A scale-free stationary state is difficult to prove. Criticisms of abundant labeling of power law dynamics require statistical tests that our data cannot satisfy, i.e., independent and dependent variables that range more than two orders of magnitude each (Stumpf and Porter 2012). Additionally, many statistical tests for the goodness of fit of power law distributions for binned data additionally require a longer tail than our data provides (Virkar and Clauset 2014). We have followed the statistical suggestions put forward by fervent critics of the search for power laws (Clauset et al. 2009) as well as we could. We, therefore, fit a discrete version of power law by maximizing the likelihood and perform two statistical tests, Vuong’s and Kolmogorov-Smirnov, to determine whether the distribution tends to be a lognormal or power law. Future research may explore this hypothesis by either increasing the number of geo-scales (expansion of scaling sample size) or consider extending the statistical tests to other exponential families.

Lastly, no two disasters are the same, either in terms of the damage caused or in terms of the affected society. The generalizability of the potential power law relationships and the distributions of Twitter activity to other cities and other disasters should be investigated.

6 Conclusion

Spatial scale does not constitute geography alone; it is directly tied to a wide variety of demographic, infrastructural, and analytical variances. Accounting for these confounding factors is a necessary component of crisis analytics. Within our paper, we hope to have provided a roadmap for crisis informatics researchers using social media to better understand how the chosen scale of their analysis will affect their results. Many of the potential power law relationships identified in this paper indicate an exponential tradeoff between the geographic specificity of smaller scales and the statistical certainty that an identified social media behavior represents an endangered population. Crises within a disaster context happen to individuals and communities; it is important to work toward using social media data (in addition to other forms of data generated by humans-as-sensors such as 311-Data) to improve our ability to correctly assess the severity and magnitude of an identified emergency. Both of those factors depend on geographic scale relative to the surrounding areas and that area’s own history. Future research should investigate different facets of scaling across alternative sources of data in analyses of crisis response and examine how various scales may relate to applications in practice. How we ascertain and contextualize our data,

then, is heavily influenced by how we structure our analysis, and we need to be wary of what dependencies might be tipping the scales.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grants No. 1760645 and 1837021. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contributions Conceptualization: Rachel Samuels, John E. Taylor; Methodology and Analysis: all authors; Writing—original draft preparation: Rachel Samuels; Writing—review and editing: all authors; Funding acquisition: John E. Taylor; Supervision: Neda Mohammadi, John E. Taylor.

Funding This material is based upon work supported by the National Science Foundation under Grants No. 1760645 and 1837021. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

- Ashktorab Z, Brown C, Nandi M, Culotta A (2014) Tweedr: mining twitter to inform disaster response. Proc 11th Int ISCRAM Conf 11:354–358
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- Bian R, Wilmot CG (2015) Spatiotemporal population distribution method for emergency evacuation. *Transp Res Record: J Transp Res Board* 2532(1):99–106. <https://doi.org/10.3141/2532-12>
- Boyd D, Crawford K (2012) Critical questions for Big Data. *Inform, Commun Soc* 15:662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Carr D, Olsen A, White D (1992) Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartogr Geogr Inf Syst* 19:228–236. <https://doi.org/10.1559/152304092783721231>
- Chen C, Neal D, Zhou M (2013) Understanding the evolution of a disaster—a Framework for Assessing Crisis in a System Environment (FACSE). *Nat Hazards* 65:407–422. <https://doi.org/10.1007/s11069-012-0371-6>
- Chen Y, Wang Q, Ji W (2020) Rapid assessment of disaster impacts on highways using social media. *J Manage Eng* 36(5):04020068. [https://doi.org/10.1061/\(asce\)me.1943-5479.0000836](https://doi.org/10.1061/(asce)me.1943-5479.0000836)
- Choe S, Park J, Han S et al (2017) A study on the real-time management and monitoring process for recovery resources using Internet of Things. *Int Res J Eng Technol* 4:2634–2639
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703. <https://doi.org/10.1137/070710111>
- Fan C, Esparza M, Dargin J et al (2020) Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters. *Comput Environ Urban Syst* 83:101514. <https://doi.org/10.1016/j.compenvurbysys.2020.101514>
- Gillespie CS (2014) Fitting heavy tailed distributions: thepowerlawpackage. *J Statist Softw.* <https://doi.org/10.18637/jss.v064.i02>
- Grubestic TH, Matisziw TC (2006) On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *Int J Health Geogr* 5:1–15. <https://doi.org/10.1186/1476-072X-5-58>
- Guan X, Chen C (2014) Using social media data to understand and assess disasters. *Nat Hazards* 74:837–850. <https://doi.org/10.1007/s11069-014-1217-1>
- Harris County Universal Services GIS Open Data. Harris County Universal Services GIS Open Data. <https://geoharriscounty.opendata.arcgis.com/>
- Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency. *ACM Comput Surv* 47:1–38. <https://doi.org/10.1145/2771588>

- Imran M, Castillo C, Lucas J et al (2014) AIDR: artificial intelligence for disaster response. Proc companion Publ 23rd Int Conf World wide web companion 159–162. <https://doi.org/10.1145/2567948.2577034>
- Indiana University (2018) Botometer® by OSoMe. In: Obs. Soc. Media
- Java A, Song X, Finin T, Tseng B. Why we twitter (2007) Proc 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis - WebKDD/SNA-KDD '07. ACM Press. <https://doi.org/10.1145/1348549.1348556>
- Jelinski DE, Wu J (1996) The modifiable areal unit problem and implications for landscape ecology. *Landsc Ecol* 11:129–140. <https://doi.org/10.1007/BF02447512>
- Jessop B, Brenner N, Jones MS (2008) Theorizing sociospatial relations. *Environ Plan D Soc Sp* 26:389–401. <https://doi.org/10.1039/d042246>
- Jiang B (2018) Trends in spatial analysis and modelling. Springer International Publishing, Cham
- Jongman B, Wagemaker J, Romero B, de Perez E (2015) Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals. *ISPRS Int J Geo-Inform* 4:2246–2266. <https://doi.org/10.3390/ijgi4042246>
- Kendall MG (1938) A new measure of rank correlation. *Biometrika*. <https://doi.org/10.2307/2332226>
- Kryvasheyev Y, Chen H, Moro E et al (2015) Performance of social network sensors during hurricane sandy. *PLoS ONE* 10:e0117288. <https://doi.org/10.1371/journal.pone.0117288>
- Kryvasheyev Y, Chen H, Obradovich N et al (2016) Rapid assessment of disaster damage using social media activity. *Sci Adv* 2:e1500779. <https://doi.org/10.1126/sciadv.1500779>
- Lieberman-Cribbin W, Liu B, Schneider S et al (2017) Self-reported and FEMA flood exposure assessment after hurricane sandy: association with mental health outcomes. *PLoS ONE* 12:1–15. <https://doi.org/10.1371/journal.pone.0170965>
- Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc*. <https://doi.org/10.1080/01621459.1951.10500769>
- Mislove A, Lehmann S, Ahn Y et al (2011) Understanding the demographics of twitter users. Proc Int AAAI Conf on Web and Social Media 5(1):554–557
- Morstatter F, Liu H (2017) Discovering, assessing, and mitigating data bias in social media. *Online Soc Netw Media* 1:1–13. <https://doi.org/10.1016/j.osnem.2017.01.001>
- Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46.5:323–351
- Polisciuc E, Maças C, Assunção F, Machado P (2016) Hexagonal gridded maps and information layers: a novel approach for the exploration and analysis of retail data. In: Proc SIGGRAPH ASIA 2016 Symposium on Visualization on - SA '16. ACM Press, New York, New York, USA, pp 1–8. <https://doi.org/10.1145/3002151.3002160>
- Potter KM, Koch FH, Oswalt CM, Iannone BV (2016) Data, data everywhere: detecting spatial patterns in fine-scale ecological information collected across a continent. *Landsc Ecol* 31:67–84. <https://doi.org/10.1007/s10980-015-0295-0>
- Qadir J, Ali A, ur Rasool R et al (2016) Crisis analytics: big data-driven crisis response. *J Int Humanit Action* 1:12. <https://doi.org/10.1186/s41018-016-0013-9>
- Raue S, Azzopardi L, Johnson CW (2013) #Trapped! social media search system requirements for emergency management professionals. Proc 36th Int ACM SIGIR Conf Res Dev Inf Retr - SIGIR '13 1073–1076. <https://doi.org/10.1145/2484028.2484184>
- Reibel M, Agrawal A (2007) Areal interpolation of population counts using pre-classified land cover data. *Popul Res Policy Rev* 26:619–633. <https://doi.org/10.1007/s11113-007-9050-9>
- Reuter C, Kaufhold MA (2018) Fifteen years of social media in emergencies: a retrospective review and future directions for crisis Informatics. *J Conting Cris Manag* 26:41–57. <https://doi.org/10.1111/1468-5973.12196>
- Roesslein J (2020) Tweepy: Twitter for Python. <https://Github.Com/Tweepy/Tweepy>
- Saib M-S, Caudeville J, Carre F et al (2014) Spatial relationship quantification between environmental, socioeconomic and health data at different geographic levels. *Int J Environ Res Public Health* 11:3765–3786. <https://doi.org/10.3390/ijerph110403765>
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. Proc 19th Int Conf World Wide Web 851–860. <https://doi.org/10.1145/1772690.1772777>
- Samuels R, Taylor JE (2020) Deepening the divide: crises disproportionately silence vulnerable populations on social media. *J Manag Eng*. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000848](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000848)
- Samuels R, Taylor JE, Mohammadi N (2020) Silence of the Tweets: incorporating social media activity drop-offs into crisis detection. *Nat Hazards*. <https://doi.org/10.1007/s11069-020-04044-2>

- Shelton T, Poorthuis A, Graham M, Zook M (2014) Mapping the data shadows of Hurricane Sandy: uncovering the sociospatial dimensions of “big data.” *Geoforum* 52:167–179. <https://doi.org/10.1016/j.geoforum.2014.01.006>
- Shelton T, Poorthuis A, Zook M (2015) Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information. *Landsc Urban Plan* 142:198–211. <https://doi.org/10.1016/j.landurbplan.2015.02.020>
- Stumpf MPH, Porter MA (2012) Critical truths about power laws. *Science* 335:665–666. <https://doi.org/10.1126/science.1216142>
- Tapia AH, Moore K (2014) Good enough is good enough: overcoming disaster response organizations’ slow social media data adoption. *Comput Support Coop Work CSCW Int J* 23:483–512. <https://doi.org/10.1007/s10606-014-9206-1>
- Toepke SL (2018) Minimum collection period for viable population estimation from social media. 138–147. <https://doi.org/10.5220/0006803901380147>
- U.S. Census Bureau (2016) Decennial Census 2010
- Virkar Y, Clauset A (2014) Power-law distributions in binned empirical data. *Ann Appl Stat* 8:89–119. <https://doi.org/10.1214/13-AOAS710>
- Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*. <https://doi.org/10.2307/1912557>
- Wang Q, Taylor JE (2016a) Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS ONE* 11:1–14. <https://doi.org/10.1371/journal.pone.0147299>
- Wang Q, Taylor JE (2016b) Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms. *J Comput Civ Eng* 30:04015004. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000469](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000469)
- Wang Y, Taylor JE (2019) DUET: data-driven approach based on latent dirichlet allocation topic modeling. *J Comput Civ Eng* 33:04019023. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000819](https://doi.org/10.1061/(asce)cp.1943-5487.0000819)
- Wang Y, Wang Q, Taylor JE (2017) Aggregated responses of human mobility to severe winter storms: an empirical study. *PLoS ONE* 12(12):e0188734. <https://doi.org/10.1371/journal.pone.0188734>
- Wurman J, Kosiba K (2018) The role of small-scale vortices in enhancing surface winds and damage in Hurricane Harvey (2017). *Mon Weather Rev* 146:713–722. <https://doi.org/10.1175/MWR-D-17-0327.1>
- Yang K-C, Varol O, Davis CA et al (2019) Arming the public with AI to counter social bots. *Hum Behav Emerg Technol*. <https://doi.org/10.1002/hbe2.115>
- Zaman T, Fox EB, Bradlow ET (2014) A Bayesian approach for predicting the popularity of tweets. *Ann Appl Statist*. <https://doi.org/10.1214/14-aos741>
- Zhang C, Zhou G, Yuan Q, Zhuang H, Zheng Y, Kaplan L, et al. GeoBurst (2016) Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval. *ACM*. <https://doi.org/10.1145/2911451.2911519>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.