MIT Open Access Articles

Quantifying human mobility resilience to extreme events using geo-located social media data

Massachusetts Institute of Technology

REGULAR ARTICLE   **Open Access**

Quantifying human mobility resilience to extreme events using geo-located social media data

Kamol Chandra Roy[1], Manuel Cebrian[2] and Samiul Hasan[1*]

*Correspondence:
samiul.hasan@ucf.edu
[1]Department of Civil,
Environmental, and Construction
Engineering, University of Central
Florida, Orlando, USA
Full list of author information is
available at the end of the article

Abstract

Mobility is one of the fundamental requirements of human life with significant societal impacts including productivity, economy, social wellbeing, adaptation to a changing climate, and so on. Although human movements follow specific patterns during normal periods, there are limited studies on how such patterns change due to extreme events. To quantify the impacts of an extreme event to human movements, we introduce the concept of mobility resilience which is defined as the ability of a mobility system to manage shocks and return to a steady state in response to an extreme event. We present a method to detect extreme events from geo-located movement data and to measure mobility resilience and transient loss of resilience due to those events. Applying this method, we measure resilience metrics from geo-located social media data for multiple types of disasters occurred all over the world. Quantifying mobility resilience may help us to assess the higher-order socio-economic impacts of extreme events and guide policies towards developing resilient infrastructures as well as a nation's overall disaster resilience strategies.

Keywords: Human mobility; Resilience; Geo-location data; Social media

1 Introduction

Increased population growth and interdependent infrastructure systems have made our cities and communities more vulnerable to extreme events [1, 2]. Natural disasters are responsible for a global $520 billion losses and moving 26 million people to poverty in every year [3]. To deal with such extreme events, a shift from reactive to pro-active policies focusing on disaster resilience is needed [4]. Resilience is commonly used to indicate the ability of a system or entity to return to its normal state after a disruption due to a disaster event [5]. Community resilience has been described as a process of linking to a network of adaptive capabilities that help to adapt after a disruptive event [6]. To assess resilience, depending on the fields and events, both qualitative [7–9] and quantitative [10–12] approaches exist. While it has been widely studied for physical infrastructure systems, resilience of socio-economic systems is hard to quantify. Human mobility is a key factor to understand the impacts of disasters to our social and economic activities since socio-economic development is strongly associated with mobility [13].

Human mobility analysis has drawn much attention in many research fields for its wide applications. Most of the studies have modeled mobility as probability distributions of the length of the traveled distance and the waiting time between any two displacements. Analyzing a wide range of data sets, studies have established that human mobility is not random rather it follows some specific patterns [14–17]. For instance, human mobility has been studied using large-scale trajectory datasets including bank notes [16], taxi data [18, 19], GPS observations [20], Wi-Fi [15], cell phone call recordings [21, 22], and social media posts [23–27]. These studies have found that mobility follows power-laws [14, 16, 18, 21, 22, 28–33], log-normal [19, 20], exponential distribution [34–39] or a combination of power-law and exponential distributions [35, 38].

During extreme events, human mobility goes through a significant perturbation compared to regular periods. People are less likely to move the same way in emergency situations, such as a hurricane, typhoon, earthquake and other natural or manmade extreme events, as they do in normal conditions. Understanding this perturbation will increase the effectiveness of disaster preparedness, information communication, reduce fatalities, and minimize economic losses [40, 41]. Despite its importance, few studies have investigated human mobility under disasters. Although studies have investigated how individuals behave during an extreme event [42–48], they are mainly based on post-disaster surveys with limited sample size. Based on these survey data, it is impossible to compare pre and post disaster human movements and measure mobility resilience at a system scale. Alternatively, analyzing mobile phone data Lu et al. [49] shows that the predictability of people's trajectory remains high during the three month period after the earthquake in Haiti in 2010. Social media data can also offer a promising direction in observing human movements during extreme events. Guan et al. [50] proposed a method to track the dynamics of social and infrastructure networks using Twitter and taxi and subway operations data. However, this study mainly focuses on the dynamic nature of certain properties of the networks during a disaster without quantifying resilience of those systems. A method that can quantitatively measure perturbations and recovery times of human mobility will greatly impact disaster management as well as in policy making towards building disaster resilient infrastructures, communities, and cities.

While disaster resilience has been studied in many fields, quantifying human mobility resilience under disasters is still unexplored. Donovan et al. [51] have studied transportation system resilience for the New York City using taxi GPS data for multiple disasters. Recent studies [40, 41, 52] have shown that under disaster events human mobility goes through perturbation but still follows the same distributions similar to the ones in a steady state, and the shift in the center of mass and radius of gyration in a perturbed state are correlated with the steady state radius of gyration. Although, these studies have suggested that human mobility is somewhat resilient to disasters, a quantitative assessment of mobility resilience is still missing in the literature. Furthermore, these studies did not explore the expected correlations of mobility resilience across different types of extreme events.

Previously, several concepts of resilience have been proposed. Hosseini et al. [5] have reviewed the methods of defining and quantifying resilience in various fields. Bruneau et al. [10] developed a framework for measuring resilience considering four dimensions: (i) *robustness* reflecting the strength or ability of the system to reduce the damage; (ii) *rapidity* representing the rate or speed of recovery; (iii) *resourcefulness* reflecting the ability to apply materials and human resources by prioritizing goals when an event occurs; and (iv)

redundancy representing the capacity to achieve goals by prioritizing objective to restrain loss and future disruptions. They have also proposed the following equation to measure resilience loss of infrastructures of a community due to an earthquake:

$$\text{RL} = \int_{t_0}^{t_1} \big[100 - Q(t) \big] \, dt, \tag{1}$$

where, RL denotes resilience loss, $Q(t)$ denotes a quality function for infrastructure service at time t and $(t_1 - t_0)$ is the recovery time. This formula forms the basis of a resilience triangle. Although this metric was originally proposed for an earthquake, it can be applied to many other contexts [5]. In addition to conceptualizing the linkage between vulnerability, resilience and adaptive capacity Cutter et al. proposed a place based model for understanding community resilience [53]. Hosseini et al. proposed a Bayesian network based framework to quantify infrastructure resilience mainly considering the absorptive, adaptive, and restorative capacity perspectives [54–57]. But this approach needs many variables, interconnected with resilience, which are difficult to collect in the context of human mobility because it involves a large geographical area.

However, measuring resilience, in a mobility context, has been difficult due to the lack of appropriate metrics over longer time periods. Geo-location data from social media can offer a solution to this problem. In this study, by analyzing user displacements from a pre-disaster period to a post-disaster one, we measure perturbation and recovery time for multiple types of disaster. To validate our results, we have used one-month of taxi data from the New York City recording taxi movements before, during, and after hurricane Sandy. Quantifying the loss of resilience and recovery time from disruptions in response to an extreme event can help understanding the broader socio-economic impacts of disasters. Furthermore, these resilience metrics will help in making policy towards building resilient cities and communities.

This paper makes several contributions. First, it defines the concept of mobility resilience and develops methods to detect extreme events in mobility data and to measure required metrics to measure resilience and transient loss of resilience from movement data. Second, it applies the proposed method of measuring resilience to geo-located data collected from Twitter for multiple disasters. Thus, this paper shows that geo-located social media data can be effectively used to measure human mobility resilience to extreme events.

2 Data and methods

To measure mobility resilience, we have used geo-tagged tweets from several types of disaster (Table 1). The data sets have been collected from Dryad digital repositories http://datadryad.org/resource/doi:10.5061/dryad.88354 [58], originally collected by Wang et al. [41] and https://datadryad.org//resource/doi:10.5061/dryad.15fv2, collected by Kryvasheyeu et al. [59].

To validate our approach of using social media data, we collected New York City taxi data which includes taxi movement for the period same as the hurricane Sandy twitter data. The data was collected from a repository hosted by New York City Taxi and Limousine Commission (http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml). In the data, each observation represents a trip and there were total 12,892,877 trips in the study period.

Table 1 Data description

Type	Disaster Name	Disaster Location	No. of Tweets	No. of Users
Hurricane	Sandy (all tweets)	USA	52,493,130	13,745,659
	Sandy (geo-tagged tweets)	USA	24,149,780	5,981,012
Earthquake	Bohol (Bohol)	Bohol, Philippines	114,606	7942
	Iquique (Iquique)	Iquique, Chile	15,297	1470
	Napa (Napa)	Napa, USA	38,019	1850
Typhoon	Wipha (Tokyo)	Tokyo, Japan	849,173	73,451
	Halong (Okinawa)	Okinawa, Japan	166,325	5,124
	Kalmaegi (Calasiao)	Calasiao, Philippines	21,698	1,063
	Rammasun (Manila)	Manila, Philippines	408,760	27,753
Winter storm	Xaver (Norfolk)	Norfolk, Britain	115,018	8498
	Xaver (Hamburg)	Hamburg, Germany	15,054	2745
	Storm (Atlanta)	Atlanta, USA	157,179	15,783
Thunder storm	Storm (Phoenix)	Phoenix, USA	579,735	23,132
	Storm (Detroit)	Detroit, USA	765,353	15,949
	Storm (Baltimore)	Baltimore, USA	328,881	14,582
Wildfire	New South Wales (1)	New South Wales, Australia (1)	64,371	9246
	New South Wales (2)	New South Wales, Australia (2)	34,157	4147

Hurricane Sandy data have tweets from several places including USA, Canada, Mexico and other countries. For measuring resilience for a city or a state in response to hurricane Sandy, we have applied appropriate location filters. For example, a trip can be made within the New York City or having only an origin or destination in it. Since displacements are calculated in six-hour periods, when calculating resilience for the New York City, if a location filter is applied, only the displacement within the New York City will be considered in a six-hour period. If a location filter is not applied, both displacements within the New York City and having origins or destinations at the New York City will be considered in a six-hour period. Except hurricane Sandy data, the rest of the data consist city-specific tweets where those cities were subject to a disruptive event. Thus, a location filter or constraint is not required for these cases.

In this study, we apply the concept of resilience for understanding human mobility under a disaster. Following the basic definition of resilience, we define *mobility resilience* as the ability of a mobility infrastructure system responsible for the movement of a population to manage shocks and return to a steady state in response to an extreme event. These events include a hurricane, earthquake, terrorist attack, winter storm, wildfire, flood, and others. We propose a simple method based on human movement data using *normalized per user displacement* as a key indicator of human mobility. Comparing the difference between per user displacements from typical displacements, the proposed method can detect a disruptive event from movement data and calculate the maximum deviation from normal conditions and the recovery time. Finally, applying the concept of resilience triangle, we estimate resilience and transient loss of resilience for an event detected by the method. The proposed method can take any kind of movement data as inputs including coordinates from mobile phone call recordings, GPS observations, social media posts and many others. In this paper, we present our resilience analysis based on social media data from multiple types of disasters.

2.1 Extracting location time series of a user

First, the coordinates of a user are sorted in an ascending order by timestamps. If there are not enough users for an hourly based analysis, we can divide each day in 4 periods such

as 12 AM to 6 AM, 6 AM to 12 PM, 12 PM to 6 PM and 6 PM to 12 AM. From the sorted time series, locations (i.e., latitude and longitude) of each user are extracted in six-hour interval for each day.

$$P_u^{d,t} = \{(x,y)|(x,y) \in (\text{latitude}, \text{longitude of a region})\}, \tag{2}$$

where, $P_u^{d,t}$ denotes the set of locations of a user u in day d at period t $d\varepsilon$ (days in the dataset), $t\varepsilon$ (periods in a day), $u\varepsilon$ (users in dataset).

2.2 Displacement metric

From the set of locations of a user, distances between two consecutive points are calculated using the Haversine formula [60] shown in Equation (3). Given a pair of points (latitude and longitude), Haversine formula calculates the great-circle distance between two points. Although the most appropriate distance will be the actual traveled distance (distance along the traveled road or air path) by a user, it is impossible to obtain this actual distance from social media data due to the lack of trajectory information. Euclidean distance is the shortest distance between any two points which is often not the case in real road or air distance. We adopted Haversine distance because it considers the curvature of the earth and for small distance almost similar to Euclidian distance. Thus, Haversine displacement is better than the Euclidian distance and the most suitable for air distance. The Haversine displacement is adopted by many previous studies [40, 41, 61, 62] related to human mobility. To the best of our knowledge, Canberra distance is not suitable for human mobility analysis because it tends to calculate the distance in a higher dimensional space.

For calculating displacements, a user must have at least two locations within a six-hour interval. Otherwise, the user is not considered in that interval.

$$C = 2r \times \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right)}\right), \tag{3}$$

where r is radius of earth, ϕ is latitude and φ is longitude. Displacement between two consecutive points will be calculated for each user at every six-hour interval.

The average of the displacements for an interval is calculated by dividing the sum of the displacements by the total number of unique users contributing to that displacements. Thus,

$$D^{d_i, t_p} = \left\{ \frac{\sum_{d_i, t_p}^{d_i, t_{p+\Delta t}} C^{d,t}}{\sum_{d_i, t_p}^{d_i, t_{p+\Delta t}} u^{d,t}} \right\}, \tag{4}$$

where D^{d_i, t_p} represents the average displacements from period t_p to $t_{p+\Delta t}$ for day d_i. The term $\sum_{d_i, t_p}^{d_i, t_{p+\Delta t}} C^{d,t}$ indicates the summation of the displacements for all users from t_p to $t_{p+\Delta t}$ for day d_i and the term $\sum_{d_i, t_p}^{d_i, t_{p+\Delta t}} u^{d,t}$ represents the total number of users contributing to these displacements within this period. Here Δt is the time interval to calculate human mobility. In this study, $\Delta t = 6$ hours is chosen considering the availability of enough users within the interval.

If a user has more than two observations at a given period, we calculate all the displacements between the consecutive points. To calculate the average displacement at a given period on a given day, we sum up all the displacements of all the users and divide it by the total number of unique users having displacements at that given period on that given day. However, we do not normalize it by the number of displacements observed for a user. During a disaster, human mobility can be affected by both the distance and the frequency of displacements. For example, let us consider, before a disaster, an individual used to make 4 displacements or trips each having a distance of 2 miles, in a given 6-hour period. And, during a disaster, the same individual makes 2 trips each having a distance of 2 miles in a 6-hour period. Now, if we normalize by the number of observations (i.e., calculate displacement per trip per user), then we will determine displacement of this user 2 miles/trip for both pre-disaster and during disaster periods, although in this case individual mobility was significantly decreased during the disaster. Thus, when calculating the average displacement, we do not normalize by the number of observations so that we can capture the effect of a disaster on both trip frequencies and distances.

2.3 Extraction of typical and actual displacements time series

The mobility dataset to be used for a resilience analysis should cover pre-disaster, disaster, and post disaster periods. Using the average displacements value in the pre-disaster period, we can make four sets of typical values for the four periods considered in a day. These four typical values are calculated separately for weekdays and weekends.

$$D_{\text{weekday}}^{t} = \left\{ D^{d,t} \text{ where}, d \in (\text{pre-disaster weekdays}) \right\}, \qquad (5)$$

$$D_{\text{weekend}}^{t} = \left\{ D^{d,t} \text{ where}, d \in (\text{pre-disaster weekend days}) \right\}, \qquad (6)$$

where D_{weekday}^{t} represents the set of displacements at period t considering only weekdays in the pre-disaster period. Similarly, D_{weekend}^{t} represents the set of displacements at period t considering only weekends in the pre-disaster period. For instance, if we have 4 periods per day, and if we select first 7 days as a pre-disaster period, for each period, we have a set of 5 values of displacement for weekdays and a set of 2 values for weekends. The mean and standard deviation of these sets of displacement are used to compare the actual displacement at the corresponding periods of a day to check whether the displacement is typical or not. To capture this effect, we can compute standardized displacement, Z score, for each actual displacement using the equation given below:

$$Z^{d,t} = \left\{ \begin{array}{c} \frac{D^{d,t} \text{-mean of } D_{\text{weekday}}^{t}}{\text{standard deviation of } D_{\text{weekday}}^{t}} \text{ if } d \in (\text{week days}) \\ \text{else } \frac{D^{d,t} \text{-mean of } D_{\text{weekend}}^{t}}{\text{standard deviation of } D_{\text{weekend}}^{t}} \end{array} \right\}, \qquad (7)$$

where $Z^{d,t}$ represents the Z score at day d and period t. If d is a weekday, typical displacements for weekdays are used to compare; and if d is a weekend day, typical displacements for weekends are used.

2.4 Extreme event detection

An extreme event can disrupt human mobility by either increasing mobility or decreasing mobility. We consider two parameters for detecting an extreme event: a threshold z

score (α) and the number of time intervals (τ). The first parameter checks the amount of deviation from typical values and the second parameter checks how long this deviation persists.

$$\text{Event}_{d_i,t_p}^{d_j,t_q} = \left\{ Z^{d,t} : Z^{d,t} \leq \alpha_l \text{ and } \sum_{d_i,t_p}^{d_j,t_q} d,t \geq \tau \right\} \tag{8}$$

or,

$$\text{Event}_{d_i,t_p}^{d_j,t_q} = \left\{ Z^{d,t} : Z^{d,t} \geq \alpha_u \text{ and } \sum_{d_i,t_p}^{d_j,t_q} d,t \geq \tau \right\}. \tag{9}$$

Equation (8) and (9) represent the event detection for decreased and increased mobility, respectively; where $\text{Event}_{d_i,t_p}^{d_j,t_q}$ represents an extreme event from day d_i period t_p to day d_j period t_q; $d_i, d_j \in$ (days in data set) and $t_p, t_q \in$ (periods in a day); α_l, α_u represent the lower and upper threshold of Z score; and τ represents the threshold number of periods when Z score is above or below the threshold Z score. These parameters (α, τ) can be selected to identify shorter or longer extreme events depending on the type of a disaster and the area affected by it.
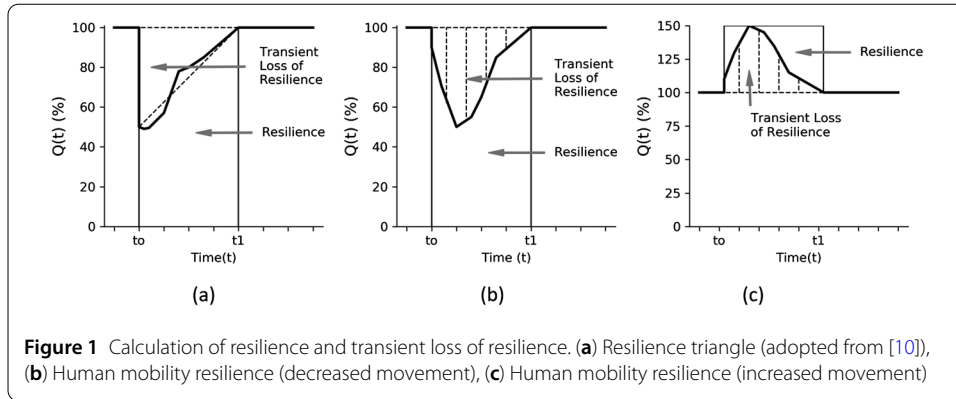
We recommend selecting the threshold values based on a decision maker's need. For instance, a small threshold on event duration (τ) can capture mobility resilience due to events (such as rainfall, thunderstorm, special events etc.) that last short periods. In contrast, a longer duration threshold is recommended if we want to calculate resilience only for events that last longer period such as hurricane, typhoon etc.

On the other hand, the threshold on z score captures the amount of deviation occurred due to an event. The lower threshold (α_l) and upper threshold (α_u) values are used for capturing the events due to decreased and increased mobility, respectively. Again, selection of the thresholds depends on the decision maker's need. For example, if we want to capture only the events that make a huge deviation from normal condition, a very small α_l and very big α_u should be chosen.

A threshold is meant to separate the condition when the level of human mobility deviates significantly ($< \alpha_l$ or $> \alpha_u$) from normal condition for a significant amount of time ($\geq \tau$). In our study, we selected the threshold values that best capture the actual timing of the events (landfall time, earthquake strike time etc.). Another approach could be to fit a distribution of the predisaster data and choose a threshold which is significantly different from the normal condition at 90% or 95% confidence level. But in our case, there are not enough data for fitting a distribution for most of the cases. We consider variability between weekdays and weekends only.

2.5 Resilience calculation

Once an extreme event has been detected, maximum deviation and recovery time can be easily calculated. We define human mobility resilience as the ability of a mobility infrastructure system responsible for the movement of the population of a community to manage shocks and return to a steady state in response to an extreme event. Bruneau et al. [10] introduced an equation for calculating resilience loss as shown in Equation (1). As applied to infrastructures of a community, Bruneau's equation [10] computes the loss of

Figure 1 Calculation of resilience and transient loss of resilience. (**a**) Resilience triangle (adopted from [10]), (**b**) Human mobility resilience (decreased movement), (**c**) Human mobility resilience (increased movement)

resilience by the size of degradation of the expected quality of an infrastructure over time. But community resilience as a whole should be calculated with respect to all the extreme events possible. When applied to people and its environment, Norris et al. [6] used the term resilience as a metaphor where a transient dysfunction occurs during a crisis due to the degradation of quality of life. A resilient community can adapt to the situation after the event while a vulnerable community goes through a persistent dysfunction [6]. Here, using Bruneau's approach, we calculate the loss of resilience which is equivalent to the size of dysfunction/degradation of human mobility. But mobility resilience represents a long-term property of a community in response to all the possible crisis events. Since we determine the loss of resilience in response to a single crisis only, we call the size of the degradation as the transient loss of resilience, defined as:

$$\text{Transient loss of resilience, TLR} = \int_{t_0}^{t_1} \big[100 - Q(t)\big]\, dt,$$

where, TLR is the transient loss of resilience which is the area (see Fig. 1(a)) between the horizontal line from 100 and the quality curve Q(t) from t_0 to t_1 which is the recovery period for any event.

A schematic representation of this equation (see Fig. 1(a)) is known as a resilience triangle. From this triangle, the transient loss of resilience in any extreme event can be calculated as the area formed by the dashed lines and the vertical line (see Fig. 1(a)). Inspired from the resilience triangle, we represent the resilience by dividing this area into smaller trapezoids (see Fig. 1(b) and 1(c)) having height equal to the increment of time (six hours) considered in the analysis. This assumption is required since, unlike an idealized quality function, a real-world quality function indicating human mobility gradually drops from and improves to its typical values. Thus, assuming smaller trapezoids will minimize the loss in calculation.

In our analysis, we assume human mobility level as a proxy of the quality of the mobility infrastructure system. Thus, we define $Q(t)$ as the ratio of average actual displacements to average typical displacements of a population at a time period t. If an actual displacement is equal to a typical displacement, the value of the quality function is 100 or the ratio is 1. In our case, $Q(t)$ at a period t represents how much different the level of human mobility is compared to a typical value of the level of human mobility at the same period before the disaster. We obtain the recovery time (t_0 to t_1) (see Fig. 1) from the extreme event detection phase. Here, t_0 is the starting point of the detected extreme event and t_1 is the end point of

the detected event. The duration between t_0 and t_1 is the recovery time. The summation of the areas of all the small trapezoids is the transient loss of resilience (indicated by transient loss of resilience in Fig. 1(b) and 1(c)). The residual area (indicated by resilience in Fig. 1) represents the value of resilience during the recovery period. For increased mobility area considered in resilience calculation are defined by the maximum quality percentage/ratio (see Fig. 1(c)).

The selection of the thresholds on α and τ have effects on event detection and resilience calculation. For a given α value, the threshold on τ determines whether a deviated state of human mobility will be considered as a disruptive event or not. But once an event is detected, τ has no effect on the calculated resilience value. The selection of α threshold will directly affect the duration of an event which affects the resilience calculation. For instance, bigger α_u and smaller α_l will reduce the duration of an event and hence calculated resilience loss will be less. Although the resilience calculation depends on the selection of the thresholds, the ranking of events with respect to resilience will not change given that the same threshold value is chosen for all the events.

When interpreting the resilience and transient loss of resilience values, we should consider certain aspects of the resilience metric. At a fundamental level, the proposed resilience metric measures the impact of a disruption to the mobility of a population and its infrastructures. The greater the impact of a disruption has on human mobility, the greater the transient loss of resilience will be. We consider both the increase and decrease of mobility level similarly, through calculating transient loss of resilience, since both situations indicate impacts to the typical level of population mobility. However, loss of resilience calculated from decreased mobility should not be compared with the same due to increased mobility. For increased mobility, our resilience calculation is limited as we have an unbounded scenario (see Fig. 1(c)). Furthermore, an increased mobility does not necessarily indicate a better performance of the mobility infrastructure system. It is more likely that in such a situation the infrastructure system has collapsed forcing people to displace further.

3 Results

The approach to calculate resilience has been applied over location-based social datasets (see Table 1). During these events, we observe two types of responses in the mobility function which either significantly drops (decreased mobility) or significantly rises (increased mobility). To represent both types of events, two thresholds z scores (α values) have been used for detecting an extreme event. For decreased mobility cases, a threshold z score value of 40 percentile ($\alpha_l = 40$) and for increased mobility cases, a threshold z score of 90 percentile ($\alpha_u = 90$) have been chosen to detect an extreme event. However, when no event was detected with these thresholds, $\alpha_l = 60$ percentile have been chosen; this relaxes the lower threshold of z score. As the threshold duration of the extreme event when the z value is below α_l has been chosen as 7 time periods (i.e., $\tau = 7$ or 42 hours) and when the z value is above α_u has been chosen as 3 time periods (i.e., $\tau = 3$ or 18 hours).

Figure 2 shows the major steps in calculating resilience for three types of disasters namely: Hurricane Sandy (Fig. 2(a)), earthquake at Bohol (Fig. 2(b)) and a thunder storm at Phoenix, Arizona (Fig. 2(c)). Table 2 presents the results of resilience calculation for multiple types of disasters along with the threshold values used to detect the events. Events detected by 60 percentile thresholds are not comparable with the events detected by 40 percentile thresholds. The 40 percentile events are more severe than the 60 percentile

Figure 2 Resilience and transient loss of resilience for multiple disasters. Each figure has three panels; the first panel shows the actual and typical values; the second panel shows the event detection by z score; and the third panel shows the resilience and transient loss of resilience. Note: DPU = Displacements Per User (Kilometer), TF = Trip Frequency

events. Among 40 percentile events, the highest recovery time was found 144 hours for hurricane Sandy for the state of New York and the highest transient loss of resilience was found 344.89 for earthquake Iquique. We have also calculated the ratio between transient loss of resilience and resilience ($\frac{\text{TLR}}{R}$). The highest ratio of resilience loss over resilience has been found as 2.73 for the state of New York for hurricane Sandy. Among the 60 percentile events, the state of New Jersey during hurricane Sandy had the highest recovery time, transient resilience loss and transient loss of resilience over resilience ratio. These metrics indicate the magnitude of the impact of hurricane Sandy on the mobility systems of the sates of New York and New Jersey.

In addition to Twitter data, we have used taxi trips data to calculate the resilience metrics. Figure 2(d) shows the resilience and recovery time for taxi movements in the New

Table 2 Comparison of resilience, transient loss of resilience and recovery time for multiple types of events occurred in different location

Disaster Name	Location	Threshold Z Score	Location Filter	Start Time	Recovery Time (hr.)	Max Deviation	Resilience (R)	Transient Loss of Resilience (TLR)	Ratio (TLR/R)
Sandy	New York City	$\alpha_l = 60$	Y	2012-10-26 00:00	132	0.540	106.730	19.260	0.181
		$\alpha_l = 40$	N	2012-10-28 12:00	66	0.010	17.620	42.370	2.404
	New York State	$\alpha_l = 40$	Y	2012-10-28 12:00	48	0.260	21.860	20.100	0.920
		$\alpha_l = 40$	N	2012-10-28 06:00	144	0.087	36.400	101.400	2.730
	New Jersey State	$\alpha_l = 60$	Y	2012-10-28 00:00	120	0.176	52.000	55.00	1.057
		$\alpha_l = 60$	N	2012-10-27 12:00	168	0.001	21.179	140.820	6.648
				2012-11-06 00:00	48	0.018	8.907	33.090	3.715
	Pennsylvania State	$\alpha_l = 60$	Y	2012-10-28 00:00	120	0.180	58.930	49.060	0.833
		$\alpha_l = 60$	N	2012-10-26 06:00	144	0.003	12.600	125.390	9.949
				2012-11-02 06:00	72	0.015	13.970	52.026	3.720
Earthquake	Bohol, Philippines	$\alpha_u = 90$	NA	2013-10-15 00:00	54	9.330	120.470	162.310	1.340
				2013-10-19 18:00	24	11.035	64.956	115.680	1.780
	Iquique, Chile	$\alpha_u = 90$	NA	2014-04-02 18:00	48	38.167	519.058	344.890	0.664
	Napa, USA	$\alpha_u = 90$	NA	2014-08-23 18:00	18	6.416	27.490	37.503	1.360
Wild Fire	NSW1, Australia	$\alpha_u = 90$	NA	2013-10-18 12:00	18	8.257	58.860	28.230	0.480
		$\alpha_l = 40$	NA	2013-10-19 06:00	48	0.188	22.440	19.550	0.870
	NSW2, Australia	$\alpha_l = 40, 60$	NA	NO TLR					
Winter Storm	Xaver, Norfolk, Britain	$\alpha_l = 40$	NA	2013-12-02 12:00	48	0.339	25.370	16.629	0.655
	Xaver, Hamburg, Germani	$\alpha_l = 40$	NA	2013-12-04 18:00	48	0.035	24.817	17.182	0.690
		$\alpha_u = 90$	NA	2013-12-13 12:00	36	4.306	55.704	43.480	0.780
	Atlanta, USA	$\alpha_l = 40$	NA	2014-01-28 12:00	54	0.261	20.450	27.545	1.346
Rain Storm	Phoenix, USA	$\alpha_l = 40$	NA	2014-09-06 18:00	60	0.329	40.000	13.000	0.413
	Detroit, USA	$\alpha_l = 40$	NA	Not Enough Pre-Disaster Data					
	Baltimore, USA	$\alpha_l = 40, 60$	NA	NO TLR					
Typhoon	Wipha, Tokyo, Japan	$\alpha_l = 40, 60$	NA	NO TLR					
	Halong, Okinawa, Japan	$\alpha_l = 40$	NA	2014-07-29 06:00	96	0.616	74.000	10.000	0.135
	Kalmaegi, Philippines	$\alpha_l = 40$	NA	2014-09-08 12:00	96	0.005	42.568	42.000	0.990
		$\alpha_l = 40$		2014-09-23 12:00	54	0.003	24.188	23.811	0.980
	Rammasun, Philippines	$\alpha_l = 40, 60$	NA	NO TLR					

Note: NA = Not Applicable, Y = Yes, N = No.

York City. For measuring resilience in taxi data, taxi trips have been used instead of the taxi trip distance. Most of the trips in taxi occurred between some frequently visited places and thus, the average traveled distances per trip were almost same for the disrupted days although there were significantly less number trips in those days. For taxi trips, the maximum deviation at the landfall day is found as 0.052 which means only 5.2 percent of the typical trips occurred at the landfall day of hurricane Sandy; the recovery time is found 96 hours. A recent study [51] measuring transportation system resilience by taxi data using pace as a quality indicator found recovery time as 132 hours for hurricane Sandy. From Table 2, we can see that human mobility recovery time and transient loss of resilience for New York city is 66 hours and 42.37, respectively. The two results between taxi resilience and human mobility resilience is not directly comparable because taxi is just one of the modes of human mobility.

During hurricane Sandy, among the states, the state of New York suffered the highest transient loss of resilience followed by the states of New Jersey and Pennsylvania. For hurricane Sandy both recovery time and transient loss of resilience are higher when a location constraint is not applied. Except hurricane Sandy data, typhoon, winter storm and rain storm data are location constrained. Thus, transient losses of resilience for these events are lower compared to hurricane Sandy's unconstrained transient loss of resilience. This finding is consistent with previous findings [52] that during these types of disasters, short trips are less affected compared to long trips. These events discussed above faced a significant amount of decrease in mobility from a typical mobility function.

However, in an earthquake, instead of a decreasing mobility function, we observe a significant increase in human mobility, probably due to the long-distance migration of people forced by severe infrastructure damages. Figure 2(b) shows the resilience calculation for an earthquake happened at Bohol, Philippines in 2013. The recovery time and transient loss of resilience for this event are 54 hours and 162.31, respectively. Our method has detected one more event after around 3 days. This event may represent the increased mobility when displaced people returned to their places as studies found that natural disaster like earthquake cause human migration. Table 2 shows the other earthquake resilience and recovery time results. Among the earthquakes analyzed in this study, Iquique had the highest deviation and transient loss of resilience, 38.167 and 344.89, respectively and Napa had the lowest transient loss of resilience and deviation. A study [41] on the same data for measuring human mobility pattern found that although human mobility during most of typhoon, rainstorms, winter storms and Napa earthquake can be predicted by established patterns, mobility during earthquakes Bohol and Iquique cannot be predicted. Instead of decreased mobility, a significant increase in mobility with large transient loss of resilience during these events may explain this result.

4 Discussions

In this paper, we present a method to compute resilience metrics using geo-location data from social media. The proposed method can detect an extreme event from human movements, measure the recovery time and the maximum deviation from a steady state mobility indicator, and assess the values of resilience and transient loss of resilience. Applying this method on multiple disaster data, we find that human movements within a geographic area (e.g., trips only within a city) is less affected compared to all the movements associated with the area (e.g., trips from, to, and within the city). Disasters such as hurricane, typhoon,

winter storm decrease human mobility and the amount of perturbation depends on the location and severity of the disaster. However, an earthquake increases human mobility causing a significant transient loss of resilience. This is probably because an earthquake is unpredictable while for the other disasters people had warnings lasting over multiple days.

The findings of this study are very important for understanding the nature and amount of perturbation and the subsequent transient loss of resilience in human mobility due to a disaster. Thus, it will help understanding the higher-order impacts of a disruptive event in human society and national economy. It can also help in policy making, as resilience assessment is critical for building a resilient transportation system.

However, there are some limitations in the metric used here. We do not have any measurement of at what levels the infrastructure should be performing before a disruption and after the recovery efforts. Therefore, we assume the pre-disruption mobility level as a proxy of infrastructure quality and expect that after recovery population mobility should reach to the pre-disruption level. We also assume that the pre-disruption mobility level is the best possible condition (100%). This may not be true as a community may not have access to proper mobility infrastructures even before a disaster. Furthermore, after the recovery activities, mobility level may not return to its optimal condition. The proposed metric cannot detect events less than six hours long because a minimum period of six hours is chosen. Also, in a pre-disaster period, variations among weekdays and variations between weekend days are not considered due to the lack of enough pre-disaster data. Movements of social media users may not represent well the actual population movement during a disaster.

Quantifying mobility resilience is difficult due to its complex interactions with many interconnected systems. We choose a simple metric from [10] to determine transient resilience loss in mobility due to an extreme event so that the approach can be applied to different types of disasters without considering many dimensions. This study is one of the first empirical studies to quantify mobility resilience from mobility data. Availability of comprehensive infrastructure and mobility data will lead to a more robust and complete resilience metric.

Abbreviations
R, Resilience; TLR, Transient Loss of Resilience; DPU, Displacements Per User (Kilometer); TF, Trip Frequency.

Availability of data and materials
The datasets supporting the conclusions of this article are available in the Dryad digital repositories http://datadryad.org/resource/doi:10.5061/dryad.88354 [58], originally collected by Wang et al. [41] and https://datadryad.org//resource/doi:10.5061/dryad.15fv2, collected by Kryvasheyeu et al. [59].

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Study conception and design: SH KCR; Acquisition of data: MC; Analysis and interpretation of data and results: KCR SH; Drafting of Manuscript: KCR; Revision and editing: SH MC; Funding acquisition: SH; Supervision: SH. All authors read and approved the final manuscript.

Author details
^{1}Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, USA. ^{2}Media Laboratory, Massachusetts Institute of Technology, Cambridge, USA.

Publisher's Note

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

References

1. Huppert HE, Sparks RSJ (2006) Extreme natural hazards: population growth, globalization and environmental change. Philos Trans R Soc A, Math Phys Eng Sci 364:1875–1888
2. Hasan S, Foliente G (2015) Modeling infrastructure system interdependencies and socioeconomic impacts of failure in extreme events: emerging R&D challenges. Nat Hazards 78:2143–2168
3. The World Bank (2016) Natural Disasters Force 26 Million People into Poverty and Cost $520bn in Losses Every Year, New World Bank Analysis Finds. http://www.worldbank.org/en/news/press-release/2016/11/14/natural-disasters-force-26-million-people-into-poverty-and-cost-520bn-in-losses-every-year-new-world-bank-analysis-finds. Accessed 15 Jun 2017
4. Cutter SL, Ahearn JA, Amadei B et al (2013) Disaster resilience: a national imperative. Environ Sci Policy Sustain Dev 55:25–29
5. Hosseini S, Barker K, Ramirez-Marquez JE (2016) A review of definitions and measures of system resilience. Reliab Eng Syst Saf 145:47–61
6. Norris FH, Stevens SP, Pfefferbaum B et al (2008) Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. Am J Community Psychol 41:127–150
7. Alliance R (2007) Assessing resilience in social-ecological systems—a workbook for scientists. Transformation 22:1–53
8. Speranza CI, Wiesmann U, Rist W (2014) An indicator framework for assessing livelihood resilience in the context of social-ecological dynamics. Glob Environ Chang 28:109–119
9. Kahan JH, Allen AC, George JK (2009) An operational framework for resilience. J Homel Secur Emerg Manag 61
10. Bruneau M, Chang SE, Eguchi RT et al (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. Earthq Spectra 19:733–752
11. McCallum I, Liu W, See L et al (2016) Technologies to support community flood disaster risk reduction. Int J Disaster Risk Sci 7:198–204
12. Nicholson CD, Barker K, Ramirez-Marquez JE Vulnerability analysis for resilience-based network preparedness. Manuscr Revis
13. Pappalardo L, Pedreschi D, Smoreda Z, Giannotti F (2015) Using big data to study the link between human mobility and socio-economic development. 2015 IEEE Int Conf Big Data (Big Data), 871–878
14. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. 453
15. Alessandretti L, Sapiezynski P, Lehmann S, Baronchelli A (2017) Multi-scale spatio-temporal analysis of human mobility. PLoS ONE 122:e0171686
16. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462–465
17. Jurdak R, Zhao K, Liu J et al (2015) Understanding human mobility from Twitter. PLoS ONE 10:1–16
18. Yao CZ, Lin JN (2016) A study of human mobility behavior dynamics: a perspective of a single vehicle with taxi. Transp Res, Part A, Policy Pract 87:51–58
19. Wang W, Pan L, Yuan N et al (2015) A comparative analysis of intra-city human mobility by taxi. Phys A, Stat Mech Appl 420:134–147
20. Tang J, Liu F, Wang Y, Wang H (2015) Uncovering urban human mobility from large scale taxi GPS data. Phys A, Stat Mech Appl 438:140–153
21. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. Nat Phys 6:818–823
22. Deville P, Song C, Eagle N et al (2016) Scaling identity connects human mobility and social interactions. Proc Natl Acad Sci 113:7047–7052
23. Hasan S, Zhan X, Ukkusuri SV (2013) Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. Proc 2nd ACM SIGKDD Int Work Urban Comput—UrbComp '13
24. Rashidi TH, Abbasi A, Maghrebi M et al (2017) Exploring the capacity of social media data for modelling travel behaviour: opportunities and challenges. Transp Res, Part C, Emerg Technol 75:197–211
25. Hasan S, Ukkusuri SV (2014) Urban activity pattern classification using topic models from online geo-location data. Transp Res, Part C, Emerg Technol 44:363–381
26. Roy KC, Hasan S (2018) Quantifying human mobility resilience to extreme events using geo-located social media data. In: Proceedings of transportation research board 97th annual meeting
27. Roy KC (2018) Understanding crisis communication and mobility resilience during disasters from social media. M.Sc. Thesis, University of Central Florida, https://stars.library.ucf.edu/etd/6200/
28. Zhao K, Musolesi M, Hui P et al (2015) Explaining the power-law distribution of human mobility through transportation modality decomposition. Sci Rep 5:9136
29. Noulas A, Scellato S, Lambiotte R et al (2012) A tale of many cities: universal patterns in human urban mobility. PLoS ONE 7
30. Beiró MG, Panisson A, Tizzoni M, Cattuto C (2016) Predicting human mobility through the assimilation of social media traces into mobility models. EPJ Data Sci 5
31. Vaca C, Aiello LM, Jaimes A, Milano P (2014) Modeling dynamics of attention in social media with user efficiency. EPJ Data Sci 31(5)
32. Han XP, Hao Q, Wang BH, Zhou T (2011) Origin of the scaling law in human mobility: hierarchy of traffic systems. Phys Rev E, Stat Nonlinear Soft Matter Phys 83:2

33. Hawelka B, Sitko I, Beinat E et al (2014) Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci 41:260–271
34. Liu Y, Sui Z, Kang C, Gao Y (2014) Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PLoS ONE 9
35. Gallotti R, Bazzani A, Rambaldi S (2016) Towards a statistical physics of human mobility. Int J Mod Phys C 23(09):125
36. Wu L, Zhi Y, Sui Z, Liu Y (2014) Intra-urban human mobility and activity transition: evidence from social media check-in data. PLoS ONE 9
37. Liang X, Zheng X, Lv W et al (2012) The scaling of human mobility by taxis is exponential. Phys A, Stat Mech Appl 391:2135–2144
38. Liu H, Chen YH, Lih JS (2015) Crossover from exponential to power-law scaling for human mobility pattern in urban, suburban and rural areas. Eur Phys J B 88:1–7
39. Zhao K, Chinnasamy MP, Tarkoma S (2015) Automatic city region analysis for urban routing. 2015 IEEE Int Conf Data Min Work 1136–1142
40. Wang Q, Taylor JE (2014) Quantifying human mobility perturbation and resilience in hurricane sandy. PLoS ONE 9:1–5
41. Wang Q, Taylor JE (2016) Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. PLoS ONE 11:1–14
42. Mesa-Arango R, Hasan S, Ukkusuri SV et al (2013) Household-level model for hurricane evacuation destination type choice using hurricane ivan data. Natural Hazards Review 14:11–20
43. Hasan S, Ukkusuri S, Gladwin H, Murray-Tuite P (2011) Behavioral model to understand household-level hurricane evacuation decision making. J Transp Eng 137:341–348
44. Hasan S, Mesa-Arango R, Ukkusuri S (2013) A random-parameter hazard-based model to understand household evacuation timing behavior. Transp Res, Part C, Emerg Technol 27:108–116
45. Sadri AM, Ukkusuri SV, Murray-Tuite P, Gladwin H (2014) Analysis of hurricane evacuee mode choice behavior. Transp Res, Part C, Emerg Technol 48:37–46
46. Sadri AM, Ukkusuri SV, Murray-Tuite P, Gladwin H (2015) Hurricane evacuation routing strategy from Miami beach: choice of major bridges. Transp Res Rec, 1–24
47. Sadri AM, Ukkusuri SV, Murray-Tuite P (2013) A random parameter ordered probit model to understand the mobilization time during hurricane evacuation. Transp Res, Part C, Emerg Technol 32:21–30
48. Roy KC, Hasan S (2019) Modeling the dynamics of hurricane evacuation decisions from real-time Twitter data. In: Proceedings of transportation research board 98th annual meeting
49. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. Proc Natl Acad Sci 109:11576–11581
50. Guan X, Chen C, Work D (2016) Tracking the evolution of infrastructure systems and mass responses using publically available data. PLoS ONE 11:e0167267
51. Donovan B, Work DB (2017) Empirically quantifying city-scale transportation system resilience to extreme events. Transp Res, Part C, Emerg Technol 79:333–346
52. Wang Q, Taylor JE (2014) Quantifying, comparing human mobility perturbation during Hurricane Sandy, Typhoon Wipha, Typhoon Haiyan. Procedia Econ Financ 18:33–38
53. Cutter SL, Barnes L, Berry M et al (2008) A place-based model for understanding community resilience to natural disasters. Glob Environ Chang 18:598–606
54. Hosseini S, Barker K (2016) Modeling infrastructure resilience using Bayesian networks: a case study of inland waterway ports. Comput Ind Eng 93:252–266
55. Hosseini S, Barker K (2016) A Bayesian network model for resilience-based supplier selection. Int J Prod Econ 180:68–87
56. Hosseini S, Al Khaled A, Sarder MD (2016) A general framework for assessing system resilience using Bayesian networks: a case study of sulfuric acid manufacturer. J Manuf Syst 41:211–227
57. Hosseini S (2016) Modeling and measuring resilience: applications in supplier selection and critical infrastructure
58. Wang Q, Taylor JE (2016) Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. In: Dryad Digit. Repos. http://datadryad.org/resource/doi:10.5061/dryad.88354
59. Kryvasheyeu Y, Chen H Performance of social network sensors during Hurricane Sandy. PLoS ONE 102 e0117288
60. Robusto CC (1957) The cosine-haversine formula. Am Math Mon 64:38–40
61. Gong Y, Deng F, Sinnott RO (2015) Identification of (near) real-time traffic congestion in the cities of Australia through Twitter. In: Proc ACM first int work underst city with urban informatics—UCUI '15, pp 7–12. https://doi.org/10.1145/2811271.2811276
62. Laylavi F, Rajabifard A, Kalantari M (2016) A multi-element approach to location inference of Twitter: a case for emergency response. ISPRS Int J Geo-Inf 5:56