Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after Hurricane Maria

Takahiro Yabe[1], Satish V. Ukkusuri[1]* and P. Suresh C. Rao[1,2]

*Correspondence:
sukkusur@purdue.edu
[1]Lyles School of Civil Engineering,
Purdue University, 550 Stadium Mall
Drive, West Lafayette, IN, USA
Full list of author information is
available at the end of the article

Abstract

Recent disasters have shown the existence of large variance in recovery trajectories across cities that have experienced similar damage levels. Case studies of such events reveal the high complexity of the recovery process of cities, where inter-city dependencies and intra-city coupling of social and physical systems may affect the outcomes in unforeseen ways. Despite the large implications of understanding the recovery processes of cities after disasters for many domains including critical services, disaster management, and public health, little work have been performed to unravel this complexity. Rather, works are limited to analyzing and modeling cities as independent entities, and have largely neglected the effect that inter-city connectivity may have on the recovery of each city. Large scale mobility data (e.g. mobile phone data, social media data) have enabled us to observe human mobility patterns within and across cities with high spatial and temporal granularity. In this paper, we investigate how inter-city dependencies in both physical as well as social forms contribute to the recovery performances of cities after disasters, through a case study of the population recovery patterns of 78 Puerto Rican counties after Hurricane Maria using mobile phone location data. Various network metrics are used to quantify the types of inter-city dependencies that play an important role for effective post-disaster recovery. We find that inter-city *social* connectivity, which is measured by pre-disaster mobility patterns, is crucial for quicker recovery after Hurricane Maria. More specifically, counties that had more influx and outflux of people prior to the hurricane, were able to recover faster. Our findings highlight the importance of fostering the social connectivity between cities to prepare effectively for future disasters. This paper introduces a new perspective in the community resilience literature, where we take into account the inter-city dependencies in the recovery process rather than analyzing each community as independent entities.

Keywords: Disaster recovery, Human mobility, Mobile phone data, Inter-city networks, Social networks

Introduction

Following the recent series of natural hazards with unprecedented severity and magnitude, including Hurricanes Harvey, Irma and Maria, the concept of urban resilience has gained significant global attention (Kull et al. 2016; Gitay et al. 2013). For many cities, it is of utmost importance to minimize economic loss and maintain the wellbeing of their citizens (Eakin et al. 2017; Kousky 2014). Recent disasters have shown the existence

of large variance and disparities in recovery trajectories across communities that have experienced similar damage levels (Finch et al. 2010; Aldrich 2012). We have witnessed manifold cases where cities experience significant loss of population even after sufficient recovery of infrastructure systems (Myers et al. 2008; McCaughey et al. 2018). These case studies reveal the high complexity of the recovery of cities, where both the inter-city dependencies and intra-city coupling of social and physical systems may affect the recovery outcomes in an unforeseen and non-linear manner (Klammler et al. 2018; Aerts et al. 2018).

Post disaster population displacement and recovery patterns on the city scale have long been studied based on surveys and census data (McCaughey et al. 2018; Cutter et al. 2008; Gray and Mueller 2012; Finch et al. 2010; Fussell et al. 2014; DeWaard et al. 2016; Sadri et al. 2018; Piguet et al. 2011). Such studies reveal factors that are associated with displacement and recovery; however, they often neglect the effect of interdependencies between cities. With the increase in the availability of large mobility datasets including mobile phone call detail records, Global Positioning System (GPS) logs, and social media posts, it has become possible to observe individual mobility patterns at an unprecedented high spatio-temporal granularity (Deville et al. 2014; Jiang et al. 2016; Wardrop et al. 2018; Gonzalez et al. 2008; Schneider et al. 2013). In the context of disasters, several studies have focused on analyzing human mobility during and after earthquakes (Lu et al. 2012; Song et al. 2014; Yabe et al. 2016), hurricanes (Wang and Taylor 2014; 2016) and other anomalous events (Bagrow et al. 2011). Most recently, cross comparative analysis of population recovery using large scale mobile phone GPS datasets from multiple disasters have unraveled general patterns of population recovery (Yabe et al. 2019). Moreover, the heterogeneity in initial and long-term displacement rates across various communities were explained by a set of key universal factors including the community's median income level, population size, housing damage rate, and the connectivity to other regions. Despite such progress, the importance of inter-city dependencies on recovery is under-studied in the literature. The aforementioned work (Yabe et al. 2019) only focused on the proximity of regions to other regions on the road network. The proximity of region i to other regions was defined as $d_p(i) = \frac{\sum_{j \in S(i)} N_j}{N_i}$, where N_i denotes the number of households in region i, and $S(i)$ denotes the set of regions that can be reached within 1 hour by vehicles from region i. This metric fails to capture the social dimension of the connectivity between cities.

In this paper, we investigate how inter-city dependencies in both physical as well as social forms contribute to the recovery of cities after disasters. We investigate this problem through a case study of the population recovery patterns of 78 Puerto Rican counties after Hurricane Maria. We analyze mobile phone location datasets, which include the GPS location data of more than 50,000 unique users from over 6 months before and after the Hurricane. Various metrics from the spatial networks literature are used to quantify the node-level characteristics of Puerto Rican counties on social and physical inter-city networks to understand the types of inter-city dependencies that play an important role for effective post-disaster recovery. We find that inter-city *social* connectivity, which is measured using pre-disaster mobility patterns, is crucial for quicker recovery after Hurricane Maria. More specifically, counties that had more influx and outflux of people prior to the hurricane were able to recover faster. Our findings highlight the importance of fostering the social connectivity between cities as well as strengthening the

physical infrastructure, to prepare effectively for future disasters. This paper introduces a new perspective in the community resilience literature, where we take into account the inter-city dependencies in the recovery process rather than analyzing each community as independent entities.

The remainder of the paper is organized as follows. First, we describe the datasets (mobile phone data, socio-economic data, and disaster damage data) used in this study. Next, we introduce our methods for 1) estimating the time until recovery using mobile phone data, 2) computing the node level network statistics that reflect the physical and social connectivity, and 3) spatial regression analysis. Then, we present the empirical results using data collected from Puerto Rico after Hurricane Maria in 2017. Finally, the results are discussed and the study is concluded in the final sections.

Data

Mobile phone data

Mobile phone location data were collected by Safegraph Inc.[1] during Hurricane Maria. GPS data were obtained from mobile phones of individuals who agreed to provide their location data for research purposes, and all information were anonymized to protect the identity and security of users. Each observation includes the spatial and temporal information (time, longitude, latitude) of mobile phones as well as a hashed unique identifier (ID), as shown in Table 1. The example ID is masked to protect privacy issues.

Figure 1 shows the probability density of the number of observations per day for a user in the location dataset. As shown in panel **A)**, the probability density of the number of GPS logs per user follows a power law distribution. The average number of logs per user was 83 and the standard deviation was 262. For our analysis of population recovery on the county scale, it is important to account for the biases in population samples. To ensure that the data are not spatially biased, we plot the number of samples and sample rate (= observed IDs / population) proportional to census population in Fig. 1 for Puerto Rico. We can see that the number of mobile phone samples are highly correlated with census population for each county, with Pearson's correlation coefficient of 0.985 Fig. 1b. Moreover, Fig. 1c shows that the sample rates have low correlations with census population. Thus, it is shown that the mobile phone data have little to no spatial bias in sample distributions.

However, after the Hurricane, the spatial bias in mobile phone user samples increases due to various reasons, including damages to mobile phone towers cutting off mobile phone users from the mobile network, and power outages causing people to not be able to recharge their phones. As explained more in depth in the "Methods" section, we overcome this issue by using data from only the temporal periods where the spatial bias of mobile phone user sampling rate was minimal.

Socio-demographic data

In this study, population and income data of each county were used for regression analysis, in accordance with the findings from previous studies (Yabe et al. 2019). Population data were obtained from the US National Census[2], and median income data were obtained from the American Community Survey[3].

---

[1]https://www.safegraph.com/
[2]https://www.census.gov/
[3]https://www.census.gov/programs-surveys/acs

Table 1 Fields and content of mobile phone location data

Field	Description	Sample
User ID	Hashed unique user ID for each mobile phone user	###HA9K8FH6R
Timestamp	Unix time of observation	1505771147
Longitude	Longitude coordinate of the GPS observation	-66.4304
Latitude	Latitude coordinate of the GPS observation	18.2041

Figure 2a and b show spatial plots of the population and income distribution of counties in Puerto Rico, respectively. We can observe that in Puerto Rico, the majority of the population is concentrated in the metro areas including San Juan County. The income disparity is less significant compared to the population disparity.

Disaster damage data

The tropical cyclone Hurricane Maria developed on September 16, 2017 in the Atlantic Ocean to the northeast of South America, and made landfall in southeastern Puerto Rico on September 20, 2017 with wind speeds of 155 miles per hour. Damage to Puerto Rico was severe and widespread following the Hurricane, with heavy rainfall, flooding, storm surge, and high winds causing considerable damage (Pasch et al. 2018). Various infrastructure systems were heavily damaged, causing power outages and water shortages for the entire island for months (United States Department of Energy and Restoration 2017). Fatalities as a consequence of Maria are still under investigation, however the most recent estimates suggest between 793 to 8,498 excess deaths occurred following the storm (Kishore et al. 2018). Total economic losses are estimated to be $ 92 billion.
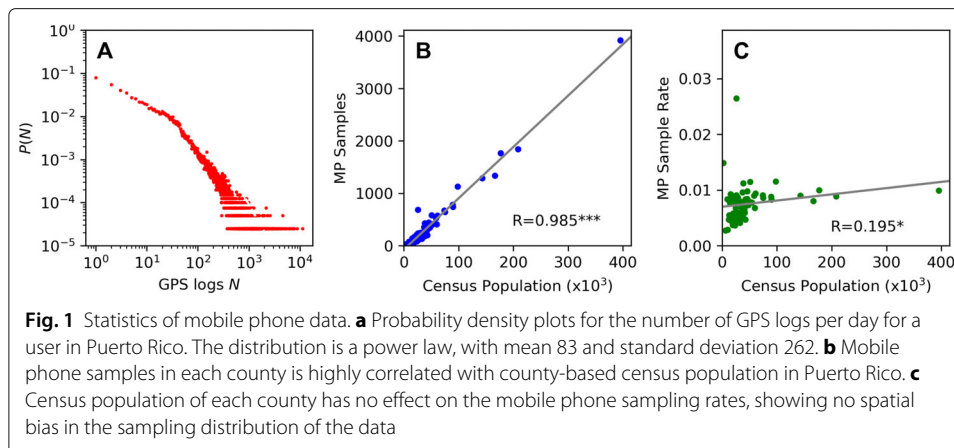
Physical damage caused by the hurricane are measured by the housing damage rates in each county, which was provided through the "Housing Assistance Data" provided by the Federal Emergency Management Agency (FEMA). The raw data can be found through the link[4]. We defined "housing damage rate" for each county as the total number of houses that were inspected to have had more than $ 10,000 worth of damage due to the target hurricane, divided by the number of households in that county. Figure 2c shows the spatial distribution of the housing damages in Puerto Rico. They gray dotted lines show the trajectory of the hurricane, and we can observe higher housing damage rates near the trajectory. We can observe that many of the counties in Puerto Rico experienced high housing damage rates, between 20% and 60%.

Methods

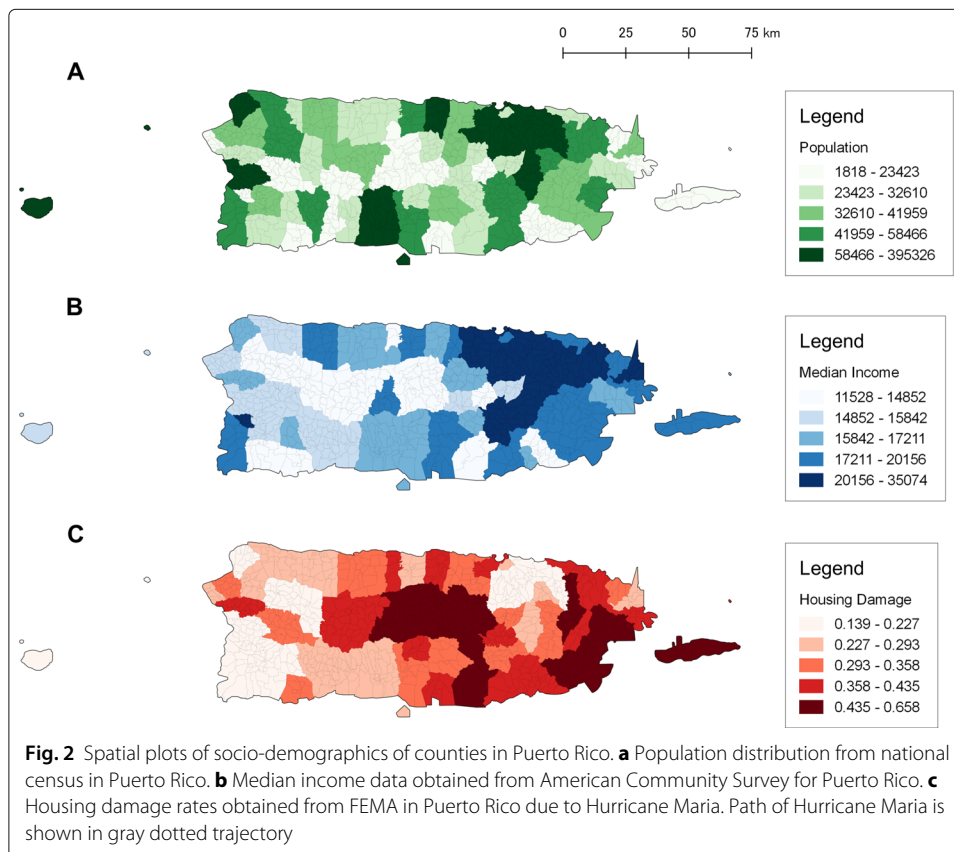#### Estimating population recovery from fragmented Mobile phone data

To estimate the recovery time of each county, the population displacement dynamics were estimated from mobile phone observations. First, the home locations of individuals were estimated from the mobile phone data observed prior to the hurricane. It is well known that human mobility trajectories show a high degree of temporal and spatial regularity, each individual having a significant probability to return to a few highly frequented locations, including his/her home location (Gonzalez et al. 2008). Due to this characteristic, it has been shown that home locations of individuals can be estimated with high accuracy by clustering the individual's stay point locations over night (Calabrese et al. 2011). Home locations of each individual were estimated by applying mean-shift clustering to

---
[4]https://www.fema.gov/media-library/assets/documents/34758

Fig. 1 Statistics of mobile phone data. **a** Probability density plots for the number of GPS logs per day for a user in Puerto Rico. The distribution is a power law, with mean 83 and standard deviation 262. **b** Mobile phone samples in each county is highly correlated with county-based census population in Puerto Rico. **c** Census population of each county has no effect on the mobile phone sampling rates, showing no spatial bias in the sampling distribution of the data
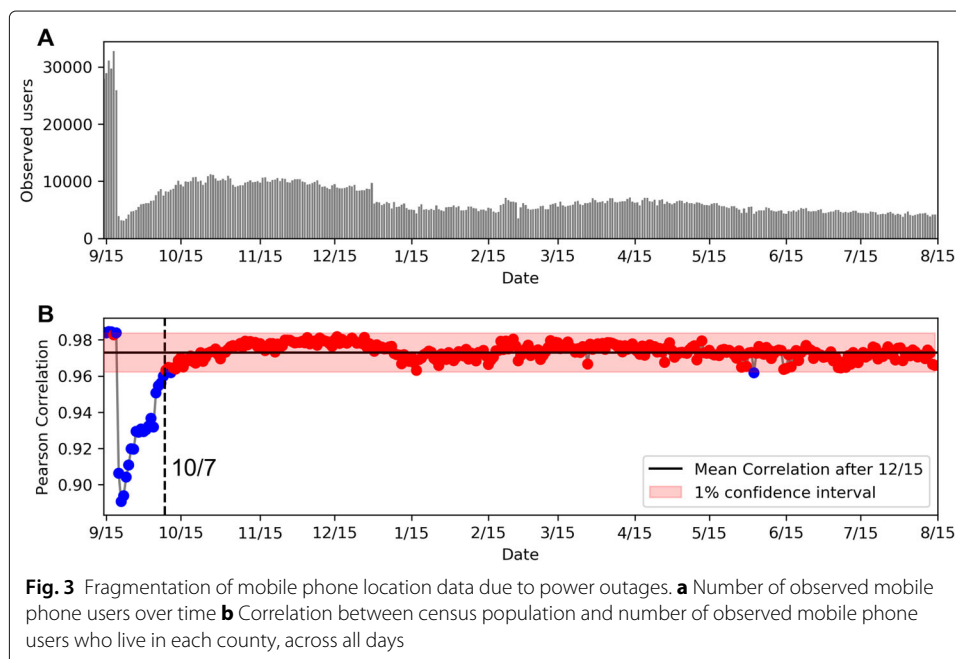
the nighttime stay points (observed between 8PM and 6AM), weighted by the duration of stays in each location (Ashbrook and Starner 2003; Kanasugi et al. 2013). Mean shift clustering was implemented using the `scikit-learn` package on Python. An individual was detected to be displaced if the individual is estimated to be staying in a location outside the city of his/her estimated home location. Such home location estimates were used to check the representativeness of mobile phone user samples in Fig. 1.

Second, using the longitudinal observations during and after the hurricane, the displacement rate of each county was estimated. The night-time stay points of each

Fig. 2 Spatial plots of socio-demographics of counties in Puerto Rico. **a** Population distribution from national census in Puerto Rico. **b** Median income data obtained from American Community Survey for Puerto Rico. **c** Housing damage rates obtained from FEMA in Puerto Rico due to Hurricane Maria. Path of Hurricane Maria is shown in gray dotted trajectory

individual were estimated from their mobile phone location data observed during night time, using mean-shift clustering of mobile phone observation points weighted by the length of stay at each point. We spatially aggregated the number of people estimated to have stayed the night in each county to obtain the daily population count for each county. However, we found that there was a significant decrease in the number of observed users during and after Hurricane Maria. The number of mobile phone users observed before, during and after Hurricane Maria are shown in Fig. 3a. We observe a significant decrease in the number of users over time, which is assumed to be due to various reasons including: 1) no signals due to damage to mobile phone towers, 2) dormant users due to lack of power to charge their phones, and 3) users who uninstalled their app which collected location data. However, since our interest is in the displacement rate of the users in each county, the absolute number of observed mobile phone users is less important in this study. Rather, whether or not the sampling rate of mobile phone users (# of mobile phone users estimated to be residents of county i / # of census population of county i) is uniform across all counties is important for our analysis. The Pearson Correlation between each county's census population and observed mobile phone users for each day, are plotted in Fig. 3b. The red plots indicate days where the correlation is within 1% confidence interval from the mean correlation after December 15th, where the sample rates are stable. The blue plots indicate days where the correlation is significantly low, where sample rates of mobile phone user samples are biased across counties. We observe that until October 7th, the Pearson Correlation is significantly lower than usual, indicating that there is spatial inequality in the sampling rates of mobile phone users. However, after October 7th, which is 16 days from the Hurricane, the spatial bias becomes minimal (high correlation), and shows that we are able to construct the population recovery dynamics without significant sampling bias. Thus, we use population displacement data observed after October 7th to estimate the recovery time to avoid the effects of power outages and mobile phone tower outages.

Fig. 3 Fragmentation of mobile phone location data due to power outages. **a** Number of observed mobile phone users over time **b** Correlation between census population and number of observed mobile phone users who live in each county, across all days

The observed population dynamics are fitted with an exponential function. We use the exponential model instead of the linear recovery function due to better fit to data (see "Appendix" section). We define "time until recovery", denoted by $\tilde{T}_i^\alpha$, as the timing of the first time the city's population recovers to $M_i^\alpha = M_i^\infty \cdot \alpha$, which is characterized by the long term normalized population after the disaster $\left(M_i^\infty\right)$ and a constant parameter $0 < \alpha < 1$. In cases where the long term population exceeds the original population $\left(M_i^{bef} = 1 < M_i^\infty\right)$, the amount of time after exceeding the original population should not be considered as time used for recover (rather, the population is growing after the disaster). In these cases, $\tilde{T}_i^\alpha$ is defined as the timing of the first time the city's population recovers to α, which is the population of pre-disaster level multiplied by α. In sum, time until recovery is defined as the following:

$$\tilde{T}_i^\alpha = argmin_t \left(M_i(t) \geq \min(\alpha M_i^\infty, \alpha)\right) \tag{1}$$

where $\min(x, y)$ returns the smaller value between x and y. $\alpha = 0.95$ is used as the threshold parameter in the further results, since time until recovery are not affected by the parameter value significantly (see "Appendix" section). For simple notations, we write $\tilde{T}_i^{0.95}$ as T_i and call the variable "Time Until Recovery" in the following analysis. Figure 4 shows the estimated recovery times T_i of each county. We observe a general trend where the counties around San Juan recover quickly (yellow), and recovery spreads to coastal cities, then interior regions (red).
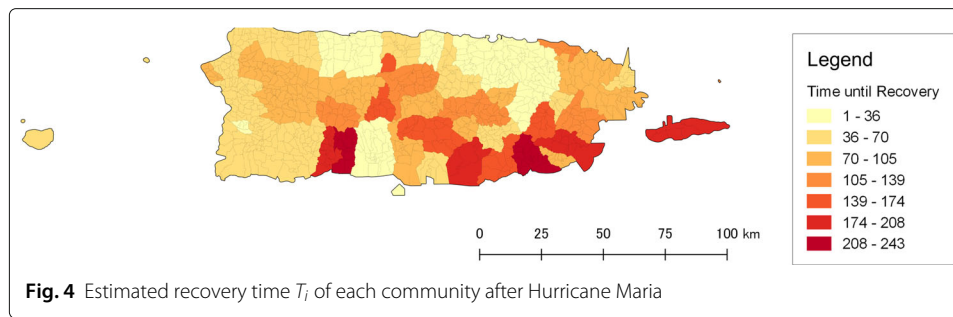
Node-level network statistics

Distance metrics for edge weights

To investigate the effect of various types of inter-city dependencies on post-disaster recovery, we construct multiple networks based on various edge weights between nodes (78 counties in Puerto Rico). Table 2 lists the 5 distance metrics that were used as edge weights to build the inter-city network $\mathcal{N}$ in this study. Given a distance metric x, we denote the network built using distance metric x as $\mathcal{N}_x$. The distance metrics (edge weights) are Euclidean distance e, travel time TT, road distance RD, mobility flow F, and the number of overnight stays S. Euclidean distance is an undirected metric, calculated by $e_{ij} = \sqrt{(x_i - x_j)^2 + (y_j - y_j)^2}$ given center points of two counties (x_i, y_i) and (x_j, y_j). Travel time and road distance between counties i and j were calculated using Google Maps API[5] in the usual conditions (prior to the disaster). Although these metrics are almost symmetrical, there are differences in travel times and fastest routes depending on the direction of travel, thus would produce directed networks. The two social distance metrics are mobility flow F and the number of overnight stays S, which are both observed using mobile phone data from prior to the disaster. Mobility flow F_{ij} is defined as the inverse of the average number of travelers from node i to node j in the usual state (prior to disaster). The number of overnight stays S_{ij} is similar to mobility flow, and is defined as the inverse of the average number of visitors from node i who stay overnight at node j in the usual state (prior to disaster).

Figure 5 visualizes the networks built using the 5 distance metrics. Node sizes are proportional to the pre-disaster population of each county prior to the disaster. All of these metrics can be defined for all origin-destination pairs, however for clarity of the figures, we only show the edges with the 500 highest edge weights. The weight of each edge is

---

[5]https://developers.google.com/maps/documentation/

Fig. 4 Estimated recovery time T_i of each community after Hurricane Maria

shown as the width of each edge. For the direct networks (**B-E**), sum of the weights on both directions are visualized. From visual inspection, we can see a large difference in the structure of networks built based on physical (**A-C**) and social metrics (**D,E**). In particular, in **D)** and **E)**, we can see that some cities on the coast have large direct weights between San Juan, which mean that despite the large physical distance, there are many people who travel and/or stay overnight between these cities.

Table 3 shows the element-wise Pearson correlation between edge weights based on different distance metrics. We can see that Euclidean distance $\mathcal{N}_e$ has high correlations with road distance $\mathcal{N}_{RD}$ and travel times $\mathcal{N}_{TT}$, thus we exclude Euclidean distance from our analysis, and focus on networks constructed by the four remaining distance metrics.

Network statistics

Using the four different networks built from physical and social distance metrics defined above, we attempt to understand what type of network statistic on these networks can explain the time until recovery well in Puerto Rico. Identifying the most highly correlated network statistic and distance metric could provide insights into the underlying process that dictates the recovery of counties after disasters. In a similar manner as the distance metrics, we test various node-level statistics to compute the importance of each node in each network. Table 4 lists the node-level statistics that we will test on the four different networks $\mathcal{N}_x$. The six network statistics for node i in network $\mathcal{N}_x$ are: weighted in, out, and total degrees ($WI(\mathcal{N}_x)_i$, $WO(\mathcal{N}_x)_i$, and $W(\mathcal{N}_x)_i$, respectively), weighted clustering coefficient $CC(\mathcal{N}_x)_i$, direct distance from the source of recovery (San Juan in the case of Puerto Rico) $DSJ(\mathcal{N}_x)_i$, and shortest path distance from the source of recovery $SPSJ(\mathcal{N}_x)_i$.

Such network statistic measures were proposed in the literature of weighted directed and spatial networks to quantify the characteristics of each node (Barthélemy 2011; Barrat

Table 2 Distance metrics x that were used as edge weights to build inter-city network $\mathcal{N}_x$

Distance metric x	Notation	Directed?	Category	Description
Euclidean distance	e_{ij}	Undirected	Physical	Euclidean distance between center of counties i and j
Travel time	TT_{ij}	Directed	Physical	Driving time from i to j in usual state (`GoogleMaps API`)
Road distance	RD_{ij}	Directed	Physical	Road distance on quickest route from i to j in usual state (`GoogleMaps API`)
Mobility flow	F_{ij}	Directed	Social	Inverse of the # of travelers from i to j in usual state (from mobile phones)
# of overnight stays	S_{ij}	Directed	Social	Inverse of the # of visitors from county i who stay overnight in county j in usual state (from mobile phones)
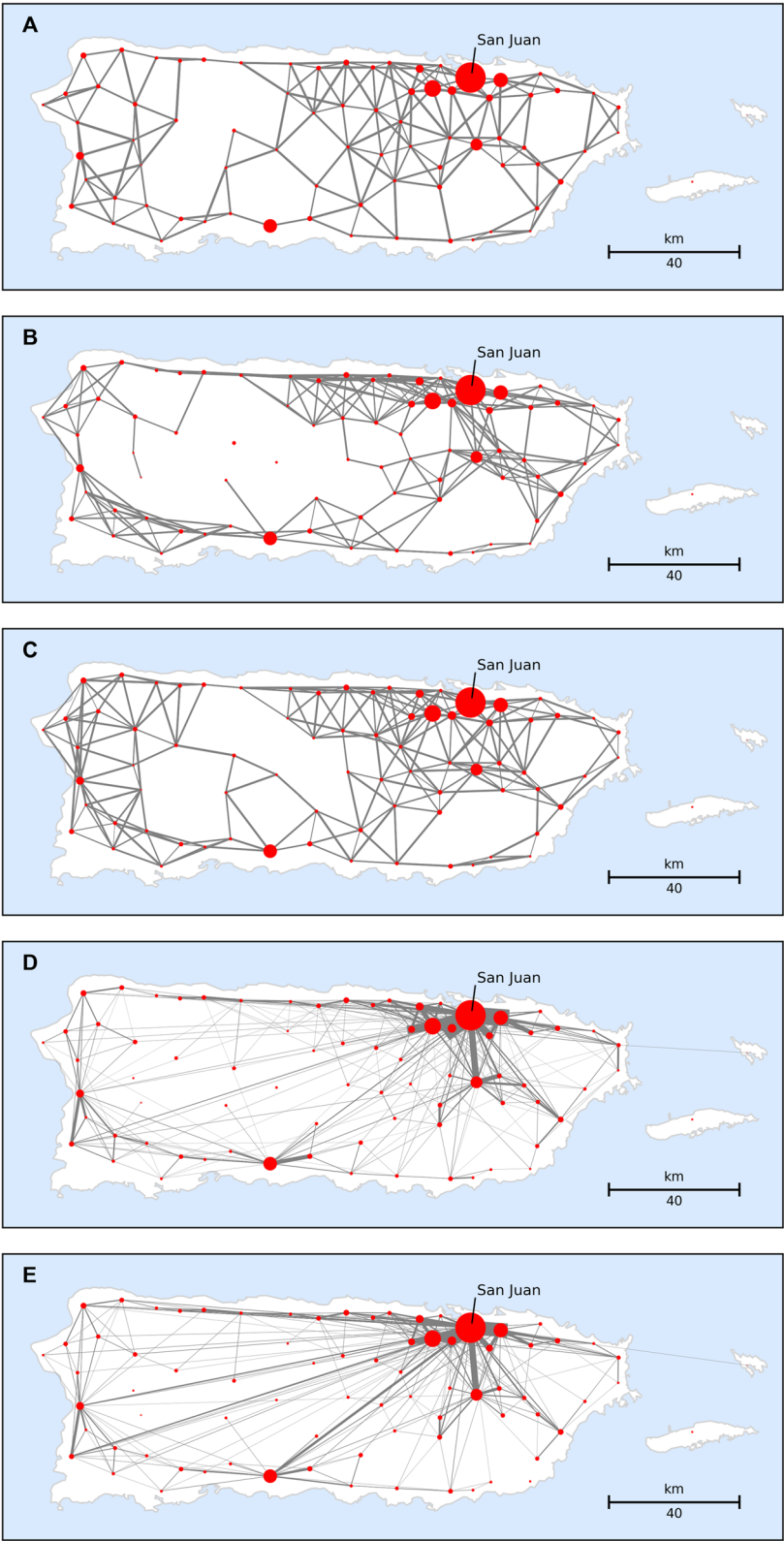
Fig. 5 Visualization of distance weights of each network metric

Table 3 Correlation between edge weights based on distance metrics $\mathcal{N}_x$

	$\mathcal{N}_e$	$\mathcal{N}_{TT}$	$\mathcal{N}_{RD}$	$\mathcal{N}_F$	$\mathcal{N}_S$
$(\mathcal{N}_e)$	1.000	0.605	0.933	0.271	0.293
$\mathcal{N}_{TT}$		1.000	0.923	0.357	0.341
$\mathcal{N}_{RD}$			1.000	0.307	0.304
$\mathcal{N}_F$				1.000	0.509
$\mathcal{N}_S$					1.000

et al. 2005; Barrat et al. 2004; Opsahl et al. 2010; Fagiolo 2007). Such metrics are applied in various domains to assess the importance of nodes in weighted networks, including urban networks (Zhong et al. 2014), urban traffic networks (De Montis et al. 2007), airport networks (Li and Cai 2004), and metabolic processes (Almaas et al. 2004).

A weighted network $\mathcal{N}_x$ can be described with a $N \times N$ adjacency matrix W, where w_{ij} denotes the $(i.j)$-th element and the weight assigned to edge i to j. The weighted in, out, total degrees, weighted clustering coefficient, and weighted betweenness centrality of node i were calculated for each weighted network $\mathcal{N}_x$. The weighted in-degree of node i is calculated by $WI(\mathcal{N}_x)_i = \sum_{j \in \mathcal{V}(i)} w_{ji}$, where $\mathcal{V}(i)$ is the set of nodes connected to node i. Similarly, the weighted out-degree of node i is calculated by $WO(\mathcal{N}_x)_i = \sum_{j \in \mathcal{V}(i)} w_{ij}$. The weighted total degree of node i is the sum of in and out degrees $W(\mathcal{N}_x)_i = WI(\mathcal{N}_x)_i + WO(\mathcal{N}_x)_i$. The weighted clustering coefficient measures the statistical level of cohesiveness around node i. In general, this value decays with respect to degrees, shown in both airplane network and authors network (Barrat et al. 2004). This is because low degree nodes are connected to highly connected communities, while large degree nodes are connected to many nodes that are not directly connected. It is computed by $CC(\mathcal{N}_x)_i = \frac{(\hat{W} + \hat{W}^T)^3_{ii}}{2T_i^D}$, where $\hat{W} = \left\{ w_{ij}^{\frac{1}{3}} \right\}$ and $T_i^D = d_i^{tot}\left(d_i^{tot} - 1\right) - 2d_i^{\leftrightarrow}$, $d_i^{\leftrightarrow} = W_{ii}^2$. The direct distance from San Juan of node i is the simply the distance metric between node i and San Juan. San Juan is chosen as the destination because it was the major source of recovery after Hurricane Maria in Puerto Rico. Similarly, we also measure the shortest path distance from San Juan on network $\mathcal{N}_x$.

Thus, in summary, we construct networks based on four different distance metrics, and define six node-level statistics for each defined network. This gives us 24 different network metrics that each quantify the physical and/or social characteristics of each node from different aspects. In the next section, we test whether these network metrics can explain the variance in recovery speed across counties in Puerto Rico after Hurricane Maria.

Table 4 Node-level statistics of node i given inter-city network $\mathcal{N}_x$

Network statistic	Notation	Description
Weighted in-degree	$WI(\mathcal{N}_x)_i$	Total weighted in-degree of i on $\mathcal{N}_x$
Weighted out-degree	$WO(\mathcal{N}_x)_i$	Total weighted out-degree of i on $\mathcal{N}_x$
Weighted total degree	$W(\mathcal{N}_x)_i$	Total weighted degree of i on $\mathcal{N}_x$
Weighted clustering coefficient	$CC(\mathcal{N}_x)_i$	Weighted clustering coefficient of i on $\mathcal{N}_x$
Direct distance from San Juan	$DSJ(\mathcal{N}_x)_i$	Direct distance to i from San Juan on $\mathcal{N}_x$
Shortest path distance from San Juan	$SPSJ(\mathcal{N}_x)_i$	Distance of shortest path to i from San Juan on $\mathcal{N}_x$

Spatial regression models

First, ordinary least squares (OLS) method is used to estimate the parameters of the general regression model specified below.

$$y = x\beta + \epsilon \tag{2}$$

where, y is an $n \times 1$ vector representing the objective variable (recovery time), x is an $n \times k$ matrix of the independent variables, and β is a $k \times 1$ vector of the coefficients. Here, the error term ϵ is assumed to be an i.i.d. normal. When there is spatial dependence in the error term, the i.i.d. normal assumption is violated. Two approaches are taken to deal with spatial correlation (Ward and Gleditsch 2018). First is the spatial lag model, where the error term is decomposed into a spatially lagged term for the dependent variable and an independent error term, $\epsilon = \rho Wy + e$, where W is the matrix that reflects the spatial proximity between areas which is commonly defined by encoding the k-nearest neighbors. This gives us the spatial lag model, described by the following equation:

$$y = \rho Wy + x\beta + \epsilon \tag{3}$$

where $\epsilon \sim N(0, \sigma^2 I)$. The parameters can be estimated using maximum likelihood estimation.

The other approach is to assume that the error is spatially correlated, instead of the objective variables affecting the objective variables of neighboring areas. We can write the spatial error model as follows:

$$y = x\beta + \lambda W\xi + \epsilon \tag{4}$$

where $\epsilon \sim N(0, \sigma^2 I)$. In the following regression analysis, we first test the general regression model and also test for significance in spatial lag ρ and spatial error λ terms. Then, we apply the appropriate spatial regression analysis accordingly.

Results

Table 5 shows the statistics of the independent and objective variables. In the regression models, recovery time T_i is set as the objective variable. The variables in the second block (county population, median income, and housing damage rate) and one variable from the third block (network statistics) are used as independent variables for each of the regression models. The network statistic variables are normalized by the following equation:

$$\hat{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{5}$$

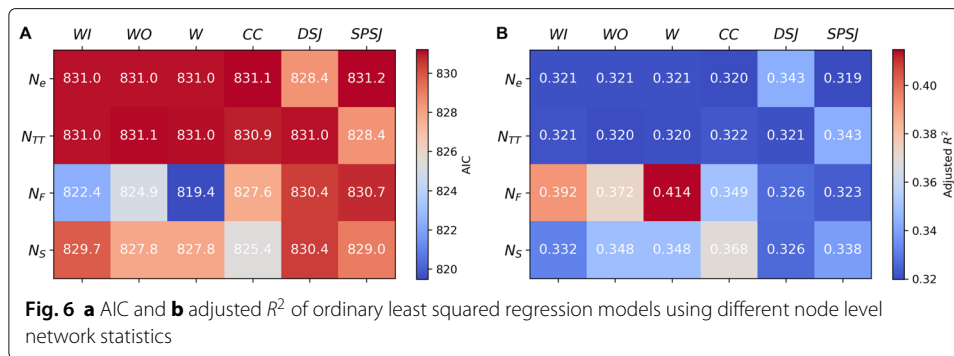where $\max(x)$ and $\min(x)$ are maximum and minimum values of variable x.

The Pearson correlations between recovery time and each independent variable are shown in the right column of Table 5. We observe that all socio-demographic variables (county population, median income, and housing damage rate) have moderate correlations with recovery time, as shown in past studies (Yabe et al. 2019). Among the network statistic variables, all of the node-level statistics of the mobility flow based network and overnight stay network had significant correlations with recovery time, indicating the significant effect of inter-city social connectivity on post-disaster recovery. In contrast to social connectivity, node-level statistics computed from physical networks were less significantly correlated with recovery time in Puerto Rico. To examine the collinearity effect of these network statistics, we test the regression models using each node-level

Table 5 Statistics of independent and objective variables

Variable	Min.	Max.	Med.	Mean	Std. Dev.	Corr. with T_i
Recovery time T_i (days)	1	243	74	80.2	58.2	–
County population	1,818	395,326	34,154	46,862	53,884	-0.476***
Median income ($)	11528	35074	17054.5	18152	4466	-0.489***
Housing damage rate	0.138	0.657	0.339	0.349	0.119	0.384***
$WI(\mathcal{N}_{TD})$	0	1.0	0.183	0.232	0.197	0.054
$WI(\mathcal{N}_{TT})$	0	1.0	0.585	0.581	0.152	0.105
$WI(\mathcal{N}_F)$	0	1.0	0.356	0.386	0.208	-0.55***
$WI(\mathcal{N}_S)$	0	1.0	0.396	0.404	0.201	-0.43***
$WO(\mathcal{N}_{TD})$	0	1.0	0.184	0.232	0.198	0.054
$WO(\mathcal{N}_{TT})$	0	1.0	0.578	0.580	0.151	0.084
$WO(\mathcal{N}_F)$	0	1.0	0.310	0.352	0.205	-0.564***
$WO(\mathcal{N}_S)$	0	1.0	0.341	0.350	0.204	-0.498***
$W(\mathcal{N}_{TD})$	0	1.0	0.184	0.232	0.198	0.054
$W(\mathcal{N}_{TT})$	0	1.0	0.582	0.580	0.151	0.094
$W(\mathcal{N}_F)$	0	1.0	0.349	0.400	0.217	-0.602***
$W(\mathcal{N}_S)$	0	1.0	0.359	0.382	0.220	-0.511***
$CC(\mathcal{N}_{TD})$	0	1.0	0.182	0.221	0.185	0.098
$CC(\mathcal{N}_{TT})$	0	1.0	0.832	0.811	0.151	0.011
$CC(\mathcal{N}_F)$	0	1.0	0.361	0.373	0.150	0.533***
$CC(\mathcal{N}_S)$	0	1.0	0.509	0.509	0.176	0.570***
$DSJ(\mathcal{N}_{TD})$	0	1.0	0.367	0.451	0.279	0.092
$DSJ(\mathcal{N}_{TT})$	0	1.0	0.431	0.487	0.272	0.187*
$DSJ(\mathcal{N}_F)$	0	1.0	0.084	0.185	0.243	0.282**
$DSJ(\mathcal{N}_S)$	0	1.0	0.185	0.288	0.263	0.397***
$SPSJ(\mathcal{N}_{TD})$	0	1.0	0.036	0.064	0.154	0.198*
$SPSJ(\mathcal{N}_{TT})$	0	1.0	0.367	0.451	0.279	0.092
$SPSJ(\mathcal{N}_F)$	0	1.0	0.117	0.143	0.144	0.326***
$SPSJ(\mathcal{N}_S)$	0	1.0	0.173	0.205	0.179	0.482***

network statistics. Figure 6 shows the Akaike Information Criterion (*AIC*) and adjusted R^2 of each regression model. We observe that the regression model using the weighted total degree of the mobility flow network ($W(\mathcal{N}_F)$) has the lowest *AIC* and adjusted R^2 value. Table 6 shows the detailed regression results of the two regression models with and without the $W(\mathcal{N}_F)$ variable. The estimated regression coefficients and their significance levels are shown in stars. Using the network statistic, both the *AIC* and adjusted R^2 improve significantly, and we observe that the population variable becomes insignificant when considering the inter-city network variable. Moreover, the results show that the network metric variable negatively affects recovery time, meaning that the more the influx and outflux mobility flow before the disaster, the quicker the recovery.

The spatial dependence of recovery time is tested using various metrics in Table 7. All metrics including the Moran's I, the lag Lagrange multiplier ρ, and the error Lagrange multiplier λ are significant, showing significant spatial dependence. Robust tests of both Lagrange multipliers show that the error Lagrange multiplier λ is more significant. Thus, we test the Spatial Error Model and compare the results with the OLS Model in Table 8. Results show that the *AIC* is lower in the Spatial Error Model, indicating that spatial dependence in the error term explains the heterogeneity in recovery time across the counties in Puerto Rico. In both models, income levels and the network metric ($W(\mathcal{N}_F)$)

Fig. 6 a AIC and **b** adjusted R^2 of ordinary least squared regression models using different node level network statistics

have significant effects on the recovery time. Housing damage rates, however, become insignificant under the spatial error model.

Discussions

Our analyses based on observational data from Puerto Rico after Hurricane Maria confirmed that inter-city network metrics, namely the pre-disaster mobility flow, has a significant positive influence on the speed of recovery. The results imply that the more socially connected an area is to other counties, the more easier it is for people living in those communities to receive support and to recovery quickly. This paper introduces a new perspective in the community resilience literature, where we take into account the inter-city dependencies in the recovery process rather than analyzing each community as independent entities. These insights encourage communities to prepare for future hazards by not only preparing its physical infrastructure (e.g. roads), but also by strengthening their social connectivity with other cities, to have greater chances of receiving support in case of emergencies.

Now, we discuss future research opportunities that this study enables. First, Puerto Rico is a unique case study because of its island geography. It is valuable to examine whether the same rules apply to other regions with different geographical characteristics. We will start collecting additional data from other disaster events to test the generalizability of our method between different disaster events. The Haiti Earthquake is an example where a large disaster struck a low-income island region. Another example where we observe severe damage is the Tohoku Tsunami (Japan) in 2011, where the coastal cities of the east coast of the Tohoku region are still struggling to recovery from the disaster. Comparing the analysis presented in this study across different disaster instances would be an interesting topic for future research.

Table 6 Ordinary Least Squares Regression Results with $W(\mathcal{N}_F)$ as network metric

	Socio-demographic only	Socio-demographic + network metric
Constant	297.21***	22.48
ln(Population)	-19.18**	10.54
Income	-4.06***	-2.78**
Housing Damage	140.6**	140.1**
Network Metric	–	-145.5***
Adjusted R^2	0.329	0.416
AIC	829.21	819.48

*** $p < 0.01$, ** $p < 0.05$

Table 7 Spatial dependence tests

Moran's I	7.433***
Lag multiplier ρ	24.82***
Error multiplier λ	31.60***
Robust lag multiplier ρ	0.747
Robust error multiplier λ	7.53***

*** $p < 0.01$

Second, modeling the underlying process of recovery was not in the scope of this study. This work was limited to testing the statistical significance of network metrics using econometric models. To better understand, predict and control the recovery of communities after disasters, there is a need to model the underlying process that dictates the population recovery. Developing agent based models and system dynamics models for predicting community recovery based on the insights obtained from this study will be the next steps in our research.

Thirdly, the reliability of the results in this study could improve if we could increase the diversity of datasets to quantify the social connectivity between counties. One candidate would be Twitter data, where we can use text-mining to determine counties that have frequent contacts via messaging or retweeting. Also, call records of mobile phones would be a good data source to quantify social connections between two counties. In Puerto Rico, social connectivity measured by the mobility of people was shown to explain recovery times. It would be valuable to investigate whether the social measures used in this study would apply to other regions with different social norms, such as Japan or mainland US. Using more datasets to investigate such questions on the inter-city dependencies in the recovery process would be the focus of future studies.

Conclusions

Using large scale mobile phone data collected from Puerto Rico, we revealed the importance of inter-city social connectivity on disaster recovery after Hurricane Maria. More specifically, we showed that observing the mobility patterns between counties prior to the disaster can increase the predictability of time until recovery of communities. These insights highlight the importance of communities and policy makers to invest more into developing the social networks across counties or nearby cities through the interaction of people prior to the disaster to prepare for future disasters, as well as investing into the physical infrastructure networks.

Table 8 Regression results of spatial error model

	Non-spatial	Spatial error model
Constant	22.48	78.40
ln(Population)	10.54	12.31
Income	-2.78**	-3.17**
Housing damage	140.1**	10.86
Network metric	-145.5***	-155.49***
λ	–	0.78***
AIC	819.48	796.84

*** $p < 0.01$, ** $p < 0.05$

Appendix

On the choice of the exponential recovery model

The normalized population data were fitted using two different functions to denoise the observations. We test two functions: 1) linear model and 2) exponential model.

The exponential model is defined as follows:

$$M_i\left(t|\tau_i, M_i^0, M_i^\infty\right) = M_i^\infty - \left(M_i^\infty - M_i^0\right)\exp\left(-\frac{t - T_0}{\tau_i}\right) \tag{6}$$

where, τ_i is the recovery speed of city i, M_i^0 is the initial population rate of city i, and M_i^∞ is the long term population rate of city i. $t = T_0$ denotes the timing of initial population rate. Figure 7a shows an illustration of the exponential model.

The alternative is the linear model, defined as the following:

$$M_i\left(t|\beta_i, M_i^0, M_i^\infty\right) = \begin{cases} M_i^0 + \beta_i(t - T_0)(t \le T_c) \\ M_i^\infty(t > T_c) \end{cases} \tag{7}$$

where, β_i is the coefficient of the linear recovery speed of city i, M_i^0 is the initial population rate of city i, and M_i^∞ is the long term population rate of city i. $t = T_c$ denotes the timing where the linear recovery line reaches M_i^∞. Figure 7b shows an illustration of the linear model.

We chose the model by testing which model is able to fit the empirical data obtain from Puerto Rico after Hurricane Maria. Parameters (τ_i, β_i and M_i^∞) of the models were fitted using least squares method. M_i^0, which is the initial population rate, is defined as $M_i^0 = \min_t M_i(t)$. Squared error (SE) and Pearson's correlation coefficient ρ between the observed and fitted time series were used as evaluation metrics. SE is defined as:
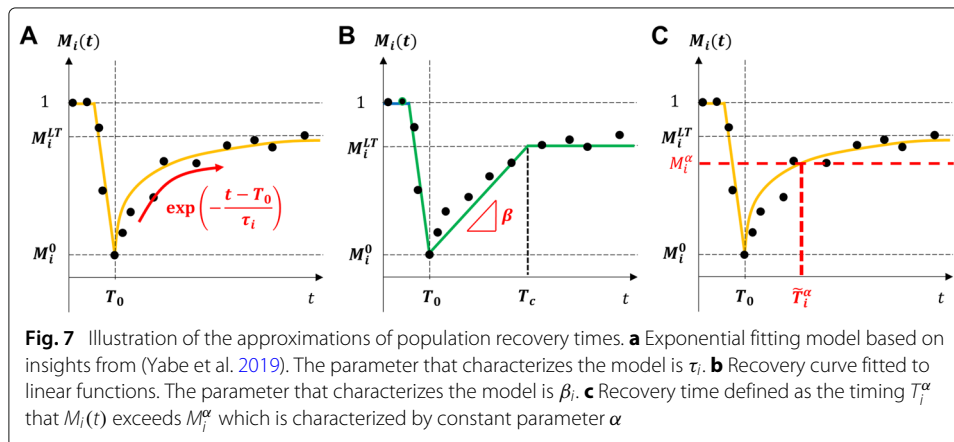
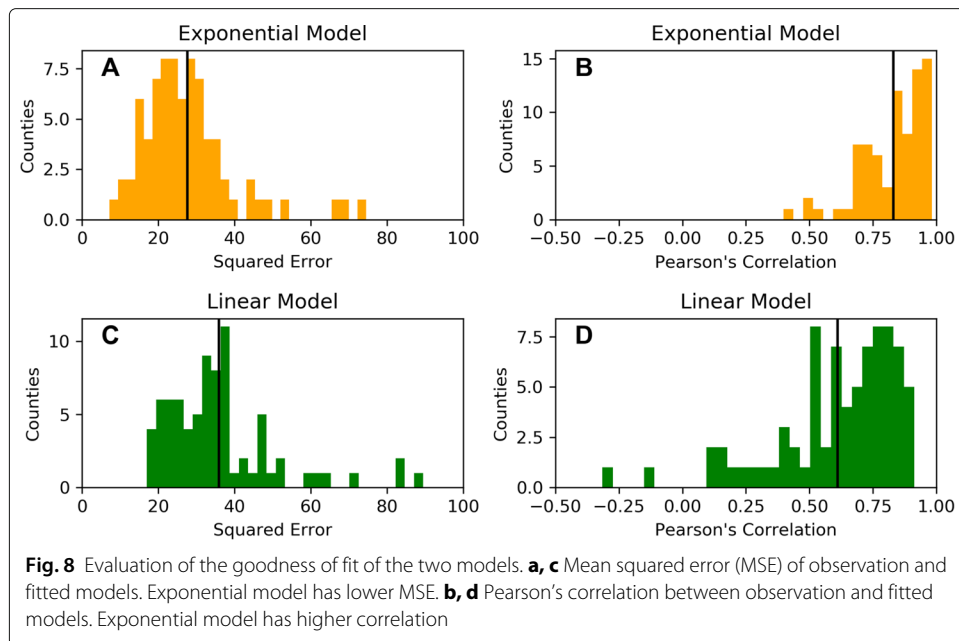$$\text{SE}\left(M(t), \hat{M}(t)\right) = \sqrt{\left(M(t) - \hat{M}(t)\right)^2} \tag{8}$$

where $M(t)$ and $\hat{M}(t)$ are observed and fitted values of the time series, and T is the total number of time steps. Pearson's correlation coefficient ρ is defined as:

$$\rho\left(M(t), \hat{M}(t)\right) = \frac{Cov(M(t), \hat{M}(t))}{\sigma_{M(t)}\sigma_{\hat{M}(t)}} \tag{9}$$

where $Cov(x, y)$ is the covariance between x and y, σ_x is the standard deviation of x.

Figure 8 shows the histograms of the two metrics for the two models, to evaluate the goodness of fit of the two models. We can observe that the exponential model has lower

Fig. 7 Illustration of the approximations of population recovery times. **a** Exponential fitting model based on insights from (Yabe et al. 2019). The parameter that characterizes the model is τ_i. **b** Recovery curve fitted to linear functions. The parameter that characterizes the model is β_i. **c** Recovery time defined as the timing T_i^α that $M_i(t)$ exceeds M_i^α which is characterized by constant parameter α

Fig. 8 Evaluation of the goodness of fit of the two models. **a, c** Mean squared error (MSE) of observation and fitted models. Exponential model has lower MSE. **b, d** Pearson's correlation between observation and fitted models. Exponential model has higher correlation
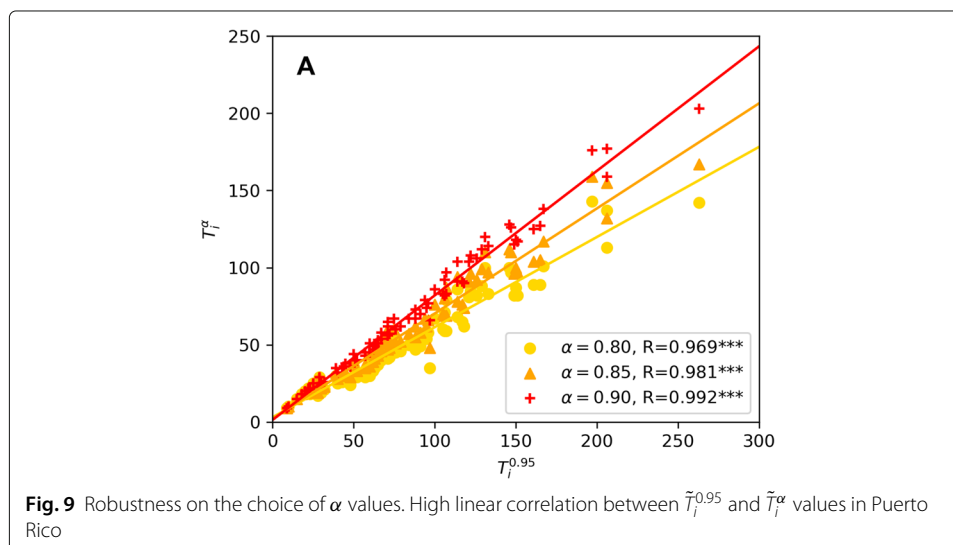
SE and higher correlation, meaning that it is able to better approximate the observations. Thus, in our further analysis, we use the results obtained by fitting the observations to the exponential model.

On the robustness against parameter α for recovery time estimation

As shown in Fig. 9, Pearson's correlation coefficients between $\tilde{T}_i^{0.80}$, $\tilde{T}_i^{0.85}$, $\tilde{T}_i^{0.90}$, $\tilde{T}_i^{0.95}$ are all very high ($\rho > 0.96$). This implies that the $\tilde{T}_i^{\alpha}$ values do not depend highly on the values of α. Thus, we use $\alpha = 0.95$ as a parameter for our further analysis. For notational simplicity, we write $\tilde{T}_i^{0.95}$ as $\tilde{T}_i$ and call the variable "time until recovery" in the analysis.

Fig. 9 Robustness on the choice of α values. High linear correlation between $\tilde{T}_i^{0.95}$ and $\tilde{T}_i^{\alpha}$ values in Puerto Rico

Author details
[1]Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN, USA. [2]Department of Agronomy, Purdue University, 915 W. State Street, West Lafayette, IN, USA.

References
Aerts J, Botzen W, Clarke K, Cutter S, Hall JW, Merz B, Michel-Kerjan E, Mysiak J, Surminski S, Kunreuther H (2018) Integrating human behaviour dynamics into flood disaster risk assessment. Nat Clim Chang 8(3):193
Aldrich DP (2012) Building Resilience: Social Capital in Post-disaster Recovery. University of Chicago Press, Chicago
Almaas E, Kovacs B, Vicsek T, Oltvai Z, Barabási A-L (2004) Global organization of metabolic fluxes in the bacterium escherichia coli. Nature 427(6977):839
Ashbrook D, Starner T (2003) Using gps to learn significant locations and predict movement across multiple users. Pers Ubiquit Comput 7(5):275–286
Bagrow JP, Wang D, Barabasi A-L (2011) Collective response of human populations to large-scale emergencies. PLoS ONE 6(3):17680
Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Nat Acad Sci 101(11):3747–3752
Barrat A, Barthélemy M, Vespignani A (2005) The effects of spatial constraints on the evolution of weighted complex networks. J Stat Mech: Theory Exper 2005(05):05003
Barthélemy M (2011) Spatial networks. Phys Rep 499(1-3):1–101
Calabrese F, Di Lorenzo G, Liu L, Ratti C (2011) Estimating origin-destination flows using opportunistically collected mobile phone location data from one million users in boston metropolitan area. IEEE Perv Comput 10(4):36–44
Cutter SL, Barnes L, Berry M, Burton C, Evans E, Tate E, Webb J (2008) A place-based model for understanding community resilience to natural disasters. Glob Environ Change 18(4):598–606
De Montis A, Barthélemy M, Chessa A, Vespignani A (2007) The structure of interurban traffic: a weighted network analysis. Environ Plan B: Plan Des 34(5):905–924
Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ (2014) Dynamic population mapping using mobile phone data. Proc Nat Acad Sci 111(45):15888–15893
DeWaard J, Curtis KJ, Fussell E (2016) Population recovery in new orleans after hurricane katrina: exploring the potential role of stage migration in migration systems. Popul Environ 37(4):449–463
Eakin H, Bojórquez-Tapia LA, Janssen MA, Georgescu M, Manuel-Navarrete D, Vivoni ER, Escalante AE, Baeza-Castro A, Mazari-Hiriart M, Lerner AM (2017) Opinion: urban resilience efforts must consider social and political forces. Proc Nat Acad Sci 114(2):186–189
Fagiolo G (2007) Clustering in complex directed networks. Phys Rev E 76(2):026107
Finch C, Emrich CT, Cutter SL (2010) Disaster disparities and differential recovery in new orleans. Popul Environ 31(4):179–202
Fussell E, Curtis KJ, DeWaard J (2014) Recovery migration to the city of new orleans after hurricane katrina: a migration systems approach. Popul Environ 35(3):305–322
Gitay H, Bettencourt S, Kull D, Reid R, McCall K, Simpson A, Krausing J, Ambrosi P, Arnold M, Arsovski T, et al. (2013) Building resilience: Integrating climate and disaster risk into development-lessons from world bank group experience. The World Bank Experience
Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779
Gray CL, Mueller V (2012) Natural disasters and population mobility in bangladesh. Proc Nat Acad Sci 109(16):6000–6005
Jiang S, Yang Y, Gupta S, Veneziano D, Athavale S, González MC (2016) The timegeo modeling framework for urban mobility without travel surveys. Proc Nat Acad Sci 113(37):5370–5378

Kanasugi H, Sekimoto Y, Kurokawa M, Watanabe T, Muramatsu S, Shibasaki R (2013) Spatiotemporal route estimation consistent with human mobility using cellular network data. In: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference On. IEEE, San Diego. pp 267–272

Kishore N, Marqués D, Mahmud A, Kiang MV, Rodriguez I, Fuller A, Ebner P, Sorensen C, Racy F, Lemery J, et al. (2018) Mortality in puerto rico after hurricane maria. New Engl J Med 379(2):162–170

Klammler H, Rao P, Hatfield K (2018) Modeling dynamic resilience in coupled technological-social systems subjected to stochastic disturbance regimes. Environ Syst Decis 38(1):140–159

Kousky C (2014) Informing climate adaptation: A review of the economic costs of natural disasters. Energy Econ 46:576–592

Kull D, Gitay H, Bettencourt S, Reid R, Simpson A, McCall K (2016) Building resilience: World bank group experience in climate and disaster resilient development. In: Climate Change Adaptation Strategies–An Upstream-downstream Perspective. Springer International Publishing, Switzerland. pp 255–270

Li W, Cai X (2004) Statistical analysis of airport network of china. Phys Rev E 69(4):046106

Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 haiti earthquake. Proc Nat Acad Sci 109(29):11576–11581

McCaughey JW, Daly P, Mundir I, Mahdi S, Patt A (2018) Socio-economic consequences of post-disaster reconstruction in hazard-exposed areas. Nat Sustain 1(1):38

Myers CA, Slack T, Singelmann J (2008) Social vulnerability and migration in the wake of disaster: the case of hurricanes katrina and rita. Popul Environ 29(6):271–291

Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: Generalizing degree and shortest paths. Soc Netw 32(3):245–251

Pasch RJ, Penny AB, Berg R (2018) National hurricane center tropical cyclone report: Hurricane maria:1–48. https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf

Piguet E, Pécoud A, De Guchteneire P (2011) Migration and climate change: An overview. Refug Surv Quart 30(3):1–23

Sadri AM, Ukkusuri SV, Lee S, Clawson R, Aldrich D, Nelson MS, Seipel J, Kelly D (2018) The role of social capital, personal networks, and emergency responders in post-disaster recovery and resilience: a study of rural communities in indiana. Nat Hazards 90(3):1377–1406

Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. J Royal Soc Int 10(84):20130246

Song X, Zhang Q, Sekimoto Y, Shibasaki R (2014) Prediction of human emergency behavior and their mobility following large-scale disaster. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York. pp 5–14

United States Department of Energy IS, Restoration E (2017) Hurricanes Nate, Maria, Irma, and Harvey Situation Reports. https://www.energy.gov/ceser/downloads/hurricanes-nate-maria-irma-and-harvey-situation-reports. Accessed 17 Sept 2018

Wang Q, Taylor JE (2014) Quantifying human mobility perturbation and resilience in hurricane sandy. PLoS ONE 9(11):112608

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):0147299

Ward MD, Gleditsch KS (2018) Spatial Regression Models, vol. 155. Sage Publications, United States of America

Wardrop N, Jochem W, Bird T, Chamberlain H, Clarke D, Kerr D, Bengtsson L, Juran S, Seaman V, Tatem A (2018) Spatially disaggregated population estimates in the absence of national population and housing census data. Proc Nat Acad Sci 115(14):3529–3537

Yabe T, Tsubouchi K, Fujiwara N, Sekimoto Y, Ukkusuri SV (2019) Universal population recovery patterns after disasters. In Rev. arXiv preprint arXiv:1905.01804

Yabe T, Tsubouchi K, Sudo A, Sekimoto Y (2016) A framework for evacuation hotspot detection after large scale disasters using location data from smartphones: case study of kumamoto earthquake. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, California. p 44

Zhong C, Arisona SM, Huang X, Batty M, Schmitt G (2014) Detecting the dynamics of urban structure through spatial network analysis. Int J Geograph Inf Sci 28(11):2178–2199

Publisher's Note

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