International Journal of Disaster Risk Reduction 108 (2024) 104537

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Hazard exposure heterophily in socio-spatial networks contributes to post-disaster recovery in low-income populations

Xiangpeng Li *, Yuqin Jiang, Ali Mostafavi

UrbanResilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, 77843, USA

ARTICLE INFO

Keywords:
Human mobility
Socio-spatial networks
Hazard-exposure heterophily
Socio-demographic characteristics
Social connectedness
Community resilience

ABSTRACT

Despite recognition of the influence of socio-spatial networks on disaster recovery, limited data-driven evidence exists regarding the nature of this relationship. To address this knowledge gap, this study investigated the duration of the human mobility recovery process and its relationship with hazard-exposure heterophily (as a measure for the resourcefulness of social ties) and various other socio-demographic characteristics. We applied the concept of hazard-exposure heterophily to the possibility of resource sharing, through social-spatial networks, between communities with different hazard exposures. We used human movement patterns to measure recovery duration. To understand the disparities during recovery, we examined the correlations between recovery duration and the social demographic characteristics of race and income caused by Hurricane Harvey which made landfall in Harris County, Texas, on August 25, 2017. We discerned high/low hazard-exposure heterophily through the use of a threshold-classification method, and we got four clusters according to the flood-inundation percentage and social connectedness values. Then we applied correlation analysis to analyze the relations between hazard-exposure heterophily/income and recovery weeks among different income groups. The results revealed that hazard-exposure heterophily considerably accelerates human mobility recovery in low-income areas, as the hazard-exposure heterophily allows for the increased exchange of knowledge and resources between members of diverse groups, ultimately accelerating the recovery process. The results of this study could benefit local agents to better allocate recovery resources.

1. Introduction

Post-disaster human mobility recovery metrics are, in essence, virtual timestamps from which the length of time from the onset of a natural hazard until the community returns to its pre-disruption state can be measured [1,2]. By analyzing the large-scale mobility data, researchers can infer the length of time for a community to return to its pre-disruption state following a natural hazard from the mobility patterns of a population [3]. Many factors influence a community's ability to return to pre-disaster status, for example, income [4] and social disparities [5]. Hong et al. [5] found that low-income individuals were disparately affected by Hurricane Harvey, and they are often the least able of any income group to evacuate to safer areas. Social networks provide both communication avenues between users and, in the aggregate, proxy measures of disaster intensity. Most disaster management studies examined how social connections are related to people's behaviors during the disaster, such as decision-making (Martin et al., 2017; [6,7,8]) and requests for help. For example, during Hurricane Harvey, Hurricane Irma, and Hurricane Sandy, persons experiencing flooding/power outages requested help via Twitter [9–11]. In addition, studies have found that social media community groups can affect people's reactions to

* Corresponding author.
E-mail addresses: xplli@tamu.edu (X. Li), yuqinjiang@txstate.edu (Y. Jiang), amostafavi@civil.tamu.edu (A. Mostafavi).

https://doi.org/10.1016/j.ijdrr.2024.104537
Received 17 November 2023; Received in revised form 20 April 2024; Accepted 6 May 2024
Available online 9 May 2024

an upcoming hazard (Shook and Turner, 2016; [6,8]). Social media platforms, such as Facebook and X (formerly known as Twitter), can play a vital role in disaster management by providing platforms for affected people to build connections across space [12].

Most post-disaster recovery studies, however, measured each spatial unit as an isolated recovery event. In other words, such measurements considered the recovery progress of a location as an independent process. Such characterization ignored the socio-spatial networks that would influence disaster recovery through social connections and resource sharing across areas. While the disaster literature recognizes the importance of social connections in post-disaster recovery, there is a dearth of data-driven characterization of the nature of the relationship between the properties of socio-spatial networks embedded in the community and the length of post-disaster recovery. Currently, there is a limited understanding of the underlying mechanisms associated with the length of post-disaster recovery from the perspective of socio-spatial networks. To tackle this issue, we applied hazard-exposure heterophily and socio-demographic features to capture network-based measurements to determine whether these factors substantially impact human mobility recovery duration in communities.

The general term of heterophily is defined as strong connections between a pair of communities with various characteristics (Liu & Mostafavi, n.d.; [13,8]). Hazard-exposure heterophily captures the extent to which social ties would be resourceful to both communities when a disaster occurs (Liu & Mostafavi, 2022; [13,8]). The main premise of the concept of hazard exposure heterophily is that if two areas with strong social ties are similarly impacted by a disaster, social ties would not be a resource to either. On the other hand, if the hazard exposure of two areas is dissimilar, the social ties would be resourceful with people in low-hazard areas helping those in high-hazard areas to expedite the recovery of the highly exposed community. In this study, we measured the strength of social connections between two communities using Facebook (Meta) Social Connectedness Index data and accordingly quantified the resourceful tie rate based on the concept of hazard-exposure heterophily.

Although research on disasters acknowledges the vital role of social relationships and links in recovery after a disaster, there is a notable lack of empirical analysis exploring the connection between the features of community-based socio-spatial networks and the pace of post-disaster recuperation. To bridge this gap, this study examined the extent of resourceful ties using the concept of hazard-exposure heterophily and evaluated its relationship with post-disaster recovery duration using data from the 2017 Hurricane Harvey. Specifically, this study asked this research question: To what extent does the relationship between hazard-exposure heterophily and the length of time for human mobility activities to resume pre-disruption status vary for different races or income backgrounds? To answer this research question, we used social links from Facebook (now known as Meta) and flood inundation data to calculate hazard-exposure heterophily and then examined the relationship with location-based human mobility data in Harris County, Texas, after Hurricane Harvey. Our central hypothesis is that a greater degree of hazard-exposure heterophily can significantly shorten the duration of human mobility recovery. Statistical tests were conducted to learn if significant correlations exist between interactions among communities with different socio-demographic features that affect human mobility recovery. This study used Zip Code Tabulation Areas (ZCTA) as the spatial units within Harris County and investigated their recovery durations based on hazard-exposure heterophily and socio-demographic categories (race and income).

2. Background

The recovery process following a disaster can be categorized either as short-term or long-term [14,15]. Long-term recovery is characterized by such factors as the economic viability of businesses. Short-term recovery, as measured by human mobility patterns, refers to the period of time starting immediately after a disaster impact until the time local residents resume pre-disaster normal lifestyle patterns, the assumption being that resuming human mobility patterns indicates that essential infrastructures have been repaired (Coleman et al., 2021). The ability of people to move around during a disaster is often limited due to road and bridge damage making it difficult for people to get necessary supplies or to evacuate the area [2,13,8,16].

One method to quantitatively measure post-recovery progress is to compare post-disaster patterns of visitation to points of interest (POIs) with pre-disaster activity to determine the number of days before visits returned to pre-disaster visitation patterns [17]. Analysis of POI visitation over time provides valuable insights into how lifestyles are affected by changes in the built environment and the status of businesses [8,17,18]. Thus, patterns of visits to POIs can be a proxy measurement of recovery progress [1,8,17,18,19]. By tracking visits to POIs, researchers can measure recovery duration at distinct locations and thus identify places that may need increased external resources during recovery. In addition, studies also show that the recovery duration varies with business type, and that essential businesses, such as gas stations and grocery stores, recovered more quickly than non-essential businesses, such as entertainment [1,8,18,20]. Following the framework of existing studies, we adopted visits to POIs as our post-disaster recovery measurement. We measured the duration of impact in terms of recovery duration, the end of which is indicated by the return of human activity to pre-disaster levels. Analyzing such recovery time after hazards can better identify communities that may need external help, and hence can improve resource allocation in recovery [1,21].

Aggregated and de-identified social media data sources reveal social connections both qualitatively and quantitatively. Due to its timeliness, social media data have been widely utilized in disaster management [9,22,23]. By tracking social media posts and conversations, researchers can gain insights into how people respond to disasters. For example, Facebook social links were used to study heterophily in Winter Storm Uri [13,8]. In the aftermath of a disaster, social media allows people to connect with others across space and time, enhancing their ability to cope with bad situations [24]. In general, communities with strong social connections with their neighbors are more likely to receive help from other communities, which accelerates a community's recovery [25]. Therefore, building strong social connections with other communities can be especially helpful during disasters, when a community needs to rely on others for sharing information, providing support, and connecting people in need with services [13,8]. For example, Lee et al. (2022) found that hazard-exposure heterophily predicts that communities in high-hazard-exposure areas seek support from well-connected low-impact areas. Hazard-exposure heterophily refers to the phenomenon in which communities with varied degrees of hazard exposure allocate their resources to help other communities during post-disaster recovery. In disaster management, hazard-exposure heterophily is a vital indicator for assessing community recovery duration, as higher hazard-exposure heterophily communities are able to receive more assistance after a disaster. Besides the hazard-exposure heterophily, other socio-demographic characteristics were associated with a community's post-disaster recovery duration, such as race and income (Liu & Mostafavi, 2022; [26,27]). Social disparity is also found to affect post-disaster recovery [28]. Income is positively correlated with the recovery duration [21,29]. In this study, we examine human mobility recovery duration and its relationships with hazard-exposure heterophily and socio-demographic characteristics.

3. Study area and data materials

3.1. Study area

During Hurricane Harvey, wind gusts reaching 150 miles per hour resulted in localized destruction and intense rainfall. To mitigate the risk of catastrophic dam failure, controlled water releases from the Addicks and Barker Reservoirs were carefully executed. Although this action led to additional flooding in certain areas, it was a necessary measure to avert more severe consequences associated with potential dam breaches. This strategic decision was vital to prevent an even greater scale of flooding and infrastructure damage across the impacted regions [30,31]. Although its center passed south of the Houston metropolitan area, heavy rains fell near a stationary front on the north and east sides of the storm, causing flooding in northeast Houston [32]. More than 100 cm of precipitation fell in the heavily populated Houston region of Texas, causing severe flooding and damage to onshore industrial facilities [33]. Harris County, core of the Houston metropolitan area, had a population of 4.7 million in 2017, making it the most populous county in Texas (US Census Bureau, 2022). Hurricane Harvey caused a significant displacement of low-income rental housing residents, resulting in an 18 % increase in homelessness one year after Hurricane Harvey [34]. Harvey caused residential structural damage of more than $77.2 million, while residential damage was valued at more than $36.9 million in Houston (Hicks & Burton, 2017). This study uses Harris County during post-Hurricane Harvey recovery as a case study.

3.2. Data description

The datasets used in this study are shown in Table 1 and Fig. 1. Our research is analyzed based on ZCTA level, and the ZCTA dataset is obtained from the U.S. Census Bureau [35], contains generalized representations of United States Postal Service Zip Code

Table 1
Data source description.

Data	Source
ZIP Code Tabulation Areas map	U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing
Flood inundation	Federal Emergency Management Agency maps
Socio-demographic characteristics	U.S. Census Bureau table data
Points of interest (POI)	Safegraph, Inc.
Meta Social Connectedness Index	Data from Meta (formerly Facebook)

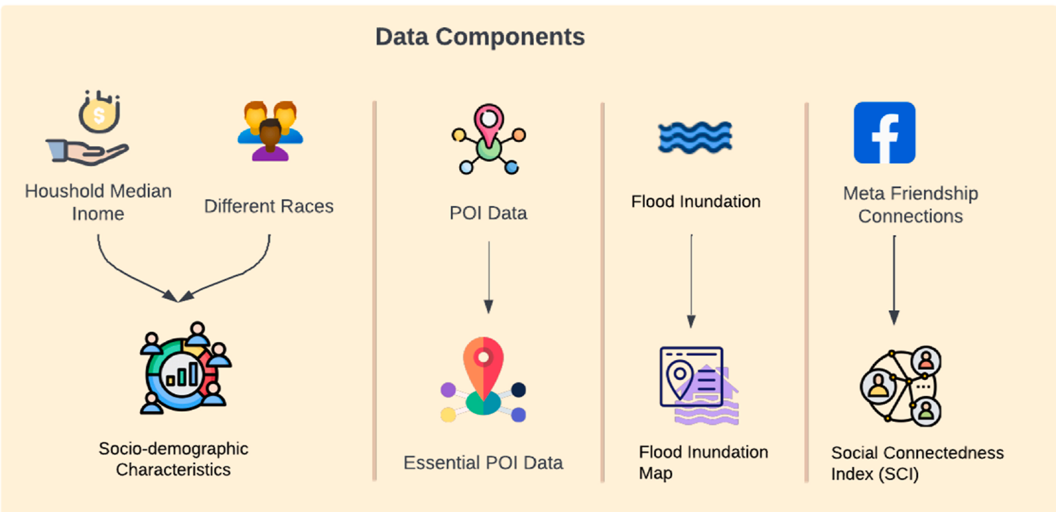

Fig. 1. Data Components of the Study. Four datasets were used for further analysis: 1. Socio-demographic characteristics, encompassing household median income and various racial demographics. 2. Essential points-of-interest (POI) data, derived from POI datasets. 3. Flood inundation map, created by aggregating flood inundation data onto a ZIP Code Tabulation Area (ZCTA) map. 4. Social Connectedness Index, calculated based on Meta friendship connections.

service regions [35]. Each polygon in the map represents a Zip Code unit, the spatial unit used in this study. According to this dataset, Harris County contains 144 ZCTAs [35].

We merged flood inundation, POIs, and socio-demographic characteristics into the ZCTA map. The socio-demographic characteristics data (US Census Bureau, 2022) contained racial composition and median income spatial units in 2017. Meta Social connectedness Index is from meta datasets.

3.2.1. Flood inundation map

Flood inundation data, provided by the Federal Emergency Management Agency, was downloaded from Hydroshare (FEMA, 2020). This dataset provides gridded depth information at a resolution of 3-m by 3-m. FEMA provides information on the severity of floods based on the effect of depth and velocity [36], as well as flood-inundation maps based on high-water marks. Flood-inundation maps delineate the areas affected by flooding and their maximum depths during a natural hazard event (USGS, 2022). Fig. 2 is the flood inundation map overlain with ZCTA boundaries.

3.2.2. Socio-demographic characteristics

3.2.2.1. Household median income. Fig. 3 displays the distribution of median household income across Harris County, Texas. It is apparent that communities situated to the north and southeast of the downtown area are predominantly lower-income, whereas regions to the east of downtown, as well as those bordering the county's perimeter, are typically associated with higher income levels.

3.2.2.2. Race composition. We considered race as another component of socio-demographic characteristics. We considered the percentage of White and Black populations in each ZCTA area as one feature of socio-demographic characteristics.

3.2.3. Essential POI data

The human mobility recovery time is considered an indicator for purposes of post-disaster human mobility recovery metrics, which are drawn from POI datasets. This POI dataset was obtained from Safegraph, a database company that provides detailed business information about geographic coordinates, street addresses, and the North American Industrial Classification System business classification code of the business.

Following the methods established by existing studies [17], we categorized POIs as essential or non-essential for recovery. In general, essential POIs provide services without which local people's normal life would be disrupted, including sixteen types of businesses

Fig. 2. Flood inundation map of Hurricane Harvey with ZCTA boundaries in Harris County, Texas.

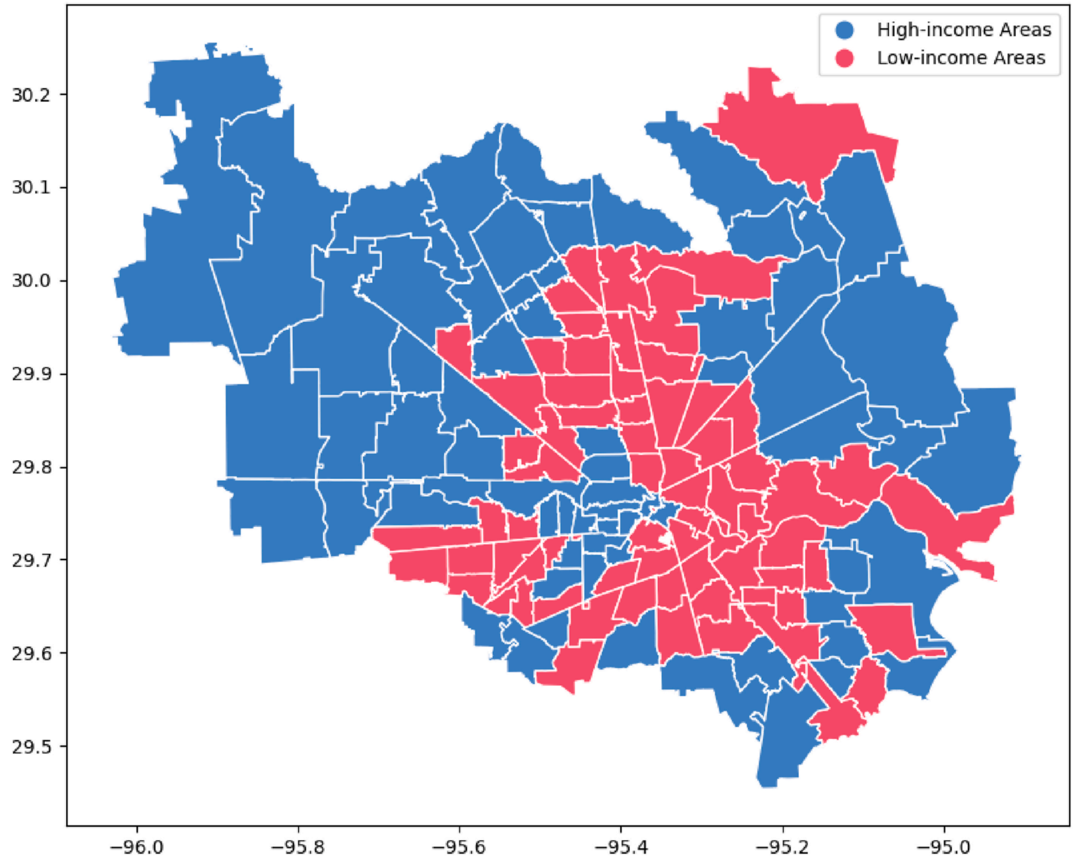

Fig. 3. Income Level distribution based on median income in Harris County, Texas. Low- and high-income regions have been delineated according to the countywide median household income, which stands at $55,855. High-income zones are represented in blue, while red indicates areas with lower income. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

for emergency preparedness, emergency response, lifestyle and well-being, and recovery activity (Fig. 4). Recovery duration was determined by the length of time visit patterns to these essential POIs returned to pre-disaster levels at the ZCTA level in Harris County after Hurricane Harvey.

3.2.4. Social Connectedness Index (SCI)

The Facebook (now known as Meta) Social Connectedness Index (SCI) is based on social connections on Facebook [37]. The connectivity between two locations was measured by the intensity of local users' activities on Facebook. The location information of each

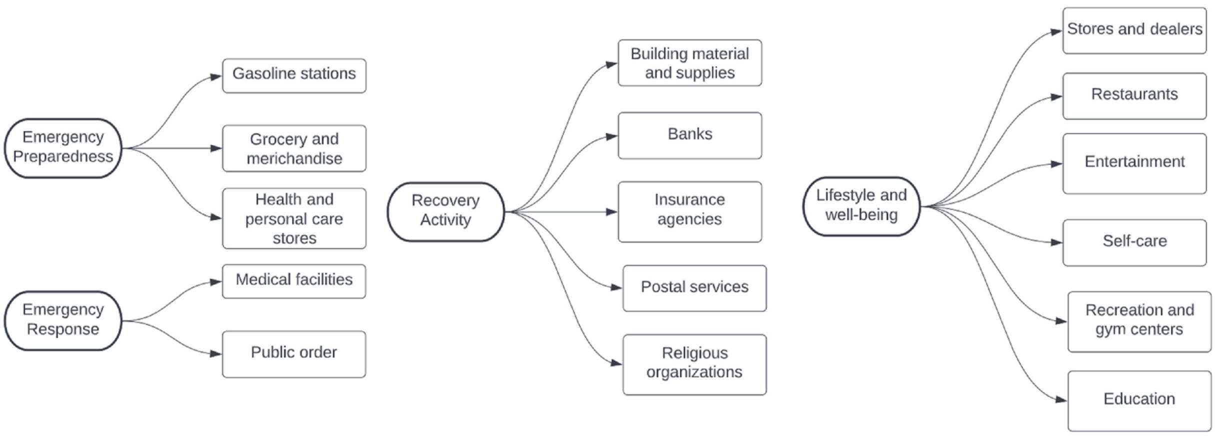

Fig. 4. Four types of sixteen essential POI categories.

user was based on their Facebook profile information, activities, location information, and device and connection data [37]. Equation (1) shows the calculation for the SCI:

$$Social\_Connectedness_{i,j} = \frac{FB_{Connections\,i,j}}{FB_{User\,i} \times FB_{User\,j}} \tag{1}$$

where i and j are two regions. FB_User_i and FB_User_j reflect the numbers of Facebook users in locations i and j, respectively. $FB\_Connections_{i,j}$ denotes the number of Facebook friendship ties between i and j. $Social\_Connectedness_{i,j}$ evaluates the probability that a user in location i and a user in location j are connected in Facebook activities. In this study, we adapted this measurement and applied it to measure social connectedness between two ZCATs.

4. Methodology

4.1. Research approach

Fig. 5 presents a synopsis of the research methodology in this investigation. Initially, we handled four distinct datasets: socio-demographic traits, critical POI data, flood inundation mapping, and SCI (Social Connectedness Index). In the data processing phase, we consolidated the essential POI information to deduce the recovery time of human mobility. Calculation of the proportion of flooded areas was determined at the Zip Code Tabulation Area level. The rate of resourceful connections was derived from the SCI. For the analytical phase, we used threshold-based classification using the flood inundation proportion and resourceful tie rate to categorize the ZCTAs into four disparate strata of hazard-exposure heterophily. Ultimately, statistical approaches were employed to

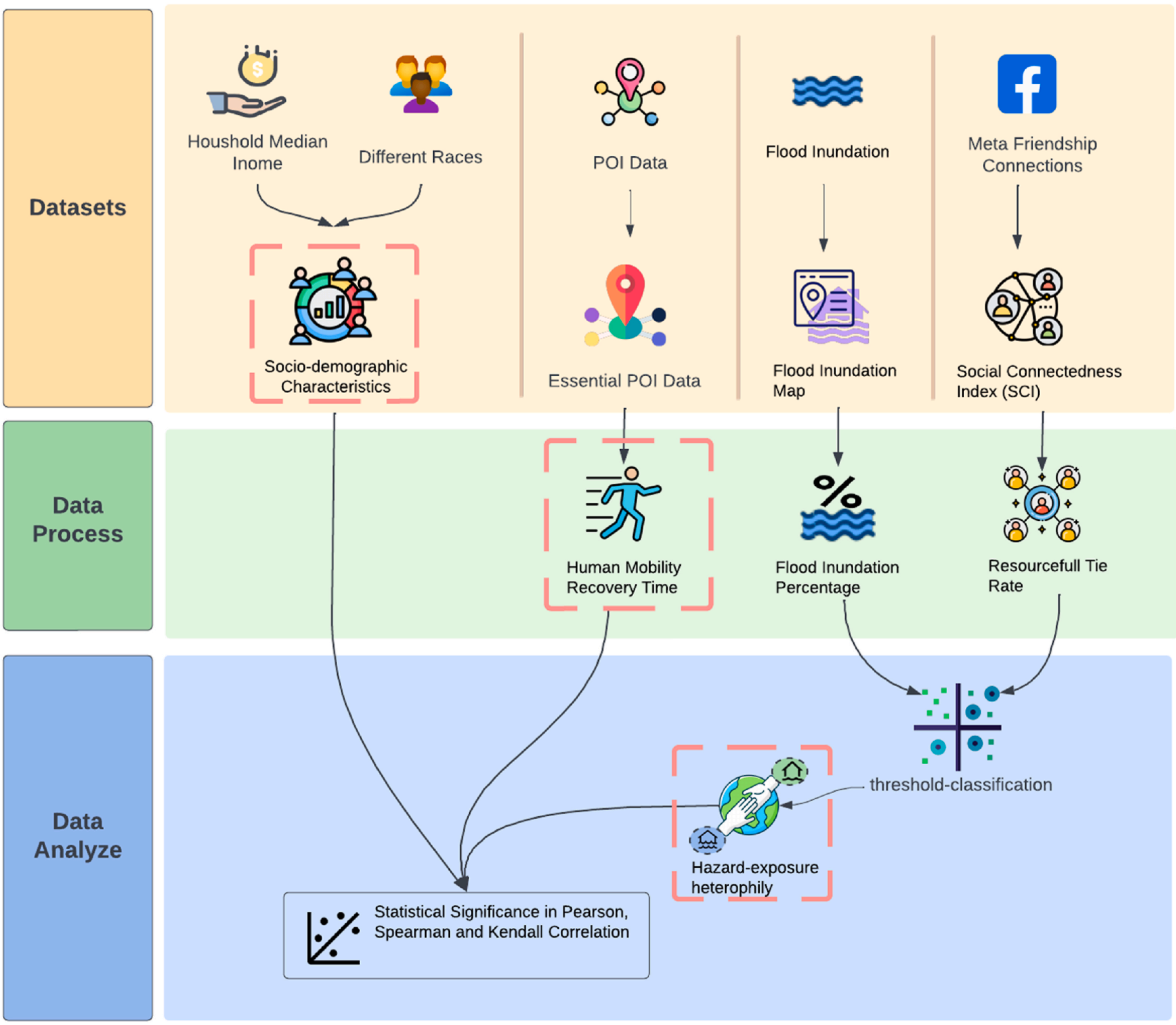

Fig. 5. Overview of the research approach.

probe the dynamics among hazard-exposure heterophily, timelines for the resumption of human mobility, and socio-demographic factors. This examination aims to reveal the influence of hazard-exposure heterophily and socio-demographic features along the length of human mobility restoration, using trend line assessment and analysis of statistical significance.

4.2. Flood inundation percentage

We used ArcGIS Pro to calculate the flood inundation area percentage of each Harris County ZCTA. The histogram plot shows the flood inundation percentage (Fig. 6) to be between 0 % and 45 %, with a median value 4.7 %. We grouped ZCTAs into high and low flood inundation exposure based on their rank above or below this median value. Fig. 7 illustrates the geographical distribution of ZCTAs. Most high flood inundation areas (blue) are located in eastern Harris County due to its proximity to the coast and its complicated networks of rivers and wetlands. Most low flood inundation areas (green) are located in the western part of Harris County, located inland and distant from the coast.

4.3. Human mobility recovery time

We incorporated essential POI data as defined by Podesta et al. [17] and followed the methodology of Coleman et al. (2023) to calculate daily visits from each home Census Block Group (CBG) to various POIs, as illustrated in Fig. 8. Our human mobility data was sourced from Spectus, which specializes in providing highly accurate and frequently updated location data. To categorize each POI, we used SafeGraph's dataset, which aligns with the North American Industry Classification System (NAICS) codes. This dataset includes detailed spatial coordinates and pertinent information about POIs, such as brand names, operating hours, and exact geographical details. Building upon SafeGraph's foundational data, we further refined the spatial definition of each POI using Microsoft's open-building footprint dataset, which offers precise polygon geometries for buildings in the Houston area derived from satellite imagery.

Recovery time is measured based on the number of weeks needed to return to the baseline levels of visits to essential POIs [1,17]. We defined August 1 through August 21, 2017, as the pre-hurricane period. The baselines of the 16 categories were calculated as the seven-day moving average of the number of daily visits to several categories of POIs during the pre-hurricane period. This baseline provides a standard reference point for pre-hazard POI visitation patterns. After the disaster, the trend is a slow resumption of pre-disaster patterns. The systemic point of inflection shown in Fig. 9 is the point of lowest percentage change of POIs visits [38]. The percentage change reveals the difference between the pre- and post-hurricane daily visits for each POI category. In Equation (2), the daily value represents the observed daily visits to the POI. The base value is the standard reference value for normal activity in the pre-hazard period. The change percentage then represents changes in the number of daily visits for each category. We used a seven-day rolling average so that the average percent change was be taken for seven-day periods:

$$7-\text{day } Rolling_{avg} = \sum_{d=1}^{7}\left(\frac{Daily\ value - Base\ value}{Base\ value}\right)\times 100\% \tag{2}$$

The resilience curves (aggregated in Fig. 9) for each POI category were calculated by plotting the seven-day rolling averages. The duration of recovery in our study dataset is the duration of impact shown in Fig. 8. We calculated the average amount of time required for each area to recover. The recovery week occurs when the seven-day rolling average reaches 90 % of the baseline values. We aggregated the POI datasets from the census block group (CBG) level to the ZCTA Level in ArcGIS Pro. If the centroid of a CBG is located inside a ZCTA, we consider this CBG as belonging to this ZCTA.

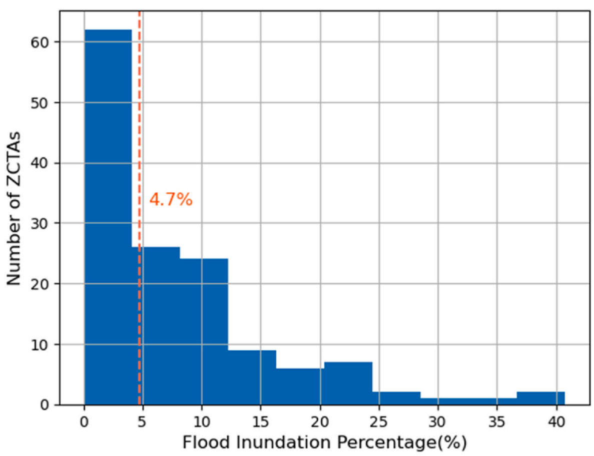

Fig. 6. Histogram plot of flood inundation percentage. The vertical red line represents the median value of flood percentage, 4.7 %. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

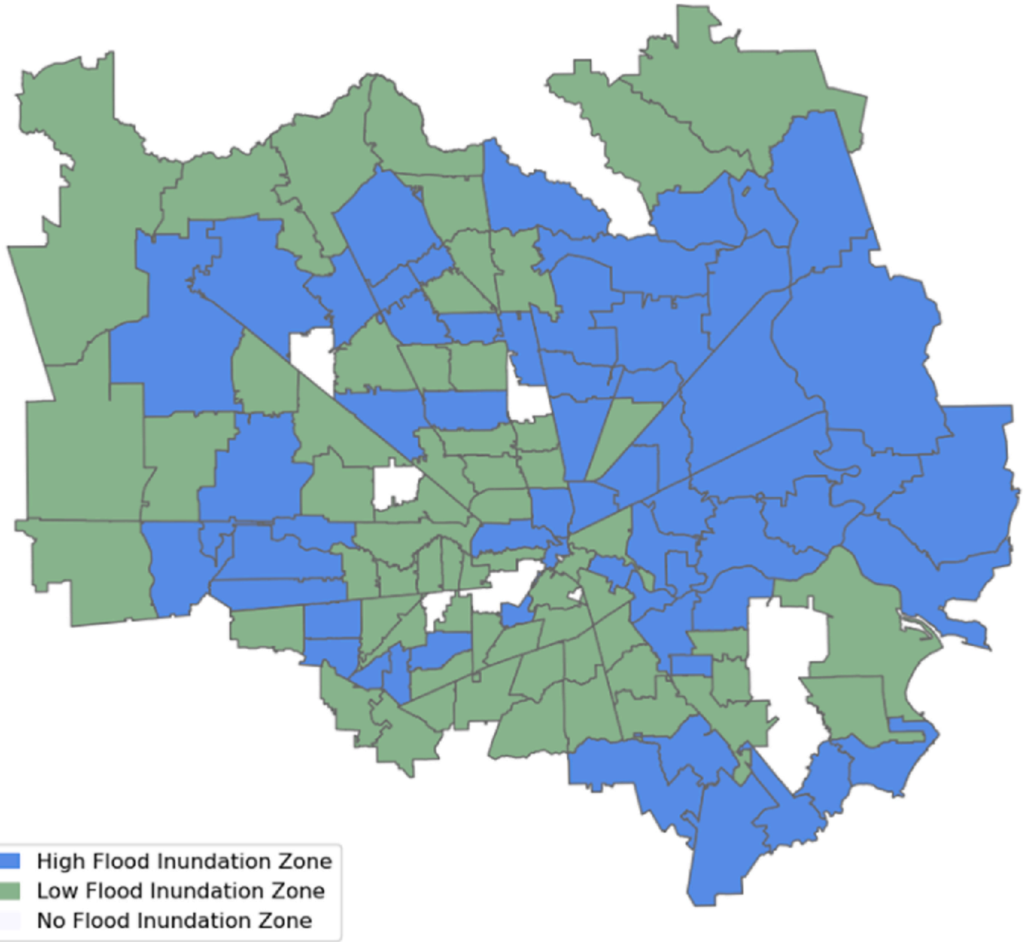

Fig. 7. High/low flood inundation map of Harris County, Texas. Blue represents high flood inundation zone, green represents low flood inundation zone, and white represents no inundation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

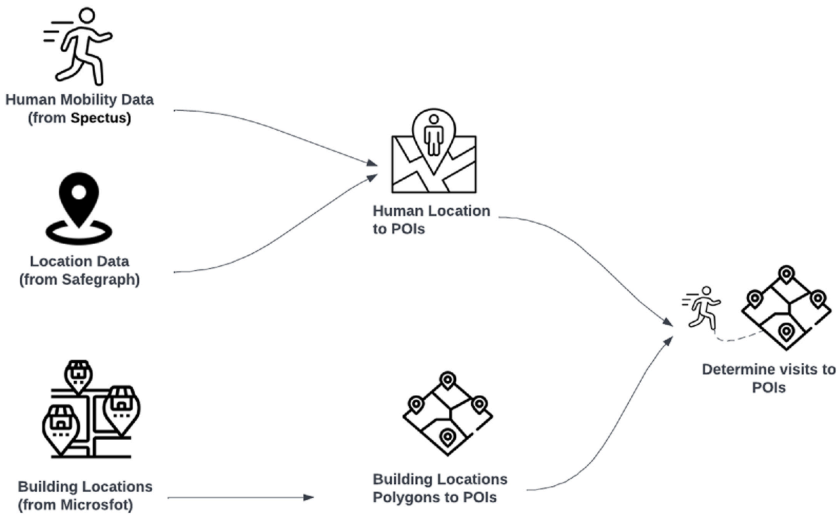

Fig. 8. Visits to POIs.

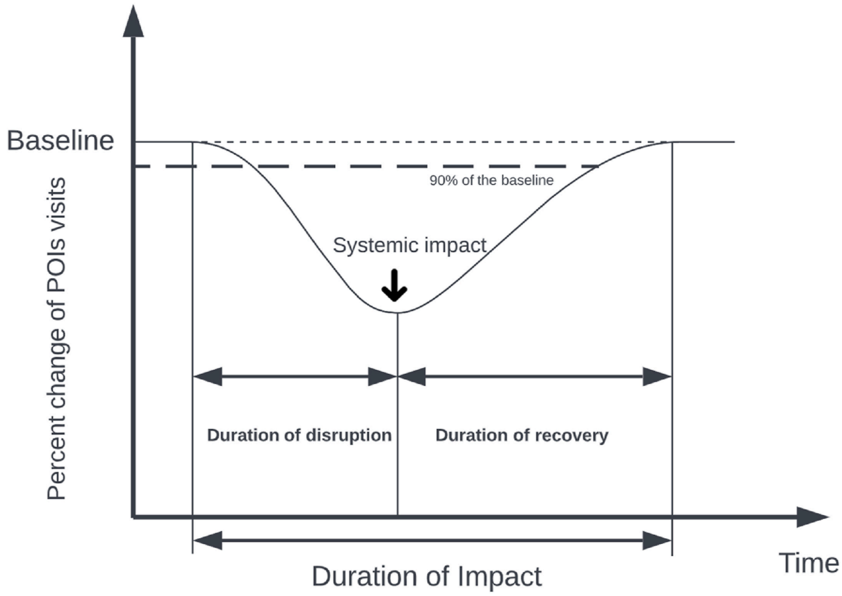

Fig. 9. Duration of impact based on the frequency of POIs visited.

4.4. Hazard-exposure heterophily calculation

In this study, hazard-exposure heterophily was calculated using the Facebook Social Connectedness Index (SCI) and flood inundation percentages. As described in Section 4.2, we grouped all the ZCTAs into high and low flood inundation zones based on the median flood inundation percentage. In Fig. 10, blue-shaded areas represent high-flood inundation areas, while green-shaded areas denote low-flood inundation areas.

The concept of hazard-exposure heterophily describes the social connections among populations residing in spatial areas with different levels of hazard exposure. Conversely, social connections among population sharing similar hazard exposure are representative of hazard exposure homophily (Liu & Mostafavi, n.d.; [13,8]). When considering flood inundation as a feature of hazard exposure within a community, heterophily is found between high-flood inundation communities and low-flood inundation areas, while homophily is found between pairs of high-flood inundation or low-flood inundation areas. When strong social ties exist between areas with varying levels of flood impact, the area with lower exposure to flood can effectively offer help to areas in high-flood zones. It is understood that those in regions with less exposure can offer prompt support for recovery and play a significant role in the sustained reconstruction of more heavily affected communities.

— : Heterophily: Social ties between dissimilar hazard-exposure areas
— : Homophily: Social ties between similar hazard-exposure areas

The SCI quantifies the strength of social connection between two Meta users [37]; however, this study necessitates the aggregation of individual connection intensity to the ZCTA level. To achieve this, we adopted the concept of "resourceful tie rate," as introduced by Liu and Mostafavi [39], to capture the social connectedness strength between two communities in a given social-spatial network. The resourceful tie rate measures the extent of connections a high-hazard-exposure region could receive from low-hazard-exposure regions under the assumption that low-hazard-exposure regions can provide resources for post-disaster recovery. Each ZCTA is defined as a node in the social-spatial network, with the resourceful tie rate acting as a node attribute. Let ρi (i = 1, 2, …, n) be the resourceful tie rate for the *i*th ZCTA, and ρ_i is calculated by Equation (3):

$$\rho_i = \frac{\sum_{j \epsilon L} Social_{Connectedness\, i,j}}{\sum_{j=1}^{n} Social_{Connectedness\, i,j}} \tag{3}$$

where ρ_i represents the resourceful tie rate of ZCTA i, quantifying the extent to which area i can receive resources from low-hazard-exposure areas (L). As shown in Eq. (3), ρ_i is calculated as the ratio of social connectedness originating from low-flood-inundation areas (L) in relation to the total social connectedness (n). This ratio shows the degree to which a ZCTA can draw resources from low-hazard-exposure areas.

The resourceful tie rate can function as an indicator of community resilience within socio-spatial networks, as the connection of high-hazard-exposure areas with low-hazard-exposure areas enables communities to access external resources during post-disaster recovery. We first calculated the resourceful tie rate flood inundation levels then derived the hazard exposure heterophily metric. High-hazard-exposure heterophily is present when an area experiences significant flood inundation and a high resourceful tie rate, indicating its ability to receive resources from low-flood-inundation areas. The ρ_i value represents the extent of hazard-exposure heterophily.

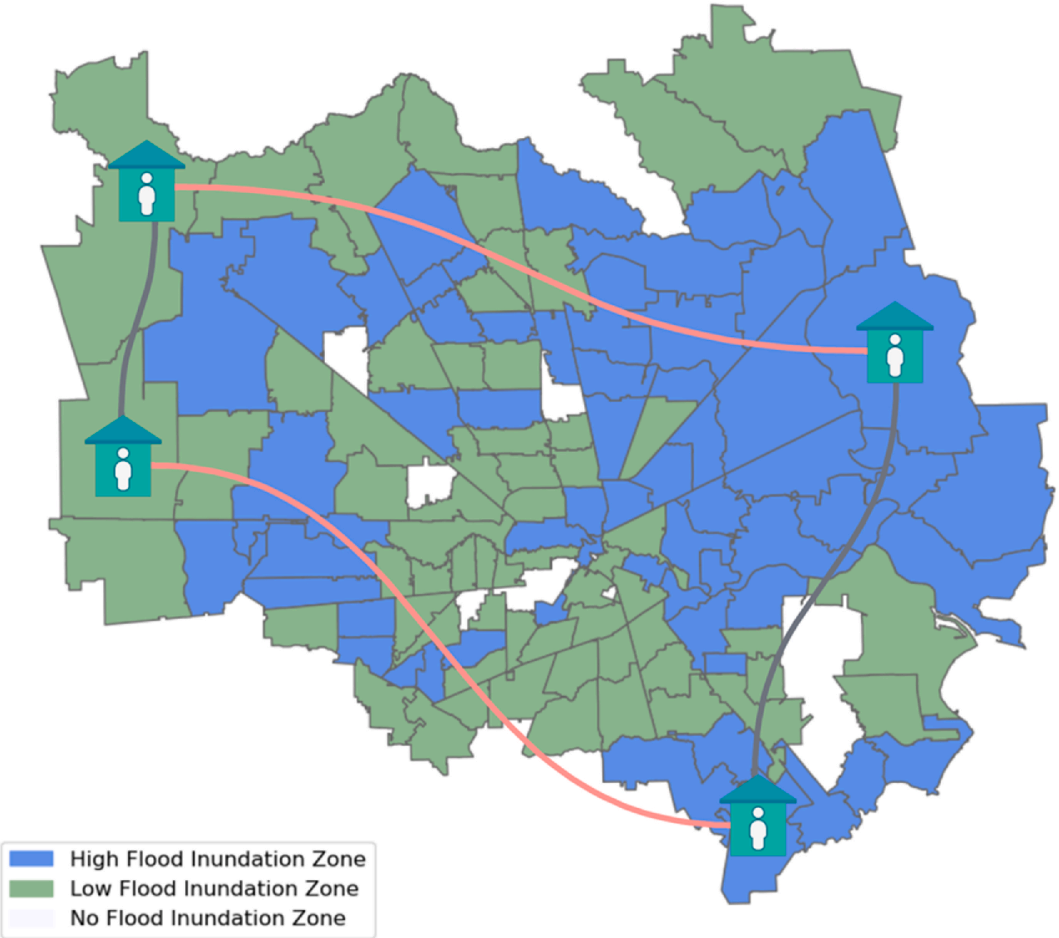

Fig. 10. Illustration of hazard-exposure heterophily.

For ZCTAs affected by flood inundation, a higher ρi implies a greater degree of hazard-exposure heterophily within a region, signifying capacity to establish connections with individuals in other areas and obtain the information and assistance necessary during natural hazard events.

4.5. Statistical tests analysis

We ran Pearson, Spearman, and Kendall statistical tests to investigate the relationships among income, heterophily, and recovery weeks durations. Pearson's correlation coefficient [40], denoted r, describes the direction and degree of the linear relationship between two continuous variables (Chok, n.d.; [41]). For example, a positive correlation coefficient means a linearly increasing relationship relates to the variables, and a negative correlation indicates a linearly decreasing relationship between variables [41]. Spearman's rank-order correlation coefficient [42], denoted ρ, is a rank-based version of Pearson's correlation coefficient. This non-parametric measure of correlation evaluates how well a monotonic function can describe a relationship between two different variables [41]. Like Spearman's rank-order correlation coefficient, Kendall's tau correlation coefficient, denoted τ, aims to capture the relation between two ordinal variables [41,43,44]. In all three methods, the coefficients range from −1 to +1. The closer the coefficient is to 0, the weaker the relationship is between the two variables. Pearson's correlation coefficient assesses the linear relationship between two variables; Kendall's and Spearman's correlation coefficients evaluate the monotonic relationship. In contrast to linear connections, in which two variables move together at a constant rate, monotonic relationships quantify the likelihood that two variables will move in the same direction but not necessarily at the same rate.

5. Results and discussion

5.1. Recovery measurement based on trips to POIs

Fig. 11 shows the recovery duration in each ZCTA area after Hurricane Harvey in Harris County, Texas. For a ZCTA, the darker the hue of the area on the map, the longer the area took to recover. This map reveals that the northern part of Harris County has the longest recovery duration. In contrast, the downtown area and the west side of Harris County had the shortest duration of recovery.

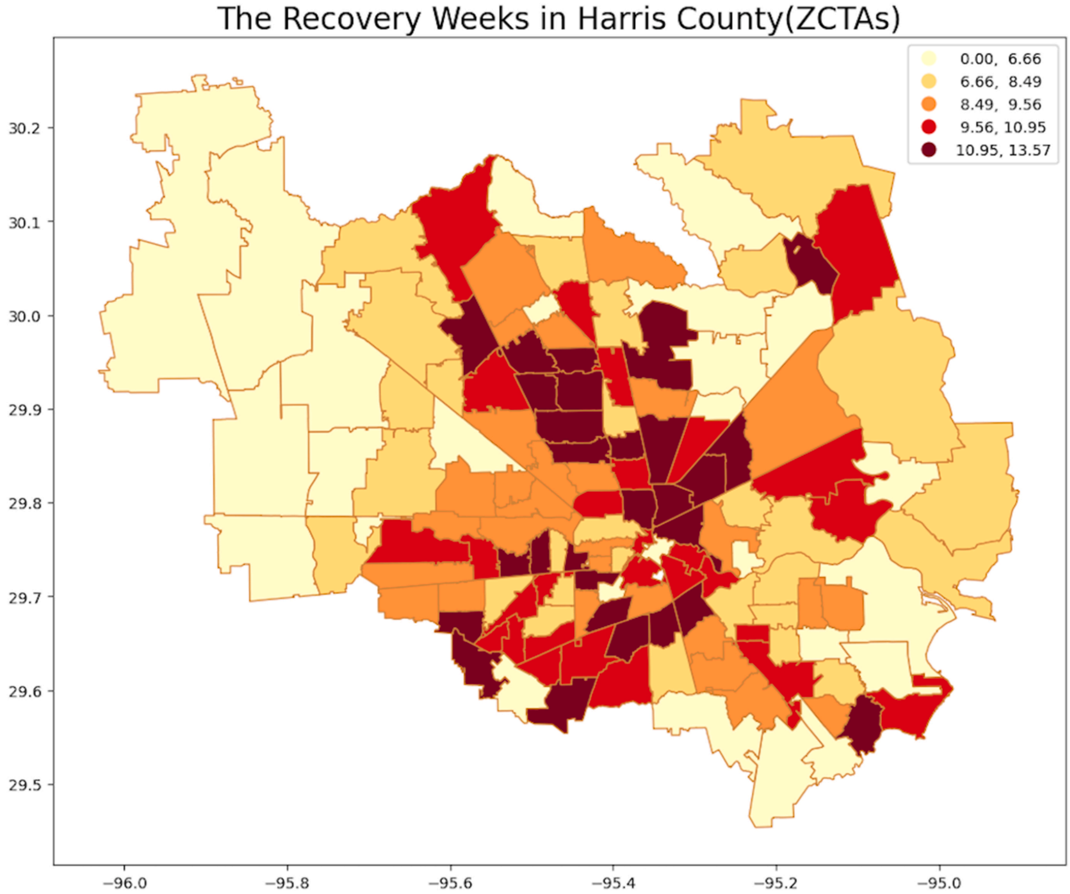

Fig. 11. The recovery weeks in each ZCTA area after Hurricane Harvey in Harris County, Texas. The darker the color of the area on the map, the longer the area took to recover. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

One possible cause is that those communities are located near major riverways, which are more vulnerable to flooding, causing extensive damage and requiring a longer recovery period.

Table 2 shows the median recovery weeks in each cluster. We can see that cluster 3 has the fewest recovery weeks, which is 8.53. Cluster 1 has the most recovery weeks, which is 10.15.

5.2. Hazard-exposure heterophily

In this study, we used the threshold-classification method to group the ZCTAs into four clusters according to flood inundation percentage and resourceful tie rate using median values as the criteria (Fig. 12). The threshold is the median value of the flood inundation percentage (6.48 %) and resourceful tie rates (0.44 %). Clusters 1 and 2 are assigned to low flood inundation areas and clusters 3 and 4 are assigned to high flood inundation areas. The right panel of Fig. 12 illustrates the spatial distribution of four clusters. Most of the ZCTAs in the eastern part of Harris County are within clusters 3 and 4. The southwest part of Harris County, which is the urban center, is mainly composed of ZCTAs that fall into clusters 1 and 2.

Cluster 1 has a low flood-inundation percentage and higher resourceful tie rates, which means it is less affected by the flooding. It is strongly connected with other low-flood-inundation areas, however, meaning that its population is less likely to help high-flood-inundation communities. This suggests that cluster 1 has low hazard-exposure heterophily; in other works, cluster 1 communities were strongly connected with similarly affected groups.

Table 2
Median recovery weeks in each cluster.

Cluster	Median recovery (weeks)
1	10.15
2	9.32
3	8.53
4	9.54

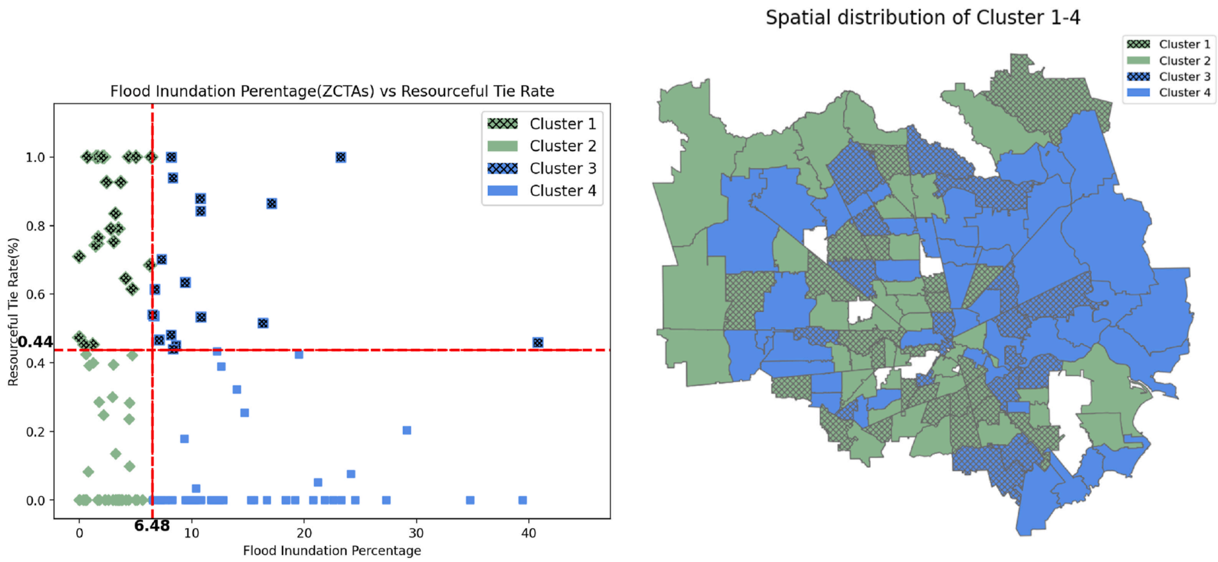

Fig. 12. Left: Four clusters based on flood-inundation and resourceful tie rates. The vertical red line is the median flood-inundation percentage; the horizontal red line is the median resourceful tie rate. Right: Spatial distribution of the four clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

In cluster 2, the flood-inundation percentage is lower than the median percentage (4.7 %), and the resourceful tie rates are also lower, meaning cluster 2 is less affected by the flooding and receives less help from low-hazard-exposure areas.

Cluster 3 and cluster 4 are higher flood-inundation percentage areas, which are higher than the median value. In cluster 3, the flood-inundation percentage is higher, and so are the resourceful tie rates. This means that communities in this cluster were heavily flooded, but they were strongly connected with low-flooded communities. Therefore, they were likely to receive external resources from low-hazard-exposure areas, giving cluster 3 communities a high heterophily.

Cluster 4 has a high flood-inundation percentage but lower resourceful tie rates, meaning it experiences high floods, but residents were less connected with low-flooded communities. This indicates that Cluster 4 is a lower hazard-exposure heterophily community.

5.3. Correlation analysis

In this research, the connection between recovery duration and hazard-exposure heterophily across different income groups was interrogated using Pearson, Spearman, and Kendall correlation analyses. Displayed in Fig. 13, the heatmaps demarcate the correlations with corresponding p-values, where blue indicates a negative relationship and red a positive one. Deeper shades connote stronger correlations.

For instance, the Pearson correlation in low-income areas reveals a negative coefficient (−0.34) between hazard-exposure heterophily and the duration of recovery weeks, implying that increased heterophily correlates with shorter recovery times. This pattern is supported by all three correlation metrics (Spearman $r = -0.31$, Pearson $\rho = -0.34$, Kendall $\tau = -0.22$), underscoring a statistically significant negative association in low-income zones. Conversely, in high-income areas, there is a notably negative correlation between median income and recovery weeks (Spearman $r = -0.49$, Pearson $\rho = -0.40$, Kendall $\tau = -0.36$), indicating that higher median incomes correspond with faster recovery. Notably, a significant positive relationship exists between the percentage of Black residents and recovery weeks, as demonstrated by the Spearman and Kendall correlations; this relationship, however, is not captured by the Pearson correlation.

P-values are denoted by asterisks: "*for* $p < 0.01$*(***), for* $p < 0.05$ *(**), and for* $p < 0.1$*(*),*" From Fig. 13, we can see that the correlation between hazard-exposure heterophily and recovery weeks in low-income areas is −0.34, with a p-value of 0.089, which means this correlation is statistically significant. While in high-income communities, hazard exposure heterophily and duration of recovery weeks have a p-value of 0.22, which does not show a statistically significant correlation. On the other hand, the correlation between median income and recovery weeks are all negative with significant P-values ($r = -0.49$, p-value $= 0.012$; $\rho = -0.4$, p-value $= 0.041$; $\tau = -0.36$, $p < 0.01$). This indicates that median income promotes human mobility recovery in high-income regions. In general, income is a factor correlated with recovery duration, which is consistent with existing literature [45]. Moreover, in low-income communities, hazard-exposure heterophily is more significantly correlated with recovery duration than that in high-income communities. One possible explanation is that people with higher incomes tend to have more resources to rebuild their lives and more access to services and support networks that can help them recover from natural hazards. High-income communities are typically better prepared for an oncoming hurricane by stocking more essential supplies. In low-income areas, however, residents may not have sufficient supplies and thus need to ask for help from external resources [5]. In summary, while income consistently influences recovery duration, the impact of hazard-exposure heterophily is more pronounced in low-income communities, potentially due to their reliance on external support and resources for post-disaster recovery.

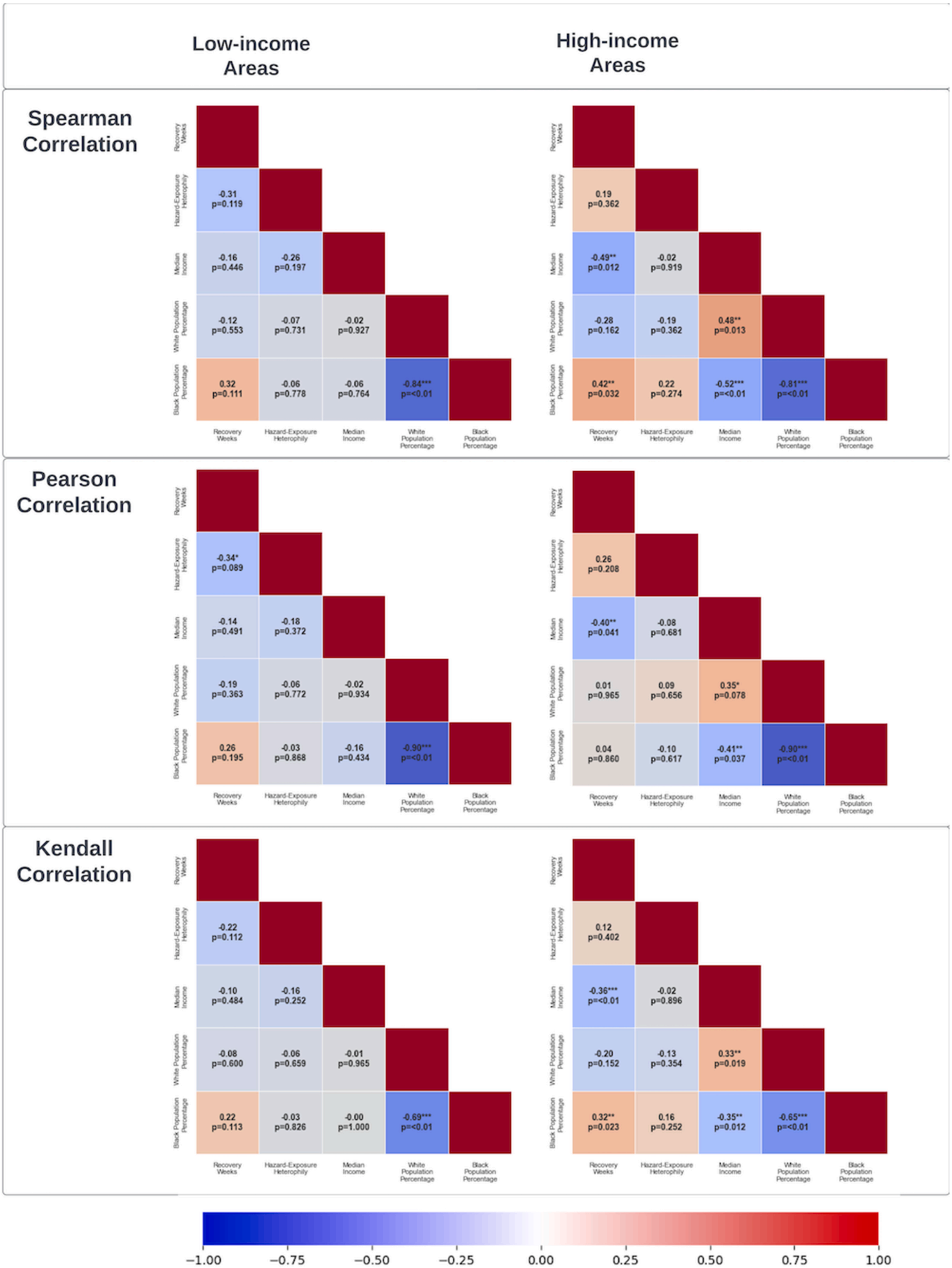
Low-income Areas
High-income Areas
Spearman Correlation
Pearson Correlation
Kendall Correlation
Recovery Weeks
Hazard-Exposure Heterophily
Median Income
White Population Percentage
Black Population Percentage
-0.31 p=0.119
-0.16 p=0.446
-0.26 p=0.197
-0.12 p=0.553
-0.07 p=0.731
-0.02 p=0.927
0.32 p=0.111
-0.06 p=0.778
-0.06 p=0.764
-0.84*** p=<0.01
0.19 p=0.362
-0.49** p=0.012
-0.02 p=0.919
-0.28 p=0.162
-0.19 p=0.362
0.48** p=0.013
0.42** p=0.032
0.22 p=0.274
-0.52*** p=<0.01
-0.81*** p=<0.01
-0.34* p=0.089
-0.14 p=0.491
-0.18 p=0.372
-0.19 p=0.363
-0.06 p=0.772
-0.02 p=0.934
0.26 p=0.195
-0.03 p=0.868
-0.16 p=0.434
-0.90*** p=<0.01
0.26 p=0.208
-0.40** p=0.041
-0.08 p=0.681
0.01 p=0.965
0.09 p=0.656
0.35* p=0.078
0.04 p=0.860
-0.10 p=0.617
-0.41** p=0.037
-0.90*** p=<0.01
-0.22 p=0.112
-0.10 p=0.484
-0.16 p=0.252
-0.08 p=0.600
-0.06 p=0.659
-0.01 p=0.965
0.22 p=0.113
-0.03 p=0.826
-0.00 p=1.000
-0.69*** p=<0.01
0.12 p=0.402
-0.36*** p=<0.01
-0.02 p=0.896
-0.20 p=0.152
-0.13 p=0.354
0.33** p=0.019
0.32** p=0.023
0.16 p=0.252
-0.35** p=0.012
-0.65*** p=<0.01
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Fig. 13. Correlation of features from low-income and high-income areas. We included the Recovery Weeks, Hazard-exposure Heterophily, Median Income, White Population Percentage, and Black Population Percentage as our features. Red represents a positive relationship; blue represents a negative relationship. The darker the hue, the stronger the relationship. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

5.4. Discussion

The results of our research emphasize the significance of social connectivity in the context of disaster recovery, particularly within economically disadvantaged neighborhoods. Our study reveals that the strength of social connections across different exposure levels—what we term as hazard-exposure heterophily—has a more pronounced impact on the duration of recovery in low-income areas than the high-income areas. This highlights the importance of resourceful social ties, especially in low-income areas. Low-income populations typically have limited access to resources that could help them prepare for the storm and its aftermath. Many low-income communities rely on outside assistance to rebuild and recover from the devastation caused by disasters [46]. As a result, external assistance plays a crucial role in helping poorer communities recover from disasters. From this perspective, hazard-exposure heterophily can help plan community-to-community resource allocation in post-disaster recovery.

Expanding on this insight, our analysis delineates four distinct clusters, using the threshold-based classification method, each with unique characteristics and needs that can inform specific disaster response and recovery strategies. Cluster 1 and Cluster 2 all exhibit lower flood inundation percentages. Cluster 1, however, experienced the highest number of recovery weeks, which could be due to its lowest hazard-exposure heterophily. After all, although it experiences a lower area flood-inundation percentage, it still has higher resourceful tie rates. Cluster 2 has the second shortest number of recovery weeks, and it has lower resourceful tie rates. For these communities, prevention strategies can focus on leveraging existing social networks to distribute resources and information during disasters. These communities can act as resource hubs, providing supplies, shelter, and support to more heavily impacted areas. In Cluster 3, the combination of high impact from flooding and strong external connections, as well as the shortest number of recovery weeks, suggests that reinforcing these networks and improving infrastructure could significantly reduce disaster impacts. Cluster 4, facing high floods with few resourceful ties, requires targeted efforts to build community resilience and connect with external support networks. Addressing the unique characteristics of each cluster can lead to more effective disaster response, leveraging the strengths of well-connected communities and supporting those in greater need, ultimately enhancing overall community resilience and resource distribution.

6. Conclusion

The intent of this study was to investigate the relationship between socio-spatial networks and the duration of post-disaster recovery. Recognizing the significant role social ties play in the aftermath of disasters, this research sought to fill a gap in the empirical analysis of how socio-spatial network characteristics impact recovery times with different socio-characteristics. We focused on hazard-exposure heterophily as a key feature of these networks, examining how it interacts with socio-demographic factors to influence the duration of human mobility recovery. We first used resourceful tie rate to measure the possibility of receiving external help from low-hazard exposure areas. Then we identified four clusters with different levels of hazard-exposure heterophily based on their hazard exposure and resourceful tie rate. Finally, we considered socio-demographic characteristics and analyzed how these features were correlated with human mobility recovery.

In light of the findings, this research offers significant insights into the intricate dynamics of post-disaster recovery, particularly the role of socio-spatial networks. The results revealed that hazard-exposure heterophily considerably accelerates human mobility recovery in low-income areas, as the hazard-exposure heterophily allows for the increased exchange of knowledge and resources between members of various groups, ultimately accelerating the recovery process. In addition, communities with higher median income experience significantly faster recovery.

It is worth noting that this research has some limitations. First, we considered only race and income for the socio-demographic characteristics. Other socioeconomic or demographic characteristics, however, were also found to be associated with recovery duration: age, ethnicity, and education ([47,48]; Aldrich et al., 2012). Future studies could further examine correlations with other characteristics to achieve a better understanding of the nuances of disaster recovery. Second, we used only essential POI datasets to calculate recovery weeks. Early-warning signals detected from the essential POIs visits appeared earlier than those from non-essential POIs [19,20]. Communities can have different recovery priorities in terms of essential or non-essential businesses [8,18]. Third, our analysis was limited to connections within Harris County, potentially restricting the calculation of resources beyond the study area. Future studies could broaden the scope to assess resource tie rates more comprehensively. Fourth, we considered flood-exposure percentage to define hazard-exposure heterophily, which will possibly limit the real outcomes of a hazard. Future studies can adopt other datasets that consider other impact data. Finally, the social connection measurement was based on Facebook connections. Although social media platforms, such as Facebook, have provided good measurements for social connections, the issue of representativeness remains unsolved [6,8,49,]. Future research could focus more on how to leverage multiple data sources to achieve a better population sample for social connection measurements.

In summary, this study presents a nuanced understanding of the interplay between socio-spatial networks and recovery duration, offering valuable insights for disaster management, especially regarding the optimization of recovery strategies in low-income areas.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

CRediT authorship contribution statement

Xiangpeng Li: Writing – original draft, Methodology, Formal analysis. **Yuqin Jiang:** Writing – review & editing. **Ali Mostafavi:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing interests.

Data availability

Data will be made available on request.

References

[1] Y. Jiang, F. Yuan, H. Farahmand, K. Acharya, J. Zhang, A. Mostafavi, Data-driven tracking of the bounce-back path after disasters: Critical milestones of population activity recovery and their spatial inequality, International Journal of Disaster Risk Reduction 92 (2023) 103693. https://doi.org/10.1016/j.ijdrr.2023.103693.
[2] F. Yuan, C. Fan, H. Farahmand, N. Coleman, A. Esmalian, C.-C. Lee, F.I. Patrascu, C. Zhang, S. Dong, A. Mostafavi, Smart Flood Resilience: Harnessing Community-Scale Big Data for Predictive Flood Risk Monitoring, Rapid Impact Assessment, and Situational Awareness (arXiv:2111.06461), arXiv, 2021. http://arxiv.org/abs/2111.06461.
[3] T. Yabe, S.V. Ukkusuri, P.S. C Rao, Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after Hurricane Maria, Applied Network Science 4 (1) (2019) 98, https://doi.org/10.1007/s41109-019-0221-5.
[4] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A.L. Schmidt, C.M. Valensise, A. Scala, W. Quattrociocchi, F. Pammolli, Economic and social consequences of human mobility restrictions under COVID-19, Proc. Natl. Acad. Sci. USA 117 (27) (2020) 15530–15535, https://doi.org/10.1073/pnas.2007658117.
[5] B. Hong, B.J. Bonczak, A. Gupta, C.E. Kontokosta, Measuring inequality in community resilience to natural disasters using large-scale mobility data, Nat. Commun. 12 (1) (2021) 1870, https://doi.org/10.1038/s41467-021-22160-w.
[6] Y. Jiang, Z. Li, S.L. Cutter, Social network, activity space, sentiment, and evacuation: what can social media tell us? Ann. Assoc. Am. Geogr. 109 (6) (2019) 1795–1810.
[7] Y. Jiang, Z. Li, S.L. Cutter, Social distance integrated gravity model for evacuation destination choice, International Journal of Digital Earth 14 (8) (2021) 1004–1018.
[8] C.-C. Lee, A. Rajput, C.-W. Hsu, C. Fan, F. Yuan, S. Dong, A. Esmalian, H. Farahmand, F.I. Patrascu, C.-F. Liu, B. Li, J. Ma, A. Mostafavi, *Quantitative Measures for integrating Resilience into transportation planning practice: Study in Texas* (arXiv:2205.02758), arXiv (2022). http://arxiv.org/abs/2205.02758.
[9] B. Zhou, L. Zou, A. Mostafavi, B. Lin, M. Yang, N. Gharaibeh, H. Cai, J. Abedin, D. Mandal, VictimFinder: harvesting rescue requests in disaster response from social media with BERT, Comput. Environ. Urban Syst. 95 (2022) 101824, https://doi.org/10.1016/j.compenvurbsys.2022.101824.
[10] N. Pourebrahim, S. Sultana, J. Edwards, A. Gochanour, S. Mohanty, Understanding communication dynamics on Twitter during natural disasters: a case study of Hurricane Sandy, Int. J. Disaster Risk Reduc. 37 (2019) 101176.
[11] A. Kumar, J.P. Singh, Y.K. Dwivedi, N.P. Rana, A deep multi-modal neural network for informative Twitter content classification during emergencies, Ann. Oper. Res. 319 (1) (2022) 791–822, https://doi.org/10.1007/s10479-020-03514-x.
[12] B.R. Lindsay, Social Media and Disasters: Current Uses, Future Options, and Policy Considerations, 2011.
[13] C.-C. Lee, S. Namburi, X. Xiao, A. Mostafavi, Homophilic and heterophilic characteristics shaping community formation in human mobility networks during extreme weather response, arXiv (2022) arXiv:2205.04981). http://arxiv.org/abs/2205.04981.
[14] B.D. Phillips, Disaster Recovery, 2nd ed., 74, CRC Press, Boca Raton, 2015, p. 76. https://doi.org/10.1201/b19328.
[15] G.R. Webb, K.J. Tierney, J.M. Dahlhamer, Predicting long-term business recovery from disaster: a comparison of the loma prieta earthquake and hurricane andrew. Global environmental change Part B, Environ. Hazards 4 (2) (2002) 45–58.
[16] D. Serre, B. Barroca, M. Balsells, V. Becue, Contributing to urban resilience to floods with neighbourhood design: the case of Am Sandtorkai/Dalmannkai in Hamburg: contributing to urban resilience to floods with neighbourhood design, Journal of Flood Risk Management 11 (2018) S69–S83, https://doi.org/10.1111/jfr3.12253.
[17] C. Podesta, N. Coleman, A. Esmalian, F. Yuan, A. Mostafavi, Quantifying community resilience based on fluctuations in visits to points-of-interest derived from digital trace data, J. R. Soc. Interface 18 (177) (2021) 20210158, https://doi.org/10.1098/rsif.2021.0158.
[18] N. Coleman, X. Gao, J. DeLeon, A. Mostafavi, Human activity and mobility data reveal disparities in exposure risk reduction indicators among socially vulnerable populations during COVID-19 for five U.S. metropolitan cities, Sci. Rep. 12 (1) (2022) 15814, https://doi.org/10.1038/s41598-022-18857-7.
[19] Q. Li, Z. Tang, N. Coleman, A. Mostafavi, Detecting early-warning signals in time series of visits to points of interest to examine population response to COVID-19 pandemic, IEEE Access 9 (2021) 27189–27200, https://doi.org/10.1109/ACCESS.2021.3058568.
[20] F. Yuan, A. Esmalian, B. Oztekin, A. Mostafavi, Unveiling spatial patterns of disaster impacts and recovery using credit card transaction fluctuations, Environ. Plan. B Urban Anal. City Sci. 49 (9) (2022) 2378–2391, https://doi.org/10.1177/23998083221090246.
[21] B. Rouhanizadeh, S. Kermanshachi, T.J. Nipa, Exploratory analysis of barriers to effective post-disaster recovery, Int. J. Disaster Risk Reduc. 50 (2020) 101735, https://doi.org/10.1016/j.ijdrr.2020.101735.
[22] Z. Wang, X. Ye, Social media analytics for natural disaster management, Int. J. Geogr. Inf. Sci. 32 (1) (2018) 49–72.
[23] M. Yu, C. Yang, Y. Li, Big data in natural disaster management: a review, Geosciences 8 (5) (2018) 165.
[24] M. Jurgens, I. Helsloot, The effect of social media on the dynamics of (self) resilience during disasters: a literature review, J. Contingencies Crisis Manag. 26 (1) (2018) 79–88, https://doi.org/10.1111/1468-5973.12212.
[25] C. Page-Tan, An analysis of social media use and neighbor-assisted debris removal in Houston following Hurricane Harvey, Int. J. Disaster Risk Reduc. 63 (2021) 102450, https://doi.org/10.1016/j.ijdrr.2021.102450.
[26] W.G. Peacock, S. Van Zandt, Y. Zhang, W.E. Highfield, Inequities in long-term housing recovery after disasters, J. Am. Plann. Assoc. 80 (4) (2014) 356–371, https://doi.org/10.1080/01944363.2014.980440.
[27] T. Yabe, P.S.C. Rao, S.V. Ukkusuri, Regional differences in resilience of social and physical systems: case study of Puerto Rico after Hurricane Maria, Environ. Plan. B Urban Anal. City Sci. 48 (5) (2021) 1042–1057, https://doi.org/10.1177/2399808320980744.
[28] N. Seltzer, J. Nobles, Post-disaster fertility: hurricane katrina and the changing racial composition of new orleans, Popul. Environ. 38 (4) (2017) 465–490, https://doi.org/10.1007/s11111-017-0273-3.
[29] D.C. Downey, Disaster recovery in Black and white: a comparison of New Orleans and gulfport, Am. Rev. Publ. Adm. 46 (1) (2016) 51–74, https://doi.org/10.1177/0275074014532708.
[30] E.S. Blake, D.A. Zelinsky, National hurricane center, NATIONAL HURRICANE CENTER TROPICAL CYCLONE REPORT HURRICANE HARVEY, https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf, 2018.
[31] A. Kiaghadi, H.S. Rifai, Physical, chemical, and microbial quality of floodwaters in Houston following hurricane Harvey, Environ. Sci. Technol. 53 (9) (2019) 4832–4840, https://doi.org/10.1021/acs.est.9b00792.
[32] Watson, K. M., Harwell, G. R., Wallace, D. S., Welborn, T. L., Stengel, V. G., & McDowell, J. S. (2018). Characterization of peak streamflows and flood inundation

of selected areas in southeastern Texas and southwestern Louisiana from the August and September 2017 flood resulting from Hurricane Harvey (Report 2018–5070; Scientific Investigations Report, p. 56). USGS Publications Warehouse. https://doi.org/10.3133/sir20185070.
[33] R. Qin, N. Khakzad, J. Zhu, An overview of the impact of Hurricane Harvey on chemical and process facilities in Texas, Int. J. Disaster Risk Reduc. 45 (2020) 101453, https://doi.org/10.1016/j.ijdrr.2019.101453.
[34] A. Ward, Homeless after Harvey: for Some, the Historic Flooding in Houston washed Away Shelter and Security, 2018 Houston Chronicle.
[35] U.S.C. Bureau, ZIP code tabulation areas (zctas). Census.gov, Retrieved November 21, 2022, from. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html, 2022, May 16.
[36] A. Bonner, Guidance for Flood Risk Analysis and Mapping, 2020.
[37] M. Bailey, R. Cao, T. Kuchler, J. Stroebel, A. Wong, Social connectedness: measurement, determinants, and effects, J. Econ. Perspect. 32 (3) (2018) 259–280, https://doi.org/10.1257/jep.32.3.259.
[38] E.D. Vugrin, D.E. Warren, M.A. Ehlen, R.C. Camphouse, A framework for assessing the resilience of infrastructure and economic systems, in: K. Gopalakrishnan, S. Peeta (Eds.), Sustainable and Resilient Critical Infrastructure Systems: Simulation, Modeling, and Intelligent Engineering, Springer Berlin Heidelberg, 2010, pp. 77–116, https://doi.org/10.1007/978-3-642-11405-2_3.
[39] Chia-Fu Liu, Ali Mostafavi, Revealing hazard-exposure heterophily as a latent characteristic of community resilience in social-spatial networks, Sci. Rep. 13 (2023), https://doi.org/10.1038/s41598-023-31702-9.
[40] K. Pearson, Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia, Phil. Trans. 373 (1896) 253–318.
[41] S.D. Bolboaca, L. Jäntschi, Pearson versus Spearman, Kendall's tau correlation analysis on structure-activity relationships of biologic active compounds, Leonardo J. Sci. 5 (9) (2006) 179–200.
[42] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1904) 72–101, https://doi.org/10.2307/1412159.
[43] J. Durbin, A. Stuart, Inversions and rank correlation coefficients, Journal of the Royal Statistical Society Series B. Methodological 13 (1951) 303–309.
[44] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Stat. 19 (1948) 293–325, https://doi.org/10.1214/aoms/1177730196.
[45] A.M. Sadri, S.V. Ukkusuri, S. Lee, R. Clawson, D. Aldrich, M.S. Nelson, J. Seipel, D. Kelly, 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) (2018) 1377–1406, https://doi.org/10.1007/s11069-017-3103-0.
[46] P.K. Freeman, Allocation of post-disaster reconstruction financing to housing, Build. Res. Inf. 32 (5) (2004) 427–437.
[47] J.M. Marcelin, M.W. Horner, E.E. Ozguven, A. Kocatepe, How does accessibility to post-disaster relief compare between the aging and the general population? A spatial network optimization analysis of hurricane relief facility locations, Int. J. Disaster Risk Reduc. 15 (2016) 61–72, https://doi.org/10.1016/j.ijdrr.2015.12.006.
[48] G.Y. Reinhardt, Race, trust, and return migration: the political drivers of post-disaster resettlement, Polit. Res. Q. 68 (2) (2015) 350–362, https://doi.org/10.1177/1065912915575790.
[49] M. Malik, H. Lamba, C. Nakos, J. Pfeffer, Population bias in geotagged tweets, proceedings of the international AAAI conference on web and social media 9 (4) (2015) 18–27.