

ARTICLE OPEN



Equality of access and resilience in urban population-facility networks

Chao Fan¹✉, Xiangqi Jiang², Ronald Lee² and Ali Mostafavi¹✉

While conceptual definitions have provided a foundation for measuring inequality of access and resilience in urban facilities, the challenge for researchers and practitioners alike has been to develop analytical support for urban system development that reduces inequality and improves resilience. Using 30 million large-scale anonymized smartphone-location data, here, we calibrate models to optimize the distribution of facilities and present insights into the interplay between equality and resilience in the development of urban facilities. Results from ten metropolitan counties in the United States reveal that inequality of access to facilities is due to the inconsistency between population and facility distributions, which can be reduced by minimizing total travel costs for urban populations. Resilience increases with more equitable facility distribution by increasing effective embeddedness ranging from 10% to 30% for different facilities and counties. The results imply that resilience and equality are related and should be considered jointly in urban system development.

npj Urban Sustainability (2022)2:9; <https://doi.org/10.1038/s42949-022-00051-3>

INTRODUCTION

Urban systems are composed of urban facilities (such as grocery stores and healthcare facilities), road networks, and human populations¹. The complex interactions between human populations and facilities in urban systems play a key role in the well-being of residents^{2,3}. Urban systems, however, are suffering from significant risks of natural and man-made disasters, such as extreme weather events and road closures, as well as facility shutdown⁴. Such shocks could lead to disruptions of population's access to facilities and degrade the quality of urban life⁵. Improving the resilience of urban systems is of importance to sustain life activities and satisfy life demands in urban populations⁶.

Resilience of a system is defined based on the ability to absorb stresses and retain functionality in disasters^{7–10}. Given the essential role of population-facility interactions for urban life¹¹, in this study, we examine the resilience of urban systems as the ability to retain access to facilities and the extent to which the facilities in service can satisfy the demand of urban populations. A vast number of existing studies^{12,13} have been conducted to characterize and measure the resilience of various individual urban systems, such as transportation networks^{14,15}, business facilities¹⁶, and communities¹⁷. These studies¹⁸ presented useful methods and insights into the understanding of adverse impacts of disasters and how quickly urban systems recover from disruptions. One drawback of these existing works¹⁹, however, is that they focus on the resilience of physical systems, and the consideration of complex interactions between population and facilities for equality of access is still scarce. A few studies²⁰, recently, have highlighted these challenges, including risk disparity and inequality in population-facility interactions in the context of disasters. These studies²¹ pointed out the inequality of disruption impacts among urban populations, which undermines the resilience of urban systems during crises. Integrating equality considerations into resilience analysis of urban systems is an

essential step in the science and practice of urban resilience planning²².

Current studies²³ in inequality research in urban science have focused on definitions²⁴, metric development²⁵, and quantification methods²⁶. Inequality in the context of urban studies is defined as the metric to quantify the differences in any parameter of interest based on social or demographic stratifications. Based on the definition, existing efforts include the specification of inequality indicators in different aspects of urban populations, such as neighborhood isolations²⁷; mobility segregations in physical-social space²⁸, and differential impacts of emergencies²⁹. These studies mainly examine the inequality in different parameters of interest (such as mobility) based on social and demographic stratification³⁰. For example, in the context of disasters, one recent study³¹ found clear socioeconomic and racial disparities in evacuation patterns among Houston communities during Hurricane Harvey. Another study evaluated and found disparate disaster impacts on cities using human-generated information³². More recently, in addition to socio-demographic factors, spatial inequality, which is based solely on the geography of the cities, has also been studied and emphasized³³. Examples of relevant studies include geographic variations in US metro³⁴ and effects of population size on spatial inequality³⁵. The results and findings reported in these studies allow us to measure the degree of inequality in various phenomena. Despite growing attention, the tools and approaches to reducing inequality in urban systems are still rather limited.

One important constraint that impedes the development of inequality-mitigation solutions is the limited granularity of available data. Prior studies were conducted using longitudinal surveys, randomized experiments³⁶, physical sensors, and social media data³⁷. For example, one study used questionnaire surveys to examine the digital divide and inequalities of communities regarding access to information facilities³⁸. These data sets lack details regarding spatial and temporal resolutions of human movement activities and facility locations. Although physical

¹Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA. ²Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA. ✉email: chfan1993@gmail.com; amostafavi@civil.tamu.edu

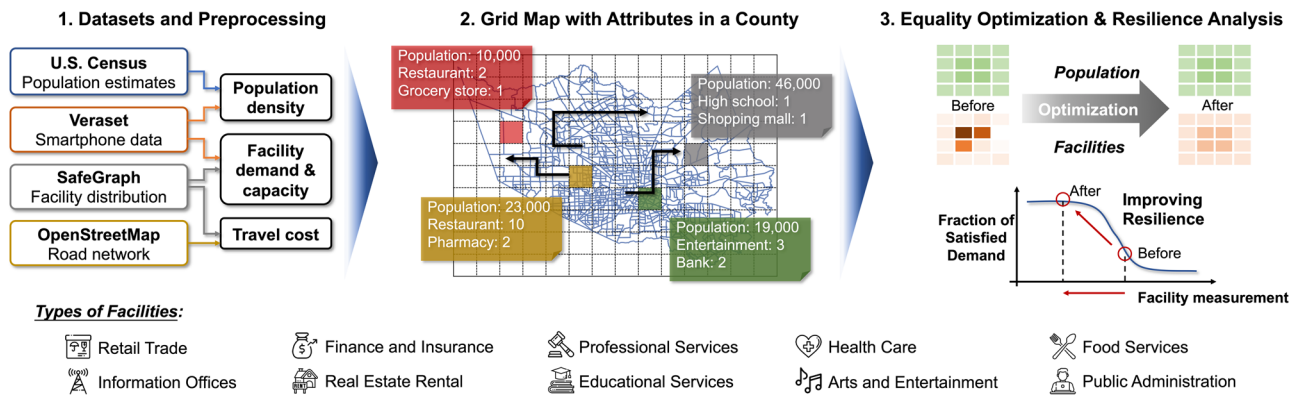


Fig. 1 | Illustration of datasets, preprocessing steps, and the framework of this study. This study employed four large-scale datasets (census data, smartphone location data, facility information, and road networks) to generate the distribution of population, demand and capacity of the facilities, and travel costs. Then we created a grid map for each county, and each grid is associated with a set of attributes, such as population, facilities, and demand of residents. Finally, optimization and resilience models were developed to optimize the distributions of facilities and quantify the change in the resilience of a city. This study focuses on ten types of facilities in each county, such as retail trade, finance and insurance, information offices, as shown in the bottom of this figure.

sensors and social media data may have the geographical information capturing the locations of facilities and human visits, the data itself contain a lot of noise. In prior studies, the limits of this information created challenges in characterizing activities of dispersed populations, how these variations intersect with accessibility, and the influence of resilience of urban systems in short-term shocks³⁹. The prevalence of smartphones has made available, large-scale human activity data. These data sets record the temporal and spatial information of human activities in a very detailed manner, allowing researchers to estimate time and place of device transmissions during the study periods⁴⁰. Through measuring the number of devices visiting a facility in the course of a month, we could infer the lifestyle patterns and facility demand¹¹.

To explore the opportunities of smartphone location data, a great number of advanced analytical methods have been developed in existing related studies. One stream of the studies⁴¹ focuses on investigating mobility patterns of the population to evaluate the accessibility in regular and disaster conditions. For example, Marini et al.⁴² introduced particle swarm optimization methods to address spatial accessibility optimization problems. The approach was further adopted and applied to optimize the residential care facility locations in a study to promote equal access among different age groups of urban populations⁴³. These optimal allocation problems have been studied in mathematics and statistics for decades, and multiple applications have demonstrated their superiority and effectiveness. Notable applications include the optimal allocation of cellular networks⁴⁴ and planning of bicycle infrastructures⁴⁵, which allow maximizing equity of accessibility. In addition, another stream of analytical methods⁴⁶ for use of smartphone data is the mathematical and statistical models of system resilience. For example, Gao et al.¹⁰ proposed a universal resilience model for complex systems which have guided the modeling and design of ecological, biological, economic, and technological systems resilient to both internal failures and environmental changes. Recent studies⁴⁷ have proved the applications of the method in modeling the resilience of urban systems, including social and physical systems. Despite the advancement of analytical methods, studies that reveal the interplay between inequality and resilience are limited. Therefore, existing researchers and practitioners, such as urban planners and emergency responders, have few analytic tools to support the joint considerations and decisions for improved resilience and mitigation of inequality.

To contribute to this research direction, we expand the body of work by a granular and generalizable model which connects inequality to the resilience of urban systems through an effective distance variable. We conducted a multicounty study that examines the accessibility of residential grid cells to different types of facilities by measuring travel distances on road networks. The intent of the optimization model employed in this study is to minimize the total travel distances of the residential populations to facilities and to maximize the equality of access to facilities. The output of the model allows us to identify proper grid cells for developing new facilities and redistributing existing facilities. Further, we created a model to differentiate the resilience of the systems in optimal and empirical scenarios based on the change of facility distributions. Finally, having experimentally validated our measure following short-term disruptions, we understand the impact of facility disruption and the extent to which the improved equality of access could increase the resilience capacity of the urban systems.

RESULTS

Patterns of existing facility distributions

For ten types of facilities (see Fig. 1), we analyzed the patterns of facility and population distributions in ten metropolitan counties in the United States (see Fig. 2b). Here, we take the example of Suffolk County (Boston) and four types of facilities: retail trade, education, health care, and arts and entertainment to illustrate the results. The results for the rest of the counties and facilities are provided in Supplementary Figs. 7–9. Figure 2a depicts the distribution of the urban populations in Suffolk County. The map of the county is divided into grids, and the size of the population in each grid is characterized as the population density in one area. As the figure shows, populations are distributed across the majority of the grid cells in a county. For example, some areas, such as the center, north and east of the county are slightly more populated than other areas. A sharp slope is observed for the Complementary Cumulative Density Function (CCDF) of the population distribution in a county in Fig. 2b, indicating that the variations of population sizes are slight among grid cells. Compared the distributions of population and facilities based on the CCDF in Fig. 2b, d, we observed that the slopes of the CCDFs for facilities are sharper than the CCDF for populations (see numerical values of slopes in Supplementary Table 6). Hence, we consider that the distribution of urban populations is more polycentric than facility distribution in a county. Such a

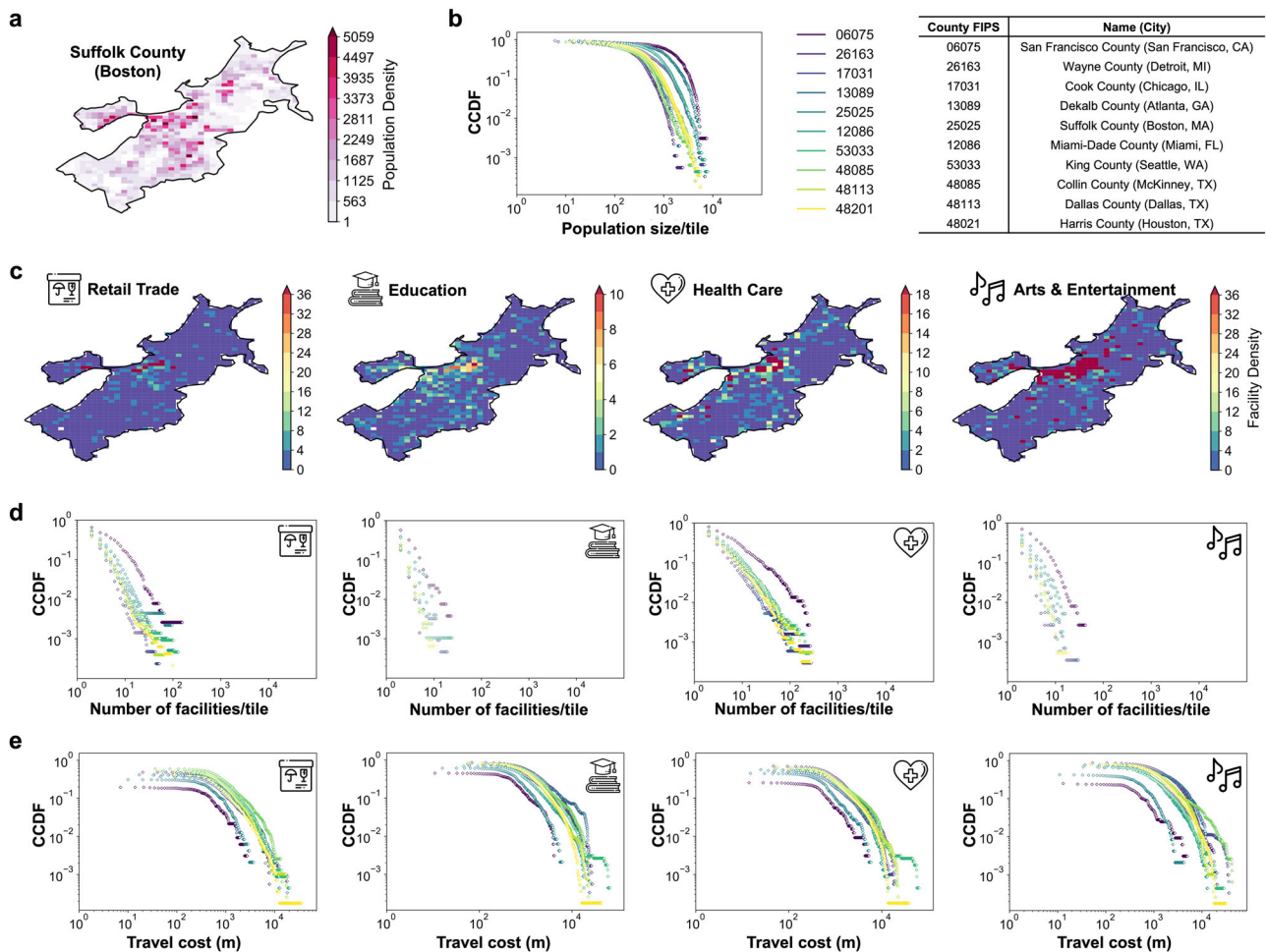


Fig. 2 Patterns of the existing population and facility distributions. **a** Population density in the grid map of Suffolk County (Boston). The population distributions in other counties are provided in Supplementary Fig. 7. **b** the complementary cumulative density function (CCDF) of population density for ten metropolitan counties. The counties are labeled by FIPS code, and the names of the counties are provided in the table next to the plot. **c** The density of four types of facilities, as examples, in the grid map of Suffolk County. The plots for other counties and facilities are provided in Supplementary Figs. 8, 9. **d** CCDF of facility density for ten metropolitan counties for four exemplary types of facilities; **e** CCDF of travel cost from each residential grid cell (tile) to their nearest facility grid cell. The people living in the same grid cells have the same travel cost in accessing the same facility grid cell.

polycentricity pattern is also demonstrated by the Urban Centrality Index (UCI). In Fig. 3c, the UCIs for populations are around 0.2 for all counties, which are the lowest values compared to the UCI of other facilities in a county (see numerical values of UCIs in Supplementary Table 5). This quantitative evidence suggests that the distributions of facilities are spatially variant, compared to the distribution of populations. This pattern can also be observed in other metropolitan counties such as Cook County, King County, and Dallas County. This finding is general among the ten studied metropolitan counties, regardless of the scales, shapes, locations, and population sizes of these counties.

The distributions of facilities, however, indicate a significantly different spatial pattern. Figure 2c shows the spatial distribution of four example facilities in Suffolk County. It is obvious that the majority of the facilities are distributed in the center of the county, especially for retail trade, health care, and entertainment facilities. The rest of the areas in the county have a smaller number of facilities. Such spatial variations in facility density distributions in a county have led to monocentricity of the facility distributions. To further examine the generality of this finding, we plotted the CCDF of the number of facilities in each grid cell for the four example facilities (Fig. 2d). The slopes of all the curves are

particularly sharp, close to a power-law distribution. From these figures, we can infer that the majority of facilities occupy a very few areas, and the rest of the areas only have very few facilities. This is the property of the power-law distribution, which is generally observed from the facility distributions in ten counties.

Counties tend to be polycentric in the spatial distributions of populations, but monocentric in facility distributions. This inconsistency between the distributions of population and facilities creates a challenge for equitable access to facilities. People living in different areas of a county have to travel different distances to access the facility and use the services. As shown in Fig. 2e, the travel costs (distances) people have to spend in accessing the facilities vary significantly. For example, people living in some areas need to travel more than 10 kilometers to access health care facilities. The long travel distance could have dire social, economic, and environmental effects. We observed the unequal travel costs from the distribution of travel costs in Fig. 2e for ten studied counties, which is a general phenomenon in these counties. Specifically, people living in some areas may not need to travel outside of their grid cells, but some people need to travel for a long distance to access the grid cells where these facilities are concentrated. These unequal travel costs are the results of the

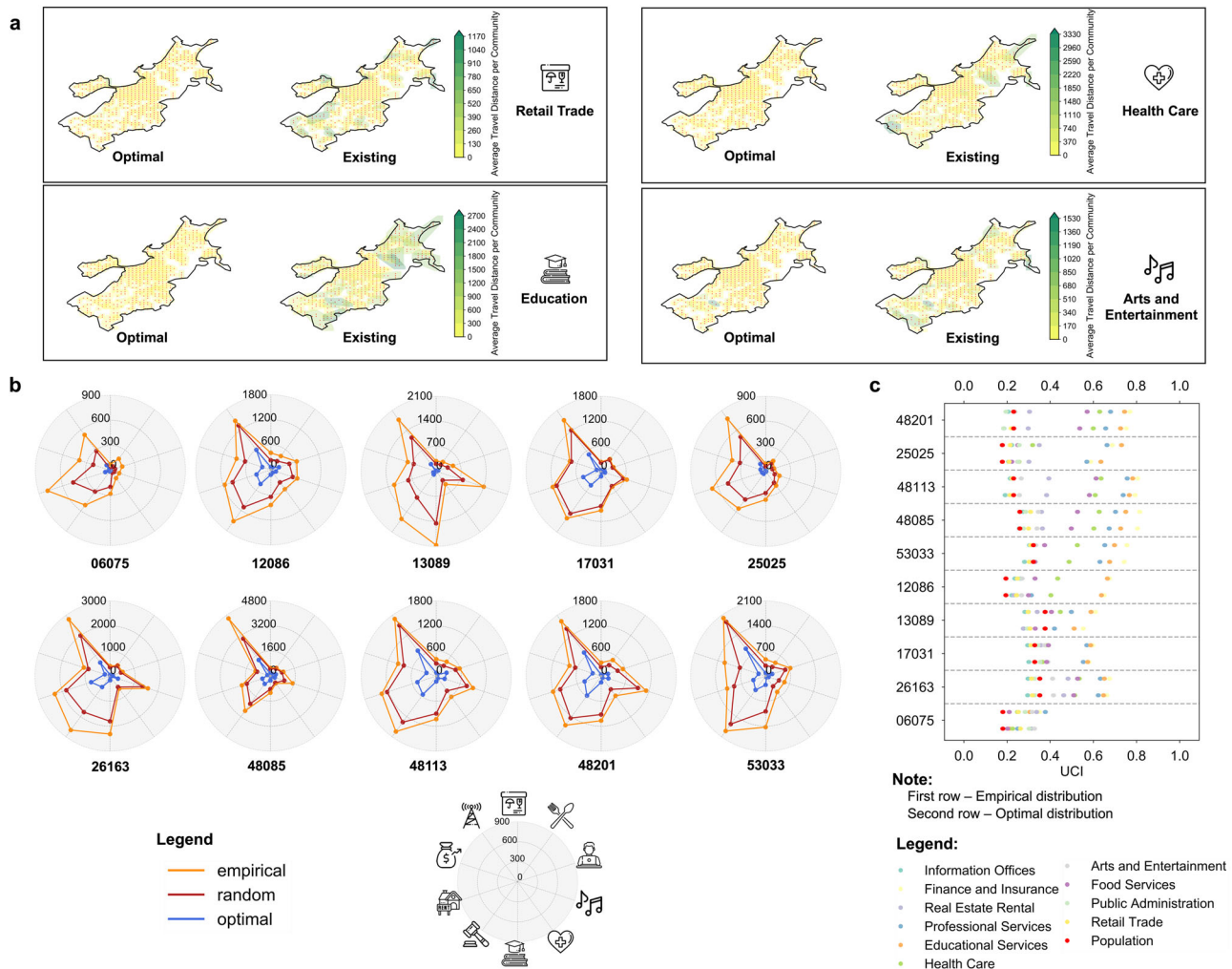


Fig. 3 The optimal distribution of facilities and the resulted travel costs (distances). **a** Service communities of different types of facilities in the existing and optimal scenarios in Suffolk County. The dots represent residential grid cells, and the areas are the service communities. Due to a limited number of facilities in some areas, the people living in grid cells of some service communities have to travel for a long distance to access certain facilities. Hence, some communities are in darker green. The results for the rest of the facilities and counties are provided in Supplementary Fig. 10. **b** The average travel costs in optimal, existing, and random scenarios for ten types of facilities in ten counties. The numbers on each circle on the charts represent the travel costs (in meters). **c** The UCIs of the spatial distributions of facilities and population in counties. The first row of dots for each county represents the UCIs of the existing facility distribution, and the second row of the dots for each county represents the UCIs of the optimal facility distribution. The numeric values of UCIs are provided in Supplementary Table 5.

spatially-variant distribution of facilities, which further lead to grand challenges for population-facility interactions in cities such as traffic jams in facility-concentrated areas, and heavy expenses for people with fewer accesses to facilities. Hence, it is important to optimize the distributions of facilities and improve the equality of travel costs of people living in different areas.

Optimizing the equality of access to facilities

We implemented the optimization algorithms on ten facilities in ten counties. Through urban development and facility redistribution, the enhancement in equality of access to facilities among urban populations can be achieved. Here, we assign each residential grid cell to its nearest facility grid cell. The group of residential grid cells serviced by the same facility is defined as the service community of the facility⁵. We calculated the average travel distances among the residential grid cells in each service community and displayed the results for four types of facilities in Suffolk County in Fig. 3a. In the existing facility distributions, service communities in the northeast and southwest of the county

are in shown in dark green, indicating that people living in these service communities have to travel for a much longer distance than the people living in other areas of the county. It is obvious that the existing pattern of distributed facilities cause an unequal distribution of travel costs among urban populations in accessing facilities. In optimal scenarios where we minimize the total travel distances, populations would have generally more equitable access than those in the existing scenario. Not only would the travel costs be reduced in general, but also people would gain more equal accessibility to these essential facilities.

Figure 3b depicts the extent to which the optimal distributions of facilities reduce the average travel distances of populations to access the facilities in each county. To test the performance of the optimal scenario, we conducted experiments in which we randomly distributed new facilities in the grid map of a county one hundred times and computed the average travel distances for all random experiments. Here, we randomly distributed the facilities across grids in a county. By comparing the average travel distances among three scenarios, we found the optimal

facility distribution that achieves the best performance in reducing the travel distances for urban populations. In addition, the shapes of the radar maps in these three scenarios are almost the same, indicating that, while the optimal scenario could improve the accessibility of the population to facilities, but it does not change the spatial structure of facilities in a county. That is, the average travel distance to a type of facility (such as information offices and professional services) remains relatively long in random and optimal scenarios, if that is the case in the existing scenario. Changing the spatial structure of cities is rather impractical; also, it is important to maintain the existing human lifestyle patterns. Hence, the optimal scenarios would lead to improving equality of access to facilities without significantly altering the spatial structure of cities.

To further quantify the improvement of equality and measure the consistency between facility and population distributions, we adopted the urban centrality index (UCI). This index integrates inequality measurements based on the number of facilities and their spatial relationships. A county has a monocentric structure when its UCI is close to 1. The county is polycentric when UCI is close to 0. Figure 3c displays the results of UCI for the population and facilities in ten counties. The red dots represent the UCI of the populations, and other colored dots represent different types of facilities. It is obvious that the UCI values of facilities span from 0.2 to 0.8. For example, facilities like educational services, finance and insurance services, and professional services, for which the UCI values are close to 0.8 for the majority of the counties, are extremely monocentric. The UCI of the populations surrounding at 0.2, however, tends to be the smallest value among all the UCIs in a county, implying that the distribution of the populations is polycentric. These patterns are generally present in all ten counties. In addition, we observed that the UCIs of some facilities, such as information offices, are close to the UCI of populations, indicating polycentricity of their spatial distributions in a county. In the optimal scenarios, the UCI values for all facilities are reduced, and they get close to the UCIs of populations. The optimal scenarios promote a balanced distribution of facilities and enable a more equal access for the populations to facilities based on travel costs. The results demonstrate the performance of the optimization algorithms and provide an important approach to improve the equal accessibility of the population to facilities.

Quantifying resilience through enhanced accessibility

The improved equality of access to facilities raises an important question regarding the resilience of population-facility (PF) networks in a county. Our hypothesis is that enhanced equality of access also improves resilience in PF networks during crises. In this analysis, we defined the resilience of our PF network as the ability to achieve and retain the high fraction of satisfied demand when facilities are disrupted or shut down in shocks. This approach to model resilience is inspired by the theories of network dynamics. The interactions between entities in every mutualistic system, like ecological systems¹⁰ and job markets⁴⁸, are dynamic, which is due to the intrinsic dynamics of entities as well as their mutualistic interactions. The resilience of a mutualistic network is determined based on its ability to retain basic functionality when disruptions and failures threaten the persistence of the mutualistic interactions⁴⁹. To model the resilience of the PF network, we start by considering a bipartite network consisting of two types of nodes: population grid cells and facility grid cells (Fig. 4a). Populations interact with facilities to satisfy demand. Due to the nature of dynamic human activities and facility supplies ($F(S_i)$), when the demand of populations is be satisfied at one facility grid cell, populations may access to another facility grid cell to satisfy their demand $G(S_i, U_j)$. The connectivity between two facility grid cells, as a measure of the probability people could change from one facility to another when the facility

is disrupted, is captured by the distance, w_{ij} . In short-term shocks, geographically co-located facilities have high probabilities of disruption, while distant facilities can serve as the substitutes to the disrupted facilities. Hence, the larger effective distance would enable the transition of people from disrupted facilities to substitutes to satisfy demand.

Before examining the resilience of the PF networks, we first collected statistics for the facility networks to have a basic understanding of the network structures. Figure 4b plots the networks of four types of facilities in Suffolk County. Most of the networks, especially the networks of retail trade and health care facilities, have core components where the nodes are densely connected with each other due to great weights on the links. Since the weights of the links are defined by the distance between two facility grid cells, the densely connected component indicates large distances among these facilities. The presence of core components could contribute to reducing the risks of large-scale failures among the facilities in the county and also allow people to access the facilities for their lifestyle needs. Hence, the weights of the links in the conceptual facility networks could imply the smoothness of people's transition from one facility to another during disasters. In addition, some nodes are isolated from the network, meaning that these nodes are physically located near the majority of other facilities. In crises like flooding, these nodes may also get perturbed since they are located close to affected areas. Hence, the population's demand may not be satisfied if they access these facilities during disasters.

As we model the resilience from the ecological view of facilities, the resilience of the PF network is influenced by the intrinsic dynamics; the rate of population pattern change (λ); and the mutualistic dynamics, in other words the transition between facilities based on the spatial distributions ($G(S_i, U_j)$). To study the impact of spatial distribution, we first examined the intrinsic dynamics through the parameter λ in the intrinsic dynamics function $F(S_i)$. In Fig. 4c, we show the fraction of satisfied demand for each rate of population pattern change due to the change of population's demand on facilities. For cities with various total capacities of facilities, the rate of pattern change has different effects on the access to facilities and demand satisfaction. No matter the types of facilities, a higher total capacity of a type of facility in a county usually improves the resilience of PF networks. For example, when λ gets close to 1, the fraction of satisfied demand could still remain at a high level. We observe that, for the same type of facility, the facilities with the highest total capacity do not always retain the highest fraction of satisfied demand when λ approaches 1. This result indicates that total capacity does not always contribute to the county resilience and raises the need for the investigation of other factors like the mutualistic dynamics based on facility distribution.

By controlling the parameters λ and α , the effective distance embeddedness w_{eff} is the only factor influencing the fraction of satisfied demand in the model. In Fig. 5a, we plotted the solution derived from a normal status where the changing rate of satisfied demand equals zero. Different counties with a variety of capacities of facilities collapse to the middle of the curve. That is because these counties have similar effective distance embeddedness of corresponding facilities. The effective distance embeddedness encodes the spatial location of the facilities as well as the inequality of the distribution in a county. All counties share a similar distribution pattern for different facilities since the dots are concentrated in the middle of the curve. As shown in Fig. 5a, for example, we see that the distributions of retail trade stores in the ten counties share similar patterns due to the close values of w_{eff} .

Furthermore, to examine the resilience of the facilities in a county, we looked into the change of w_{eff} in both optimal and existing scenarios through the process adopted from the theory of network resilience. In the experiments, we removed f_n fraction of facilities from a county and observed the effects of the

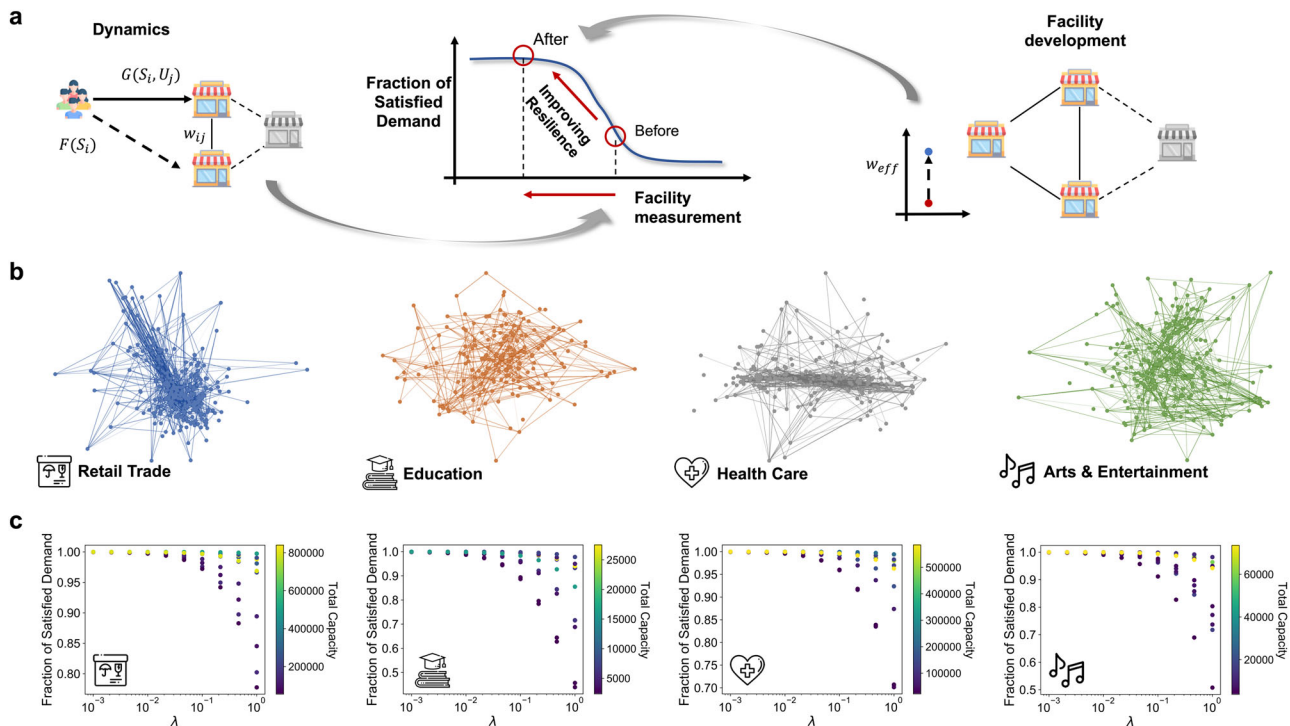


Fig. 4 Characterization and resilience analysis in population-facility (PF) networks. **a** A schema for our formalization of the dynamic process of population-facility interactions. The sub-figure on the left side shows the internal dynamics of the population ($F(S_i)$) and the dynamic interactions with facilities ($G(S_i, U_j)$) in the empirical distribution of facilities. The sub-figure on the right side shows the change of facility distribution in the optimal scenario. The sub-figure in the center illustrates the impacts of changing facility distributions on the fraction of satisfied demand of the population in counties. **b** The facility connectivity characterized by distances for visualizing the urban facility networks. Here, we show example types of facilities in Suffolk county. The nodes are grid cells that have facilities, and the links between grid cells are created based on the connectivity through the road network. The links are visualized when the w_{ij} is greater than a threshold. **c** The steady-state solutions of the resilience model based on the existing distributions of four example types of facilities with varying λ values. The total capacity is obtained by summing up all the capacities of existing facilities in a county. The color bar representing total capacities indicates different counties with different total capacities. More results are provided in Supplementary Fig. 11.

perturbation on the effective distance embeddedness. The perturbations were introduced randomly to the system. That means the removed facilities we selected each time were completely random. Random perturbations in facilities could occur due to crises such as floods, hurricanes, and power outages. One hundred runs of each perturbation simulation were performed. Similar perturbations were implemented on the networks of different facilities. Figure 5b shows the results of the effective distance embeddedness for four facilities in two counties when subjected to perturbations. It can be observed that the values of w_{eff} in the optimal scenarios are usually 10% to 30% higher than those in the empirical scenario. For example, as we can see, the effective distances for retail trade stores in Suffolk and San Francisco counties in optimal scenarios were about 20% higher than the effective distance in empirical scenario. The effective distance decreases along with the removal of facilities in the county. The effective distance in optimal scenarios, however, always stay on top of the effective distance in empirical scenario. By plugging the w_{eff} back to the curve in Fig. 5a, it is evident that the fraction of satisfied demand tends to be higher for greater w_{eff} . This result indicates the ability of the facility network to sustain the fraction of satisfied demand of the population in a county. The optimal scenario can always retain high fractions due to the high effective distance, implying the resiliency of the PF network. Using the measure of effective distance embeddedness in our model, we show the extent to which resiliency of the PF networks can be enhanced when the equality of access is improved. The findings also demonstrate that equality and resiliency of access to facilities are correlated with each other.

DISCUSSION

In this study, we present and calibrate models that connect the resilience and inequality of urban systems based on the dynamic and complex interactions between population and facilities. We obtained optimal facility distribution by maximizing the equality of access and minimizing the travel distances of total populations in a county. Through developing new facilities and redistributing a small set of existing facilities, the travel costs of the populations in accessing facilities would be reduced and balanced among people living in different urban areas. By applying the resilience model in both optimal and empirical scenarios, we found that the optimal scenario is significantly associated with the resilience of the population-facility networks. The interplay between population and facility in ten metropolitan counties in the United States reveals that resilience and equality are correlated with each other. The findings imply that the spatial structure of facility distribution in a county plays an important role in both equality and resilience of access to facilities. This study has significant implications for facility investments, public and private transportation policies, and resilience planning.

First, the patterns of existing facility distribution presented in this study are general across the ten counties and are inconsistent with population distribution in the counties, which leads to significant differences in travel distances to access facilities. The findings in this work highlight the importance of the complex interactions of population, facilities, and road networks which serve as the foundation of urban life. Human activities are time-varying and location-specific. Facilities and infrastructure could influence the activities of humans and the outcomes of their activities, such as demand satisfaction and infrastructure usage.

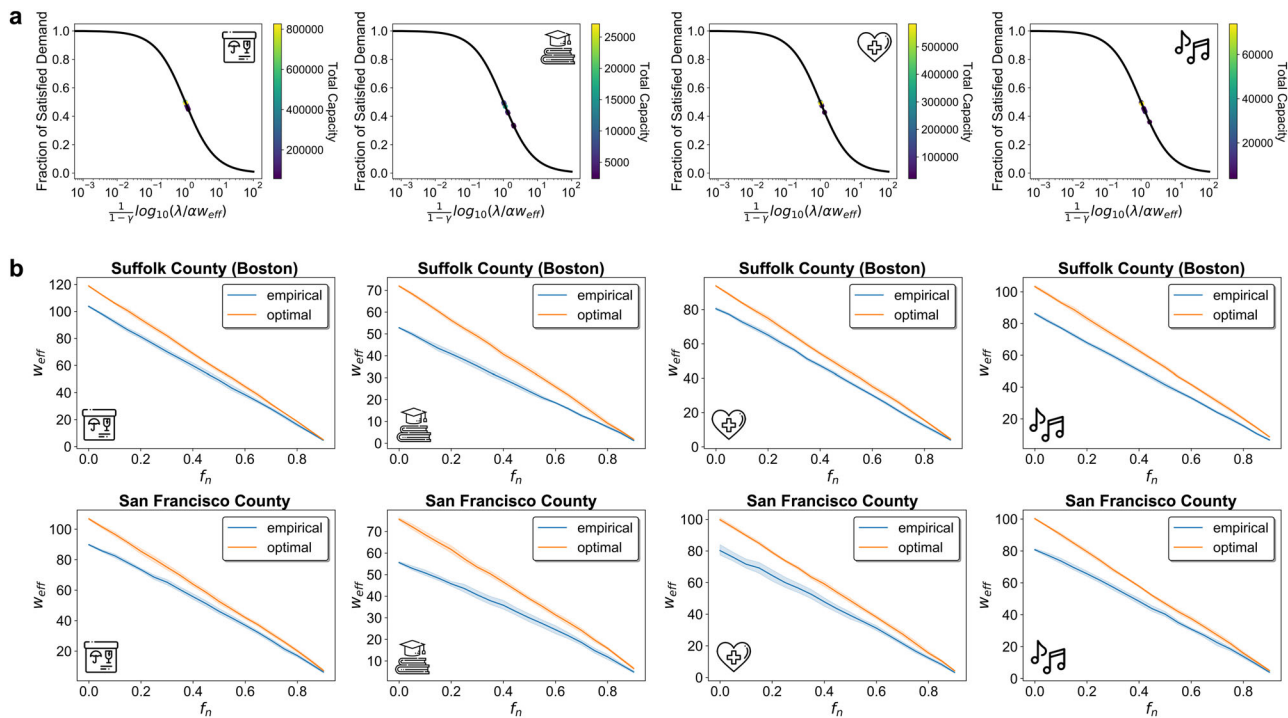


Fig. 5 The quantification of resilience improvement in both empirical and optimal scenarios. **a** The equilibrium solutions for the resilience model based on existing resolutions. ($\gamma = 0.5$, and $\lambda = 0.0464$ in these cases). **b** The relationship between effective distance and removed proportion of facilities to quantitatively describe the resilience of the PF network. The change of the effective distance, w_{eff} , due to the change of facility distributions in optimal and existing distributions. When a proportion of facilities (f_n) are randomly removed from the counties, the effective distance also changes. The results for four example types of facilities for two counties, Suffolk and San Francisco counties, are displayed here. More results are provided in Supplementary Figs. 12, 13.

Analysis and optimizations without considering other entities in urban systems may lead to disruptive impacts. For example, the concentration of facilities in the center of counties may benefit economic activities and operation efficiency. The resulting traffic congestion, however, is more due to the higher density of people in the center of the counties⁵⁰. In our study, we consider the optimal scenario of facility distribution in terms of the travel distances for people living in different areas of cities. Hence, under this consideration, the current facility distribution is suboptimal, compared with the distribution obtained from the optimization model. The current locations of facilities, however, are a result of market and societal forces, which might be an optimal solution for each facility. For example, facilities may collocate in a group since they want to get attention from the consumers of other facilities in the group. This intention of location selection benefits for the profit of facilities themselves. However, this process usually cannot achieve the optimization for the travel costs of the populations in a county. Hence, we consider the current facility distribution is suboptimal in terms of equality and resilience considerations. Problems such as traffic jams, vehicle exhaust emissions, and disparity of life quality are imposed on large cities. The problems of inequality in facility access becomes even more dire during crises. Thus, in the new era of urban development, having an integrative perspective to consider the complex interactions between populations and facilities in cities could inform about effective solutions to current challenges in cities.

Second, while resilience and equality are being studied and evaluated separately in urban science and planning domains, they are, to some extent, correlated and should be considered jointly as important metrics for improving urban facility distributions. Existing studies have shown that inequality is amplified during crises. That case is very likely to be observed when inequality is present in regular conditions. For example, people, living in areas

where facilities are scarce must travel a long distance to access facilities. Assuming the likelihood of disaster impacts are the same for all road segments, the longer the travel distance, the higher the likelihood of getting impacted by disasters⁵¹. Hence, inequality increases the vulnerability of people who have less access to facilities in regular situations and undermines resilience in disasters. Our study only considers the resilience of urban systems to facility shutdown in short-term shocks. We also want to bring to the attention the impacts on equality and resilience caused by road disruptions alone, such as road damage and construction without shutdown of surrounding facilities. To simulate such scenarios, an important component in the model is to characterize the impacts of road disruptions on human decisions on the trade between accessing old facilities with increased travel expenses and selecting new facilities. This simulation requires computing the travel expenses to different facilities. Future research with more empirical data could look into these challenges, and create models to examine the resilience of population-facility networks in more scenarios. Concurrent consideration of resilience and equality is essential and would lead to plans and decisions that not only improve urban function during normal times, but also make urban systems more resilient in the case of crises.

Third, although there is a growing interest in urban facility development that enhances the equality and resilience of access to facilities, few analytical methods and tools are available to guide new investments in achieving equality and resilience goals. The models presented in this study have important practical contributions and implications for the planning of urban facilities. Of particular importance, our study helps urban decision makers develop data-driven resilience and equality plans, which account for localized variations in urban populations. In practice, the relocation cost and land availability may limit the capability of redistributing facilities in cities. To deal with this practical issue,

in our model, we also consider adding new facilities and also optimize the locations of these new facilities in the phase of urban planning and development. Hence, the results from our analyses could provide suggestions to the urban planners and city managers and governments to take into account the equality and resilience in the process of urbanization. Potential applications could be evaluating the locations of new facilities, identifying a sequence of locations for investment prioritization, and analyzing the contribution of the facilities to the resilience of a county. Such applications and information are essential for counties with limited resources and allow planners to quantify the benefits of facility development on equality and resilience. In addition, our model involves the travel distance which highly depends on the road networks in cities. Planning and construction of road networks for increasing transportation efficiency and reducing travel costs could also benefit the equality and resilience of the population-facility networks. Hence, the outcomes of our models could inform stakeholders regarding planning and development of road networks. The methods in this study are also scalable to large areas such as metropolitan statistical areas which include tens of populated counties. In that case, the equality and resilience of urban-rural regions could be further explored using proposed methods.

Finally, this work is among the first attempts to optimize facility distribution for inequality mitigation and resilience improvement. We made a couple of assumptions in this study that did not capture the effects of social backgrounds of the residents and also the effects of transportation systems. For example, low-income, elderly and minority people may need more support, such as lower expenses, to maximize their access to facilities. Equal distribution may still fail to satisfy their life demand. Hence, integrating fairness in equality and considering the equity problem in urban system development is of vital importance and necessity. In addition, this study only considered travel distances as the measure of accessibility to facilities. Transportation tools, such as public transportation systems, private cars and ride-sharing, are diverse and developed in metropolitan counties. The traffic conditions on different roads vary as well. Taking into account these factors as measures of travel costs could lead to more accurate and realistic solutions. Despite these limitations, the models and findings from this study provide tools and evidence for improving equality and resilience in complex population-facility networks of urban systems and could also serve as a foundation for future research avenues.

METHODS

Data collection and preprocessing

This study involves multiple datasets, such as census demographics, anonymized smartphone data, facility locations, and road networks, at the county scale but with different spatial resolutions. To enable consistency across all datasets, we defined a grid map that divides a county into grids of similar size (about 0.5 km × 0.5 km) based on methods suggested in existing studies¹⁵. Each grid could be associated with the characteristics of populations demographics, facilities, and road networks.

Population distribution estimation. Demographic data collected by the US Census provide an accurate estimation of the population at the census tract level⁵². It should be noted that, census blocks are very fine-scale and high resolution. However, the census data such as population size are not publicly available at the census block level. Hence, we adopted the grid map and distribute the populations to grid cells. The U.S. Census Bureau's 2014–2018 American Community Survey forms the basis of population density estimation. Census tracts, however, are shaped irregularly and are much larger than the grid we defined. Estimating the population in each grid requires additional information. Here, we adopted anonymized mobile phone data from Veraset Inc., a company that collects location data across a number of applications on mobile phone devices after the user consents to the use of their anonymized data. Although this data contains only the location information for a subset of

the population, prior studies^{53,54} using these data have found them valid representations of demographics in the United States. The data was shared under a strict contract with Veraset through their collaborative program in which they provide access to de-identified and privacy-enhanced mobility data for academic research only. All researchers processed and analyzed the data under a non-disclosure agreement and were obligated to not share data further or to attempt to re-identify data.

Of the more than 30 million smartphone devices with 2–3 billion location data per day in the US sample, we filtered the data for ten metropolitan counties, including King County (Seattle), Washington, Harris County (Houston), Texas, and Suffolk County (Boston), Massachusetts. The data set spans from January 27, 2020, to February 23, 2020, (4 weeks). Estimating population density requires knowledge of the home grids of the devices, which are not directly reported by Veraset. Based on successful cases in home location estimation from prior studies²⁷, density-based spatial clustering of applications with noise (DBSCAN) is a commonly used, time-efficient and scalable method that can accurately identify clusters of data points with noise. Hence, we employed DBSCAN and the location points from 10 p.m. to 7 a.m. local time to estimate the home grids of our sample devices. The algorithm allows us to remove dwell points it designates as noise, and to cluster geospatial coordinates, given a minimum threshold of samples (we used three in this study) and a minimum distance (we used 0.0004 degree for the differences of longitude and latitude). We consider that the home grids of the devices are most likely within the clusters with most data points. As such, the grid containing the average of data points in the highest-density cluster is treated as the home grid of a device. For devices with a number of data points below the minimum threshold, we directly took the average of the coordinates of closed data points as the home locations of these devices.

The estimation of the number of devices in grids provides a reference to a basic understanding of the population distribution across grids in a census tract. Using this data, we computed the proportion of devices in a grid among all the grids in a census tract. Then we allocated the population from the census survey in each census tract to grids as an estimate of people living in a specific grid. By doing so, we computed the ratio of smartphone devices to the census population, called “amplification factor” (α). This factor will be further used to estimate the demand of people for facilities and the capacity of the facilities in the county.

Facility distribution. Information such as categories and locations of urban facilities were provided by SafeGraph, a location intelligence data company that has documented the geographical and business information about physical places in the United States⁵⁵. Each facility, also called as “point of interest,” is a dedicated geographic entity that people may visit for different needs or interests. Examples of facilities include restaurants, retail stores, and grocery stores. In the data set, each facility is also associated with geographic coordinates to geographically identify the facility, the industry category code, and so on. Industry categories are defined by the North American Industry Classification System (NAICS)⁵⁶. We selected the ten most common categories of facilities in our analysis, such as retail trade, finance and insurance, health care, and recreation (see Fig. 1). The complete set of facility categories, codes, and examples is provided in Supplementary Note 1.

Demand and capacity estimation. People's demands for different facilities are vary naturally based on the types of facilities and their lifestyles. Such variation results in complex interactions between people and facilities and influences travel costs for accessing facilities. To take into consideration facility demands of people living in urban areas, we investigated facility visit patterns from our smartphone location data. We computed the number of times a device visited a specific type of facility during a 4-week period. By aggregating the number of times the devices, whose home locations are in the same grid, visited a specific type of facility, we can make a basic estimate of the demand of the devices for the facility. To estimate the demand of the actual population living in a grid cell, the basic demand estimates are further rescaled by the amplification factor α . By this process, we have estimates of demand on facilities by different urban subpopulations.

Considering the balance between the demand of people on the facilities and the capacity of the facilities, the capacity of all facilities in the same category is usually equivalent to the demand of people in a county in normal conditions. Hence, we approximated the capacity of a type of facility in a county based on the estimated demand of the total population for this facility. In addition, to further simplify the calculation, the capacity

of each facility is assumed to be stable and equal during the study period. We distributed the total estimated capacity of the facilities in a county equally to each facility in the same category. The number of facilities in a grid, hence total capacity, however, varies across the county. For example, one grid cell has two facilities, while the other has one facility. Then the capacity of the first grid cell is twice as the capacity of the other grid cell.

Road networks and distance estimation. The road network data was obtained from OpenStreetMap, which provides geodata underlying the maps of the world⁵⁷. The data set includes road categories, speed limits, and the coordinates of the start and end of a lane. We used the roads in the categories of highway and primary roads to construct the road networks, which are adopted to compute the distance between each pair of grid cells in a county and further estimate the travel cost people have to spend to move from one grid cell to another. In this study, the grids that are geographically co-located are connected by the roads. Counties, especially large counties like Harris County, are divided into a large number of grids, which results in high computation costs to identify the shortest path between any randomly selected pair of grids. To this end, we simplified the process of calculating the grid distance by summing up travel costs between geographically neighbored grids along the way from one grid to another. That is, we first, calculated the travel cost for pairs of grids that are geographical neighbors, sharing a border or corner. The travel cost is derived by randomly selecting the intersection of roads in each grid, identifying the path connecting the selected intersections in two nearby grids, and calculating the length of the path using the adopted road networks. Then, we identified continuous lists of grids on all possible routes for any selected pair of grids that are not geographical neighbors. By summing up the travel costs between the grids in all lists, we can finally find the shortest path for any selected pair of grids, which is used to approximate the actual shortest path between two grids. This method is relatively efficient and saves significant amount of computational costs in implementing the optimization algorithms in the following analyses.

Urban centrality index

To measure the distributions of urban populations and facilities in counties, this study employed urban centrality index (UCI)⁵⁸. UCI can reflect the extent to which the facilities or populations are unequally distributed in urban areas. The scale of UCI ranges from 0 to 1, representing the transition from poly-centrality to mono-centrality. The UCI is the product of two elements, local coefficient (LC) and proximity index (PI). The local coefficient measures the inequality of the distribution in urban areas using:

$$LC = \frac{1}{2} \sum_{i=1}^N \left| k_i - \frac{1}{N} \right| \quad (1)$$

where, N is the total number of areas (grids), k_i is the share of quantity (here, we refer to population or facilities) in the area (grid) i . When LC gets close to 0, the quantity gets to an even distribution. The distribution becomes more concentrated when LC gets larger.

The proximity index measures the spatial separation of the quantity across urban areas, which is formulated by:

$$V = \mathbf{K}' \times \mathbf{D} \times \mathbf{K} \quad (2)$$

$$PI = 1 - \frac{V}{V_{\max}}$$

where, \mathbf{K} is a vector of k_i , \mathbf{D} is a distance matrix for all pairs of grids, with zeros on its diagonal. V , Venables index, indicates the degree of spatial separation, but the scales may vary in different spatial settings due to various magnitudes of distance elements in \mathbf{D} . Hence, the proximity index normalizes the value of V . V_{\max} represents the maximum attainable value of V in a defined urban area. Since the shapes of counties tend to be irregular, we approximate the V_{\max} by uniformly distributing the quantity on the grids along the boundary of a county. Then, the theoretical range of PI can be (0, 1).

Finally, the UCI is obtained using Eq. (3),

$$UCI = LC \times PI \quad (3)$$

Large UCI values indicate that the distribution of population or facilities is more monocentric, while small UCI values show polycentric distributions.

Equality of access to facilities

Cities are expanding in both population size and facilities. The distributions of populations and development of facilities, however, are inconsistent,

leading to inequality of access to facilities in urban areas. Finding optimal distributions of facilities to reduce inequality is essentially a location specification problem. In this study, we consider the two cases, development of new facilities and redistribution of existing facilities, in conjunction during urban expansion. Hence, the problem, here, is formalized as: given a set of grid cells, N , in a county, a set of residential cells $R \in N$ which are populated by humans, how can we distribute the facilities in grid cells M to minimize the travel costs (distances) from residential cells to facility cells for all urban populations.

Here, we consider the grid cells for facilities are composed of existing facility cells and new facility cells (around 10% of existing facility cells) which are randomly distributed in a county at the beginning of this procedure. Then, we form this problem as an integer program:

$$\begin{aligned} & \text{minimize} \quad \sum_{i \in F, j \in R} d_{ij} r_{ij} p_j \\ & \text{subject to} \quad \forall j \in R: \sum_{i \in F} r_{ij} = 1 \\ & \quad \quad \quad \forall i \in F, j \in R: r_{ij} \leq y_i \\ & \quad \quad \quad \sum_{i \in F} y_i = M \\ & \quad \quad \quad \forall i \in F, j \in R: r_{ij} \in \{0, 1\} \\ & \quad \quad \quad \forall i \in F: y_i \in \{0, 1\} \end{aligned} \quad (4)$$

where, i is the index of the grid cells of facilities, j is the index of grid cell of residential population, p_j is the size of the population in the grid cell j , R and F are sets of grid cells of residents and facilities. In this problem, we are given a metric space with resident cells and facility cells, travel costs d_{ij} of residents living in grid cell j , and accessing facility i . y_i indicates if facility i is open and r_{ij} indicates if residents from grid cell j are accessing facility i . The constraints on r_{ij} in the equation means that residents from a grid cell can only be assigned to one nearest facility grid cell, and in the meantime, the facility must be open. This problem is defined as a p-median problem. We applied a fast interchange heuristic based on swap-based local search to solve this problem⁵⁹. That is, given a certain option for inserting a facility f_i , the approach would compute the profit $profit(f_i, f_r)$ based on the gain and loss of distance change, as well as extra distance differences in accessing facility f_i and f_r . Detailed algorithms and formulae are provided in Supplementary Note 2.

Resilience model of population-facility networks

As described earlier, interactions between population and facilities are dynamic and mutualistic. First, the population's life activity patterns for their demand on facilities may change over time. People may not stick to a specific facility for satisfying their life demands. Second, people's demand for facilities could be satisfied by accessing facilities in different urban areas, which provide similar services and supplies. From the perspective of system resilience, the prominent role in evaluating resilience is the structure of networks¹⁰. Here, we applied analogous measures to facility networks in a county to model its resilience due to perturbations in road networks or facilities themselves.

To model dynamic human activities in accessing facilities, the model has to treat facilities in each grid as a basic unit serving a certain amount of population. Here, we introduce a dynamical model on the population-facility (PF) network, a bipartite network representing the dynamic and mutualistic interactions between population and urban facilities. The resilience model of the PF network is composed of two components. The first component describes the dynamics of demand satisfaction in a specific facility due to the change of people's tendencies in accessing facilities. The second component describes the spillover effects between facilities in different grids due to the disruptions of facilities in certain grids, which can influence the resilience of the PF network. When the facilities in one grid cannot satisfy the demand of the population, people may change to access facilities in other grids.

This process requires a function $G(U_i, S_j)$ that captures the change of accessing behaviors between facilities in grids i and j ⁴⁸. Here, S_j represents satisfied demand by the facility at location j , and U_i represents unsatisfied demand by the facility at location i . Most prior studies^{48,60} model this process as $G(U_i, S_j) \propto w_{ij} S_j^\gamma U_i^{1-\gamma}$, where $\gamma \in [0.5, 0.7]$. The parameter, w_{ij} , is indicative of how easily the population can switch between facilities in different grids. Since we are studying the context of disasters, such as flooding and power outage, it is important to consider disruptions in facilities. We consider that the more dispersed the distribution of facilities, the lower the risks that the facilities suffer from the same perturbation. That is, distant facilities can serve as the substitutes for the facilities which

are disrupted in crises. For example, if facilities in some grid cells are disrupted, then facilities in other grid cells which are far away from the impacted grids could support unsatisfied demand to maintain the normal level of life activities among residents. Essentially, our hypothesis is that cities with more decentralized distributions of facilities would be more resilient to extreme events. Although both equality and resilience models in population-facility networks incorporate distance metrics as parameters, the metrics of distance in the two models are different. The equality of access to facilities is measured based on the travel distance from people's residential areas to the locations of facilities. The resilience model is derived from existing theories of network resilience and percolation¹⁰. The spread probability of failure is in proportional to the distances between the nodes in a network⁴⁷. Hence, the resilience model incorporates the distance as a parameter. The resilience model, however, is created based on the facility network solely. The distance metric involved in the resilience model is the distance between any of the two facilities. Moreover, the distance between two facilities is not essentially related to the travel distances of the populations. In other words, decentralized distribution of facilities does not necessarily mean the optimal scenario where facility distributions enable equality of access for the populations. The hypothesis for the resilience model to be reasonable, which does not serve as a strong assumption for the model.

Accordingly, we measure the distances between facilities in different grids based on the road networks and define the values of distances as w_{ij} in the formula. We obtain the equations as follow:

$$\begin{aligned} \frac{dS_i}{dt} &= -\lambda S_j + \alpha \sum_{i \in F} w_{ij} S_j^y U_i^{1-y} \\ \frac{dU_i}{dt} &= \lambda S_j - \alpha \sum_{i \in F} w_{ij} S_j^y U_i^{1-y} \end{aligned} \quad (5)$$

where, λ represents the rate of pattern change, meaning the rate that people change their demand on a facility, α is a fitting parameter, w_{ij} represents the road distance between facilities in location i and j , t here represents the time. Here, $\sum_{j \in F} (S_j + U_j)$ is a constant, which represents the total demand of the population on a type of facility in a county. A significant natural disaster could indeed change the consumption patterns of the population, like how often people go to grocery stores and the number of commodities that people would buy once. Despite that, the demand of the population should be constant with what people need in regular conditions. Our model does not account for the frequencies of facility accesses and other consumption patterns. The extreme cases will not influence the results significantly. Based on the definition and assumptions, this model enables studying the dynamics of human-facility interactions during crises like facility shutdown, power outage, or flooding. The long-term effects, such as travel costs and economic saving, are included in the optimization model with enhancement of equitable access.

Modeling facility networks in counties requires high-dimensional representations. For example, in a city with M grid cells for a type of facility, the matrix of \mathbf{W} will have $\binom{M}{2}$ dimensions. Such high dimensionality increases computational costs. To reduce its dimensionality, we simplify the model to an effective one-dimensional model, based on a methods from prior studies⁶¹. The simplification method takes the average of the values in all dimensions for a grid cell, $w_j = \sum_{i \in F} w_{ij}$, which can be called grid cell j 's embedding. Let $\mathbf{W} = \sum_{i,j \in F^2} w_{ij}$, to aggregate the embeddings of all grid cells. Then, we further calculate the effective average for all variables:

$$S_{\text{eff}} = \left(\sum_{j \in F} S_j w_j \right) / W, U_{\text{eff}} = \left(\sum_{j \in F} U_j w_j \right) / W, w_{\text{eff}} = \left(\sum_{j \in F} (w_j)^2 \right) / W \quad (6)$$

where, S_{eff} is the effective satisfied demand, U_{eff} is the effective unsatisfied demand, and w_{eff} is the effective distance, serving as low-dimensional representative variables with an aggregation of the same type of variables in the network. These effective variables are similar to a form of weighted-degree average, and this form is robust against noise and parameter variations, and the computation can complete in a reasonable amount of time⁶². Using the effective variables, we can reduce the dimensionality of the resilience model and reformulate it as:

$$\begin{aligned} \frac{dS_{\text{eff}}}{dt} &= -\lambda S_{\text{eff}} + \alpha w_{\text{eff}} S_{\text{eff}}^y U_{\text{eff}}^{1-y} \\ \frac{dU_{\text{eff}}}{dt} &= \lambda S_{\text{eff}} - \alpha w_{\text{eff}} S_{\text{eff}}^y U_{\text{eff}}^{1-y} \end{aligned} \quad (7)$$

This one-dimensional system is solvable. We consider the normal status of the facilities in a city, where the fraction of satisfied demand is stable given grid cell representations \mathbf{W} and an undisturbed environment. That is, the changing rate of S_{eff} equals to zero. In this status, the fraction of satisfied demand can be obtained by:

$$\hat{S}_{\text{eff}} = \frac{1}{1 + (\lambda / (\alpha w_{\text{eff}}))^{1/(1-y)}} \quad (8)$$

where, \hat{S}_{eff} is the solution of satisfied demand under the condition of certain effective distance. The derivation for Eq. (8) is shown in Supplementary Note 3. This model allows us to study the resilience of the populations-facilities (PF) network in a universal manner, regardless of population size, geography, and lifestyles in different cities. The main influential factor influencing the resilience and dynamics of the PF network is the effective distance w_{eff} which embedded the spatial topology of the facility distribution in a city. The extent of resilience enhancement based on the optimal distribution of facilities could be quantified based on the effective distance. To support the main claim in this study, we provide an example (see Supplementary Note 4) to indicate that the distance metrics to evaluate the resilience performance of the population-facility network is not correlated to the travel distance of the populations in the optimization model.

As we defined in the second step of the method, each grid cell includes facilities and road infrastructure. The grid cells in a county are connected by road segments. We consider the failure of the entire grid cell when simulating disruptions of facilities. That is, all facilities and road infrastructure in the grid cell would fail. Hence, individuals cannot pass through the roads in disrupted grid cells. In this case, our simulation with a consideration of other urban systems is consistent with real-world scenarios such as flooding.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

All data were collected through a CCPA and GDPR compliant framework and utilized for research purposes. The data that support the findings of this study are available from Veraset Inc., but restrictions apply to the availability of these data, which were used under license for the current study. The data can be accessed upon request submitted on veraset.com. Other data we use in this study are all publicly available.

CODE AVAILABILITY

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

Received: 16 July 2021; Accepted: 16 February 2022;

Published online: 31 March 2022

REFERENCES

- Liu, X. et al. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* **3**, 564–570 (2020).
- Bettencourt, L. M. A. Urban growth and the emergent statistics of cities. *Sci. Adv.* **6**, eaat8812 (2020).
- Fan, C. & Mostafavi, A. Metanetwork framework for performance analysis of disaster management system-of-systems. *IEEE Syst. J.* <https://doi.org/10.1109/JSYST.2019.2926375> (2019).
- McDaniels, T., Chang, S., Cole, D., Mikawoz, J. & Longstaff, H. Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation. *Glob. Environ. Chang.* **18**, 310–318 (2008).
- Xu, Y., Olmos, L. E., Abbar, S. & González, M. C. Deconstructing laws of accessibility and facility distribution in cities. *Sci. Adv.* **6**, eaab4112 (2020).
- Elmqvist, T. et al. Sustainability and resilience for transformation in the urban century. *Nat. Sustain.* **2**, 267–273 (2019).
- Cimellaro, G. P., Reinhorn, A. M. & Bruneau, M. Framework for analytical quantification of disaster resilience. *Eng. Struct.* **32**, 3639–3649 (2010).
- Ganin, A. A. et al. Resilience and efficiency in transportation networks. *Sci. Adv.* **3**, e1701079 (2017).

9. Hung, H.-C., Yang, C.-Y., Chien, C.-Y. & Liu, Y.-C. Building resilience: Mainstreaming community participation into integrated assessment of resilience to climatic hazards in metropolitan land use management. *Land use policy* **50**, 48–58 (2016).
10. Gao, J., Barzel, B. & Barabási, A.-L. Universal resilience patterns in complex networks. *Nature* **530**, 307–312 (2016).
11. Di Clemente, R. et al. Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nat. Commun.* **9**, 3330 (2018).
12. Ganin, A. A. et al. Resilience in intelligent transportation systems (ITS). *Transp. Res. Part C Emerg. Technol* **100**, 318–329 (2019).
13. Meerow, S., Newell, J. P. & Stults, M. Defining urban resilience: a review. *Landsc. Urban Plan.* **147**, 38–49 (2016).
14. Li, D. et al. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proc. Natl. Acad. Sci.* **112**, 669–672 (2015).
15. Fan, C., Jiang, X. & Mostafavi, A. A network percolation-based contagion model of flood propagation and recession in urban road networks. *Sci. Rep.* **10**, 1–12 (2020).
16. Yabe, Takahiro, Zhang, Yunchang & Ukkusuri, Satish, V. Quantifying the economic impact of disasters on businesses using human mobility data: a Bayesian causal inference approach. *EPJ Data Sci.* **9**, 36 (2020).
17. Cutter, S. L. et al. A place-based model for understanding community resilience to natural disasters. *Glob. Environ. Chang* **18**, 598–606 (2008).
18. Ouyang, M. & Fang, Y. A Mathematical framework to optimize critical infrastructure resilience against intentional attacks. *Comput. Civ. Infrastruct. Eng.* **32**, 909–929 (2017).
19. Kontokosta, C. E. & Malik, A. The resilience to emergencies and disasters index: applying big data to benchmark and validate neighborhood resilience capacity. *Sustain. Cities Soc.* **36**, 272–285 (2018).
20. Natalie, C., Amir, E. & Ali, M. Equitable resilience in infrastructure systems: empirical assessment of disparities in hardship experiences of vulnerable populations during service disruptions. *Nat. Hazards Rev.* **21**, 4020034 (2020).
21. Esmalian, A., Dong, S., Coleman, N. & Mostafavi, A. Determinants of risk disparity due to infrastructure service losses in disasters: a household service gap model. *Risk Anal.* **41**, 2336–2355 (2021).
22. Yabe, T. & Ukkusuri, S. V. Effects of income inequality on evacuation, reentry and segregation after disasters. *Transp. Res. Part D Transp. Environ.* **82**, 102260 (2020).
23. Mirza, M. U., Xu, C., Bavel, B., van, van Nes, E. H. & Scheffer, M. Global inequality remotely sensed. *Proc. Natl. Acad. Sci. USA* **118**, 18 (2021).
24. Acemoglu, D. & Robinson, J. Foundations of Societal Inequality. *Science*. **326**, 5953, 678–679 (2009).
25. Gazzotti, P. et al. Persistent inequality in economically optimal climate policies. *Nat. Commun.* **12**, 3421 (2021).
26. Emrich, C. T., Tate, E., Larson, S. E. & Zhou, Y. Measuring social equity in flood recovery funding. *Environ. Hazards* **19**, 228–250 (2020).
27. Wang, Q., Phillips, N. E., Small, M. L. & Sampson, R. J. Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc. Natl. Acad. Sci. USA* **0**, 201802537 (2018).
28. Xu, Y., Belyi, A., Santi, P. & Ratti, C. Quantifying segregation in an integrated urban physical-social space. *J. R. Soc. Interface* **16**, 20190536 (2019).
29. Millett, G. A. et al. Assessing differential impacts of COVID-19 on black communities. *Ann. Epidemiol.* **47**, 37–44 (2020).
30. Frigerio, I. & De Amicis, M. Mapping social vulnerability to natural hazards in Italy: a suitable tool for risk mitigation strategies. *Environ. Sci. Policy* **63**, 187–196 (2016).
31. Hong, B., Bonczak, B. J., Gupta, A. & Kontokosta, C. E. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* **12**, 1870 (2021).
32. Kryvasheyev, Y. et al. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2**, 1–11 (2016).
33. Achten, S. & Lessmann, C. Spatial inequality, geography and economic activity. *World Dev.* **136**, 105114 (2020).
34. Florida, R. & Mellander, C. The geography of inequality: difference and determinants of wage and income inequality across US metros. *Reg. Stud.* **50**, 79–92 (2016).
35. Kondo, N. et al. Income inequality and health: the role of population size, inequality threshold, period effects and lag effects. *J. Epidemiol. Community Heal* **66**, e11–e11 (2012).
36. Chetty, R., Hendren, N. & Katz, L. F. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am. Econ. Rev.* **106**, 855–902 (2016).
37. Dargin, J. S., Fan, C. & Mostafavi, A. Vulnerable populations and social media use in disasters: uncovering the digital divide in three major U.S. hurricanes. *Int. J. Disaster Risk Reduct* **54**, 102043 (2021).
38. Bornman, E. Information society and digital divide in South Africa: results of longitudinal surveys. *Information. Commun. Soc* **19**, 264–278 (2016).
39. Yao, Y., Hong, Y., Wu, D., Zhang, Y. & Guan, Q. Estimating the effects of “community opening” policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses. *Comput. Environ. Urban Syst.* **70**, 163–174 (2018).
40. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
41. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci. USA* **109**, 11576–11581 (2012).
42. Marini, F. & Walczak, B. Particle swarm optimization (PSO). A tutorial. *Chemom. Intell. Lab. Syst.* **149**, 153–165 (2015).
43. Tao, Z., Cheng, Y., Dai, T. & Rosenberg, M. W. Spatial optimization of residential care facility locations in Beijing, China: maximum equity in accessibility. *Int. J. Health Geogr.* **13**, 33 (2014).
44. Wu, S., Yang, J., Peng, R. & Zhai, Q. Optimal design of facility allocation and maintenance strategy for a cellular network. *Reliab. Eng. Syst. Saf.* **205**, 107253 (2021).
45. Olmos, L. E. et al. A data science framework for planning the growth of bicycle infrastructures. *Transp. Res. Part C Emerg. Technol* **115**, 102640 (2020).
46. Ouyang, M. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliab. Eng. Syst. Saf.* **121**, 43–60 (2014).
47. Wang, W., Yang, S., Stanley, H. E. & Gao, J. Local floods induce large-scale abrupt failures of road networks. *Nat. Commun.* **10**, 2114 (2019).
48. Moro, E. et al. Universal resilience patterns in labor markets. *Nat. Commun.* **12**, 1972 (2021).
49. Zhang, H., Liu, X., Wang, Q., Zhang, W. & Gao, J. Co-adaptation enhances the resilience of mutualistic networks. *J. R. Soc. Interface* **17**, 20200236 (2020).
50. Diao, M., Kong, H. & Zhao, J. Impacts of transportation network companies on urban mobility. *Nat. Sustain* **4**, 494–500 (2021).
51. Markhvida, M., Walsh, B., Hallegatte, S. & Baker, J. Quantification of disaster impacts through household well-being losses. *Nat. Sustain* **3**, 538–547 (2020).
52. Bureau, U. C. 2019 National and State Population Estimates. Available at: <https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html> (2019).
53. Chen, M. K., Chevalier, J. A. & Long, E. F. Nursing home staff networks and COVID-19. *Proc. Natl. Acad. Sci. USA* **118**, e2015455118 (2021).
54. Chen, M. K. & Rohla, R. The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020 LP–1021024 (2018).
55. SafeGraph. SafeGraph: Places Data & Foot-Traffic Insights. SafeGraph. Available at: <https://www.safegraph.com/> (2020).
56. United States Census Bureau. North American Industry Classification System (NAICS). United States Census Bureau (2017). Available at: <https://www.census.gov/naics/> (2021).
57. Open Street Map. OpenStreetMap. Available at: <https://www.openstreetmap.org/> (2021).
58. Pereira, R. H. M., Nadalin, V., Monasterio, L. & Albuquerque, P. H. M. Urban centrality: a simple index. *Geogr. Anal.* **45**, 77–89 (2013).
59. Resende, M. G. C. & Werneck, R. F. A fast swap-based local search procedure for location problems. *Ann. Oper. Res.* **150**, 205–230 (2007).
60. Douglas, P. H. The Cobb-Douglas production function once again: its history, its testing, and some new empirical values. *J. Polit. Econ.* **84**, 903–915 (1976).
61. Xu, M., Radhakrishnan, S., Kamarthi, S. & Jin, X. Resiliency of mutualistic supplier-manufacturer networks. *Sci. Rep.* **9**, 13559 (2019).
62. Jiang, J. et al. Predicting tipping points in mutualistic networks through dimension reduction. *Proc. Natl. Acad. Sci. USA* **115**, E639–E647 (2018).

ACKNOWLEDGEMENTS

This material is based in part upon work supported by the National Science Foundation under Grant CMMI-1846069 (CAREER), the Texas A&M University X-grant 699, and the Microsoft Azure AI for Public Health Grant. The authors also would like to acknowledge the data support from Veraset, Inc. and SafeGraph, Inc. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Microsoft Azure, SafeGraph or Veraset, Inc.

AUTHOR CONTRIBUTIONS

C.F., X.J., R.L., and A.M. designed the study. X.J., C.F., and R.L. implemented the method and empirical case study. A.M. support with data acquisition. C.F. and A.M. wrote the main manuscript. All authors reviewed and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42949-022-00051-3>.

Correspondence and requests for materials should be addressed to Chao Fan or Ali Mostafavi.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022