ARTICLE　　OPEN

Check for updates

Spatiotemporal heterogeneity reveals urban-rural differences in post-disaster recovery

Sangung Park[1✉], Tong Yao[2✉] and Satish V. Ukkusuri [1✉]

A post-disaster recovery process necessitates significant financial and time investment. Previous studies have found the importance of post-disaster spatial recovery heterogeneity, but the recovery heterogeneity has not been extended to the directed recovery relationships despite the significance of sequential recovery plans. Identifying a causal structure between county-level time series data can reveal spatial relationships in the post-disaster recovery process. This study uses a causal discovery method to reveal the spatiotemporal relationships between counties before, during, and after Hurricane Irma in 2017. This study proposes node aggregation methods at different time scales to obtain internally validated causal links. This paper utilizes points of interest data with daily location information from mobile phones and county-level daily nighttime light data. We find intra-regional homogeneity, inter-regional heterogeneity, and a hierarchical structure among urban, suburban, and rural counties based on a network motif analysis. Subsequently, this article suggests county-level post-disaster sequential recovery plans using the causal graph methods. These results help policymakers develop recovery scenarios and estimate the corresponding spatial recovery impacts.

INTRODUCTION

Natural disasters suspend normal human activities and cause severe community-level disruptions that exceed capacity in multiple ways[1]. Recovering the current conditions to pre-disaster conditions in many perspectives is called a post-disaster recovery (PDR) process[2]. There are two reasons why the PDR process needs high budgets. At first, large-scale catastrophes and subsequent high damages require high budgets[3] and workforce[4] for the PDR process. The other reason for demanding high budgets and workforce is that boosting the PDR process is affected by various costly but influential factors: (1) a functional state of physical infrastructure[5–8], (2) a functional state of social vulnerability[1,9–13], and (3) socioeconomic attributes[14–17].

The PDR process is usually modeled as a complex system due to the interconnected influential factors[5,7,9,18–22]. Previous studies have constructed the PDR process consisting of the physical network and the social network interwoven by influential elements[5–7]. However, it is hard to capture the functionality of each layer and the linkages across them. Instead, previous studies have used the proxies of two layers. The proxies of physical infrastructure are the moonlight-adjusted nighttime light (NTL) data[23,24], electricity data[5], and water deficit data[19]. The other proxy of social infrastructure is normalized visit density (NVD) to certain types of points of interest (POIs) using mobile phone location data[19,25,26] coming from the number of visits to POIs. NVD of POIs serves as an indicator of the community recovery, as community activity necessitates visits to POIs[27]. Previous studies have noted that the aggregation of human mobility is a key factor for the urban vitality index[28], which in turn is related to community activity, a main building block of the effective disaster recovery of communities[29]. Thus, the number of visits to POIs determines the community recovery. This study builds upon the literature and models the PDR process using two proxies of influential factors, NTL data and NVDs of POIs.

Previous studies have focused on correlations between influential factors but did not model causal relationships between influential factors[1,8,11–13]. The limitation of correlation analysis is that it often gives spurious outcomes in the complex systems. Spurious results occur when predicted results and tested counterfactual scenarios do not come from causal relationships[30–32] in the complex system[33]. This is mainly because correlation analysis does not capture confounders[32]. The confounder is a variable that simultaneously affects both a dependent and an independent variable. If we only revealed the associations among factors, we could not determine whether the associated factors are causally related due to the confounding effect. Consequently, we obtain ambiguous relationships among factors that produce spurious results. On the other hand, a causal relationship represents which node affects the other neighbors of the nodes and explains the directed relationship from the cause to the effect. The confounder can be captured by revealing directed relationships among influential factors. For example, in the second section of Fig. 1a, a confounder node X_{i1} only reveals the undirected node associations, but we can easily observe the confounding effects in the causally related nodes with directed links. Recent studies have revealed that confounding analysis is indispensable in the era of big data[30] and complex systems[33], especially in urban areas[34], due to many interconnected factors.

Through the causal relationship in the complex system, previous studies in several fields have inferred regionally heterogeneous outcomes, such as earth systems[35–37], human body[35,38], local sustainable development[39], and urban heat island[40]. These studies provided perspectives on the usage of causal discovery methods. Causal discovery is a method of revealing the causal relationship between various datasets without randomized controlled experiments, used in the machine learning community[30,41,42]. Causal discovery considers multivariate features in the time series data with causal graphs

[1]Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette 47907 IN, USA. [2]Elmore Family School of Electrical and Computer Engineering, Purdue University, 501 Northwestern Ave., West Lafayette 47907 IN, USA. ✉email: park1202@purdue.edu; yao127@purdue.edu; sukkusur@purdue.edu
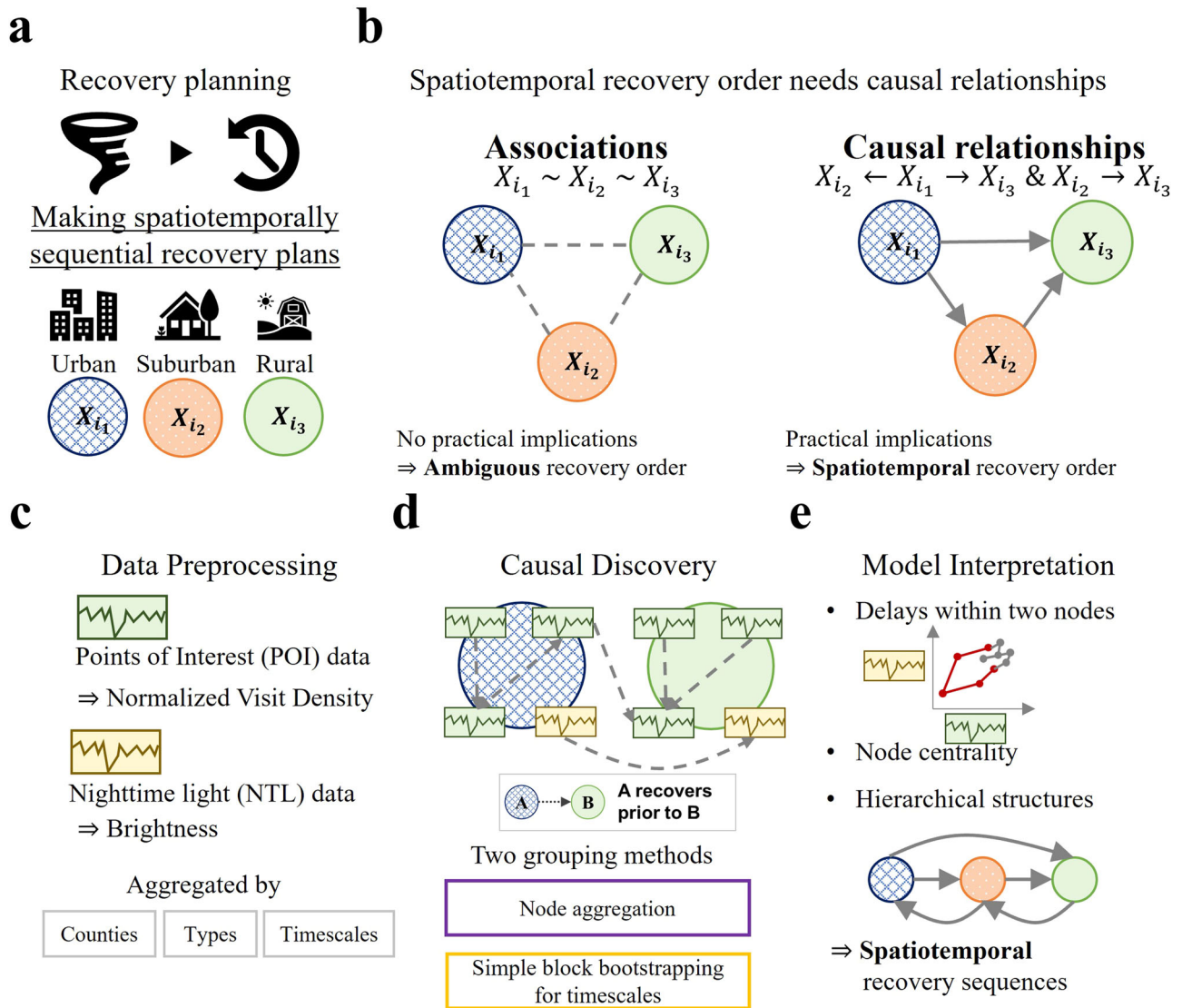
a

Recovery planning

Making spatiotemporally
sequential recovery plans

Urban   Suburban   Rural

X_{i_1} X_{i_2} X_{i_3}

b

Spatiotemporal recovery order needs causal relationships

Associations
$X_{i_1} \sim X_{i_2} \sim X_{i_3}$

X_{i_1} X_{i_3}

X_{i_2}

No practical implications
⇒ **Ambiguous** recovery order

Causal relationships
$X_{i_2} \leftarrow X_{i_1} \rightarrow X_{i_3}$ & $X_{i_2} \rightarrow X_{i_3}$

X_{i_1} X_{i_3}

X_{i_2}

Practical implications
⇒ **Spatiotemporal** recovery order

c

Data Preprocessing

Points of Interest (POI) data
⇒ Normalized Visit Density

Nighttime light (NTL) data
⇒ Brightness

Aggregated by

| Counties | Types | Timescales |

d

Causal Discovery

A ⇢ B A recovers prior to B

Two grouping methods

Node aggregation

Simple block bootstrapping
for timescales

e

Model Interpretation

• Delays within two nodes

• Node centrality

• Hierarchical structures

⇒ **Spatiotemporal**
recovery sequences

Fig. 1 Spatiotemporally sequential recovery plans by causal discovery methods and a framework of this study. a Descriptions of spatiotemporal recovery policy. A spatiotemporal recovery policy is suitable for disproportional recovery impacts between different types of counties. **b** Descriptions of associations and causal relationships. Spatiotemporal recovery policy requires causal relationships to gain the practical implications of the spatiotemporal recovery policy. **c** Data preprocessing of POI and NTL data. This study uses the POI dataset and NTL dataset to obtain the NVDs and brightness aggregated by counties, types of POIs, and different time scales. **d** Descriptions of causal discovery and two grouping methods. We utilize the causal discovery method to obtain causal links. We use two grouping methods to obtain internally validated causal links, the POI-level and county-level node aggregation and simple block bootstrapping merging causal graphs at different time scales. **e** Model interpretations. We interpret the graph union from three different perspectives: (1) interpreting delays within two nodes to validate the causal links, (2) node centrality to identify the spatial heterogeneity, and (3) hierarchical structures to reveal the causal relationships between different types of counties. Finally, we suggest spatiotemporally sequential recovery sequences.

represented by nodes (variables in time series data) and directed links (causal relationships between nodes). Through the causal graphs, we can trace the direct and indirect impacts of the nodes, as causal relationships inherently include the time-delayed relationship among regions[35,43]. This also holds true when we subsample or aggregate the time series data[43–45]. Previous studies have revealed potential causal links between regions and assessed the performance of causal discovery methods by comparing true positive correlation rates. Various methods have been employed, such as Granger causality for human brain regions[38], the PC algorithm for country-level air temperature and the elements of the human heart[35], and the graphical Lasso regression for continent-level precipitation[37] and country-level finance interactions[46].

Although different areas explored regional properties through causal discovery methods, previous studies in the PDR management area have not compared urban and rural properties with respect to heterogeneous spatiotemporal government PDR policies. Government PDR policy is a policy for the recovery action or effort from the local or federal governments to mitigate disaster damages. Through the government PDR policy, governments adjust the timing and intensity of the support efforts to meet the effective, efficient, and equitable recovery process[2]. Some examples missed the appropriate recovery period and suffered recovery failure[47–49].

Another critical aspect of government recovery policy is the recovery sequences between regions. In this paper, the PDR sequence is defined by the recovery sequence to determine the recovery order by county. Currently, the urban area owns the top

priority of the recovery due to the damage difference between urban and rural areas. Specifically, the physical and social infrastructure recovery cost of urban areas is higher than that of rural areas[50–52]. Sometimes, rural areas began the recovery with insufficient resources[50,53] by disaster resilience index-based approaches[25]. This leads to slower recovery in rural areas compared to urban areas, a discrepancy inherently arising from economic inequality[54]. Despite the importance of timing, intensity, and spatially sequential recovery plans, previous studies have not modeled the sequential and spatial impact of government policies on the PDR process.

Therefore, the spatiotemporal sequential recovery plan should be supported by the directed spatial relationships to consider confounding impacts. The sequential recovery plan is important in different types of areas, including urban, suburban, and rural counties. However, associations between regions cannot represent spatiotemporally sequential recovery plans, while causal relationships can represent spatial directions. In this work, we verify the effects of recovery sequences between urban, suburban, and rural areas with causal analysis. The general logic of the spatiotemporally sequential recovery plan is expressed in Fig. 1a.

In summary, this work has two motivations. First, methodologically, previous studies regarding PDR did not consider the causal relationships between various factors. Previous studies have focused only on intra-level associations between influential factors. Second, practically, previous studies have not focused on spatial heterogeneity in the disaster recovery process based on a data-driven spatiotemporal propagation analysis of PDR impacts.

To bridge these gaps, this study uses causal discovery methods to infer the PDR sequential recovery plans in the PDR process with disaster-related datasets from Hurricane Irma, 2017, for six months from the first day of July (see Fig. 2a). This article finds the causal graphs of different PDR-related datasets: (1) the NVD of certain types of POIs (see Fig. 2c) using mobile phone location data and (2) moonlight-adjusted nightlight data (see Fig. 2d). We use Structural Analytical Modeling (SAM), one of the causal discovery techniques, to infer the directed causal graph. Subsequently, this study uses two grouping methods, POI-level and county-level node aggregation and simple block bootstrapping at different time scales, to obtain the internally validated causal links of the different causal graphs (see Fig. 2b). Based on these results, we reveal the recovery directions between different types of POIs and find how spatiotemporal impacts affect the PDR process. We also obtain data-driven proof of the spatially sequential recovery plans by uncovering the existence of hierarchical structures in a causal graph union and extract the spatiotemporal propagation of the PDR impacts. This framework is illustrated in Fig. 1b.

RESULTS

We use SAM to model a causal graph $\mathcal{G}_{tp_m} = (\mathcal{V}_{tp}, \mathcal{E}_{tp})$ denoted by a set of nodes $\mathcal{V}_{tp}$ and a set of links $\mathcal{E}_{tp}$. Note that a node represents a county, and a link represents a causal relationship from one node to the other. Specifically, this paper defines the link from node 1 to node 2 as the recovery sequence from node 1 to node 2, meaning that node 1 begins to recover before node 2, and changes in node 1 directly affect node 2 (see Results section for details). We obtain causal graphs at different time scales in Fig. 3b–f. The upper part of each panel describes the results of county-level aggregation, and the lower part depicts the results of POI-level aggregation. This study uses aggregation of POI nodes at the county level and aggregation of county nodes at the POI level. The county-level node aggregation groups ten different POIs' NVDs into one county node. On the other hand, the POI-level node aggregation groups seven different county nodes into one POI node. In the aggregation step, we obtain internally validated links that imply the recovery sequence after Hurricane Irma, where the direction of a link represents the recovery sequence from a

node that recovers first to a node that recovers later. An internally validated link from the grouping methods is defined as an integrated causal link from one node to the other within different causal graphs representing repeated robust results. We draw the internally validated link based on graph union if and only if there are at least five links with the same node and relationships. Threshold five is selected to determine the internally validated link, eliminating 30.6% (56) of the causal links.

We observe two general patterns of internally validated links using POI types. The first pattern is the county-level intra-regional homogeneity which explains the homogeneity of POI trends in one county. We find that a causal graph union has 18.2% of the intra-regional internally validated links, and all causal graphs have more than 10% of the intra-regional internally validated links within eleven nodes (ten POIs and one brightness) from Fig. 3g. Note that the probability of drawing the intra-regional link in the random graph is 13.1%. The high proportions of intra-regional internally validated links indicate that the patterns of NVDs within one county are similar to each other. For further analysis, see Supplementary Table 1 in the supplementary material to check the spatial homogeneity by the correlation coefficients and Supplementary Table 2 in the supplementary material to check the Granger causality test results between types of POIs.

The second pattern is the consistent relationships between different types of POIs, (1) from school POIs to childcare services POIs, and (2) from restaurant POIs to entertainment POIs in Fig. 3a to f. The first type of relationship is interpretable since, during and after the natural disasters, school POIs can substitute childcare services POIs, but childcare services POIs cannot take the role of school POIs supported by the current role of childcare services in some public schools[55,56]. Therefore, policymakers recovered schools before childcare services. On the other hand, the possible explanation for the interdependence from restaurant to entertainment POIs is that restaurant POIs are an important type of POIs, while entertainment POIs are not necessary during and immediately after Hurricane Irma. Note that restaurant POIs are the necessary social capital for living, and entertainment POIs do not have the first priority for the disaster recovery process of the social capital. This relationship also holds before and after Hurricane Irma (See Supplementary Figs. 2 and 3).

Figure 4b also supports the delayed relationship between restaurant POIs and entertainment POIs during and after the disaster. The red trajectory represents the period around Hurricane Irma from September 8th to September 15th. This red line forms a polygon whose upper part shows the period before and during Hurricane Irma and whose lower part shows the period after Hurricane Irma. Specifically, if the red line has a linear relationship, it represents the prompt and same responses between restaurant and entertainment POIs before, during, and after Hurricane Irma. However, the formation of a polygon indicates a relationship change between two different periods: the pre-disaster period and the post-disaster period. Specifically, in Fig. 4b, the two types of POIs simultaneously decrease in the pre-disaster period, which can be attributed to the equal impact of the hurricane on the two types of POIs. Subsequently, in the post-disaster period, the recovery of restaurant POIs is faster than that of entertainment POIs, suggesting a delayed relationship between restaurant and entertainment POIs. Figure 3a to f further support the delayed relationship between restaurant and entertainment POIs.

We obtain universal patterns of the spatial relationships between county nodes based on the graph union (Fig. 3a) over different periods (Fig. 3b–f). We obtain 13 internally validated links from the graph union in Fig. 3a. These internally validated links represent causal relationships in different counties and describe the propagation of the recovery process between counties.

The county nodes are classified into three different types by population density, (1) urban, (2) rural, and (3) suburban counties
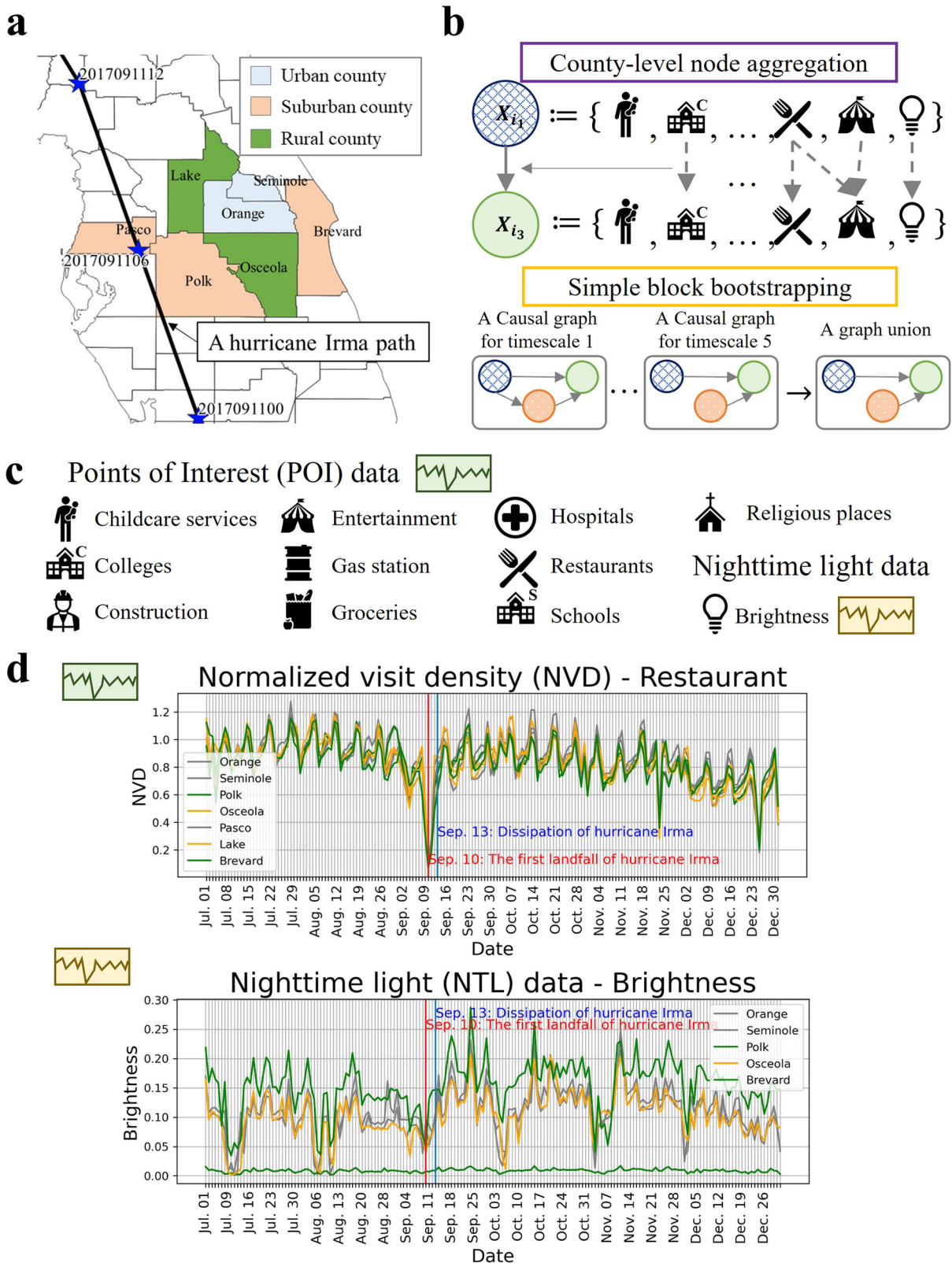
Fig. 2 Descriptions of datasets and grouping methods. a Descriptions of the spatiotemporal scope of Hurricane Irma, 2017. The black line represents the path of Hurricane Irma. Colored regions identify different types of counties, urban, suburban, and rural counties. The coordinate of the center of this image is N 28°32'1" E 278°33'5" with CRS EPSG:4269 - NAD83. **b** Explanations of two grouping methods. We use two grouping methods. One is the county-level node aggregation to obtain internally validated links within one county. The other method is the simple block time series bootstrapping to have internally validated links between different counties. If the number of the same links is not less than 5, we treat it as an internally validated link for stable results. **c** 10 focused types of POIs and brightness. We select ten different types of POIs and brightness related to disaster recovery. **d** County-level time series data. The upper figure describes county-level time series data of normalized visit density on restaurants. The lower plot describes county-level time series data of brightness.
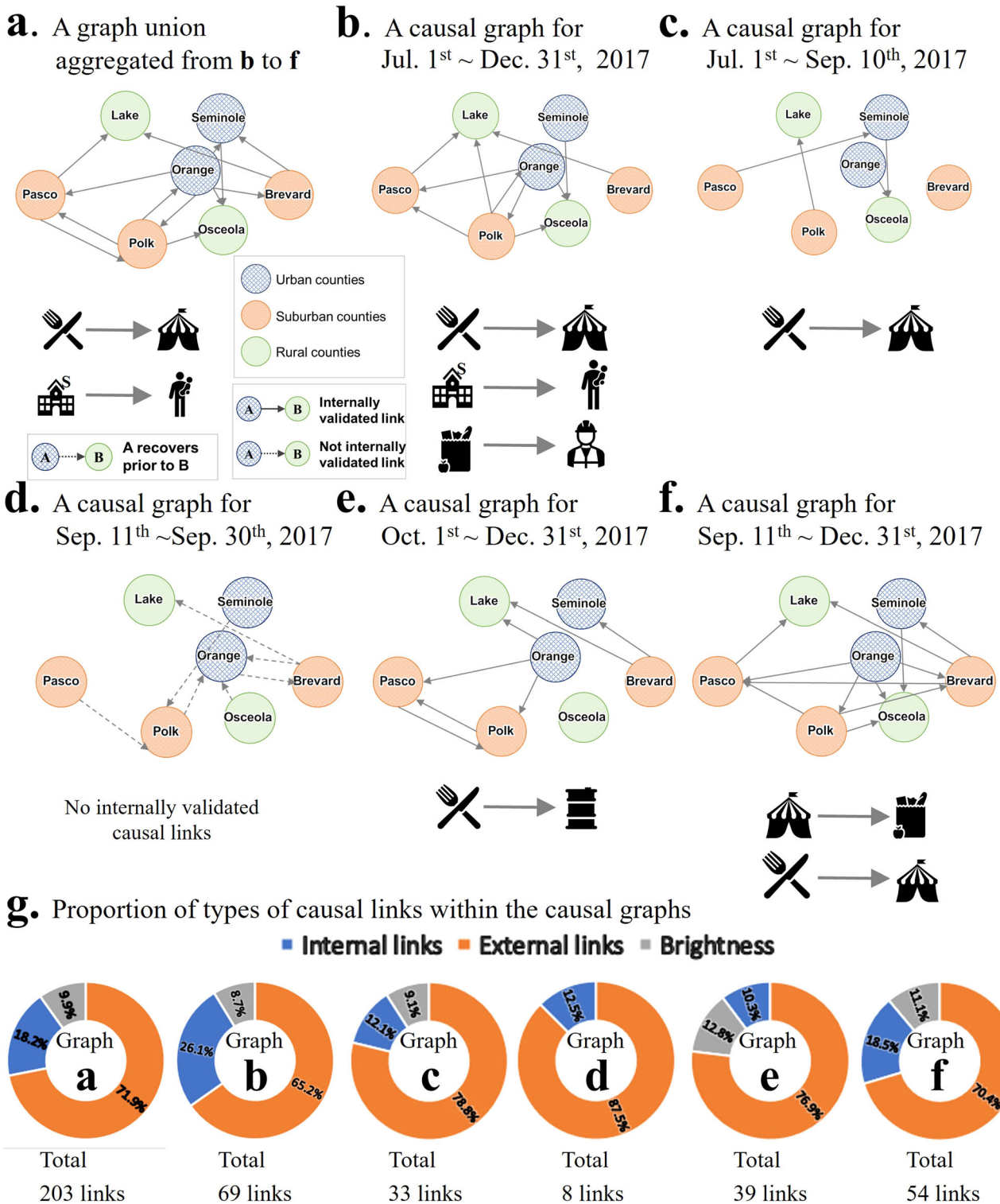
Fig. 3   Causal graphs at different time scales. a A causal graph union aggregated from **b** to **f**. **b** A causal graph for July 1st, 2017 - December 31st, 2017 (tp_1). **c** A causal graph for July 1st, 2017 - September 10th, 2017 (tp_2). **d** A causal graph for September 11th, 2017 - September 30th, 2017 (tp_3). **e** A causal graph for October 1st, 2017 - December 31st, 2017 (tp_4). **f** A causal graph for September 11th, 2017 - December 31st, 2017 (tp_5). Panels (**b**) to (**f**) in this figure have two parts. The upper part shows the county-level aggregation results, and the lower part describes the POI-level aggregation results. **g** The proportion of types of causal links of causal graphs from panels (**a**) to (**f**). Intra-regional links represent the links within a certain county. On the other hand, inter-regional links represent the links between two counties.

(see Methods section). We can clearly see the heterogeneity by the types of counties. For example, Fig. 3a shows the interdependence of Orange County with Seminole, Osceola, Brevard, Polk, and Pasco counties. Specifically, we observe the spatial pattern that urban counties' out-degree centrality is greater than or equal to the in-degree centrality, while rural counties have a higher in-degree centrality than out-degree centrality in Fig. 4a and c also support the spatial heterogeneity of county-level
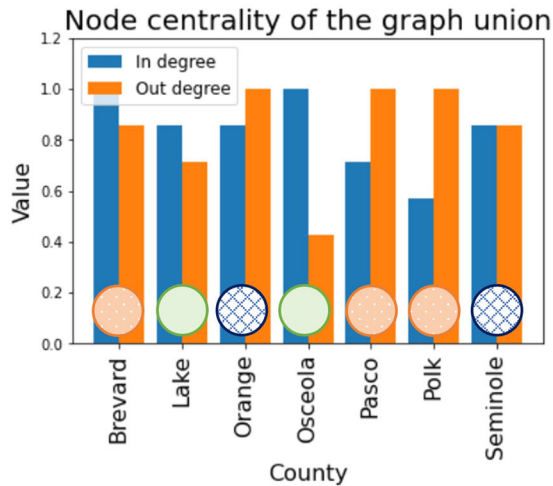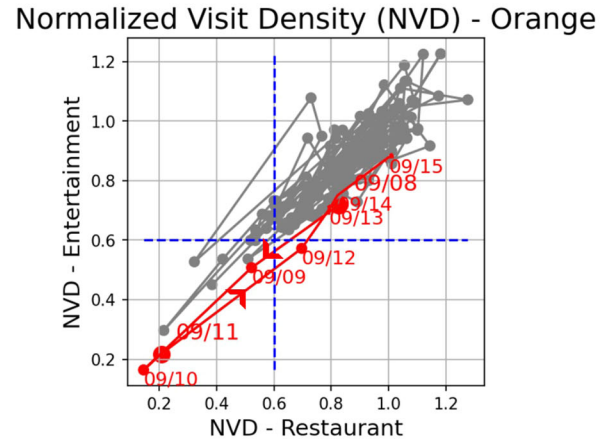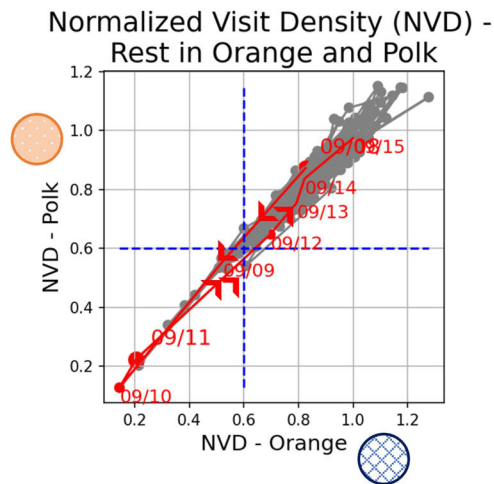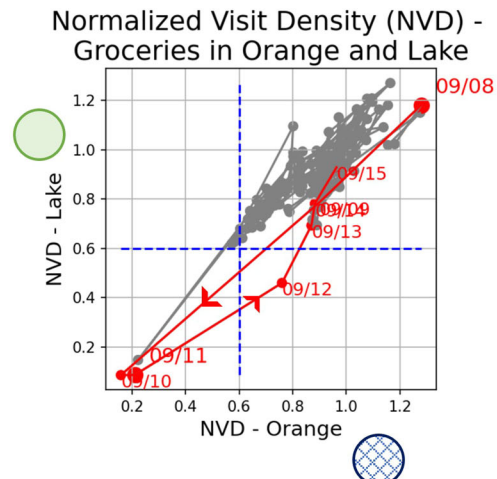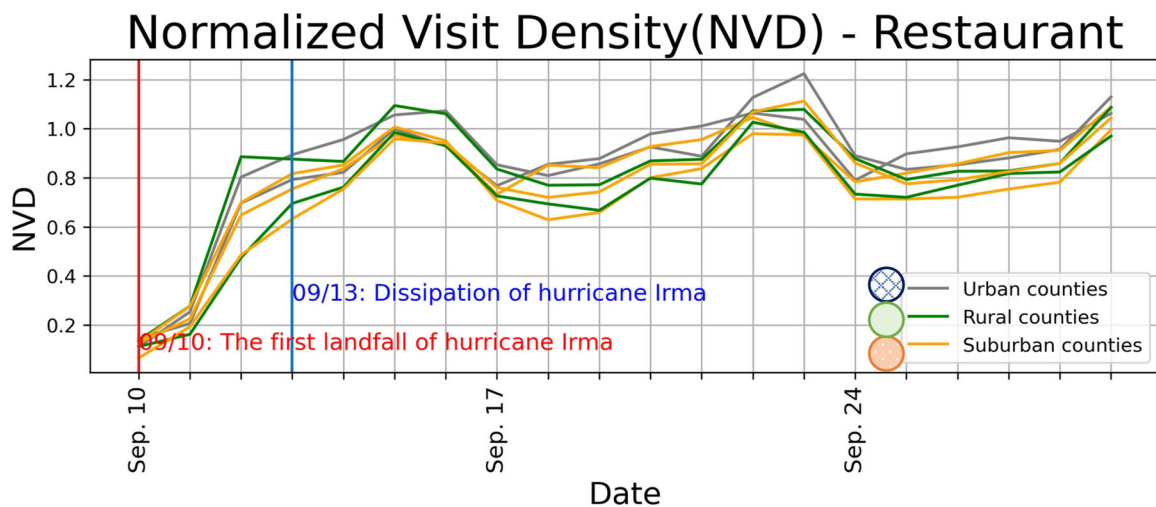
Fig. 4 Model interpretations. a Node centrality of the graph union in Fig. 3a. We compare the in-degree and out-degree of the graph union. **b** A trajectory plot over time between NVD on restaurants and entertainment in Orange County. We compare two different types of POIs in Orange County to validate the causal relationship. A red line represents the points during and right after the disaster. **c** A trajectory plot over time between NVD on restaurants in Orange County (Urban) and Polk County (Suburban). **d** A trajectory plot over time between NVD on restaurants in Orange County (Urban) and Lake County (Rural). **e** NVD time series data on restaurants by three different types of counties.

recovery patterns over time. Also, the polygons in Fig. 4c and d explain the interdependence between two counties, like the explanations in the 4b.

The last pattern we observe is the independence between the NTL data and the POI data. The correlation coefficients between POI and NTL data are less than or equal to 0.2 (see Supplementary Table 1 in the supplementary material). Also, the Granger causality test between POI and NTL data does not make any causal links between them (see Supplementary Table 2 in the supplementary material). It implies that the NTL and POI data are uncorrelated. Likewise, the causal discovery methods do not give any implications between NTL data and POI data. This independence implies that the proxies of the physical and social infrastructure are irrelevant. The contrasting independence can be explained for two reasons. The first reason is that the NTL dataset is not a good indicator of representing daily changes in physical infrastructure. For example, the NTL dataset only visualizes the intensity of daily night light, so the data can be unstable due to cloudy conditions and cannot quickly represent the recovery states.

The other reason for the contrasting independence is that Hurricane Irma did not severely damage the county compared to the other hurricanes. For example, counties in Florida overcame most of the power outages caused by Hurricane Irma in one week, compared to Puerto Rico counties that were severely damaged by Hurricane Irma and Maria consequently[23]. We have discussed the two reasons in detail in the Discussion section.

The above findings on spatial heterogeneity of recovery only compared the in-degree and out-degree centrality between different types of counties. However, they did not give us the overall spatially sequential recovery plans among the types of regions. We find the hierarchical structure of the causal graph union by network motif analysis to reveal the overall spatially sequential recovery plans. Network motif analysis is used to find structural properties by looking at the statistically frequent triads of the entire network[57–59]. Previous studies have stated that the high frequencies of certain types of network motifs imply the existence of a hierarchical structure in the network[57]. There are three network motifs positively related to the hierarchical structures in Fig. 5a: (1) double-dominant (DD), (2) double-subordinate (DS), and (3) transitive triad (TT)[57]. On the other hand, frequent pass-along (PA) network motifs support the nonexistence of the hierarchical structure.

We compare the frequencies of the network motifs with the mean frequencies of the network motifs of 100 null graphs generated from the directed Erdos-Renyi random network model with 7 nodes and link probabilities of 0.2[60]. From[61], we use the normalized z-score to compare the frequencies of network motifs as follows statistically:

$$Z_i = \frac{F_{target,i} - \mu_{null,i}}{\sigma_{null,i}}, \qquad (1)$$

where Z_i is the normalized z-score of the i-th network motif, $F_{target,i}$ is the frequency of the i-th network motif on the target graph, $\mu_{null,i}$ is the mean of network motifs on the null graph, and $\sigma_{null,i}$ is the standard deviation of network motifs on the null graph model. The results of network motif analysis in Fig. 5b show the high z-score of DD and DS and the low z-score of PA. It supports the existence of a hierarchical structure. Figure 5c is an example of a hierarchical structure in the graph union. The graph union can be classified into urban, suburban, and rural counties. The urban county class emits the causal link to the other classes of counties, suburban and rural counties. The rural county class receives the link from the urban and suburban classes. The suburban county class interacts with both classes. The overall direction of the graph union goes from urban counties to suburban and rural counties.

(1) Intra-regional homogeneity, (2) inter-regional heterogeneity, and (3) the hierarchical structure revealed from the causal graph

union motivate us to establish the data-driven spatiotemporal PDR sequences and propagation of spatiotemporal recovery policy. First, Orange and Seminole counties are the most important counties for PDR based on the hierarchical structure in the graph union. Changes in Orange and Seminole counties directly affect almost all nodes except Lake, and all trends are caused by Orange in this study. Second, the recovery order of Osceola and Lake is relatively low compared to the other counties. Third, Pasco, Polk, and Brevard counties interacted with urban and rural counties. Therefore, we can set the recovery sequence as {Orange, Seminole, (Pasco, Polk, Brevard), Osceola, and Lake}. Note that a brace represents a sequence that has an order within elements. For example, {A, B, C} means that A starts recovery first, and B and C start recovery sequentially. Note that the elements of the brace in the inner parentheses can be interchangeable. On the other hand, we can also set up the recovery sequences by types of POIs, {(Restaurants, Schools), (Entertainment, Childcare services)} since changes in restaurants affect entertainment POIs and changes in schools highly influence childcare service POIs.

DISCUSSION

This study utilized SAM, one of the causal discovery methods, to reveal the spatiotemporal relationships between counties. This study extracted a causal graph of NVD to certain types of POIs and NTL data for periods before, during, and after Hurricane Irma in 2017. This study has shown (1) intra-regional similarity, (2) inter-regional dissimilarity, and (3) a hierarchical structure of PDR between regions to help policymakers set up spatiotemporally sequential recovery plans in PDR.

First, results show the intra-regional similarity in the county-level aggregation. Universal trends of POIs in each county are similar, although they have different properties by the types of POIs. Specifically, the intensity and timing of the PDR process by NVD within the same county are similar in Fig. 4b, supported by the previous studies[27,62]. Second, we found the inter-regional similarity by types of POIs. Each type of POIs has similar patterns between all the selected counties. For example, restaurant POIs between different counties share similar patterns in Fig. 2d. Universal POI interdependence from restaurant to entertainment POIs and from school to childcare service POIs, is observed in Fig. 3a to f and 4b. It is also supported by previous studies in different disasters, e.g., Hurricane Harvey in 2017[27]. The universal interdependence among POIs implies that the prioritization and necessity of various types of POIs are consistent in the PDR process, as illustrated by the recovery sequences for different types of POIs in Results section. The interdependence can serve as a foundation for developing disaggregated spatiotemporal recovery plans tailored to specific types of POIs.

Third, we found the regional dissimilarity between urban, suburban, and rural counties in the PDR process. This regional dissimilarity is a basis for explaining the urban-rural differences of the PDR process. We observed the regional dissimilarity of in-degree and out-degree by types of counties in Fig. 4a. Also, it is supported by Fig. 4b and e by explaining the spatial heterogeneity and the delays of county-level recovery patterns. It can be identified by the distinct dissimilarity of socioeconomic attributes between urban, suburban, and rural counties in Table 1, such as the populations and budgets they allocated for Hurricane Irma. This observation can be assisted by the previous studies and perspectives of urban-rural differences[53,63,64]. The delayed relationships among three types of counties extend the interdependence among them to the regional dissimilarity in the PDR period, shown by Fig. 4c and d. Therefore, the regional dissimilarity and the delayed relationships support the internally validated causal graphs in the PDR process, indicating that the NVDs in urban counties began recovering before the NVDs in rural counties.

a. Network motifs of hierarchical structures
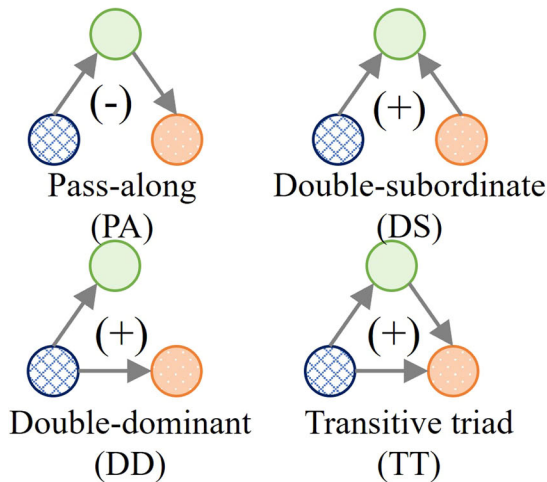
b. Normalized z-score on network motif analysis
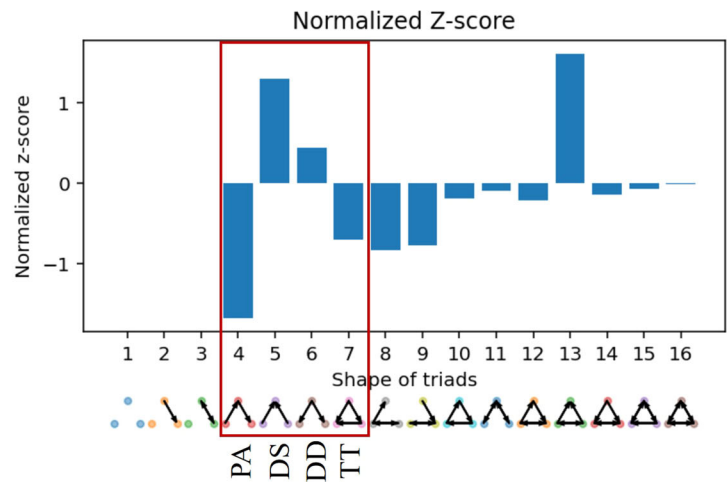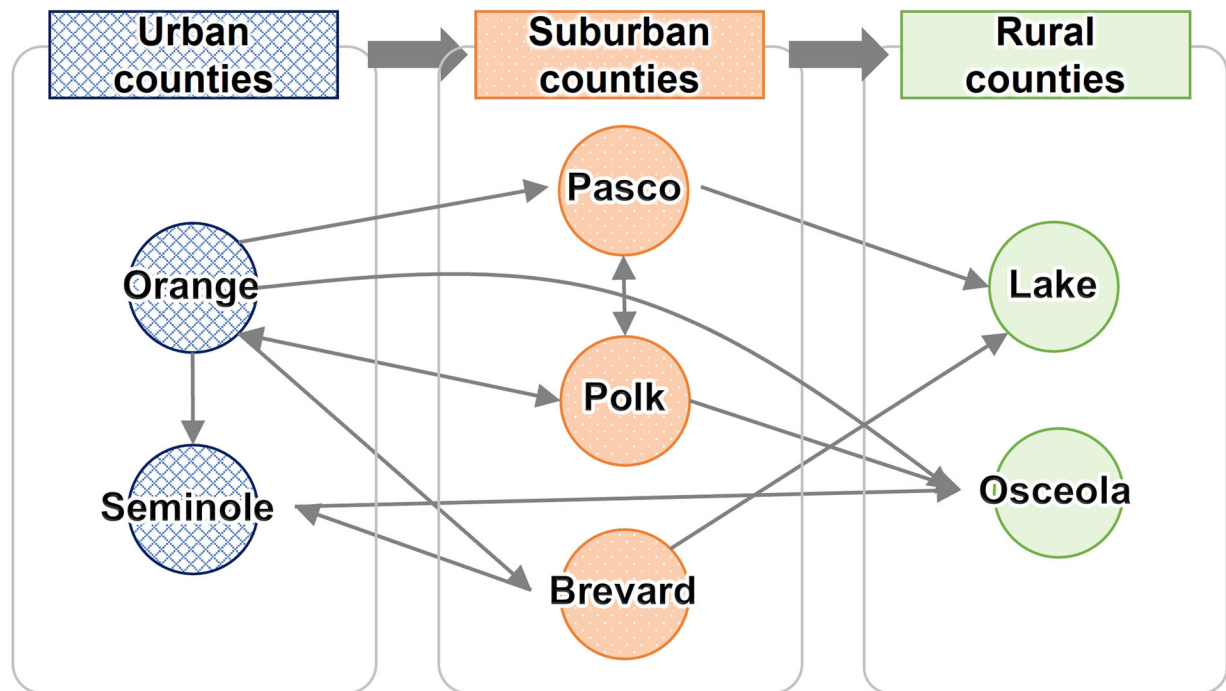

c. A hierarchical structure in a graph union

Fig. 5 A hierarchical structure in the causal graph union. a Network motifs of hierarchical structures. There are four types of network motifs related to the hierarchical structure. PA is negatively related to the hierarchical structure, and the others are positively related to the hierarchical structure. **b** Normalized z-score on network motif analysis. We extract all types of triad network motifs and compared them to the networks generated by the Erdos-Renyi model. Red boxes are network motifs related to the hierarchical structure. **c** A hierarchical structure in a graph union. We relocate all nodes by different types of counties and observe the hierarchical structure with few exceptions in the suburban counties.

Furthermore, this study has shown the existence of the hierarchical structure in the causal graph union. The network motif analysis in Fig. 5b inferred the existence of the hierarchical structure. Figure 5c showed a candidate of the hierarchical structure classified by urban, suburban, and rural county levels. The graph union also has internally validated causal links from urban to suburban and rural counties. This direction implies the temporal dependence of POI recovery between county types.

The independence between the NTL data and the POI data can be explained for the following reasons. Since Hurricane Irma did

not cause severe damage to the physical infrastructure system of selected counties due to disaster preparedness, the selected counties did not suffer severe power outages and quickly recovered the original electrical functionality. Therefore, NTL data, proxies of electricity outage, appear independent of POI data, a proxy of social infrastructure. Regarding hurricanes with severe damage, we can observe the evident interdependence between the NTL data and the disaster recovery process[1,23].

The findings of this study have significant implications for setting up spatiotemporally sequential PDR plans. The recovery

Table 1. Sociodemographic attributes of selected counties.

Name	μ (thousands)	A (square mile)	μ / A	Damage (millions)	Budget (millions)	Budget / μ
Seminole(U)	422.7	310	1366	6.62	9.60	22.7
Orange(U)	1146.0	903	1269	15.97	20.75	18.1
Pasco(S)	464.7	748	622	6.24	9.78	21
Brevard(S)	543.4	1015	535	16.42	16.25	29.9
Polk(S)	602.1	1797	335	22.58	24.73	41.1
Lake(R)	297.1	941	316	5.66	7.02	23.6
Osceola(R)	268.7	1328	202	7.70	6.01	22.4

The alphabet letters after the name represent the types of counties, U for urban counties, S for suburban counties, and R for rural counties. μ denotes the population of one county, and A represents the area of one county (square mile). Note that damage and budget come from housing assistance program data for owners collected by FEMA.

sequence plan is a good foundation for efficient and fast recovery policies. Targeted recovery policies can be developed when policymakers allocate the budget to a group of the nodes affecting a certain county (parents). For example, an increase in NVD in Orange and Polk counties promotes the recovery of Pasco County. Furthermore, Osceola County is directly affected by Orange, Seminole, and Polk counties, but not Brevard or Pasco County. Accordingly, to boost the recovery of Osceola County, we can promote the recovery of interconnected counties - Orange, Seminole, and Polk counties, including Osceola County itself.

This study suggests two different directions of spatiotemporal recovery plans from the above findings. The first direction is to build upon the current recovery policies with special considerations for rural areas. In other words, this policy starts the PDR from urban counties to suburban and rural counties. Simultaneously, this policy allocates PDR support to the selected suburban and rural areas that are highly likely that insufficient recovery actions will be provided. Adjusting the recovery speed by types of counties is necessary to alleviate the consequent dangers of hurricanes between regions. The results of this study, the causal graph unions, identify the suburban or rural areas that are not spatially related to the urban areas in the PDR process. Those suburban or rural areas will be candidates to be considered with special considerations for rural areas. Also, policymakers can establish recovery priorities when the resilience index and the extent of damage in suburban and rural areas are similar. The amounts of budgets to be allocated for the suburban or rural areas are the future direction. Merging considerations for rural areas with current recovery policies is straightforward as policymakers already allocate the budgets based on the fundamental disaster-related statistics and vulnerability indices of counties, such as population density, hurricane damage estimates, and social vulnerability index[65], and it naturally yields the urban area first recovery policy. Currently, this pattern can be managed through proactive strategies in the suburban and rural areas during the pre-disaster preparedness[66,67], but no such strategies exist in the PDR process. Adding considerations for rural areas to the current PDR process might simplify its implementation.

However, there exists room for more efficient recovery strategies than the first direction. For example, we do not know the holistic recovery impact of one county from the other counties through the first direction. When a certain county is severely damaged, we usually focus on the county itself and ignore the impact of other counties. However, certain types of POIs are interconnected within the county and the other neighboring counties. It implies that supporting the recovery of other neighboring counties indirectly helps the county boost the recovery.

Therefore, we propose an alternative to enhance the PDR of the targeted county by leveraging the indirect recovery impacts from neighboring counties. To boost the recovery of the targeted county, policymakers should initiate the recovery of counties directly affecting the targeted county, similar to the spatiotemporally sequential PDR plans. There are three steps to establishing the spatiotemporal sequential PDR plans. The first step is to identify the counties impacting the recovery of POIs by using the causal graph union. Second, policymakers select the county in urgent need of the PDR process. The final step is to allocate the recovery budgets for the targeted county as well as counties affecting the targeted counties. By considering the indirect impacts of neighboring counties, we can boost the PDR process of the target county. This approach will help policymakers mitigate the recovery equity issues in PDR[68-70].

This work is suitable for organizing the county-level PDR process in an era of frequent natural hazards. Recent Intergovernmental Panel on Climate Change (IPCC) reports have warned of the concurrent and repeated natural hazards of climate change driven by human activities[71]. Simultaneously, IPCC stated that the impacts of natural hazards disproportionately affect people and cities vulnerable to natural hazards[71], leading to the regional difference among various types of regions and countries. United Nations Office for Disaster Risk Reduction (UNDRR) also highlighted that disadvantaged groups disproportionately incur inequality due to the vicious cycle of climate-related disasters[72], making disadvantaged groups vulnerable to natural hazards. However, IPCC and UNDRR cannot organize the fast PDR process in sustainable and resilient infrastructure systems and communities at the county level. This study provides spatiotemporally sequential PDR plans which can boost the PDR process by using mobile phone location data to estimate the state of POIs. Since frequent natural hazards require stakeholders to achieve a faster PDR process than before, spatiotemporally sequential PDR plans can serve as an alternative to current recovery policies.

This study elucidates the new usage of mobile phone location datasets that have not been used for understanding the hierarchy of the PDR. Current opportunities of mobile phone location datasets for the natural hazards are population displacements and evacuation modeling, analyses on migration and recovery, and damage estimates[73]. This study inspects the spatiotemporal relationships of POIs by counties, which are easily overlooked but are still crucial in the complex dynamical systems of the PDR process. The usage of the mobile phone location dataset empowers revealing the spatiotemporal relationships that cannot be detected before.

This study scientifically contributes to understanding spatiotemporal heterogeneity and how urban counties affect suburban and rural counties in PDR. In particular, state-of-the-art causal discovery methods are used to identify causal relationships between regions. The causal discovery method is also applicable to modeling any inter-regional dependence from the mobile

phone location data. Using this method, we can identify how the PDR process spatially disseminates over time. This work also addresses the importance of selecting suburban counties to explain spatial heterogeneity clearly. Furthermore, this study proposes two node grouping methods to clarify the causal relationships and obtain internally validated causal links. These grouping methods help us manage large time series datasets and raw causal links. They also enable the creation of causal graph unions, categorized by different POIs and time scales. Moreover, we reveal the hierarchical structure in the PDR process by network motif analysis. We derive suggestions for spatiotemporally sequential PDR plans from the causal graph union and the hierarchical structure.

We summarize the validation of this study from similar case studies. First, we found the county-level intra-regional and POI-related patterns from Hurricane Harvey[27]. The relationships between childcare services and schools are supported by public school cases[55,56]. Second, a systematic review of urban-rural differences in PDR supports the regional recovery heterogeneity among types of counties[53]. Third, the recovery priority from urban to rural areas is emerging due to the damage difference affected by hurricanes[50–52].

There are still some limitations to this study. First, this study lacks a measure for model effectiveness as it does not include a ground truth for the causal graph. Using data fusion on the survey dataset regarding the causal relationships between POI types, we can obtain a sample of a ground-truth causal graph and estimate the model. Second, this study focuses solely on counties with minor damage, excluding those with severe or complete damage to POIs and housing. Although Hurricane Irma was one of the most damaging storms, Orlando and its neighboring counties were not severely damaged due to their preparedness for hurricanes based on previous critical experiences on hurricanes. As the objective of this study is to explain the urban-rural differences, we limited our dataset to counties with minor damages. With a dataset on severely damaged urban and rural counties, we can further examine whether the interdependence exists between physical infrastructure and social networks from the causal discovery method. Third, we do not model and quantify the intra-regional and inter-regional recovery impacts. This work only suggests a candidate for the spatiotemporally sequential PDR plans based on time series data. Future research will focus on developing metrics for regional interdependence and integrating them into data-driven PDR planning models.

This research serves as one of the building blocks of the inter-regional PDR dynamics to evaluate the spatiotemporally sequential recovery plans between regions quantitatively. Furthermore, this work will help policymakers develop spatiotemporally sequential PDR plans with limited budgets and a limited workforce.

METHODS

Disaster information and spatiotemporal scopes

We select Hurricane Irma as a target disaster to explain the causal relationships between urban and rural counties. Hurricane Irma was a Category 5 hurricane that hit Florida in the United States on September 10, 2017, and dissipated on September 13, 2017. FEMA provided a housing assistance program by Hurricane Irma in Florida for homeowners worth $414 million[3]. We handle datasets covering the six months from July to December 2017 to investigate the impacts of Hurricane Irma. We point out that the six-month period is sufficient to describe the PDR process from the mobility patterns of POIs[27]. Mobility patterns from the previous disasters returned to normal within two months - see Supplementary Figure 1.

We select a cluster of urban and rural areas to analyze the inter-regional recovery dependence. Orlando is the largest city affected by Hurricane Irma in central Florida. Therefore, seven counties are selected including Orlando and its surrounding counties: Polk, Orange, Osceola, Seminole, Pasco, Lake, and Brevard counties.

We choose urban, suburban, and rural counties based on the population density of 1,000 people per square mile shown in Table 1[74]. If the population density of a county is greater than 1,000, that county is considered an urban county. Rural counties are selected when the population density is less than 500. The urban counties are Seminole and Orange counties, and the rural counties are Lake and Osceola. Suburban counties in this paper are defined as counties that have both urban and rural counties' characteristics. We especially treat Pasco and Brevard counties as suburban counties since most of these counties are densely populated, but the county's population density is less than 1,000. On the other hand, Polk County is also selected as the suburban county since most of the county is sparsely populated, but part of this county contains a highly densely populated area.

Datasets

Daily moonlight-adjusted nighttime light (NTL) data represent the daily nighttime brightness of the surface by Suomi National Polar-orbiting Platform/Visible Infrared Imaging Radiometer Suite (SNPP/VIIRS), preprocessed by National Aeronautics and Space Administration[23,24]. This dataset contains daily NTL monitoring results at a 500-meter resolution. We utilize NTL data as the brightness of certain regions at night and as an indicator of electricity after the disaster for the physical network. The lower part of Fig. 2d shows the county-level aggregated NTL data over time.

We use the daily visit dataset of POIs to illustrate the social network in the multilayer network structure. The POI visit dataset is provided by Safegraph Inc. (https://www.safegraph.com). POI datasets are commonly used to infer daily human activity[26,75]. The POI visit dataset is made up of the GPS location information dataset of mobile phones and smartphones in the United States through various apps. It consists of detailed information on POIs, including name, location, types, and brand of POIs based on North American Industry Classification System (NAICS) codes, and the number of visits by day based on GPS location data by counting the number of visits only if the visitor stays there for more than four minutes.

We select 10 different classes of POI to represent PDR: child care services (children), colleges, construction, entertainment, gasoline, groceries, hospitals, religious places (religion), restaurants, and schools[19,27]. We use J to represent ten different types of POIs.

We convert the number of visits of devices by the census block group to the normalized visit density (NVD) by county. First, we normalize the number of daily visits of mobile devices by county and POI information:

$$X_{ijk}^{Target} = \frac{X_{ijk}}{\sqrt{A(i)}} \times \frac{P(l)}{D(l)}, \tag{2}$$

where X_{ijk}^{Target} is the visit density of the k-th POI in the j-th POI class in the i-th county, X_{ijk} is a sequence of the number of daily visits of mobile devices to the k-th POI in the j-th POI class in the i-th county, $A(k)$ is the area of the k-th POI, $P(l)$ is the population in the l-th census block group, $D(l)$ represents the number of devices in the l-th census block group.

Subsequently, we normalize the visit density (X_{ijk}^{Target}) by the POI classes:

$$X_{ijk}^{Norm} = \frac{X_{ijk}^{Target}}{\sum_{j \in J} X_{ijk}^{Base} / n(J)}, \tag{3}$$

where X_{ijk}^{Norm} is a sequence of NVD of the k-th POI in the j-th POI class in the i-th county, X_{ijk}^{Base} is the normalization factor for X_{ijk}^{Norm} by the j-th class of POIs in the i-th county, $n(j)$ is the number of POIs in the j-th class, and J is a set of the selected classes of POIs. Normalization is composed of an aggregation of data from the j-th class of POIs in the i-th county in March, July, and August 2017.

Finally, we use aggregated POIs X_{ij}^{Norm} by types of POIs:

$$X_{ij} := X_{ij}^{Norm} = \sum_{k \in K} \frac{X_{ijk}^{Norm}}{n(k)}. \qquad (4)$$

Here, X_{ij}^{Norm} is the mean visit density to j-th POI class in the i-th county. We simply suggest X_{ij} as the abbreviation of X_{ij}^{Norm}. We let $K = \{k_1, k_2, k_3, \ldots, k_{10}\}$ represent a set of POIs of the same class.

Causal discovery (CD) methods

Causal discovery (CD) methods are defined as an alternative to forward causal inference. Conventional forward causal inference aims to quantify causal effects using average treatment effects ($E[y|t = 1] - E[y|t = 0]$), where $t = 1(0)$ denotes a treatment (control). The input data for forward causal inference have features (X), treatment (t), and outcomes (y). Furthermore, forward causal inference does not necessarily have ground-truth causal graphs regarding the dataset.

On the other hand, the CD method aims to select causal links among X by checking whether X_i causes X_j, $X_i \rightarrow X_j$, $\forall X_i, X_j \in X$. CD can infer causal graph $\mathscr{G}$ with high-dimensional observational time series data without experimenting with treatment t. It would be helpful to give a ground-truth causal graph to evaluate the model, but it is not required to acquire that graph. This study aims to select causal links among various PDR-related datasets, so we choose the CD method to obtain causal links between regions and POI types.

There are four assumptions for the CD methods to rigorously match the causal graph with the real causal structures[41]. The first assumption is the causal Markov assumption, meaning that the noise variables of all nodes are assumed to be independent of each other except for the direct causes (called parents) of each node. The second assumption is the causal faithfulness assumption, implying that causal graphical structures must include all conditional independence relationships. The third assumption is the causal sufficiency assumption, which means that there is no external confounder outside of the causal graph. The last assumption is the acyclicity of the causal graph. The causal graph must be a directed acyclic graph that does not have a cycle in the causal graph.

In this paper, we assume the underlying generative model of the data is a Functional Causal Model (FCM)[32]. Let $X = \{X_1, \ldots, X_d\}$ denote a set of continuous random variables whose causal relationships are of interest. We let FCM be defined by a pair $(\mathscr{G}, f)$. Here, $\mathscr{G}$ is a directed acyclic graph (DAG) with nodes denoted by X and $f = (f_1, \ldots, f_d)$ is a set of d causal mechanisms. In FCM, f can be non-linear or non-parametric. For each variable $X_j \in X$, we assume it follows the distribution

$$X_j = f_j(X_{Pa(j;\mathscr{G})}, E_j), \qquad (5)$$

where $Pa(j; \mathscr{G})$ is the set of parents of X_j in $\mathscr{G}$ and E_j is a random noise variable characterizing the effect of non-observed variables. If X_i is a parent of X_j, i.e., $X_i \in Pa(j; \mathscr{G})$ and $X_i \rightarrow X_j$, variable X_i (the cause) has a causal influence on X_j (the effect) and vice versa.

To learn the DAG representing the acyclic FCM from observational data, we use SAM with the Causal Discovery Toolbox containing the latest CD algorithms[41,42]. Based on adversarial neural networks, SAM leverages both conditional independencies and distributional asymmetries induced by the causal direction, achieving trade-offs between model complexity and data fitting. The input of SAM is high-dimensional time series data representing each node of the causal graph, and the output of SAM is the weighted adjacency matrix to represent the averaged strength of the causal relationship[41,76]. The global score for this model has a minimax penalized optimization problem considering model complexity and data fitting loss. This score is executed by a stochastic gradient algorithm with respect to the set of parameters and structural gates to compensate for both model complexity and data fitting loss problems that are commonly raised in the rest of the models. Finally, the algorithm produces an acyclic causal graph (for more details, see[41] and Supplementary Figure 4 for the parameter selection process of SAM).

A significant difference of SAM from the other CD methods is the utilization of a generative adversarial neural network to fit the loss function. Putting the variable in the structural gates, SAM designs the conditional generative neural network for the causal mechanisms of a set of nodes. SAM is adequate and accurate in identifying nonlinear causal structures between continuous variables from observational data.

Node aggregation

There are two types of node aggregation, county-level node aggregation and POI-level node aggregation. The upper side of Fig. 2b describes the county-level node aggregation. The raw results of the causal graph yield 11 different nodes in one county, including 10 types of POIs and brightness. These raw results show unclear relationships between regions since some links only represent specific causal relationships between nodes. Therefore, we aggregate all types of POIs and brightness into one county node and draw internally validated links.

First, we use mathematical expressions to explain county-level node aggregation. Define $X_{i,all}$ as the node representing the i-th county and $X_{i,all} = \{X_{ij}, \forall j \in J\}$, where i indicates the i-th county and j means the j-th POI class or brightness, and J is the set of different classes of POIs. When we see at least two links from X_{1j} to X_{2j} for any $j \in J$, we draw an internally validated link from $X_{1,all}$ to $X_{2,all}$.

On the other hand, POI-level node aggregation uses the same process as county-level node aggregation but utilizes seven different county nodes into one POI node. Specifically, we aggregate all county nodes into one POI node. Taking mathematical expressions, $X_{all,j}$ is defined by a node to represent j-th POI class or brightness, and $X_{all,j} = \{X_{ij}, \forall i \in I\}$, where I is the set of counties. When we see at least two links from X_{i1} to X_{i2} for any $i \in I$, we draw an internally validated link from $X_{all,1}$ to $X_{all,2}$.

We use this node aggregation method in two ways. First, we use this step to visualize the results of the causal graph. Fig. 3b to f represent the visualized results. We choose 2 as the threshold for the number of links to draw the internally validated causal links. On the other hand, we apply this step to the graph union in Fig. 3a. Note that the threshold value 5 is used for the graph union in Fig. 3a. Details of the threshold value are expressed in Simple block bootstrapping of causal graphs at different time scales subsection.

Simple block bootstrapping of causal graphs at different time scales

We perform a simple block bootstrapping step to obtain internally validated results. The lower side of Fig. 2b explains how we aggregate raw results using different time scales to obtain internally validated links. In Fig. 3b, we include the result with data from the entire period. We split the datasets into five different time scales in Fig. 3b to f: (1) the entire period related to Hurricane Irma (from July 1st to December 31st, 2017) expressed by time period tp_1, (2) the pre-disaster period (from July 1st to September 10th, 2017) represented by time period tp_2, (3) the short-term PDR period from September 11th to September 30th, 2017 by time period tp_3, (4) the mid-term PDR period from October 1st to December 31st, 2017 by time period tp_4, and (5) the aggregated PDR period from September 11th to December

31st, 2017 by time period tp_5. These results are not internally validated when we do not divide the datasets into different time scales. Specifically, it is highly likely that we will obtain causal links that are only applicable to specific time scales or types of POIs.

Let $G_{tp_m} = (\mathcal{V}_{tp_m}, \mathcal{E}_{tp_m})$ be the causal graph for the m-th time period tp_m, denoted by a set of nodes $\mathcal{V}_{tp_m}$ and a set of links $\mathcal{E}_{tp_m}$. Note that each causal graph contains the node aggregation at the county level and POI level, that is, $\mathcal{V}_{tp_m} = \{X_{i,all}, \forall i \in I\} \cup \{X_{all,j}, \forall j \in J\}$.

In Fig. 3a, we present the graph union of the raw model output graphs at all time scales. Let $G_U = (\mathcal{V}_U, \mathcal{E}_U)$ be the causal graph union, where $\mathcal{V}_U = \mathcal{V}_{tp_1}$ and $\mathcal{E}_U = \{e_{i_1,i_2,tp_m} | \sum_{m \in M} n(e_{i_1,i_2,tp_m}) \geq 5, \forall i_1, i_2 \in I, e_{i_1,i_2,tp_m} \in E_{tp_m}, m \in M := \{1,2,3,4,5\}\}$. This means that when we observe not less than five links from v_{i_1} to v_{i_2} in the causal graphs by different time periods, we add the link e_{i_1,i_2} to the causal graph union. We set 5 as the threshold for the occurrence of links from v_{i_1} to v_{i_2} to fortify internal validation. Note that the threshold ($thr_{\mathcal{E}_U}$) for the occurrence of links is calculated by $thr_{\mathcal{E}_U} = 1 + median(Occ(\mathcal{E}_U))$, where $Occ(\mathcal{E}_U)$ means the occurrence of links in $\mathcal{E}_U$, and $median(S)$ represents the median of the occurrence of links S. Fig. 3a represents the causal graph union grouped by simple block bootstrapping[77].

Reporting summary

DATA AVAILABILITY

We use two types of time series datasets. First, the daily moonlight-adjusted nighttime light (NTL) data are publicly available at Earthdata (https://www.earthdata.nasa.gov). Second, based on the request, the daily visit dataset from mobile phone locations is available at Safegraph (https://www.safegraph.com).

CODE AVAILABILITY

The source code and results are available on GitHub: https://github.com/Sangung/Causal_Discovery_Regional_dissmiliarity.git.

REFERENCES

1. Aldrich, D.P. Building Resilience: Social Capital in Post-disaster Recovery. University of Chicago Press, Chicago (2012).
2. U.S. Department of Homeland Security: National Disaster Recovery Framework. U.S. Department of Homeland Security Washington, DC (2016).
3. Spending Explorer of Recovery Support Function Leadership Group (RSFLG). Recovery Support Function Leadership Group (RSFLG). https://recovery.fema.gov/spending-explorer.
4. Federal Emergency Management Agency: 2017 Hurricane Season FEMA After-Action Report. Federal Emergency Management Agency Washington, DC (2018).
5. Danziger, M. M. & Barabási, A.-L. Recovery coupling in multilayer networks. *Nat. Commun.* 13, 1–8 (2022).
6. Esmalian, A. et al. Disruption tolerance index for determining household susceptibility to infrastructure service disruptions. International Journal of Disaster Risk Reduction, 102347 (2021).
7. Yang, Y., Ng, S. T., Zhou, S., Xu, F. J. & Li, H. A physics-based framework for analyzing the resilience of interdependent civil infrastructure systems: A climatic extreme event case in hong kong. *Sustain. Cities Soc.* 47, 101485 (2019).
8. Miranda, J. J., Ishizawa, O. A. & Zhang, H. Understanding the impact dynamics of windstorms on short-term economic activity from night lights in central america. *Econ. Disasters Climate Change* 4, 657–698 (2020).
9. Zarghami, S. A. & Dumrak, J. A system dynamics model for social vulnerability to natural disasters: Disaster risk assessment of an australian city. *Int. J. Disaster Risk Red.* 60, 102258 (2021).
10. Williams, B. D. & Webb, G. R. Social vulnerability and disaster: understanding the perspectives of practitioners. *Disasters* 45, 278–295 (2021).
11. Lee, S., Sadri, A. M., Ukkusuri, S. V., Clawson, R. A. & Seipel, J. Network structure and substantive dimensions of improvised social support ties surrounding households during post-disaster recovery. *Nat. Hazards Rev.* 20, 04019008 (2019).
12. Sadri, A. M. et al. 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, 1377–1406 (2018).
13. Cagney, K. A., Sterrett, D., Benz, J. & Tompson, T. Social resources and community resilience in the wake of superstorm sandy. *PLoS One* 11, 0160824 (2016).
14. Ukkusuri, S., Seetharam, K., Morgan, P. & See, L. Resilience of cities to external shocks: Analysis, modeling and economic impacts. SAGE Publications Sage UK: London, England (2021)
15. 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).
16. Sadiqi, Z., Trigunarsyah, B. & Coffey, V. A framework for community participation in post-disaster housing reconstruction projects: A case of afghanistan. *Int. J. Project Manag.* 35, 900–912 (2017).
17. Olshansky, R. B. Planning after hurricane katrina. *J. Am. Planning Assoc.* 72, 147–153 (2006).
18. Liu, X. et al. Network resilience. *Phys. Rep.* 971, 1–108 (2022).
19. Yabe, T., Rao, P. S. C. & Ukkusuri, S. V. Resilience of interdependent urban socio-physical systems using large-scale mobility data: Modeling recovery dynamics. *Sustain. Cities Soc.* 75, 103237 (2021).
20. Dong, S., Esmalian, A., Farahmand, H. & Mostafavi, A. An integrated physical-social analysis of disrupted access to critical facilities and community service-loss tolerance in urban flooding. *Comp. Environ. Urban Syst.* 80, 101443 (2020).
21. Sutley, E. J. & Hamideh, S. An interdisciplinary system dynamics model for post-disaster housing recovery. *Sustain. Res. Infrastruct.* 3, 109–127 (2018).
22. Pribadi, K., Kusumastuti, D., Sagala, S. & Wimbardana, R. Disaster recovery. Used or misused development opportunity. R. Shaw, ed (2013).
23. Román, M. O. et al. Satellite-based assessment of electricity restoration efforts in puerto rico after hurricane maria. *PloS One* 14, 0218883 (2019).
24. Shi, K. et al. Evaluating the ability of npp-viirs nighttime data to estimate the gross domestic product and the electric power consumption of china at multiple scales: A comparison with dmsp-ols data. *Remote Sens.* 6, 1705–1724 (2014).
25. Yabe, T., Rao, P. S. C., Ukkusuri, S. V. & Cutter, S. L. Toward data-driven, dynamical complex systems approaches to disaster resilience. *Proc. Nat. Acad. Sci.* 119, 2111997119 (2022).
26. Fan, C., Jiang, X., Lee, R. & Mostafavi, A. Equality of access and resilience in urban population-facility networks. *npj Urban Sustain.* 2, 1–12 (2022).
27. Podesta, C., Coleman, N., Esmalian, A., Yuan, F. & Mostafavi, A. Quantifying community resilience based on fluctuations in visits to points-of-interest derived from digital trace data. *J. Roy. Soc. Interface* 18, 20210158 (2021).
28. Montgomery, J. Making a city: Urbanity, vitality and urban design. *J. Urban Des.* 3, 93–116 (1998).
29. Coles, E. & Buckle, P. Developing community resilience as a foundation for effective disaster recovery. *Austr. J. Emerg. Manag.* 19, 6–15 (2004).
30. Guo, R., Cheng, L., Li, J., Hahn, P. R. & Liu, H. A survey of learning causality with data: Problems and methods. *ACM Comput. Surv. (CSUR)* 53, 1–37 (2020).
31. Pearl, J. & Mackenzie, D. The Book of Why: the New Science of Cause and Effect. Basic books, New York (2018).
32. Pearl, J. Causality. Cambridge university press, Cambridge (2009).
33. Stavroglou, S. K., Pantelous, A. A., Stanley, H. E. & Zuev, K. M. Unveiling causal interactions in complex systems. *Proc. Nat. Acad. Sci.* 117, 7599–7605 (2020).
34. Caragliu, A. & Del Bo, C. F. Smart cities and the urban digital divide. *npj Urban Sustain.* 3, 43 (2023).
35. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* 5, 4996 (2019).
36. Runge, J. et al. Inferring causation from time series in earth system sciences. *Nat. Commun.* 10, 1–13 (2019).
37. Arbia, G., Bramante, R., Facchinetti, S. & Zappa, D. Modeling inter-country spatial financial interactions with graphical lasso: An application to sovereign co-risk evaluation. *Regional Sci. Urban Econ.* 70, 72–79 (2018).
38. Hesse, W., Möller, E., Arnold, M. & Schack, B. The use of time-variant eeg granger causality for inspecting directed interdependencies of neural assemblies. *J. Neurosci. Methods* 124, 27–44 (2003).
39. Jain, G. & Espey, J. Lessons from nine urban areas using data to drive local sustainable development. *npj Urban Sustain.* 2, 1–10 (2022).
40. Wang, Z.-H. Reconceptualizing urban heat island: Beyond the urban-rural dichotomy. *Sustain. Cities Soc.* 77, 103581 (2022).
41. Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D. & Sebag, M. Structural agnostic modeling: Adversarial learning of causal graphs. *J. Mach. Learn. Res.* 23, 1–62 (2022).

42. Kalainathan, D., Goudet, O. & Dutta, R. Causal discovery toolbox: Uncovering causal relationships in python. *J. Mach. Learn. Res.* **21**, 1406–1410 (2020).
43. Glymour, C., Zhang, K. & Spirtes, P. Review of causal discovery methods based on graphical models. *Front. Gen.* **10**, 524 (2019).
44. Gong, M., Zhang, K., Schoelkopf, B., Tao, D. & Geiger, P. Discovering temporal causal relations from subsampled data. In: *International Conference on Machine Learning*, 1898–1906 (PMLR, 2015).
45. Gong, M., Zhang, K., Schölkopf, B., Glymour, C. & Tao, D. Causal discovery from temporally aggregated time series. In: *Uncertainty in Artificial Intelligence: Proceedings of The… Conference. Conference on Uncertainty in Artificial Intelligence*, vol. 2017 (NIH Public Access, 2017).
46. Nowack, P., Runge, J., Eyring, V. & Haigh, J. D. Causal networks for climate model evaluation and constrained projections. *Nat. Commun.* **11**, 1415 (2020).
47. Falk, W. W., Hunt, M. O. & Hunt, L. L. Hurricane katrina and new orleanians'sense of place: Return and reconstitution or "gone with the wind"? *Du Bois Rev.* **3**, 115 (2006).
48. Bryner, N. S. et al. Washed out: Policy and practical considerations affecting return after hurricane katrina and superstorm sandy. *J. Asian Dev.* **3**, 73–93 (2017).
49. Groen, J. A. & Polivka, A. E. Going home after hurricane katrina: Determinants of return migration and changes in affected areas. *Demography* **47**, 821–844 (2010).
50. Laska, S. Louisiana's Response to Extreme Weather: A Coastal State's Adaptation Challenges and Successes. Springer, Cham (2020).
51. Xu, J. & Qiang, Y. Spatial assessment of community resilience from 2012 hurricane sandy using nighttime light. *Remote Sensing* **13**, 4128 (2021).
52. Mitsova, D., Esnard, A.-M., Sapat, A. & Lai, B. S. Socioeconomic vulnerability and electric power restoration timelines in florida: the case of hurricane irma. *Nat. Hazards* **94**, 689–709 (2018).
53. Safapour, E., Kermanshachi, S. & Pamidimukkala, A. Post-disaster recovery in urban and rural communities: Challenges and strategies. *Int. J. Disaster Risk Red.* **64**, 102535 (2021).
54. Lenzi, C. & Perucca, G. Economic inequalities and discontent in european cities. *npj Urban Sustain.* **3**, 26 (2023).
55. PreKindergarten Programs. The U.S. Department of Health and Human Services, Administration for Children and Families, Office of Child Care (OCC). https://childcare.gov/consumer-education/prekindergarten-programs.
56. School-Age Child Care and Camp Programs. The U.S. Department of Health and Human Services, Administration for Children and Families, Office of Child Care (OCC). https://childcare.gov/consumer-education/school-age-child-care.
57. Shizuka, D. & McDonald, D. B. The network motif architecture of dominance hierarchies. *J. Roy. Soc. Interface* **12**, 20150080 (2015).
58. Lei, D. et al. Inferring temporal motifs for travel pattern analysis using large scale smart card data. *Transp. Res. Part C: Emerging Technol.* **120**, 102810 (2020).
59. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
60. Erdős, P. et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–60 (1960).
61. Gal, E., Perin, R., Markram, H., London, M. & Segev, I. Neuron geometry underlies a universal local architecture in neuronal networks. bioRxiv https://doi.org/10.1101/656058 (2019).
62. Lee, C.-C., Maron, M. & Mostafavi, A. Community-scale big data reveals disparate impacts of the texas winter storm of 2021 and its managed power outage. *Human. Social Sci. Commun.* **9**, 1–12 (2022).
63. Wang, Y., Kyriakidis, M. & Dang, V. N. Incorporating human factors in emergency evacuation–an overview of behavioral factors and models. *Int. J. Disaster Risk Red.* **60**, 102254 (2021).
64. Cutter, S. L., Ash, K. D. & Emrich, C. T. Urban–rural differences in disaster resilience. *Annals Am. Assoc. Geogr.* **106**, 1236–1252 (2016).
65. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. & Lewis, B. A social vulnerability index for disaster management. *J. Homeland Sec. Emerg. Manag.* **8**, 0000102202154773551792 (2011).
66. Xiao, Y., Olshansky, R., Zhang, Y., Johnson, L. A. & Song, Y. Financing rapid community reconstruction after catastrophic disaster: Lessons from the 2008 wenchuan earthquake in china. *Nat. Hazards* **104**, 5–30 (2020).
67. Peng, Y., Shen, L., Tan, C., Tan, D. & Wang, H. Critical determinant factors (cdfs) for developing concentrated rural settlement in post-disaster reconstruction: a china study. *Nat. hazards* **66**, 355–373 (2013).
68. Sobhaninia, S. & Buckman, S. T. Revisiting and adapting the kates-pijawka disaster recovery model: A reconfigured emphasis on anticipation, equity, and resilience. *Int. J. Dis. Risk Red.* **69**, 102738 (2022).
69. Domingue, S. J. & Emrich, C. T. Social vulnerability and procedural equity: Exploring the distribution of disaster aid across counties in the united states. *Am. Rev. Public Adm.* **49**, 897–913 (2019).
70. Berke, P. R., Kartez, J. & Wenger, D. Recovery after disaster: Achieving sustainable development, mitigation and equity. *Disasters* **17**, 93–109 (1993).
71. IPCC 2023: Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]). IPCC, Geneva (2023).
72. UNDRR The Report of the Midterm Review of the Implementation of the Sendai Framework for Disaster Risk Reduction 2015-2030. UNDRR, Geneva (2023).
73. Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comp. Environ. Urban Syst.* **94**, 101777 (2022).
74. Ratcliffe, M., Burd, C., Holder, K. & Fields, A. Defining rural at the us census bureau. *Am. Commun. Surv. Geogr. Brief* **1**, 1–8 (2016).
75. Xiong, C., Hu, S., Yang, M., Luo, W. & Zhang, L. Mobile device data reveal the dynamics in a positive relationship between human mobility and covid-19 infections. *Proc. Nat. Acad. Sci.* **117**, 27087–27089 (2020).
76. Kang, Q. et al. Machine learning-aided causal inference framework for environmental data analysis: a covid-19 case study. *Environ. Sci. Technol.* **55**, 13400–13410 (2021).
77. Politis, D. N. The impact of bootstrap methods on time series analysis. *Stat. Sci.* 219–230 (2003).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

S.P., T.Y., and S.V.U. proposed the question and designed the research. S.P. and S.V.U. performed a causal discovery analysis and wrote the article. T.Y. proposed the causal discovery methods and edited the paper.

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/s42949-023-00139-4.

Correspondence and requests for materials should be addressed to Sangung Park, Tong Yao or Satish V. Ukkusuri.

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.