

Supplementary Information for Behavior-based dependency networks between places shape economic resilience of cities

Takahiro Yabe^{1,*}, Bernardo García Bulle Bueno¹, Morgan R. Frank^{2,3,4},
Alex Pentland^{1,4}, Esteban Moro^{1,4,5,*}

¹Institute for Data, Systems, and Society, Massachusetts Institute of Technology, USA

²Department of Informatics and Networked Systems, University of Pittsburgh, USA

³Digital Economy Lab, Institute for Human-Centered AI, Stanford University, Stanford, CA, USA

⁴Media Lab, Massachusetts Institute of Technology, USA

⁵Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid, Spain

*Corresponding authors: tyabe@mit.edu, emoro@mit.edu

Supplementary Notes

1	Mobility data analytics and representativeness	7
1.1	Home estimation and stop detection	7
1.2	Safegraph POI data and visit attribution	8
1.3	Data representativeness	9
1.4	Post-stratification of mobility data	9
2	Behavior-based dependency networks	15
2.1	Measuring behavior-based dependency between places	15
2.2	Qualitative characteristics of networks	15
2.3	Comparison of networks under different co-visit detection parameters	16
3	Statistical robustness of behavior-based dependency networks	30
3.1	Robustness against choice of time period	30
3.2	Computing quartiles of w_{ij} via bootstrap method	30
3.3	Robustness against post-stratification processing	32
4	Modeling dependency network weights	38
4.1	Gravity-based null networks	38
4.2	Regression model of w_{ij}	39
4.3	Robustness against different network instances	39
4.4	Robustness against co-visit detection parameters	45
4.5	Robustness against different time periods	45

5	Impacts of behavior-based dependency during the COVID-19 pandemic	48
5.1	Analysis of visitation losses in cities	48
5.2	Model specification and estimation results	48
5.3	Robustness against co-visit detection parameters	51
5.4	Results using changes in visits during different time periods	55
5.5	Robustness of regression results when using different time periods to generate dependency networks	61
5.6	Robustness of regression results when using only short non-work stays	66
5.7	Case study on school closure period	66
6	Cascading impacts of hypothetical shocks	74
6.1	Leontief open model	74
6.2	Cascading effects of individual place closures	74
7	Software	82

List of Figures

S1	Estimates of number of visits to POIs under different maximum spatial threshold parameters	11
S2	Sample rates of mobile phone location data across the five metropolitan areas.	12
S3	Comparison of census population and smartphone users for census block groups in the five metropolitan areas	13
S4	Comparison of census block group median income and the sample rate of mobile phone smartphone users in the five metropolitan areas	14
S5	Distribution of time and step differences between two POIs visited by the same individual on the same day	16
S6	POI subcategories with the largest average in- and out-weights	17
S7	Proportion of edge weights among POI category pairs	19
S8	POI subcategories with the largest average in- and out-weights, when $T_c = 1$ hour . .	20
S9	POI subcategories with the largest average in- and out-weights, when $T_c = 3$ hour . .	21
S10	Dependency network in New York under different co-visit detection parameters . . .	22
S11	Dependency network in Boston under different co-visit detection parameters	23
S12	Dependency network in Seattle under different co-visit detection parameters	24
S13	Dependency network in Los Angeles under different co-visit detection parameters . .	25
S14	Dependency network in Dallas under different co-visit detection parameters	26
S15	Network diagram showing the average dependencies between POI subcategories in New York and Boston	27
S16	Network diagram showing the average dependencies between POI subcategories in Seattle and Los Angeles	28
S17	Network diagram showing the average dependencies between POI subcategories in Dallas	29
S18	Comparison of in- and out-weights of each POI across different data collection periods	31
S19	Comparison of in- and out-weights of each POI across different data collection periods, contd.	32
S20	Comparison of category pairwise weight proportion across different data collection periods	33
S21	Comparison of category pairwise weight proportion across different data collection periods	34
S22	Out- and in-weights of POIs when applying bootstrap method to remove smaller weights	35
S23	Out- and in-weights of categories when applying bootstrap method to remove insignificant weights	36
S24	Out- and in-weights of POIs when applying post-stratification to correct for biases in mobile phone data	37
S25	Correlation between the observed n_{ij} values and the gravity component g_{ij} under different model parameters	38
S26	Comparison of actual and null networks in New York and Boston	40
S27	Comparison of actual and null networks in Seattle and Los Angeles	41
S28	Comparison of actual and null networks in Dallas	42
S29	Comparison between real and null network on the proportion of edge weights among POI category pairs	42
S30	Histograms of edge distances between actual and null networks.	43

S31	Regression results for dependency weights when using different co-visit detection parameters, contd.	46
S32	Change in visits to POIs in New York, Boston, Seattle, Los Angeles, and Dallas during the pandemic period.	49
S33	Correlation between change in visits of ego and the weighted change in visits of alters by category pairs	50
S34	Regression results for dependency weights when using different co-visit detection parameters, contd.	54
S35	Change in visits to POIs in New York, Boston, Seattle, Los Angeles, and Dallas during the fall semester period.	70
S36	Cascading impacts of a 50% visit reduction to colleges.	75
S37	Cascading impacts of a 100% visit reduction to colleges.	76
S38	Cascading impacts of a 25% visit reduction to colleges.	77
S39	Total cascading impact of closing places on other locations in New York and Boston .	79
S40	Total cascading impact of closing places on other locations in Seattle and Los Angeles	80
S41	Total cascading impact of closing places on other locations in Dallas	81

List of Tables

S1	Number of places in the Safegraph dataset in the five core-based statistical areas (CBSAs) analyzed in this study.	8
S2	Description of the four core-based statistical areas (CBSAs) analyzed in this study. .	10
S3	Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10}w_{ij}$) using physical factors	44
S4	Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10}w_{ij}$) selected using the bootstrap method with physical factors	44
S5	Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10}w_{ij}$) selected using only short stays (< 4 hrs) with physical factors . . .	45
S6	Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10}w_{ij}$) using network from 2019 January - April data	47
S7	Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10}w_{ij}$) selected using network from 2019 May - August data	47
S8	Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in New York	52
S9	Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Boston	52
S10	Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Seattle	53
S11	Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Los Angeles	53
S12	Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Dallas	55
S13	Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in New York	56

S14	Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Boston	56
S15	Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Seattle	57
S16	Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Los Angeles	57
S17	Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Dallas	58
S18	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York	58
S19	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston	59
S20	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle	59
S21	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles	60
S22	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas	60
S23	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from data between 2019 January to April	61
S24	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from data between 2019 January to April	62
S25	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from data between 2019 January to April	62
S26	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from data between 2019 January to April	63
S27	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from data between 2019 January to April	63
S28	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from data between 2019 May to August	64
S29	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from data between 2019 May to August	64
S30	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from data between 2019 May to August	65
S31	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from data between 2019 May to August	65

S32	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from data between 2019 May to August	66
S33	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from stays shorter than 4 hours	67
S34	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from stays shorter than 4 hours	67
S35	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from stays shorter than 4 hours	68
S36	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from stays shorter than 4 hours	68
S37	Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from stays shorter than 4 hours	69
S38	Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in New York	71
S39	Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Boston	71
S40	Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Seattle	72
S41	Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Los Angeles	72
S42	Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Dallas	73

1 Mobility data analytics and representativeness

1.1 Home estimation and stop detection

In this study, we utilize an anonymized location dataset of mobile phones and smartphone devices provided by Spectus Inc., a location data intelligence company which collects anonymous, privacy-compliant location data of mobile devices using their software development kit (SDK) technology in mobile applications and ironclad privacy framework. Spectus processes data collected from mobile devices whose owners have actively opted in to share their location and requires all application partners to disclose their relationship with Spectus, directly or by category, in the privacy policy. With this commitment to privacy, the data set contains location data for roughly 15 million daily active users in the United States. Through Spectus’ Data for Good program, Spectus provides mobility insights for academic research and humanitarian initiatives. All data analyzed in this study are aggregated to preserve privacy¹. Each entry in the data table comprises anonymized device ID, location coordinates, start time, and dwell time of the stop for the device.

To define the type of location (Home or Work), different variables are used, including the number of days spent in a given location in the last month, the daily average number of hours spent in that location, and the time of the day spent in the location (nighttime/daytime). To estimate the home position of a user, the algorithm combines the three variables and creates a score that represents the probability that the position points to the home. The more days and the average number of hours spent in the position, the higher the score is. Higher scores will also be assigned to the most common places during the night. The location that maximizes this score is defined as the home of the device.

Once the location of the home location is identified, the algorithm looks for the work position. Note that the algorithm requires the work location to be located at least 100 meters apart from the home location. The same variables used for the detection of the home location are used, but a higher score is given to daytime locations for the work location rather than nighttime locations. Spectus runs the algorithm every week in order to confirm or update the inferred home and work locations as we observe new data. We will only consider devices that have been present in Spectus’ dataset for at least 15 days. Spectus tightly restricts access to the inferred precise home and work locations of devices. Furthermore, it is used as input into various downstream processes to create more privacy-protected versions of Spectus datasets. For example, we only expose home and work datasets in Spectus Workbench associated with standard Census Block Groups, created by the U.S. Census Bureau, rather than the precise locations. This offers a good balance between utility and privacy: according to the U.S. Census Bureau, there are between 600 and 3000 people living in each block group. Each block group is an aggregate of contiguous U.S. blocks sharing similar socio-demographic characteristics. The representativeness of this data has been tested and corrected in Section 1.3 in the Supplementary Material. The stops, which are location clusters where individual users stay for a given duration, are estimated using the Sequence Oriented Clustering approach [17].

In this study, we selected New York, Boston, Seattle, Los Angeles, and Dallas as the five CBSAs given constraints on data collection. The five cities were selected with respect to the diversity of characteristics in terms of geographical locations (2 Northeast, 1 Northwest, 1 Southwest, and 1 Southern), sociodemographic details (population ranges from 3.5M in Seattle to 19.8M in New York), political inclinations, and weather characteristics.

¹<https://spectus.ai/privacy/privacy-policy/>

Table S1: Number of places in the Safegraph dataset in the five core-based statistical areas (CBSAs) analyzed in this study.

Place category	CBSAs				
	New York	Boston	Seattle	Los Angeles	Dallas
Arts and Museums	4,580	1,183	799	3,697	1,073
City and Outdoors	7,201	5,311	2,792	4,817	2,174
Coffee and Tea	14,799	3,481	3,392	11,395	4,360
College	3,737	1,266	766	2,661	942
Entertainment	2,296	657	584	2,101	745
Food	66,348	12,687	9,924	39,907	16,795
Grocery	14,763	2,942	1,964	7,636	3,792
Health	20,654	4,799	3,913	14,212	5,806
Office	7,840	3,256	2,628	10,936	6,128
Service	114,581	33,525	23,920	104,253	45,520
Shopping	92,211	20,302	14,089	79,024	29,069
Sports	12,383	5,892	2,464	10,509	3,782
Transportation	41,729	11,114	13,245	32,141	14,056
All places	403,669	107,081	81,140	326,332	136,686

1.2 Safegraph POI data and visit attribution

To measure the visitation patterns of individuals in urban environments, we attribute the stops of individual users to specific places in the city. To study the stops at different places, we use stops that are longer than 10 minutes but shorter than ten hours. In our study, we use location data of places collected by Safegraph ². To protect the users’ privacy, we have removed various privacy-sensitive places from our places database. Sensitive places include health-related places, places where the vulnerable population are located, military-related, religious facilities, places that are related to sexual-orientation, and adult-oriented places ³. As a result, we have a total of 403,669 places in New York, 107,081 places in Boston, 81,140 places in Seattle, 326,332 places in Los Angeles, and 136,686 places in Dallas. The breakdown of the number of places by the place category is shown in Table S1. To attribute a stop to a place, we simply attribute each stop the closest place in our dataset. To avoid attributing a stop to place far away, we attribute the stop to a place within $d_{max} = 100$ meters from the observed location of the stop. If the stop is further away than 100 meters from any place in the dataset, the stop is discarded from our dataset and not used for computing the diversity of encounters.

The robustness of our estimation of the number of visits to different locations has been tested using different spatial thresholds of d_{max} . Different levels of d_{max} could change how individuals’ stays are attributed to the places and thus could affect our estimates of the income diversity of physical encounters. Figure S1 compares the estimated number of visits to places over a 1 month period in the five cities when we use different levels of d_{max} (y-axis) with our default parameter $d_{max} = 100m$ (x-axis). For all values $d_{max} = \{50, 150, 200\}$, the Pearson correlation of the number of visits are extremely high, in most cases above $\rho > 0.95$. This robustness check shows that the estimated

²<https://www.safegraph.com/>

³<https://spectus.ai/privacy/spoi-policy/>

visitation patterns do not depend on the choice of the spatial threshold parameter for visit attribution.

1.3 Data representativeness

The location data used in our study is collected from smartphones via various apps and services. Although a significant portion (85% according to 2021 data⁴) of the US population owns a smartphone, one could question the representativeness of the 1.16 million user samples across geographical regions and income quantiles. Studies have reported the digital divide and smartphone usage gaps across sociodemographic groups in the US [14]. In this section, we test whether our group of users in the mobility data is representative of the total population, and further employ post-stratification techniques to correct for any potential biases in the sampling rates across places and socioeconomic status and to test whether the results on income diversity dynamics are robust to such uncertainties concerning data representativeness.

The sampling percentage of the mobility data ($100\% \times \text{number of observed mobile phone users divided by the total population from the census data}$) is around 5% to 10% across all census block groups (CBGs) in the metropolitan regions, as shown in Figure S2. Sample rates are calculated by dividing the total number of users who were observed across a three-month period in the dataset by the census block group population obtained from the American Community Survey [2]. There is a large variation in the sample rates, within and across metropolitan areas. To test whether the users in the location data are representative of the entire population, first, we compare the population detection in our mobility data and the 2019 ACS data for each of the CBGs in the cities. The panels in Figure S3 show the comparison between the census population (x-axis) and the number of observed smartphone users (y-axis) on the CBG scale in the months of September 2019 to January 2020 in the five metropolitan areas. The correlation is moderately high, between around $\rho = 0.55$ and $\rho = 0.78$, showing that despite the use of such small census areas and potential bias in the smartphone usage patterns, we are able to obtain a good representation of the population. To overcome this bias, in Section 1.4, we use post-stratification techniques to correct for such differences in the sample percentages across CBGs. Later, in Section 3.3, we assess whether our estimates of visitation patterns are affected by the representativeness of the data.

In addition to the differences in sampling rates across CBGs, differences in representativeness across income groups are important to ensure the findings of the study are not biased toward specific sociodemographic groups. To measure the representativeness across income quantiles, we plot the correlation between the median income of the census block group and the sample rate (shown in Figure S2) for the five metropolitan areas, as shown in Figure S4. A balanced dataset would have a nonsignificant correlation between the two metrics. However, we observe a small (below $\rho = 0.2$) but significant positive correlation between income and sample rates for New York, Los Angeles, and Dallas. This indicates that in these metropolitan areas, higher income groups are over-represented in the dataset. In Section 1.4, we use post-stratification techniques to correct for such differences in the sample percentages across CBGs and income groups.

1.4 Post-stratification of mobility data

To correct for, and to understand the effects of the varying sampling rates across CBGs and income groups on our estimation of visitation patterns to places, we apply a post-stratification technique,

⁴https://www.statista.com/topics/2711/us-smartphone-market/#topicHeader_wrapper

Table S2: Description of the four core-based statistical areas (CBSAs) analyzed in this study.

CBSA	Population	# users (monthly)	# stays (monthly)	# places
New York-Newark-Jersey City	19.95M	1,456K	36.67M	403,669
Boston-Cambridge-Newton	4.64M	144K	2.34M	107,081
Seattle-Tacoma-Bellevue	3.55M	141K	2.23M	81,140
Los Angeles-Long Beach-Anaheim	13.05M	452K	10.00M	326,332
Dallas-Fort Worth-Arlington	6.70M	425K	8.83M	136,686
Total	47.89M	2.62M	60.07M	1,054,908

which is used in previous studies [10, 18]. Post-stratification is a well know sampling tool [11] and is typically used to study the impact of sampling biases in mobile phone location data [6] or (geolocated) social media data [16] on various downstream tasks and analyses. Following the methods employed in Moro et al. [10], we denote w_g the expansion factor, which is the ratio of the population of census block group g to the population detected in our mobility data. We then weight the visits completed by people from census block group g spends at place α by

$$\hat{n}_{gi} = w_g n_{gi}$$

where the assumption is that $n_{g\alpha}$ is proportional to the number of people visiting the place. Using this method, we could increase (decrease) the visits to places by people coming from census block groups that are under-estimated (over-estimated).

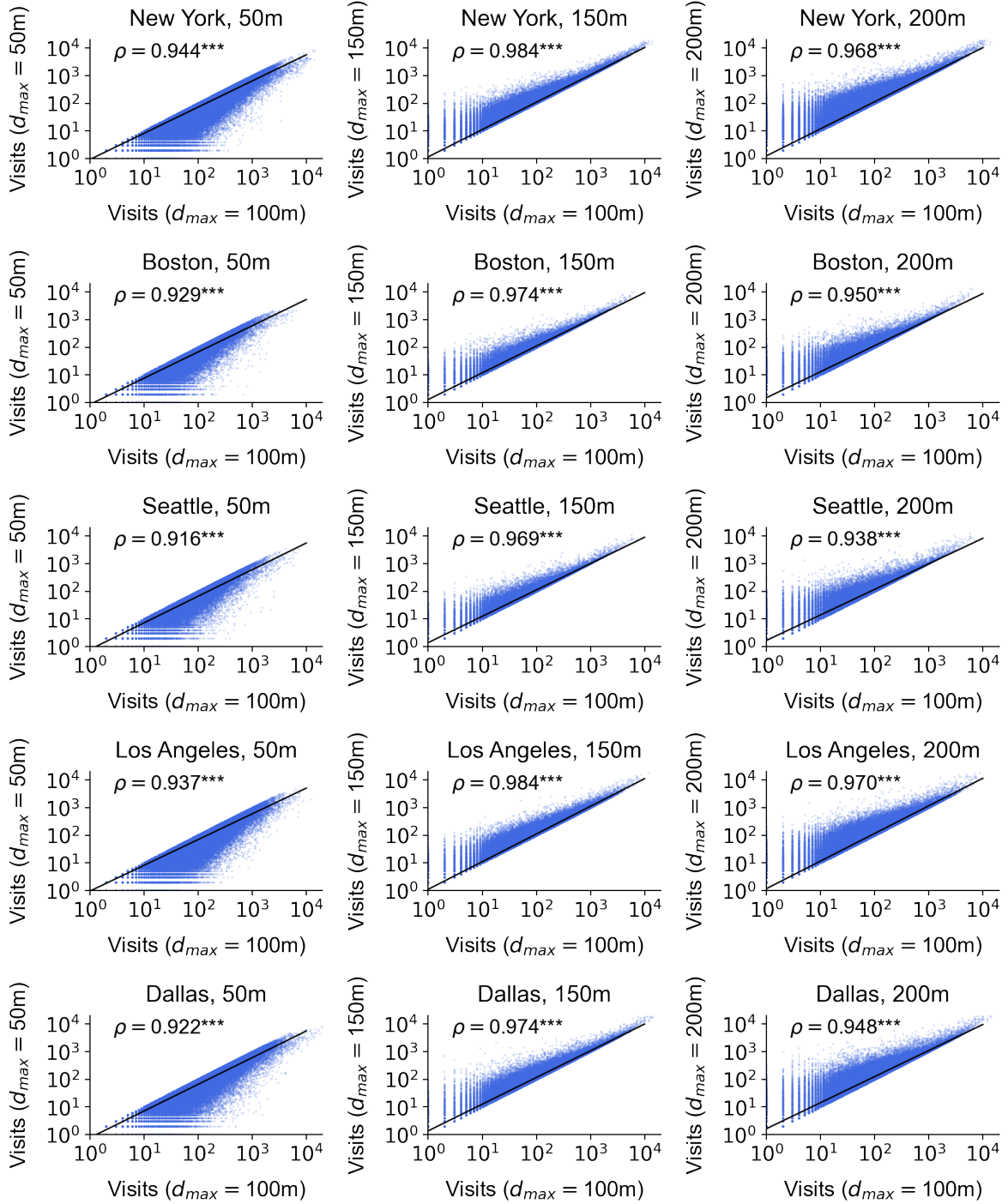


Figure S1: Estimates of number of visits to POIs under different maximum spatial threshold parameters. Pearson correlation of estimated number of visits to different POIs between different spatial threshold parameters are extremely high across all cities.

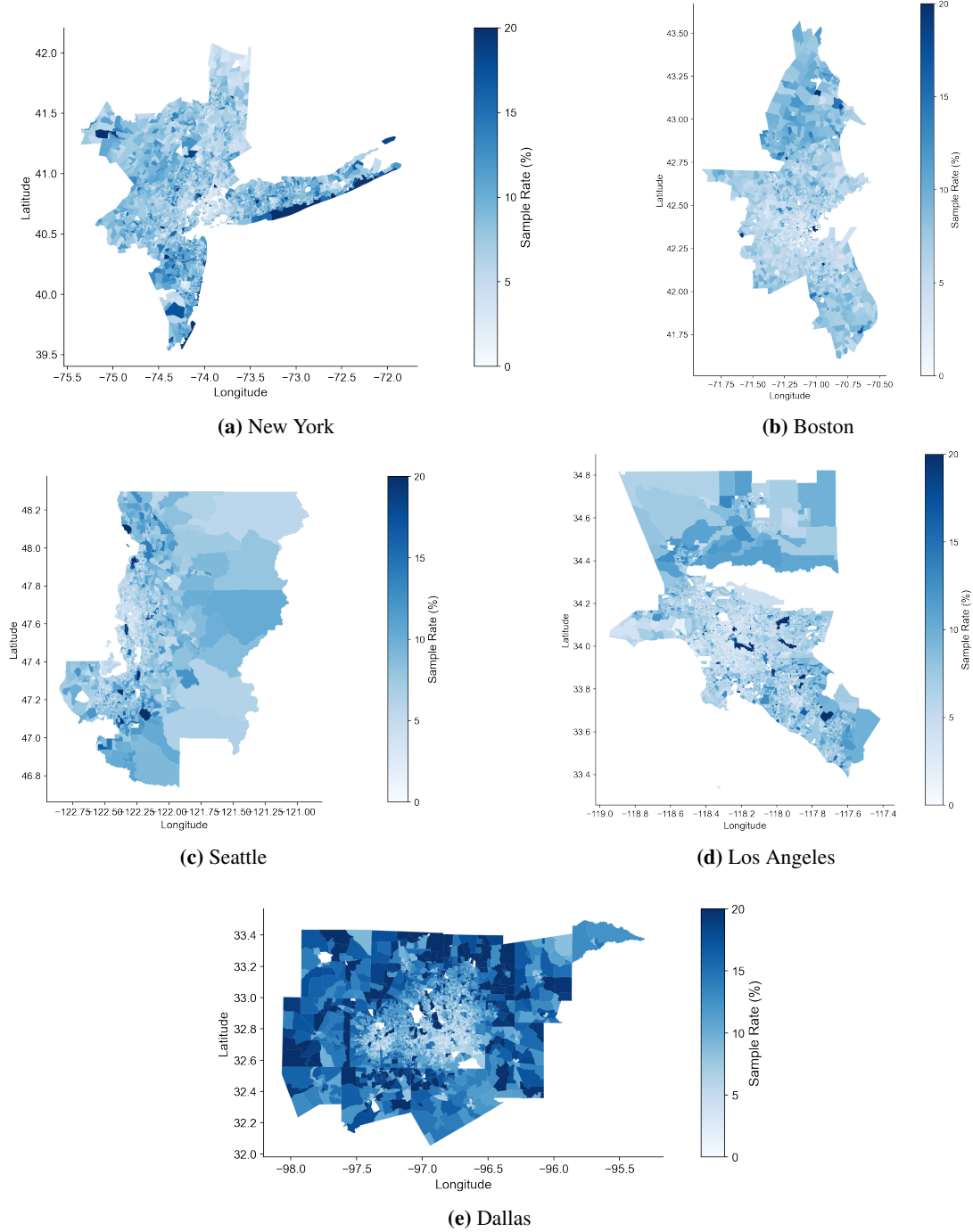


Figure S2: Sample rates of mobile phone location data across the five metropolitan areas. Sample rates are calculated by dividing the total number of users who were observed across a three month period in the dataset by the census block group population obtained from the American Community Survey [2]. Maps were produced in Python using the TIGER shapefiles from the U.S. Census Bureau [15]

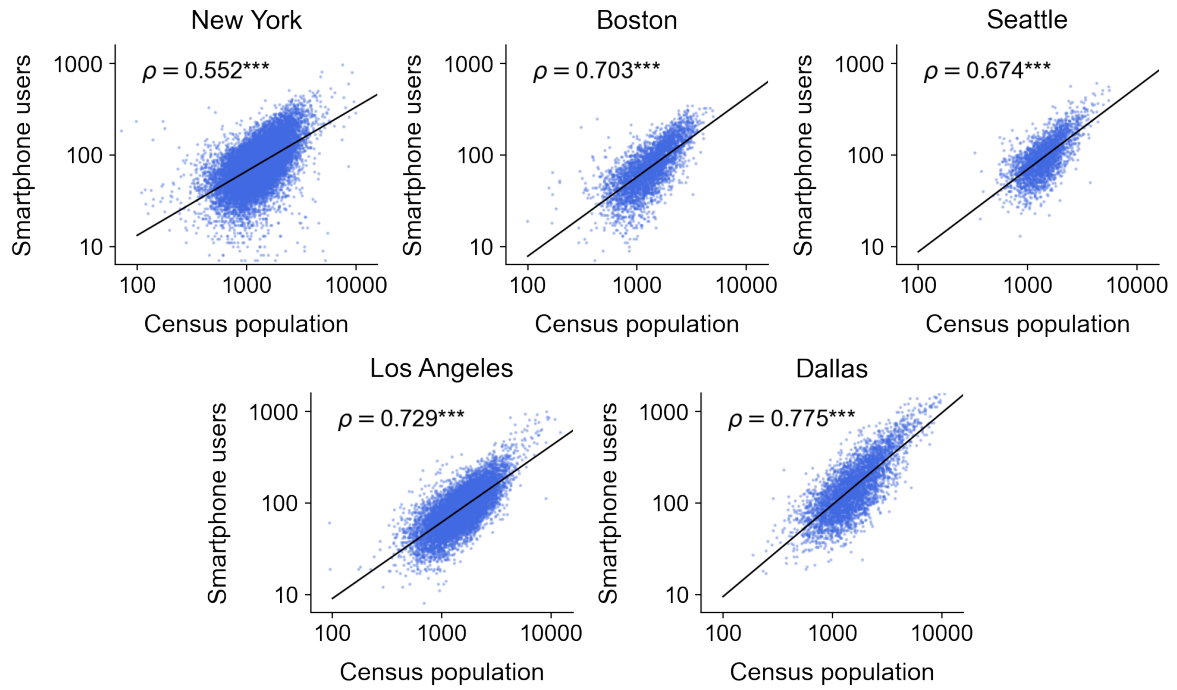


Figure S3: Comparison of census population and smartphone users for census block groups in the five metropolitan areas. The correlation is moderately high, between around $\rho = 0.55$ and $\rho = 0.78$, showing that despite the use of such small census areas and potential bias in the smartphone usage patterns, we are able to obtain a good representation of the population. This bias is corrected using post-stratification techniques.



Figure S4: Comparison of census block group median income and the sample rate of mobile phone smartphone users in the five metropolitan areas. The correlation is small (below $\rho = 0.2$) but significantly positive in New York, Los Angeles, and Dallas. This indicates that in these metropolitan areas, higher income groups are over-represented in the dataset. This bias is corrected using post-stratification techniques.

2 Behavior-based dependency networks

2.1 Measuring behavior-based dependency between places

Using the large and longitudinal dataset of GPS location records in five major US metropolitan areas introduced in Supplementary Note 1, we construct the behavior-based dependency network at the level of points-of-interest (POIs; e.g., businesses and amenities) in cities. Within the mobility dataset, we identified stays at places that were detected to be between 10 minutes and 10 hours and we spatially matched those stays with the closest place locations within 100 meters to infer visits to specific POIs, as described in Supplementary Note 1. The representativeness of the data across regions and income levels was ensured via post-stratification techniques.

In this study, the dependence of a POI i on another POI j is defined as

$$w_{ij} = \frac{n_{ij}}{n_i} \quad (1)$$

where n_i denotes the number of visits to POI i and n_{ij} denotes the number of ‘co-visits’ between POIs i and j . A co-visit is defined as an instance in which POIs i and j were visited by the same individual:

- on the same day,
- within T_c hours from exiting POI i (j) to entering POI j (i), and
- within T_s intermediate POIs

Because the denominator is based on the number of visits to the target POI, $w_{ij} \neq w_{ji}$. This simple but intuitive measure considers the asymmetric nature of dependencies between POIs. As shown in Figure S5, the distributions of the time difference or number of intermediate steps between any given two POIs within a movement trajectory follow a long-tailed distribution, where the majority of the pairs of POIs can be included within 6 hours and 1 step difference. Therefore, as a baseline parameter setting, we use $T_c = 6$ hours and $T_s = 1$ POIs (which indicates direct visitation from POI i to POI j , or vice-versa). The characteristics of the dependency network when using different co-visit detection threshold parameters T_c and T_s are tested and discussed in the following sections such as Supplementary Note 2.3 and 4.2. By computing the dependency weights $w_{ij} \forall i, j$, we obtain the behavior-based dependency matrix $W \in \mathbb{R}^{N \times N}$ where N is the total number of POIs present in the CBSA.

2.2 Qualitative characteristics of networks

To obtain a better understanding of the generated behavior-based dependency network, the POI subcategories with the largest average in- and out- weights were plotted in Figure S6. Supercenters, malls, department malls, and airports had the largest in-weights, indicating that these places were depended by many other POIs. On the other hand, places such as art dealers and various types of retail stores had the largest average out-weight, indicating that such places had high dependency on other places. The subcategories that appeared in both rankings were consistent across the five metropolitan areas. As shown in Figures S8 and S9, the POI subcategories that appeared in each ranking were similar when we changed the co-visit detection parameter to $T_c = 1$ and $T_c = 3$, respectively.

To further observe the dependency relationships between categories, we computed the proportion of edge weights among POI category pairs, shown in Figure S7. The proportion of edge weights was

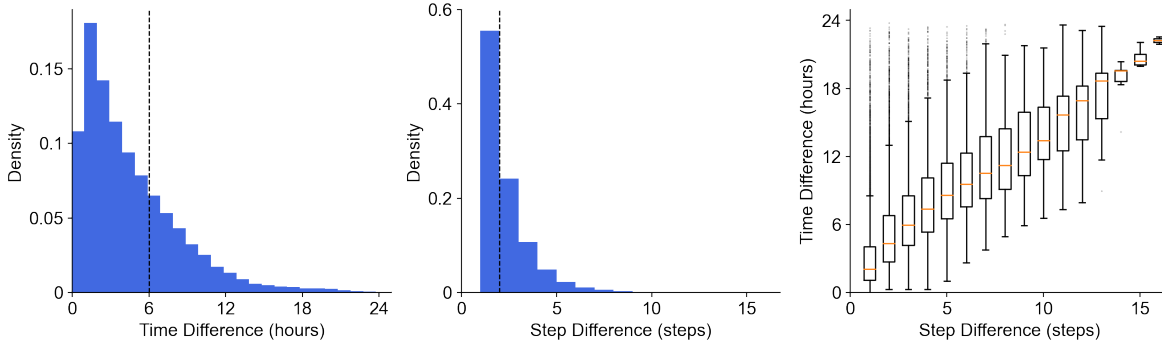


Figure S5: Distribution of time and step differences between two POIs visited by the same individual on the same day. The majority of the co-visits can be captured by using a threshold of 6 hours time difference and 1 step difference. The dependency networks were computed using co-visits that occurred within 6 hours and 1 step.

computed by taking the sum of weights that connect the vertical and horizontal categories and dividing that by the total weights that exist in the network, more formally, by $\sum_{i \in A, j \in B} w_{ij} / \sum_{i, j} w_{ij}$. As the matrices in Figure S7 show, the patterns of dependency are consistent across cities, with Food, Service, Shopping, and Transport playing a big role in both depending on and being depended by others.

2.3 Comparison of networks under different co-visit detection parameters

How much do the co-visit detection parameters (T_s, T_c) affect the structural properties of the dependency networks? To investigate this, we generated the dependency network using 25 pairs of parameters ($T_s = [1, 2, 3, 5, \text{unlimited}]$ and $T_c = [1, 3, 6, 12, 24]$ hours). The following analysis compares the following characteristics of the dependency networks across different co-visit detection parameters:

- POI subcategories with largest in- and out-weights (Figures S8 and S9)
- Distributions of in-weights and out-weights per POI (panels (a) and (b) in Figures S10 to S14)
- Relationship between average weight and physical distance between the POIs (panels (a) and (b) in Figures S10 to S14)
- Difference in the proportion of weights for each category pair (positive means baseline network has more weights between the category pairs) (panel (c) in Figures S10 to S14)

As shown in Figures S8 and S9, the POI subcategories that appeared in each ranking were similar when we changed the co-visit detection parameter to $T_c = 1$ and $T_c = 3$, respectively. Under all parameters, supercenters, malls, department malls, and airports had the largest in-weights, while places such as art dealers and various types of retail stores had the largest average out-weight.

Panels (a) in Figures S10 to S14 show that although the distribution obviously shifts to the right (more steps would include more links, thus more in- and out- weights) the maximum step difference parameter does not make a big difference on the distributions of the in-weights, out-weights, or the decay of average weights with the physical distance.

On the other hand, Panels (b) in Figures S10 to S14 show that the distributions are significantly affected by the choice of time difference parameter. Obviously, the longer the temporal threshold we

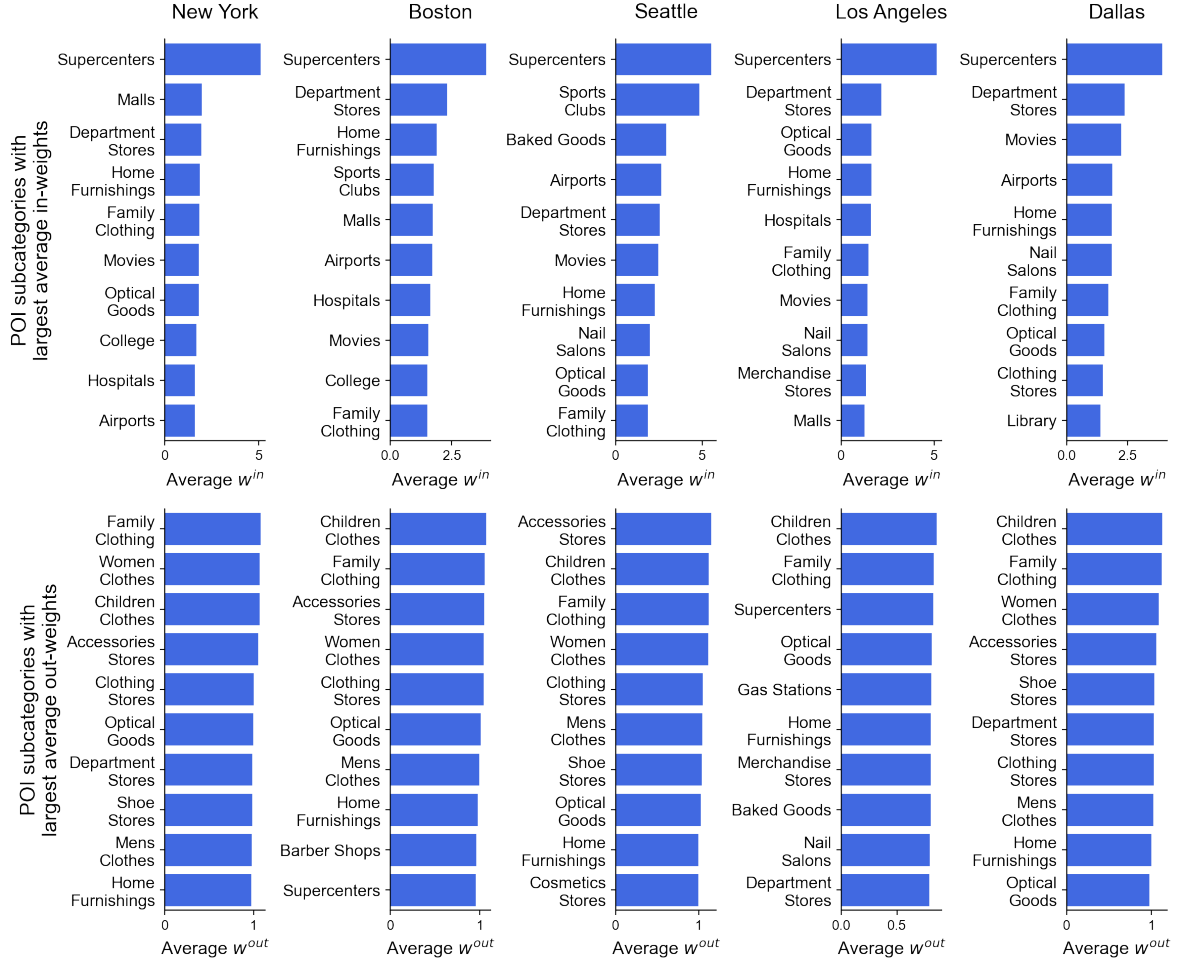


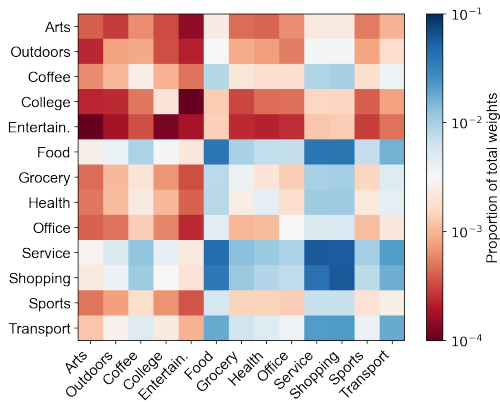
Figure S6: POI subcategories with the largest average in- and out-weights Supercenters, malls, department malls, and airports had the largest in-weights, while places such as art dealers and various types of retail stores had the largest average out-weight.

use, the distributions shift to the right (more time difference would include more links, thus more in- and out- weights). Especially, the dependency networks with the 1 hour threshold present a significantly sparser network with lower in-weights, out-weights, and faster decay of distance with respect to physical distance. The $T_c = 6$ threshold that we chose as the baseline parameter generates network characteristics that are closely in between $T_c = 3$, and $T_c = 12$ and $T_c = 24$.

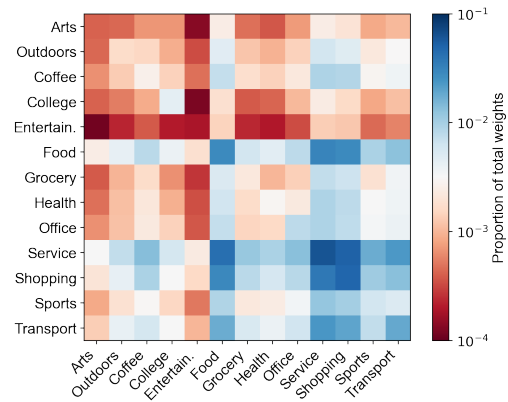
Panels (c) in Figures S10 to S14) show that despite the changes in co-visit detection parameters, the proportion of weights among category pairs are not affected significantly. Note that positive differences (blue color) mean that the baseline network ($T_s = 1, T_c = 6$) contains more weight in the specific category pair. The only exception is $T_c = 1$ (similar to the in-weights, out-weights, and distance decay curves), where we observe much fewer weights among shopping POIs in the baseline network. This indicates that when we limit the time difference to 1 hour, we capture significantly more of the shopping-to-shopping weights compared to other category pairs.

Network diagrams in Figures S15 to S17 show the average dependencies between POI subcate-

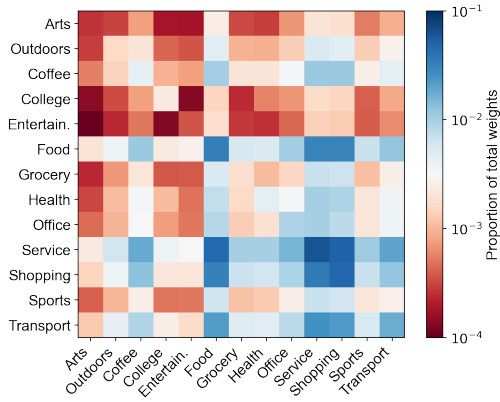
gories in the five cities. Each node represents a POI subcategory (there are 96 of them in the dataset) and the three largest outgoing dependency edges are shown for each node. Node sizes show the in-degree of the constructed network (i.e., how many other POI categories depend on that node). Many shopping subcategories including supercenters, department stores, malls, and clothing stores, and colleges, cafes, restaurants are depended by many other subcategories.



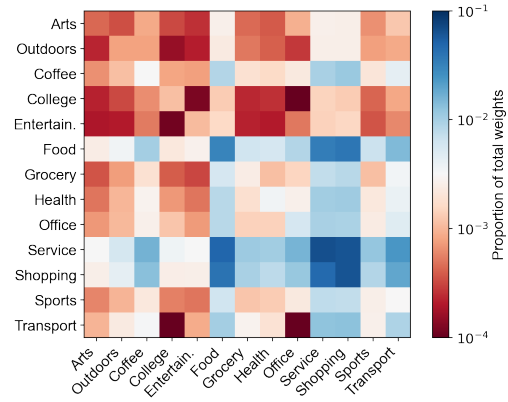
(a) New York



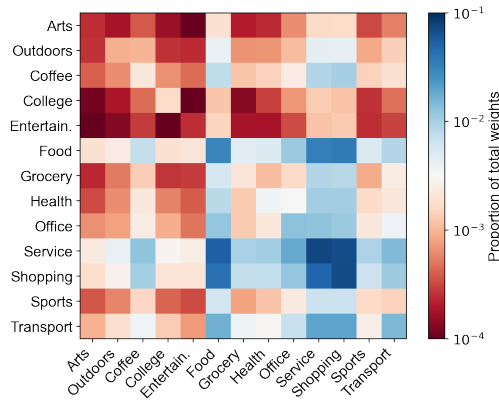
(b) Boston



(c) Seattle



(d) Los Angeles



(e) Dallas

Figure S7: Proportion of edge weights among POI category pairs. The proportion of edge weights was computed by taking the sum of weights that connect the vertical and horizontal categories and dividing that by the total weights that exist in the network. The patterns of dependency are consistent across cities, with Food, Service, Shopping, and Transport playing a big role in both depending on and being depended by others.

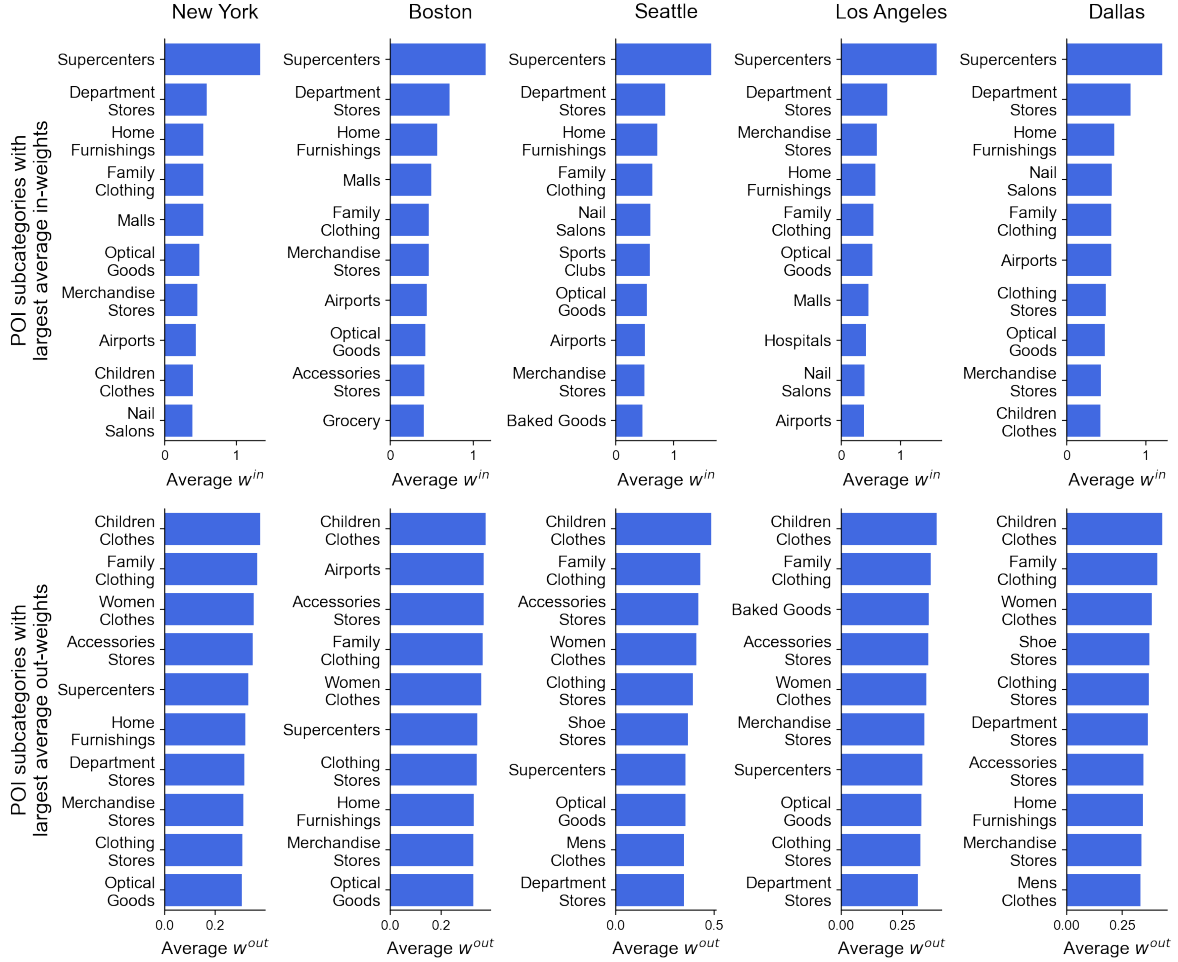


Figure S8: POI subcategories with the largest average in- and out-weights, when $T_c = 1$ hour Supercenters, malls, department malls, and airports had the largest in-weights, while places such as art dealers and various types of retail stores had the largest average out-weight. Results were similar to the baseline parameter when $T_c = 6$ hours.

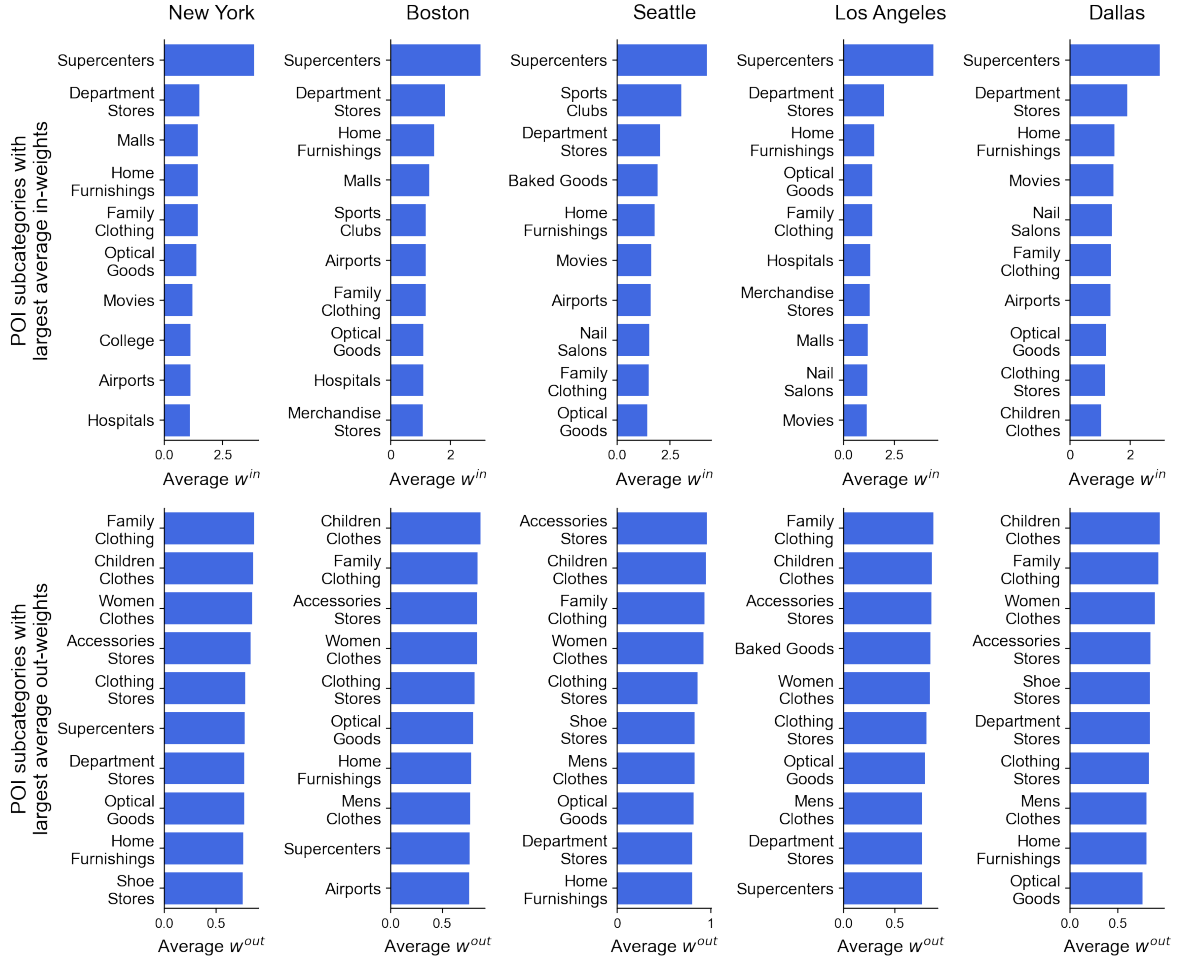
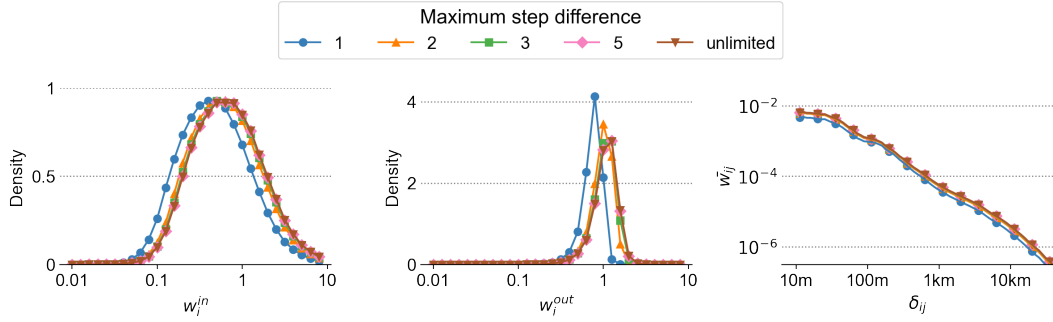
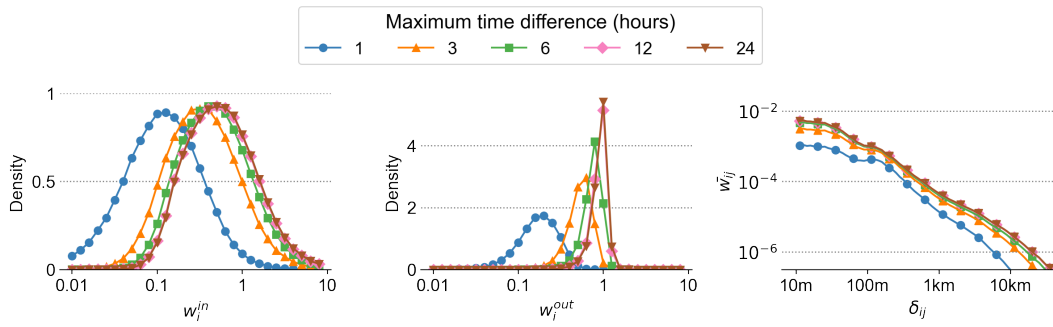


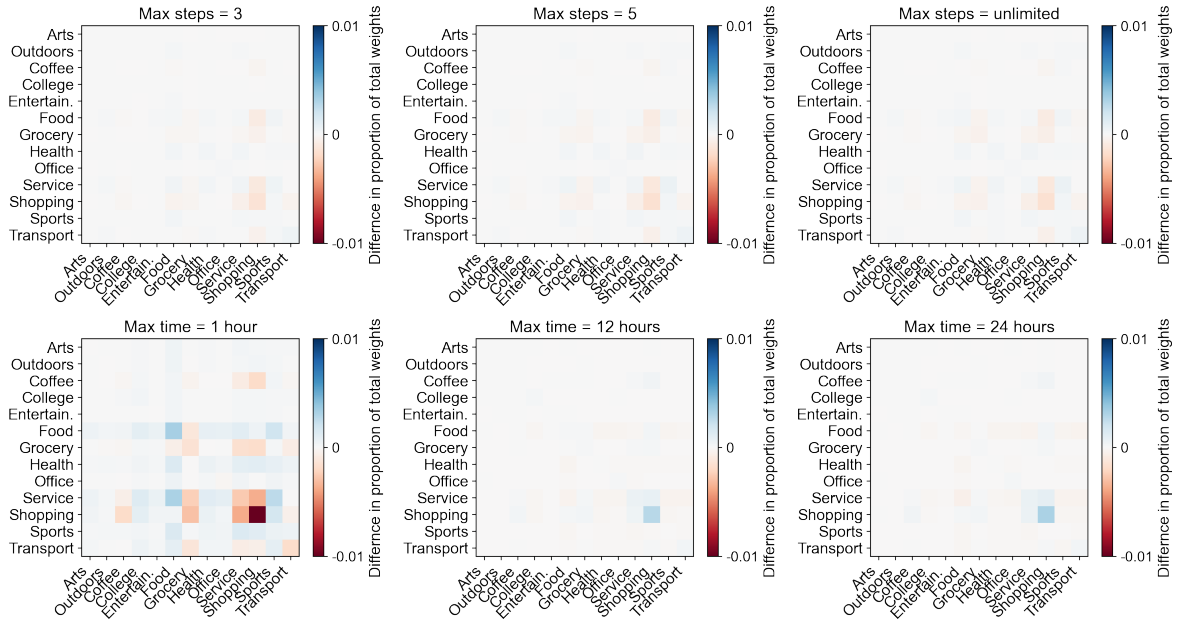
Figure S9: POI subcategories with the largest average in- and out-weights, when $T_c = 3$ hour Supercenters, malls, department malls, and airports had the largest in-weights, while places such as art dealers and various types of retail stores had the largest average out-weight. Results were similar to the baseline parameter when $T_c = 6$ hours.



(a) Network characteristics under different maximum step difference T_s

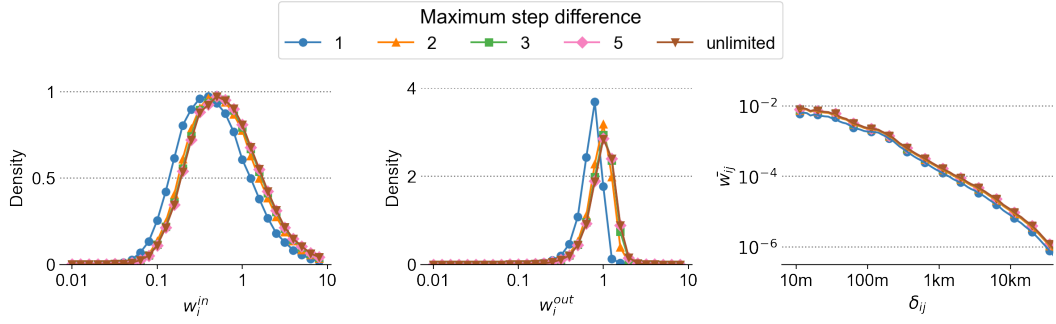


(b) Network characteristics under different maximum time difference T_c

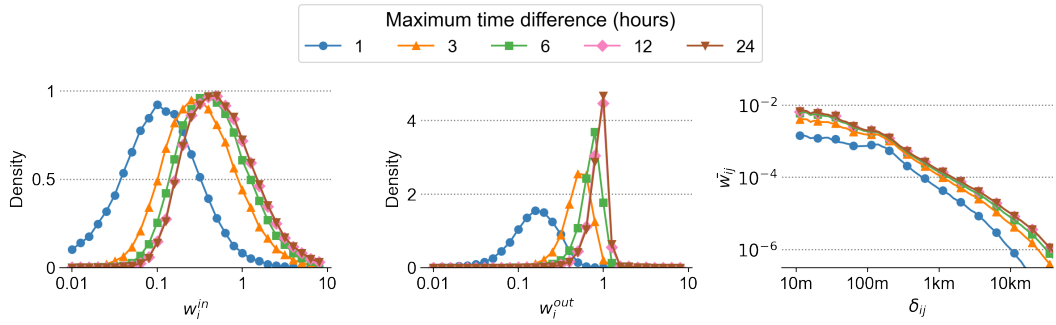


(c) Difference in category pairwise dependency weights compared to baseline network

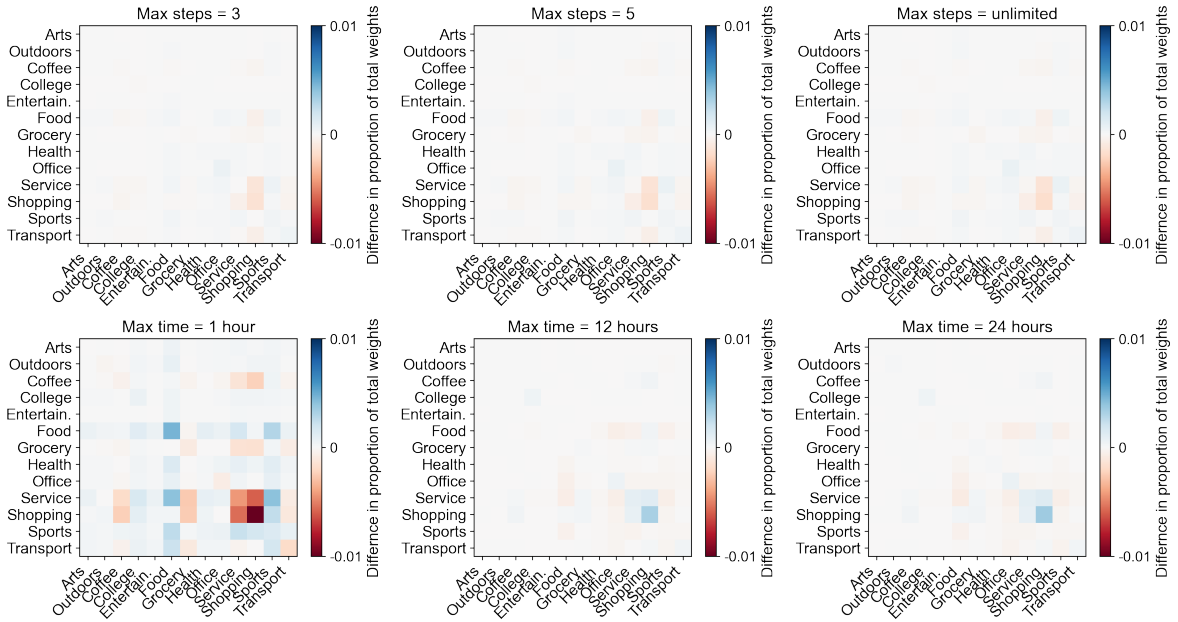
Figure S10: Dependency network in New York under different co-visit detection parameters.



(a) Network characteristics under different maximum step difference T_s

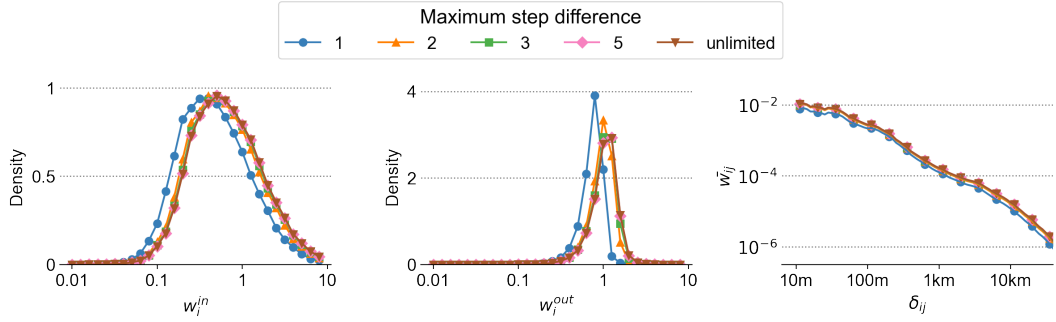


(b) Network characteristics under different maximum time difference T_c

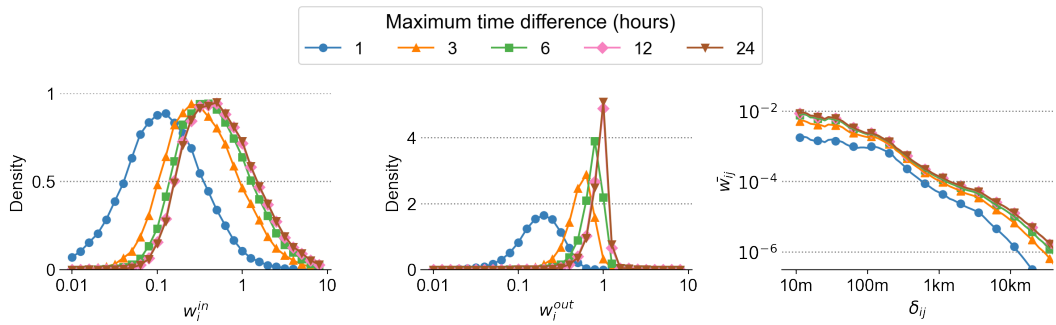


(c) Difference in category pairwise dependency weights compared to baseline network

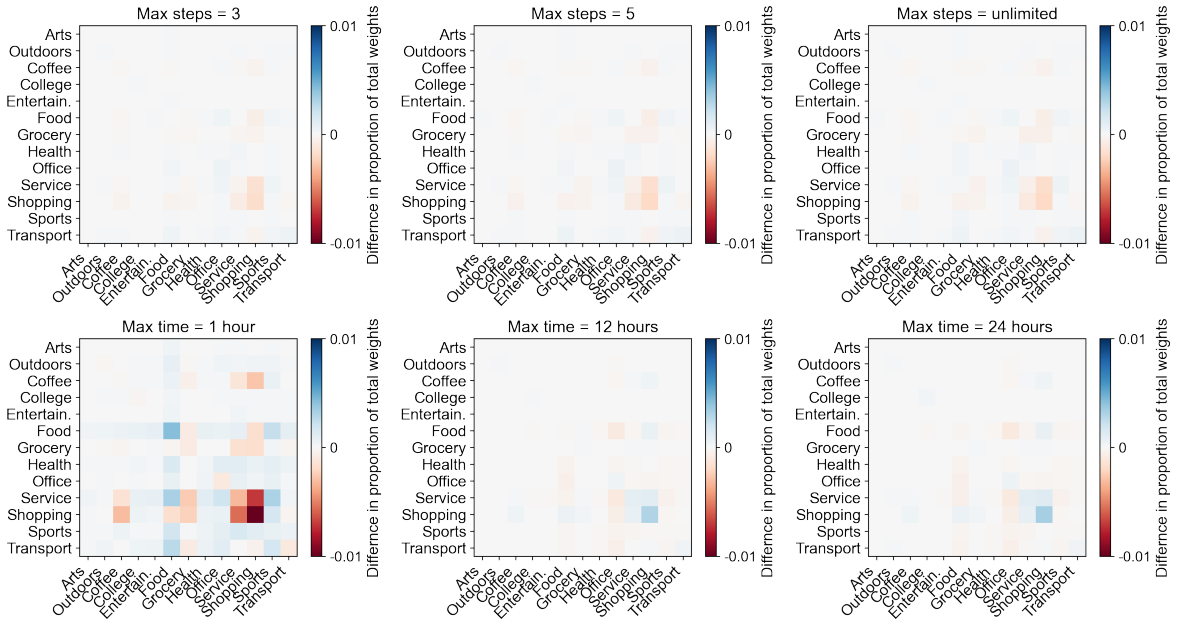
Figure S11: Dependency network in Boston under different co-visit detection parameters.



(a) Network characteristics under different maximum step difference T_s

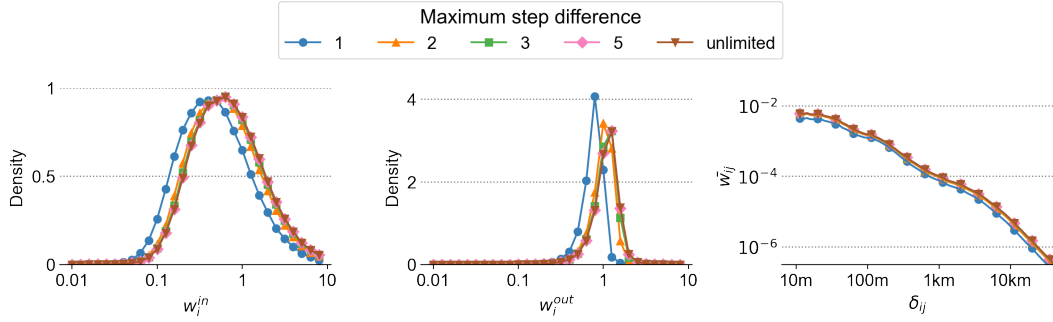


(b) Network characteristics under different maximum time difference T_c

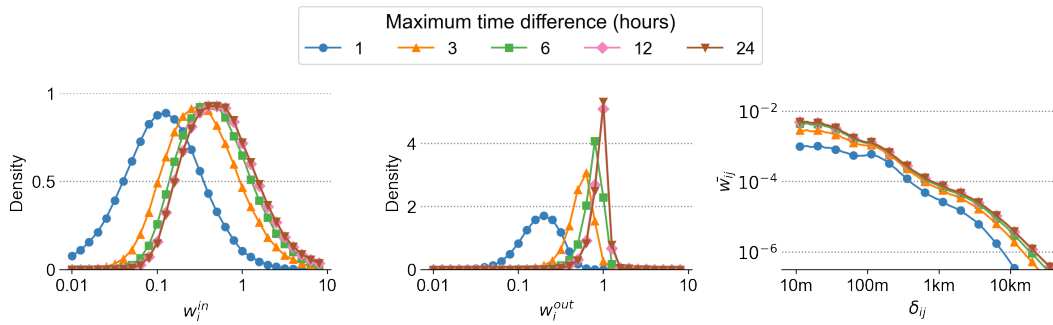


(c) Difference in category pairwise dependency weights compared to baseline network

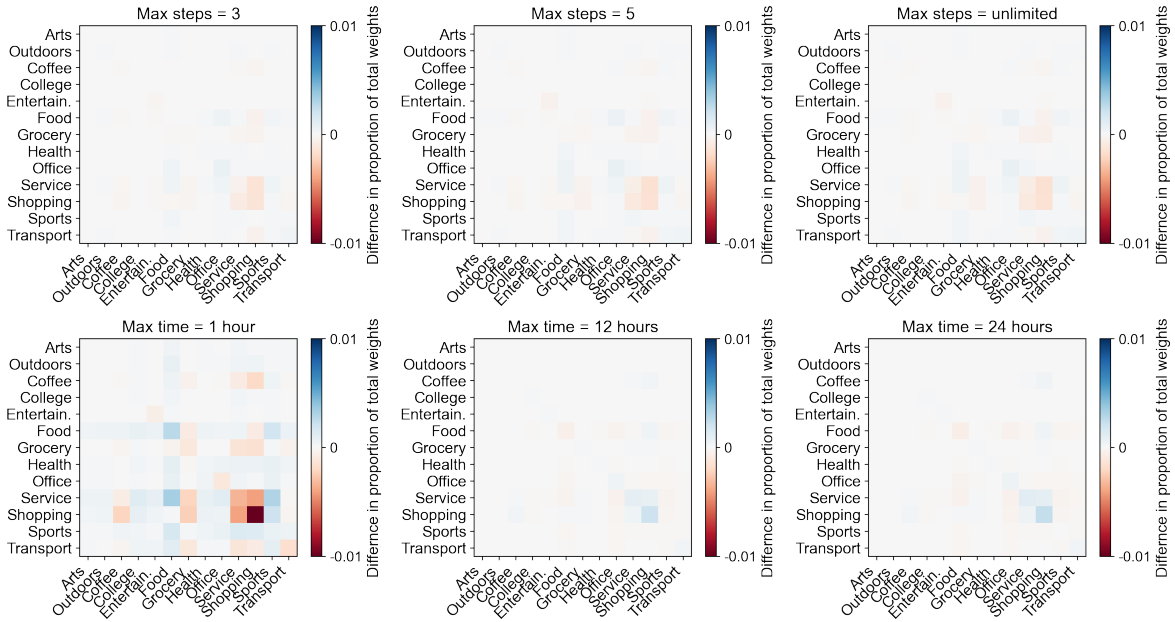
Figure S12: Dependency network in Seattle under different co-visit detection parameters.



(a) Network characteristics under different maximum step difference T_s

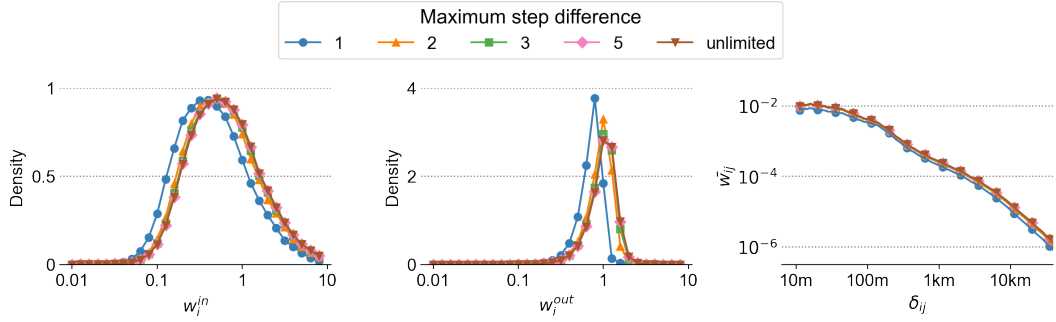


(b) Network characteristics under different maximum time difference T_c

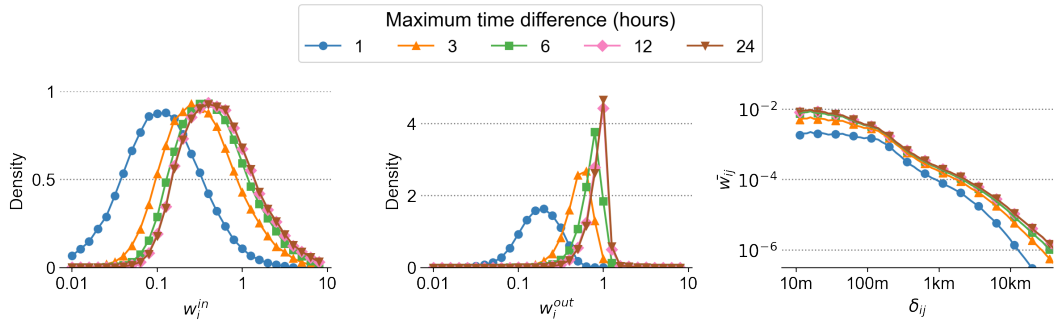


(c) Difference in category pairwise dependency weights compared to baseline network

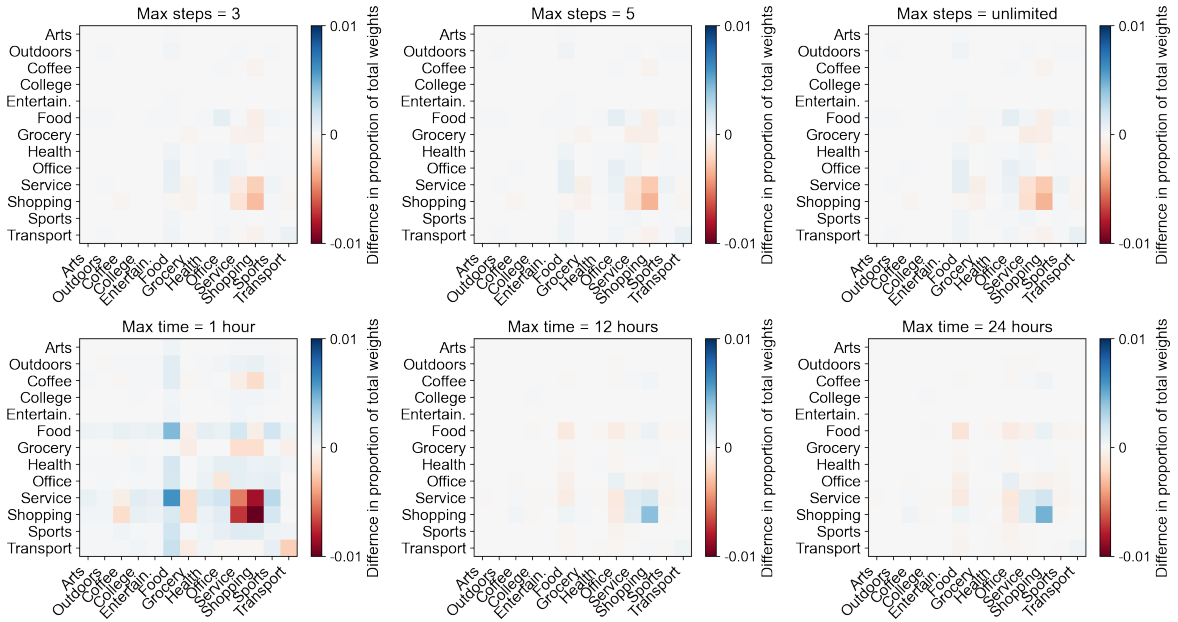
Figure S13: Dependency network in Los Angeles under different co-visit detection parameters.



(a) Network characteristics under different maximum step difference T_s



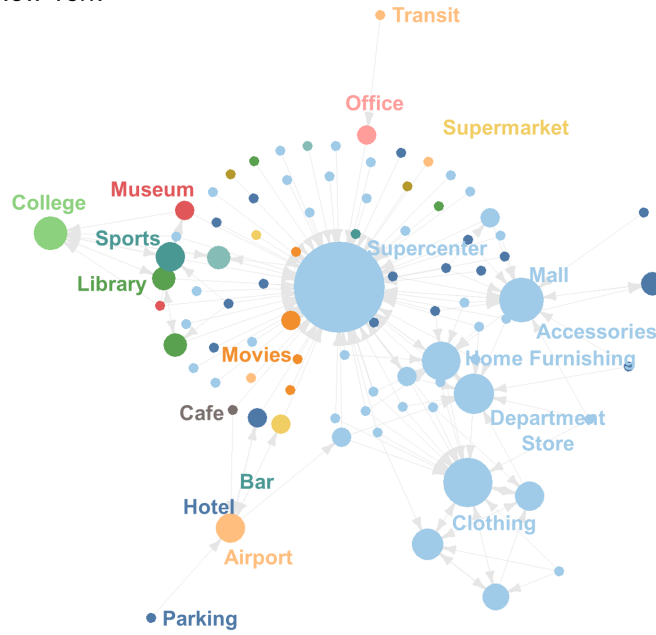
(b) Network characteristics under different maximum time difference T_c



(c) Difference in category pairwise dependency weights compared to baseline network

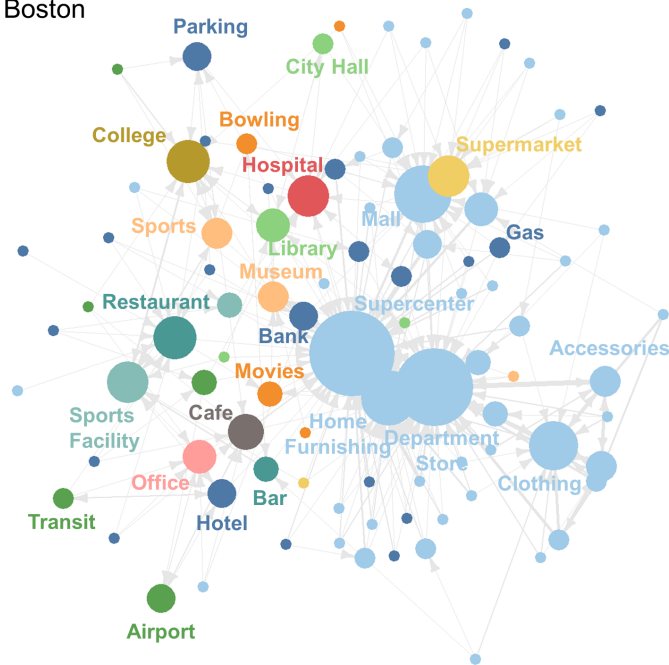
Figure S14: Dependency network in Dallas under different co-visit detection parameters.

New York



(a) New York

Boston



(b) Boston

Figure S15: Network diagram showing the average dependencies between POI subcategories in New York and Boston. Each node is a POI subcategory and the three largest outgoing dependency edges are shown for each node. Node sizes show the in-degree of the constructed network.

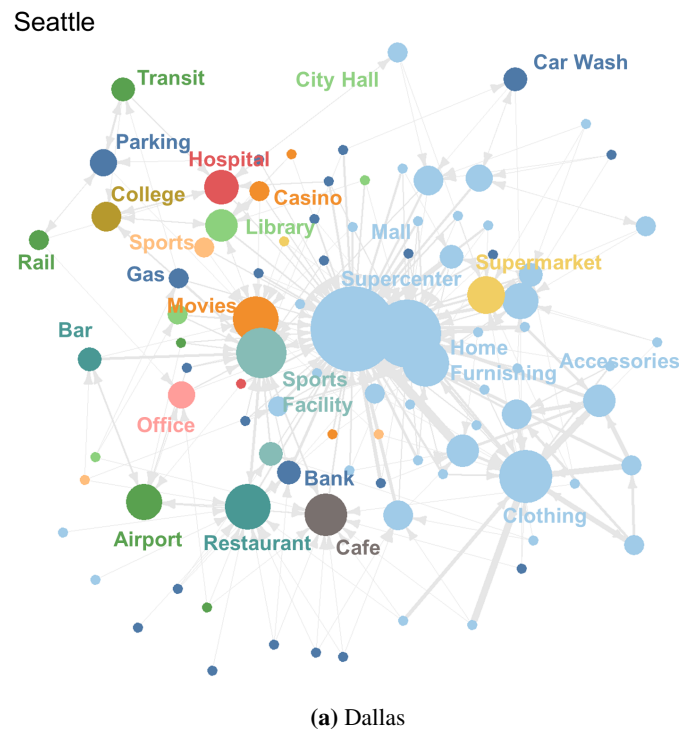


Figure S17: Network diagram showing the average dependencies between POI subcategories in Dallas. Each node is a POI subcategory and the three largest outgoing dependency edges are shown for each node. Node sizes show the in-degree of the constructed network.

3 Statistical robustness of behavior-based dependency networks

Apart from co-visit detection parameters T_c and T_s , many factors and decision parameters could affect the characteristics of the dependency network and its weights. In this section, we test the statistical robustness of the edge weights and network characteristics by comparing the dependency network with several variants. More specifically, we will make comparisons with the following networks:

- Dependency network computed using data from a different time period, for example, January – April 2019, instead of September – December 2019.
- Dependency network where edges with small weights are removed. This is done by computing the 25th percentile of the weights w_{ij} using bootstrap method.
- Dependency network calculated without the post-stratification procedure.

The following subsections show the details of how each type of network is calculated and comparisons of network statistics with the empirical dependency network. In this section, comparisons of edge-based and category-aggregated edge weights are compared between the empirical and alternative dependency networks. Further robustness checks on the modeling of network weights, the predictability of shocks during COVID-19, and resilience to future hypothetical shocks are performed in Supplementary Notes 4, 5, and 6, respectively.

3.1 Robustness against choice of time period

The baseline dependency network was generated using data collected during the period of September to December, 2019. To assess whether the dependency networks are dynamic across time, we compared the in- and out-weights of each POI across different data collection periods. Figures S18 and S19 shows that the in- and out-weights across different time periods (2019 May – August and 2019 January – April) are highly correlated ($\rho > 0.7$) with the baseline time period (2019 September – December). Moreover, the category pairwise weight proportions were compared across different data collection periods, as shown in Figures S20 and S21. The category pairs that are highly dependent on each other are consistent across different time periods (2019 May – August and 2019 January – April) with the baseline time period (2019 September – December).

3.2 Computing quartiles of w_{ij} via bootstrap method

As shown in Figure 1 of the main manuscript and other figures in the Supplementary Material (e.g., Figures S10 - S14), edge weights w_{ij} are distributed across 4 orders of magnitude, and there exist edges with very small weights (i.e., $w_{ij} < 0.001$). To test the robustness of the results presented in this paper to such extremely small edge weights, we estimate the confidence interval of each edge weight w_{ij} using bootstrap resampling.

To obtain bootstrapped samples of the dependency weights w_{ij} , the co-visits were sampled randomly with replacement from the original co-visit dataset. For each bootstrap sample the bootstrap replication of the dependency weights w_{ij}^b where $b = 1, 2, 3, \dots, 30$ were computed. The 25th percentile of the distribution of the weights, denoted by w_{ij}^{p25} were computed from the sample distribution, and only edges with $w_{ij}^{p25} > 0$ were used in the 'bootstrap network'.

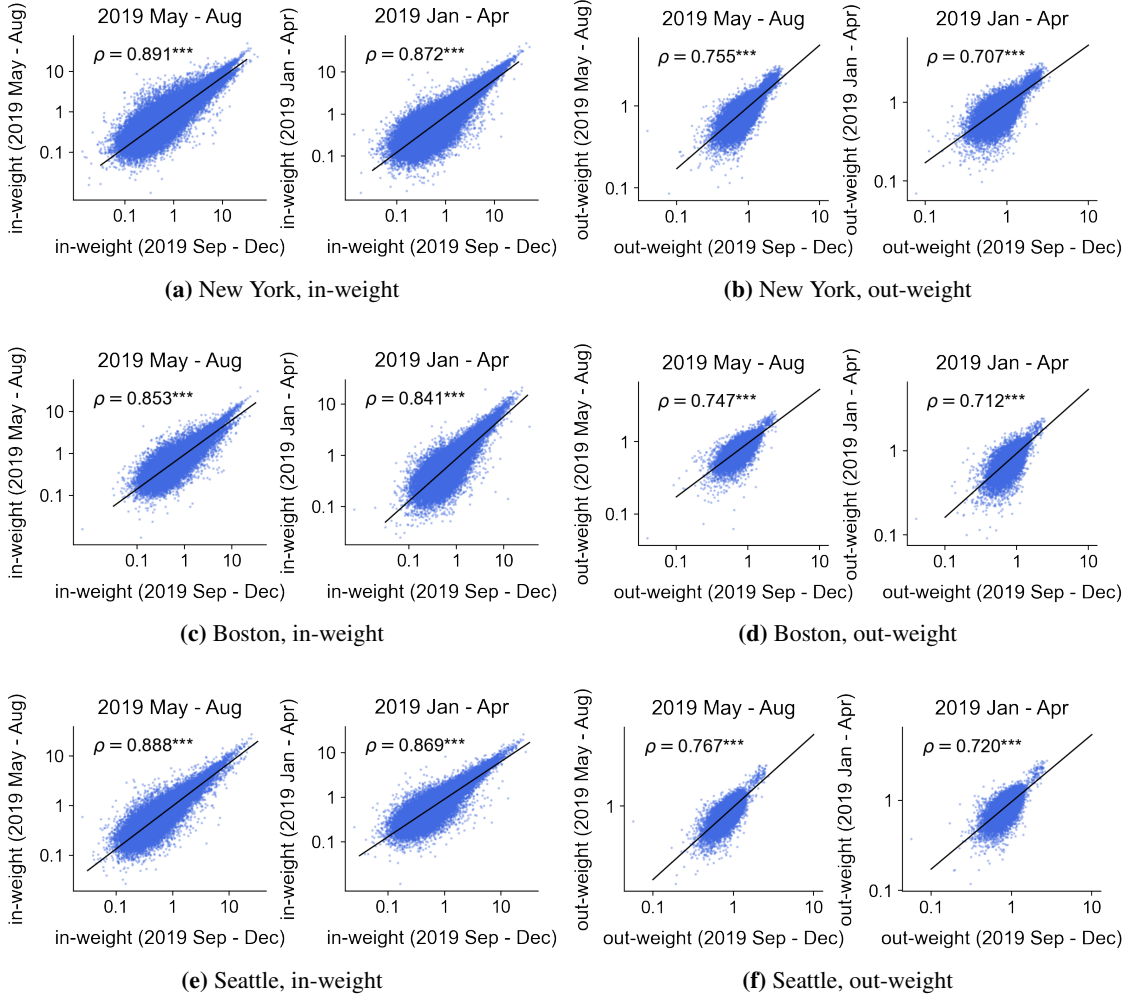


Figure S18: Comparison of in- and out-weights of each POI across different data collection periods for New York, Boston, and Seattle. In- and out-weights across different time periods (2019 May – August and 2019 January – April) are highly correlated ($\rho > 0.7$) with the baseline time period (2019 September – December).

Figure S22 shows the in- and out- weights of POIs before and after applying the bootstrap method to select statistically significant edges. The correlation is generally moderate ($\rho = 0.6$) for out-weight and high ($\rho = 0.9$) for in-weights. The variance becomes larger especially for the nodes with out-weights below 1. Figure S23 shows the out- and in-weights of categories when applying bootstrap method to remove insignificant weights. The correlation is extremely high ($\rho > 0.99$) for both out- and in-weights when aggregated to POI categories, indicating that bootstrapping does not have a significant effect when the analysis is aggregated to the category level.

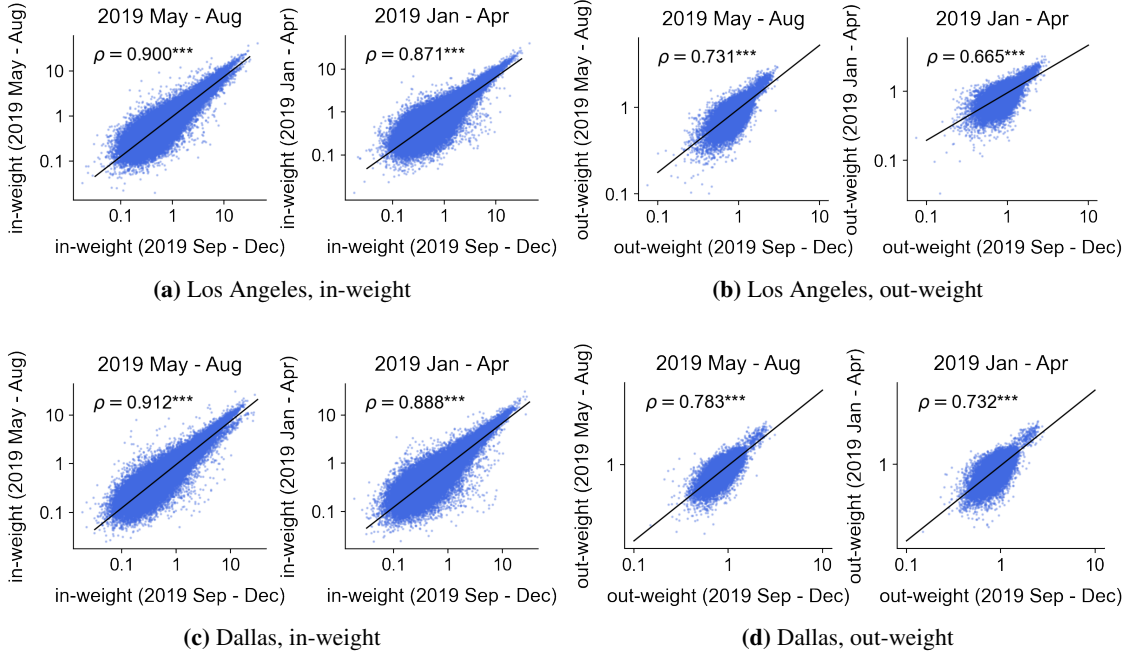


Figure S19: Comparison of in- and out-weights of each POI across different data collection periods for Los Angeles and Dallas. In- and out-weights across different time periods (2019 May – August and 2019 January – April) are highly correlated ($\rho > 0.7$) with the baseline time period (2019 September – December).

3.3 Robustness against post-stratification processing

In Supplementary Note 1, analyses showed significant biases in sampling rates of mobile phone users across census block groups and income ranges. To correct for such bias, we conducted post stratification method (Supplementary Note 1.4) and used the corrected data to estimate and analyze the dependency networks. Here, we measure how different the estimates of out- and in-weights of POIs would be if we did not apply the post-stratification technique to correct for biases in mobile phone data. In Figure S24, we observe that the correlation is high ($\rho > 0.8$) for both out- and in-weights even though there was significant bias in sample rates across CBGs and income groups.

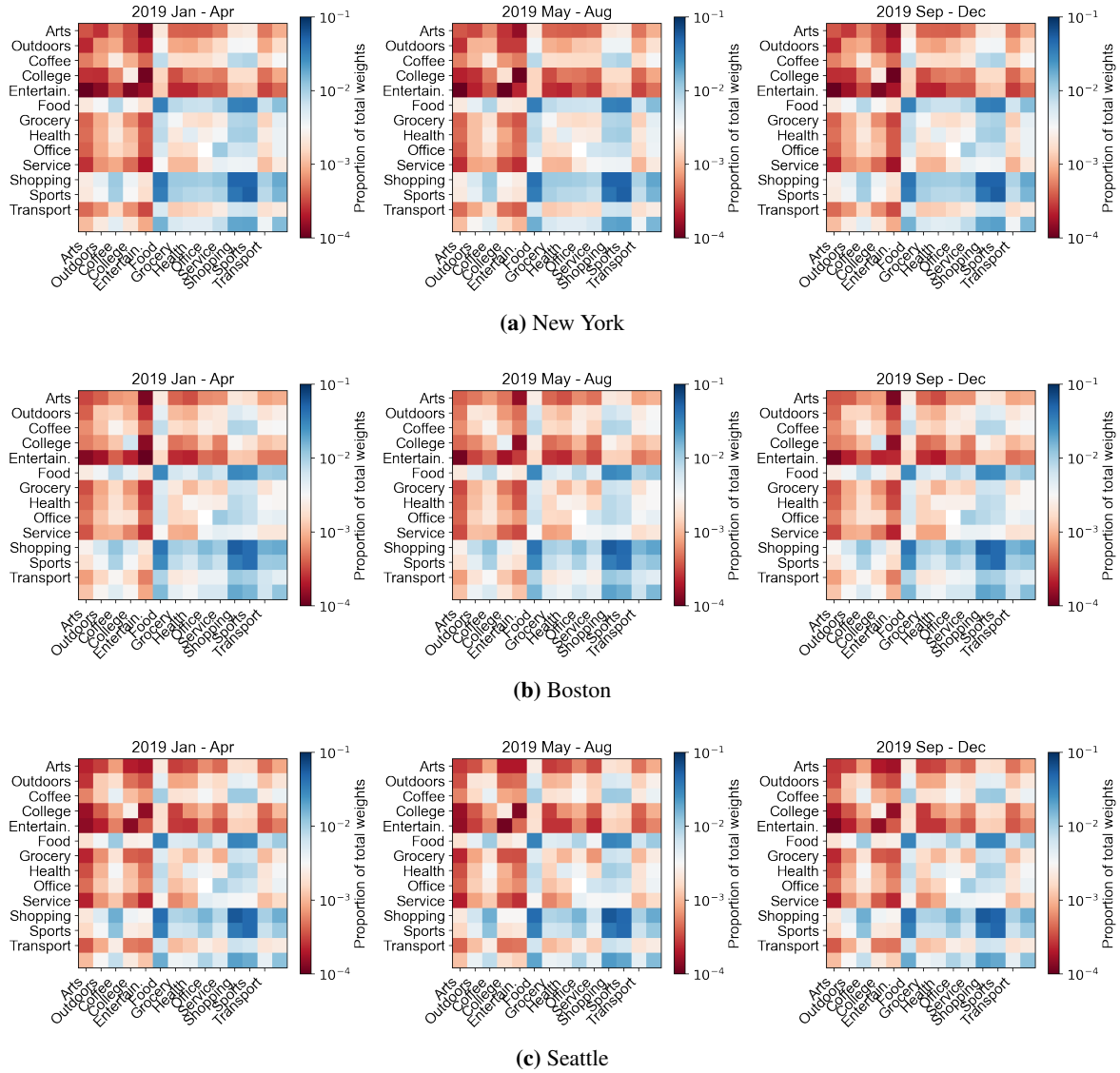


Figure S20: Comparison of category pairwise weight proportion across different data collection periods, for New York, Boston, and Seattle. The category pairs that are highly dependent on each other are consistent across different time periods (2019 May – August and 2019 January – April) with the baseline time period (2019 September – December).

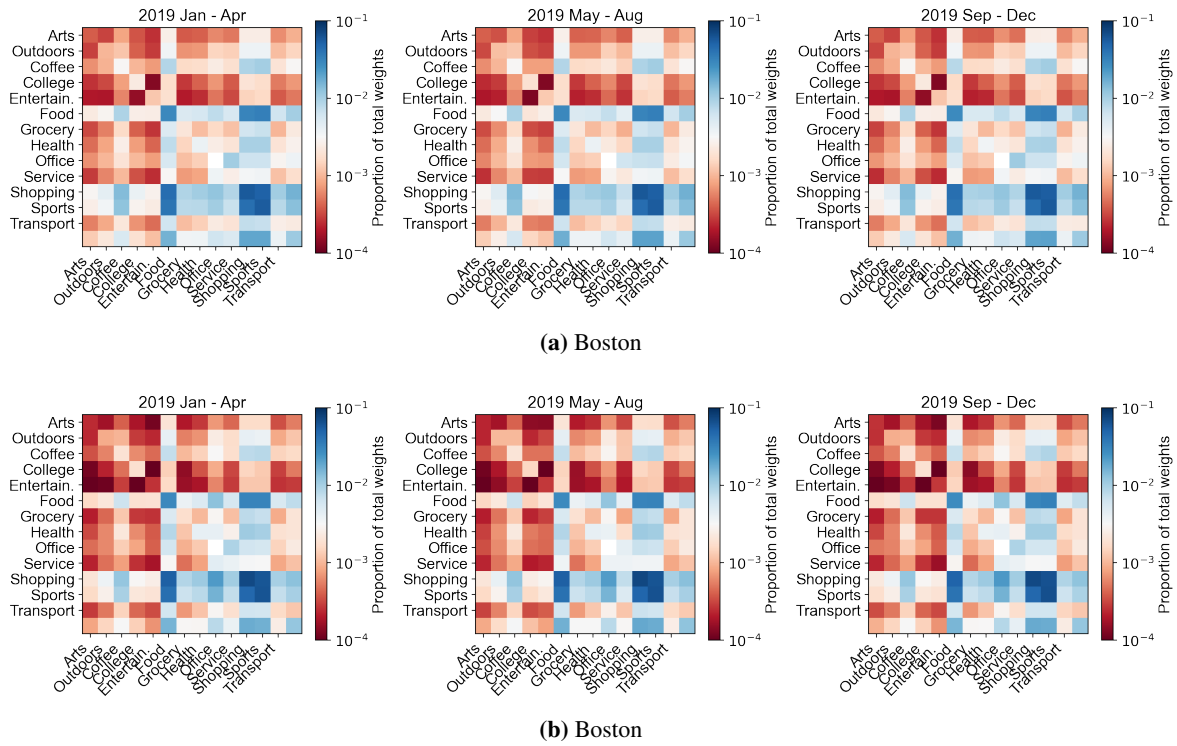


Figure S21: Comparison of category pairwise weight proportion across different data collection periods, for Los Angeles and Dallas. The category pairs that are highly dependent on each other are consistent across different time periods (2019 May – August and 2019 January – April) with the baseline time period (2019 September – December).

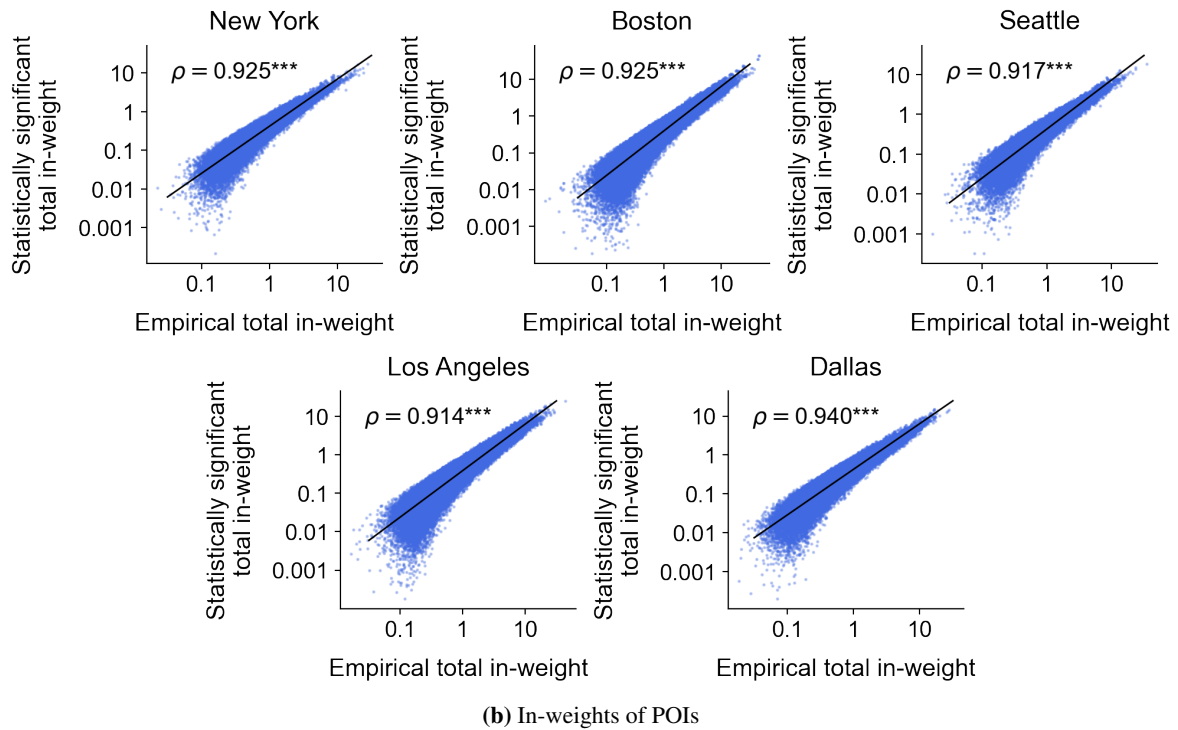
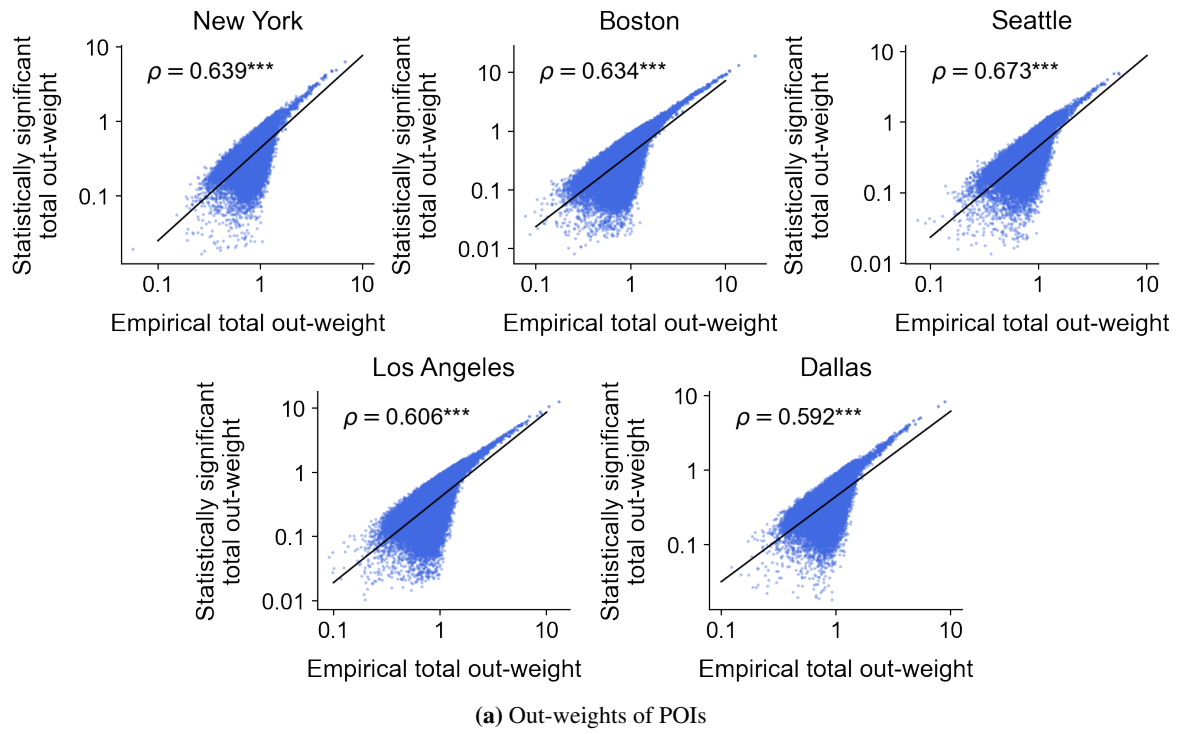
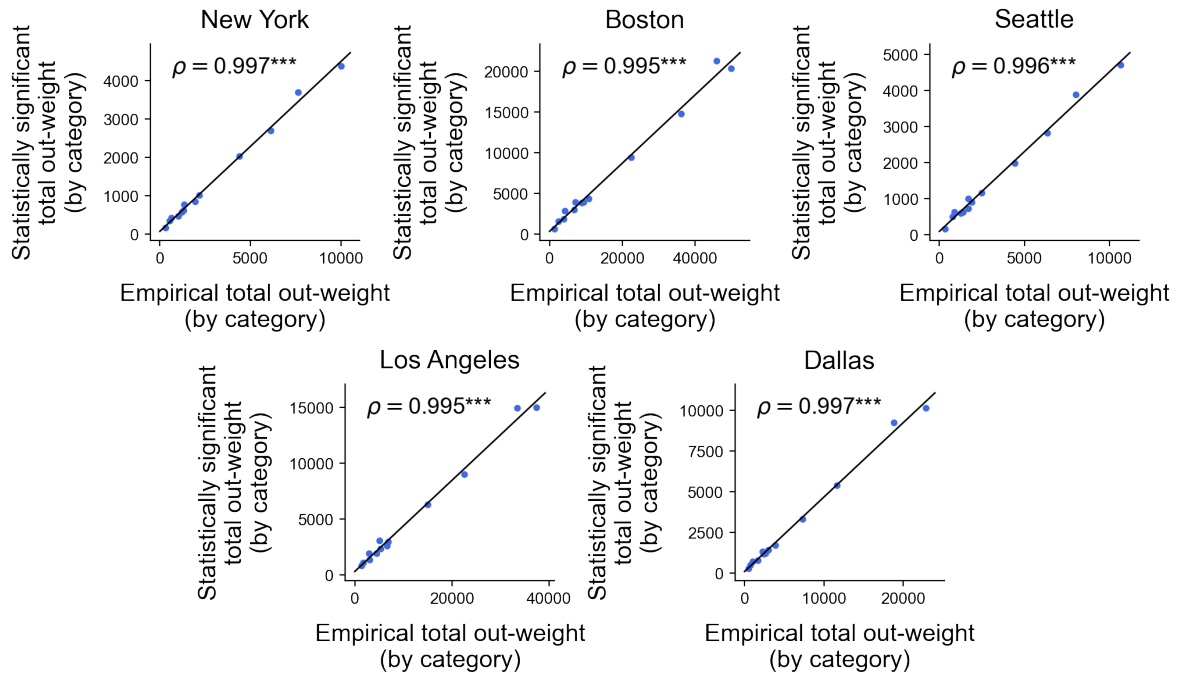
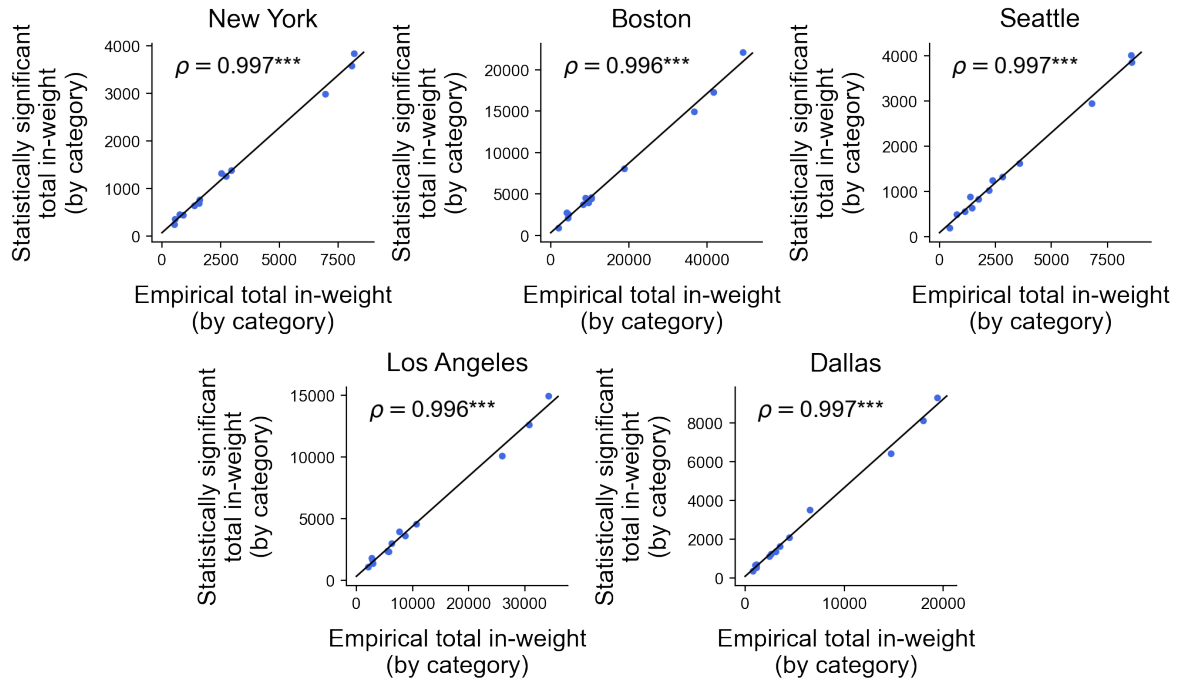


Figure S22: Out- and in-weights of POIs when applying bootstrap method to remove smaller weights. The correlation is generally moderate ($\rho = 0.6$) for out-weight and high ($\rho = 0.9$) for in-weights.



(a) Out-weights of categories



(b) In-weights of categories

Figure S23: Out- and in-weights of categories when applying bootstrap method to remove insignificant weights. The correlation is extremely high ($\rho > 0.99$) for both out- and in-weights when aggregated to POI categories.

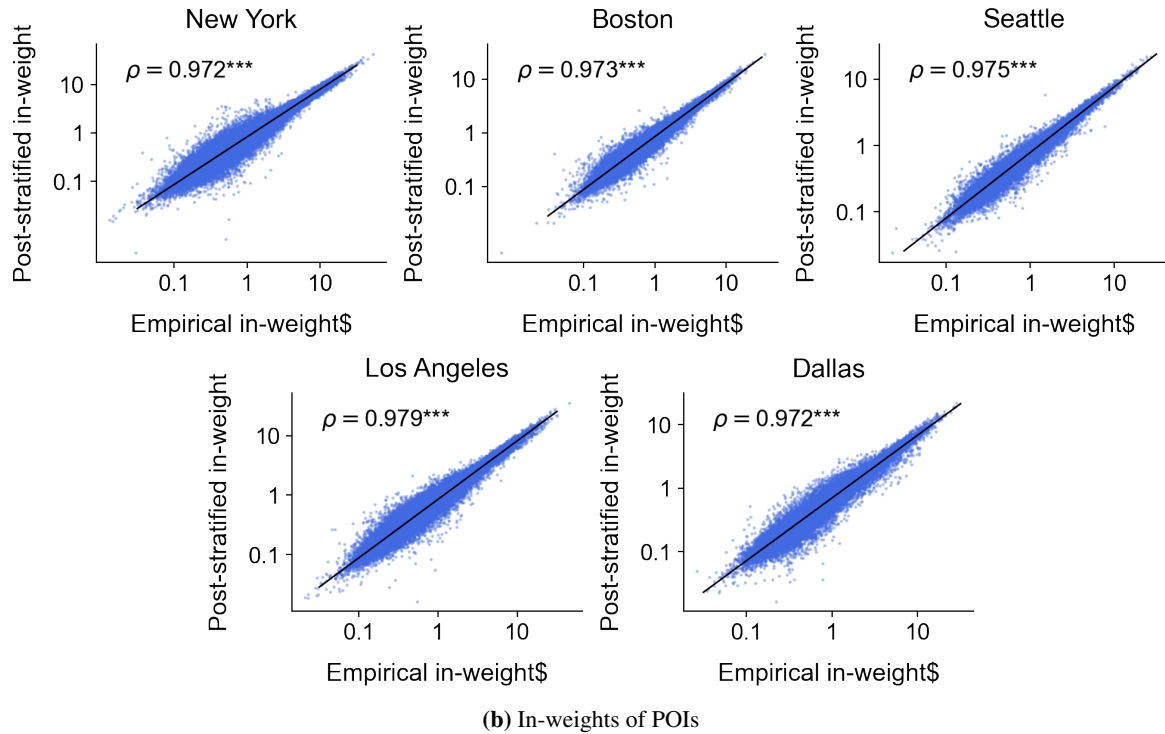
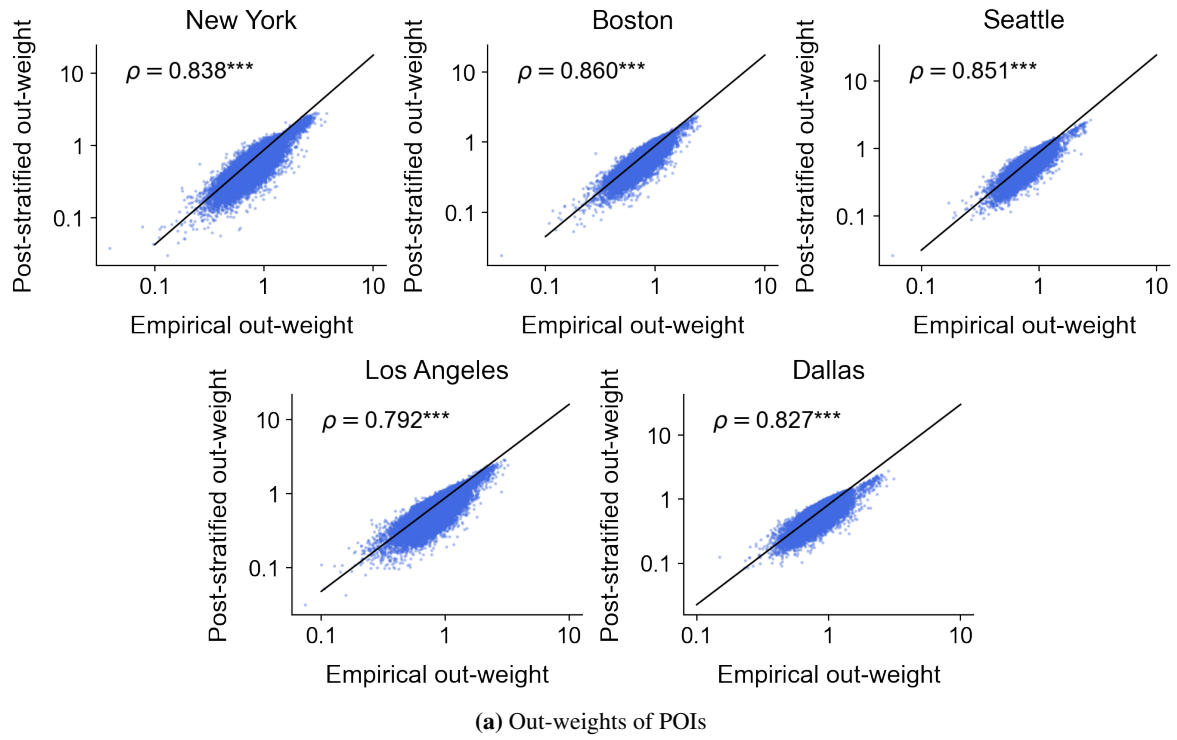


Figure S24: Out- and in-weights of POIs when applying post-stratification to correct for biases in mobile phone data. The correlation is high ($\rho > 0.8$) for both out- and in-weights even though there was significant bias in sample rates across CBGs and income groups.

4 Modeling dependency network weights

4.1 Gravity-based null networks

To further understand the structural properties of the dependency networks, we generated null networks that satisfy the following properties:

- the weight w_{ij} decays with physical distance
- the in-weight w_{ij} is larger for nodes with larger visitation n_i

To generate such null networks that preserve such qualities, we utilized the generalized gravity law as the theoretical starting point, where we can model the number of common visitors as the gravity component g_{ij} , which is the product of the total number of visitors to places i and j , n_i and n_j , divided by a function of the distance between i and j , d_{ij} , which takes the form $f(d_{ij}) = (d_0 + d_{ij})^\gamma$:

$$g_{ij} = n_i n_j / (d_0 + d_{ij})^\gamma \quad (2)$$

where n_i and n_j are the total number of visits to POIs i and j , d_{ij} is the physical distance between POIs i and j , d_0 is the distance cutoff parameter, and γ is the exponent parameter of the gravity model. Parameters d_0 and γ were fitted empirically to maximize the correlation between g_{ij} and n_{ij} , which is the total number of common visitors between POIs i and j , and as shown in Figure S25, $d_0 = 0.2$ and $\gamma = 1.5$ showed the highest correlation when the values were logged and not logged.

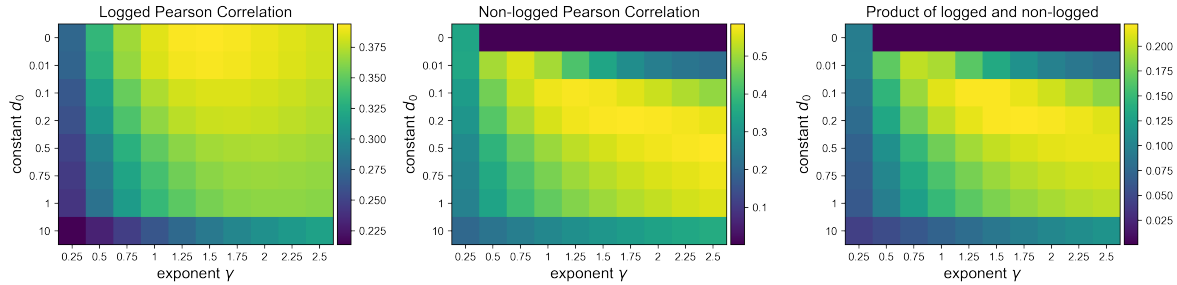


Figure S25: Correlation between the observed n_{ij} values and the gravity component g_{ij} under different model parameters

Using the fitted gravity model $w_{ij} = n_{ij}/n_i \sim n_j/(d_0 + d_{ij})^\gamma$, we now construct null networks that mimic the physical properties of the actual dependency networks. To enable a fair comparison between the actual dependency network and the null network, we constructed the null networks while fulfilling the following two properties:

1. the in-degree of each node are kept consistent in the null network
2. the in-weight of each node are kept consistent in the null network

To achieve this, first we construct a linear relationship between the gravity component $g_{ij} = n_j/(d_0 + d_{ij})^\gamma$ and w_{ij} , $g_{ij} = \phi(w_{ij})$. Then, for each edge in the actual dependency network connecting i and j with a dependency weight w_{ij} , we compute its theoretical gravity component using the linear relationship ϕ and select an alternative node with the same level of corresponding gravity component from its 10,000 closest nodes, and is assigned the same weight w_{ij} . This algorithm enables us to

construct a null network where we 1) maintain the linear relationship between w_{ij} and g_{ij} , 2) the same number of in-edges are selected for each node, and 3) the total in-weight for each node is kept consistent.

Figures S26 to S28 show the map visualizations of the actual (left) and null (right) networks for the five cities. In all cities, we can observe the null network being more locally clustered around large POIs, while the actual network is more dispersed and contains more long-distance connections. Note that the in-degrees and in-weights of all nodes are equivalent across the two networks. Figure S29 shows the histogram of physical distances of all edges in the actual and null networks. Figure S30 compares the proportion of edge weights among POI category pairs across the actual and null networks. The proportion of edge weights was computed by taking the sum of weights that connect the vertical and horizontal categories and dividing that by the total weights that exist in the network. The patterns of dependency are consistent across cities, with Service and Transport node depending more on other nodes in the real network compared to the null network.

4.2 Regression model of w_{ij}

Results comparing the relationship between dependency weights and physical distance suggest that the dependency weights could be dictated by physical factors. Here, we model dependency weights w_{ij} with various physical factors to investigate how much of the variance we observe in dependency weights can be explained by such factors. We build simple linear regression models of the form:

$$\log_{10} w_{ij} \sim \log_{10} d_{ij} + n_j + \eta_i + \eta_j + \theta_i + \theta_j \quad (3)$$

where $\log_{10} w_{ij}$ denotes the logged dependency weight of POI i on POI j , and:

- $\log_{10} d_{ij}$ is the logged Haversine distance between POIs i and j .
- n_j is the number of visitors that POI j receives. The premise is that popular POIs are depended more by other POIs. This variable was standardized $((x - \mu)/\sigma)$ prior to the analysis for comparison across cities.
- η_i is the fixed effect (FE) for POI i 's subcategory. There are 97 subcategories in the POI dataset, and examples are shown in Figure S6.
- θ_i is the fixed effect for POI i 's located Public Use Microdata Area (PUMA).

The regression results are shown in Table S3. To save computation time, 1 million edges were randomly sampled to include in the regression. The resulting R^2 or the coefficients were robust against different sampling runs. Overall, the adjusted R^2 was relatively low (between 0.10 and 0.16 for all cities), suggesting that physical factors determine a small % of the variance of dependency weights. The coefficients are similar across cities, with the logged Haversine distance showing a slower decay compared to the decay observed in Figure S10, which is due to the samples in the regression being limited to ones that have non-zero weights.

4.3 Robustness against different network instances

Table S4 shows the same regression results using the selected network edges using the bootstrap method introduced in Supplementary Note 3.2. The bootstrap method removed edges with weights that were almost zero (if the 25th percentile of the bootstrap sample distribution was below zero), and

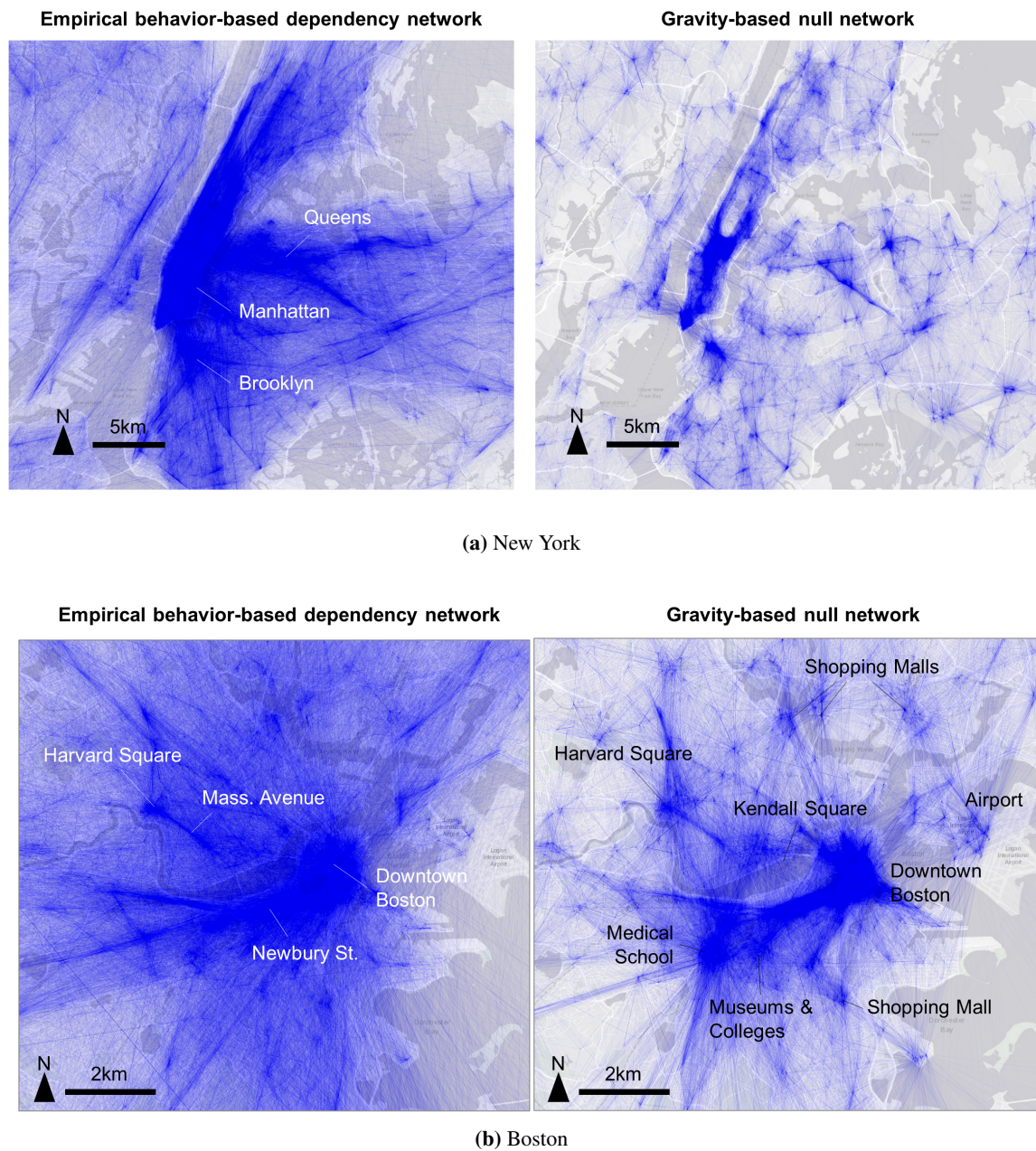
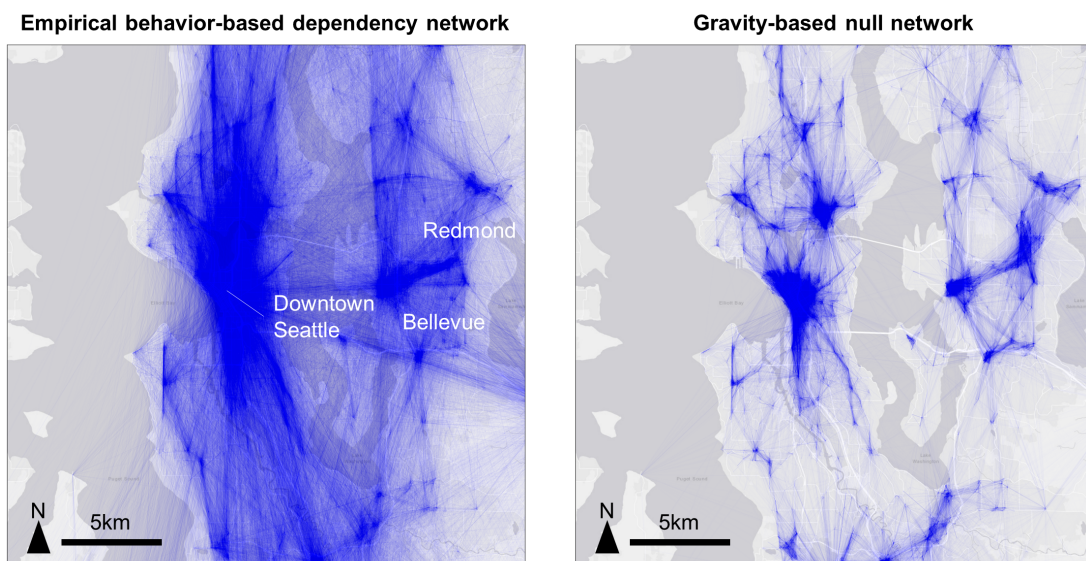
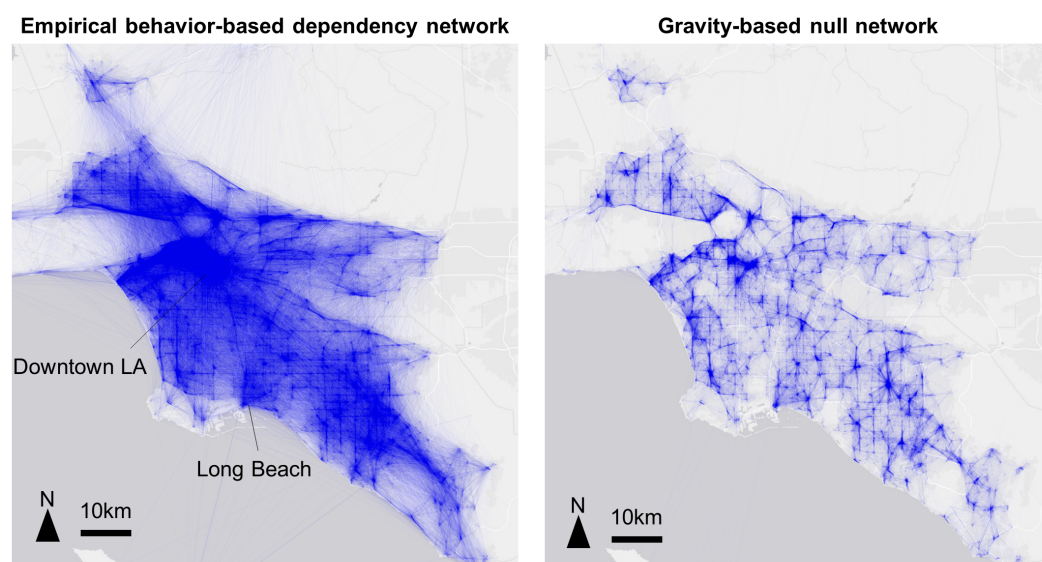


Figure S26: Comparison of actual and null networks in New York and Boston. For visual purposes, around 1 million links are shown for both the actual and null networks.



(a) New York



(b) Boston

Figure S27: Comparison of actual and null networks in Seattle and Los Angeles. For visual purposes, around 1 million links are shown for both the actual and null networks.

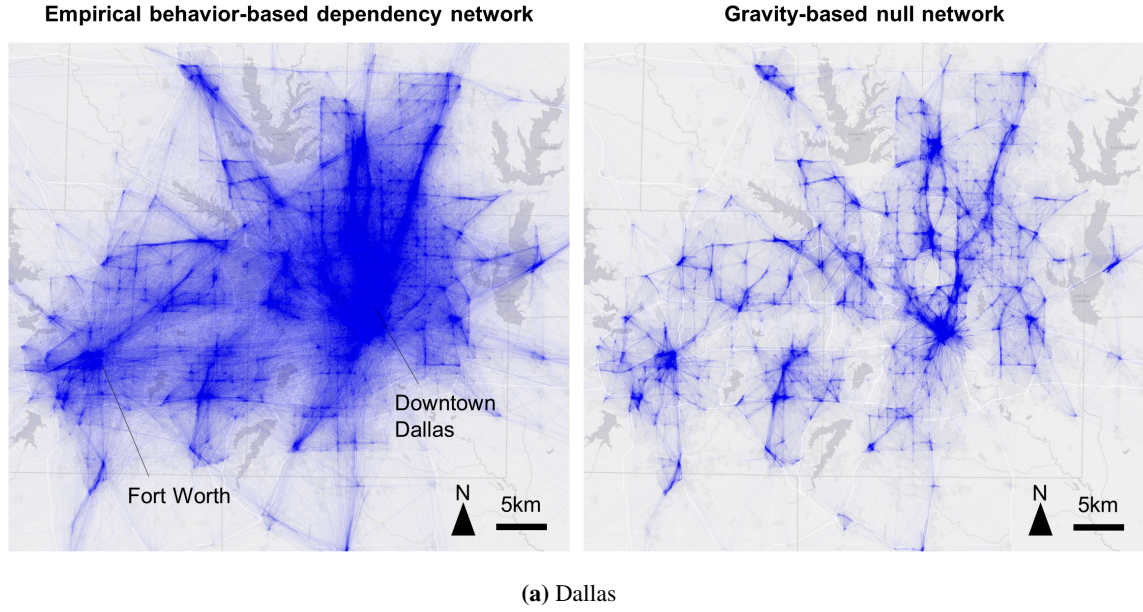


Figure S28: Comparison of actual and null networks in Dallas. For visual purposes, around 1 million links are shown for both the actual and null networks.

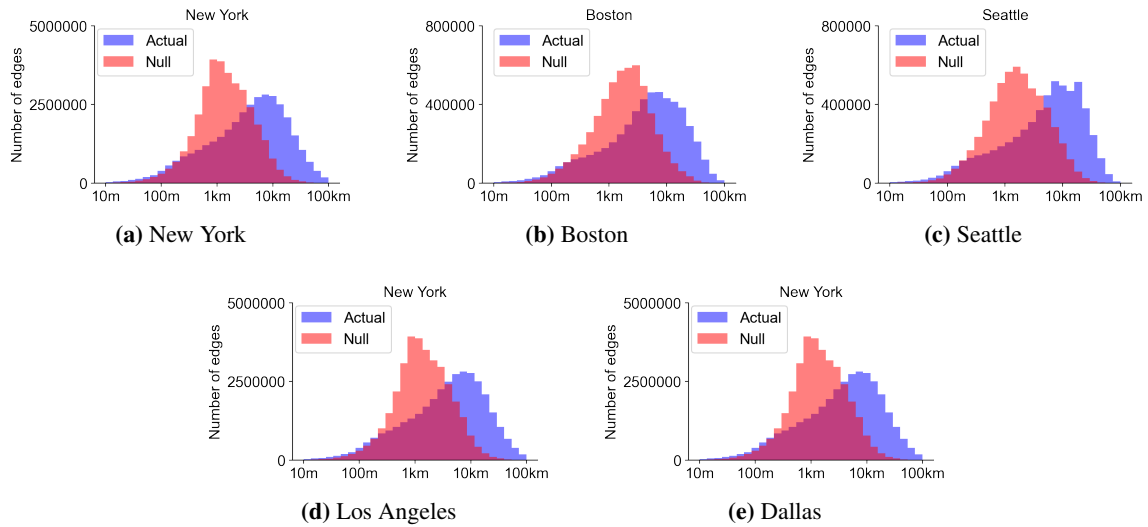


Figure S29: Comparison between real and null network on the proportion of edge weights among POI category pairs. The proportion of edge weights was computed by taking the sum of weights that connect the vertical and horizontal categories and dividing that by the total weights that exist in the network. The patterns of dependency are consistent across cities, with Service and Transport node depending more on other nodes in the real network compared to the null network.

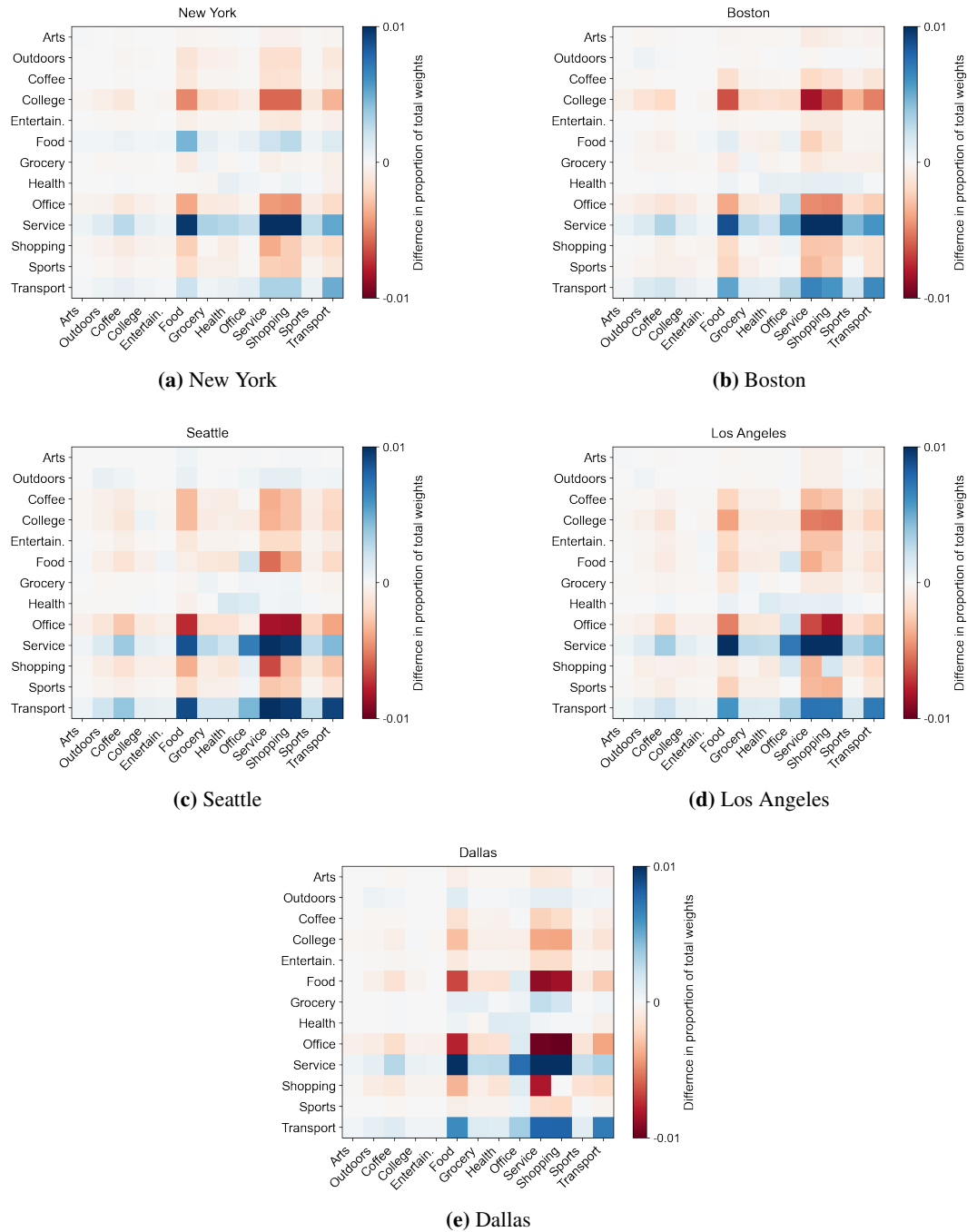


Figure S30: Histograms of edge distances between actual and null networks. The actual networks contain more long distance edges than the null networks.

Table S3: Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10} w_{ij}$) in US cities using the logged Haversine distance ($\log_{10} d_{ij}$), the popularity in visits to place j (v_j), and the subcategory fixed effects and PUMA fixed effects of POIs i and j . Variables were centered and standardized prior to analysis.

Variable	Dependent Variable: $\log_{10} w_{ij}$				
	New York	Boston	Seattle	Los Angeles	Dallas
Constant	-2.268 ***	-2.291 ***	-2.315 ***	-2.312 ***	-2.408 ***
$\log_{10} d_{ij}$	-0.129 ***	-0.147 ***	-0.134 ***	-0.125 ***	-0.143 ***
v_j (standardized)	0.022 ***	0.026 ***	0.027 ***	0.020 ***	0.022 ***
Subcategory FE for i and j	Y	Y	Y	Y	Y
PUMA FE for i and j	Y	Y	Y	Y	Y
Observations (sampled)	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Adj. R^2	0.089	0.117	0.122	0.108	0.115

as a result, as shown in Figure S22(a), the POIs with smaller weights (especially out-weights) experienced a significant decrease in edge weights. The regression results show a much higher R^2 compared to the original results using all of the edges, explaining around 20% to 30% of the variance in the dependency weights. The results are once again consistent across cities, with regression coefficients in the same range.

Table S5 shows the same regression results using the dependency network computed using only short stays (< 4 hours). The results of the regression (adjusted R squared and the model coefficients) are consistent with the results from the full network shown in Table S3.

Table S4: Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10} w_{ij}$) in US cities which were selected using the bootstrap method, using the logged Haversine distance ($\log_{10} d_{ij}$), the popularity in visits to place j (v_j), and the subcategory fixed effects and PUMA fixed effects of POIs i and j . Variables were centered and standardized prior to analysis.

Variable	Dependent Variable: $\log_{10} w_{ij}$				
	New York	Boston	Seattle	Los Angeles	Dallas
Constant	-2.067 ***	-1.978 ***	-2.100 ***	-2.249 ***	-2.005 ***
$\log_{10} d_{ij}$	-0.163 ***	-0.199 ***	-0.186 ***	-0.188 ***	-0.232 ***
v_j (standardized)	0.025 ***	0.033 ***	0.032 ***	0.030 ***	0.032 ***
Subcategory FE for i and j	Y	Y	Y	Y	Y
PUMA FE for i and j	Y	Y	Y	Y	Y
Observations (sampled)	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Adj. R^2	0.136	0.192	0.172	0.215	0.244

Table S5: Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10} w_{ij}$) in US cities which were selected using only short stays (≤ 4 hrs), using the logged Haversine distance ($\log_{10} d_{ij}$), the popularity in visits to place j (v_j), and the subcategory fixed effects and PUMA fixed effects of POIs i and j . Variables were centered and standardized prior to analysis.

Variable	Dependent Variable: $\log_{10} w_{ij}$				
	New York	Boston	Seattle	Los Angeles	Dallas
Constant	-2.182 ***	-2.220 ***	-2.202 ***	-2.280 ***	-2.259 ***
$\log_{10} d_{ij}$	-0.131 ***	-0.149 ***	-0.134 ***	-0.126 ***	-0.145 ***
v_j (standardized)	0.022 ***	0.029 ***	0.028 ***	0.021 ***	0.024 ***
Subcategory FE for i and j	Y	Y	Y	Y	Y
PUMA FE for i and j	Y	Y	Y	Y	Y
Observations (sampled)	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Adj. R^2	0.096	0.125	0.131	0.117	0.121

4.4 Robustness against co-visit detection parameters

The regression analysis showed two main findings: 1) the physical factors can only explain a fraction of the variance in dependency weights, and 2) the coefficient of the logged Haversine distance is moderate compared to distance exponents used in traditional models such as the Gravity Model (i.e., -2). To assess the robustness of these findings of the regression model, the regression was performed using dependency networks generated using different co-visit detection parameters (also used in Supplementary Note 2.3). The 10 pairs of parameters ($T_s = [1, 2, 3, 5, \text{unlimited}]$ and $T_c = [1, 3, 6, 12, 24]$ hours) were used to generate the dependency networks for each city, and the regression results (adjusted R^2 and coefficient for Haversine distance) are shown in Figure S31. The matrix plots across the five metropolitan areas show similar patterns, both the adjusted R^2 and coefficient for Haversine distance are relatively different when the maximum time difference is 1 hour. In the rest of the parameter pairs, we obtain a similar result compared to Table S3.

4.5 Robustness against different time periods

To check the robustness of the results with respect to the time periods used to generate the dependency network, the same regression experiments were performed using 2019 January – April data and 2019 May – August data, as shown in Tables S6 and S7. We can observe that the R^2 values and the regression coefficients are similar with those obtained in Table S3.

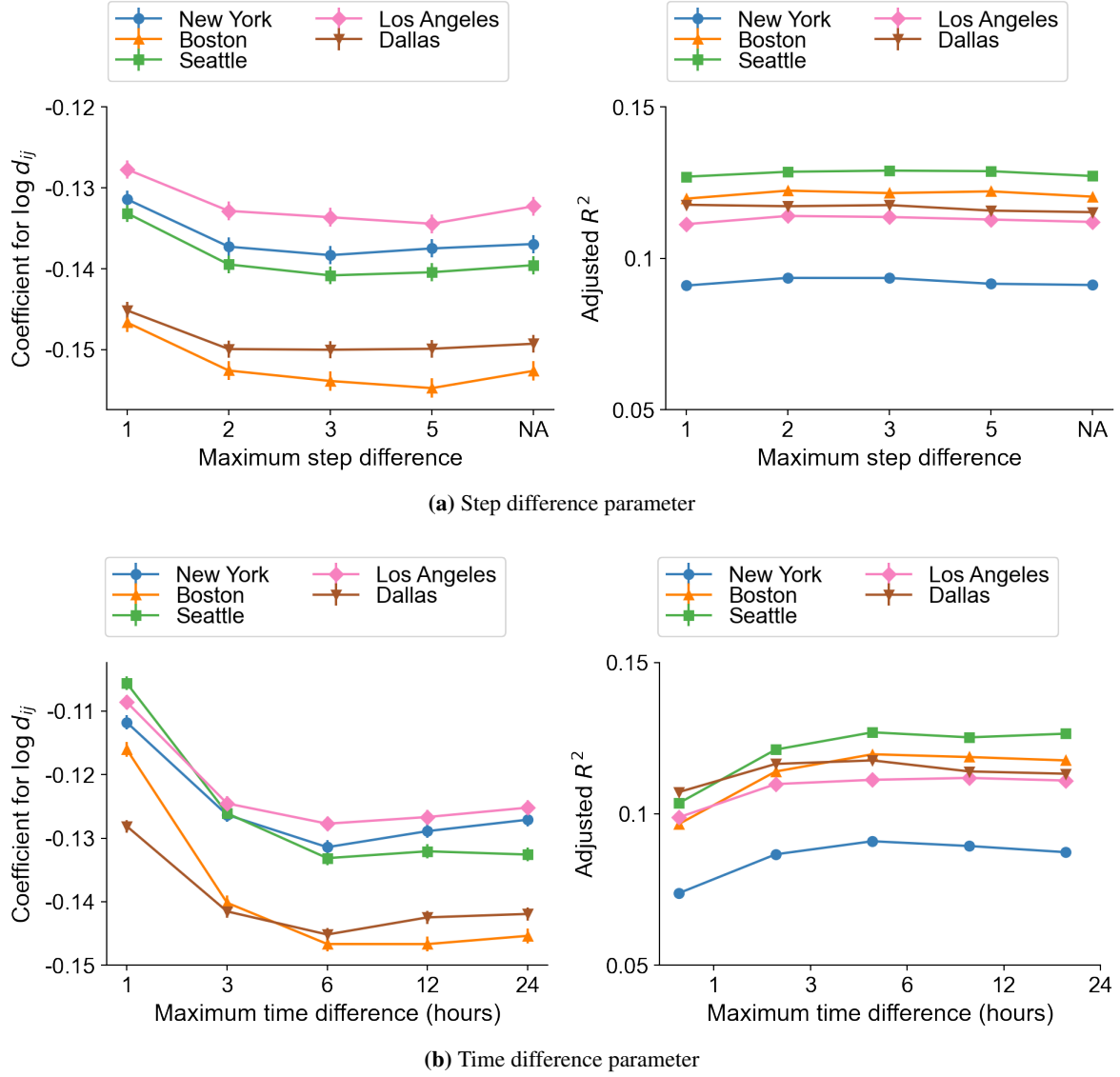


Figure S31: Regression results for dependency weights when using different co-visit detection parameters, continued for Los Angeles and Dallas. Except for the networks generated using maximum time difference of 1 hour, the results (adjusted R^2 and coefficient for Haversine distance) are relatively similar as the results in Table S3.

Table S6: Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10} w_{ij}$) in US cities using network from 2019 January - April data.

Variable	Dependent Variable: $\log_{10} w_{ij}$ (2019 January - April data)				
	New York	Boston	Seattle	Los Angeles	Dallas
Constant	-2.157 ***	-2.242 ***	-2.273 ***	-2.264 ***	-2.268 ***
$\log_{10} d_{ij}$	-0.128 ***	-0.146 ***	-0.132 ***	-0.127 ***	-0.144 ***
v_j (standardized)	0.019 ***	0.029 ***	0.026 ***	0.019 ***	0.021 ***
Subcategory FE for i and j	Y	Y	Y	Y	Y
PUMA FE for i and j	Y	Y	Y	Y	Y
Observations (sampled)	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Adj. R^2	0.092	0.120	0.121	0.112	0.117

Table S7: Linear regression models predicting the logged dependency weights between POIs i and j ($\log_{10} w_{ij}$) in US cities using network from 2019 May - August data.

Variable	Dependent Variable: $\log_{10} w_{ij}$ (2019 May - August data)				
	New York	Boston	Seattle	Los Angeles	Dallas
Constant	-2.312 ***	-2.292 ***	-2.320 ***	-2.333 ***	-2.421 ***
$\log_{10} d_{ij}$	-0.130 ***	-0.146 ***	-0.134 ***	-0.127 ***	-0.143 ***
v_j (standardized)	0.020 ***	0.023 ***	0.024 ***	0.019 ***	0.021 ***
Subcategory FE for i and j	Y	Y	Y	Y	Y
PUMA FE for i and j	Y	Y	Y	Y	Y
Observations (sampled)	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Adj. R^2	0.090	0.114	0.118	0.108	0.114

5 Impacts of behavior-based dependency during the COVID-19 pandemic

5.1 Analysis of visitation losses in cities

To investigate the usefulness of the dependency network for understanding the resilience of businesses, we construct a regression modeling framework that predicts the change in visitation patterns to a POI using information about the change in visitation patterns to its alters and the dependency network. The observed change in visits to different places is computed by:

$$\tilde{v}_i = \left(\frac{v_i^{after}}{v_i^{before}} - 1 \right) * 100(\%) \quad (4)$$

where v_i^{before} and v_i^{after} denote the number of visits to place i before the pandemic (September - December 2019) and during the pandemic period (March - June 2020), respectively. The left side panels in Figure S32 plots the distribution of the change in visits $\tilde{v}_i$ in the five metropolitan areas. Overall, as expected, we observe a peak at around $\tilde{v}_i = -75$ for all cities, suggesting the overall substantial decrease in visits during the initial stages of the pandemic. The right side panels in Figure S32 plots the average change in visits across different POI categories. We observe heterogeneity across POI categories, where Arts, Colleges, Food, and Sports POIs experienced the largest losses during the initial stages of the pandemic. On the other hand, Outdoor and Grocery places had a minor impact compared to the other POI types. Moreover, prior studies have found that POIs located in urban areas were affected more compared to rural areas (e.g., [18]). Therefore, we take into consideration the effects of the category of the places and the area in which the POIs are located to predict the visitation losses to POIs.

To test the hypothesis that the change in visitation patterns to a POI can be modeled using information about the change in visitation patterns to its alters and the dependency network, we first investigate the simple correlation between the two metrics. Figure S33 shows the partial correlation $\rho(\tilde{v}_i, \sum_j w_{ij} \tilde{v}_j)$, where POI i and j 's categories are A (vertical) and B (horizontal), respectively. Overall, several category pairs exhibit significant and strong correlation. For example, coffee POIs' changes in visits are strongly correlated with the weighted changes in visits to connected food, service, and shopping POIs across the five metropolitan areas. This significant and strong correlation suggests that the changes in the visitation of the ego may be predicted using the information of the alter POIs connected via the dependency network.

5.2 Model specification and estimation results

Using the correlations revealed in the previous section, we build a simple linear regression model of the form:

$$\tilde{v}_i \sim \sum_j w_{ij} \tilde{v}_j + \sum_j \delta_{ij} \tilde{v}_j + \eta_i + \theta_i \quad (5)$$

where $\tilde{v}_i$ denotes the change in visitations to POI i during the initial stages of the pandemic (March 1st - May 21st, 2020), and:

- $\sum_j w_{ij} \tilde{v}_j$ is the sum of the neighbors' (POIs j) change in visitations ($\tilde{v}_j$) weighted by the dependency network weights w_{ij} .

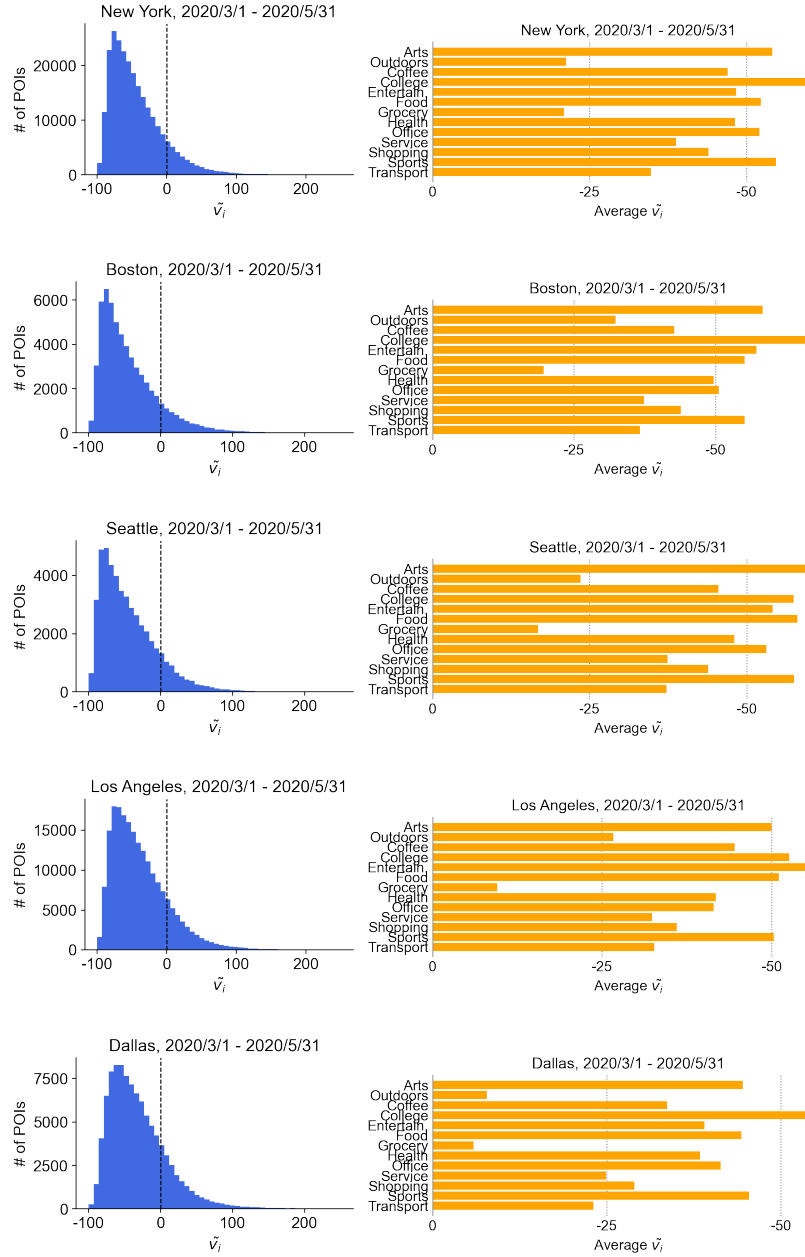


Figure S32: Change in visits to POIs in New York, Boston, Seattle, Los Angeles, and Dallas during the pandemic period. The left panels show the distribution of the change in visits, and the right panels show the average effects per POI category.

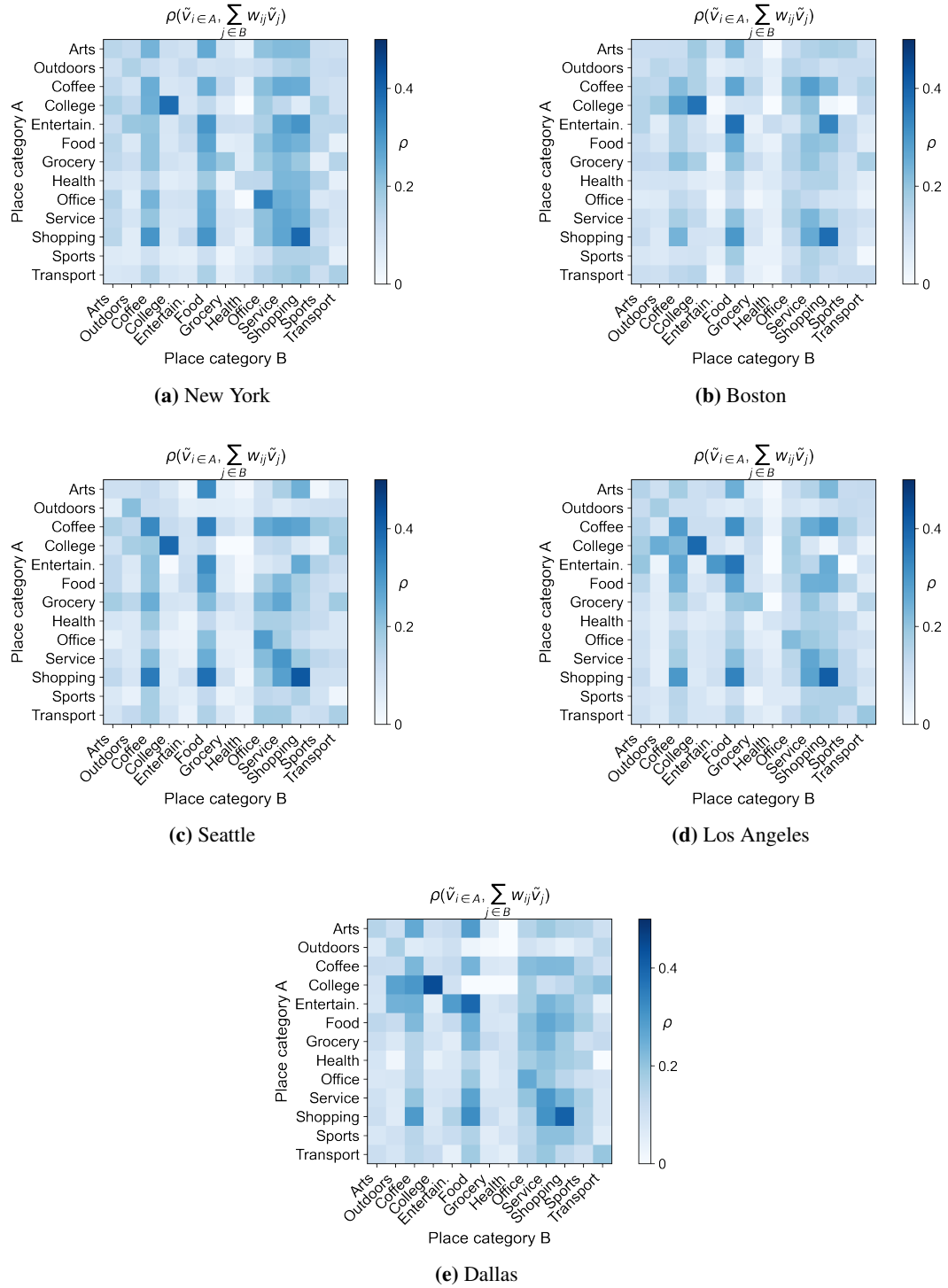


Figure S33: Correlation between change in visits of ego and the weighted change in visits of alters by category pairs. Significant correlation between several category pairs suggest predictability of $\tilde{v}_i$ using information of alters.

- $\sum_j \delta_{ij} \tilde{v}_j$ is the sum of the neighbors' (POIs j) change in visitations ($\tilde{v}_j$) weighted by the distance based network weights δ_{ij} . The distance based network weights mimic the patterns observed between physical distance and dependency weights, which are constant until 100 meters, but decays as the inverse squared of the actual distance beyond 100 meters.
- η_i is the fixed effect (FE) for POI i 's subcategory. There are 97 subcategories in the POI dataset, and examples are shown in Figure S6.
- θ_i is the fixed effect for POI i 's located Public Use Microdata Area (PUMA).

Tables S9 to S12 show the regression results for New York, Boston, Seattle, Los Angeles, and Dallas, under the following four models:

- (1) using only the subcategory and PUMA fixed effects,
- (2) using the distance-based network effects $\sum_j \delta_{ij} \tilde{v}_j$ in addition to the subcategory and PUMA fixed effects,
- (3) using the behavior-based dependency network effects $\sum_j w_{ij} \tilde{v}_j$ in addition to the subcategory and PUMA fixed effects, and
- (4) using all of the factors introduced in the full model.

Both the distance-based network effects and behavior-based dependency network effects were standardized (by subtracting the mean and dividing by the standard deviation) before analysis, therefore the magnitude of the coefficients can be compared.

In all cities, the model performance significantly increases when we use the behavior-based dependency network effects in model (3), compared to the baseline model (1) and distance-based network effects model (2). Including both distance-based network effects and the behavior-based network effects slightly increases the R^2 , however, the estimated coefficients for $\sum_j w_{ij} \tilde{v}_j$ are substantially larger (5 to 15 fold) than the coefficients of $\sum_j \delta_{ij} \tilde{v}_j$, indicating the significance of the dependency network effects.

5.3 Robustness against co-visit detection parameters

To test the robustness of the prediction power of the models using the dependency-based network effects, we test the adjusted R^2 of the models using dependency networks generated using different parameters for the co-visit detection algorithm. Similar to Supplementary Note 4.3, we test the regression models under 25 pairs of parameters ($T_s = [1, 2, 3, 5, unlimited]$ and $T_c = [1, 3, 6, 12, 24]$ hours). Figure S34 shows the adjusted R^2 values of the models using dependency networks generated with the corresponding co-visit detection parameters. Across all five metropolitan areas, we observe the general trend of the performance of the models increasing with longer maximum time differences and with smaller number of between-visit steps. This suggests that, for detecting the impact of the COVID-19 pandemic on businesses, we need to observe 1-step (thus direct) visitation patterns between POIs, regardless of the interval time between the entry and exit.

Table S8: Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in New York.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-40.9*** (5.345)	-41.6*** (5.110)	-41.7*** (5.110)	-42.2*** (5.089)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.278*** (0.085)		3.410*** (0.082)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.59*** (0.103)	14.26*** (0.103)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	213,140	213,140	213,140	213,140
R^2	0.158	0.168	0.230	0.237
Adjusted R^2	0.157	0.167	0.230	0.236

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S9: Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Boston.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-20.1*** (6.512)	-22.1*** (6.253)	-33.7*** (6.253)	-34.9*** (6.221)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.919*** (0.182)		3.946*** (0.176)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			13.00*** (0.201)	12.60*** (0.200)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	48,848	48,848	48,848	48,848
R^2	0.145	0.158	0.213	0.221
Adjusted R^2	0.143	0.155	0.210	0.218

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S10: Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Seattle.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-21.5** (10.58)	-22.2** (10.06)	-26.3*** (10.06)	-26.6*** (10.00)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		5.021*** (0.187)		3.993*** (0.179)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			13.72*** (0.205)	13.32*** (0.205)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	41,451	41,451	41,451	41,451
R^2	0.168	0.182	0.249	0.258
Adjusted R^2	0.166	0.180	0.247	0.256

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S11: Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Los Angeles.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-17.3*** (4.820)	-17.8*** (4.554)	-21.2*** (4.554)	-21.4*** (4.534)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.860*** (0.095)		3.531*** (0.091)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.30*** (0.100)	13.89*** (0.100)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	168,385	168,385	168,385	168,385
R^2	0.137	0.150	0.229	0.236
Adjusted R^2	0.136	0.149	0.228	0.235

Note:

*p<0.1, **p<0.05, ***p<0.01

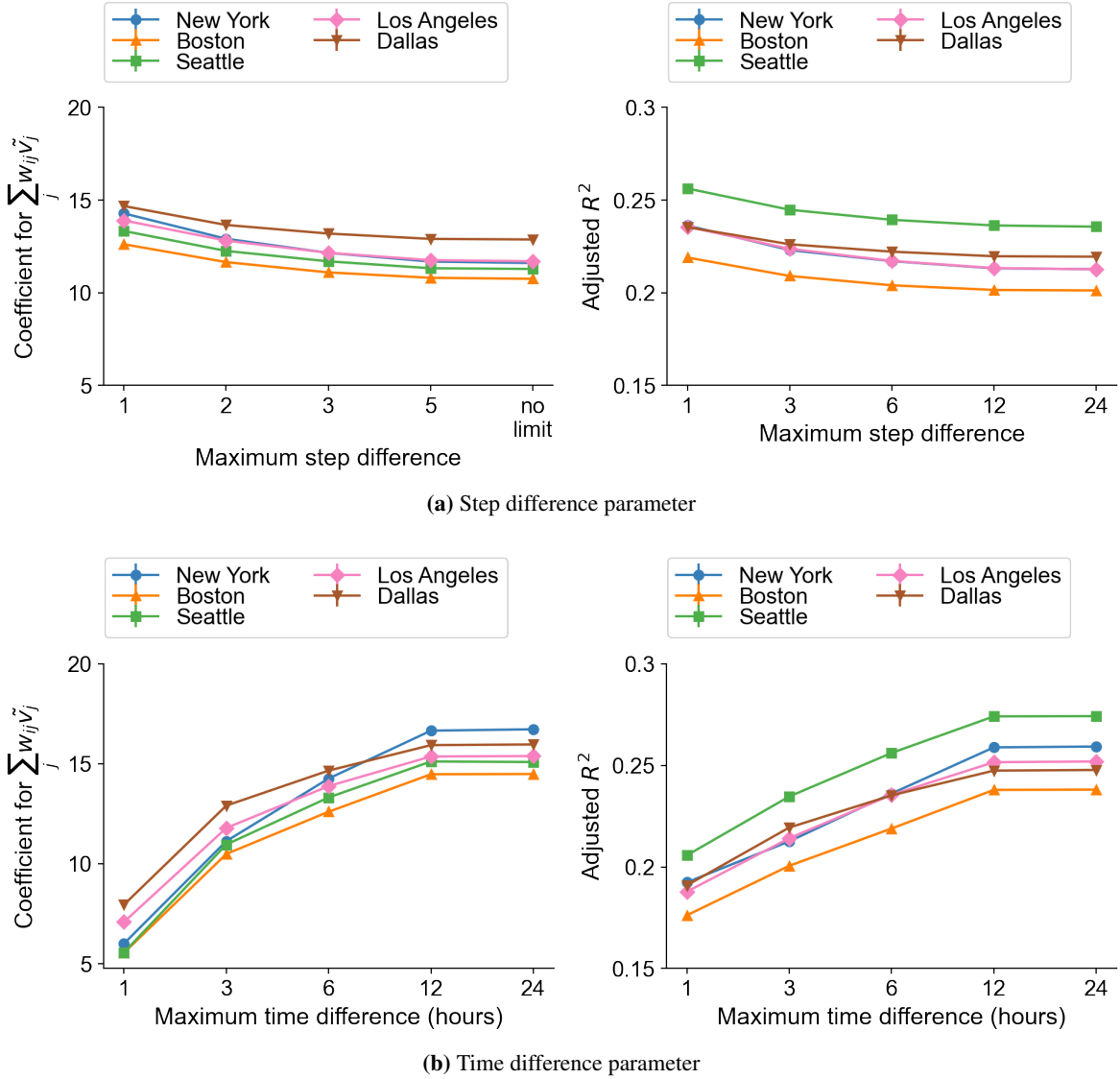


Figure S34: Regression results for dependency weights when using different co-visit detection parameters, continued for Los Angeles and Dallas. Except for the networks generated using maximum time difference of 1 hour, the results (adjusted R^2 and coefficient for Haversine distance) are relatively similar as the results in Table S3.

Table S12: Linear regression models predicting the change in visits to POIs during the pandemic (2019 March - May) in Dallas.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	3.944 (10.25)	1.362 (9.730)	-11.3 (9.730)	-12.6 (9.671)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		5.985*** (0.146)		4.427*** (0.140)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			15.28*** (0.160)	14.66*** (0.160)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	81,457	81,457	81,457	81,457
R^2	0.140	0.158	0.227	0.236
Adjusted R^2	0.139	0.156	0.225	0.235

Note:

*p<0.1; **p<0.05; ***p<0.01

5.4 Results using changes in visits during different time periods

To test whether the behavior-based dependency network effects significantly improves the prediction performance of visitation losses to POIs regardless of the time period, we test the same regression models with dependent variables calculated from different periods during the pandemic. Tables S13 to S17 show the regression models for predicting the change of visits during 2020/6/1 and 2020/8/31, compared to the pre-pandemic period (2019/9/1 to 2019/11/30). Thus, the dependent variable is computed as:

$$\tilde{v}_i = \left(\frac{v_i^{2020\text{June}\sim 2020\text{August}}}{v_i^{2019\text{September}\sim 2019\text{November}}} - 1 \right) * 100(\%) \quad (6)$$

In all metropolitan areas, the adjusted R^2 using the behavior-based dependency network effects (model 3) outperforms models 1 and 2 by 50% to 100%, showing substantial improvement in predictability. The regression coefficients for the behavior-based dependency network effects is significantly larger (30 fold) than the distance based effects, similar to the findings from the experiments in Tables S9 to S12.

Similarly, Tables S18 to S22 show the regression models for predicting the change of visits during 2020/9/1 and 2020/11/30, compared to the pre-pandemic period (2019/9/1 to 2019/11/30). In all metropolitan areas, the adjusted R^2 using the behavior-based dependency network effects (model 3) outperforms models 1 and 2 by 50% to 100%, showing substantial improvement in predictability. The regression coefficients for the behavior-based dependency network effects is significantly larger (30 fold) than the distance based effects, similar to the findings from the previous experiments. Overall the magnitude of the R^2 across all 4 models are smaller than those in the previous experiments. This suggests that the dependency network observed during one year prior to the dependent variable loses predictive power due to structural changes in the urban form. Nevertheless, experiments using change

in visits during different time periods during the pandemic showed that the behavior-based dependency network effects substantially improve predictability compared to using the distance-based network effects.

Table S13: Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in New York.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-20.6 (17.00)	-21.1 (15.66)	-15.8 (15.66)	-16.3 (15.62)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		26.41*** (0.275)		10.12*** (0.275)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			57.97*** (0.297)	53.96*** (0.316)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	213,140	213,140	213,140	213,140
R^2	0.072	0.111	0.213	0.218
Adjusted R^2	0.071	0.110	0.212	0.217

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S14: Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Boston.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-19.6 (18.67)	-24.9 (17.52)	-33.7* (17.52)	-35.2** (17.38)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		25.83*** (0.519)		14.70*** (0.521)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			44.40*** (0.546)	39.43*** (0.570)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	48,848	48,848	48,848	48,848
R^2	0.062	0.107	0.173	0.187
Adjusted R^2	0.059	0.104	0.171	0.185

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S15: Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Seattle.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	3.507 (16.98)	2.605 (16.36)	-1.16 (16.36)	-1.55 (16.27)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		8.777*** (0.303)		6.400*** (0.296)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			19.09*** (0.338)	17.98*** (0.340)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	41,451	41,451	41,451	41,451
R^2	0.106	0.124	0.170	0.180
Adjusted R^2	0.104	0.121	0.168	0.177

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S16: Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Los Angeles.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	8.252 (7.296)	7.685 (6.885)	3.501 (6.885)	3.390 (6.857)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		8.702*** (0.145)		5.212*** (0.140)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			21.46*** (0.149)	20.42*** (0.151)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	168,385	168,385	168,385	168,385
R^2	0.080	0.099	0.180	0.187
Adjusted R^2	0.079	0.098	0.179	0.186

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S17: Linear regression models predicting the change in visits to POIs during the pandemic (2019 June - August) in Dallas.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	38.63** (18.73)	33.88* (18.16)	12.24 (18.16)	10.28 (18.11)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		8.472*** (0.269)		5.673*** (0.265)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			20.41*** (0.283)	19.47*** (0.285)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	81,457	81,457	81,457	81,457
R^2	0.062	0.073	0.118	0.123
Adjusted R^2	0.060	0.071	0.116	0.121

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S18: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-11.7 (9.700)	-12.4 (9.413)	-11.4 (9.413)	-11.9 (9.382)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		9.426*** (0.163)		6.109*** (0.162)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			22.57*** (0.196)	21.12*** (0.199)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	213,140	213,140	213,140	213,140
R^2	0.081	0.095	0.135	0.140
Adjusted R^2	0.080	0.094	0.134	0.139

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S19: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-13.4 (15.04)	-15.8 (14.94)	-19.8 (14.94)	-21.2 (14.90)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		8.396*** (0.427)		6.962*** (0.429)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			11.95*** (0.469)	10.84*** (0.472)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	48,848	48,848	48,848	48,848
R^2	0.081	0.088	0.093	0.097
Adjusted R^2	0.078	0.085	0.090	0.095

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S20: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	8.718 (16.05)	7.900 (15.56)	4.898 (15.56)	4.468 (15.48)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		7.389*** (0.286)		5.628*** (0.281)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			16.64*** (0.323)	15.80*** (0.324)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	41,451	41,451	41,451	41,451
R^2	0.071	0.086	0.127	0.136
Adjusted R^2	0.069	0.083	0.125	0.133

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S21: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	9.110 (7.274)	8.471 (7.055)	4.076 (7.055)	3.920 (7.024)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		7.655*** (0.144)		5.468*** (0.142)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			15.62*** (0.151)	14.69*** (0.152)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	168,385	168,385	168,385	168,385
R^2	0.053	0.069	0.109	0.117
Adjusted R^2	0.052	0.068	0.108	0.116

Note:

*p<0.1, **p<0.05, ***p<0.01

Table S22: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	17.86 (19.67)	14.28 (19.18)	-4.65 (19.18)	-6.44 (19.14)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		7.407*** (0.282)		5.340*** (0.278)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			18.95*** (0.292)	18.28*** (0.293)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	81,457	81,457	81,457	81,457
R^2	0.037	0.045	0.084	0.089
Adjusted R^2	0.035	0.043	0.083	0.087

Note:

*p<0.1, **p<0.05, ***p<0.01

5.5 Robustness of regression results when using different time periods to generate dependency networks

The robustness checks conducted in Supplementary Notes 3.1 and 4.4 showed high correlation in edge weights between different time periods. Nevertheless, in this section, we tested whether the time periods used to generate the behavior-based dependency networks affects the performance of the regression models. Tables S23 to S27 show the regression models for predicting the change of visits during 2020/3/1 and 2020/5/31, using the dependency network computed from data between 2019/1/1 and 2019/4/30.

In all metropolitan areas, the adjusted R^2 using the behavior-based dependency network effects (model 3) outperforms models 1 and 2 by around 50%, showing substantial improvement in predictability. The regression coefficients for the behavior-based dependency network effects is significantly larger (around 5 fold) than the distance based effects, similar to the findings from the previous experiments.

Similarly, Tables S28 to S32 show the regression models for predicting the change of visits during 2020/3/1 and 2020/5/31, using the dependency network computed from data between 2019/5/1 and 2019/8/31. Again, in all metropolitan areas, the adjusted R^2 using the behavior-based dependency network effects (model 3) outperforms models 1 and 2 by around 50%, showing the robustness against the choice of time period to generated the dependency networks. In summary, the results on the predictability of shocks using the behavior-based dependency networks are robust against the choice of the time periods used to generate the dependency networks.

Table S23: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from data between 2019 January to April.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-38.3*** (5.251)	-38.9*** (5.033)	-40.3*** (5.033)	-40.8*** (5.009)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.089*** (0.084)		3.614*** (0.081)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			13.63*** (0.100)	13.44*** (0.100)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	206,108	206,108	206,108	206,108
R^2	0.166	0.175	0.234	0.241
Adjusted R^2	0.165	0.174	0.233	0.240

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S24: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from data between 2019 January to April.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-22.0*** (6.518)	-24.0*** (6.277)	-33.1*** (6.277)	-34.6*** (6.242)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.779*** (0.185)		4.102*** (0.178)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			12.30*** (0.202)	12.00*** (0.201)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	46,643	46,643	46,643	46,643
R^2	0.147	0.159	0.210	0.219
Adjusted R^2	0.145	0.157	0.208	0.217

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S25: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from data between 2019 January to April.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-27.2** (11.03)	-27.5** (10.52)	-30.1*** (10.52)	-30.3*** (10.46)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.856*** (0.191)		3.937*** (0.183)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			13.12*** (0.211)	12.75*** (0.210)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	39,582	39,582	39,582	39,582
R^2	0.173	0.186	0.247	0.256
Adjusted R^2	0.171	0.184	0.245	0.253

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S26: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from data between 2019 January to April.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-18.6*** (4.828)	-19.3*** (4.563)	-20.9*** (4.563)	-21.3*** (4.542)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.651*** (0.097)		3.561*** (0.092)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.26*** (0.102)	13.91*** (0.102)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	162,861	162,861	162,861	162,861
R^2	0.138	0.150	0.230	0.237
Adjusted R^2	0.137	0.149	0.229	0.236

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S27: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from data between 2019 January to April.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	7.027 (10.15)	4.521 (9.647)	-8.57 (9.647)	-9.93 (9.585)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		5.827*** (0.147)		4.493*** (0.141)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.91*** (0.162)	14.37*** (0.161)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	78,222	78,222	78,222	78,222
R^2	0.145	0.161	0.228	0.238
Adjusted R^2	0.143	0.160	0.227	0.237

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S28: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from data between 2019 May to August.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-38.2*** (5.249)	-38.7*** (5.016)	-40.5*** (5.016)	-40.9*** (5.002)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		3.398*** (0.085)		2.733*** (0.081)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.70*** (0.104)	14.50*** (0.103)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	210,540	210,540	210,540	210,540
R^2	0.164	0.170	0.236	0.240
Adjusted R^2	0.163	0.169	0.235	0.240

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S29: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from data between 2019 May to August.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-19.8*** (6.580)	-21.4*** (6.324)	-33.8*** (6.324)	-34.9*** (6.302)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		3.816*** (0.185)		3.244*** (0.178)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			13.20*** (0.208)	13.00*** (0.207)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	47,995	47,995	47,995	47,995
R^2	0.146	0.153	0.212	0.217
Adjusted R^2	0.144	0.151	0.210	0.215

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S30: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from data between 2019 May to August.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-21.9** (10.61)	-22.5** (10.07)	-26.3*** (10.07)	-26.6*** (10.03)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.667*** (0.190)		3.622*** (0.181)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.01*** (0.210)	13.64*** (0.210)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	40,800	40,800	40,800	40,800
R^2	0.172	0.184	0.253	0.260
Adjusted R^2	0.169	0.181	0.251	0.258

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S31: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from data between 2019 May to August.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	-17.6*** (4.839)	-17.8*** (4.561)	-20.5*** (4.561)	-20.6*** (4.543)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.597*** (0.096)		3.372*** (0.091)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.72*** (0.102)	14.37*** (0.102)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	166,292	166,292	166,292	166,292
R^2	0.136	0.148	0.233	0.239
Adjusted R^2	0.135	0.147	0.232	0.238

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S32: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from data between 2019 May to August.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	8.334 (10.10)	5.849 (9.581)	-8.60 (9.581)	-9.90 (9.530)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		5.621*** (0.149)		4.216*** (0.143)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			15.68*** (0.164)	15.16*** (0.165)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	80,431	80,431	80,431	80,431
R^2	0.141	0.156	0.228	0.236
Adjusted R^2	0.140	0.154	0.227	0.235

Note:

*p<0.1, **p<0.05, ***p<0.01

5.6 Robustness of regression results when using only short non-work stays

In this section, we test whether the behavior-based dependency network generated using only short (i.e., non-work) stays yields similar performance metrics in the regression models. Tables S33 to S37 show the regression models for predicting the change of visits during 2020/3/1 and 2020/5/31, using the dependency network computed with stays shorter than 4 hours. In all metropolitan areas, the adjusted R^2 using the behavior-based dependency network effects (model 3) outperforms models 1 and 2, showing improvement in predictability. The regression coefficients for the behavior-based dependency network effects are significantly larger (around 5 fold) than the distance based effects, similar to the findings from the previous experiments. This results suggest that our estimation results are robust against the limiting the analysis to the use of short non-work trips.

5.7 Case study on school closure period

To further test the effectiveness of using the dependency network for predicting the cascade of shocks, we used school holidays as a separate case study. The assumption is that, places that are connected to schools (in our case, college places) tend to have an additional decrease in visits than other places that have less dependency on college POIs. To compute the change in visits, we used the school semester period (September to November 2019) as the post period and the summer break period (June to August 2019) as the pre- period.

$$\tilde{v}_i = \left(\frac{v_i^{2019\text{September}\sim 2019\text{November}}}{v_i^{2019\text{June}\sim 2019\text{August}}} - 1 \right) * 100(\%) \quad (7)$$

Table S33: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in New York when using dependency network generated from stays shorter than 4 hours.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-41.1*** (5.230)	-41.8*** (4.943)	-42.7*** (4.943)	-43.2*** (4.928)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.217*** (0.084)		2.896*** (0.080)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			16.66*** (0.105)	16.26*** (0.106)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	208,329	208,329	208,329	208,329
R^2	0.165	0.174	0.254	0.258
Adjusted R^2	0.164	0.173	0.253	0.257

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S34: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Boston when using dependency network generated from stays shorter than 4 hours.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-20.6*** (6.562)	-22.6*** (6.261)	-36.6*** (6.261)	-37.5*** (6.236)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.844*** (0.183)		3.456*** (0.176)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.16*** (0.205)	13.68*** (0.206)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	47,466	47,466	47,466	47,466
R^2	0.148	0.160	0.225	0.231
Adjusted R^2	0.145	0.158	0.223	0.229

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S35: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Seattle when using dependency network generated from stays shorter than 4 hours.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-21.7** (10.50)	-22.4** (9.905)	-28.8*** (9.905)	-29.0*** (9.860)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.933*** (0.188)		3.439*** (0.179)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			14.94*** (0.211)	14.44*** (0.212)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	40,135	40,135	40,135	40,135
R^2	0.171	0.185	0.263	0.270
Adjusted R^2	0.169	0.183	0.261	0.268

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S36: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Los Angeles when using dependency network generated from stays shorter than 4 hours.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>				
	(1)	(2)	(3)	(4)
Constant	-18.5*** (4.834)	-19.0*** (4.544)	-22.9*** (4.544)	-23.1*** (4.527)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		4.905*** (0.096)		3.241*** (0.091)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			15.07*** (0.102)	14.61*** (0.103)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	164,195	164,195	164,195	164,195
R^2	0.139	0.152	0.239	0.245
Adjusted R^2	0.138	0.151	0.238	0.244

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S37: Linear regression models predicting the change in visits to POIs during the pandemic (2019 September - November) in Dallas when using dependency network generated from stays shorter than 4 hours.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during the pandemic)</i>			
	(1)	(2)	(3)	(4)
Constant	3.572 (10.11)	0.977 (9.521)	-13.2 (9.521)	-14.2 (9.476)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		5.910*** (0.146)		3.843*** (0.140)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			16.34*** (0.161)	15.65*** (0.163)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	78,929	78,929	78,929	78,929
R^2	0.144	0.162	0.242	0.250
Adjusted R^2	0.143	0.160	0.241	0.248

Note:

*p<0.1, **p<0.05, ***p<0.01

Figure S35 shows the changes in visits to POIs in the five cities during the fall semester period compared to the summer break period. As the right panels clearly show, College POIs have a significant (25% to 60%) increase in the number of visits during the semester compared to the breaks. The same model specification was used to test the effectiveness of using the dependency network. Tables S38 to S42 show the regression results. Although the predictability of the change in visits is generally lower than the pandemic results (due to the smaller impacts of summer holidays compared to the pandemic), in all five cities, the model performance is highest when we use the behavior-based dependency network compared to when we use the distance-based null network.

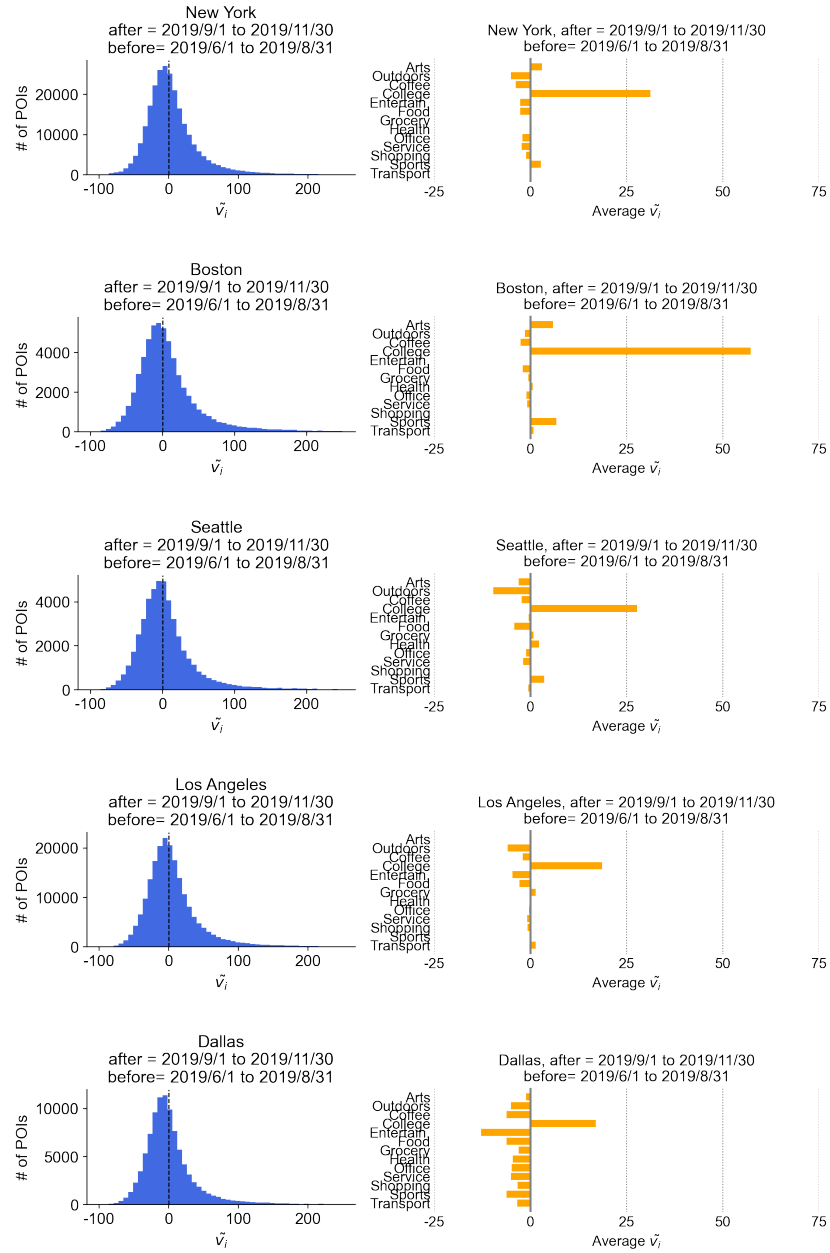


Figure S35: Change in visits to POIs in New York, Boston, Seattle, Los Angeles, and Dallas during the fall semester period. The left panels show the distribution of the change in visits, and the right panels show the average effects per POI category.

Table S38: Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in New York.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during school semester)</i>				
	(1)	(2)	(3)	(4)
Constant	-5.92 (14.83)	-2.48 (14.42)	-4.70 (14.42)	-3.06 (14.39)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		15.11*** (0.240)		7.714*** (0.247)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			26.13*** (0.235)	23.74*** (0.247)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	212,008	212,008	212,008	212,008
R^2	0.015	0.033	0.069	0.073
Adjusted R^2	0.014	0.032	0.068	0.072

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S39: Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Boston.

<i>Dependent variable: $\tilde{v}_i$ (Change in visits during school semester)</i>				
	(1)	(2)	(3)	(4)
Constant	12.09 (16.29)	17.42 (15.45)	18.28 (15.45)	20.11 (15.39)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		20.47*** (0.456)		9.685*** (0.475)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			33.63*** (0.458)	29.87*** (0.492)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	48,652	48,652	48,652	48,652
R^2	0.049	0.087	0.144	0.151
Adjusted R^2	0.046	0.084	0.141	0.148

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S40: Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Seattle.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during school semester)</i>			
	(1)	(2)	(3)	(4)
Constant	-7.76 (56.04)	-8.32 (55.94)	-8.75 (55.94)	-9.05 (55.92)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		6.630*** (0.991)		4.304*** (1.011)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			12.02*** (0.974)	11.15*** (0.995)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	41,294	41,294	41,294	41,294
R^2	0.004	0.005	0.008	0.008
Adjusted R^2	0.001	0.002	0.005	0.005

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S41: Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Los Angeles.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during school semester)</i>			
	(1)	(2)	(3)	(4)
Constant	4.979 (12.90)	5.578 (12.63)	6.292 (12.63)	6.486 (12.61)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		11.14*** (0.253)		5.416*** (0.260)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			21.40*** (0.251)	19.81*** (0.262)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	166,784	166,784	166,784	166,784
R^2	0.009	0.020	0.050	0.052
Adjusted R^2	0.008	0.019	0.049	0.051

Note:

*p<0.1; **p<0.05; ***p<0.01

Table S42: Linear regression models predicting the change in visits to POIs during the school semester (2019 September - November) compared to summer break (2019 June - August) in Dallas.

	<i>Dependent variable: $\tilde{v}_i$ (Change in visits during school semester)</i>			
	(1)	(2)	(3)	(4)
Constant	15.96 (45.82)	32.60 (41.70)	25.89 (41.70)	33.77 (40.70)
Distance effect $\sum_j \delta_{ij} \tilde{v}_j$		65.77*** (0.598)		40.32*** (0.634)
Dependency effect $\sum_j w_{ij} \tilde{v}_j$			75.00*** (0.579)	57.42*** (0.629)
Subcategory FE	Y	Y	Y	Y
PUMA FE	Y	Y	Y	Y
Observations	81,103	81,103	81,103	81,103
R^2	0.006	0.135	0.177	0.216
Adjusted R^2	0.004	0.134	0.175	0.214

Note:

*p<0.1; **p<0.05; ***p<0.01

6 Cascading impacts of hypothetical shocks

6.1 Leontief open model

To explore what the behavior-based dependency network can tell us about other types of future shocks, we apply the network effects regression model to simulate the spatial cascades of such shocks in different cities. More specifically, re-writing and reorganizing the regression model in matrix form, we obtain

$$\vec{v} = W\vec{v} + \vec{f} \quad (8)$$

where $\vec{v}$ is a vector of $\tilde{v}_i$ for all N places, W is an $N \times N$ matrix where each element is $\tilde{w}_{ij} = \hat{\beta}_W w_{ij}$, and vector $\vec{f}$ is an aggregation of all fixed effects β_0 , η_i , and θ_i . This model specification is known as the Leontief Open Model, which is a simplified and linear economic model for an economy in which input equals output [8]. To predict the propagation of shocks throughout places in the city, the shocks are modeled in the fixed effect vector f (e.g., all colleges experience an external shock of -50% visits reduction due to uptake of online education), and the production vector $\vec{v}$ is computed by solving the linear system $\hat{\vec{v}} = (I - W)^{-1} \vec{f}$ via the generalized minimal residual iteration method.

The shift to online education which occurred during the pandemic is reported to have a continuing effect, with roughly 20 percent of school systems planning to or have already started online school programs [12]. Previous studies [1] as well as analysis in Figure 2 in the main manuscript have pointed out that college campuses have a substantial impact on the local economy. If online learning and remote education were permanent and increased with the help of advanced technology (e.g., augmented reality), what impacts would it have on other businesses and amenities?

Figure S36a shows the simulated effects of a 50% reduction in visits to college POIs (gray points) on nearby non-college POIs (red points; darker red indicates larger negative impacts). Impacted POIs are not limited to those in proximity to college POIs, but also in locations that are popular with college students, for example, Massachusetts Avenue which connects MIT and Harvard University. For comparison, we simulated the shocks to non-college POIs using the physical distance network $\hat{W}$, where $\hat{w}_{ij}$ is used as the matrix elements instead of behavior-based dependency w_{ij} in Figure S36c. Comparing the simulation results using the dependency network and the null network shows that neglecting the behavior-based dependencies results in a substantial underestimation of the effects on POIs that are located further away from colleges. Model parameters, $\beta_W = 0.70$ and $\beta_{null} = 0.15$ were derived from prior regressions on estimation of changes in visits during pre-pandemic periods.

The effects of online education were heterogeneous for different place categories located at different distances from colleges. Figure S36b shows the 90th percentile of impacts on POIs by category and distance (log scaled). While most significant impacts occur within 0.5 km, places such as arts and museums, food, and service places experience substantial long-distance impacts. Furthermore, simulations assuming different levels of visit decrease to colleges (e.g., -100% , -25%) show a similar long-distance cascade of shocks (Figures S37 and S38). These persistent spatial cascades emphasize the importance of considering behavior-based dependency relationships between places to grasp the holistic impact of such urban shocks for resilient urban planning.

6.2 Cascading effects of individual place closures

Further leveraging the network model, we are able to simulate the impacts of POI closure scenarios and identify the seed nodes (POIs) that have the largest cascading effects on other POIs if inflicted by other urban shocks. For each node, we simulate the cascading impacts of a 100% visit change to a

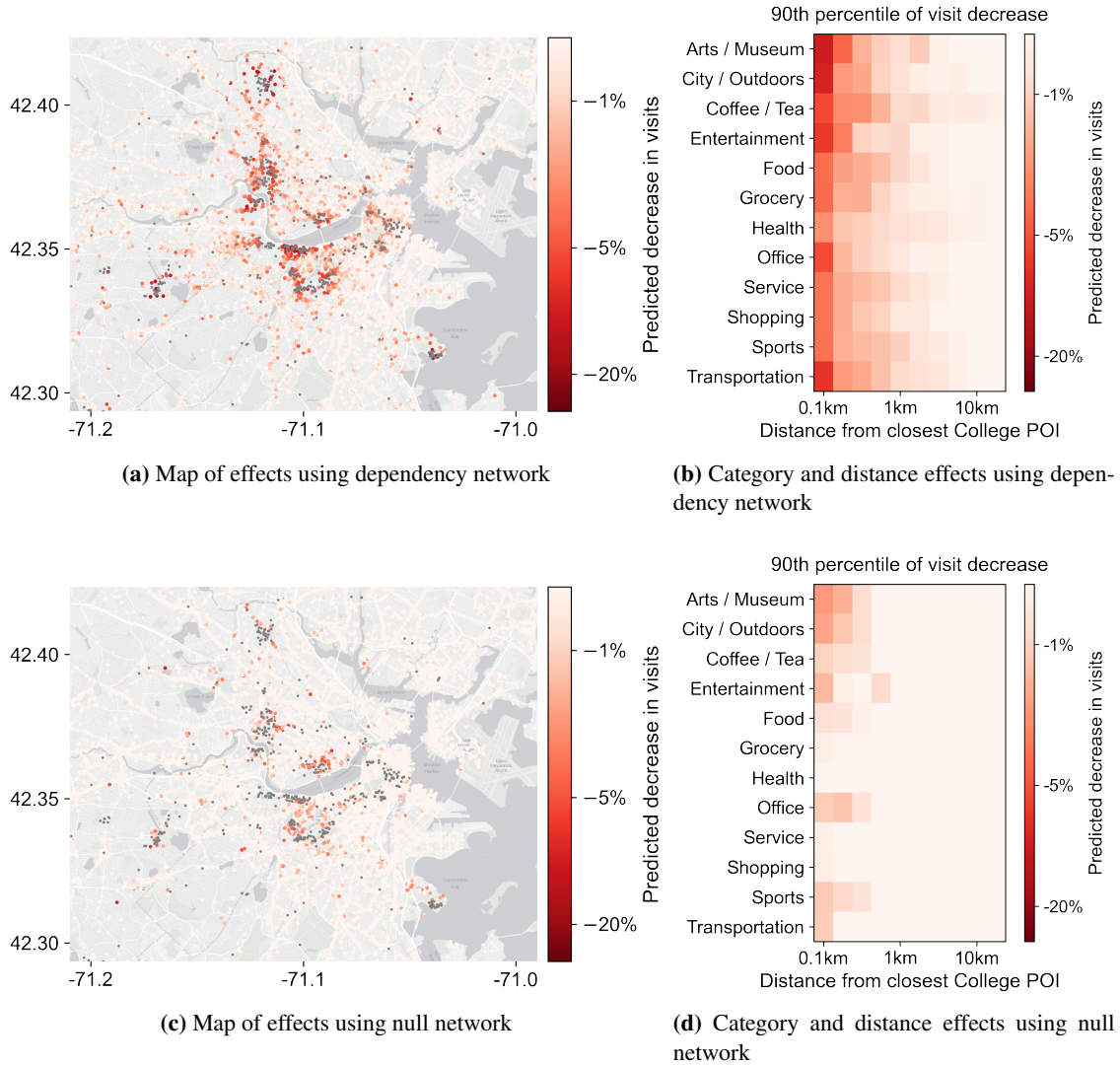


Figure S36: Cascading impacts of a 50% visit reduction to colleges. a. Simulated effects of a 50% reduction in visits to college POIs (gray points) on nearby non-college POIs (red points; darker red indicates larger negative impacts), using the fitted Leontief Open Model. Impacted POIs are not limited to those in proximity to college POIs. b. Impacts of the 50% visit reduction to colleges on places by category and distance (90th percentile decrease in visits are shown, log scaled). c. and d. show results using the null networks.

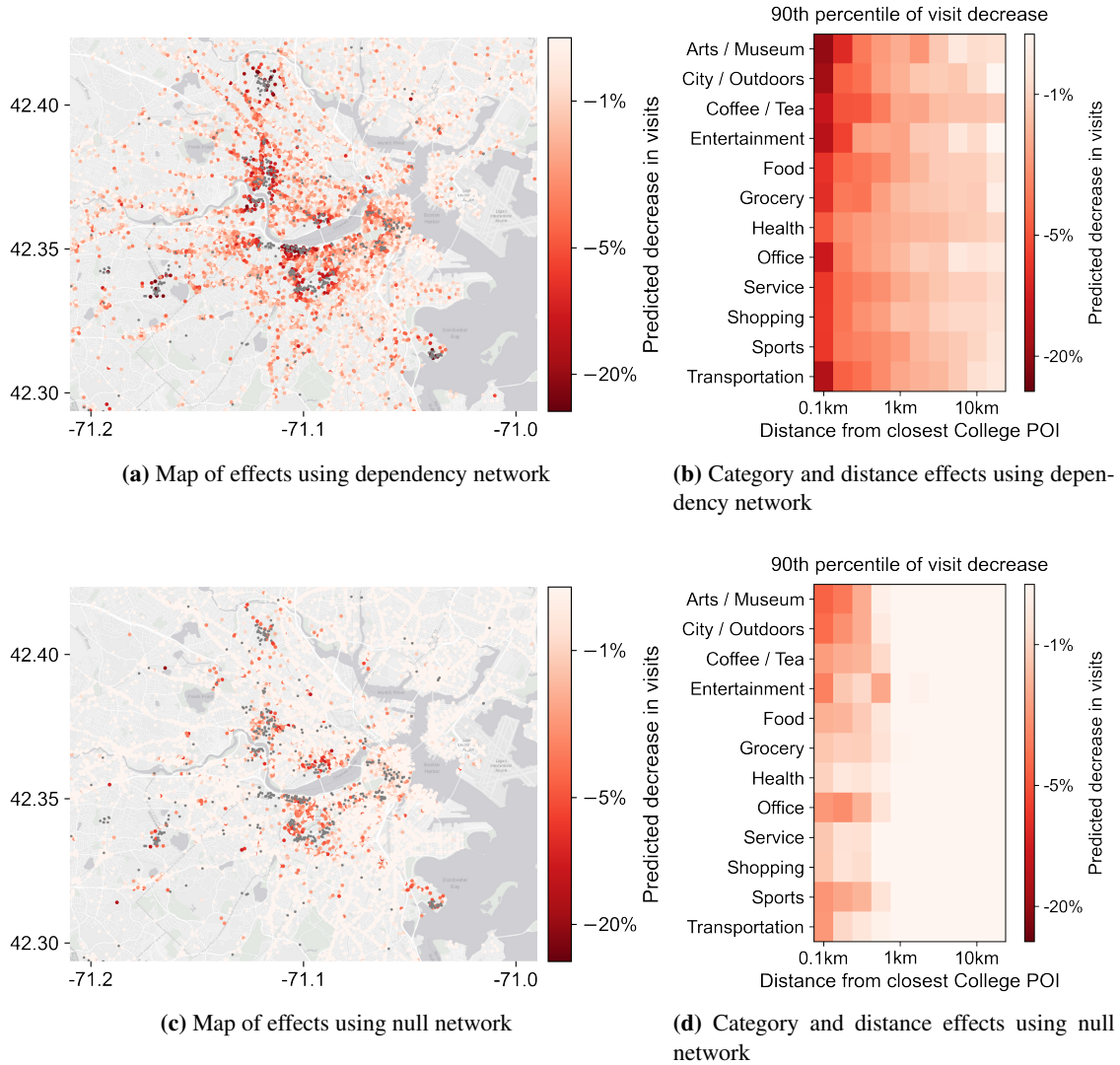


Figure S37: Cascading impacts of a 100% visit reduction to colleges. a. Simulated effects of a 50% reduction in visits to college POIs (gray points) on nearby non-college POIs (red points; darker red indicates larger negative impacts), using the fitted Leontief Open Model. Impacted POIs are not limited to those in proximity to college POIs. b. Impacts of the 100% visit reduction to colleges on places by category and distance (90th percentile decrease in visits are shown, log scaled). c. and d. show results using the null networks.

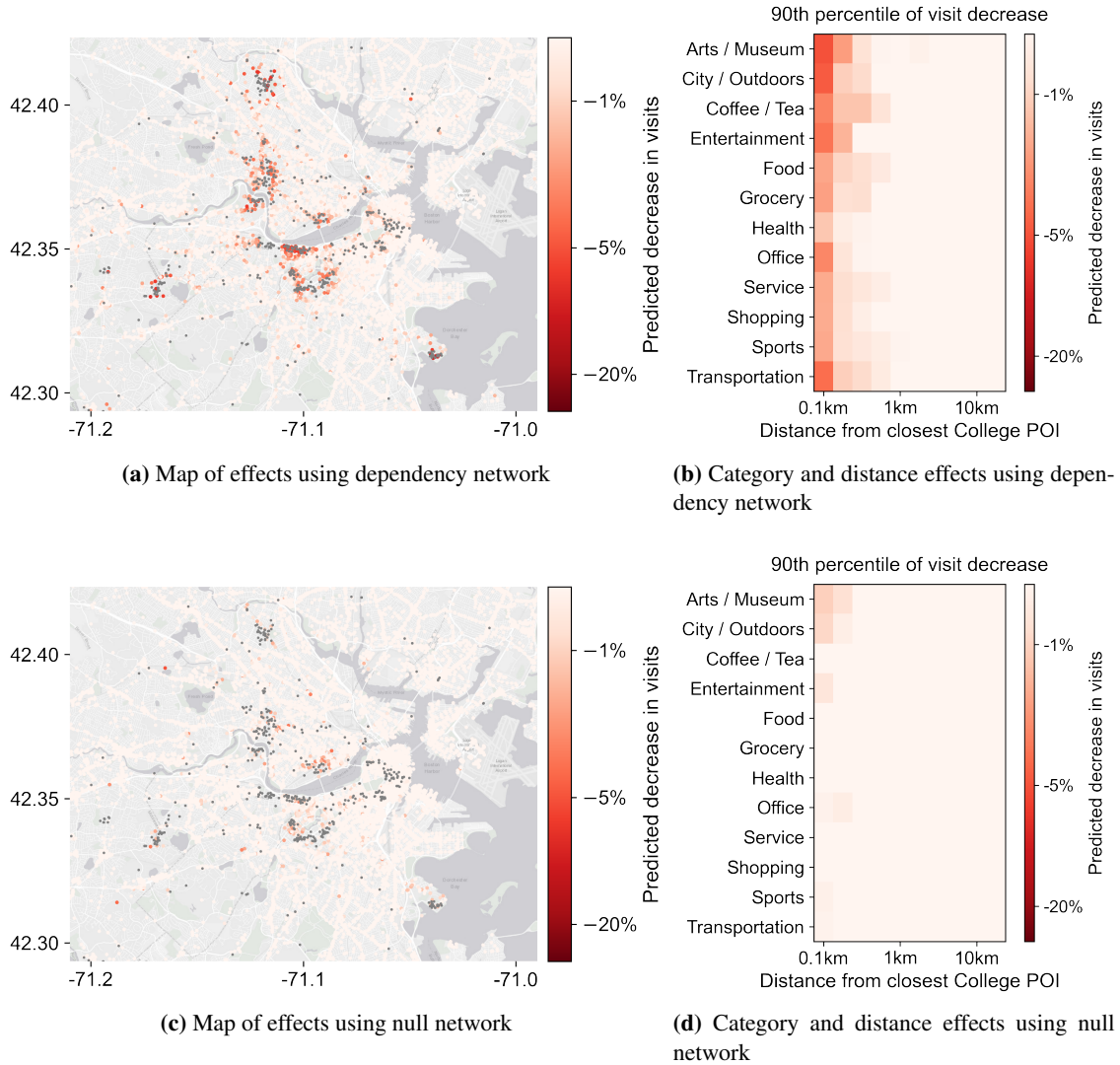


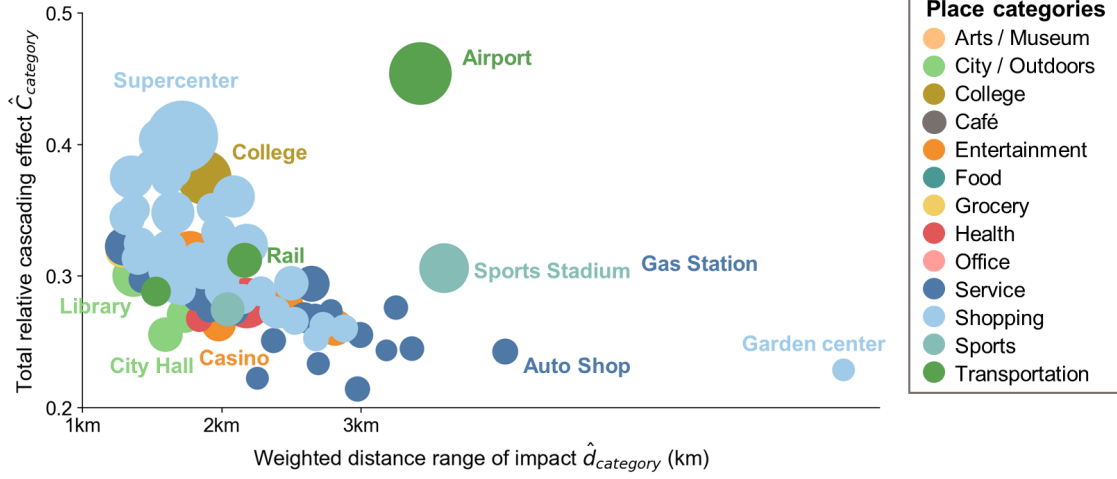
Figure S38: Cascading impacts of a 25% visit reduction to colleges. a. Simulated effects of a 25% reduction in visits to college POIs (gray points) on nearby non-college POIs (red points; darker red indicates larger negative impacts), using the fitted Leontief Open Model. Impacted POIs are not limited to those in proximity to college POIs. b. Impacts of the 50% visit reduction to colleges on places by category and distance (90th percentile decrease in visits are shown, log scaled). c. and d. show results using the null networks.

single node i , by computing $\hat{\vec{v}}^{(i)} = (I - W)^{-1} \vec{e}^{(i)}$, where $\vec{e}^{(i)}$ is a one-hot encoding vector of the initial shock that assigns a change in visits of +1 to node i and 0 otherwise, and $\hat{\vec{v}}^{(i)}$ is the resulting vector of the cascading impacts, where each element measures the impacts of the initial shock to all nodes. The total impacts of changes in the number of visits to all nodes can be computed by multiplying $\hat{\vec{v}}^{(i)} = (\hat{v}_1^{(i)}, \dots, \hat{v}_N^{(i)})$ with the vector of total visits to each POI, $\vec{n} = (n_1, \dots, n_N)$. Thus, the total impacts of the initial shock to node i can be computed by $C_i = \sum_{j; j \neq i} \hat{v}_j^{(i)} n_j$. By further scaling the impact to its own size n_i , we obtain the total relative cascading effect as $\hat{C}_i = C_i / n_i$. $\hat{C}_i = 0.3$ indicates that increasing the number of visits to node i by 100% ($= n_i$) results in a total of 30%* n_i increase in visits across all other nodes. The mean relative cascading impacts of each POI category, $\hat{C}_{category}$ are shown in the y-axis of Figures S39 to S41. POI categories such as airports, supercenters, colleges, furniture stores, theme parks, railway stations, and sports stadiums have a high impact on other POIs in urban areas propagated through behavior-based dependency networks.

When implementing policies to close down certain POIs for emergency response (e.g., lockdowns during pandemics), it is important to understand the spatial extent of the cascade. To quantify this, we defined the distance range of the cascade by computing the average distance to impacted nodes, weighted by the magnitude of the impacts. More specifically, we compute the weighted distance range of POI i by $\hat{d}_i = \sum_{j; j \neq i} \hat{v}_j^{(i)} d_{ij} / \sum_{j; j \neq i} \hat{v}_j^{(i)}$.

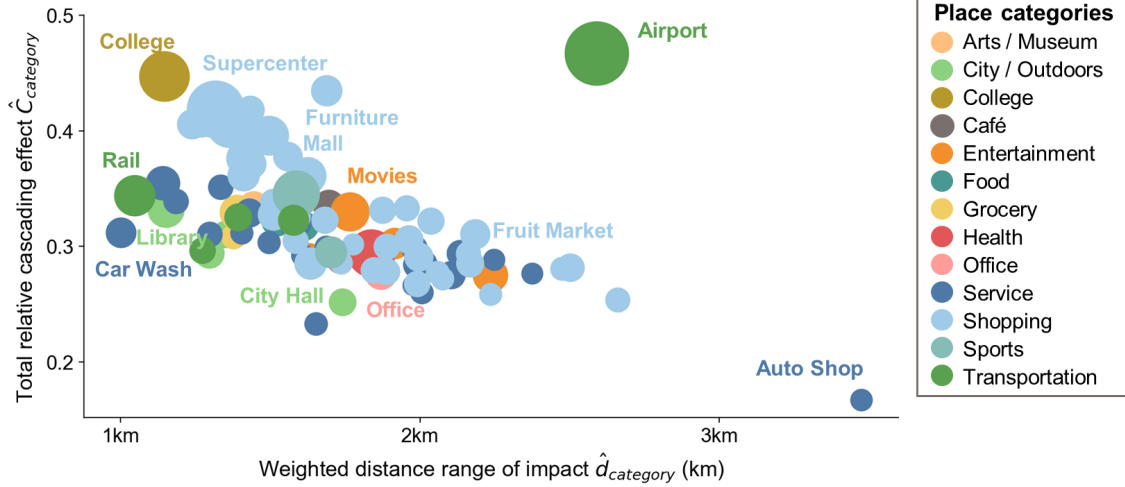
The weighted distance range of impact for each POI category, $\hat{d}_{category}$ are shown in the x-axis of Figures S39 to S41. Supercenters and colleges have high cascading effects but are focused locally ($\sim 1.5\text{km}$ around the POI). On the other hand, the impacts of airports, stadiums, theme parks, and gas stations are both large and far-reaching ($\sim 2.5\text{km}$ to 3.5km). Understanding the magnitude and spatial extent of the cascading effects could be applied to design emergency management policies to effectively close places while minimizing economic losses. The large magnitude of the spatial cascades that occur due to behavior-based dependency networks calls for new urban policy-making approaches that balance the benefits of mobility restriction measures (e.g., preventing the spread of diseases) while minimizing the total cascading economic impacts to urban places and amenities.

New York



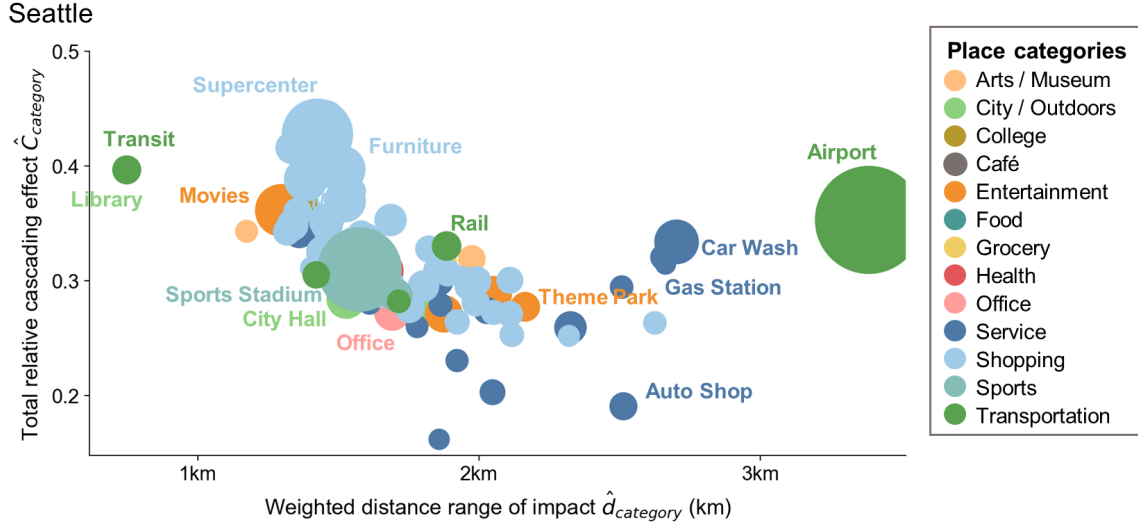
(a) New York

Boston

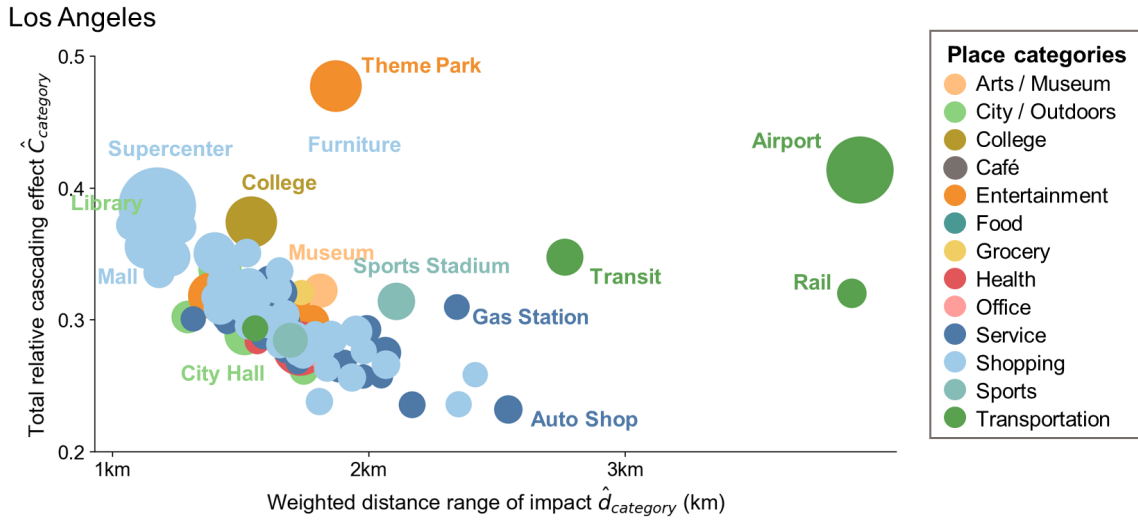


(b) Boston

Figure S39: Total cascading impact of closing places on other locations. Total cascading impact of closing places on other locations, relative to its own size (x-axis) and the weighted distance range of the impact (y-axis) for different POI subcategories in New York and Boston.

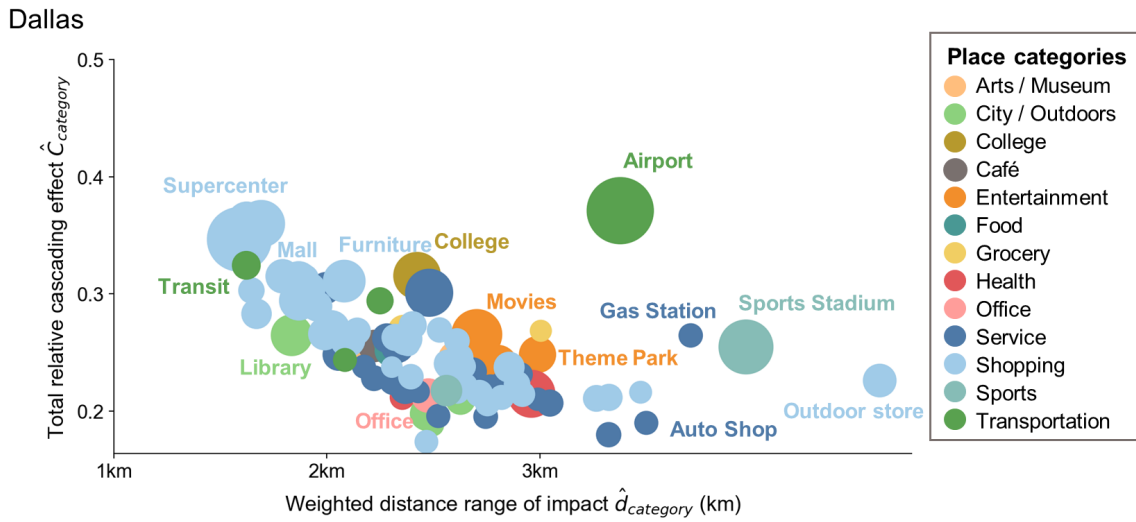


(a) Seattle



(b) Los Angeles

Figure S40: Total cascading impact of closing places on other locations. Total cascading impact of closing places on other locations, relative to its own size (x-axis) and the weighted distance range of the impact (y-axis) for different POI subcategories in Seattle and Los Angeles.



(a) Dallas

Figure S41: Total cascading impact of closing places on other locations. Total cascading impact of closing places on other locations, relative to its own size (x-axis) and the weighted distance range of the impact (y-axis) for different POI subcategories in Dallas.

7 Software

Analysis was conducted using Python, Jupyter Lab, and the following libraries and software:

- NumPy [4] for general computation on Python.
- Pandas [9] for loading, transforming, and analyzing data tables.
- Matplotlib [5] for creating plots and figures.
- GeoPandas [7] for spatial analysis and plotting map figures.
- networkx [3] for network analysis.
- Statsmodels [13] for statistical modeling and econometric analysis.
- A Python implementation of the R `Stargazer` multiple regression model creation tool⁵ was used to create the regression tables.

⁵<https://github.com/mwburke/stargazer>

References

- [1] Ernest R Bonner. The economic impact of a university on its local community. *Journal of the American Institute of Planners*, 34(5):339–343, 1968.
- [2] United States Census Bureau. American community survey. <https://www.census.gov/programs-surveys/acs>. Accessed: 2019-12-14.
- [3] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.
- [4] Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.
- [5] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(03):90–95, 2007.
- [6] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.
- [7] K Jordahl. Geopandas: Python tools for geographic data. URL: <https://github.com/geopandas/geopandas>, 3, 2014.
- [8] Wassily Leontief. *Input-output economics*. Oxford University Press, 1986.
- [9] Wes McKinney et al. pandas: a foundational python library for data analysis and statistics. *Python for high performance and scientific computing*, 14(9):1–9, 2011.
- [10] Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large us cities. *Nature Communications*, 12(1):1–10, 2021.
- [11] Matthew J Salganik. *Bit by bit: Social research in the digital age*. Princeton University Press, 2019.
- [12] Heather L. Schwartz, David Grant, Melissa Kay Diliberti, Gerald P. Hunter, and Claude Messan Setodji. *Remote Learning Is Here to Stay: Results from the First American School District Panel Survey*. RAND Corporation, Santa Monica, CA, 2020.
- [13] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61. Austin, TX, 2010.
- [14] Eric Tsetsi and Stephen A Rains. Smartphone internet access and use: Extending the digital divide and usage gap. *Mobile Media & Communication*, 5(3):239–255, 2017.
- [15] United States Census Bureau. Tiger data products guide. <https://www.census.gov/programs-surveys/geography/guidance/tiger-data-products-guide.html>. Accessed: 2023-02-27.

- [16] Qi Wang, Nolan Edward Phillips, Mario L Small, and Robert J Sampson. Urban mobility and neighborhood isolation in america's 50 largest cities. *Proceedings of the National Academy of Sciences*, 115(30):7735–7740, 2018.
- [17] Longgang Xiang, Meng Gao, and Tao Wu. Extracting stops from noisy trajectories: A sequence oriented clustering approach. *ISPRS International Journal of Geo-Information*, 5(3):29, 2016.
- [18] Takahiro Yabe, Bernardo García Bulle Bueno, Xiaowen Dong, Alex Pentland, and Esteban Moro. Behavioral changes during the covid-19 pandemic decreased income diversity of urban encounters. *Nature Communications*, 14(1):2310, 2023.