

# Supplementary Information to Urban highways are barriers to social ties

Luca Maria Aiello<sup>a,b,\*</sup>, Anastassia Vybornova<sup>a</sup>, Sándor Juhász<sup>c,d,e</sup>, Michael Szell<sup>a,b,c,f</sup>, and Eszter Bokányi<sup>g</sup>

<sup>a</sup>IT University of Copenhagen, Copenhagen, 2300, Denmark; <sup>b</sup>Pioneer Centre for AI, Copenhagen, 1350, Denmark; <sup>c</sup>Complexity Science Hub, Vienna, 1080, Austria;

<sup>d</sup>Corvinus University of Budapest, Budapest, 1093, Hungary; <sup>e</sup>HUN-REN CERS, Budapest, 1097, Hungary; <sup>f</sup>ISI Foundation, Turin, 10126, Italy; <sup>g</sup>University of Amsterdam, Amsterdam, 1018WV, The Netherlands

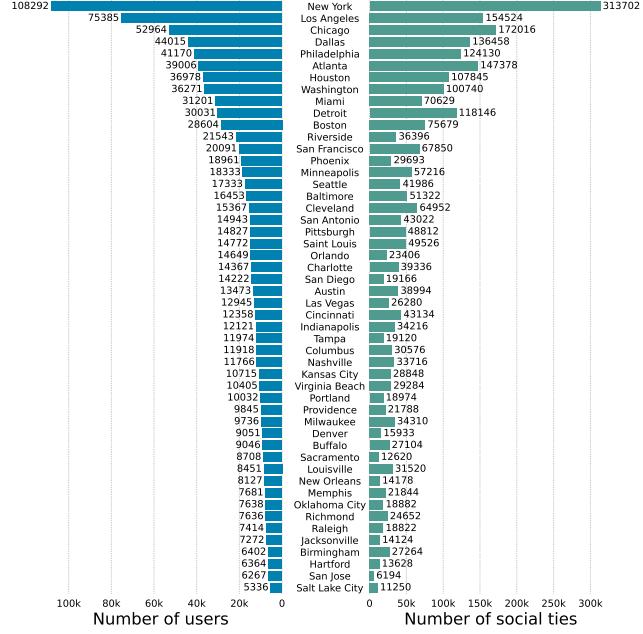
**A. Social network data.** Social connections within cities are mapped through a large social network snapshot from Twitter with precise geolocation information. This dataset provides a remarkable context for studying spatial network patterns inside cities, as individual-level social connections are rarely available with precise geographic location at large scales.

Mutual follower relations on Twitter may not always represent actual friendship connections in real life. However, multiple studies have used Twitter as a proxy for social interactions to study urban segregation (1–4). Moreover, there is recent evidence that the structure of online social networks follows closely that of register-based social networks, which represent formal connections of people such as family, work, and school relationships (5), especially at local spatial scales, such as a metropolitan area in our paper.

To enhance the quality of our Twitter data as much as possible, we adopt a two-fold approach to exclude bots, commercial, and political Twitter accounts from the sample. First, we use the geographic information and timestamp of tweets to exclude users whose consecutive tweets indicate unrealistic mobility behavior. Specifically, we calculate the speed of users between subsequent tweets based on the two locations and the time difference, and exclude users who changed their position faster than 800 km/h at least once, had at least 10 consecutive tweets more than 2 km apart, and at least 5% of these tweets had a speed greater than 120 km/h. These detailed criteria were necessary to avoid the exclusion of high-speed users for technical reasons, such as location hopping between cell phone towers or users occasionally flying on airplanes with wifi access. The same sample construction method was used previously in (6) and (7). Second, we rely on the number of users' outgoing (following) and incoming (follower) ties to exclude those who potentially do not use Twitter to maintain real social connections, i.e., typically commercial or political profiles. We exclude users who had an extreme number of connections, i.e., more than 5000 outgoing or more than 5000 incoming connections. Note that the data is from the time period 2012–2013, when to have more than 2000 outgoing contacts, users had to surpass a threshold of at least 2000 followers (see (8)). In addition, only 1415 users (0.1% of the total) followed more than 5000 other users at that time; in our work, we focus only on mutual followership ties.

Fig. SI1 shows the network size in terms of nodes and edges in all 50 metropolitan areas under study. Edges are based on mutual followership. Besides presenting the raw numbers of social network nodes and edges for each metropolitan area, Table SI1 reports the bounding box coordinates we used for our modeling.

Fig. SI2A illustrates that the distribution of the number of users across cities closely follows the distribution of population size from the census. In addition, we infer the socioeconomic



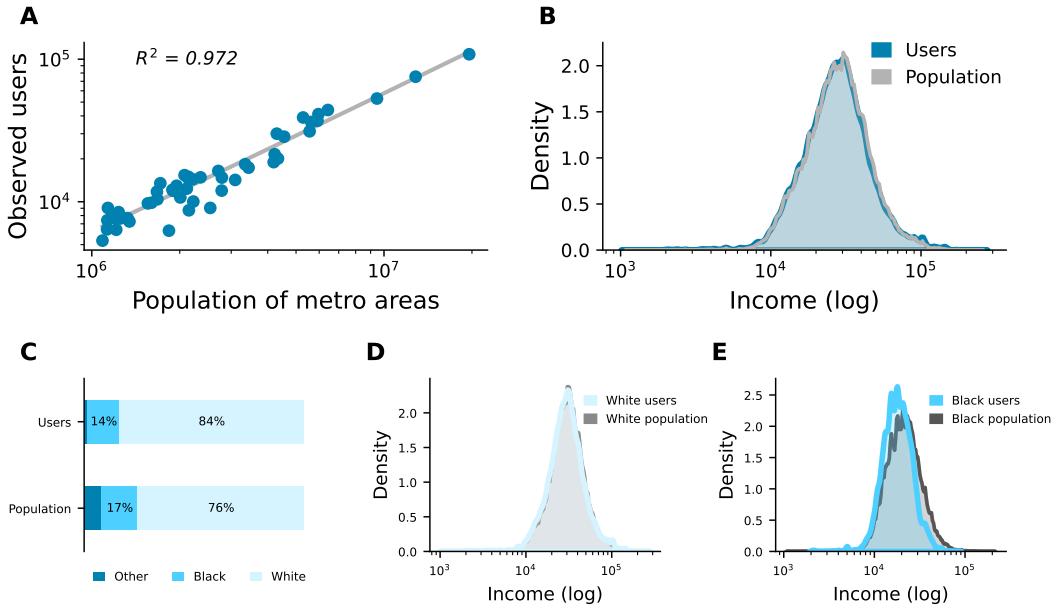
**Fig. SI1. Social network size.** Number of nodes and edges in the Twitter social network for the 50 largest metropolitan areas of the US.

status and demographics of users by linking their home location to census tract-level data on income and racial composition from the 2012 American Community Survey. While it is common practice in working with online social networks or mobility data to use estimated home locations for assigning socio-economic characteristics to users (1, 9, 10), this inference is not free from bias.

Fig. SI2B compares the users' inferred income distribution with the income distribution across all census tracts in metropolitan areas weighted by population. There is a near exact overlap between the two distributions. Additionally to the visual evidence, a permutation test with 1000 permutations does not reject the hypothesis that the two income distributions derive from the same underlying distribution (observed mean difference = 94 USD, significant at  $p > 0.05$ ).

Figure SI2C compares the racial composition of our sample with that of the general population, revealing a slight under-representation of minority groups in our sample (Black: 14% vs. 17% and Other: 2% vs. 7%) together with a slight over-representation of white users (84% vs. 76%). Figure SI2D-E breaks down the income distributions into the two largest

\*To whom correspondence should be addressed. E-mail: luai@itu.dk



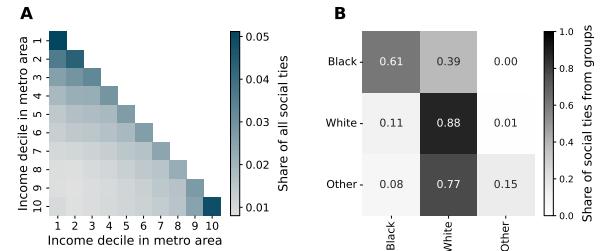
**Fig. SI2. Twitter data representativeness across various dimensions.** **A.** Correlation between population and the number of Twitter users in the 50 largest US metropolitan areas. **B.** Income distribution of users by income level in the census tracts of their home location and income distribution of the population of the 50 largest metropolitan areas. The population income distribution is derived from the population-weighted census tract level income. **C.** Comparison of the racial composition of users in our sample and in population. Users are assigned to racial groups based on the most populous racial group in the census tract of their home location. **D.** Comparison of the income distribution of white users and white population. **E.** Comparison of the income distribution of Black users and Black population.

racial groups, revealing an underrepresentation of the wealthier Black population in our sample. These discrepancies may result from two factors: a potential racial bias in our Twitter sample, and/or inaccuracies in estimating user race based on census data, as we assign users to the majority racial group of their home census tract.

Some of these representation disparities also manifest in social connectivity patterns. Figure SI3A indicates that we observe slightly more mutual followership ties within lowest and highest income deciles. Figure SI3B illustrates the pattern of connections by racial group, and shows a tendency for intra-group connections in the two most populous groups (white 84%, Black 14% of all users). In previous studies, assortative connection patterns were observed across both income and racial groups in large US metropolitan areas (1, 3, 6, 11, 12). In particular, some studies found that the level of homophily observed in social ties tends to be even more pronounced for the lower and upper ends of the income distribution, as we find also in our data (13, 14).

**B. Home Location Estimation.** We start with the friend-of-friend algorithm (15) to identify users' home locations from geocoded tweets, at a precision below census tract level on grid cells of  $100 \times 100$  m. This algorithm starts by identifying the three densest spatial clusters of geocoded datapoints for each user. Two geotagged tweets of the same user are considered to belong to the same spatial cluster if they are at most 1 km apart. To eliminate outliers, we iteratively filter out from each cluster the datapoints that are most distant from the cluster centroid, until all points are at most within a  $3\sigma$  radius from the centroid. After this trimming process, the three highest cardinality clusters per user are retained.

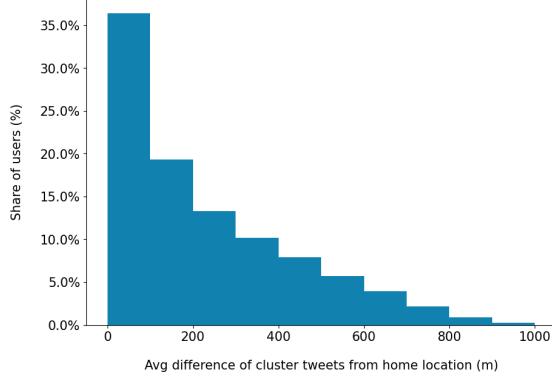
We only keep users who have at least two of their three



**Fig. SI3. Social connections between income deciles and across racial groups.** **A.** Distribution of social ties across income deciles. Deciles are calculated separately for each metropolitan area and users are assigned to deciles by their census tract of home location. **B.** Connection patterns across racial groups. The matrix is row normalized to illustrate the proportion of ties from one racial group (rows) to all other groups (columns).

clusters falling within the same US metropolitan area which both contain at least 50 tweets posted during weekdays over the 2-year period covered by the dataset. In line with established practices (6, 16, 17), we label the cluster with the most tweets between 8PM and 8AM as the user's home location.

To gauge the robustness of the home location estimation, we calculated the average distance of users' tweets from their home location, using the geolocated messages that have been labelled as belonging to one cluster by the friend-of-friend algorithm and the iterative  $3\sigma$  trimming. Figure SI4 shows the share of Twitter users with the average distance of tweets from their home location in 100 m bins. Altogether, for slightly more than 35% of users, this average distance is below 100 m, and for an additionally almost 20%, between 100 and 200 m. Note that the average might still be affected by outliers even after



**Fig. SI4. Robustness of the home location estimation.** Distribution of the average distance of tweets labelled as belonging to one cluster after the friend-of-friend algorithm and the iterative  $3\sigma$  trimming process. The vertical axis shows the share of users from our sample per 100 m bins.

the trimming process, and that this gives little information on the actual shape of the spatial distribution of tweets around the home location.

**C. Street network data.** We download the street network data from OpenStreetMap (OSM) (18) using the OSMnx Python package v.1.2.1 (19). Specifically, we use the `osmnx.graph.graph_from_bbox` function with the parameters `network_type = all_private`, `simplify = False`, `retain_all = True`, `truncate_by_edge = True`, `clean_periphery = False`. For each city, we construct the network comprising all the streets within the bounding box of their Metropolitan Statistical Area (MSA) as defined by the US Census Bureau (Table SI1).

All OSM street segments are labeled with the OSM attribute `highway`, which identifies the type of street that the segment represents. For each city, we extract the network of highways, freeways and major transportation roads such as interstates, by considering segments labeled as `highway=motorway`, `highway=trunk`, `highway=motorway_link`, or `highway=trunk_link`.

We simplify the resulting graphs with the OSMnx function `simplification.simplify_graph` to remove interstitial nodes. The resulting street network for each city is a graph with edges representing street segments and nodes representing the intersections between them. For the qualitative studies of the 9 cities presented in the case study (Fig. 4 in the main text), we use the QGIS software (20) to additionally simplify the graphs, manually removing truncated street segments and merging small consecutive highway segments between intersections into larger segments.

**D. Spatial null model detailed description.** The *Directed Configuration Model* (DCM) (21) is a network null model suited for directed social graphs, and has been used extensively in network science. To randomize the connections in an existing directed graph, DCM converts all incoming and outgoing edges of a node into in- and out-stubs, namely ‘dangling’ edges attached to the node. Then, each out-stub is matched with an in-stub selected uniformly at random to form a directed edge. This method generates a new random network characterized by the same degree sequence as the original network. However, when dealing with a network in which nodes occupy a position

in physical space, i.e., a spatial network, using a standard configuration model is not sufficient, as it does not consider the spatial constraints which can heavily influence connectivity patterns. In particular, social connectivity on commonly modeled by the *gravity model* (22, 23), an empirical relationship stating that the volume  $w_{ij}$  of social connections between two geographical areas  $i$  and  $j$  is proportional to the total number of possible connections between them (calculated as the product between the populations in the two areas  $N_i \cdot N_j$ ), and inversely proportional to a power of their Euclidean distance  $d_{ij}^\gamma$ :

$$w_{ij} = \frac{N_i^\alpha N_j^\beta}{d_{ij}^\gamma} \quad [\text{SI1}]$$

The exponents of the gravity formula can be estimated from real data by fitting it to a linear regression using the Ordinary Least Squares (OLS) method:

$$\log(w_{ij}) = \alpha \log(N_i) + \beta \log(N_j) - \gamma \log(d_{ij}) \quad [\text{SI2}]$$

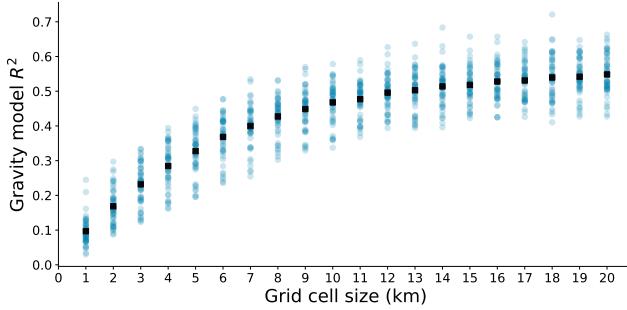
We verify empirically that the geographic arrangement of the nodes and ties in our Twitter data is compatible with the gravity model. Fig. SI5 shows the goodness of fit of a gravity model that estimates the number of social connections between two spatial cells in a regular grid overlaid on the city map. Only cells connected by at least one social tie are considered. The poorer fit at high spatial granularity is mainly given by data sparsity. When considering small cells, the average number of home locations within a cell is very low, which causes the majority of cells to contribute a noisy signal to the gravity model fit. To illustrate this point, we re-calculated the gravity model fit by randomly sampling users (Fig. SI6). As the sample size grows, the quality of the fit increases across all spatial resolutions, but much more so for larger cell sizes. Such variations in  $R^2$  growth correlate with the growth of the average number of users per cell. For example, doubling the user sample from 25% to 50% brings the number of users per cell from 20 to 32 for the 15 km resolution, while it increases the number users in 1 km cells only from 1.4 to 1.6 — an insufficient statistic to reliably estimate the gravity effect of social connections.

We further checked the robustness of our data’s fit with the Gravity Law with additional model specifications to illustrate the robustness of gravity models in our intra-city case. Following the approach of Silva and Tenreyro’s “Log of Gravity” (24), we experiment with both OLS and PPM (Poisson Pseudo-Maximum Likelihood), and by discarding (Tab. SI2) or considering (Tab. SI3) pairs of cells with zero-flow. We repeated the same analysis using census tracts instead of square cells (Tab. SI4). These regressions confirm that distance has a significant negative relations with the number of social ties between census tracts of metropolitan areas.

Our null model follows the algorithm of the configuration model and extends it with the gravity model, such that both degree sequences and spatial connectivity patterns are preserved. First, all ties in the network are turned into in- and out-stubs. Then, an out-stub  $i$  is selected at random. Let  $j$  be the stub that was originally connected to  $i$ , and  $d_{ij}$  be the Euclidean distance between them. The set of candidate in-stubs for the random rewiring of  $i$  is now restricted to those that are approximately at distance  $d_{ij}$  from  $i$ , namely the set of stubs  $S = \{k | d_{ij} - \epsilon \leq d_{ik} \leq d_{ij} + \epsilon\}$ . We set empirically

Cbsacode	City	State	#Nodes	#Edges	Bounding box				City center	
					West	South	East	North	Lat	Lon
12060	Atlanta	GA	39006	147378	-85.387	32.845	-83.269	34.618	33.757	-84.388
12420	Austin	TX	13473	38994	-98.298	29.631	-97.024	30.906	30.270	-97.743
12580	Baltimore	MD	16453	51322	-77.312	38.711	-75.748	39.722	39.291	-76.614
13820	Birmingham	AL	6402	27264	-87.422	32.660	-86.044	34.260	33.520	-86.810
14460	Boston	MA	28604	75679	-71.899	41.566	-70.323	43.573	42.361	-71.056
15380	Buffalo	NY	9046	27104	-79.312	42.438	-78.460	43.635	42.882	-78.875
16740	Charlotte	NC	14367	39336	-81.538	34.458	-79.848	36.059	35.226	-80.843
16980	Chicago	IL	52964	172016	-88.942	40.737	-86.929	42.670	41.881	-87.627
17140	Cincinnati	OH	12358	43134	-85.299	38.473	-83.673	39.729	39.101	-84.513
17460	Cleveland	OH	15367	64952	-82.348	40.988	-81.002	42.252	41.498	-81.695
18140	Columbus	OH	11918	30576	-83.653	39.362	-82.024	40.713	39.961	-82.997
19100	Dallas	TX	44015	136458	-98.067	32.052	-95.859	33.434	32.788	-96.800
19740	Denver	CO	9051	15933	-106.210	38.693	-103.706	40.044	39.752	-104.999
19820	Detroit	MI	30031	118146	-84.158	42.028	-82.334	43.327	42.318	-83.038
25540	Hartford	CT	6364	13628	-73.030	41.178	-72.099	42.039	41.767	-72.673
26420	Houston	TX	36978	107845	-96.622	28.765	-94.353	30.630	29.774	-95.361
26900	Indianapolis	IN	12121	34216	-87.015	39.048	-85.576	40.380	39.768	-86.158
27260	Jacksonville	FL	7272	14124	-82.460	29.622	-81.151	30.830	30.327	-81.656
28140	Kansas City	MO	10715	28848	-95.188	38.026	-93.477	39.789	39.089	-94.588
29820	Las Vegas	NV	12945	26280	-115.897	35.002	-114.043	36.854	36.167	-115.148
31080	Los Angeles	CA	75385	154524	-118.952	32.750	-117.413	34.823	34.059	-118.253
31140	Louisville	KY	8451	31520	-86.330	37.806	-84.867	38.784	38.257	-85.760
32820	Memphis	TN	7681	21844	-90.589	34.424	-89.184	35.652	35.148	-90.051
33100	Miami	FL	31201	70629	-80.886	25.137	-79.974	26.971	25.775	-80.193
33340	Milwaukee	WI	9736	34310	-88.542	42.842	-87.069	43.544	43.038	-87.918
33460	Minneapolis	MN	18333	57216	-94.262	44.196	-92.135	46.247	44.983	-93.271
34980	Nashville	TN	11766	33716	-87.567	35.408	-85.779	36.652	36.161	-86.779
35380	New Orleans	LA	8127	14178	-90.964	28.855	-88.758	30.712	29.952	-90.074
35620	New York	NY	108292	313702	-75.359	39.475	-71.777	41.602	40.767	-73.979
36420	Oklahoma City	OK	7638	18882	-98.313	34.681	-96.619	36.165	35.472	-97.498
36740	Orlando	FL	14649	23406	-81.958	27.642	-80.861	29.277	28.543	-81.380
37980	Philadelphia	PA	41170	124130	-76.233	39.290	-74.39	40.609	39.951	-75.156
38060	Phoenix	AZ	18961	29693	-113.335	32.501	-110.448	34.048	33.449	-112.071
38300	Pittsburgh	PA	14827	48812	-80.519	39.721	-78.974	41.173	40.444	-79.996
38900	Portland	OR	10032	18974	-123.786	44.886	-121.514	46.389	45.519	-122.674
39300	Providence	RI	9845	21788	-71.907	41.096	-70.752	42.096	41.824	-71.413
39580	Raleigh	NC	7414	18822	-78.995	35.255	-78.007	36.266	35.774	-78.640
40060	Richmond	VA	7636	24652	-78.241	36.708	-76.645	38.008	37.540	-77.438
40140	Riverside	CA	21543	36396	-117.803	33.426	-114.131	35.809	33.983	-117.373
40900	Sacramento	CA	8708	12620	-122.422	38.018	-119.877	39.316	38.571	-121.480
41180	Saint Louis	MO	14772	49526	-91.419	38.003	-89.138	39.523	38.628	-90.183
41620	Salt Lake City	UT	5336	11250	-114.047	39.904	-111.553	41.077	40.758	-111.895
41700	San Antonio	TX	14943	43022	-99.603	28.613	-97.631	30.139	29.424	-98.491
41740	San Diego	CA	14222	19166	-117.611	32.529	-116.081	33.505	32.725	-117.157
41860	San Francisco	CA	20091	67850	-123.174	37.054	-121.469	38.321	37.784	-122.437
41940	San Jose	CA	6267	6194	-122.203	36.197	-120.597	37.485	37.331	-121.886
42660	Seattle	WA	17333	41986	-122.853	46.728	-120.907	48.299	47.605	-122.335
45300	Tampa	FL	11974	19120	-82.909	27.571	-82.054	28.695	27.951	-82.454
47260	Virginia Beach	VA	10405	29284	-77.502	36.029	-75.709	37.603	36.842	-76.134
47900	Washington	DC	36271	100740	-78.453	37.991	-76.322	39.720	38.901	-77.039

**Table SI1. Summary of the 50 cities with their bounding box coordinates and size of the Twitter social network (nodes and edges).**

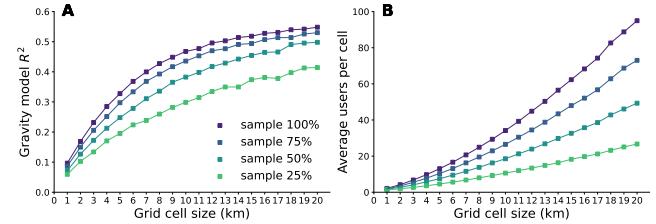


**Fig. SI5. Gravity model of Twitter data.**  $R^2$  goodness of fit of a linear regression to estimate the number of social connections between two areas from their geographical distance and their respective number of Twitter users' home locations (Equation SI2). A good fit indicates that the geographical patterns of the social connections are compatible with a gravity law. Each point on the plot represents the result of a regression ran on a single city and considering the tiles of a regular grid with a fixed granularity. The black squares represent the average of all the realizations for the given grid granularity. The fit stabilizes at around  $R^2 = 0.5$  when considering tiles of 10 km of side.

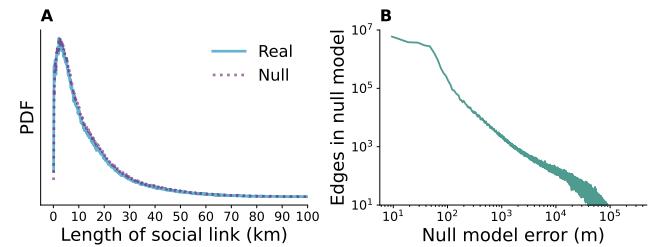
$\epsilon = 50m$ . The matching in-stub  $k$  is randomly selected among all candidates with probability that is proportional to their local density  $N_k$ , namely the number of nodes in the area surrounding the node with that stub. Such an area is empirically defined as a circle of radius 500m centered around the candidate node. This selection based on local density reflects the gravity law in Equation SI1. From that Equation, our algorithm can disregard both  $N_i$ , because the out-node  $i$  is fixed, and  $d_{ij}$  because all candidate nodes are selected to be at approximately distance  $d_{ij}$  from  $i$ . Last, the re-wired stubs are removed from the data. The algorithm iterates over the remaining in-stubs until none are left.

As long as the set of candidate in-stubs  $S$  is not empty at any iteration of the algorithm (i.e., at least one candidate in-stub is found at the desired distance), the rewired social network produced by this algorithm is a null model that exhibits strong properties; not only it reproduces the degree sequence and spatial connectivity patterns of the original network, but it also preserves the length distribution of social ties departing from *each* individual node. However, this last property is not always strictly guaranteed, since the iterative nature of the algorithm may lead to the exhaustion of the set  $S$  if all suitable in-stubs for the current out-stub have been used in prior iterations. When such an event occurs, the algorithm incrementally increases the buffer size  $\epsilon$  until  $S$  contains at least one element. In the extreme case where  $\epsilon \rightarrow \infty$ , all available stubs are considered with selection probability proportional to their local density, reflecting a global gravity law.

Our empirical observations indicate that the difference between the length  $d_{ij}$  of an original social tie and the length  $d_{ik}$  of its randomly rewired counterpart is minimal. The distributions of lengths in the real and null models are indistinguishable according to a paired Kolmogorov-Smirnov test ( $p = 0.0$ ). The median error is 26 m, and it is at most 100 m for 92% of the ties. In relative terms, 90% of the null ties have a length within a 2% error compared to their corresponding real ties. The distributions of tie lengths and null model errors are provided in Fig. SI7. Notably, the length distribution in Fig. SI7A, where most ties are beneath the length of 25–30 km, is in line with previous findings from empirical studies on



**Fig. SI6. Gravity model of sampled Twitter data.** **A.**  $R^2$  goodness of fit of a linear regression to estimate the number of social connections between two areas from their geographical distance and their respective number of Twitter users' home locations. Each point represents the average fit of the 50 cities. Different curves represent the fit obtained by random subsampling the users within each city. **B.** The number of users per cell, averaged across the 50 cities, for different subsamples.



**Fig. SI7. Null model error.** **A.** Probability density function of the geographical length of Twitter social ties in the real data and in the null model. The two distributions are indistinguishable according to a paired Kolmogorov-Smirnov test ( $p = 0.0$ ). **B.** Distribution of the absolute error on the geographical length of social ties introduced by the null model. The error is calculated as the absolute difference between the length of a real tie and the length of the corresponding tie in the null model.

social tie length (25) and face-to-face interaction frequencies (26).

**E. Statistical significance of Barrier Scores.** Figure SI8 shows an heatmap displaying Barrier Scores with non-significant values crossed-out ( $p > 0.01$ ). Multiple instances of negative Barrier Scores are non-significant, especially at short distances, confirming our initial intuition.

**F. Barrier Scores for longer distances.** The Barrier Scores for distances up to 20 km are shown in Figure SI9. In most cities, the scores approach zero at distances longer than 10 km. The confidence intervals of the mean shows high uncertainty in a few cases at very short distances, due to the scarcity of short-range social ties in some cities. Notably, the uncertainty is highest for those cities that have negative barrier scores at short distances, suggesting that those negative values are likely artifacts of data scarcity.

**G. Negative Barrier Scores at short distances.** Significant negative scores are rare across all distances and cities, and especially, most occurrences of negative Barrier Scores are non-significant. To develop a fine-grained understanding for negative values of the Barrier Score at short distances in a local context, we focus on the city of Jacksonville, which we identify as the most significant outlier. We visualize the Barrier Scores for shorter-distance ties on aggregated highway segments (see Fig. SI10). This closer look shows that highway segments with negative Barrier scores correspond to bridges that cross a natural barrier (St. John's river). Notably, all bridges con-

Estimator	OLS	OLS	OLS	OLS	PPML	PPML	PPML	PPML
Dependent variable	$\log(t_{ij})$	$\log(t_{ij})$	$\log(t_{ij})$	$\log(t_{ij})$	$t_{ij} > 0$	$t_{ij} > 0$	$t_{ij} > 0$	$t_{ij} > 0$
Grid cell size	1 km	5 km	10 km	20 km	1 km	5 km	10 km	20 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance (log)	-0.056*** (0.008)	-0.413*** (0.023)	-0.775*** (0.029)	-1.189*** (0.035)	-0.242*** (0.001)	-1.882*** (0.004)	-3.089*** (0.008)	-3.886*** (0.016)
Population <sub>i</sub> (log)	0.059*** (0.008)	0.191*** (0.007)	0.309*** (0.009)	0.453*** (0.013)	0.301*** (0.003)	1.144*** (0.003)	1.526*** (0.005)	1.673*** (0.008)
Population <sub>j</sub> (log)	0.057*** (0.005)	0.183*** (0.006)	0.306*** (0.008)	0.446*** (0.011)	0.287*** (0.001)	1.080*** (0.003)	1.453*** (0.005)	1.684*** (0.007)
Metro fixed effect	Yes							
Observations	1,128,349	420,630	156,908	41,948	1,128,349	420,630	156,908	41,948
R <sup>2</sup>	0.148	0.347	0.506	0.650	-	-	-	-

**Table S12. OLS and PPML regressions illustrate the effectiveness of gravity models in predicting social ties.** OLS stands for Ordinary Least Squares, while PPML stands for Poisson pseudo-maximum-likelihood. The models are fit on the data from the 50 largest metropolitan areas, and are tested at different grid cell sizes. The sample of each model consists of all grid cell pairs that are connected by at least one social tie. All model specifications indicate that social connections are more likely between geographically close locations. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

Estimator	OLS	OLS	OLS	PPML	PPML	PPML
Dependent variable	$\log(1 + t_{ij})$	$\log(1 + t_{ij})$	$\log(1 + t_{ij})$	$t_{ij}$	$t_{ij}$	$t_{ij}$
Grid cell size	5 km	10 km	20 km	5 km	10 km	20 km
	(1)	(2)	(3)	(4)	(5)	(6)
Distance (log)	-0.082*** (0.006)	-0.201*** (0.010)	-0.407*** (0.013)	-2.916*** (0.003)	-3.725*** (0.006)	-4.627*** (0.014)
Population <sub>i</sub> (log)	0.031*** (0.001)	0.065*** (0.002)	0.120*** (0.005)	1.696*** (0.002)	1.704*** (0.003)	1.687*** (0.007)
Population <sub>j</sub> (log)	0.030*** (0.001)	0.065*** (0.002)	0.117*** (0.004)	1.646*** (0.002)	1.648*** (0.003)	1.701*** (0.007)
Metro fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,642,246	1,352,109	192,858	8,642,246	1,352,109	192,858
R <sup>2</sup>	0.133	0.211	0.262	-	-	-

**Table S13. OLS and PPML regressions considering all grid cell pairs to illustrate the effectiveness of gravity models in predicting social ties.** OLS stands for Ordinary Least Squares, while PPML stands for Poisson pseudo-maximum-likelihood. The models are fit on the data from the 50 largest metropolitan areas, and are tested at different grid cell sizes. The sample of each model includes every grid cell combination, even without any observed social ties. All model specifications indicate that social connections are more likely between geographically close locations. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

necting the two river banks are part of the highway network; the relevance of bridges as connecting elements (“bottlenecks”) of a street network is thus further exacerbated by the fact that there are no alternative (e.g., pedestrian) crossings. This finding provides a plausible explanation for the phenomenon.

**H. Barrier Scores remain positive after considering other physical barriers.** To check for physical barriers other than highways that could confound our estimation of the Barrier Scores, we perform two additional pieces of analysis that account for barrier types other than highways. We decided which barrier types to include based on previous studies placing a strong emphasis on transportation infrastructure playing a crucial role for urban morphology, in particular with respect to the physical barriers that it introduces in the urban space (27). Previous literature concurs on considering transportation infrastructure as the union of major roads for motorized trans-

port (highways); railways; and waterways (all water features such as rivers, canals and lakes) (28–30). These categories of urban features are commonly used in recent studies that explore societal impacts of transportation infrastructure (31–33). Therefore, in addition to highways, we consider railways and waterways.

We base our experiments on manually curated OpenStreetMap (OSM) (18) data for railways and waterways within each of the 50 metropolitan areas. The manual curation was needed due to data quality, tag diversity in crowd-sourced OSM data, and interestingly, also because a substantial number of features marked as “railways” in the data have been converted to hiking trails (see e.g., the initiatives by the Rails To Trails Conservancy (34)).

First, we re-calculated the Barrier Scores by limiting the analysis to urban areas that do not contain any other major physical barriers than highways. To do that, we consider only

Estimator	OLS $\log(1 + t_{ij})$ (1)	OLS $\log(t_{ij} > 0)$ (2)	OLS $\log(1 + t_{ij})$ (3)	PPML $t_{ij}$ (4)	PPML $t_{ij} > 0$ (5)
Dependent variable	$\log(1 + t_{ij})$	$\log(t_{ij} > 0)$	$\log(1 + t_{ij})$	$t_{ij}$	$t_{ij} > 0$
Distance (log)	-0.098*** (0.012)	-0.163*** (0.014)	-0.026*** (0.008)	-1.511*** (0.004)	-0.710*** (0.003)
Population <sub>i</sub> (log)	0.017*** (0.005)	0.031*** (0.008)	0.005** (0.002)	0.402*** (0.007)	0.178*** (0.005)
Population <sub>j</sub> (log)	0.004 (0.005)	-0.001 (0.009)	0.008*** (0.003)	0.015*** (0.004)	-0.041*** (0.003)
Metro fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	2,669,688	672,846	26,998,504	2,669,688	672,846
R <sup>2</sup>	0.040	0.091	0.042	-	-

**Table SI4. OLS and PPML regressions illustrate the effectiveness of gravity models in predicting social ties between census tracts.** OLS stands for Ordinary Least Squares, while PPML stands for Poisson pseudo-maximum-likelihood. Models 1 and 4 are the same as our main models, taking into account the pairs of census tracts with no (zero) social ties in a selected sample. Model 3 includes census tract pairs with zero observed ties to account for all possible pairs, and Models 2 and 5 are restricted to pairs with more than zero ties. All model specifications indicate that social connections are more likely between geographically close locations. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

the subset of social ties that are fully contained within regions that are not crossed by any railway or waterway, as illustrated in Figure SI11 for the city of Austin. The resulting Barrier Scores are shown in Fig. SI12. Note that in this method, by focusing on smaller regions rather than the whole metropolitan area, long-distance ties are by definition underrepresented. Nevertheless, the Barrier Scores obtained for short and medium distances confirm the same trends observed when using the full dataset, especially at short distances.

Second, we propose an alternative approach to adjust our null model for the presence of physical barriers when computing the Barrier Score. Unlike the previous solution, this method considers the whole metropolitan area, without partitioning it into smaller sections. Specifically, we modify the null model to discount the contribution of social ties for which the role of highway as barrier is potentially confounded by other physical barriers. The rationale behind this method is explained in Figure SI13. Essentially, we reduce the Barrier Score contribution from “confounded” null ties that intersect both highways and other barriers. We emphasize the *very conservative* nature of this adjustment, because of two reasons. Firstly, this adjustment assumes that the barrier effect of highways on social ties is *always nullified* when null ties intersecting a highway also intersect other types of barriers. Secondly, the contribution of those type of ties is discarded *irrespective of the proximity* between the highway and the other physical barrier they intersect. We find that even after applying this new, more stringent null model, Barrier Scores remain positive up to a distance of 8 km (Fig. SI14).

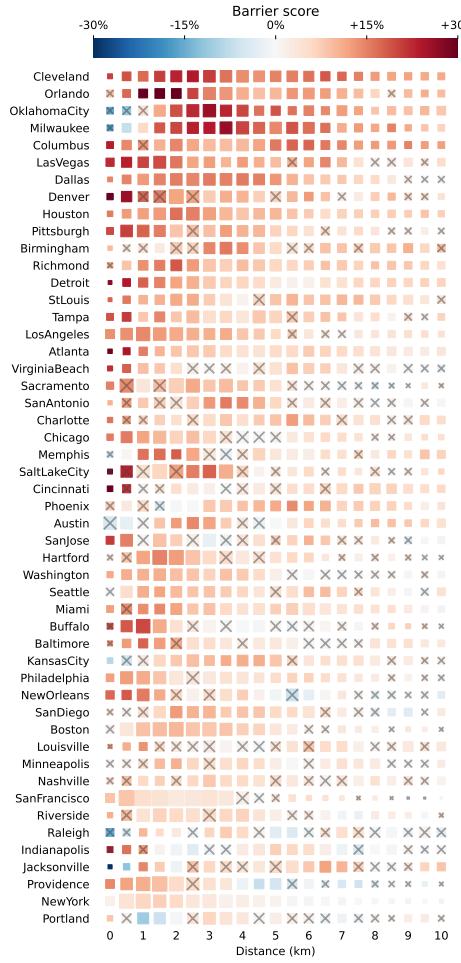
**I. Sensitivity analysis of city-level regression.** The regression results presented in Fig. 3 (main text) refer to a model which predicts a Barrier Score considering social ties with lengths up to 10 km ( $B_D$ , with  $D = 10\text{km}$ ). Fig. SI15 shows the regression coefficients and adjusted  $R^2$  for all values of  $D$  ranging from 1km to 50 km. The regression results hold in the range between 5 km and 30 km.

**J. Alternative regression models at the census tract pair level.** To test the robustness of the tract-level regression results, we first experiment with different ways of selecting the sample

of tract pairs to include in the regression, as summarized in Table SI5. The first three columns in Table SI6 compare different sampling criteria.

There are two common methodologies for selecting pairs of locations when constructing a gravity model (24). The first method involves selecting only those pairs of locations that are connected by at least one social tie. Model 2 in Tab. SI6 adheres to this criterion by including pairs linked by a minimum of one real social tie (672,846 observations). Consistent with our expectations, this model yields a negative and statistically significant coefficient for the number of highways crossed. The second approach involves the selection of *all* possible location combinations, irrespective of connections between them. Model 3 implements this strategy, thus obtaining 26,998,504 possible tract pairs. This is the only model for which we obtain a positive coefficient for the number of highways. This sign flip is mostly explained by sparsity. About 98% of the nearly 27 million possible tract pairs considered have no social ties connecting them. Given this heavy skew in the data, the OLS model infers that tracts pairs with non-zero number of ties between them are predominantly located within the densely populated urban cores of metropolitan areas, where Twitter user density is also higher. For example, the densely populated urban areas, as defined by the US Census, occupy 20% of the total metropolitan areas and contain 80% of our Twitter population. They also contain 52% of the highway infrastructure, with a highway density that is 8.4 times higher than suburban areas. As a result, urban tracts with dense social connections are much more commonly intersected by highways than the peripheral tracts with sparse Twitter penetration, resulting in a positive regression coefficient for the number of highways. Ultimately, this finding points to the fact that, while our experiments offer robust evidence on the barrier effect in urban areas characterized by a high density of ties and highways, they do not provide sufficient evidence for such an effect existing in suburban or rural areas.

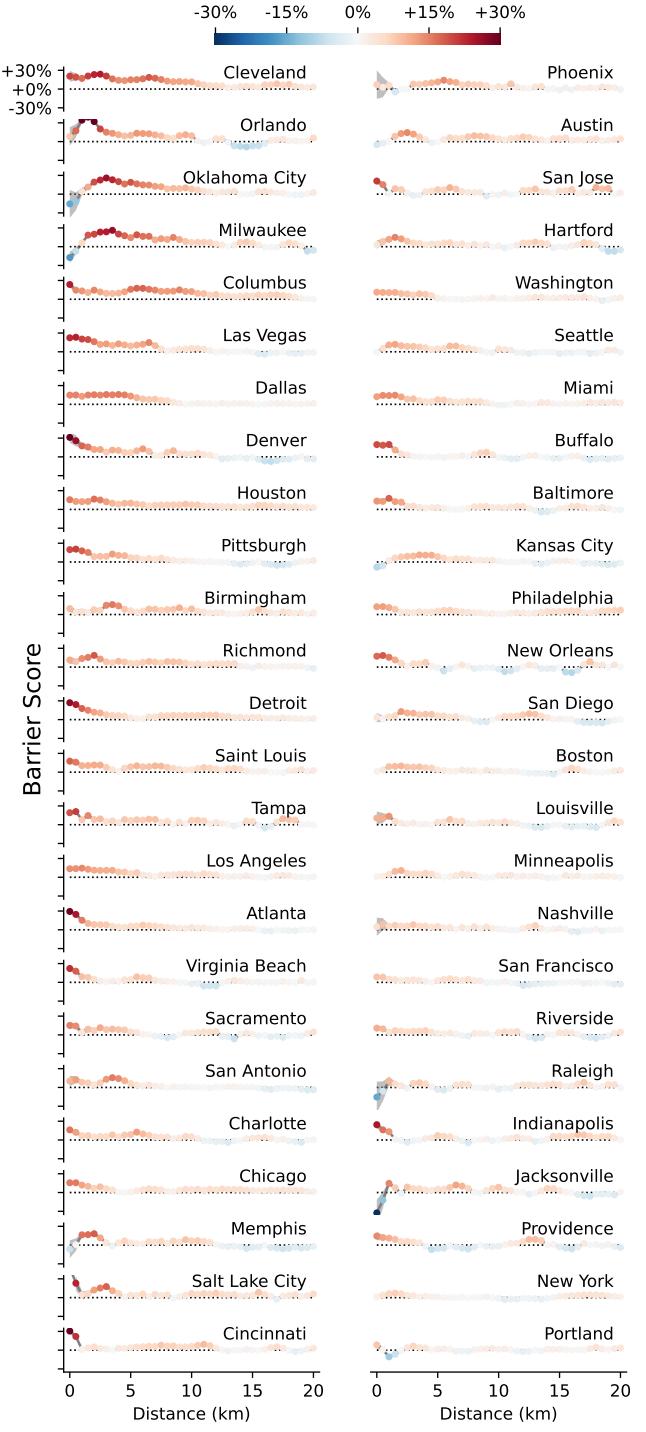
To overcome the issue of data sparsity and strike a good balance between the two sampling methods mentioned above, we leverage the information from our null models. We do so by constructing a sample of census tract pairs connected by



**Fig. S18. Statistical testing reveals most negative Barrier Scores are non-significant.** Heatmap of all Barrier Scores  $B(d)$  grouped into 0.5 km bins of social tie distance. Values marked with an  $\times$  are non-significant according to T-tests ( $p > 0.01$ ).

at least one real social tie *or* one null tie created in any of the random realizations of our null model. This approach effectively balances the two sampling strategies previously described, considering the 2,669,688 tract pairs that are either connected in the observed Twitter social network (672,846 observations), or that could plausibly be connected based on proximity and population (1,996,842 observations). This is a natural choice of sampling, as it includes all pairs of tracts that can be potentially connected according to our null model, while excluding pairs of locations that are too sparsely populated and far apart to exhibit even a minimal level of social connectivity. Model 1 in Tab. SI6 implements this strategy and is our preferred OLS regression presented in the main text. Its coefficients are again in line with expectations.

To further illustrate the robustness of our results, in Tab. SI6 we report two alternative models to our log-linear, OLS-based specifications using PPML (Poisson pseudo-maximum likelihood) regressions. PPMLs expect a Poisson distribution

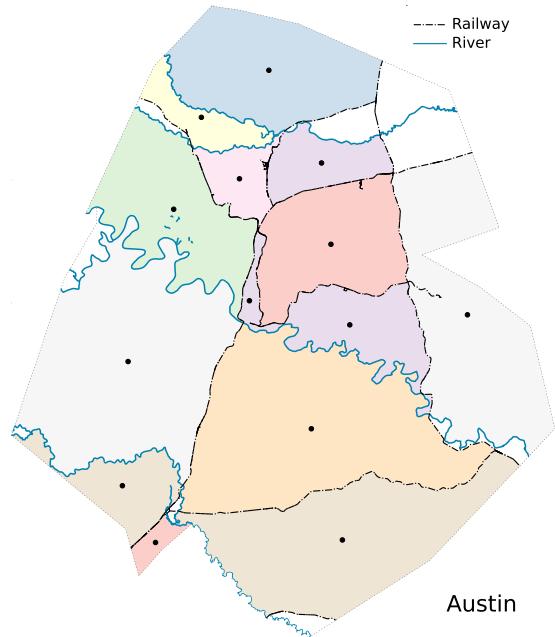


**Fig. S19. Barrier scores up to 20 km of distance.** The shaded gray areas represent the 95% confidence intervals.

and count data for the dependent variable, so they fit our data better. Model 4 based on our selected sample and Model 5 based on the sample of all observed connections show identical results to OLS models in terms of sign and significance for all of our variables. Due to computational demand, it was not possible to fit PPML models to observations from all possible census tract pairs, but we would expect results similar to the OLS setting.



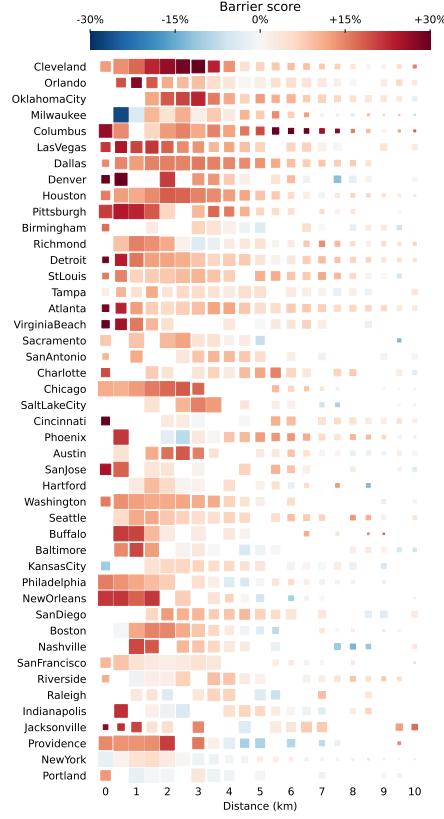
**Fig. SI10. Bridges explain negative Barrier Scores in Jacksonville.** Highway Barrier Scores for the highways in central Jacksonville. Red represents positive scores, blue represents negative scores. The bridges connecting downtown to the south bank facilitate the connection between two densely populated areas, resulting in negative Barrier Scores. The Acosta Bridge and the Main Street Bridge highlighted in the map are those that contribute the most to the negative Barrier Scores  $B(d)$  for values of  $d$  below 1.5 km.



**Fig. SI11. Polygonization of the metropolitan area of Austin based on railways and waterways.** Each polygon is delimited by railways, waterways, or the metropolitan area boundaries. Polygons containing an insufficient number of users are discarded. The remaining ones (colored and marked with a dot) are retained, and the Barrier Score is recalculated by considering only social ties that are fully within the polygons. The purpose of the polygonization is to measure Barrier Scores in geographical areas that are not confounded by other major physical barriers.

	No null model ties	Null model ties	Total
No social ties	24,328,816	1,996,842	26,325,658
Observed social ties	159,426	513,420	672,846

**Table SI5. Sample composition behind our main models.** Our null model is leveraged to construct the sample for our census tract pair level regressions. Light grey colors indicate the selected sample behind our preferred specifications.



**Fig. SI12. Barrier Scores in urban areas with no railways nor waterways.** Heatmap of all Barrier Scores  $B(d)$  grouped into 0.5 km bins of social tie distance. The estimated scores correspond to Barrier Score estimation that limits the analysis to social ties fully contained in areas that are not crossed by any railway or waterway. Non-significant values are not displayed.

Our controlled correlations in Table 1 (main text) and all the above models are pooled regression models with fixed effects at the metropolitan area level. This means that census tract pair level observations from different cities are combined, and a dummy control variable is added to account for metropolitan area specificities. We also compute Model 5 from Table 1 (main text) for each of the 50 metropolitan areas separately, and report the coefficients for the number of highways crossed in Fig. SI16. The results are in general consistent with the aggregated findings, with only 4 cities exhibiting positive and significant coefficients for that variable.

Lastly, we check the robustness of our regression results by changing its dependent variable. First, we fit the regression on the Barrier Score between two tracts as the dependent variable, simply defined as the ratio between the number of real and null ties connecting them. In this model, the sign and significance of the regression model's coefficients is in line with the other models (Table SI7). Second, as a sanity check, we fit a regression model on the number of null ties, and obtained non-significant coefficients for the number of highways crossed, which is expected since the null model is oblivious to highways (Table SI8).

#### K. Controls and re-weighting in regression models to account for race biases.

To account for the biases in the representation of different

Estimator	OLS $\log(1 + t_{ij})$ (1)	OLS $\log(t_{ij} > 0)$ (2)	OLS $\log(1 + t_{ij})$ (3)	PPML $t_{ij}$ (4)	PPML $t_{ij} > 0$ (5)
Dependent variable					
Nr highways crossed (log)	-0.013*** (0.003)	-0.038*** (0.006)	0.013*** (0.002)	-0.326*** (0.007)	-0.182*** (0.005)
Nr. of railways and waterways crossed (log)	-0.013*** (0.003)	-0.009* (0.005)	-0.004*** (0.001)	-0.108*** (0.004)	-0.011*** (0.003)
Income abs difference	-0.018*** (0.001)	-0.006** (0.003)	-0.002* (0.001)	-0.154*** (0.003)	-0.001 (0.002)
Racial similarity	0.023*** (0.003)	0.016*** (0.004)	0.003*** (0.001)	0.276*** (0.004)	0.055*** (0.003)
Distance (log)	-0.081*** (0.015)	-0.133*** (0.015)	-0.033*** (0.009)	-1.243*** (0.006)	-0.607*** (0.005)
Population (product log)	0.019** (0.008)	0.035*** (0.013)	0.009*** (0.003)	0.350*** (0.005)	0.157*** (0.004)
Metro fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	2,669,688	672,846	26,998,504	2,669,688	672,846
R <sup>2</sup>	0.051	0.098	0.044	-	-

**Table S16. Alternative regressions at the level of the census tract pairs to support our main models.** OLS stands for Ordinary Least Squares, while PPML stands for Poisson pseudo-maximum-likelihood. All independent variables are standardized in the same way for all models. Models 1 and 4 take into account the pairs of census tracts with no (zero) social ties but at least one null model tie connecting them. Model 3 includes all possible pairs of census tracts, and models 2 and 5 are restricted to pairs connected by at least one real tie. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

	Ratio of social ties to null model ties				
	(1)	(2)	(3)	(4)	(5)
Nr. of highways crossed (log)	-0.235*** (0.031)				-0.178*** (0.021)
Nr. of rail- and waterways crossed (log)		-0.090*** (0.005)			-0.040 (0.026)
Income abs. difference			-0.139*** (0.013)		-0.125*** (0.011)
Racial similarity				0.195*** (0.020)	0.177*** (0.018)
Distance (log)	-0.358*** (0.030)	-0.432*** (0.033)	-0.487*** (0.022)	-0.500*** (0.020)	-0.347*** (0.034)
User population (product log)	0.098*** (0.018)	0.099*** (0.019)	0.091*** (0.018)	0.077*** (0.018)	0.041** (0.017)
Metro fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	2,669,688	2,669,688	2,669,688	2,669,688	2,669,688
R <sup>2</sup>	0.009	0.010	0.010	0.011	0.012

**Table S17. Ordinary least squares regression models on the ratio of social ties and null model based ties between pairs of census tracts including spatial and socio-demographic features.** All the models include the metropolitan area as fixed effect. Crucially, the number of social ties between two tracts decreases with the number of highways that are crossed, after controlling for distance, user population, and socio-economic differences between the tracts. All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

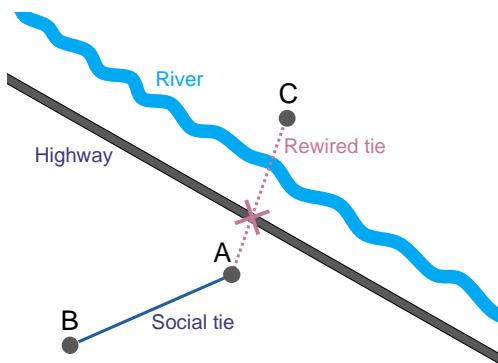
socioeconomic strata in our data (Fig. S12), we fit alternative regression models to those presented in Table 1 of the main text. These refined models incorporate both the income level of each tract and a new measure of racial similarity, defined as the cosine similarity of tract-level vectors that represent the population share of white, Black, and other racial groups. Compared to the regression model in Table 1 that uses solely a dummy variable for the majority racial group, these variables more accurately capture the nuanced relationship between social connectivity and socioeconomic factors. The resulting

coefficients suggest that the implementation of these enhanced measures of socioeconomic similarity does not impact our results, as detailed in Table S19 – the coefficient of the highways crosses remains negative and significant.

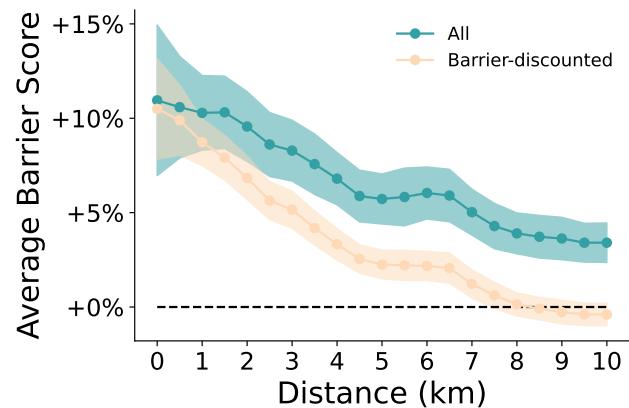
Third, while we cannot identify the income and race of individual users beyond their likely classification based on census tract of residence, we introduce a new analysis to test the robustness of our findings through *stochastic racial assignment*, a method employed in various studies in the literature on urban segregation (35, 36). This approach randomly assigns users to

	Nr social ties (log)	Nr null model ties (log)
	(1)	(2)
Nr highways crossed (log)	-0.013*** (0.003)	-0.003 (0.004)
Nr railways and waterways crossed (log)	-0.013*** (0.004)	0.005** (0.002)
Income abs. difference	-0.018*** (0.001)	-0.004* (0.002)
Racial similarity	0.023*** (0.003)	0.004* (0.002)
Distance (log)	-0.081*** (0.015)	-0.105*** (0.010)
User population (product log)	0.019** (0.0008)	0.029*** (0.009)
Metro fixed effect	Yes	Yes
Observations	2,669,688	2,669,688
R <sup>2</sup>	0.051	0.115

**Table SI8. Ordinary least squares regressions on the number of null model based connections between pairs of census tracts.** All the models include the metropolitan area as fixed effect. Model (1) is the same as our final model in the main text, considering real social connections as a dependent variable and is used for comparative purposes. Model (2) is based on the same sample as Model (1). All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ .



**Fig. SI13. Null model discounting the confounding effect from other physical barriers.** Let  $A$  be a user located close to a highway that was built alongside a river.  $A$  is connected through a social tie with user  $B$ , who is located on the same side of the highway as  $A$ . Suppose that the null model rewrites the social tie  $(A, B)$  to form a new null tie  $(A, C)$ , where user  $C$  is located on the other side of the highway and the river. In this illustrative example, the number of highways crossed in the null model ( $c_E^{null}$ ) is higher than the number of crosses in the real model ( $c_E$ ), thus contributing towards a positive Barrier Score. Given the null model assumptions, the null tie is interpreted as a connection that would have occurred in an hypothetical scenario where connectivity is not influenced by highways. However, since the null tie crosses both an highway and a river, it is uncertain whether the absence of the tie  $(A, C)$  in the real data is due to the highway or the river. To account for this uncertainty, we discard the effect of the tie  $(A, C)$  on the Barrier Score by simply reverting the null tie back to its original real tie  $(A, B)$ , so that the number of highways crossed in the two models is the same ( $c_E^{null} = c_E$ ). This revised null model applies this correction on every instance of null tie that crosses at least one highway and one barrier of a different type (waterway or railway).



**Fig. SI14. Average Barrier Score after discounting for other barriers** The Barrier Score vs. distance calculated with the regular null model and with an alternative null model that discounts the contributions to the Score given by ties crossing both highways and other barriers (waterways or railways), as illustrated in Figure SI13. Despite the discounting, the Barrier Score is still positive up to 8 km of distance.

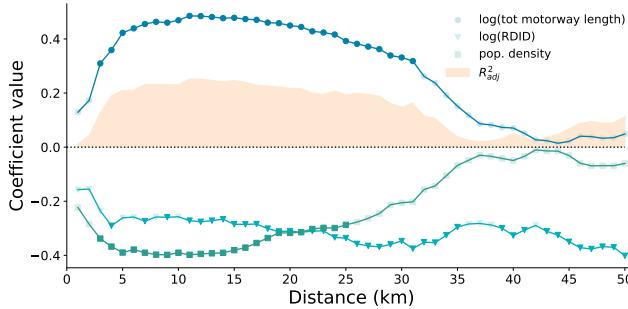
racial groups according to the actual demographic distribution of the population in the given tract. By doing so, we ensure that the racial composition among users reflects that of the population. This allows us to use racial similarity measures derived from these random assignments to calculate the proportion of same-race social connections between tracts. This exercise of randomized assignment was conducted 20 times for

robustness, with Table SI10 presenting an example regression from this procedure, and Figure SI17 illustrating the coefficients across the 20 regressions. The coefficient for the share of same-race ties was found to be positive and significant in all models ( $p < 0.05$ ), while our primary variable of interest, the number of highways crossed, consistently remained negative and significant across all models.

Finally, we implement a reweighting procedure to address income and racial biases. Initially, we categorize census tracts by their majority race (white, Black, or other) and income deciles (1 to 10), and count the number of users and the total population within these resulting 30 categories. We then input these counts into the Iterative Proportional Fitting (IPF) method (37), which iteratively calculates category weights while preserving the marginal population distributions

	Number of social ties (log)		
	(1)	(2)	(3)
Nr. of highways crossed (log)		-0.012*** (0.003)	
Nr. of railways and waterways crossed (log)		-0.014*** (0.004)	
Racial similarity (cosine similarity)	0.057*** (0.004)	0.064*** (0.004)	
Income tract1 (log)	-0.002 (0.005)	-0.015*** (0.005)	
Income tract2 (log)	-0.016*** (0.003)	-0.027*** (0.003)	
Income abs. difference	-0.018*** (0.003)	-0.011*** (0.001)	
Distance (log)	-0.099*** (0.013)	-0.098*** (0.013)	-0.080*** (0.015)
User population (product log)	0.034*** (0.009)	0.017** (0.008)	0.027*** (0.009)
Metro fixed effect	Yes	Yes	Yes
Observations	2,669,688	2,669,688	2,669,688
R <sup>2</sup>	0.046	0.049	0.056

**Table S19. Ordinary least squares regression models on the number of social connections with alternative variables on the socio-economic similarity of census tracts.** All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ .



**Fig. SI15. Sensitivity analysis of city-level regression.** Value of  $\beta$  coefficients and  $R^2_{adj}$  for OLS regression models aimed at predicting the Barrier Score calculated considering only social ties of length up to  $d$ . Transparent bullets indicate non-significant coefficients.

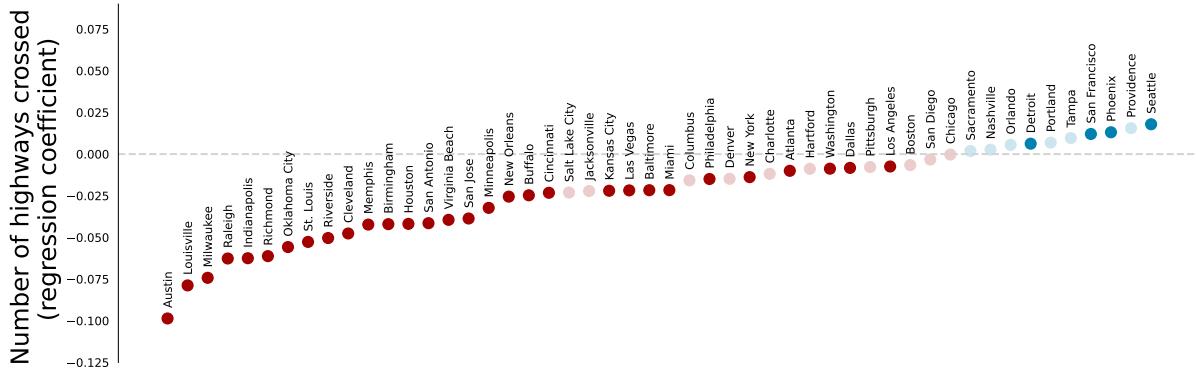
along the two selected variables – majority race and income decile. The resulting weights quantify the over- or under-representation of specific socioeconomic categories within our data (38). We assign these weights to the tracts based on their income decile and majority race. We adjust the number of social ties between each pair of census tracts by multiplying these ties by the weights associated with the respective tracts, assuming any potential bias in the observed ties is rooted in the representation biases of the underlying population groups. We use this reweighted connection count as an adjusted dependent variable, and we fit a new regression model whose the coefficients of which are presented in Table SI11. Despite accounting for potential under- and over-representation, the influence of highways on our results remained unchanged.

**L. Decreasing Barrier Score with distance.** Fig. 2 (main text) shows that the Barrier Score tends to decrease with distance. This suggests that the role of highways as barriers could change

if more distant locations are considered. To quantify this decreasing barrier effect, we rely on our census tract level regression setting (see Table 1 in main text). First, we include the interaction effect of distance and number of highways crossed to our preferred regression specification. Interactions allow us to test conditional effects of one variable (in our case, distance) on the contribution of another variable (in our case, highways crossed) to the dependent variable (number of social ties). Model 1 in Table SI12 matches our final model from the main text, while Model 2 contains the interaction term. The interaction of distance and number of highways crossed is positive and significant, while the sign and significance of all other variables remain unchanged. Models 3 and 4 show the same results, but these models are based only on census tract pairs with observed social ties (see Tab. SI6).

Second, we leverage this interaction term to visualize the changes in the coefficient of number of highways crossed in a two-way interaction term, conditional on the value of the other included variable, i.e. distance of census tracts. In other words, here we plot the estimated effect of the number of highways crossed on the number of social connections at different distances. Fig. SI18 shows that highways separating tracts within less than 20 km are associated with lower levels of social connectivity, whereas they are associated with a higher number of social ties for pairs of tracts that are farther apart. This suggests that highways might embody barriers to social ties at short and medium distances, but they might foster accessibility (hence, opportunities for social connections) at longer distances.

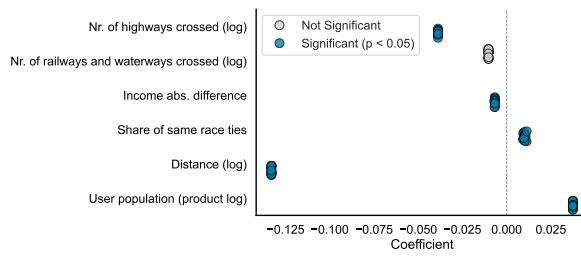
**M. Negative Highway Barrier Scores.** One explanation for the negative highway barrier scores in the maps displayed in Fig. 4 (main text) is the spatial distribution of ties crossing those highways. The colors in Fig. 4 encode the Highway Barrier Scores calculated considering *all* real and null ties crossing a



**Fig. SI16. Estimated effect of highways on social ties in each of the 50 metropolitan areas.** The coefficients are the results of separate models for each city, where the most detailed specification, Model 6 from Table 1 (main text) is used. Transparent markers indicate non-significant coefficients ( $p > 0.01$ ).

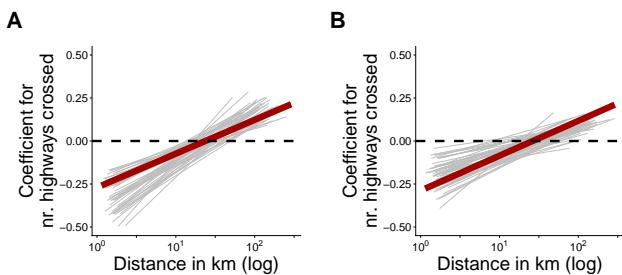
Number of social ties (log)	
	(1)
Nr. of highways crossed (log)	-0.039*** (0.006)
Nr. of railways and waterways crossed (log)	-0.010* (0.006)
Income abs. difference	-0.007** (0.003)
Share of same race ties	0.010*** (0.003)
Distance (log)	-0.133*** (0.015)
User population (product log)	0.037*** (0.013)
Metro fixed effect	Yes
Observations	672,846
R <sup>2</sup>	0.097

**Table SI10. Ordinary least squares regression models on the number of social connections after randomly assigning users to racial groups according to the underlying demographic distribution.** All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ .



**Fig. SI17. Results of 20 different regressions using stochastically assigned race to users.** Each dot represents a coefficient estimated in the same fashion as Table SI10 illustrates.

given highway. However, ties with different length tend to contribute differently to the Barrier Score, with long-distance ties having a higher tendency of yielding negative Barrier Scores (as illustrated in Fig. SI18). Therefore, when a highway is crossed by more long-distance ties than short-distance ones, its overall Highway Barrier Score can tip towards negative values. Figure SI19 illustrates this effect in the central part of Orlando,



**Fig. SI18. The effect of highways on social ties at different distances.** A. is based on Model 2 of Tab. SI12, while B. is based on Model 4. Both figures suggests that a higher number of highways between pairs of tracts is associated with a lower number of social ties between them, up to an inter-tract distance of 20 km. The red line illustrates the estimated effect and the associated confidence interval, which is very low for all distances in both model versions. The grey lines show the result of the same estimation for each city separately.

one of the cities discussed in Fig. 4. When considering only short-length social ties (Fig. SI19A), all highways considered have positive Barrier Scores, but those flip to negative when considering long-range ties only (Fig. SI19B). In aggregate,

	Nr social ties (log)	Reweighted nr social ties (log)
	(1)	(2)
Nr highways crossed (log)	-0.013*** (0.003)	-0.013*** (0.003)
Nr railways and waterways crossed (log)	-0.013*** (0.004)	-0.011*** (0.004)
Income abs. difference	-0.018*** (0.001)	-0.019*** (0.001)
Racial similarity	0.023*** (0.003)	0.026*** (0.003)
Distance (log)	-0.081*** (0.015)	-0.085*** (0.014)
User population (product log)	0.019** (0.0008)	0.039*** (0.005)
Metro fixed effect	Yes	Yes
Observations	2,669,688	2,669,688
R <sup>2</sup>	0.051	0.057

**Table S11. Ordinary least squares regressions on where the dependent variable was corrected for potential under- or overrepresentation.** All the models include the metropolitan area as fixed effect. Model (1) is the same as our final model in the main text, considers real social connections as a dependent variable and is used for comparative purposes. Model (2) uses the reweighted number of social ties as the dependent variable. All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

Estimator	OLS	OLS	OLS	OLS
Dependent variable	$\log(1 + t_{ij})$ (1)	$\log(1 + t_{ij})$ (2)	$\log(t_{ij} > 0)$ (3)	$\log(t_{ij} > 0)$ (4)
Nr highways crossed (log)	-0.013*** (0.003)	-0.270*** (0.013)	-0.038*** (0.006)	-0.288*** (0.012)
Distance (log)	-0.081*** (0.015)	-0.238*** (0.013)	-0.133*** (0.015)	-0.292*** (0.012)
Distance X Nr highways crossed		0.197*** (0.008)		0.203*** (0.006)
Nr. of railways and waterways crossed (log)	-0.013*** (0.003)	-0.013*** (0.004)	-0.009* (0.005)	-0.011** (0.005)
Income abs difference	-0.018*** (0.001)	-0.016*** (0.001)	-0.006** (0.003)	-0.004* (0.003)
Racial similarity	0.023*** (0.003)	0.021*** (0.003)	0.016*** (0.004)	0.013*** (0.003)
Population (product log)	0.019** (0.008)	0.018*** (0.007)	0.035*** (0.013)	0.035*** (0.012)
Metro fixed effect	Yes	Yes	Yes	Yes
Observations	2,669,688	2,669,688	672,846	672,846
R <sup>2</sup>	0.051	0.064	0.098	0.111

**Table S12. Controlled correlations at the level of the census tract pairs including interaction terms.** The interaction of distance and number of highways crossed in Model 2 and 4 are used to visualize the changing role of highways in the function of distance in Fig. S118. Standard errors are clustered at the metropolitan area level. \*\*\*: p<0.01, \*\*: p<0.05, \*: p<0.1.

this leads to mixed Highway Barrier Score signs (Fig. S119C). It is important to stress that not all highways exhibit this behavior, and the Barrier Score is consistently positive for many of them. In particular, the Barrier Scores of the highways that are discussed in Fig. 4 are positive across all distances.

**N. Segmentation of highways in qualitative analysis on racial segregation.** To calculate the Highway Barrier Scores depicted in Figure 4 in the main text, we divided the number of con-

nctions traversing each highway in the actual dataset by the corresponding count derived from the null model. This necessitates defining the specific highway segment over which these crossings are counted. The native OpenStreetMap data is not ideally suited for such analysis, as it delineates highways as a series of segments that can be exceedingly brief (occasionally just a few meters in length) and vary substantially in size across different cities and highways. Conversely, treating highways that extend for hundreds of kilometers as a single



**Fig. SI19. Aggregating over short and long distance ties can lead to mixed Highway Barrier Score signs.** Highway Barrier Scores for the highways in central Orlando. Red represents positive scores, blue represents negative scores. The scores are calculated considering: **A.** only short-distance ties up to 5 km; **B.** and only long-distance ties between 15 km and 20 km; **C.** highway crosses from all ties.

segment would not effectively illustrate the variation in Barrier Score as the highway intersects various parts of the city. Lacking alternative data sources to guide the segmentation, we opted for a manual approach, taking into account three criteria: intersections, sharp bends, and a relative consistency in segment length.

To verify the robustness with respect to different choices of partitioning, we recalculated the Barrier Scores for several other partitions with different maximum segment lengths. Specifically, we also conducted automated partitioning based on a two-step segmentation approach, testing robustness across different values for the maximum threshold (maximum segment length)  $t_{max}$ . To do so, we first take all LineStrings from the manually simplified highway network, and segmented them with a precision of  $t_{max}/10$ , by means of adding additional interstitial nodes along the linestring coordinates, which ensures that the maximum distance between any two directly coordinates in direct succession does not exceed  $t_{max}/10$  meters). Then, for each LineString segmented in this way, we iteratively (adding one coordinate at each iteration step) create segments of at most length  $t_{max}$ . Figure SI20 shows results for the example of Detroit where highways were automatically partitioned as described, considering eight different maximum segment lengths, with  $t_{max}$  ranging from 500m to 3km. The patterns obtained with alternative partitioning strategies are consistent with those reported in the main manuscript, confirming that our conclusions do not depend on the specific partitioning.

**O. Barrier Scores of other street types.** We compared the Barrier Score calculated on highways with Barrier Scores calculated considering other types of street included in Open Street Map's categorization. Fig. SI21 presents the distance-constrained Barrier Score  $B(d)$  for three further street types in addition to highways, by descending road hierarchy: *primary* roads, *secondary* roads, and *residential* streets (i.e., streets that provide direct access to housing). While all road types yield positive scores, the highway score is markedly higher than all others.

**P. Randomization of highways.** The observed decrease in Barrier Scores when considering streets with less vehicular traffic, such as residential streets, suggests that streets further down in the road hierarchy do not significantly impede mobility and social interactions. However, this interpretation could be confounded by the inverse correlation between the typical volume of vehicular traffic on streets of a given type and the

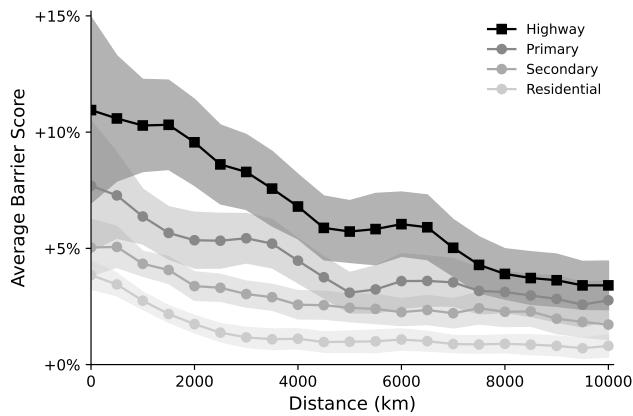
abundance of that street type in the urban network. For instance, large American cities often feature only a small number of heavily trafficked highways, contrasted by a large number of residential streets. This raises the question of whether the observed Barrier Score patterns on highways are merely a statistical artifact of their relative scarcity.

To address this concern, we recompute the Barrier Scores on a hypothetical highway network of comparable length to the real network, but with randomly placed sections. To build the randomized version of a highway network of total length  $L$ , we use an iterative approach. At each iteration  $i$ , we select two random points within the bounding box of the target city, and find their respective closest nodes on the street network. We connect the two nodes with the shortest path between them, and add the obtained path to the randomized network. Let the length  $L_i$  be the total length of the randomized network up to iteration  $i$ . To ensure a minimal residual between the real and randomized networks, we discarded any iteration  $i$  that would cause the randomized network's size to exceed the real network size by more than 1% (namely, when  $L_i \geq 1.01 \cdot L$ ). The algorithm stops  $L_i \approx L$ , which we empirically represent with the range  $[0.99 \cdot L \leq L_i \leq 1.01 \cdot L]$ . The Barrier Scores are then averaged over 50 versions of the randomized layout for each city considered.

Fig. SI22 presents the results for a selected set of cities. In all cases, the Barrier Scores of the randomized street network were significantly lower than that of the real street network, particularly for ties of length up to 6-8km. This is true for cities that are not crossed by any major natural barriers (i.e., Dallas, Atlanta, Miami) as well as cities that are built around major water bodies (i.e., Boston, Pittsburgh, Seattle). However, the randomized street model still exhibits positive Barrier Scores, with values fluctuating with distance in a manner similar to the real highway network. This suggests that the randomized highway network is subject to the same spatial constraints imposed by the city's morphology and actual street network layout. For example, most shortest paths connecting the northern and southern parts of Seattle are bound to pass through the highway bridges that cross the Lake Union connecting Washington Lake to the bay (Interstate 5 or Route 99). Consequently, the spatial patterns of the randomized highway network cannot be fully disentangled from those of the real street network. Therefore, our results should be interpreted as upper bounds of the contribution of the highway network sparsity to the Barrier Score that we observe on the real data.

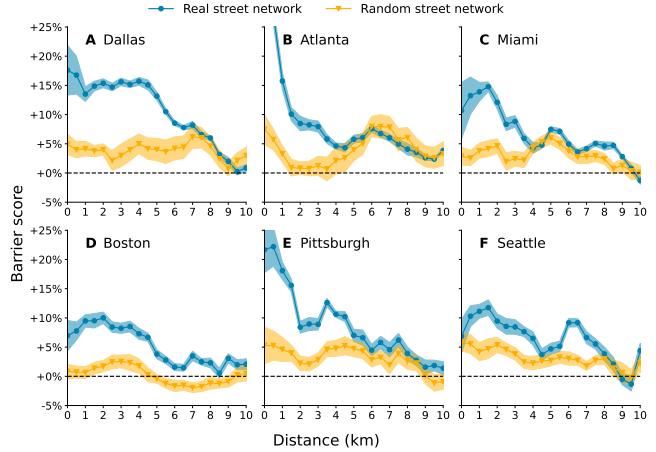


**Fig. SI20. Highway Barrier Scores in Detroit, calculated on highway segments of different length.** Highways are in color, following the color coding of Fig. 2 in the main text. Segment length used in automated partitioning is reported in the bottom-right corner of each map.



**Fig. SI21. The Barrier Score decreases with social tie distance for highways, to a lesser extent for other street types.** The distance-constrained Barrier Score  $B(d)$  across multiple distances, averaged over all cities, and calculated for different types of roads. Across all distances, streets that are higher up in the road hierarchy have higher Barrier Scores.

**Q. Barrier effect in urban vs. suburban areas.** Our results are based on data sets that have been spatially limited to metropolitan areas as defined by the US census\*. To test how population density and city center proximity affect our results, we conduct two additional experiments with more restrictive spatial limitations: first, limiting our input data to functional urban areas (FUAs) as defined by the Organisation for Economic Co-operation and Development<sup>†</sup> (OECD); and second, limiting our input data to a 20 km radius from the city center. The coordinates of the city centers we selected are reported in Table SI1. The radial areas around the city centers on average cover 10% of the metropolitan areas, and contain 55% of our Twitter population. FUAs on average cover 20% of the metropolitan areas, and contain 80% of our Twitter population. Crucially, we find that FUAs contain 52% of the highway infrastructure of metropolitan areas, and the density



**Fig. SI22. Random street layout.** Barrier Score for social ties at distance  $d$  for a model that considers the real highway network compared to a model that uses a randomized version of the highway network. Cities on top panels **A-C** are selected among cities without major natural barriers within their built environment, whereas cities on the bottom panels **D-F** are built around major water bodies.

of highways in FUAs (calculated as the total highway length per area unit) is 8.4 times higher than the highway density in suburban areas. Barrier Scores for urban areas follow closely the pattern obtained for the full dataset, whereas scores for suburban areas are generally lower and less stable (Fig. SI23A). The difference is less pronounced when comparing the 20 km radius around the center with the area outside that radius (Fig. SI23B). These findings provide additional evidence that the barrier effect of highways is more prominent in densely populated areas, and is less related to the proximity to the city center.

We also experiment with including distance to the city center in our regressions. We calculate the distance of each census tract from the identified city center and measured the distance of the census tract pairs from the center using the logarithm of the product of distances. This variable takes lower values when both census districts are close to the city

\* <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html>

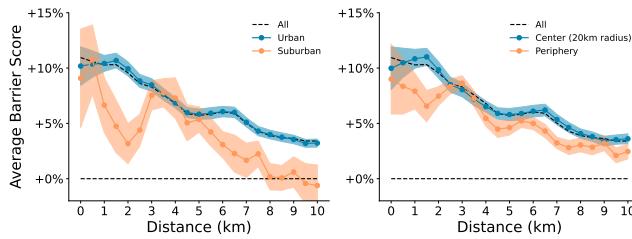
† <https://www.oecd.org/en/data/datasets/oecd-definition-of-cities-and-functional-urban-areas.html>

	Number of social ties (log)		
	(1)	(2)	(3)
Distance from center (product log)		0.012*** (0.004)	0.003 (0.003)
Nr. of highways crossed (log)	-0.013*** (0.003)		-0.012*** (0.003)
Nr. of railways and waterways crossed (log)	-0.013*** (0.004)		-0.012*** (0.003)
Income abs. difference	-0.018*** (0.001)		-0.017*** (0.002)
Racial similarity	0.023*** (0.003)		0.023*** (0.003)
Distance (log)	-0.081*** (0.015)	-0.109*** (0.012)	-0.085*** (0.013)
User population (product log)	0.019** (0.008)	0.023*** (0.008)	0.017** (0.008)
Metro fixed effect	Yes	Yes	Yes
Observations	2,669,688	2,669,688	2,669,688
R <sup>2</sup>	0.051	0.043	0.051

**Table SI13. Ordinary least squares regression models on the number of social connections between pairs of census tracts including spatial and socio-demographic features.** All the models include the metropolitan area as fixed effect. Model (1) is our final model in Table 1 of the main text and is used for comparison. The additional variable of distance from the center is significant in the simple model (2), but loses significance in a model that takes into account distance, user population and socio-economic differences between census tracts. The main variable for the number of highways crossed does not change when distance from the center is taken into account. All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ .

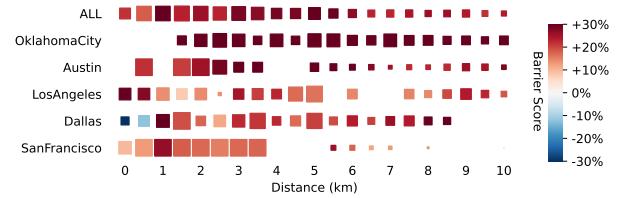
center. Table SI13 shows that considering distance from the center does not change our main results.

Last, we ran OLS regressions for pairs of tracts that occur in urban areas (Table SI14). We consider one model including tracts intersecting the urban area (Model (2)) and one model including tracts that are fully contained within the urban area (Model (3)). Both models exhibit coefficients that are in line with the regression on all tracts (Model (1)).



**Fig. SI23. Barrier Scores are higher in urban areas.** Barrier Scores versus distance calculated considering different subsets of users. **A.** Comparison between urban and suburban areas. Urban areas account for 80% of the population and for about 20% of the full metropolitan areas. **B.** Comparison between area around the city center (20 km radius) and area outside the center. The central radii contain about 55% of the population and account for 10% of the full metropolitan areas. Values are averaged over 20 randomized runs of the null model, and then macro-averaged over the 50 cities considered. 95% confidence intervals are shown. Both plots provide the values calculated on all users as reference.

**R. Alternative social network data: Gowalla.** Studying the relationship between the built environment and social connections requires large-scale social network data with fine-grained geographical information. To illustrate the robustness of our results, we repeat our computations using the dataset from the online social platform Gowalla, derived from its API (39).



**Fig. SI24. Barrier Score vs. distance in the Gowalla database.** Heatmap of Barrier Scores  $B(d)$  grouped into 0.5 km distance bands for the five cities most represented in the Gowalla dataset. Color denotes Barrier Score, areas of the squares denote relative number of links per distance band. The Barrier Scores are considerably higher than those observed on average in our Twitter dataset, across all distances.

Gowalla is a location-based social network where users share their locations by means of check-ins. Unlike in Twitter, Gowalla is designed for connecting people who know each other. The undirected friendship network from the site consist of 196,591 nodes and 950,327 edges. While the social network part of the data is similar to our Twitter network based on mutual followership, the geographic information available to determine the home location of users is different.

The Gowalla data contains 6,442,890 check-ins across 50 cities performed by 107,092 users from February 2009 to October 2010. Inspired by (39), we use the following procedure to determine the home location of users: First, as a minimum requirement, we focus on users who have at least 10 different visited locations in the dataset, and discard others. Second, we place check-ins into size 10 H3 hexagons (Uber's Hexagonal Hierarchical Spatial Index (40)). These hexagons refer to an average  $15000 m^2$  area, which is close to the buffer area of a point with a 70 m radius. Third, we identify the hexagon in which each user made the most visits as their home location

	Number of social ties (log)		
	Both tract within the metro area	Both tracts within the urban area	Both tracts intersect the urban area
	(1)	(2)	(3)
Nr. of highways crossed (log)	-0.013*** (0.003)	-0.011*** (0.003)	-0.009*** (0.003)
Nr. of railways and waterways crossed (log)	-0.013*** (0.004)	-0.010** (0.004)	-0.012** (0.003)
Income abs. difference	-0.018*** (0.001)	-0.018*** (0.002)	-0.018*** (0.002)
Racial similarity	0.023*** (0.003)	0.025*** (0.004)	0.022*** (0.003)
Distance (log)	-0.081*** (0.015)	-0.089*** (0.017)	-0.092*** (0.018)
User population (product log)	0.019** (0.008)	0.014* (0.008)	0.020** (0.009)
Metro fixed effect	Yes	Yes	Yes
Observations	2,669,688	1,400,386	2,159,802
R <sup>2</sup>	0.051	0.054	0.054

**Table SI14. Ordinary least squares regression models on the number of social connections between pairs of census tracts including spatial and socio-demographic features.** All the models include the metropolitan area as fixed effect. Model (1) is our final model in Table 1 of the main text and is used for comparison. Model (2) limits the sample behind the model to census tract pairs within the urban area of metropolitan areas, while model (3) limits the sample to census tract pairs that at least intersect with urban area inside metropolitan areas. Our main variable for the number of highways crossed does not change in the models based on limited samples. All variables indicated with (log) are transformed using  $\log_{10}(1 + x_{ij})$  to consider zero values. Standard errors are clustered at the metropolitan area level. \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$ .

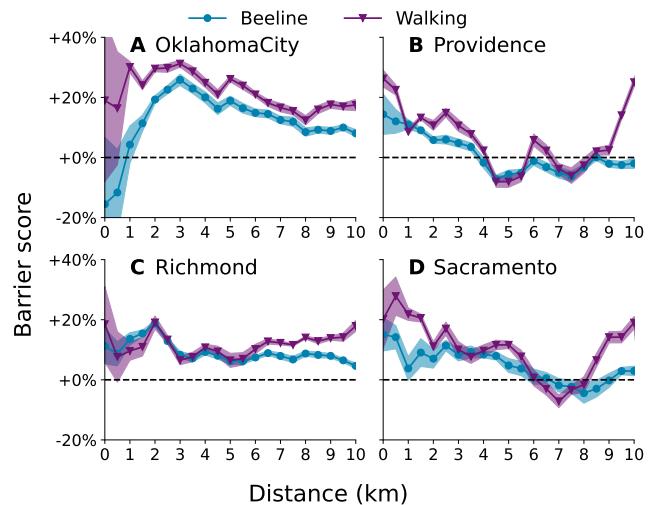
City	Users	Social links
Austin	2,050	17,756
Dallas	1,539	8,054
San Francisco	1,254	5,070
Los Angeles	1,068	2,362
Oklahoma City	481	5,084

**Table SI15. Number of geo-referenced users and social ties between them in the five most represented cities in the Gowalla dataset**

if they have at least 5 check-ins. We tested several more restrictive configurations, including time of day filters, but the results were the same for most users.

As a result of these filters, we were left with only five cities with a sufficient amount of data to extract reliable measurements (Tab. SI15). The Barrier Score for these cities are considerably higher than those obtained from the Twitter data for the same cities (Fig. SI24), which corroborates the validity of our results, and in addition suggests that the interplay between highways and social connections may be even more pronounced for stronger social ties.

**S. Beeline distance vs. walking distance.** In our spatial modeling, we conceptualize social ties as straight segments connecting nodes, with the distance between points calculated as the length of these segments in Euclidean space, a measure often referred to as ‘beeline’ distance. This approach facilitates an intuitive depiction of spatial distance and its efficient computation. To determine whether using straight segments is a good enough approximation, we recalculate the Barrier Scores using shortest walking distance rather than crow-fly (beeline) distance. This requires to recompute the null model, rewiring a social tie  $i, j$  to form a new tie  $i, k$  such that the



**Fig. SI25. Calculating Barrier Scores for beeline versus walking distance in 4 cities shows robust distance patterns.** Averages over 20 randomized runs of the null model are shown, together with 95% confidence intervals.

shortest walking distance is preserved ( $d_{ij}^{walk} = d_{ik}^{walk}$ ). Given the algorithm’s significant computational cost, we limited this analysis to a selection of 4 cities that *i*) have a relatively small social network, with less than 35k edges, and *ii*) feature natural topological barriers within the city boundaries (rivers and lakes). For each city, we executed 20 different random realizations of the null model. The results are presented in Figure SI25. The variation of the Barrier score with respect to walking distance follows a very similar pattern to the Barrier

Score calculated with crow-fly distance. Moreover, the Barrier Score is often higher when considering walking distance, suggesting that our estimates of the barrier effect of highways are conservative.

**T. Effect of  $d_{ij}$  deterrence on the Barrier Score.** The deterrence factor  $d_{ij}$  has a discounting effect that impacts mostly the component of the Barrier Score that is given by short-distance ties – namely, the values of  $B(d)$  for low values of  $d$ . This becomes obvious when comparing extreme values of  $d$ . For example, the ties that contribute to the calculation of the Barrier Score  $B(d)$  for the distance bin  $d = [10000m, 10500m]$  are all characterized by roughly the same length ( $\pm 5\%$ ), therefore, the number of highways these ties cross ( $c_{ij}$ ) will be discounted by approximately the same value of  $d_{ij}$ . Conversely, ties in the bin  $d = [0m, 500m]$  have a higher relative length variation, and the  $d_{ij}$  deterrence might impact quite differently two ties in the same bin. We included  $d_{ij}$  as a discounting factor precisely to even-out any intra-bin variations that could end up disproportionately representing the contribution of longer ties.

We tested empirically how  $B(d)$  changes when removing the  $d_{ij}$  deterrence from the formula, to exactly understand the consequences of the  $d_{ij}$  choice. Figure SI26 shows that: *i*) the impact is mostly limited to the short distance bins, *ii*) the magnitude of the impact is small, and *iii*) the length-discounted version of the score tends to *underestimate* the barrier effect compared to the non-discounted version, suggesting that our estimates of the barrier effect of highways are conservative.

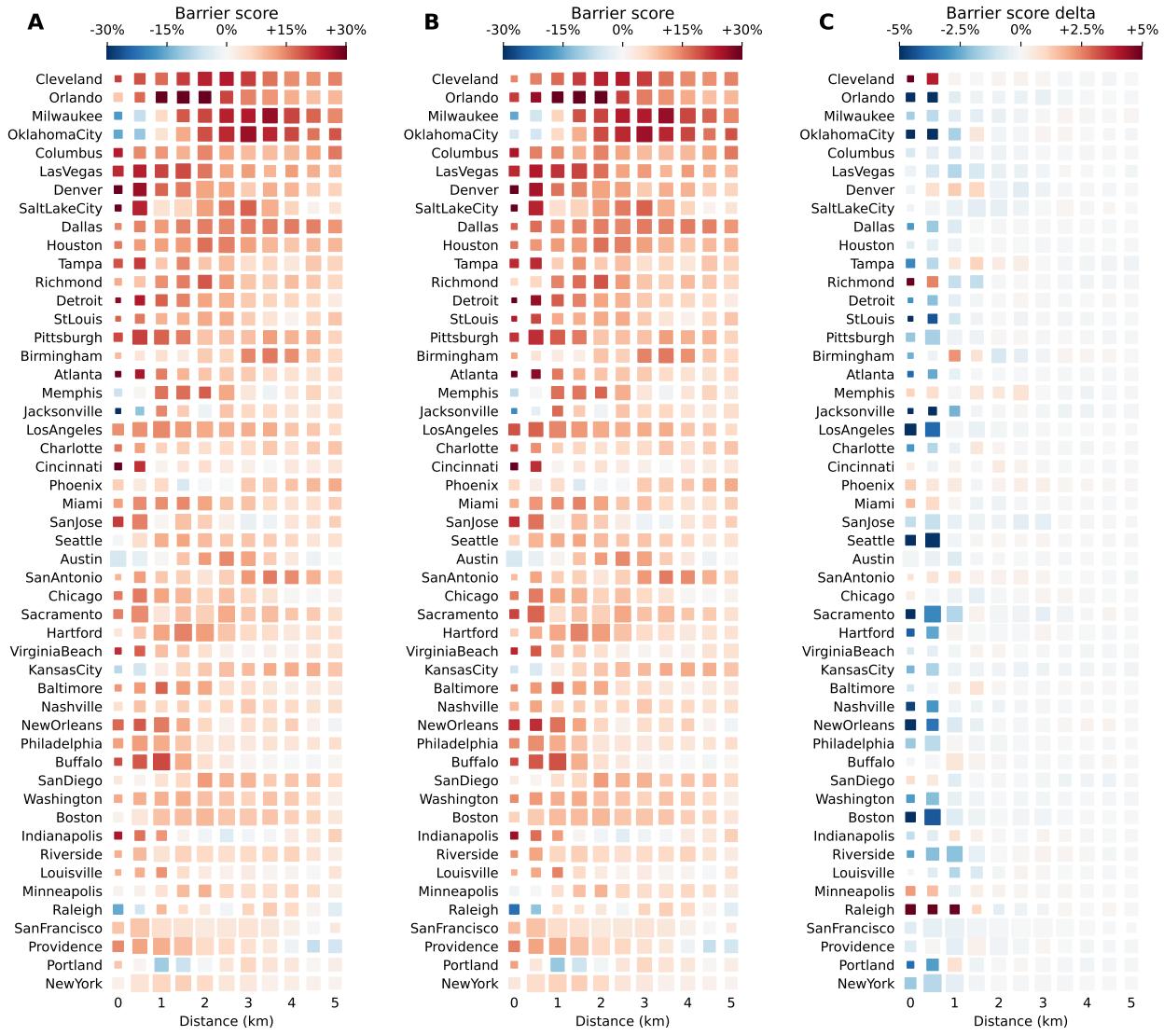
**U. The Color of Highways: The racial context of US highway construction.** Social segregation can happen along many different axes; one of the most obvious ones to scrutinize in contemporary US cities is race. Segregationist practices were ruled out by law only in the 1960s, with the Civil Rights Acts of 1964 and 1968, the former outlawing “*discrimination or segregation on the ground of race, color, religion, or national origin*” (41) and the latter explicitly expanding this principle to the provision (selling and renting) of housing (42). Prior to this, however, explicitly racist and exclusionary urban policies such as redlining and racial zoning were widely practiced (43), and their legacy is still lingering up until today (44). The Interstate Highway System (IHS) is a compelling example thereof (45, 46). From the beginning of its construction in 1956, the IHS fostered urban sprawl for decades to come and played a major role in the suburbanization of US cities, which in turn is racially biased (47–50). By the time that massive government-funded highway construction got underway throughout the country in the late 1950s, many US urban areas had majority Black inner cities. Many cities followed the advice of urban planner Robert Moses, not only on his proclaimed imperative that “most of our new (...) expressways (...) must go right through cities and not around them” (51), but also in his unambiguously racialized suggestion that “the practical solutions of the traffic problems in cities should be coordinated with slum clearance” (51), which meant building highways *through* Black neighborhoods. The choice of specific sites within a city where highways would lead through sparked a great number of protests across the country, known as Freeway Revolts (52). There are numerous, extensively studied examples of highway construction sites tearing apart or displacing historically Black communities (53–56), proverbially known as “*white roads through Black bedrooms*” (57).

Likewise, many decision-makers used such major construction projects as an instrument to get rid of so-called “ghettos” or to spatially exclude underprivileged communities (43, 58, 59). Effects of both practices are visible in today’s patterns of racial residential segregation. Below, we explore these effects and their historical context for our nine case study cities.

**Cleveland, OH** has a long history of racial exclusion (60). Ever since urban highway construction gained momentum in the late 1950s, local infrastructural decisions have often been heatedly disputed at the intersection of race, class, and suburbanization (61, 62). The two suburban neighborhoods of Shaker Heights and Cleveland Heights are well-known examples of these disputes. In the 1960s, when both neighborhoods were predominantly white and particularly wealthy, initial plans to construct a set of highways through these neighborhoods were successfully overturned. This sparked protests by local residents, supported by Carl Stokes, the first Black mayor of a major US city (63, 64). From 1976 onwards, traffic diverters were installed at the suburban fringe of Shaker Heights, dubbed by some as the “Berlin Wall” for Black people (45) and de facto keeping out a lower income minority; many inhabitants saw the traffic diversion measures as racially connotated (45, 62). As of today, both Shaker Heights and Cleveland Heights are regarded by some as a symbol of successful suburban racial integration (65). Overall, however, Cleveland is one of the poorest and most racially segregated among major US cities (60, 66, 67). At the same time, Cleveland ranks highest in our computations of the city-wide Barrier Score. Moreover, the Barrier Score is high for I-77, which separates the cluster of predominantly Black communities in the east from the rest of the metropolitan area (see Fig. 4A in the main text).

**Orlando, FL.** In 1961, the I-4 in Florida became one of the first highways supported by the Federal-Aid Highway Act of 1956 to officially celebrate its opening (68). Within Orlando, the I-4 was built parallel to Division Street (today Division Avenue), a “dividing line” (69) between white and Black neighborhoods. The I-4 further emphasized this racial division (69, 70). The neighborhood of Parramore, once a thriving Black community, suffered a particularly heavy impact from the construction, with 551 Black properties displaced. I-4 thus literally cemented an already existing racial separation, further acting as a “class barrier” (71) as it separated Parramore from downtown Orlando (72). The construction of the Expressway 408 further exacerbated this impact, isolating the Griffin Park public housing project from the rest of Parramore (71). In a 2006 report, the City of Orlando identified the 408 as a “development barrier” (73) for the neighborhood. As of today, Orlando still remains highly segregated, with the I-4 cutting through the city “like a picket line” (70). At the same time, both I-4 and Expressway 408 have high Barrier Scores in our results, as shown in Fig. 4B in the main text.

**Milwaukee, WI.** Next to Cleveland, Milwaukee is the second city which appears both in the list of 10 most segregated US cities (74) and in the top 3 of our Barrier Score ranking (Fig. 4C in the main text). In the 1960s, when the Black population of Milwaukee was still forcefully constrained to live in the North Side, the city was home to numerous protests against segregation. In 1967, for example, the Milwaukee’s NAACP



**Fig. S126. The impact of the deterrence  $d_{ij}$  on Barrier Scores.** **A.** Barrier Scores at distance  $d$ . **B.** Barrier Scores calculated without considering the distance deterrence  $d_{ij}$ . **C.** the absolute difference (delta) between the original Barrier Score and its non-discounted version – note the inflated color scale here for making the differences visible. Values of delta range between  $-5\%$  and  $+5\%$ , and they converge to zero after a distance of approximately 2 km. Negative values of delta indicate that the length-discounted version of the Barrier Score tends to underestimate the barrier effect compared to the non-discounted version. The average delta across cities for distances within 2 km is  $-0.012$ .

(National Association for the Advancement of Colored People) Youth Council marched from North to South across the 16th Street Bridge, which was jokingly called “the longest bridge in the world” for connecting Africa and Poland, given its location between the majority-Black North Side, and the (at that time) almost exclusively Polish Old South Side (75). The construction of I-43 in the 1960s severely impacted the North Side’s Bronzeville neighborhood, a formerly vibrant Black community. Many of Bronzeville’s inhabitants were displaced, and commercial areas were demolished (53, 76). I-43 also impacted the South Side, leading to housing shortage. However, while other communities within Milwaukee were dissipated as a consequence of highway construction, the Southside remained “solidly Polish” (77). In our computations for Milwaukee, we see that both the I-794 (separating the majority Black North from the majority white South), and the I-43 (the northern part of which disrupted many Black neighborhoods) have high

Barrier Scores, as illustrated in Fig. 4C in the main text.

**Oklahoma City, OK.** In contemporary Oklahoma City, the “city’s divided soul” (78) is most prominently expressed in the I-235, which separates majority Black neighborhoods in the East from majority white neighborhoods in the West. In contrast to previously mentioned cities, however, urban highway construction in Oklahoma City did not appear as immediate impact on previously thriving communities. For example, the I-235 (formerly known as the Centennial Expressway) was constructed in the late 1980s, creating a link between I-35, I-40, and I-44, and simultaneously isolating the largest historically Black neighborhood of Deep Deuce from the rest of the inner city (79, 80). However, the construction of I-235 was not so much a catalyst of Black neighborhood destruction, but rather the final blow to a neighborhood whose vitality and street life had already been eroded by several decades of car-centric plan-

ning and “urban renewal” (81). As civil right activist James Baldwin succinctly put it in an interview with Kenneth Clark in 1963, “Urban renewal (...) means Negro removal” (82). Under the auspices of OCURA (Oklahoma City Urban Renewal Authority), established in 1961, entire neighborhoods of Oklahoma City fell victim to major construction projects (83, 84). The University Medical Center urban renewal project, for example, displaced over 700 families, over 90% of which were Black (85). Lastly, in the recent two decades, Oklahoma City has witnessed a complex process of gentrification, most prominently underway in Deep Deuce (81, 86). The Barrier Scores we computed for Oklahoma City are particularly high in the inner city, for all highways mentioned above: I-35, I-40, I-44, and I-235 (see Fig. 4D in the main text).

**Detroit, MI** is known to be not only within the top 10 metropolitan areas by numbers of Black inhabitants (87), but also as one of the most racially segregated cities in the US (88). The city has a complex history in which the rise of the automobile, then deindustrialization, suburbanization and political marginalization are intertwined with a legacy of systemic and physical violence against Black people (58). The Birwood Wall (also known as Eight Mile Wall or Wailing Wall) is an infamous concrete symbol of Detroit’s predicament. It was built in 1941 by a private developer, with the aim of securing governmental funds for the construction of an all-white residential complex, which, as by the Federal Housing Administration’s requirements, had to be physically separated from the adjacent majority Black area redlined as “slum” (58, 89). As of today, the Birwood Wall is still standing, located next to the Eight Mile Road, which in turn is closely associated with racial segregation in popular culture, manifesting the divide between the Black inner city and the white suburbs (90). From the late 1940s onwards, land clearing for highway construction additionally exacerbated Detroit’s already ongoing housing crisis, displacing tens of thousands of residents, most of them Black, often on short notice and without proper assistance to find a new dwelling (58, 91). Numerous Black neighborhoods were destroyed by highway construction: the I-75 and I-375 practically erased Black Bottom, Hastings Street, Paradise Valley, and parts of the Lower East Side; while the formerly coherent urban fabric in the Western part of the city got bisected by I-94 and M10 (58, 59, 91, 92). For all highways mentioned above, the Barrier Scores in our computations are at their highest within the city center, in proximity of the historically Black (former) neighborhoods of Black Bottom and Paradise Valley, while Eight Mile Road stands out as a city limit delineation (marking the transition between counties) with a particularly high Barrier Score (see Fig. 4E in the main text).

**Austin, TX.** As of today, Austin is the most economically segregated large metropolitan area in the US (93). At the same time, as the Black Austin Coalition underlines, income inequality amongst Austin’s residents is strongly associated with race (94). The infamous Austin City Plan of 1928 suggested to “solve” the “race segregation problem” (the “problem” being *how* to segregate Black people within “constitutional” limits) by implementing a set of measures which would force Black people into moving to the East side of the city, with the East Avenue as officially proposed segregation line. The implementation of the City Plan had the foreseen consequences

of Black Austinites being forced to move to the East side of the city (94, 95). Thus, thirty years later, building the I-35 along East Avenue meant “solidifying the dividing line” (96) between two racially disparate parts of the city, notably with only a handful of crossings between East and West. Our Barrier Score computations also depict the I-35 as clear dividing line between two parts of the city (Fig. 4F in the main text). Over the following decades, a combination of systemic neglect, “urban renewal” (including the seizing of Black property and systematic funding of non-Black serving projects), and suburban sprawl fostered by highway presence severely impacted Black communities in East Austin (94, 97, 98). In the present time, gentrification is underway in East Austin, with real estate prices rising tenfold in the last 20 years. Simultaneously, Austin is currently the only large city in the US that is suffering a net decrease in Black inhabitants *in spite of* an overall rapidly growing population (99). As of 2024, plans are underway for the I-35 to be expanded from 16 to 22 lanes in Downtown (100).

**Columbus, OH** Several highways (I-70, I-71, I-670) bisect the urban fabric of Columbus, Ohio (see Fig. 4G in the main text). The alignment of these highways with former redlined areas is particularly startling (101). The chilling words of Warren Cremeen, an official of Ohio’s Department of Transport, illustrate the intentionality behind routing decisions: “[W]e married highway money and urban renewal money and wiped out (...) the worst slum in the state of Ohio” (quoted in an interview with Rose and Seely (1990) (102)). In several locations in the city, highway trajectories went directly through redlined neighbourhoods. Flytown, a diverse, but socioeconomically disadvantaged neighbourhood with many Black and Irish inhabitants, was erased from the map to make way for I-670 (103). Large parts of the Near East Side, which at the time of construction was predominantly Black, faced large-scale destruction by all three major highways in the city: I-70, I-71 and I-670. Two historically Black neighbourhoods in the Near East Side were particularly affected. I-70 was built right through Hanford Village, by that time a bustling Black community, destroying many homes and cutting off the Western part of the neighbourhood from the rest of it (103). Similarly, King-Lincoln/Bronzeville saw the construction of I-71 through its core, entailing the demolition of homes and businesses and severely hindering access to this Black community in combination with I-670 along its borders; at the same time, the predominantly white, affluent neighborhood Bexley, east of Bronzeville, was spared from the highways (104). On a larger scale, I-71 disconnected the predominantly Black Near East Side from Downtown.

**Richmond, VA** was the Capital of the Confederacy during the Civil War; in the following decades, the city became deeply segregated, with redlining practices further exacerbating racial divisions (101). The neighbourhood of Jackson Ward gained key importance for Richmond’s Black community, dubbed “Black Wall Street” and “Harlem of the South” prior to World War II. In the 1960s, however, the I-95 and the I-64/I-95 interchange were built directly through Jackson Ward, destroying numerous homes and businesses, and cutting off the neighbourhood’s northern part, Gilpin (see Fig. 4H in the main text). The public housing community of Gilpin Court was

thus cut off both from Jackson Ward and from the rest of Richmond, with only a few physical connections bridging across the highway. “Urban renewal” brought further displacement to Jackson Ward through large-scale construction projects such as the Coliseum (55). Population decline, urban neglect and continuously high poverty rates further impacted Richmond communities in the vicinity of the newly constructed highways, particularly north of I-95 (54, 105). Historically, Richmond’s segregationist policies in public housing further contributed to a “concentration of racialized poverty” (54). Today, Richmond’s inhabitants are confronted with particularly high eviction rates entangled with racialized dispossession (55), giving rise to initiatives like Residents of Public Housing in Richmond Against Mass Evictions (RePHRAME) (106). Most recently (as of 2023), the city of Richmond has obtained a grant from the US department of transportation, dedicated to “reconnecting” Jackson Ward with the rest of the city across highway division lines (107).

**Nashville, TN.** Nashville’s urban landscape is heavily fragmented by three major interstates: I-24, I-40, and I-65 (see Fig. 4I in the main text). The trajectory of I-40 displays a “kink in the road” (108) through North Nashville: When going from West to East, instead of following Charlotte Avenue, I-40 is routed one mile further north. The construction of I-40 began only in 1967; its exact route, however, had already been decided a decade prior, without duly involving the public. Initial plans with a more direct routing of I-40 had then been discarded in favor of introducing said “kink in the road”, which implied major disruptions for the majority Black North Nashville. Due to misleading information and legal process violations from the authorities’ side, the actual plans for I-40 and the destruction that it entailed only became clear to the residents of North Nashville as construction had already begun (108). In response, the Nashville I-40 Steering Committee was formed; their case *Nashville I-40 Steering Committee v. Ellington* won a temporary restraining order – the first time that highway construction had been “halted by claims of racial discrimination” (109). However, the case was ultimately lost in federal court; the I-40 ended up ripping through North Nashville as planned and “bulldozed local prosperity in the name of national economic development” (44). I-40 was constructed along and through Jefferson street, Nashville’s main Black business and cultural district, which became severely impacted and divided. 80% of Black-owned businesses were either directly demolished, or damaged through reduced accessibility for clients. The same highway also “cut in half a thriving academic cluster” (108) as it separated three major Black higher education institutions, Fisk University, Meharry Medical College, and Tennessee A. & I. University (later Tennessee State University), both from each other *and* from surrounding majority Black neighbourhoods. Real estate values dropped in the area and housing conditions quickly and severely deteriorated (108). Ultimately, the construction of I-40 and its consequences represent a decisive contribution to today’s rampant poverty rates in the area (110).

- Wang Q, Phillips NE, Small ML, Sampson RJ (2018) Urban mobility and neighborhood isolation in America’s 50 largest cities. *Proceedings of the National Academy of Sciences* 115(30):7735–7740.
- Morales AJ, Dong X, Bar-Yam Y, ‘Sandy’ Pentland A (2019) Segregation and polarization in urban areas. *Royal Society Open Science* 6(10):190573.
- Dong X, et al. (2020) Segregated interactions in urban and online space. *EPJ Data Science* 9(1):20.

- de la Prada AG, Small ML (2024) How people are exposed to neighborhoods racially different from their own. *Proceedings of the National Academy of Sciences* 121(28):e2401661121. Publisher: Proceedings of the National Academy of Sciences.
- Menyhért M, et al. (2024) Connectivity and Community Structure of Online and Register-based Social Networks. arXiv:2406.17752 [physics].
- Bokányi E, Juhász S, Karsai M, Lengyel B (2021) Universal patterns of long-distance commuting and social assortativity in cities. *Scientific Reports* 11(1):20829. Number: 1 Publisher: Nature Publishing Group.
- Kovács AJ, Juhász S, Bokányi E, Lengyel B (2022) Income-related spatial concentration of individual social capital in cities. *Environment and Planning B: Urban Analytics and City Science* p. 239980832211206.
- Saito K, Masuda N (2014) Two Types of Well Followed Users in the Followership Networks of Twitter. *PLOS ONE* 9(1):e84265. Publisher: Public Library of Science.
- Athey S, Ferguson B, Gentzkow M, Schmidt T (2021) Estimating experienced racial segregation in US cities using large-scale GPS data. *Proceedings of the National Academy of Sciences* 118(46):e2026160118.
- Moro E, Calaceti D, Dong X, Pentland A (2021) Mobility patterns are associated with experienced income segregation in large US cities. *Nature Communications* 12(1):4633. Bandiera\_abtest: a Cc\_license\_type: cc\_by Cg\_type: Nature Research Journals Number: 1 Primary\_atype: Research Publisher: Nature Publishing Group Subject\_term: Computational science;Social sciences Subject\_term\_id: computational-science;social-sciences.
- Mijs JJB, Roe EL (2021) Is America coming apart? Socioeconomic segregation in neighborhoods, schools, workplaces, and social networks, 1970–2020. *Sociology Compass* 15(6):e12884. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/soc4.12884>.
- Jackson MO (2021) Inequality’s Economic and Social Roots: The Role of Social Networks and Homophily.
- Leo Y, Fleury E, Alvarez-Hamelin JL, Sarraute C, Karsai M (2016) Socioeconomic correlations and stratification in social-communication networks. *Journal of The Royal Society Interface* 13(125):20160598. Publisher: Royal Society.
- Reme BA, Kotsadam A, Bjelland J, Sundsøy PR, Lind JT (2022) Quantifying social segregation in large-scale networks. *Scientific Reports* 12(1):6474. Publisher: Nature Publishing Group.
- Dobos L, et al. (2013) A multi-terabyte relational database for geo-tagged social network data in 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom). pp. 289–294.
- McNeill G, Bright J, Hale SA (2017) Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Science* 6(1):24.
- Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications* 387(21):5317–5325.
- {OpenStreetMap Contributors} (2024) OpenStreetMap.
- Boeing G (2017) OSRMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65:126–139.
- QGIS.org (2023) QGIS Geographic Information System.
- Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Physical Review E* 64(2):026118. Publisher: American Physical Society.
- Krings G, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment* 2009(07):L07003.
- Scellato S, Noulas A, Lambiotte R, Mascolo C (2011) Socio-Spatial Properties of Online Location-Based Social Networks. *Proceedings of the International AAAI Conference on Web and Social Media* 5(1):329–336. Number: 1.
- Silva JMCS, Tenreyro S (2006) The Log of Gravity. *The Review of Economics and Statistics* 88(4):641–658.
- Kowald M, et al. (2013) Distance patterns of personal networks in four countries: a comparative study. *Journal of Transport Geography* 31:236–248.
- Parady G, et al. (2021) A comparative study of social interaction frequencies among social network members in five countries. *Journal of Transport Geography* 90:102934.
- Fleischmann M, Romice O, Porta S (2021) Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science* 48(8):2133–2150.
- IUCN World Commission on Protected Areas (WCPA), Connectivity Conservation Specialist Group, Ament R, Clevenger A, Van Der Ree R, Center for Large Landscape Conservation (2023) *Addressing ecological connectivity in the development of roads, railways and canals*. (Center for Large Landscape Conservation) Vol. 5.
- Roy S (2023) Types and Development of Transportation Infrastructure in *Disturbing Geomorphology by Transportation Infrastructure: Problem, Prospect, and Solution*, ed. Roy S. (Springer International Publishing, Cham), pp. 19–46.
- Rodrigue JP (2024) *The Geography of Transport Systems*. (Taylor & Francis). Google-Books-ID: UVYIEQAAQBAJ.
- Mitchell R, Lee D (2014) Is There Really a “Wrong Side of the Tracks” in Urban Areas and Does It Matter for Spatial Analysis? *Annals of the Association of American Geographers* 104(3):432–443. Publisher: Routledge \_eprint: <https://doi.org/10.1080/00045608.2014.892321>.
- Kramer R (2018) Testing the role of barriers in shaping segregation profiles: The importance of visualizing the local neighborhood. *Environment and Planning B: Urban Analytics and City Science* 45(6):1106–1121. Publisher: SAGE Publications Ltd STM.
- Higgsman M, Stockton J, Anciaes P, Scholes S, Mindell JS (2022) Community severance and health – A novel approach to measuring community severance and examining its impact on the health of adults in Great Britain. *Journal of Transport & Health* 25:101368.
- to Trails Conservancy R (2024) Rails to Trails Conservancy: Building A Nation Connected By Trails.
- Liao Y, Gil J, Yeh S, Pereira RHM, Alessandretti L (2024) Socio-spatial segregation and human mobility: A review of empirical evidence. arXiv:2403.06641 [cs].
- Liao Y, Gil J, Yeh S, Pereira RHM, Alessandretti L (2024) The Uneven Impact of Mobility on

- the Segregation of Native and Foreign-born Individuals. arXiv:2407.00404 [cs].
37. Fienberg SE (1970) An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics* 41(3):907–917.
  38. Bishop YM, Fienberg SE, Holland PW (2007) *Discrete Multivariate Analysis: Theory and Practice*. (Springer Science & Business Media, New York).
  39. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11. (Association for Computing Machinery, New York, NY, USA), pp. 1082–1090.
  40. Uber (2023) h3-py: Uber's H3 Hexagonal Hierarchical Geospatial Indexing System in Python. original-date: 2018-06-12T22:39:59Z.
  41. of the United States Government GR (1964) Civil Rights Act of 1964.
  42. of the United States Government GR (1968) Civil Rights Act of 1968.
  43. Rothstein R (2017) *The Color of Law: A Forgotten History of How Our Government Segregated America*. (Liveright Publishing). Google-Books-ID: SdtDDQAAQBAJ.
  44. Arcadi T (2023) Concrete Leviathan: The Interstate Highway System and Infrastructural Inequality in the Age of Liberalism. *Law and History Review* 41(1):145–169. Publisher: Cambridge University Press.
  45. Schindler SB (2014) Architectural Exclusion: Discrimination and Segregation through Physical Design of the Built Environment. *Yale Law Journal* 124:1934.
  46. Karas D (2015) Highway to Inequity: The Disparate Impact of the Interstate Highway System on Poor and Minority Communities in American Cities. *New Visions for Public Affairs* 7.
  47. Rabin Y (1973) Highways as a Barrier to Equal Access. *The ANNALS of the American Academy of Political and Social Science* 407(1):63–77. Publisher: SAGE Publications Inc.
  48. Boustan LP (2010) Was Postwar Suburbanization “White Flight”? Evidence from the Black Migration. *Quarterly Journal of Economics* 125(1):417–443. Publisher: Oxford University Press / USA.
  49. Massey DS, Tannen J (2018) Suburbanization and segregation in the United States: 1970–2010. *Ethnic and Racial Studies* 41(9):1594–1611. Publisher: Routledge \_eprint: <https://doi.org/10.1080/01419870.2017.1312010>.
  50. Hadden Loh T, Coes C, Buthe B (2020) Separate and unequal: Persistent residential segregation is sustaining racial and economic injustice in the U.S. *Brookings*.
  51. Moses R (1954) Statement for the President's Advisory Committee on a National Highway Program, (President's Advisory Committee on a National Highway Program (Clay Committee), Washington, DC), Technical report.
  52. Mohl RA (2004) Stop the Road: Freeway Revolts in American Cities. *Journal of Urban History* 30(5):674–706.
  53. House PA (1970) Relocation of Families Displaced by Expressway Development: Milwaukee Case Study. *Land Economics* 46(1):75.
  54. Howard AL, Williamson T (2016) Reframing public housing in Richmond, Virginia: Segregation, resident resistance and the future of redevelopment. *Cities* 57:33–39.
  55. Howell K, Teresa BF (2020) Displacement, Demobilization, and Democracy: Current Eviction and Historic Dispossession in Richmond, Virginia. *Metropolitics*. Publisher: Metropolis.
  56. Mahajan A (2023) Highways and segregation. *Journal of Urban Economics* p. 103574.
  57. Drummond Ayres BJ (1967) White Roads Through Black Bedrooms. *The New York Times*.
  58. Sugrue TJ (2014) *The origins of the urban crisis: Race and inequality in postwar detroit-updated edition*. (Princeton University Press) Vol. 168.
  59. Miller J (2018) Roads to nowhere: how infrastructure built on American inequality. *The Guardian*.
  60. Yankey O, Lee J, Gardenhire R, Borawski E (2023) Neighborhood Racial Segregation Predict the Spatial Distribution of Supermarkets and Grocery Stores Better than Socioeconomic Factors in Cleveland, Ohio: a Bayesian Spatial Approach. *Journal of Racial and Ethnic Health Disparities*.
  61. Mohl RA (2001) Urban Expressways and the Racial Restructuring of Postwar American Cities. *Economic History Yearbook* 42(2):89–104.
  62. Sisley L (2017) Shaker Barricades - Reinforcing Suburban Separation.
  63. Dawson VP (2023) Saving the Shaker Lakes: How an Alliance between Two Wealthy Suburbs and Cleveland's Black Mayor Stopped the Clark Freeway. *Journal of Planning History* 22(3):241–262.
  64. Doll L (2024) Cleveland's Forgotten Freeways.
  65. Souther JM (2023) Through the Ivory Curtain: African Americans in Cleveland Heights, Ohio, before the Fair Housing Movement. *Journal of Urban History* 49(6):1312–1341.
  66. Berube A, Shah I, Friedhoff A, Shearer C (2019) Metro Monitor 2019: Inclusion remains elusive amid widespread metro growth and rising prosperity.
  67. Menendian S, Gambhir S, Gailes A (2021) The Roots of Structural Racism Project. Twenty-First Century Racial Residential Segregation in the United States.
  68. Administration FH (2017) The Greatest Decade 1956–1966.
  69. Brotemarkle BD (2006) *Crossing Division Street: An Oral History of the African-American Community in Orlando*. (Florida Historical Society Press). Google-Books-ID: EZgaAQAAIAAJ.
  70. Wolf C (2018) New map reminds us that Orlando remains incredibly segregated. *Orlando Weekly*.
  71. Gama YK (2015) Master's thesis (Rollins College, Winter Park, FL).
  72. Sherman N (2022) From Connections To Barriers: How The Interstate Highway System Both Connected And Segregated Communities. *Creating Knowledge. The LAS Journal of Undergraduate Scholarship* 15.
  73. of Orlando C (2006) Downtown Orlando Community Venues Masterplan. Part B: The Masterplan, Technical report.
  74. Institute O/B (2019) Most to Least Segregated Cities.
  75. Society WH (2024) Crossing the Line: The Milwaukee Fair Housing Marches of 1967–1968.
  76. McBride GG, Byers SR (2007) The First Mayor of Black Milwaukee: J. Anthony Josey. *The Wisconsin Magazine of History* 91(2):2–15.
  77. Lackey JF, Petrie R (2013) *Milwaukee's Old South Side*. (Arcadia Publishing). Google-Books-ID: epX9rg4qSScC.
  78. Felder B (2014) City's divided soul seen from 23rd Street. *Oklahoma Gazette*.
  79. Welge WD (2007) *Oklahoma City Rediscovered*. (Arcadia Publishing). Google-Books-ID: DiYBSFqmTSYC.
  80. Lippe S (2023) Project Report (Southern New Hampshire University, Manchester, NH).
  81. Payne AA, Greiner AL (2019) New-Build Development and the Gentrification of Oklahoma City's Deep Deuce Neighborhood. *Geographical Review* 109(1):108–130. Publisher: Routledge \_eprint: <https://doi.org/10.1111/gere.12294>.
  82. Baldwin (1963) Conversation With James Baldwin, A; James Baldwin Interview.
  83. Authority OCUR (1963) Annual Report 1963, (Oklahoma City Urban Renewal Authority, Oklahoma City), Technical report.
  84. Lackmeyer S, Kiewer A (2021) Historically Black neighborhood in Oklahoma City finds no relief from decades of explosions. *The Oklahoman*.
  85. Lab DS (2024) Family Displacements through Urban Renewal, 1950–1966 in *Renewing Inequality*. (Robert K. Nelson and Edward L. Ayers).
  86. Tierney S, Petty C (2015) Gentrification in the American heartland? Evidence from Oklahoma City. *Urban Geography* 36(3):439–456. Publisher: Routledge \_eprint: <https://doi.org/10.1080/02723638.2014.977038>.
  87. Moslimani M, Tamir C, Budiman A, Noe-Bustamante L, Mora L (2024) Facts About the U.S. Black Population.
  88. Institute O/B (2024) City Snapshot: Detroit.
  89. Einhorn E, Lewis O (2021) A segregation wall has stood in Detroit for 80 years. We found out who built it. *NBC News*.
  90. Newsroom MR (2014) 8 Mile Road is eight miles from where? *Michigan Public*.
  91. Whitaker D (2023) Understanding the Impact of I-375 Construction.
  92. Da Via C (2012) A Brief History Of Detroit's Black Bottom Neighborhood.
  93. Florida R, Mellander C (2015) *Segregated city : The geography of economic segregation in America&apos;s metros*. (Martin Prosperity Institute).
  94. Woods CT, Coalition BA, Styles M, Blanc P, Torosian C (2012) Black Austin Coalition: Building An Embassy. Town Hall Evaluation, (Black Austin Coalition, Austin, TX), Technical report.
  95. Skop E (2010) Austin: A city divided in *The African Diaspora in the United States and Canada at the Dawn of the 21st century*. (Academic Publishing, New York), pp. 109–122.
  96. Austin R (2023) History of the I-35 Corridor.
  97. Bernier N, McGlinchy A (2023) Highway to sprawl: How I-35 shapes where people live in Austin. Section: Transportation.
  98. Charpentier M, McGlinchy A (2023) Two paragraphs forced Black residents to East Austin. Exploding real estate prices forced them out. Section: Austin.
  99. Buchele M (2016) Austin's Population Is Booming. Why Is Its African-American Population Shrinking? *KUT Radio, Austin's NPR Station*.
  100. Bernier N (2024) Your ultimate guide to the I-35 expansion through Central Austin. *KUT Radio, Austin's NPR Station*.
  101. Nelson RK, Winling L, Ayers EL (2023) Mapping Inequality: Redlining in New Deal America.
  102. Rose MH, Seely BE (1990) Getting the Interstate System Built: Road Engineers and the Implementation of Public Policy, 1955–1985. *Journal of Policy History* 1(1):23–55.
  103. Smith T, Lentz E (2014) African-American Settlements and Communities in Columbus, Ohio, (Columbus Landmarks Foundation, Columbus, Ohio), Technical report.
  104. Thompson E (2020) How Black neighborhoods in Columbus were destroyed in '60s by highways. *The Columbus Dispatch*.
  105. (RRHA) RRHaHA (2023) Jackson Ward Community Plan, (Richmond, VA), Technical report.
  106. RePHRAME (2014) Residents Of Public Housing In Richmond Against Mass Eviction.
  107. Ward RJ (2023) Reconnect Jackson Ward Feasibility Study.
  108. Haynes C (2020) One Mile North. *Belmont Law Review* 8:1.
  109. Mohl RA (2002) The Interstates and the Cities: Highways, Housing, and the Freeway Revolt, (Poverty & Race Research Action Council, Washington, DC), Technical report.
  110. Perry AM, Barr A (2021) To restore North Nashville's Black middle class, local policymakers should pursue reparations, (Brookings Institution, Washington, DC), Technical report.