A Multi-Scale Analysis of 27,000 Urban Street Networks

Geoff Boeing[1]

Abstract

This paper serves illustrative purposes as it presents wide-ranging preliminary empirical findings on US urban form and street network characteristics, using OSMnx. It also demonstrates the scalability and flexibility of OSMnx as a new research tool. This paper addresses past limitations by conducting an analysis of street networks with large sample sizes, with clearly defined network definitions and extents for reproducibility, and using non-planar, directed graphs. In particular, it examines urban street networks focusing on structure, connectedness, centrality, and robustness. In total it cross-sectionally analyzes 497 urbanized areas' street networks, 19,655 cities' and towns' street networks, and 6,857 neighborhoods' street networks. These sample sizes are larger than those in any previous study and the preliminary empirical findings illustrate the use of OSMnx as a new research platform.

Introduction

On May 20, 1862, Abraham Lincoln signed the Homestead Act into law, making land across the United States Midwest and Great Plains available for free to all applicants (9). Under its auspices over the next 70 years, the federal government distributed 270 million acres of public land (10% of the entire US landmass) to private owners in the form of 1.6 million homesteads (7; 11). Towns with gridiron street networks sprang up rapidly across the Great Plains and Midwest, due to both the prevailing urban design paradigm of the day and the standardized rectilinear town plats used repeatedly to lay out instant new cities (12). Through path dependence, the spatial signatures of these land use laws, design paradigms, and planning instruments can still be seen today in these cities' urban forms and street networks.

A cross-sectional analysis of American urban form through its street networks at metropolitan, municipal, and neighborhood scales can reveal these artifacts and histories across the nation. Network analysis is a natural approach to the study of cities as growing, complex systems that self-organize to fill space with a capillary circulation network (8). The empirical literature on street network analysis is growing ever richer, but suffers from some limitations— discussed in (6) and summarized here. First, sample sizes tend to be fairly small due to data availability, gathering, and processing constraints. Second, reproducibility is difficult when the dozens of decisions that go into analysis—such as spatial extents, topological simplification and correction, definitions of nodes and edges, etc.—are ad hoc or only partly reported. Third, and related to the first two, studies frequently oversimplify to planar or undirected primal graphs for tractability, or use dual graphs despite the loss of geographic and metric information.

This paper addresses these limitations by conducting an analysis of street networks at multiple scales, with large sample sizes, with clearly defined network definitions and extents for reproducibility, and using non-planar, directed graphs. In particular, it examines urban street networks— represented here as primal, non-planar, weighted multi-digraphs with possible self-loops—focusing on structure, connectedness, centrality, and resilience. Most studies in the street network literature that conduct topological and/or metric analysis tend to have sample sizes ranging around 5 to 50 networks. This paper instead conducts a large analysis of over 27,000 urban street networks at multiple overlapping scales across the United States. Namely, it examines the street networks of every US incorporated city and town, urbanized area, and Zillow-defined neighborhood. In total, this study presented uses OSMnx (6) to download, construct, and analyze 497 urbanized areas' street networks, 19,655 cities' and towns' street networks, and 6,857 neighborhoods' street networks. It uses these street networks to conduct four analyses: at the metropolitan scale, at the municipal scale, at the neighborhood scale, and with a case study looking deeper at the neighborhood-scale street networks in the city of San Francisco.

This paper is organized as follows. In the next section, it presents the data sources, tools, and methods used to collect, construct, and analyze these street networks. Then it presents findings of the analyses at three spatial scales: metropolitan, municipal, and neighborhood. Finally, it concludes with a discussion of these findings and their implications for street network analysis, urban form complexity, and city planning.

[1]PhD Candidate, Department of City and Regional Planning, University of California, Berkeley

Corresponding author:
Geoff Boeing, Department of City and Regional Planning, University of California, Berkeley, United States
Email: gboeing@berkeley.edu

Methodology

This study uses OSMnx to download, construct, correct, analyze, and visualize street networks at metropolitan, municipal, and neighborhood scales (6). To define the study sites and their spatial boundaries, this study uses three sets of geometries. The first defines the metropolitan-scale study sites using the 2016 national TIGER/Line shapefile of US census bureau urban areas, released 19 August 2016. Census-defined urban areas comprise a set of census tracts that meet a minimum density threshold (14). In particular, we retain only the urbanized areas subset in this data set (i.e., areas with greater than 50,000 population), discarding the smaller urban clusters subset. The second set of geometries defines the municipal-scale study sites using 51 TIGER/Line shapefiles (again, 2016) of US census bureau places for all 50 states plus the District of Columbia. In particular, we discard the subset of census-designated places in this data set, instead retaining all the incorporated cities and towns in the United States.

The third and final set of geometries defines the neighborhood-scale study sites using 42 Zillow neighborhood shapefiles. These shapefiles contain the geometries of neighborhoods in 41 states plus the District of Columbia. This is a fairly new data set comprising nearly 7,000 neighborhoods in large American cities, but as Schernthanner et al. (10) point out, Zillow does not publish the methodology it uses to construct these geometries. However, despite its newness it already has a track record in the academic literature. For instance, Besbris et al. (5) used Zillow boundaries to examine neighborhood stigma, and Albrecht and Abramovitz (1) used these data to study neighborhood-level poverty in New York.

For all of these geometries, we use OSMnx to download the (drivable, public) street network contained within the geometry. First it buffers each geometry by 0.5 km, then downloads the street network within this buffered geometry. Next it constructs a street network from this data, corrects the topology, calculates street counts per node (this ensures that intersections are not considered cul-de-sacs simply because an incident edge connects to a node outside the desired polygon), then truncates the network to the original, desired polygon. OSMnx saves each of these networks to disk as GraphML and shapefiles. Finally, it automatically calculates metric and topological measures for each. However, in a planar graph, topological and metric properties are somewhat interrelated (8).

Average street length, the mean edge length in the undirected representation of the graph, serves as a linear proxy for block size and indicates how fine-grained or coarse-grained the network is. Node density is the number of nodes divided by the area covered by the network. Intersection density is the node density of the set of nodes with more than one street emanating from them (thus excluding dead-ends). The edge density is the sum of all edge lengths divided by the area, and the physical street density is the sum of all edges in the undirected representation of the graph divided by the area. These density measures all provide further indication of how fine-grained the network is. Finally, the average circuity represents the average ratio between an edge length and the straight-line distance between the two nodes it links.

The average node degree, or mean number of edges incident to each node, quantifies how well the nodes are connected on average. Similarly, but more concretely, the average streets per node measures the mean number of physical streets (i.e., edges in the undirected representation of the graph) that emanate from each intersection and dead-end. This adapts the average node degree for physical form rather than directed circulation. The statistical and spatial distributions of number of streets per intersection characterize the type, prevalence, and dispersion of intersection connectedness in the network. Connectivity measures the minimum number of nodes or edges that must be removed from a connected graph to disconnect it. The average node connectivity of a network—the mean number of internally node-disjoint paths between each pair of nodes in the graph—represents the expected number of nodes that must be removed to disconnect a randomly selected pair of non-adjacent nodes (4).

The clustering coefficient of a node is the ratio of the number of edges between its neighbors to the maximum possible number of edges that could exist between these neighbors. The weighted clustering coefficient weights this ratio by edge length and the average clustering coefficient is the mean of the clustering coefficients of all the nodes in the network. Centrality indicates the importance of nodes in a network. Betweenness centrality evaluates the number of shortest paths that pass through each node (2). The maximum betweenness centrality in a network specifies the proportion of shortest paths that pass through the most important node. This is an indicator of resilience: networks with a high maximum betweenness centrality are more prone to failure or inefficiency should this single choke point fail. Closeness centrality ranks nodes as more central if they are on average closer to all other nodes. Finally, PageRank ranks nodes based on the structure of incoming links and the rank of the source node.

In total, this study cross-sectionally analyzes 27,009 networks: 497 urbanized areas' street networks, 19,655 cities' and towns' street networks, and 6,857 neighborhoods' street networks. These sample sizes are larger than those in any previous study. In the following sections, we present the findings of these analyses at the metropolitan scale, the municipal scale, the neighborhood scale, and through a case study looking deeper at the neighborhood-scale street networks in the city of San Francisco.

Analysis of Metropolitan-Scale Street Networks

Table 1 shows the considerable variation in street network characteristics across the entire data set of 497 urbanized areas. This is unsurprising: urbanized areas span a wide spectrum of sizes, from the Delano, CA Urbanized Area's 26 km^2 to the New York–Newark, NY–NJ–CT Urbanized Area's 8,937 km^2—thus, density and count-based measures demonstrate substantial variance. Further, these urbanized areas span a wide spectrum of terrains, development eras and paradigms, and cultures.

Table 1. Central tendency and statistical dispersion for selected measures of all US urbanized areas' street networks.

measure	mean	std dev	min	median	max
Area (km2)	460.657	858.125	25.685	184.898	8937.429
Avg of the avg neighborhood degree	2.886	0.109	2.626	2.875	3.228
Avg of the avg weighted neighborhood degree	0.032	0.018	0.021	0.03	0.321
Avg circuity	1.076	0.019	1.023	1.074	1.14
Avg clustering coefficient	0.042	0.009	0.015	0.042	0.071
Avg weighted clustering coefficient	0.002	0.001	¡0.001	0.001	0.006
Intersection count	12582	26054	751	4593	307848
Avg degree centrality	0.001	0.001	¡0.001	0.001	0.007
Edge density (km/km2)	13.455	2.137	7.961	13.352	21.233
Avg edge length (m)	158.588	17.653	117.341	157.332	223.08
Total edge length (km)	6353	12625	427	2393	1.42E+08
Proportion of dead-ends	0.213	0.055	0.077	0.207	0.416
Proportion of 3-way intersections	0.593	0.046	0.444	0.591	0.778
Proportion of 4-way intersections	0.187	0.063	0.054	0.178	0.422
Intersection density (per km2)	26.469	6.256	12.469	26.029	49.423
Average node degree	5.153	0.302	4.307	5.143	6.056
m	40890	83678	2516	14955	981646
n	16032	32585	874	5830	373309
Node density (per km2)	33.628	7.641	17.675	33.071	61.655
Max PageRank value	0.001	0.001	¡0.001	0.001	0.003
Min PageRank value	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001
Self-loop proportion	0.008	0.008	¡0.001	0.006	0.071
Street density (km/km2)	7.262	1.221	4.217	7.171	11.797
Average street segment length (m)	161.331	17.765	119.573	160.288	225.92
Total street length (km)	3480	7026	222	1269	79046
Street segment count	22011	45725	1281	7868	533757
Average streets per node	2.764	0.162	2.223	2.77	3.217
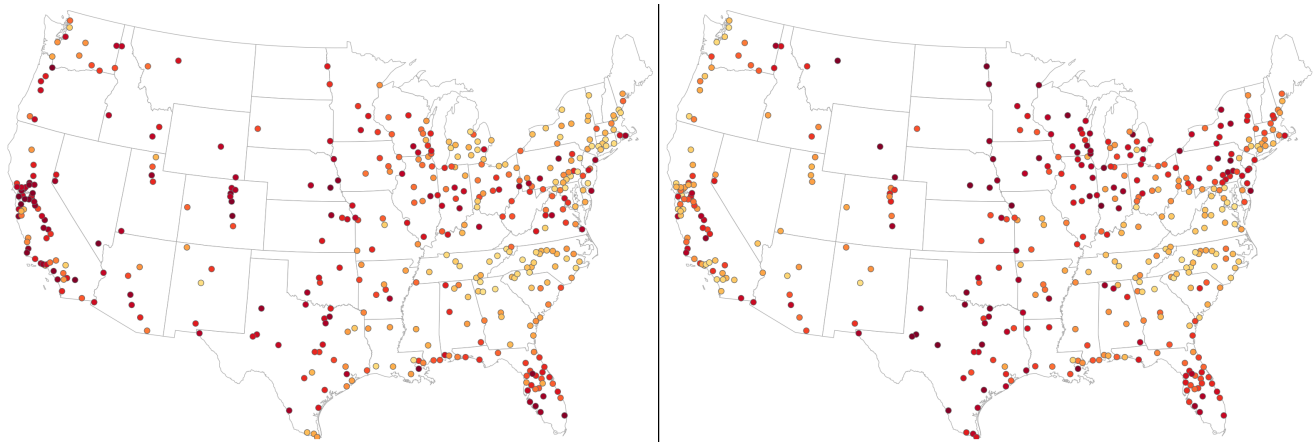
Figure 1. Intersection density per urbanized area (left) and average streets per node per urbanized area (right) from lowest (pale yellow) to highest (dark red), in the contiguous US

Nevertheless, looking across the data set provides a sense of the breadth of American metropolitan street networks. New York's urbanized area—America's largest--has 373,309 intersections and 79 million meters of linear street (or 417,570 and 83.4 million if including service roads). Delano, CA's urbanized area—America's smallest—has 874 intersections and 222,328 meters of linear street (or 964 and 231,000 meters if including service roads). The typical (Table 1, *median*) American urbanized area is approximately 185 km^2 in land area, has 5,830 intersections, and 1.3 million linear meters of street. Its street network is about 7.4% more circuitous than straight-line, as-the-crow-flies edges between nodes would be. The most circuitous network is 14% more circuitous than straight-line would be, and least is only 2%. Looking at network complexity in terms of density and connectivity, in the typical urbanized area, the average street segment length (a proxy for block size) is 160 meters. The longest average street segment is the 226-meter average of urbanized Danbury, CT. Puerto Rican cities hold the top four positions for shortest average street segment length, but among the 50 states plus DC, the shortest average street segment is the 125.3-meter average of urbanized Tracy, CA, indicating a much finer street network. The urbanized area of Portland, Oregon, with its famously compact walkable blocks, ranks second at 125.5 meters on average.
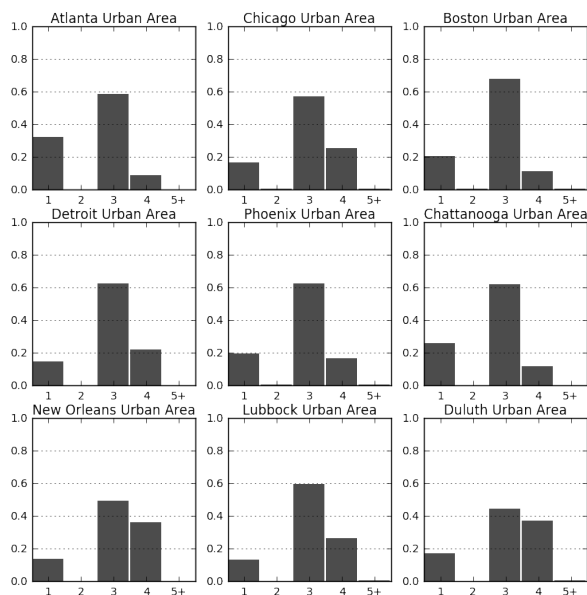
Figure 2. Distribution of node types in 9 urbanized areas, with number of streets emanating from the node on the *x*-axis and proportion of nodes of this type on the *y*-axis

The typical urbanized area has 26 intersections per km^2. Both the densest and the sparsest are in the deep south: the sparsest is 12.5 (Gainesville, GA urbanized area) and the densest is 49.4 (New Orleans urbanized area). However, New Orleans is an anomaly in the deep south. Figure 1 depicts the intersection density in each American urbanized area, from lowest density in dark red to highest density in light yellow. The map makes clear how the highest intersection densities are concentrated to the west of the Mississippi River. The lowest intersection densities are concentrated in a belt running from Louisiana, up through the Carolinas and Appalachians, and into New England. In general, only the largest cities on the east coast (e.g., Boston, New York, Philadelphia, Washington) and Florida escape this trend.

The distribution of node types (i.e., intersections and dead ends) provides a clear indicator of network connectedness. The typical urbanized area has 2.8 streets per intersection on average: lots of 3-way intersections, fewer cul-de-sacs, and even fewer 4-way intersections. The grid-like San Angelo, TX urbanized area has the most streets per node (3.2) on average, and (outside of Puerto Rico, which contains the seven lowest urbanized areas) the sprawling, disconnected Lexington Park, MD urbanized area has the fewest (2.2). These two urban areas fit the trend seen in the spatial distribution across the US in Figure 1: urbanized areas in the great plains and Midwest have particularly high numbers of streets per node on average, indicating more grid-like, connected networks. Cities in the southern and western US tend to have fewer streets per node, reflecting more dead-ends and a disconnected network. This finding is discussed in more detail in the upcoming section.

In the typical urbanized area, 18% of nodes are 4-way intersections, 59% are 3-way intersections, and 21% are dead-ends. However, this distribution varies somewhat between urbanized areas. Examining a small sample of 9 urbanized areas, chosen to maximize variance, reveals this in clearer detail. In Figure 2, urban Atlanta and

Chattanooga have very high proportions of dead-ends, each over 30% of all nodes, and very few 4-way intersections, indicating a disconnected street pattern. The urban areas of Phoenix, Boston, Detroit, and Chattanooga have particularly high proportions of 3-way intersections, each over 60%, indicating a prevalence of T-intersections. Conversely, Chicago, New Orleans, Duluth, and Lubbock have high proportions of 4-way intersections, indicating more grid-like connected networks. But what is perhaps most notable about Figure 2 is that these nine urbanized areas, despite being chosen to maximize variance, are overwhelmingly similar to each other. At the metropolitan scale, every large American urban agglomeration is characterized by a preponderance of 3-way intersections.

The relationship between fine-grained networks and connectedness/grid-ness is, however, not clear-cut. Intersection density has only a weak, positive linear relationship with the proportion of 4-way intersections in the urbanized area ($r^2 = 0.17$). But the relationship between network circuity and grid-ness is somewhat clearer. Average circuity has a negative linear relationship with the proportion of 4-way intersections in the urbanized area ($r^2 = 0.43$).

Due to the substantial variation in urbanized area size, from 25 to 9,000 km^2, the preceding analysis covers a wide swath of metropolitan place types. To better compare apples-to-apples, Table 2 focuses on the 30 largest urban areas cross-sectionally to examine how their metric and topological measures compare. This provides more consistent spatial scales and extents, while offering a window into the similarities and differences in the built forms of America's largest urban agglomerations.

Among these urban areas, Milwaukee has the least circuitous network (6% more circuitous than straight-line edges would be), and Orlando has the most (12%). San Juan and Atlanta have the fewest streets per node on average (2.36 and 2.45, respectively), while Milwaukee has the most (3.03). Cincinnati has both the lowest intersection density (18/km^2) and street density (6.1 km/km^2) while Denver has the highest intersection density (40.6/km^2) and Miami and Los Angeles have the highest street density (10.6 km/km^2, apiece). In other words, Cincinnati has a particularly coarse-grained network with few connections and paths. This can also be seen in the average street segment length, a proxy for block size: Cincinnati has the second highest (186 m), bested only by Cleveland (198 m). In contrast, the two lowest are Denver's 138-meter average and San Juan's 131-meter average.

These metropolitan-scale analyses consider trends in the built form at the scale of broad self-organized human systems and urbanized regions. However, they aggregate multiple heterogenous neighborhoods and municipalities—the scales of human life, urban design projects, and planning jurisdiction—into single units of analysis. To disaggregate and analyze finer characteristics, the following sections examine municipal- and neighborhood-scale street networks across the United States.

Analysis of Municipal-Scale Street Networks

Table 3 shows similarly great variation in street network characteristics across the entire data set of 19,655 cities and

Table 2. Selected measures of the 30 largest (by land area) urbanized areas' street networks.

urban area core city	land area km2	avg circuity	avg clustering coefficient	dead-end ratio	3-way ratio	4-way ratio	inters
New York	8937	1.06	0.04	0.18	0.62	0.2	34.44
Atlanta	6850	1.1	0.04	0.32	0.58	0.09	18.39
Chicago	6325	1.07	0.04	0.17	0.57	0.25	27.05
Philadelphia	5132	1.08	0.05	0.17	0.63	0.2	26.65
Boston	4852	1.09	0.05	0.2	0.68	0.11	24.23
Dallas	4612	1.07	0.05	0.15	0.61	0.23	34.16
Los Angeles	4497	1.06	0.03	0.21	0.56	0.22	39.45
Houston	4303	1.08	0.04	0.2	0.57	0.22	33.49
Detroit	3461	1.07	0.04	0.15	0.63	0.22	31.1
Washington	3424	1.09	0.04	0.26	0.56	0.17	31.22
Miami	3204	1.1	0.05	0.17	0.59	0.23	40.54
Phoenix	2968	1.09	0.05	0.2	0.62	0.17	35.31
Minneapolis	2647	1.08	0.05	0.19	0.57	0.23	29.54
Seattle	2617	1.07	0.03	0.3	0.54	0.16	31.57
Tampa	2479	1.1	0.05	0.2	0.58	0.21	31.35
St. Louis	2392	1.1	0.04	0.22	0.62	0.15	29.68
Pittsburgh	2345	1.09	0.04	0.23	0.6	0.16	23.57
San Juan	2245	1.11	0.02	0.36	0.56	0.08	26.57
Cincinnati	2040	1.07	0.03	0.31	0.54	0.14	17.96
Cleveland	2004	1.07	0.04	0.19	0.66	0.14	19.13
Charlotte	1920	1.08	0.04	0.3	0.57	0.11	21
San Diego	1897	1.08	0.03	0.28	0.54	0.17	28.89
Baltimore	1857	1.09	0.04	0.23	0.59	0.17	27.72
Indianapolis	1828	1.08	0.05	0.23	0.59	0.17	27.62
Kansas City	1756	1.06	0.04	0.21	0.58	0.2	32.09
Denver	1729	1.07	0.05	0.2	0.57	0.22	40.6
Orlando	1548	1.11	0.06	0.2	0.61	0.18	26.3
San Antonio	1547	1.07	0.05	0.17	0.6	0.21	28.33
Nashville	1460	1.08	0.03	0.27	0.59	0.14	19.08
Milwaukee	1413	1.06	0.06	0.14	0.55	0.3	28.27

towns. Again, this data set comprises the street networks of every incorporated city and town in the United States. Following recent work by Barthlemy and Flammini (3) and Strano et al. (13), we examine the relationship between the total street length L and the number of nodes n. The former proposed a model of cities in which L and n scale nonlinearly as $n^{1/2}$, and the latter recently confirmed this empirically with a small sample of ten European cities' street networks. However, their small sample size may limit the generalizability and interpretability of their finding. To investigate this empirically, we examine the relationship between the total street length L and the number of nodes n for 19,655 US cities and towns and find a strong *linear* relationship ($r^2 = 0.98$), as depicted in Figure 3, contradicting Strano et al. We also find a similar linear relationship at the metropolitan and neighborhood scales.

Moreover, previous findings (e.g., 8) suggest the distribution of street segment lengths in an urban street network follows a power-law distribution. However, we find that these networks instead generally follow a lognormal distribution, which makes theoretical sense as most street networks are not truly scale-free. For example, a typical street network might comprise very few very long street segments (e.g., 1 km), more medium-length segments (e.g., 250 m), and many short segments (e.g., 80 m), but very few very short segments (e.g., 10 m). This theoretical illustration suggests the lognormal distribution that this study generally

finds across municipal US street networks. An exception, of course, lies in consistently-sized orthogonal grids filling a city's incorporated spatial extents. Such distributions are extremely peaked around a single value: the linear length of a grid block.

This study finds that such cities are not uncommon, particularly between the Mississippi River and the Rocky Mountains: the Great Plains states are characterized by a unique street network form that is both orthogonal and reasonably dense. The former is partly the result of topography (flat, plains terrain that allows idealized grids) and design history (platting and development during the late nineteenth century) that favor orthogonal grids, as discussed earlier. The latter results from the fact that most towns across the Great Plains exhibit minimal suburban sprawl. Thus, the municipal boundaries snugly embrace the grid-like street network, without extending to accommodate the vast peripheral belt of twentieth century sprawl, circuity, and "loops and lollipops" (12) that characterizes cities in California that were settled in the same era but later subjected to substantial suburbanization.

For example, if we measure connectedness in terms of average number of streets per node at the city-scale and then aggregate these cities by state (Table 4), we find Nebraska, Kansas, South Dakota, Montana, North Dakota, Oklahoma, and Iowa have, in order, the highest medians (Figure 4). This indicates the most grid-like networks. If we measure

Table 3. Selected summary stats for every incorporated city and town in the US.

measure	mean	std dev	min	median	max
Area (km2)	16.703	107.499	0.039	3.918	7434.258
Avg of the avg neighborhood degree	2.94	0.297	0.4	2.953	3.735
Avg of the avg weighted neighborhood degree	0.033	0.141	¡0.001	0.029	9.357
Avg circuity	1.067	0.159	1	1.055	20.452
Avg clustering coefficient	0.048	0.041	¡0.001	0.04	1
Avg weighted clustering coefficient	0.01	0.018	¡0.001	0.005	0.524
Intersection count	324	1266	0	83	62996
Avg degree centrality	0.093	0.136	¡0.001	0.052	2.667
Edge density (km/km2)	12.654	6.705	0.006	11.814	58.603
Avg edge length (m)	161.184	80.769	25.822	144.447	3036.957
Total edge length (km)	159.067	578.521	0.052	40.986	24728.326
Proportion of dead-ends	0.192	0.093	¡0.001	0.184	1
Proportion of 3-way intersections	0.572	0.11	¡0.001	0.579	1
Proportion of 4-way intersections	0.237	0.129	¡0.001	0.217	1
Intersection density (per km2)	29.363	21.607	¡0.001	24.719	259.647
Average node degree	5.251	0.668	0.8	5.268	7.166
m	1046	3924	2	275	176161
n	401	1516	2	103	71993
Node density (per km2)	35.449	24.409	0.047	30.718	296.74
Max PageRank value	0.034	0.046	¡0.001	0.021	0.87
Min PageRank value	0.005	0.018	¡0.001	0.002	0.5
Self-loop proportion	0.005	0.015	¡0.001	¡0.001	1
Street density (km/km2)	6.528	3.435	0.003	6.109	29.302
Average street segment length (m)	162.408	81.035	25.822	145.479	3036.957
Total street length (km)	86.096	331.048	0.026	21.005	15348.008
Street segment count	558	2208	1	140	107393
Average streets per node	2.851	0.282	1	2.852	4

density and connectedness in terms of intersection density—previously identified as an emergent property of complex network organization—at the city-scale and then aggregate these cities by state, we find Rhode Island, Nebraska, New Jersey, Kansas, and Montana have, in order, the highest medians. We again see three Great Plains states near the top, alongside densely populated East Coast states. Nebraska also has the smallest block sizes (measured via the proxy of average street segment length) while the largest are concentrated in the deep South, upper New England, and Utah (Figure 4).

Municipal boundaries vary greatly in their extents around the built-up area. For example, while Rhode Island averages 56 intersections/km^2 in its cities and towns, Alaska averages only 1.3. This is an artifact of Alaska's municipal boundaries often extending thousands of square kilometers beyond the actual built-up area. In fact, Alaska has four cities (Anchorage, Juneau, Sitka, and Wrangell) with such large municipal extents that their land areas exceed that of the state of Rhode Island. These state-level aggregations of municipal-scale street network characteristics show clear variation across the country that reflect topography, economies, culture, planning paradigms, and settlement eras. But they also aggregate and thus obfuscate the variation within each state and within each city. To explore these smaller-scale differences, the following section examines street networks at the neighborhood scale.

Analysis of Neighborhood-Scale Street Networks

Thus far, we have examined every urban street network in the United States at the metropolitan and municipal scales. This analysis has focused on the complexity on the network in terms of density, resilience, and connectedness. While the metropolitan scale captures the emergent character of the wider region's complex system, and the municipal scale captures planning decisions made by a single city government, the neighborhood scale best represents the scale of individual urban design interventions into the urban form. Further, this scale more commonly reflects individual designs, eras, and paradigms in street network development than the "many hands, many eras" evolution of form at larger scales.

Table 5 presents summary statistics for this data set. Compared to the summary statistics presented at the metropolitan scale (Table 1) and the municipal scale (Table 3), here we see much greater variance. This is expected, given the smaller network sizes at the neighborhood scale. A few neighborhoods have no intersections within their Zillow-defined boundaries, resulting in a minimum intersection density of 0 across the data set. Meanwhile, the small neighborhood of Cottages North in Davis, California has the highest intersection density in the country, 444/km^2, largely as an artifact of its small area as the denominator.

Nationwide, the typical neighborhood averages 2.9 streets per intersection, reflecting the prevalence of 3-way intersections in the US, discussed earlier. The median

Table 4. Median values, aggregated by state plus DC, of selected measures of the municipal-scale street networks for every city and town in the US.

State	Intersect density (per km2)	Avg streets per node	Avg circuity	Avg street segment length
AK	1.28	2.43	1.1	223.5
AL	9.7	2.64	1.07	190.81
AR	15.75	2.78	1.06	166.32
AZ	12.45	2.77	1.08	171.8
CA	32.58	2.74	1.07	143.79
CO	29.26	2.88	1.06	136.68
CT	28.05	2.7	1.07	165.87
DC	58.91	3.26	1.04	122.23
DE	25.3	2.8	1.06	127.8
FL	26.26	2.87	1.07	150.75
GA	15.25	2.78	1.07	177.5
HI	8	2.42	1.07	177.93
IA	24.08	3.02	1.04	129.36
ID	33.85	2.91	1.06	132.08
IL	29.02	2.93	1.05	137.77
IN	35.25	2.93	1.05	125.72
KS	43.94	3.14	1.04	124.39
KY	25.12	2.68	1.07	151.28
LA	17.14	2.79	1.06	162.62
MA	32.33	2.76	1.07	135.98
MD	28.67	2.79	1.07	133.69
ME	7.69	2.67	1.07	198.93
MI	20.93	2.9	1.05	153.5
MN	18.96	2.87	1.06	152.92
MO	29.87	2.89	1.06	138.29
MS	14.76	2.75	1.06	174.86
MT	38.94	3.11	1.04	126.89
NC	19.28	2.65	1.06	166.69
ND	34.28	3.07	1.04	123.93
NE	45.89	3.16	1.04	119.79
NH	12.22	2.69	1.1	175.88
NJ	44.98	2.88	1.04	130.79
NM	18.5	2.93	1.05	152.02
NV	13.86	2.77	1.07	147.35
NY	21.89	2.75	1.06	156.88
OH	25.23	2.8	1.05	142.08
OK	28.22	3.03	1.05	139.5
OR	35.08	2.69	1.06	121.18
PA	35.69	2.87	1.05	128.34
RI	56.23	2.86	1.05	110.35
SC	18.76	2.81	1.06	169.21
SD	32.01	3.12	1.04	130.75
TN	13.62	2.71	1.07	192.83
TX	23.85	2.92	1.05	160.44
UT	12.58	2.71	1.06	191.04
VA	25.18	2.63	1.08	145.65
VT	18.91	2.55	1.08	145.18
WA	28.71	2.75	1.06	134.02
WI	17.87	2.81	1.06	156.19
WV	28.45	2.67	1.08	136.57
WY	23.48	2.92	1.06	143.63

proportions of each node type are 14.5% for cul-de-sacs, 57.4% for 3-way intersections, and 23.4% for 4-way intersections. The typical neighborhood averages 135-meter street segment lengths and 46.4 intersections per km^2. At the neighborhood scale (sample size of 6,857 in this analysis) we again find the same strong *linear* relationship between total street length and the number of nodes in a network ($r^2 = 0.98$), as seen in Figure 3, contradicting the smaller-sample findings of Strano et al. (13).

Table 5. Selected summary stats for all the neighborhood-scale street networks.

	mean	?	min	median	max
Area (km2)	5.322	15.463	0.008	1.738	323.306
Avg of the avg neighborhood degree	2.598	0.436	¡0.001	2.67	3.632
Avg of the avg weighted neighborhood degree	0.031	0.041	¡0.001	0.029	2.991
Avg circuity	1.08	0.411	1	1.044	24.29
Avg clustering coefficient	0.044	0.055	¡0.001	0.034	1
Avg weighted clustering coefficient	0.01	0.027	¡0.001	0.005	0.799
Intersection count	173	379	0	76	8371
Avg degree centrality	0.13	0.27	0.001	0.054	4
Edge density (km/km2)	17.569	7.095	0.025	18.152	59.939
Avg edge length (m)	142.279	59.182	8.447	133.848	2231.331
Total edge length (km)	71.369	166.566	0.017	29.88	3563.409
Proportion of dead-ends	0.17	0.131	¡0.001	0.145	1
Proportion of 3-way intersections	0.559	0.146	¡0.001	0.574	1
Proportion of 4-way intersections	0.275	0.176	¡0.001	0.234	1
Intersection density (per km2)	49.497	28.33	¡0.001	46.43	444.355
Average node degree	4.675	0.836	0.545	4.736	7.283
m	5201	1185	1	217	27289
n	208	459	2	90	9327
Node density (per km2)	58.677	31.802	0.063	55.626	499.9
Max PageRank value	0.055	0.086	¡0.001	0.026	0.889
Min PageRank value	0.01	0.041	¡0.001	0.002	0.5
Self-loop proportion	0.007	0.034	¡0.001	¡0.001	1
Street density (km/km2)	9.744	4.085	0.013	9.882	33.737
Average street segment length (m)	143.664	60.023	7.376	134.877	2231.331
Total street length (km)	40.049	93.987	0.009	16.248	1960.643
Street segment count	288	656	1	119	14754
Average streets per node	2.925	0.408	1	2.944	4.026

Due to the extreme values seen—resulting from the large variance in neighborhood size—we can filter the data set to examine only large neighborhoods (i.e., with area greater than the median value across the data set). In this filtered set, the five neighborhoods with the highest intersection density are all in central Philadelphia. Central neighborhoods are common at the top of this list, including Point Breeze, Philadelphia; Central Boston; Central City, New Orleans; Downtown Tampa; and Downtown Portland (Figure 5). The three neighborhoods with the lowest intersection density are on the outskirts of Anchorage, Alaska. In this filtered set, the neighborhoods with the greatest average number of streets per node tend to be older neighborhoods with orthogonal grids, such as Virginia Park, Tampa; Outer Sunset, San Francisco; and New Orleans' French Quarter. The neighborhoods with the lowest tend to be sprawling and often hilly suburbs far from the urban core, such as Scholl Canyon in Glendale, California or Sonoma Ranch in San Antonio, Texas.

For comparison, Figure 5 compares one square mile of the centers of Philadelphia, Portland, and San Francisco with one square mile of each of their suburbs. The connectedness and density of the central cities is clear, as is the disconnectivity of their suburbs. In fact, the suburbs have more in common with one another—despite being hundreds or thousands of miles apart—than they do with their central city neighbors, suggesting that land use and an era's prevailing design paradigm is paramount to geographical localism and regional context. The top row of Figure 5 represents an era of urban planning and development that preceded the automobile,

while the bottom row reflects the exclusionary zoning and mid-late twentieth century era of automobility in residential suburb design—namely the "loops and lollipops" and the "lollipops on a stick" design patterns identified by Southworth and Ben-Joseph (12).

Discussion

A complex spatial network is a network, embedded in space, that has a nontrivial topology. In other words, its structure and organization is neither fully regular nor fully random. Of the network measures, this empirical study has presented preliminary findings that emphasize street network complexity in terms of density, resilience, and connectedness. Older, denser, and more self-organized networks, such as those at the heart of Boston or lower Manhattan are complex in terms of category I complexity. Sprawling, disconnected suburban neighborhoods rank low on all measures of complexity, with the exception that their high circuity can lend itself to disorder (i.e., category I). The orthogonal grid we see in the downtowns of Portland and San Francisco have high density (i.e., intersection and street densities), connectedness (i.e., average number of streets per node), and order (based on circuity and statistical dispersion of node types), but low resilience in the presence of one-way streets, measured by maximum betweenness centralities and average node connectivity increases when switching from one-way to bi-directional edges. These latter neighborhoods are complex in terms of category II and category III complexity. Their connectedness, density, and
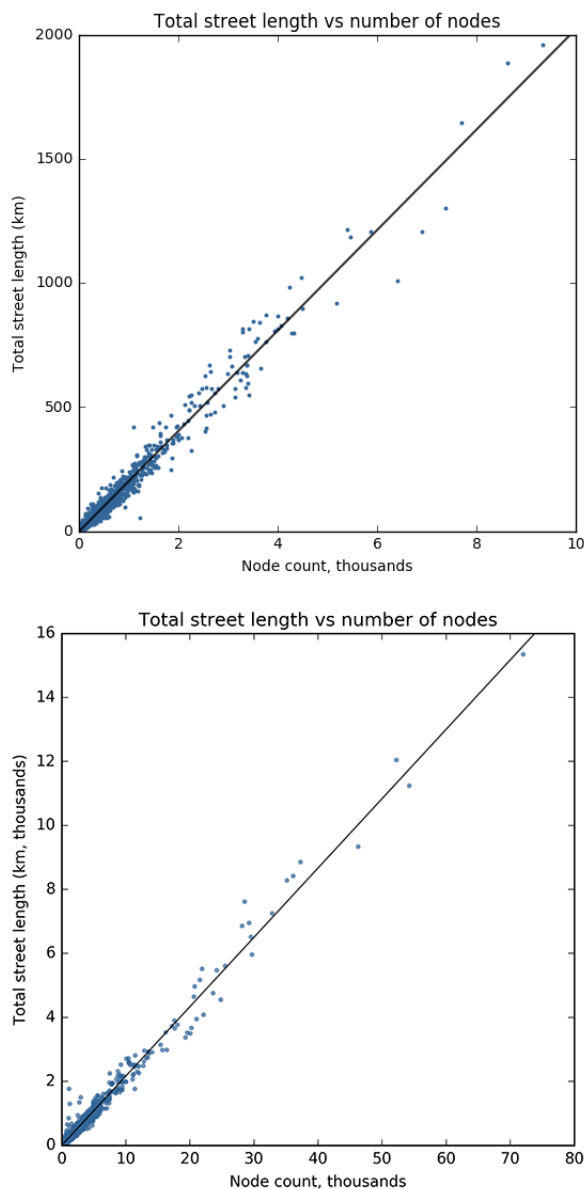
Total street length vs number of nodes

Total street length vs number of nodes

Figure 3. The linear relationship between total street length and number of nodes in the street networks of 6,857 US neighborhoods (top) and 19,655 US cities and towns (bottom).

orderliness balances at the midpoint embraced by category II. However, their extreme regularity of block sizes, streets per node, and orthogonality best represents category III.

Another critical takeaway of this analysis is that scale matters. The median average circuity is lower across the neighborhoods data set than across the municipal set, which in turn is lower than across the urbanized areas set. Conversely, the median average number of streets per node is higher across the neighborhoods data set than across the municipal set, which in turn is higher than across the urbanized areas set. The median intersection density per km^2 is about 83% higher in the neighborhoods data set than in the municipal or urbanized areas sets. These findings make sense: the Zillow neighborhood boundaries focus on large, core cities with older and denser street networks. The municipal boundaries only include incorporated cities and towns—discarding small census-designated places and

unincorporated communities. The urbanized area boundaries include far-flung sprawling suburbs.

The characteristics of an urban network for a city fundamentally depend on what city means: municipal boundaries, urbanized areas, or just certain neighborhoods? The first is a merely legal definition, but also captures the scope of city planning authority and decision-making for top-down interventions into a street network. The second captures a wider self-organized human system and its emergent built form, but tends to aggregate multiple non-homogenous built forms together into a single unit of analysis. The third captures the nature of the local built form and lived experience, but at the expense of a broad view of the urban system and metropolitan-scale trip-taking. In short, multiple scales in concert provide planners and scholars a clearer view of the urban form and the topological and metric complexity of the street network than any single scale can.

We find a strong linear relationship, invariant across scales, between total street length, L, and the number of nodes, n, in a network that contradicts some previous findings in the literature that relied on small sample sizes. We also find that most networks empirically demonstrate a lognormal distribution of street segment lengths, contradicting some previous findings in the literature, as discussed earlier. However, we believe our empirical finding makes more sense theoretically and is supported by the large-sample data at multiple scales. An obvious exception to lognormal distribution lies in those networks that exhibit substantial uniformity across the entire network. At the neighborhood scale, examples include downtown neighborhoods with similarly consistent orthogonal grids, such as that of Portland, Oregon. At the municipal scale, examples include towns in the Great Plains that have orthogonal grids with consistent block sizes, platted at one time, and never subjected to sprawl.

These findings are interesting for the practice and history of planning. The spatial signatures of the Homestead Act, successive land use regulations, urban design paradigms, and planning instruments remain clearly visible today in these cities' urban forms and street networks due to path dependence. When comparing the median municipal street networks of each state, Nebraska has the lowest circuity, the highest average number of streets per node, the second shortest average street segment length, and the second highest intersection density for similar reasons. These preliminary findings point to how street networks across the Great Plains developed all at once, but grew very little afterwards—unlike, for instance, most cities in California that were settled in the same era but later subjected to sprawl.

This finding suggests future research should incorporate a temporal analysis that this present study does not do with its cross-sectional data. This empirical analysis emphasized network structure. Expanding the study of complex urban form by examining the other types of complexity—and further linking structural complexity to the temporal complexity of dynamics and processes—lies ahead as critical future work.

In total, this paper analyzed 497 urbanized areas' street networks, 19,655 cities' and towns' street networks, and 6,857 neighborhoods' street networks. These sample sizes were larger than those in any previous study. It looked
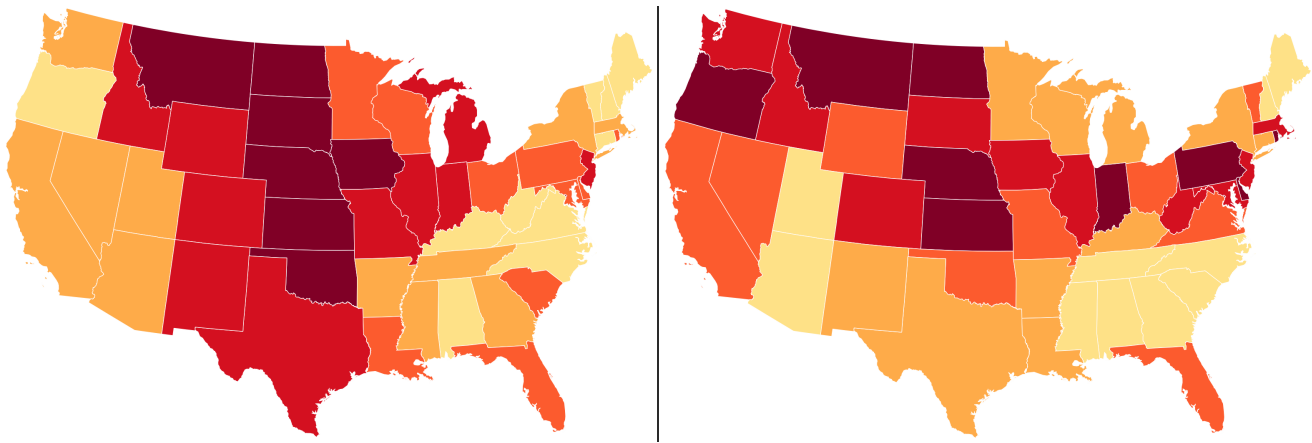
Figure 4. Left: contiguous US states by median of mean streets per node in municipal street networks, from lowest (pale yellow) to highest (dark red). Right: contiguous US states by median of mean street segment length in municipal street networks, from longest (pale yellow) to shortest (dark red).

Figure 5. Square-mile comparisons of central cities and their suburbs. Left: top, downtown Philadelphia; bottom, its suburb, King of Prussia. Middle: top, downtown Portland; bottom, its suburb, Beaverton. Right: top, downtown San Francisco; bottom, its suburb Concord.

at both metric and topological measures of the structure and complexity of these networks—particularly focusing on density, connectedness, and resilience. These preliminary empirical findings demonstrate the use of OSMnx as a new piece of research infrastructure. They suggest to urban planners new methods for acquiring and analyzing street network data, including new methods for evaluating network resilience and resilience gains with betweenness centralities and average node connectivities. Finally, this study has made all of these network datasets—for 497 urbanized areas, 19,655 cities and towns, and 6,857 neighborhoods—along with all of their attribute data and complexity measures available in an online public repository for other researchers to study and repurpose.

References

[1] Jochen Albrecht and Mimi Abramovitz. Indicator Analysis for Unpacking Poverty in New York City. Technical report, CUNY, New York, NY, 2014.

[2] Marc Barthlemy. Betweenness centrality in large complex networks. *The European Physical Journal B: Condensed Matter and Complex Systems*, 38(2):163–168, March 2004. ISSN 1434-6028. doi: 10.1140/epjb/e2004-00111-4.

[3] Marc Barthlemy and Alessandro Flammini. Modeling Urban Street Patterns. *Physical Review Letters*, 100(13), April 2008. ISSN 0031-9007. doi: 10.1103/PhysRevLett.100.138702.

[4] Lowell W. Beineke, Ortrud R. Oellermann, and Raymond E. Pippert. The Average Connectivity of a Graph. *Discrete Mathematics*, 252(1):31–45, 2002. ISSN 0012-365X. doi: 10.1016/S0012-365X(01)00180-7.

[5] Max Besbris, Jacob William Faber, Peter Rich, and Patrick Sharkey. Effect of neighborhood stigma on economic transactions. *Proceedings of the National Academy of Sciences*, 112(16):4994–4998, April 2015. ISSN 0027-8424. doi: 10.1073/pnas.1414139112.

[6] Geoff Boeing. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. (Manuscript submitted for publication), 2017. doi: 10.2139/ssrn.2865501.

[7] Lawrence Bacon Lee. *Kansas and the Homestead act, 1862-1905*. The management of public lands in the United States. Ayer Company Publishers, North Stratford, NH, 1979. ISBN 978-0-405-11341-3.

[8] A. Paolo Masucci, D. Smith, A. Crooks, and Michael Batty. Random planar graphs and the London street network. *The European Physical Journal B: Condensed Matter and Complex Systems*, 71(2):259–271, 2009. ISSN 1434-6028. doi: 10.1140/epjb/e2009-00290-4.

[9] Jason Porterfield. *The Homestead Act of 1862: A Primary Source History of the Settlement of the American Heartland in the Late 19th Century*. Rosen Publishing Group, New York, NY, 1st edition, January 2005. ISBN 978-1-4042-0178-1.

[10] Harald Schernthanner, Hartmut Asche, Julia Gonschorek, and Lasse Scheele. Spatial Modeling and Geovisualization of Rental Prices for Real Estate Portals. In Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Ana Maria A. C. Rocha, Carmelo M. Torre, David Taniar, Bernady O. Apduhan, Elena Stankova, and Shangguang Wang, editors, *Computational Science and Its Applications - ICCSA 2016*, number 9788 in Lecture Notes in Computer Science, pages 120–133. Springer International, Cham, Switzerland, 2016. ISBN 978-3-319-42110-0. DOI: 10.1007/978-3-319-42111-7_11.

[11] Michael Sherraden. *Inclusion in the American Dream: Assets, Poverty, and Public Policy*. Oxford University Press, Oxford, England, July 2005. ISBN 978-0-19-534709-8.

[12] Michael Southworth and Eran Ben-Joseph. *Streets and the Shaping of Towns and Cities*. McGraw-Hill, New York, NY, 1997. ISBN 978-1-55963-916-3.

[13] Emanuele Strano, Matheus Viana, Luciano da Fontoura Costa, Alessio Cardillo, Sergio Porta, and Vito Latora. Urban Street Networks, a Comparative Analysis of Ten European Cities. *Environment and Planning B: Planning and Design*, 40(6): 1071–1086, December 2013. ISSN 1472-3417. doi: 10.1068/b38216.

[14] U.S. Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria, 2010. URL https://www.census.gov/geo/reference/ua/urban-rural-2010.html.