A century of sprawl in the United States
Supporting Information

Chris Barrington-Leigh
Adam Millard-Ball

Contents

List of Figures

List of Tables

S1 Materials and Methods

S1.1 Data Sources

We quantify street-network sprawl using three different time series, discussed briefly in the main text. Here, we provide additional detail on each series.

1. The **TIGER/Line** series computes our measures of sprawl for all counties in the US, using four different vintages of the TIGER/Line shapefiles: 1992, 2000, 2010 and 2014. Because of lags in the data gathering and release process, we assume each vintage represents the characteristics of the street network in the previous year (1991, 1999, 2009 and 2013).[1]

2. The **Census-based** series computes our measures of sprawl for all counties in the US, using the 2014 vintage of the TIGER/Line shapefiles. We construct earlier years of the time series through assigning the median year built of residential units in each census block group (as reported in the US Census Bureau 2007-11 American Community Survey) to all streets in that block group. This yields a time series from "1939 or earlier" (the earliest category for year built that is reported) to "2005 or later".

3. The **Parcel-based** series also computes our measures of sprawl using the 2014 vintage of the TIGER/Line shapefiles. We construct earlier years of the time series by using tax records for individual land ownership parcels, which we obtain directly from county governments or from the commercial aggregator Boundary Solutions. Figure S1 shows the locations of the 226 counties for which we have parcel data, and the number of parcels from each. We match each parcel to the street network using a combination of address and geospatial data, and succeed in matching 95.1% of the 23,191,172 parcels for which we have year built information. Section S1.2 describes our matching algorithm in detail. We then assign to each street edge the year in which the earliest structure on that edge was built. In other words, we assume that a street was built at the same time as its earliest structure. For each node, we assign the year of the most recent connected edge. This yields a time series from 1920 (before which county data on the year a structure was built appear to be less reliable) to 2012, for the 226 counties that are at least partly urbanized, and for which we could obtain suitable parcel data. The 226 counties in the parcel-based series account for 9.7% of the 2,338 counties and county-equivalents in the

[1] The year in which an edge first appears in the TIGER/Line files varies depending on the Census Bureau update cycle; there is often a lag of many years between construction and incorporation into TIGER/Line. While it is technically possible to construct an annual time series from 1992-2014, a comparison to historic satellite imagery suggested that the data do not support an annual temporal resolution. Moreover, the MAF/TIGER Accuracy Improvement Project (2003-08) appears to have introduced inconsistencies into many counties which were subsequently corrected (for example, by classifying driveways as regular urban streets). We therefore restrict our time dimension to the earliest vintage, the two years of the decennial census, and the most recent vintage.

US with at least one urbanized block group; and a higher (32.7%) share of the urbanized area population.

The three different time series exhibit trends that are generally consistent (Figure 1 in the main text). However, there are differences in levels between the TIGER/Line and Census-based series (where our data include all nodes in the underlying Census Bureau files), and the parcel-based series (where our data are restricted to the subset of nodes where at least one connected edge has a parcel with year-built information). In practice, this means that a lower proportion of deadends is estimated from the parcel-based series, because (i) deadends are more likely to be service or other access roads without associated buildings; and (ii) missing data (e.g. lack of year-built information) is more likely to affect deadends, as missing data for a single edge will lead to missing data for the node. In contrast, data would need to be missing for all 3 or 4 connected edges for this to happen with a 3- or 4-degree node. Mean nodal degree of the 2013 stock was 2.73 according to the TIGER/Line series, and 2.83 according to the parcel-based series. However, these differences are unlikely to affect the analysis in this paper, because trends over time are consistent between the two series.

S1.2 Matching parcels to street edges

This section provides more details of our matching algorithm to link county assessor parcels (which provide the information on the year a structure was built) with edges (i.e., street segments).

Our matching algorithm uses two main inputs for each parcel: (i) the edges that are within 20m of the boundaries of a given parcel; and (ii) the geocoding functionality in ESRI's ArcGIS software. Figure S2 shows the process for matching parcels to edges for the 216 of 226 counties in the parcel dataset that have address data, and the percentage of matches that is obtained through each matching method. For the 10 counties where the parcel dataset omits address data (but includes year-built information), a simplified version of the algorithm is used: a parcel is matched to an edge if and only if there is a unique edge within 10m of the parcel boundary.

We calculate our measures of street-network sprawl at the level of individual nodes and edges, before aggregating (where required) to census block groups, metropolitan regions and other geographic units. Where two nodes are within 15m of each other, we treat them as a single node for purposes of calculating nodal degree. As shown in Figure S3, this procedure accounts for offset intersections (i.e. "dog-legged" or adjacent T-intersections) that functionally are the same intersection, as well as allowing for misaligned streets and other potential imperfections in the TIGER/Line geometry. The 15m distance is approximately the width of a typical two-lane urban street, including on-street parking and sidewalks. We ignore edges that are completely contained within an intersection (defined as a 7.5m radius from each constituent node), so that short edges that connect within an offset intersection, expressway ramps and similar elements of the street network do not inflate nodal degree.

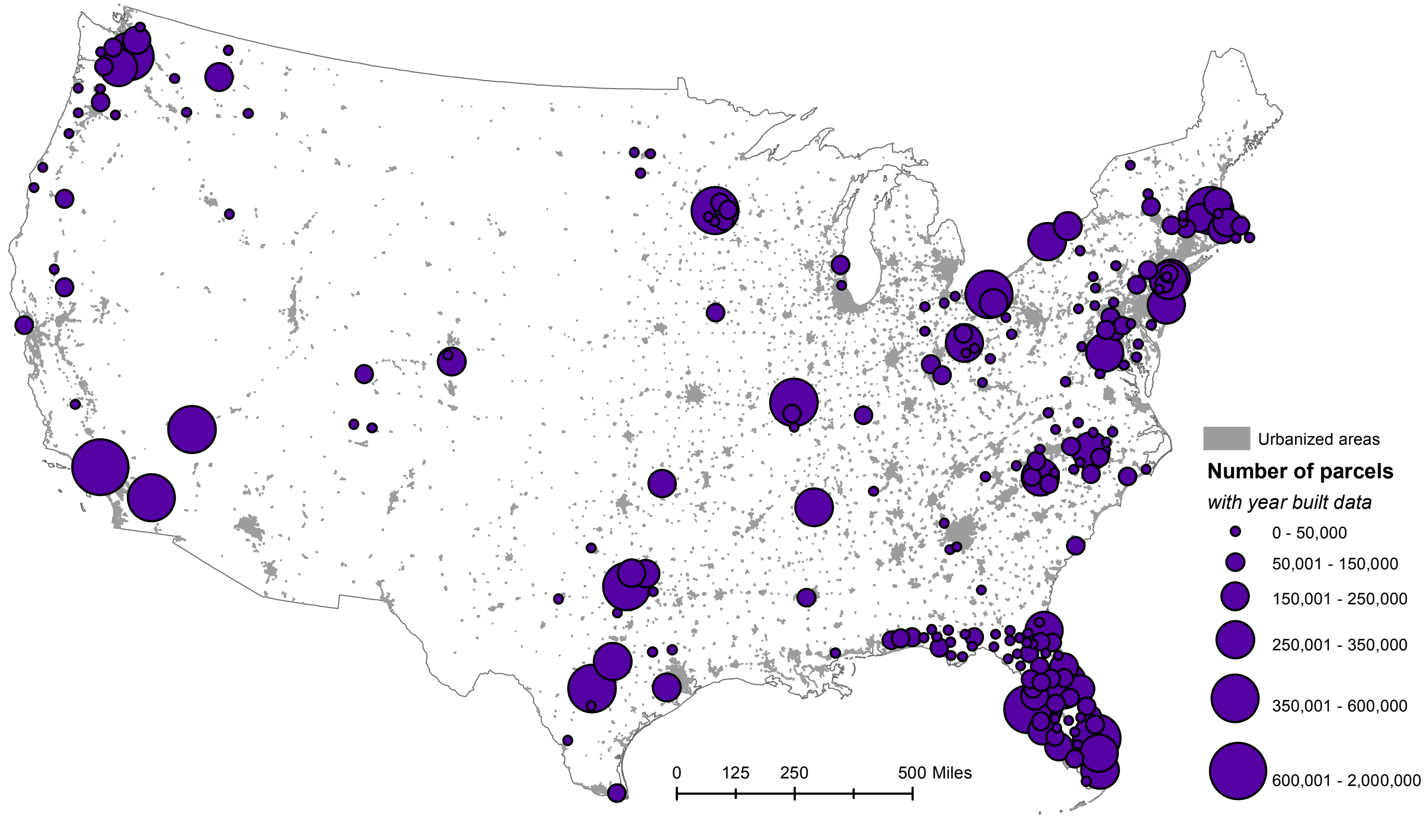

Figure S1: **Distribution of counties with parcel data.** We obtained year-built information for buildings in 226 counties with urban areas, broadly spanning the US and accounting for ∼33% of the urbanized area population.

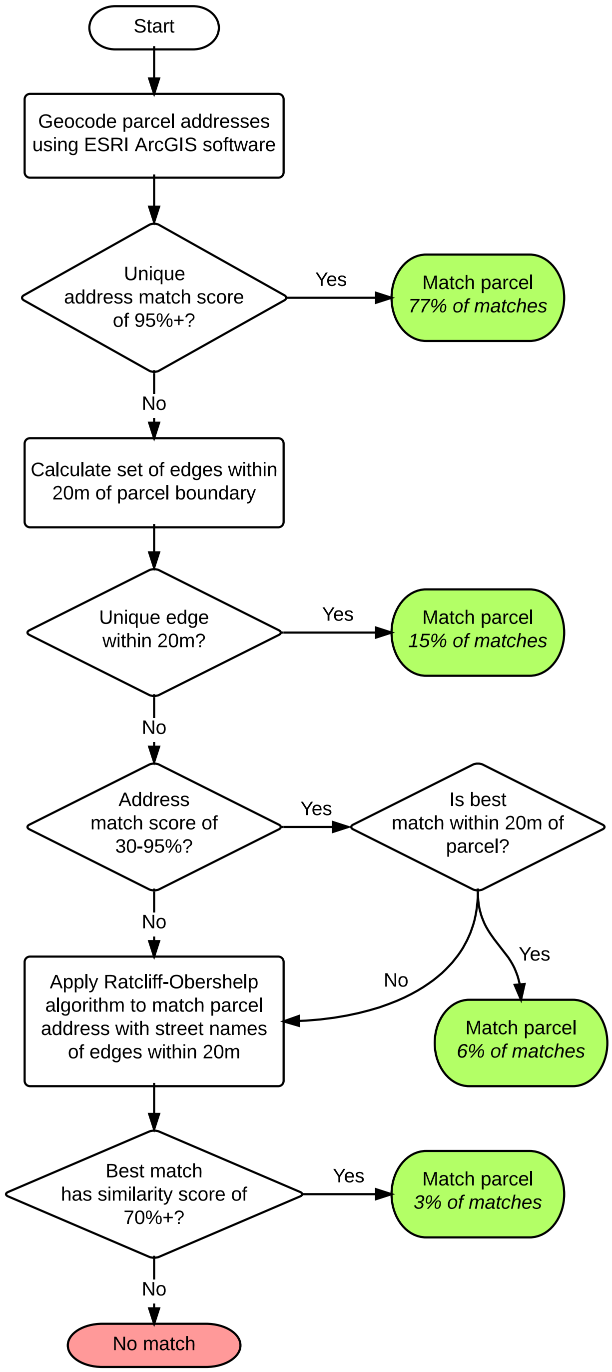

Figure S2: **Algorithm to match parcels to street edges.**

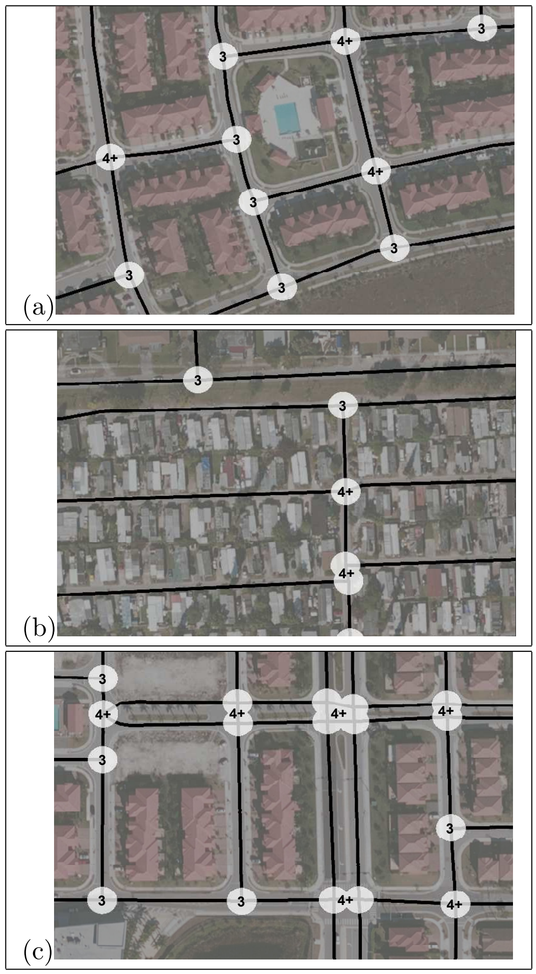

Figure S3: **Calculation of nodal degree.** Each geometric node is buffered (7.5m radius), and overlapping buffers are merged to create our dataset of nodes. In the simple case (a), calculated nodal degree is simply the number of connected edges at each geometric node. Where intersections are offset (b), our procedure merges the adjacent 3-degree nodes to create a 4-degree node. In the complex case of a divided highway (c), our procedure disregards edges that fall entirely within the overlapping buffers; this allows us to ignore freeway ramps, median connectors, and similar streets that do not functionally affect street network connectivity. Source for underlying imagery: ESRI/Digital Globe.

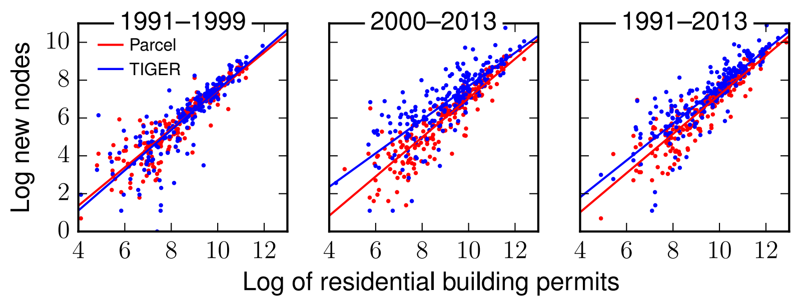

Figure S4: **Growth in nodes versus building permit issuance**. Comparison is for 1991–2013 in counties in our parcel-based series (N=224).

S1.3 Comparison to building permit data

Figure S4 compares our estimates of the growth in nodes from both the TIGER/Line series (in blue) and the parcel-based series (in red) with building permit issuance by county governments. Building permit data are from the US Census Bureau Building Permit Survey. The linear best fit is also shown. The strong correlation between permit activity and new nodes adds confidence to our methods for constructing the historical time series of intersection growth.

S2 Open data

We provide a dataset with our three measures of street-network sprawl — nodal degree, percentage of 4+ degree nodes, and percentage of deadends — for download via the journal website. Standard errors are also included. We provide annual data at the level of counties, metropolitan regions (CSAs and CBSAs), and the entire United States. Note that the data are limited to urbanized areas, defined as block groups where the majority of blocks were classified as urban in the 2010 Census. For completeness, the dataset includes the full time series from ∼1750. However, due to inaccuracies in the county assessor data, which records building ages, we caution against relying on data for the early part of this series. Accordingly, this paper focuses on the period since 1920.

We also provide the full geographic data file for the stock of streets in 2013, indicating the nodal degree of each intersection. This is provided in shapefile format, suitable for analysis with most Geographic Information Systems software, and as a graph file that describes the network.

The data are documented and archived at `http://dx.doi.org/10.5061/dryad.3k502`.

S3 Comparisons with alternative sprawl measures

Sprawl is a multi-dimensional characteristic of urban areas. Under one typology [1], there are eight distinct dimensions of land-use patterns that characterize sprawl, including density, centrality (the distance of development from the Central Business District or CBD) and nuclearity (whether a metropolitan area has a dominant urban center or is polynuclear in character). Preferred measures of urban sprawl are somewhat discipline-dependent, reflecting different policy interests and methodological traditions across disciplines. For architects such as Duany and Plater-Zyberk [2], sprawl is inherently about the rigid segregation of land uses, and urban design features such as the placement of parking in the front setback of homes. Economists, in contrast, have tended to focus on density, the scatteredness of urban development, and the size and spatial extent of metropolitan areas [3, 4, 5, 6, 7]. In large part, this reflects the intellectual history of urban economics, where the Alonso-Muth-Mills model, which posits a monocentric city where all employment is in the CBD and households choose their distance from the CBD by trading off housing and commuting costs, still has great influence [8, 7, 9].

Our street network-based measures characterize sprawl as having a low nodal degree of intersections, a high proportion of deadends, and a low proportion of intersections of degree four or more. (In graph theory, the degree of a node is the number of edges, in this case street segments, connected to the node, in this case the intersection.) Our three measures are empirically or deterministically related to similar ways to measure street connectivity, such as block length or the ratio of links to nodes [10]. Other network metrics such as the network-length linear density of nodes from each node, ratio of network-distance to geographic-distance, and treeness (dendricity) [11] are also related, but are difficult to measure in a time-series dataset such as ours where we cannot assign a year to some edges and nodes.

As noted in the main text, our measures offer several important conceptual and empirical advantages over alternatives such as density, spatial extent and centrality. First, our measures are semi-permanent. In contrast to characteristics such as density, which can change over time, the street network indicates the degree of sprawl at the time it was laid down.

Second, the connectivity of the street network shows a strong theoretical and empirical relationship with important externalities such as greenhouse gas emissions. A high proportion of deadends and a low nodal degree of intersections favor travel by the private car in several ways. Such street patterns typically increase the ratio of network distance to Euclidean distance, which reduces the generalized cost of driving relative to walking. In contrast, a gridded street network tends to be more attractive to pedestrians, is conducive to mixed land uses, allows more efficient service by public transit, and reduces travel speeds by the private car through requiring frequent stops. Low nodal degree also proxies for other factors which favor the private car, such as wider arterials and longer distances between signalized intersections. Unfortunately, these elements of walkability, and others such as sidewalk provision, cannot be measured due to a lack of comprehensive or consistent data.

In contrast, there is a tenuous externality from sprawl when measured by the amount of open space in the square kilometer surrounding a house [7]; by the size or spatial extent of metropolitan areas [3, 4, 5, 6]; or by the extent to which employment is located within a five-mile radius of the CBD [8]. Even the commonly used measure of density has a less direct relationship to the external costs of sprawl than the structure of the street network; density often proxies for other characteristics of the built environment that affect vehicle travel, and the relationship of street connectivity with total vehicle distance traveled, as measured through elasticities, is three times that of population density [12].

Third, a street network-based approach offers extremely high spatial and temporal resolution. Our units of analysis are street segments (edges) and intersections. This provides us with the ability to conduct analysis at any spatial scale, rather than being constrained by the aggregation units for census data or the resolution offered by remote sensing technologies. Our measures of sprawl vary within a city, in contrast to measures such as nuclearity and spatial extent which are a characteristic of an entire metropolitan area. Moreover, our dataset identifies the year that each street segment was built. In contrast, census-based measures such as those in [13] are limited to ten-year intervals, and the availability of remote-sensing data is even more constrained. For example, the approach in [7] is limited to two years of analysis.

Fourth, our measures of sprawl are less susceptible to issues of scale dependence than alternatives such as intersection density (the number of nodes per unit area) or residential density. Such density measures vary depending on the definition of areas; for example, whether parks, water or yet-to-be-developed land are included when measuring surface area. This presents a particular problem with time-series analysis. If the geographic units are held constant (and thus include land that is not developed in early years), such a measure will almost invariable increase over time within a given geographic unit, as more intersections or housing units are built. Thus, density-based measures are best suited for analyzing cross-sectional differences, rather than in the context of the time series that we employ here. Unlike most existing measures which correspond, ultimately, to an area density or geographically weighted average of some kind, our measure amounts to a sum over intersections, and relates to their network structure, regardless of spatial scale.

In any case, different measures of sprawl are often correlated. Figure S5 indicates the relationship between the nodal degree of intersections and three alternative measures of sprawl: residential density, the intensity of development, and a multi-dimensional sprawl index. Nodal degree, the percentage of nodes of degree 4+, and the percentage of deadends correlate with the other measures in the expected manner. The weakest relationship is with the impervious surface area, which indicates that sprawl can be built with varying degrees of impervious surface, for example depending on whether yards and public open spaces are paved. The impervious surface data are are the basis for the analysis in [7], although their measure (the extent to which development is "scattered") is constructed somewhat differently.

Hamidi & Ewing's aggregated sprawl index (which considers street connectivity as one element along with density, mix of uses, and the concentration of population and employment in defined sub-centers) [13] is one example of a composite index, often devised to rank urban areas according to their degree of sprawl. [1] use a similar approach to [13], calculating six dimensions and then summing them into a single index.

All the measures of sprawl also correlate in the expected manner with commute mode share (% of workers commuting by modes other than driving alone) and vehicle ownership. Given that urban form is one of many factors that affects vehicle ownership and travel, along with income, preferences, and so on, it is not surprising that there is considerable dispersion around the lines of best fit (estimated by lowess). However, the directionalities of the relationships are clearly evident in Figure S5.

S4 Additional results

Below are collected several figures and tables which complement or extend those given in the main text. Explanations are given in the captions and in the main text.

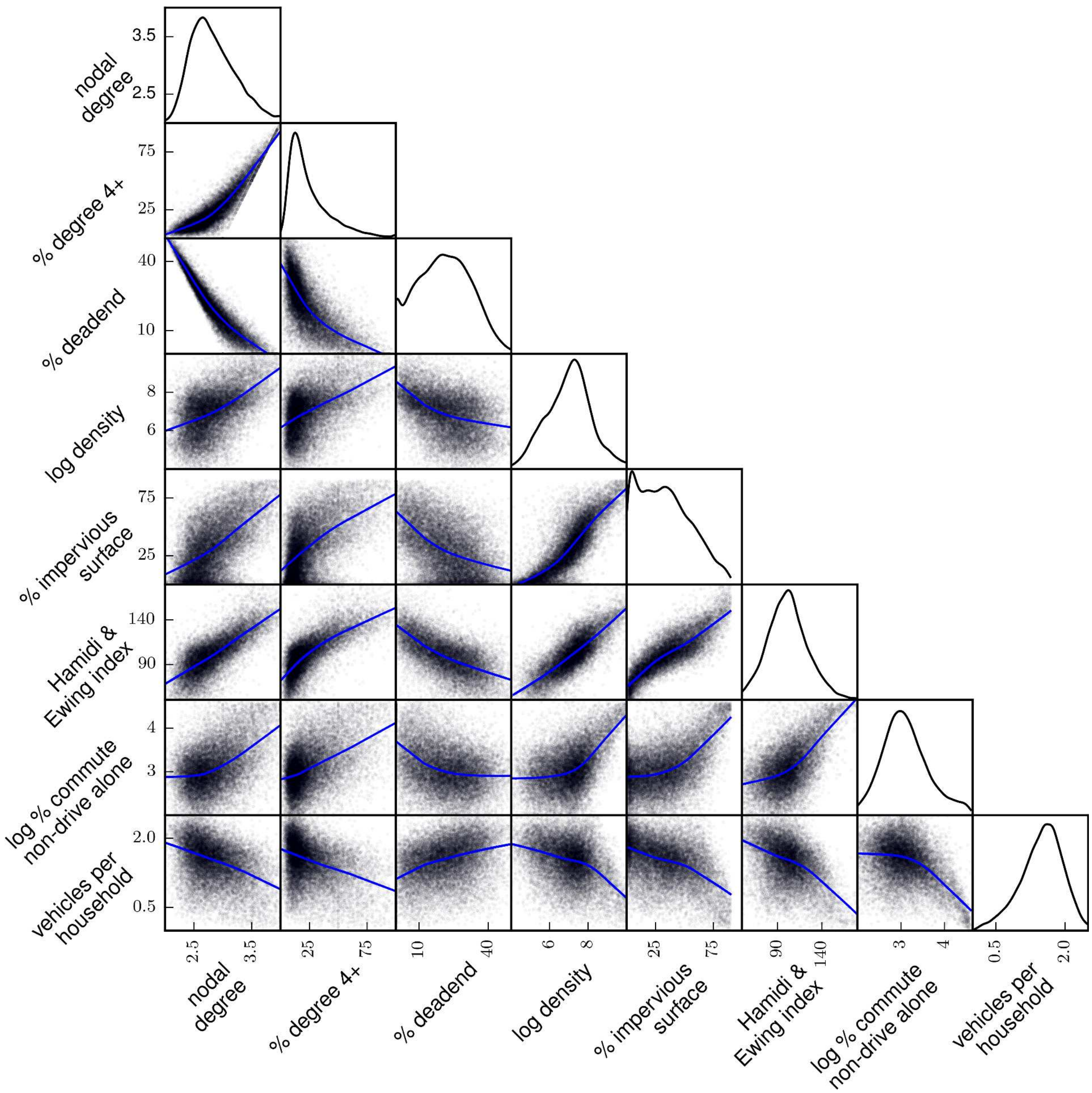

Figure S5: **Correlations between alternative measures of sprawl.** Log density, commute mode share and vehicle ownership are calculated based on the American Community Survey 2007-11. Impervious surface area is calculated based on the National Land Cover Database 2006 [14]. Hamidi & Ewing sprawl index is as reported in [13]. Diagonals provide the kernel density plot for each measure, while off-diagonals plot the relationship between different measures using a lowess smoother. A one-third random sample is used for visualization purposes; data are aggregated to the census tract level.

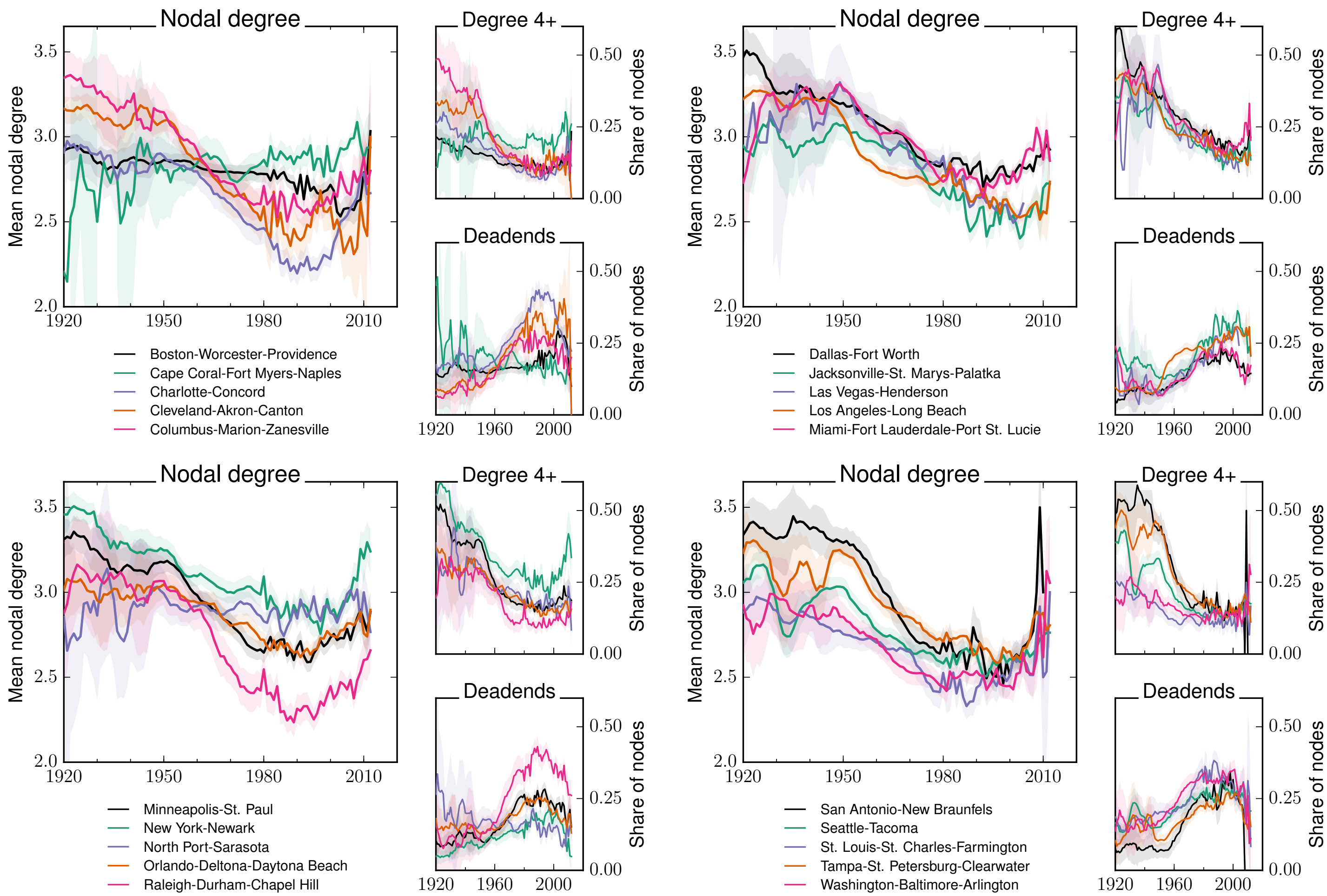

Figure S6: **Trends over time, metropolitan areas.** Figure 2 in the main text provides results for selected metropolitan regions. Here, we show results for the 20 largest metropolitan regions (Combined Statistical Areas as designated by the US Census Bureau) in our parcel-based dataset, as measured by the number of nodes. The regions shown are not necessarily the largest in the United States, as most regions are only partially covered in our dataset. Shaded areas represent 95% confidence intervals.

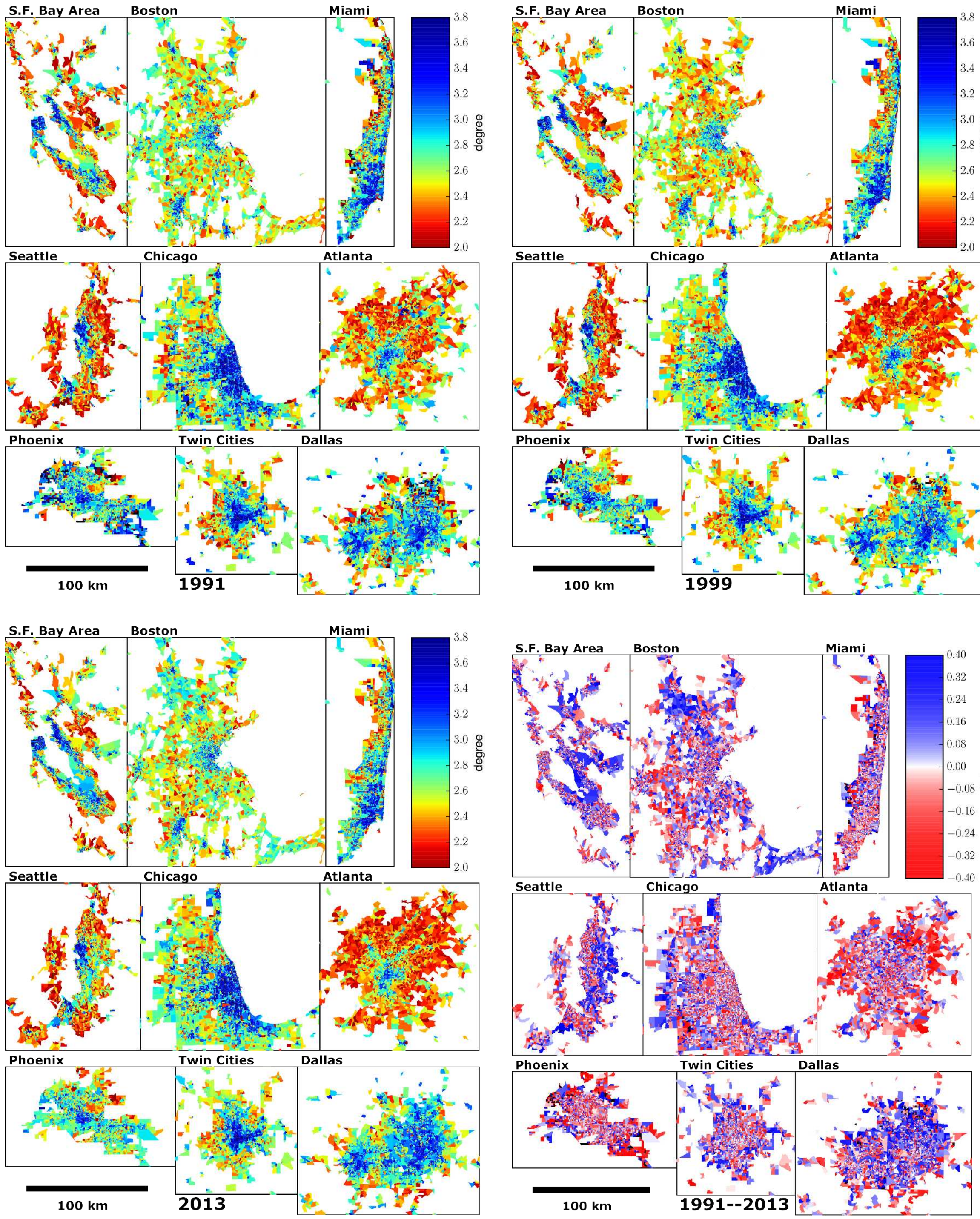

Figure S7: **Levels and changes over time in mean nodal degree in selected metropolitan areas**. Levels in mean nodal degree are shown at three points in time (top left, top right, and bottom left). The third plot (2013) is the one featured in Figure 4 of the main text. The bottom right panel shows the change in mean degree between 1991 and 2013 for block groups with significant increases in census-reported housing units. Regions are mapped to the same scale.

Metropolitan Region	Mean nodal degree					% Degree 4+					% Deadend				
	1991	1999	2009	2013	1991–2013	1991	1999	2009	2013	1991–2013	1991	1999	2009	2013	1991–2013
Urbanized US	2.80	2.75	2.74	2.74	2.60	22.9	21.7	21.4	21.5	17.9	21.3	23.2	23.9	23.5	29.0
Harrisburg-York-Lebanon, PA	2.89	2.87	2.86	2.88	2.86	25.9	25.3	25.2	25.5	24.3	18.7	18.9	19.6	18.8	19.0
Dallas-Ft Worth, TX-OK	2.91	2.88	2.86	2.87	2.81	24.4	23.4	25.4	25.7	28.0	16.9	17.8	19.8	19.2	23.3
Oklahoma City-Shawnee, OK	2.95	2.91	2.88	2.88	2.74	28.4	26.7	25.1	24.9	18.4	16.5	18.1	18.7	18.4	22.1
Birmingham-Hoover-Talladega, AL	2.71	2.65	2.71	2.72	2.73	20.5	19.1	22.5	22.4	26.1	24.7	27.2	25.6	25.3	26.4
Orlando-Deltona-Dayt. Bch, FL	2.72	2.69	2.69	2.72	2.70	18.8	17.4	17.1	18.0	16.5	23.2	24.3	23.9	23.1	23.1
Denver-Aurora, CO	2.86	2.81	2.80	2.80	2.69	25.2	24.4	23.5	23.7	20.7	19.5	21.5	22.0	21.7	25.9
San Jose-San Francisco-Oakland, CA	2.69	2.67	2.68	2.69	2.69	19.7	19.0	21.6	21.8	27.3	25.2	26.0	27.0	26.4	29.4
Milwaukee-Racine-Waukesha, WI	2.92	2.88	2.86	2.87	2.67	29.2	28.3	27.3	27.9	22.7	18.6	20.0	20.4	20.4	27.8
New Orleans-Metairie-Hammond, LA-MS	2.98	2.95	2.87	2.88	2.66	34.8	33.9	31.1	30.8	21.9	18.6	19.6	22.0	21.5	27.7
Boston-Worcester-P'dence	2.71	2.67	2.70	2.70	2.66	15.5	14.8	15.3	15.2	14.3	22.1	24.1	22.7	22.6	24.0
Austin-Round Rock, TX	2.73	2.70	2.69	2.71	2.66	18.2	17.7	18.7	18.9	20.3	22.8	24.1	24.9	24.1	27.0
Memphis-Forrest City, TN-MS-AR	2.72	2.65	2.70	2.70	2.66	19.4	17.6	19.7	19.7	20.3	23.6	26.1	25.1	24.9	27.1
Hartford-West Hartford, CT	2.66	2.64	2.65	2.66	2.66	12.8	12.4	12.4	12.7	12.4	23.2	24.4	23.8	23.3	23.4
Philadelphia-Reading-Camden, PA-NJ-DE-MD	2.90	2.87	2.85	2.85	2.65	25.2	24.5	24.0	23.9	18.5	17.6	18.8	19.5	19.3	26.5
Indianapolis-Carmel-Muncie, IN	2.82	2.75	2.75	2.76	2.65	24.7	22.7	21.7	21.9	17.1	21.4	23.8	23.4	23.2	26.1
Miami-Ft L'dale-Pt St. Lucie, FL	2.90	2.84	2.81	2.83	2.65	24.2	22.7	21.8	22.6	18.7	17.1	19.1	20.3	19.9	26.9
New York-Newark, NY-NJ-CT-PA	2.86	2.82	2.81	2.82	2.65	23.1	22.5	21.6	22.0	16.5	18.7	20.1	20.1	19.9	25.9
Minneapolis-St. Paul, MN-WI	2.87	2.81	2.78	2.80	2.64	27.2	25.3	24.0	24.8	19.9	19.9	22.0	23.0	22.5	27.9
Columbus-Marion-Zanesville, OH	2.79	2.74	2.72	2.73	2.64	22.1	20.1	19.3	19.3	14.7	21.8	23.2	23.8	23.2	25.5
St. Louis-St. Charles-F'ton, MO-IL	2.73	2.68	2.70	2.70	2.64	20.9	19.9	20.3	20.1	18.1	24.2	26.1	25.2	25.1	27.3
Chicago-Naperville, IL-IN-WI	2.94	2.90	2.86	2.87	2.61	29.8	28.4	27.0	27.1	16.7	17.9	19.3	20.3	19.9	27.7
Detroit-Warren-Ann Arbor, MI	2.94	2.88	2.85	2.86	2.59	28.6	25.9	24.6	24.9	13.0	17.1	18.9	20.0	19.5	27.1
Phoenix-Mesa-Scottsdale, AZ	2.83	2.75	2.71	2.72	2.59	20.8	18.4	16.1	15.9	10.2	19.0	21.5	22.4	22.1	25.8
Los Angeles-Long Beach, CA	2.79	2.74	2.72	2.72	2.59	22.4	21.0	21.9	21.8	20.4	21.9	23.4	24.9	24.7	30.9
Pittsburgh-New Castle-Weirton, PA-OH-WV	2.79	2.77	2.73	2.75	2.58	22.6	22.5	21.6	22.0	19.6	21.6	22.7	24.1	23.6	30.8
Salt Lake City-Provo-Orem, UT	2.69	2.65	2.63	2.64	2.58	18.4	17.1	16.4	16.3	13.6	24.7	26.3	26.6	26.0	27.9
Seattle-Tacoma, WA	2.58	2.55	2.58	2.57	2.56	20.0	19.3	18.0	17.9	13.1	31.0	32.0	30.1	30.2	28.5
Cleveland-Akron-Canton, OH	2.86	2.82	2.76	2.78	2.55	24.7	23.6	21.8	22.4	15.4	19.5	20.9	22.7	22.2	30.3
Sacramento-Roseville, CA	2.64	2.61	2.61	2.61	2.55	15.6	15.1	15.4	15.4	14.8	25.7	27.0	27.3	27.2	30.1
Virginia Beach-Norfolk, VA-NC	2.64	2.58	2.60	2.60	2.54	20.8	19.4	20.2	20.2	19.2	28.7	30.7	30.3	30.2	32.4
Washington-B'more-Arling., DC-MD-VA-WV-PA	2.62	2.58	2.57	2.59	2.54	18.3	17.5	17.5	18.2	18.2	28.0	30.0	30.2	29.8	31.9
Tampa-St. Petersburg-Clearwater, FL	2.82	2.80	2.76	2.77	2.54	22.1	21.6	21.4	21.6	19.2	19.8	20.7	22.6	22.1	32.7
Kansas City-Overland Park-Kansas City, MO-KS	2.87	2.82	2.77	2.77	2.54	25.7	24.7	23.3	23.3	17.9	19.2	21.6	23.3	23.1	32.1
Las Vegas-Henderson, NV-AZ	2.81	2.70	2.62	2.63	2.52	20.6	17.3	15.5	15.7	12.5	19.9	23.9	26.8	26.4	30.5
Portland-Vancouver-Salem, OR-WA	2.71	2.68	2.63	2.64	2.46	23.9	22.6	20.9	20.9	13.5	26.3	27.4	28.7	28.4	33.7
Grand Rapids-Wyoming-Muskegon, MI	2.82	2.76	2.71	2.71	2.46	24.4	22.3	20.3	20.3	10.0	21.4	23.2	24.9	24.5	32.2
Buffalo-Cheektowaga, NY	2.96	2.92	2.87	2.89	2.45	25.3	24.1	23.1	23.7	13.5	14.8	16.0	17.9	17.4	34.3
Houston-The Woodlands, TX	2.82	2.77	2.69	2.69	2.44	26.2	24.7	23.1	23.0	16.9	22.0	23.9	27.2	26.9	36.3
Jacksonville-St. Marys-Palatka, FL-GA	2.74	2.67	2.63	2.64	2.43	22.0	20.1	18.9	19.2	13.0	24.0	26.6	27.7	27.4	34.8
San Antonio-New Braunfels, TX	2.93	2.87	2.78	2.79	2.43	28.0	26.0	24.1	24.0	13.9	17.6	19.7	22.8	22.7	35.7
Nashville-Davidson–Murfreesboro, TN	2.67	2.59	2.57	2.58	2.42	16.0	15.1	15.6	15.7	15.3	24.5	28.2	29.2	28.7	36.5
San Diego-Carlsbad, CA	2.62	2.59	2.56	2.57	2.42	19.5	18.2	17.6	17.7	12.7	28.6	29.4	30.6	30.3	35.5
Cincinnati-Wilmington-Maysville, OH-KY-IN	2.58	2.53	2.52	2.53	2.40	18.0	17.1	16.9	17.2	15.0	30.0	31.8	32.3	32.1	37.7
Rochester-Batavia-Seneca Falls, NY	2.80	2.76	2.73	2.73	2.39	16.9	16.3	16.2	16.3	13.2	18.5	20.1	21.8	21.6	37.3
Louisville/Jefferson Co.–Eliz.–Madison, KY-IN	2.65	2.60	2.59	2.59	2.35	18.3	17.5	17.3	17.3	12.7	26.7	28.6	29.2	29.1	39.0
Raleigh-Durham-Ch. Hill, NC	2.56	2.46	2.47	2.48	2.34	15.1	13.5	14.0	14.3	13.0	29.3	33.7	33.7	33.1	39.5
Charlotte-Concord, NC-SC	2.59	2.49	2.48	2.49	2.32	14.1	12.5	12.7	12.9	10.7	27.6	31.8	32.3	32.0	39.4
Atlanta–Athens-Clarke Co–Sandy Spr., GA	2.52	2.43	2.42	2.43	2.31	11.5	10.3	10.4	10.6	9.2	29.6	33.5	34.2	33.7	39.2
Greensboro–Winston-Salem–High Point, NC	2.62	2.56	2.55	2.55	2.29	15.3	14.4	15.0	15.0	14.1	26.8	29.3	30.3	30.1	42.4
Greenville-Spartanburg-Anderson, SC	2.76	2.66	2.64	2.64	2.29	15.1	13.8	14.1	13.8	10.0	19.6	23.6	24.9	24.8	40.8

Table S1: **Rankings of 50 largest US metropolitan areas by change in nodal degree, 1991–2013.** The regions at the top of the table grew in the most connected manner, while those at the bottom grew with the most sprawl in recent years. The change from 1991–2013 is an estimate of the average for new intersections, calculated based on changes in the stock of intersections (i.e., their number and average properties).

	Mean nodal degree			% Degree 4+			% Deadend		
County	1993-97	2008-12	Change	1993-97	2008-12	Change	1993-97	2008-12	Change
All counties with parcel data	2.62	2.80	0.19	14.8	18.6	3.8	26.6	19.1	-7.5
Travis, TX	2.62	3.26	0.64	16.5	41.9	25.3	27.1	8.0	-19.1
Mecklenburg, NC	2.21	2.72	0.51	7.1	16.1	9.0	43.1	22.1	-21.0
Alachua, FL	2.56	3.00	0.44	14.7	25.9	11.3	29.3	13.0	-16.4
Iredell, NC	2.28	2.72	0.43	6.1	12.4	6.3	38.9	20.3	-18.6
Franklin, OH	2.56	3.00	0.43	9.5	17.0	7.5	26.5	8.6	-17.9
Pierce, WA	2.40	2.81	0.41	9.8	18.3	8.4	35.1	18.7	-16.3
Coweta, GA	2.23	2.64	0.41	6.4	10.8	4.4	41.6	23.4	-18.2
St Louis, MO	2.35	2.74	0.39	9.6	13.7	4.1	37.2	19.9	-17.2
Hinds, MS	2.61	2.99	0.39	17.5	20.7	3.2	28.5	10.8	-17.7
Ocean, NJ	2.66	3.04	0.39	16.4	30.9	14.5	25.4	13.3	-12.0
Broward, FL	2.67	3.05	0.38	15.9	30.7	14.8	24.7	12.9	-11.7
Orange, FL	2.63	3.01	0.37	12.8	24.4	11.6	24.8	11.9	-12.8
Union, NC	2.30	2.67	0.36	9.8	12.4	2.6	39.7	22.9	-16.8
Anoka, MN	2.68	3.04	0.36	15.2	26.6	11.4	23.7	11.4	-12.2
Leon, FL	2.54	2.88	0.34	11.7	20.3	8.6	28.6	16.0	-12.7
Clay, FL	2.56	2.90	0.34	14.0	24.8	10.7	28.8	17.3	-11.6
Thurston, WA	2.42	2.73	0.32	11.8	12.7	0.9	35.1	19.7	-15.4
Miami Dade, FL	2.83	3.14	0.31	19.3	31.2	11.9	18.4	8.6	-9.8
Harford, MD	2.56	2.86	0.30	13.7	22.2	8.6	28.9	18.1	-10.8
Fort Bend, TX	2.43	2.73	0.29	12.4	15.0	2.6	34.6	21.2	-13.4
Gaston, NC	2.36	2.64	0.29	11.7	21.8	10.1	37.9	28.6	-9.2
Escambia, FL	2.67	2.95	0.28	19.0	28.1	9.1	26.0	16.3	-9.7
Wake, NC	2.29	2.58	0.28	11.0	13.1	2.1	40.8	27.7	-13.1
Jackson, OR	2.59	2.87	0.28	13.5	26.9	13.4	27.2	19.8	-7.3
Duval, FL	2.37	2.65	0.28	10.3	14.6	4.3	36.5	24.9	-11.7
Hillsborough, FL	2.59	2.86	0.27	12.3	16.0	3.7	26.4	15.0	-11.5
Polk, FL	2.62	2.88	0.26	11.9	20.3	8.4	25.1	16.3	-8.8
Collier, FL	2.60	2.86	0.26	13.8	21.2	7.3	26.7	17.4	-9.3
Bay, FL	2.76	3.01	0.26	19.1	27.9	8.8	21.8	13.3	-8.4
Northampton, PA	2.79	3.04	0.24	17.4	27.0	9.6	19.0	11.7	-7.3
Pinellas, FL	2.62	2.85	0.23	15.5	24.8	9.3	26.7	19.8	-6.9
Seminole, FL	2.59	2.82	0.23	13.6	19.3	5.6	27.2	18.7	-8.5
Denver, CO	3.15	3.37	0.22	40.2	39.2	-1.0	12.7	1.1	-11.7
Palm Beach, FL	2.56	2.78	0.22	14.2	21.7	7.5	29.1	21.7	-7.3
Washington, MN	2.63	2.84	0.21	14.6	14.7	0.1	25.9	15.5	-10.4
Lake, FL	2.59	2.79	0.20	13.0	14.0	0.9	27.0	17.3	-9.7
Hennepin, MN	2.60	2.80	0.20	16.3	19.4	3.0	28.0	19.5	-8.5
Pasco, FL	2.61	2.80	0.19	12.1	13.1	1.0	25.7	16.4	-9.3
Onslow, NC	2.21	2.40	0.19	10.5	14.5	4.0	44.6	37.2	-7.3
Sarasota, FL	2.79	2.98	0.18	15.7	15.2	-0.5	18.2	8.7	-9.4
Okaloosa, FL	2.55	2.74	0.18	11.7	18.6	6.9	28.1	22.4	-5.7
Middlesex, MA	2.61	2.79	0.17	9.8	11.5	1.7	24.2	16.5	-7.7
Queens, NY	3.21	3.38	0.17	38.3	43.7	5.5	8.8	3.0	-5.7
Johnston, NC	2.31	2.47	0.17	8.2	14.0	5.7	38.8	33.3	-5.6
Brevard, FL	2.66	2.83	0.17	13.0	15.3	2.3	23.3	16.2	-7.1
Tarrant, TX	2.73	2.89	0.16	17.3	21.3	4.0	22.0	16.0	-6.0
Citrus, FL	2.98	3.13	0.15	20.3	22.3	2.0	11.2	4.7	-6.5
Collin, TX	2.82	2.96	0.13	22.1	20.7	-1.4	20.0	12.6	-7.4
Denton, TX	2.76	2.89	0.13	17.2	15.6	-1.6	20.8	13.4	-7.3
Volusia, FL	2.78	2.90	0.12	19.1	20.6	1.5	20.5	15.2	-5.2
Osceola, FL	2.79	2.90	0.10	19.6	17.7	-2.0	20.1	14.0	-6.1
Snohamish, WA	2.47	2.57	0.10	13.9	15.5	1.6	33.3	29.1	-4.2
Santa Rosa, FL	2.64	2.73	0.10	13.4	10.4	-3.0	24.8	18.5	-6.3
Delaware, OH	2.64	2.72	0.09	12.3	19.2	6.9	24.3	23.4	-0.9
Kitsap, WA	2.48	2.56	0.08	14.9	11.5	-3.3	33.5	27.7	-5.8
Dakota, MN	2.71	2.79	0.07	18.7	17.8	-0.9	23.6	19.5	-4.2
Polk, IA	2.78	2.84	0.07	24.6	24.8	0.1	23.5	20.3	-3.2
Alamance, NC	2.65	2.71	0.06	18.2	17.3	-0.9	26.6	23.0	-3.6
Clark, WA	2.55	2.62	0.06	15.2	17.0	1.9	29.9	27.7	-2.2
Essex, MA	2.68	2.73	0.05	10.0	17.0	7.0	21.2	22.2	1.0
St Johns, FL	2.44	2.48	0.04	13.5	13.0	-0.5	34.8	32.3	-2.5
Mesa, CO	2.62	2.66	0.04	16.0	19.6	3.6	26.9	26.7	-0.3
Riverside, CA	2.53	2.57	0.04	11.1	13.2	2.1	29.1	28.1	-1.0
Butler, OH	2.32	2.34	0.02	10.9	10.7	-0.2	39.7	38.5	-1.2
Spokane, WA	2.67	2.69	0.02	19.8	17.6	-2.2	26.6	24.5	-2.2
Charlotte, FL	3.00	3.02	0.02	16.2	19.4	3.2	8.2	8.8	0.6
Los Angeles, CA	2.68	2.69	0.01	19.0	19.3	0.3	25.5	25.0	-0.5
St Lucie, FL	3.01	3.02	0.01	19.5	25.4	5.9	9.1	11.7	2.7
Bristol, MA	2.60	2.60	0.00	10.5	9.1	-1.4	25.4	24.5	-0.9
King, WA	2.84	2.81	-0.03	17.3	23.1	5.7	16.6	20.8	4.2
Cameron, TX	2.95	2.92	-0.03	26.6	19.9	-6.6	15.7	13.8	-1.9
Kings, NY	3.51	3.48	-0.03	53.6	51.7	-1.9	1.1	1.7	0.6
Lee, FL	3.00	2.96	-0.04	21.1	24.3	3.2	10.7	14.3	3.6
Monroe, NY	2.56	2.52	-0.04	13.6	10.1	-3.5	28.8	29.2	0.4
Marion, FL	2.95	2.89	-0.06	20.7	19.2	-1.5	13.0	15.1	2.1
York, PA	2.75	2.68	-0.07	11.7	21.1	9.4	18.2	26.4	8.2
Plymouth, MA	2.80	2.73	-0.08	9.1	8.0	-1.1	14.4	17.7	3.4
Summit, OH	2.55	2.45	-0.10	13.9	11.9	-2.0	29.4	33.3	3.9
Indian River, FL	3.15	2.97	-0.18	30.4	16.9	-13.6	7.5	9.9	2.4
Manatee, FL	2.90	2.70	-0.20	21.6	15.8	-5.8	15.8	22.9	7.1

Table S2: **Rankings of counties with parcel data by recent changes.** Counties are ordered by the change in nodal degree of new development in 1993–1997 (when sprawl was at its peak) compared to 2007–12 (the most recent five-year period in our parcel-based series). Many of the counties at the top of the list, including five of the top six, have been the site of new regulations or plans to promote connected streets, at least in part of the county. Source: counties with parcel data, restricted to those with at least 100,000 population and at least 100 new nodes in each time period. An unweighted mean over individual years is used to construct the aggregated five-year periods.

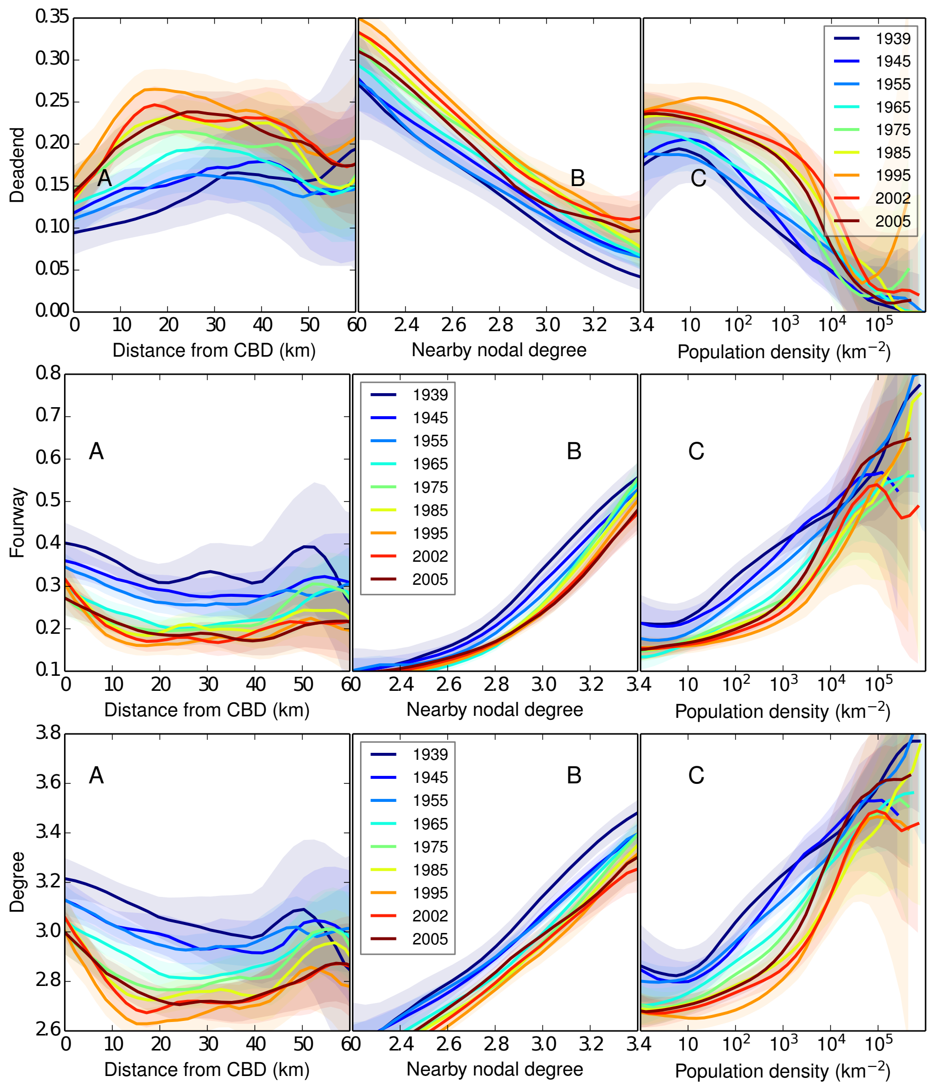

Figure S8: **Uniformity of shifts in sprawl.** Nonparametric estimates of the fraction of deadends (upper) and fraction of degree-four nodes (middle) and mean degree (lower, also in main text) as a function of (A) their distance from city center, (B) the mean nodal degree within 1 km, and (C) the local population density. Over time the relationships shift roughly uniformly and then reverse uniformly. Shaded bands show 95% confidence intervals.

S5 Online maps and video

Mean nodal degree of the entire road stock in 2013, and mean nodal degree of additions to the road stock since 1999, are plotted with census blockgroup resolution for a number of metropolitan areas, on an online supplementary data site, http://sprawl.ihsp.mcgill.ca/PNAS2015/bgmaps/.

Video animations of the street-by-street development of selected counties in our parcel dataset are available at http://sprawl.ihsp.mcgill.ca/PNAS2015.

S6 Robustness tests

Our core contribution rests on the proper identification of the construction date of each road intersection. To assess the robustness of our dating algorithm for our parcel-derived dataset, we consider a number of variations on our procedure for determining the date of nodes.

In general, we follow a two-step process. First, we assign a year to each edge, based on the year of the oldest building on that edge. Using the oldest building allows us to ignore the effects of recent development and rebuilding. Second, we assign a year to each node, based on the year of the most recent connected edge. This is because the connectivity of a node is determined by the most recent edge. For example, when a newly constructed street creates a 3-degree node by terminating at an existing road, the node did not exist prior to the construction of the most recent edge.

Possible concerns with this method are:

Measurement error: Year built information in the parcel data could be imperfect. Because we use extrema, single miscoded dates in the data we receive from counties would determine the year recorded for an edge.
We carry out sensitivity tests for the this problem by considering different points in the distribution of parcels' "year built" on each edge. Below we show values using the 2nd oldest, rather than the oldest, parcel on each edge. Similar results are obtained when using the 5th percentile.

Low parcel numbers: When edges are treated equally in determining node dates, small numbers of parcels on one edge can also cause a bias because the chance of them all being more recently rebuilt houses is higher.
We treat this issue by calculating a set of dates using only edges with five or more parcels on them.

In-fill and rebuilding: When most homes are of more recent vintage than the original road network, a reliance on parcel data becomes problematic.
We test against this third issue through our development of a time series using only TIGER vintage information. This TIGER (stock) series corroborates our main findings using the more detailed parcel-derived time series (see Figure 1 in the main text). Another rather strong test for the importance of redeveloped areas which did not affect the preexisting road

structure is to consider the oldest year built among all parcels on edges connected to a node.

Figure S9 presents national average time series for three alternative methods for calculating node age, incorporating the robustness tests described above. Also included is our baseline estimate, "Most recent", which is used in the main analyses and is shown here in black. Years calculated by the "Most recent (2nd oldest parcel)" method address the "measurement error" and "low parcel number" concerns: we drop all edges with fewer than 5 parcels, and we select the second oldest parcel on each edge to determine the date for the edge. As with our baseline method, the most recent edge is then used to characterize the construction date of the intersection.

The "Oldest" year calculation dates each node by the oldest parcel among all adjoining edges, which addresses the "infill and rebuilding" concern. However, this method is still subject to the other concerns, and also raises the extra problem that intersections created on existing roads (as in the example above) will not be dated correctly. Moreover, recent years may be biased towards deadends. For example, in 2012, the only degree-four nodes will be those where the range across edges is zero, i.e. both the oldest and most recent edge have a year of 2012.

Finally, the "3rd oldest" variant is a compromise between the latter two. It uses the date of the third oldest edge when there are at least three edges, which amounts to the same as our baseline estimate for degree-one (deadend) and degree-three intersections, but also provides a sensitivity check for the "infill and rebuilding" concern.

Figure S9 shows that all variants of our algorithm indicate a flattening out of road network sprawl in the mid/late 1990s. Moreover, there is strong consistency about the turnaround in recent years, with the exception of the last few years in our "Oldest" variant. The "3rd oldest" variant, which incorporates an extra robustness restriction, agrees closely with our baseline values. Our other qualitative observations appear also to be robust. Because deadends have only one edge, they are dated the same using the "Most recent" and "Oldest" methods. Thus, the difference in the fraction of deadends, shown in the lower right panel, reflects the difference in the denominator, driven by the number of degree-three and degree-four nodes assigned to each year.

A further robustness test involves restricting the streets considered in our analysis to those that have street names in the US Census Bureau TIGER/Line files. This can help to eliminate service roads, freeway ramps, driveways and similar streets from the dataset. Eliminating unnamed streets increases mean nodal degree by <0.05, and does not change any of the qualitative conclusions. The date of the turnaround in sprawl is unchanged.

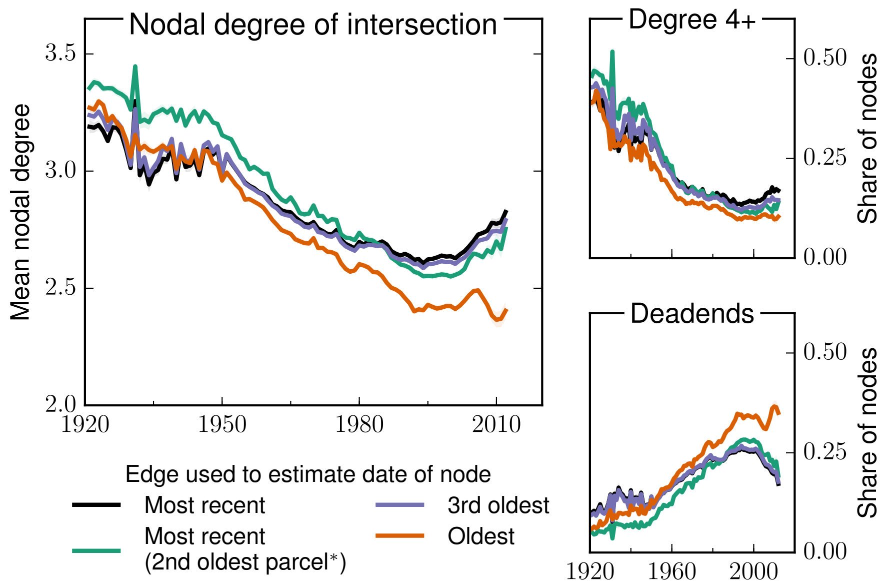

Figure S9: **Alternate methods to estimate the date of each node**. We assume that each edge was constructed at the time of the earliest parcel (building) on that edge, except as specified below. To estimate the year in which a node was built, we compare four methods. "Most recent" is our preferred measure, and is used in our analyses; the year of the most recent edge gives the year of the node. "3rd oldest" is the same as "most recent" for deadends and degree-3 nodes, but uses the year of the third oldest edge for degree 4+ nodes. "Oldest" uses the earliest year among the set of connected edges. "Most recent (2nd oldest parcel)" (*) is similar to "most recent," but the year of each edge is given by the second oldest rather than the oldest parcel. This last method only considers edges where $N_{parcels} \geq 5$, i.e. edges with at least five parcels with year-built information. See the text for an interpretation of the alternate methods.

S7 Further acknowledgements

We relied heavily on open source software tools and would like to acknowledge Matplotlib [15], Pandas [16], and Git (`http://git-scm.com/`). We are grateful to the following counties who kindly licensed free or discounted parcel data to us for research purposes: Anoka, MN; Athens, OH; Baker, FL; Bay, FL; Belmont, OH; Butler, OH; Carver, MN; Clark, WA; Clearwater, ID; Cowlitz, WA; Cumberland, NC; Dakota, MN; Defiance, OH; Delaware, OH; Denton, TX; Elmore, ID; Gaston, NC; Grant, WA; Hancock, MS; Haywood, NC; Hennepin, MN; Hillsborough, FL; King, WA; Kitsap, WA; Lake, IL; Lawrence, OH; Los Angeles, CA; Mason, WA; Milwaukee, WI; Monroe, NY; Moore, NC; Ottawa, OH; Pierce, WA; Pinellas, FL; Burke, NC; Ramsey, MN; Riverside, CA; Saratoga, NY; Scott, MN; Skamania, WA; Snohamish, WA; Spokane, WA; Spotsylvania, VA; St Louis, MO; Summit, OH; Tarrant, TX; Thurston, WA; Vanderburgh, IN; Walla Walla, WA; Warren, NY; Washington, MN; Wichita, TX; Wood, OH; Bronx, NY; Kings, NY; New York, NY; Queens, NY; Richmond, NY.

References

[1] Galster G et al. (2001) Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate* 12(4):681–717.

[2] Duany A, Plater-Zyberk E, Speck J (2001) *Suburban Nation: The Rise of Sprawl and the Decline of the American Dream*. (North Point Press, New York).

[3] Brueckner J, Fansler D (1983) The economics of urban sprawl: Theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics* 65(3):479–482.

[4] Brueckner JK, Mills E, Kremer M (2001) Urban Sprawl: Lessons from Urban Economics. *Brookings-Wharton Papers on Urban Affairs* (2001):65–97.

[5] Su Q, DeSalvo JS (2008) The Effect of Transportation Subsidies on Urban Sprawl. *Journal of Regional Science* 48(3):567–594.

[6] Song Y, Zenou Y (2009) How do Differences in Property Taxes Within Cities Affect Urban Sprawl? *Journal of Regional Science* 49(5):801–831.

[7] Burchfield M, Overman HG, Puga D, Turner MA (2006) Causes of Sprawl: A Portrait from Space. *The Quarterly Journal of Economics* 121(2):587–633.

[8] Glaeser EL, Kahn ME (2004) Sprawl and Urban Growth in *Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography*, eds. Henderson JV, Thisse JF. (Elsevier, Amsterdam) Vol. 4, pp. 2481–2527.

[9] Tanguay GA, Gingras I (2012) Gas price variations and urban sprawl: an empirical analysis of the twelve largest Canadian metropolitan areas. *Environment & Planning A* 44(7):1728–1743.

[10] Handy S, Paterson RG, Butler K (2003) *Planning For Street Connectivity: Getting from Here to There*. (APA Planning Advisory Service, Chicago).

[11] Xie F, Levinson D (2007) Measuring the Structure of Road Networks. *Geographical Analysis* 39(3):336–356.

[12] Ewing R, Cervero R (2010) Travel and the Built Environment. *Journal of the American Planning Association* 76(3):265–294.

[13] Hamidi S, Ewing R (2014) A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landscape and Urban Planning* 128:72–82.

[14] Jin S et al. (2013) A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sensing of Environment* 132:159–175.

[15] Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* 9(3):90–95.

[16] McKinney W (2010) Data Structures for Statistical Computing in Python in *Proceedings of the 9th Python in Science Conference*, eds. van der Walt S, Millman J. pp. 51 – 56.