CrossMark

A century of sprawl in the United States

Christopher Barrington-Leigh[a,1,2] **and Adam Millard-Ball**[b,1]

[a]Institute for Health and Social Policy and McGill School of Environment, McGill University, Montreal, QC, Canada H3A1A3; and [b]Environmental Studies Department, University of California, Santa Cruz, CA 95064

Edited by Susan Hanson, Clark University, Worcester, MA, and approved May 6, 2015 (received for review March 2, 2015)

The urban street network is one of the most permanent features of cities. Once laid down, the pattern of streets determines urban form and the level of sprawl for decades to come. We present a high-resolution time series of urban sprawl, as measured through street network connectivity, in the United States from 1920 to 2012. Sprawl started well before private car ownership was dominant and grew steadily until the mid-1990s. Over the last two decades, however, new streets have become significantly more connected and grid-like; the peak in street-network sprawl in the United States occurred in ∼1994. By one measure of connectivity, the mean nodal degree of intersections, sprawl fell by ∼9% between 1994 and 2012. We analyze spatial variation in these changes and demonstrate the persistence of sprawl. Places that were built with a low-connectivity street network tend to stay that way, even as the network expands. We also find suggestive evidence that local government policies impact sprawl, as the largest increases in connectivity have occurred in places with policies to promote gridded streets and similar New Urbanist design principles. We provide for public use a county-level version of our street-network sprawl dataset comprising a time series of nearly 100 y.

road network | urban sprawl | transportation | policy | climate

The planet's population is undergoing the last phase of becoming urbanized, a once-only process resulting from technological advance and the centralization of resources. However, urban development over the last century has increasingly taken the form of sprawl, characterized by low densities, spatially segregated land uses, and a street network with low connectivity. Although sprawl has been documented in Europe, Latin America, India, and China (1, 2), it is most often associated with postwar urban development in the United States.

A large body of empirical evidence links sprawl with greater vehicle travel, material use, energy consumption, and greenhouse gas emissions (3, 4). Indeed, urban economists, historically sympathetic to sprawl as a desirable market outcome, have begun to focus more on its negative externalities and on the agglomeration benefits of dense cities (5). (Other sprawl-related externalities such as a reduction in social capital may exist as well but are more contentious in the literature. See, for example, ref. 6.) To the extent that congestion, carbon, and other taxes on private vehicle travel are set inefficiently low, the private market will produce too much sprawl.

On the time scale of several decades, some characteristics of the physical layout of urban areas can change in response to infrastructure, prices, and migration. For instance, buildings can be reshaped or replaced, and new infrastructure and services can arise. However, residential roads tend to remain where they were first placed. London (1666) and San Francisco (1906) are just two examples where cities have been rebuilt on an almost identical street network following devastating fires or earthquakes (ref. 7, p. 227). As the Intergovernmental Panel on Climate Change notes, the long-lived nature of the built environment tends to lock in energy consumption and emissions once urbanization occurs (4).

Moreover, because high-density living requires more frequent access to services outside the home, low-connectivity road networks limit the extent to which residential and commercial land uses can change. As a result, areas with low-connectivity road networks will have a limited ability to adapt even in the face of rising fuel or carbon taxes. Meanwhile, there is wide variation in the degree to which extant urban areas sprawl, and understanding the influences, including possible future policies, on sprawl is key to evaluating and mitigating the possible "lock-in" effect of low-connectivity roads.

In the United States, given the doubling of fuel prices between the 1990s and mid-2014, policy efforts to promote smart growth and New Urbanism, and an apparent shift in consumer preferences toward urban living (8), one might expect an impact on new development. To date, however, the evidence has been mixed. Ramsey (9) reports that the share of infill housing construction increased in 2005–09 compared with 2000–04, and news reports announce the arrival of "peak sprawl" based on construction trends (10). In contrast, others (11) find that sprawl continued to increase, if only marginally, between 2000 and 2010. However, these studies usually rely on a comparison of just two or three time points, making it problematic to discern trends, and sprawl research in general has focused on describing and explaining cross-sectional differences in urban development in a single year.

Here, we provide a quantitative history of urban sprawl in the United States, as measured through the connectivity of the street network. We make three core contributions. First, we present, to our knowledge, the first high-resolution time series of sprawl from 1920 to 2012 based on our reconstruction of historical road networks for a substantial subset of US counties. It provides detail for small geographic areas and allows an unprecedented quantitative account of changes in urban form over the century. Using a complementary method that helps to validate our core results, we also develop a time series that covers the entire country but with lower time resolution and range. Second, we quantify the rise of sprawl in the urbanized United States since the early 20th century. We date the rise of sprawl to long before the private automobile became dominant and find that sprawl appears to have peaked in the mid-1990s. Importantly, because our measures are based on new urban streets, this turnaround is unlikely to be due to infill development on underused sites. Rather, today's newly built neighborhoods

Significance

Urban development patterns in the 20th century have been increasingly typified by urban sprawl, which exacerbates climate change, energy and material consumption, and public health challenges. We construct the first long-run, high-resolution time series of street-network sprawl in the United States. We find that even in the absence of a coordinated policy effort, new developments have already turned the corner toward less sprawl. Initial impacts on vehicle travel and greenhouse gas emissions will be modest given that the stock of streets changes slowly, but feedbacks are likely to mean that benefits compound in future years. Our publicly released data provide further opportunities for research on urban development and the social and environmental impacts of different urban forms.

Author contributions: C.B.-L. and A.M.-B. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Dryad Digital Repository, datadryad.org (dx.doi.org/10.5061/dryad.3k502).

[1]C.B.-L. and A.M.-B. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: chris.barrington-leigh@mcgill.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504033112/-/DCSupplemental.

appear to be less sprawling than their earlier counterparts. By directly identifying the properties of new construction, our metrics highlight the decisions that are driving change, even though the impact on the stock is gradual. Third, we provide evidence that urban sprawl is a persistent phenomenon—perhaps partly due to path dependencies in development decisions. There is a close correlation between the extent of sprawl in earlier time periods and that of contemporary development.

Measuring Street-Network Sprawl

We conceptualize sprawl as low connectivity in the street network. For a given geographic area, we construct measures of (*i*) mean nodal degree [i.e., the number of connected edges (incoming roads) at each intersection], (*ii*) the proportion of dead ends (i.e., nodes of degree one), and (*iii*) the proportion of nodes of degree four or more. Sprawl is characterized by a low nodal degree of intersections, a high proportion of dead ends, and a low proportion of nodes of degree four or more, all of which imply a street network with limited connectivity.

Our measures of sprawl, or related ones such as intersection density, are commonly used in urban planning and transportation research (12–15). However, the literature offers many alternative measures, such as density, contiguity of the built-up area, segregation of land uses, and urban design. Because sprawl is a multidimensional characteristic of urban areas, we discuss some of the alternative ways to operationalize it in *SI Appendix*, section S3.

We focus on street connectivity on theoretical and policy grounds. First, the connectivity of the street network is a semipermanent feature of the urban landscape and reflects decisions by cities and landowners at the time of initial development. Street rights of way are rarely vacated, so four-degree intersections usually remain that way. Opposition by homeowners fearing increased traffic, not to mention the costs of demolishing existing buildings, mean that dead-end streets also usually remain dead ends. In contrast, characteristics such as density tend to change over time in response to evolving prices, consumer preferences, and public policy. Second, street-network sprawl relates directly to important externalities such as greenhouse gas emissions and public health. Street connectivity is highly correlated with vehicle travel and modal split (3) and the incidence of diabetes, asthma, and similar health issues (16). Less connected streets increase the ratio of network distance to Euclidean distance, which reduces the generalized cost of driving relative to walking, and they are less conducive to pedestrian-oriented development and public transit service. Third, our measures of sprawl offer extremely high spatial and temporal resolution, rather than being constrained by the available geographic aggregation units, decadal gaps in census data, or the resolution offered by remote sensing technologies.

We therefore use "sprawl" as shorthand for "street-network sprawl" in the remainder of this article. *SI Appendix*, section S3 provides more analysis of how our measures of sprawl relate to

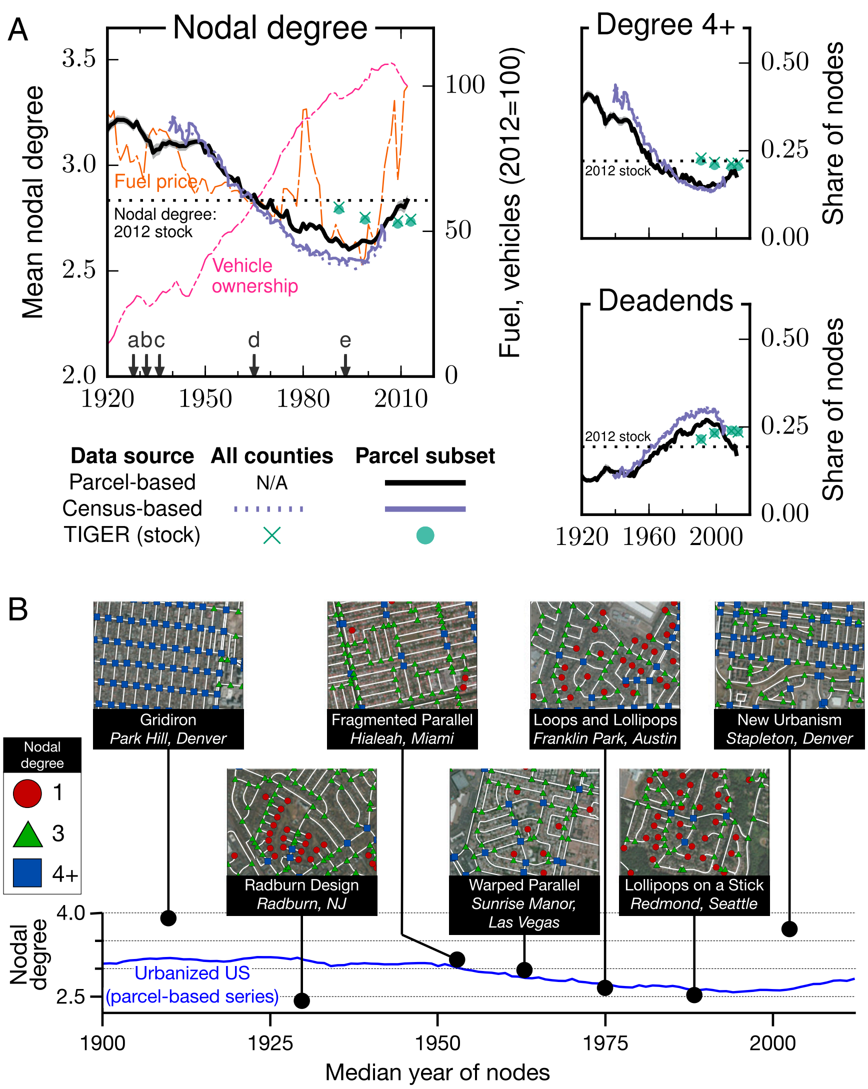

Fig. 1. Trends over time, US urbanized areas 1920–2012. Clearly evident are the rise in sprawl through most of the 20th century, the correlation with archetypal street designs, and the decline in sprawl since the mid-1990s. (*A*) The three measures of sprawl exhibit similar trends, with street networks becoming increasingly sprawl-like from 1950 through sprawl's peak in 1994. The 95% confidence intervals are shaded or too narrow to be discernible. Our preferred time series is parcel-based, represented by the solid black lines. As described in *Materials and Methods*, we validate our findings using two alternative time series, which show broad agreement. A 5-y rolling mean is used before 1950. Also indicated in the *Upper* panel are key policy events noted in ref. 17: (a) the Radburn design, (b) report by the Committee on Subdivision Layout, (c) report by the Federal Housing Administration, (d) report by the Institute for Transportation Engineers, and (e) founding of the Congress for the New Urbanism. (*B*) We identify empirical examples of the five archetypal street design patterns described in ref. 17 and show that the nodal degree of these examples generally matches the overall trends. Location names refer to the approximate neighborhood or city (e.g., Park Hill) and the metropolitan area (e.g., Denver). We also illustrate the 1928 Radburn design and the recent New Urbanist development of Stapleton, which represent opposite extremes in terms of street connectivity. A widespread move toward New Urbanism would eventually restore levels of sprawl to early 20th century levels. Underlying images courtesy of ESRI/Digital Globe.

vehicle ownership and travel, and how they correlate with alternative metrics such as residential density.

Sprawl's Rise and Decline

The *Upper* panel of Fig. 1 shows the trends over time for each of the three measures of sprawl (nodal degree, fraction of dead ends, and fraction of nodes of degree four of more). Several conclusions are immediately evident.

First, Fig. 1 indicates a rise in sprawl since the mid-1920s, with an acceleration after 1950. The early beginning of sprawl is notable, given that it predates the postwar era of mass car ownership. However, it provides quantitative evidence to confirm historical accounts that date the emergence of cul-de-sacs and similar departures from gridiron street patterns to the early to mid-20th century. Southworth and Ben-Joseph (17), for example, note the influence of the 1928 design, with cul-de-sacs prominently featured, for Radburn, New Jersey; they also point to the influence of recommendations for cul-de-sacs in reports by the Committee on Subdivision Layout (1932), Federal Housing Administration (1936), and Institute for Transportation Engineers (1965). These discrete events do not capture the more gradual evolution in street network design from the 1950s through the early 21st century, but our results closely match the archetypal patterns reported in ref. 17 and illustrated in the *Lower* panel of Fig. 1.

Second, there is a clear peaking of new sprawl construction in the mid-1990s and a subsequent decline since 2000 to the level of the 1960s. Mean nodal degree rose from ∼2.60 at sprawl's 1994 peak to ∼2.83 in 2012. Although a reversal in street connectivity trends might be expected at some point in response to changes in fuel prices, the 1994 peak predates the post-2000 rise in gasoline prices. Conversely, the ∼1980 spike in fuel prices was not associated with a similar reversal in sprawl. An alternative possibility is that, just as the 1928 Radburn design was associated with the initial rise in sprawl, the recent move toward more connected street patterns reflects the growth in New Urbanist thinking and policy since the Congress for the New Urbanism was founded in 1993. One prominent New Urbanist development, Stapleton, has a mean nodal degree of ∼3.47 (Fig. 1*B*).

Our results are in contrast to recent findings (18) that street-network sprawl continues to increase, albeit at a slower rate than before (11). However, results in refs. 11 and 18 measure only the stock of streets (which we also illustrate in Fig. 1), whereas our method is sensitive to the year-by-year developments. Our results report a major turnover and reversal in the new contributions to that stock before the turn of the century. Thus, we identify two important turning points. In ∼1994, the nodal degree of new intersections (the flow) reached its minimum. Due to the existence of cities with dense, gridded cores, the road network stock was still tending toward more sprawl until ∼2012, when the nodal degree of new intersections rose to the level of the stock.

The trends are mirrored in individual metropolitan areas. The four Combined Statistical Areas (CSAs) shown in Fig. 2 are illustrative only, but a similar pattern is evident in other metropolitan regions, reported in *SI Appendix*, Fig. S6. In all cases, nodal degree falls most rapidly from the 1950s through the mid-1990s (ending earlier in the Minneapolis–St. Paul and Washington, DC regions) and has risen since the start of the 21st century. The differences between the metropolitan areas are most evident in terms of the level of sprawl (New York and Miami being less sprawling than Seattle and Los Angeles) rather than the relative trends. As discussed in detail in *The Dynamics and Persistence of Sprawl*, there is evidence of persistence in relative levels over time. The New York–Newark region, for example, is endowed with a historic stock of highly connected streets, and additions to this stock in almost every year are less sprawl-like than the other metropolitan regions illustrated.

Spatial Patterns

With urban form quantified at the level of individual intersections, we can generate a complete account of the dynamics of sprawl over space and time. Fig. 3*A* shows three snapshots of postwar development for one illustrative region, the Minneapolis–St. Paul metropolitan area. Outside the 1950 core, degree three intersections became the dominant road form before 1980. The aggregate distribution of nodal degree values over time is shown in Fig. 3*B*, along with the overall volume of construction. Road edges connected to at least one degree four intersection are prevalent until the mid-1950s, when the proportions of dead ends and degree three nodes rise rapidly. A regrowth in the fraction of degree four nodes (at the expense of degree three and dead ends) is visible starting around 2000, before a steep decline in street construction following the housing market crash of 2007–08.

The maps in Fig. 3 do not emphasize the location of recent construction or changes in urban form. A second approach to help to understand where the changing development style is occurring, both within and between metropolitan areas, is shown in Fig. 4. It depicts the most recent levels of nodal degree, averaged to census block groups for selected major metropolitan areas. *SI Appendix* provides a similar view of new additions to the stock in recent years at the block group level, as well as snapshots of levels (stocks) in other years.

Blue areas, with high nodal degree, are characterized by the most grid-like road networks, and red and dark red represent the dead end and degree three neighborhoods characterizing sprawl. There are stark contrasts in accumulated development patterns that defy simple geographic generalizations. Many major cities have urban cores with a highly gridded structure, whereas some, like Atlanta, have very little. Most interestingly, the changes, shown in *SI Appendix*, Fig. S7, in mean nodal degree between 1991 and 2013 suggest recent trends that are not predicted simply by the stocks shown in Fig. 4, nor by a portrait of the stocks as they were in 1991 (*SI Appendix*, Fig. S7). Development in the suburbs of Seattle, the San Francisco Bay Area, and Dallas have shown significant increases in nodal degree, whereas the metropolitan area of Atlanta appears to have continued its embrace of low-connectivity, cul-de-sac road networks. In Boston, development patterns appear to have been different in the northern suburbs (higher nodal degree) than in the western areas. It should be noted that these maps disproportionately emphasize large (low-density) block groups, and much of the fine detail is not resolved in Fig. 4. Maps of a large number of urban areas are linked in *SI Appendix*.

We now turn our attention to a larger spatial scale and consider aggregate-level differences between metropolitan regions and counties. *SI Appendix*, Table S1 ranks US metropolitan areas

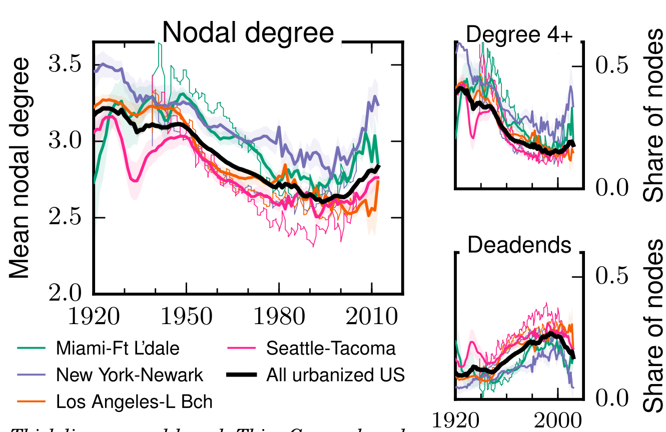

Fig. 2. Trends over time, selected metropolitan areas. Trends at the metropolitan area level largely mirror those for the United States as a whole. Data are for CSAs designated by the US Census Bureau. We focus on results from our parcel-based dataset (thicker lines, with 95% confidence intervals shaded), which only provides partial coverage of each CSA. However, similar results are obtained using our Census-based dataset (thinner lines), which is shown for comparison and covers all counties in a given CSA. Note: Before 1980, a 5-y rolling mean is used.

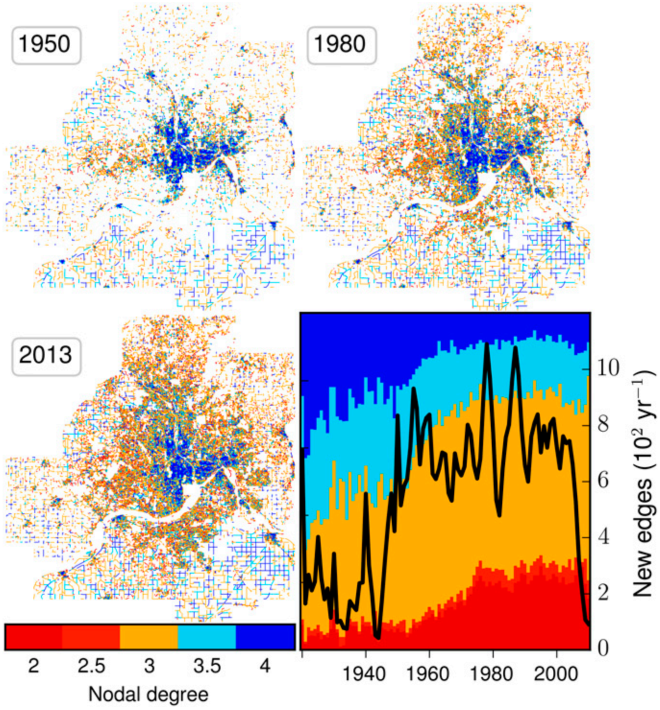

Fig. 3. Spatial and temporal patterns of sprawl in the Minneapolis–St. Paul region. Individual edges—that is, road segments bounded by two intersections—are shown at three time points. Edges are colored in five categories according to their connectivity, ranging from highly connected (gridded) in blue to cul-de-sacs in red. Connectivity is measured by the mean degree of an edge's two terminal intersections, explained in the text. Because nodes can be cul-de-sacs, degree three, or degree four-plus, there are five possible values of edge degree, ranging from 2.0 to 4.0. In 1950, the developed area is largely gridded, but growth by 1980 and by 2013 is largely of the low-connectivity kind. Rural roads also tend to be gridded. The *Lower Right* panel shows the fraction, indicated by the vertical extent of a color, of each edge type built each year. The black line shows the pace of construction, defined as the number of edges dated to each year. Dramatic drops are evident during the Depression, World War II, oil shocks, a recession in the 1970s and 1980s, and the recent Global Financial Crisis. We focus on Minneapolis–St. Paul because all seven central counties are included in our parcel-based data and because the region closely tracks national trends (*SI Appendix*, Fig. S6).

according to the change in nodal degree between 1991 and 2013. Many of the "usual suspects," such as San Francisco toward the top and Atlanta and Charlotte toward the bottom, occupy their expected positions. For example, the rankings support the impression from Fig. 4 that Atlanta has continued to pursue low-connectivity development. However, there are some surprises, most notably high rankings for Dallas–Fort Worth, Texas; Oklahoma City, Oklahoma; and Birmingham, Alabama—not normally well-known as policy environments seeking to reduce private car use.

In the case of Dallas, the rankings do provide suggestive evidence for the impact of antisprawl policies. The 1998 City of Dallas Comprehensive Plan, for example, requires residential neighborhoods to be "served by a grid street system, which minimizes the use of cul-de-sacs" (ref. 19, p. 9). Elsewhere, the rankings lack a clear link to land-use regulations, and places with long-standing (pre-2003) policies to discourage or prohibit cul-de-sacs and promote connected streets, such as Portland, Oregon; Austin, Texas; Charlotte, North Carolina; and Cary, North Carolina (the latter being in the Raleigh–Durham metropolitan area) (20, 21), lie in the middle to bottom of the rankings in *SI Appendix*, Table S1.

Most street connectivity policies, however, are undertaken at the municipal level. Absent a concerted metropolitan- or state-wide effort (such as that in Virginia, which enacted statewide standards in 2009 that strongly discourage cul-de-sacs), local-level policies are unlikely to influence the metropolitan-wide rankings. Moreover, our rankings are based on changes in the level of the stock, using our Topologically Integrated Geographic Encoding and Referencing (TIGER)-based series, which will respond to policy changes only slowly. Therefore, *SI Appendix*, Table S2 ranks counties according to the change in the nodal degree of new construction since sprawl reached its mid-1990s peak, using our parcel-based series.

Here, there is more suggestive evidence for the impacts of antisprawl policies at the local level. The county with the largest increase in nodal degree is Travis, Texas, where the principal city (Austin) has promoted more connected streets—initially through individual developments, such as the New Urbanist airport reuse plan, and more recently at a citywide level. The second-ranked county, Mecklenburg, North Carolina, is home to the city of Charlotte, which as noted above has long-standing street connectivity policies. Although the Charlotte region as a whole may still be sprawling (*SI Appendix*, Table S1), city-level regulation appears to be making a difference on a smaller scale. In Alachua, Florida (ranked third), the city of Gainesville adopted in 1999 a Traditional Neighborhood Development overlay zone that prohibits cul-de-sacs in the areas where it is applied. Gainesville is also home to several prominent New Urbanist developments such as Haile Village Center, Tioga, and Bryton. In Franklin County, Ohio (ranked fifth), the City of Columbus adopted a New Urbanist Traditional Neighborhood Development ordinance, whereas in Pierce County, Washington (ranked sixth), the largest city (Tacoma) has policies in its General Plan and development code that strongly discourage cul-de-sacs.

Such anecdotal evidence of formalized policies can be expected to represent a broader and underlying trend in design ideals and objectives, just as earlier development styles were sometimes formalized into codes and bylaws. Nevertheless, our findings here are suggestive only, and this simple analysis does not formally

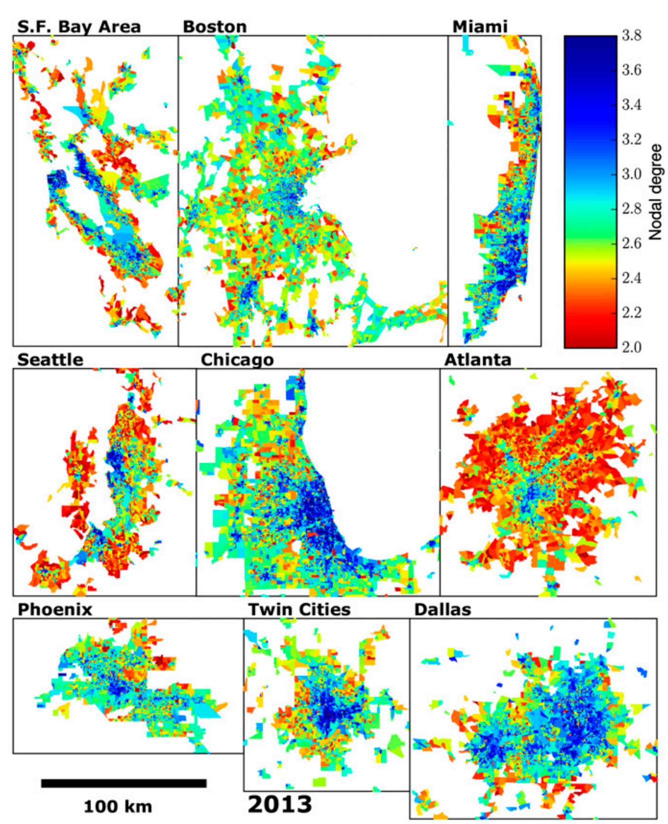

Fig. 4. Mean nodal degree in selected metropolitan areas. We find stark variation across metropolitan areas both in the stock and (shown in *SI Appendix*) in recent construction. Mean nodal degree of street networks is shown for census block groups of selected metropolitan areas in 2013.

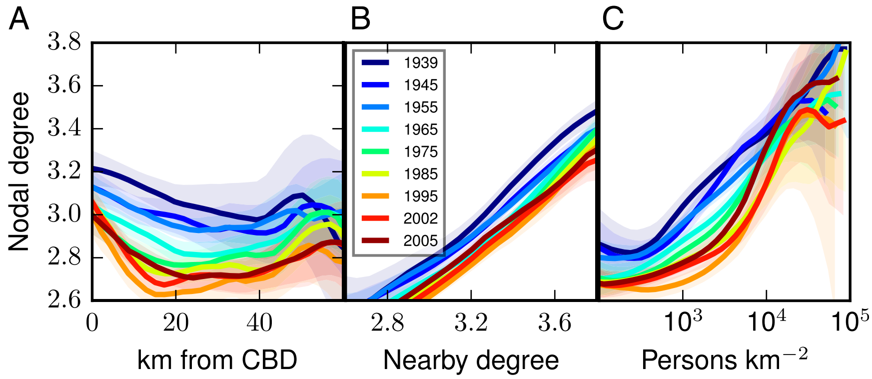

Fig. 5. Uniformity of shifts in sprawl. Nonparametric estimates of the connectivity of roads (mean degree of intersections) as a function of their distance from city center (*A*), of the mean nodal degree within 1 km in 2013 (*B*), and of the local population density (*C*). Over time the relationships fall roughly uniformly and then rise again. Shaded bands show 95% confidence intervals. Values are national averages from our parcel dataset.

quantify the role of local policies—not least, due to the lack of a comprehensive database on zoning regulations. Moreover, other factors clearly affect the connectivity of both the stock and new construction. For example, Fig. 4 and *SI Appendix*, Tables S1 and S2 suggest some persistence over time. Counties and regions that were sprawling in the past continue to develop in a similar manner. The following section explores the theoretical basis and additional empirical evidence for this phenomenon.

The Dynamics and Persistence of Sprawl

What light can the full power of a spatial time-series of urban form shed on the dynamics of sprawl? Our nation-wide data provide evidence of remarkable persistence in differences across regions, simultaneously with roughly parallel shifts in development patterns in different regions over time.

The results in Fig. 5*A*, showing that nodal degree generally falls with distance from the city center, come as no surprise, given the spatial association between sprawl and suburbia. However, it is remarkable that the spatial gradient of street connectivity has remained relatively constant since 1939. Although sprawl was rising until ~1994 and declining thereafter, similar changes have occurred in city centers as in exurbs. A similar dynamic is in evidence when considering the gradient of sprawl against nearby development (Fig. 5*B*) and residential density (Fig. 5*C*). In principle, the changes in mean degree of road networks that we find in recent years could be due to a different pattern of where new intersections are built—for instance, as more infill development occurs in dense, urban cores with connected streets in adjacent neighborhoods, as shown in ref. 9. Changes in the amount of infill notwithstanding, our findings indicate that the decline in sprawl is also due to a different style of road network being built across a range of urban contexts. In other words, the changes cannot be explained simply by a new focus on infill in the city center but rather reflect a broader shift in development patterns across the entire metropolitan area.

Fig. 5 also suggests that there is persistence in relative terms in sprawl. In other words, places that were built with a low-connectivity street network tend to stay that way, even as the network expands. We examine persistence directly in Fig. 6. Metropolitan regions that had a sprawling street-network stock in 1991 experience the greatest level of sprawl for new construction in 1999–2013 (Fig. 6*A*). In general, the most sprawling regions in 1991 such as Atlanta and Charlotte continued in that vein in more recent years, as did regions at the opposite end of the sprawl spectrum, such as Dallas–Fort Worth. An even stronger relationship is seen at the county level (Fig. 6*B*), where our parcel-based series provides better temporal resolution. Furthermore, geographic variation in development patterns is persistent across even longer time periods; the development decisions that were taken more than 50 y ago are highly predictive of contemporary new development. Near the extreme, Denver, Colorado (home to the New Urbanist Stapleton development shown in Fig. 1*B*) was largely gridded in 1992, and virtually no dead ends were built in 2008–12 (Fig. 6*B* and *SI Appendix*, Table S2). In our view, this persistence highlights the importance of the turnaround reported here, both because the turnaround is likely to be permanent and because it is despite large inertial influences.

These correlations between past and present sprawl may be due to the persistence of physical, geographic, and political factors, such as topography and political attitudes toward private car use. The correlations, however, may also indicate some path dependence. Lower density, car-oriented development offers greater returns for developers if it matches the prevailing pattern of development. Conversely, it makes less sense to build a walkable neighborhood if there is nowhere to walk to.

Conclusions

The quasi-permanence of roadways means that urban development decisions have effects that last for generations. The historic gridded centers of US cities and the narrow, winding streets of European medieval towns are still in place today (17), and the low vehicle travel and emissions of cities like San Francisco and New York are largely due to the fact that their street networks were laid down before the private car became dominant. Conversely, sprawl today—in the form of street networks with low connectivity and high proportions of dead ends—will lock in vehicle travel and emissions for decades to come.

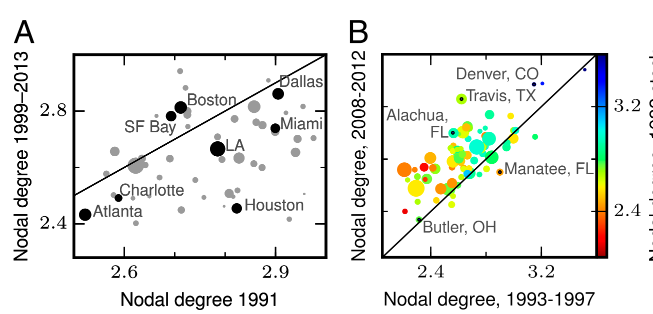

Fig. 6. Persistence of sprawl. (*A*) Nodal degree of new development, 1999–2013, against nodal degree of the stock (1991), by CSA. Labeled points are highlighted in a darker shade. Most metropolitan regions lie below the 45° line, indicating that the sprawl of the stock increased between 1999 and 2013, but as discussed in *Sprawl's Rise and Decline*, this is consistent with a turnaround in the connectivity of new construction given that the stock includes many gridded neighborhoods built before the era of mass car ownership. Data (TIGER-based series) are the same as *SI Appendix*, Table S1. (*B*) Nodal degree of new development, 2008–12 versus 1993–97. These time periods represent, respectively, the most recent years in our parcel-based dataset and the time when sprawl was at its peak in ~1993–97. Colors denote the stock of sprawl in 1992 and demonstrate the persistence of sprawl; counties that had high nodal degree in 1992, and also in the 1993–97 period, were more likely to continue to build connected streets in 2008–12. Also, almost all counties lie above the 45° line, indicating a turnaround in the connectivity of new development. Data (parcel-based series for a subset of counties) are the same as *SI Appendix*, Table S2.

In this paper, we present a unique, geographically disaggregated, long-run time series that quantifies the rise of sprawl in the United States from the early 20th century, its acceleration from the 1950s, its peak in the mid-1990s, and its subsequent decline. Although the peak and decline are apparent across the country, we find tentative signs that the decline in sprawl is most pronounced where local governments have adopted policies to improve the connectivity of the street network—for example, by prohibiting or discouraging cul-de-sacs. Moreover, we find that the connectivity of recently built streets is strongly associated with the connectivity of the earlier stock. In other words, early patterns of street connectivity may influence the nature of recent development: Sprawl begets sprawl.

The impacts of low-connectivity street networks on vehicle travel and emissions are well-documented (*SI Appendix*, Fig. S5) (3). Thus, the impacts of a turnaround in sprawl are likely already being felt, and it is notable that street-network sprawl peaked just a decade before per-capita travel demand reached a maximum in the United States in 2005 (22). In other words, peak sprawl is one potential contributor to the "peak travel" phenomenon. Moreover, the persistence and path dependence of shifts in urban form will have implications for the energetics and greenhouse gas emissions of future inhabitants of suburban neighborhoods. Just as the existing stock of locked-in sprawl from the mid- to late 20th century represents an enormous inertia, newly developed, connected street patterns will continue to affect vehicle travel and emissions for the next century and beyond. Path dependence implies that street connectivity has a secondary effect through influencing the connectivity of future streets. Thus, although we do not quantify greenhouse gas emissions impacts in this article, feedbacks are likely to mean that reductions compound in the future. Emission scenarios that adopt a short time horizon and/or fail to account for path-dependence processes are likely to underestimate the climate policy potential of land-use and transportation strategies.

The local policies—in particular, ones directly targeting the nodal degree of intersections—which we have highlighted as contributing to less sprawling construction in some areas, can be seen as just one element in a package of policies to promote denser, mixed-use, connected development patterns. Pursuit of this agenda can shape the fundamental infrastructure and incentives that guide future sustainable urban development pathways, both in the United States and in fast-growing cities around the world.

Materials and Methods

We generate three different time series of sprawl. Each series uses the most recent vintage of TIGER/Line files from the US Census Bureau to characterize the current road network but estimates the historical development of the network in a different way using (*i*) earlier vintages of the TIGER/Line files, (*ii*) the American Community Survey, or (*iii*) tax records for individual land ownership parcels. Because we are interested in urban sprawl, we limited our results to urbanized areas, defined as block groups where the majority of blocks were classified as urban in the 2010 Census. *SI Appendix*, section S1 provides more details of data sources and our algorithms for constructing the three series:

- The TIGER/Line series computes our measures of sprawl for all counties in the United States using four different vintages of the TIGER/Line shapefiles, corresponding to the street network in 1991, 1999, 2009, and 2013.
- The Census-based series is constructed through assigning the median year built of residential units in each census block group to all streets in that block group.
- The Parcel-based series is constructed from tax records for individual land ownership parcels. 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. The 226 counties in the parcel-based series account for 9.7% of the 2,338 counties and county equivalents in the United States with at least one urbanized block group and a higher (32.7%) share of the urbanized area population. *SI Appendix*, Fig. S1 shows the spatial distribution of the counties in our parcel-based series.

In short, the different time series all rely on the 2014 vintage of the TIGER/Line files but use different data sources to reconstruct the historical development of the street network through estimating the year in which each network edge was built. In general, we use the parcel-based series to report our main results, given the high resolution and length of the series, and because (unlike the Census-based series) it does not make assumptions about homogeneity of construction dates within a census block group. We rely on the TIGER/Line and Census-based series to validate our findings and assess the extent to which the parcel-based series provides results that are representative of the entire United States.

Note that the TIGER/Line series reflects the characteristics of the stock of streets in a given year. The other two series reflect the construction of new streets in a given year—that is, additions to the stock.

Data series in tabular, geographic, and graph theoretic formats; an interactive map explorer of summary data; and road evolution videos are available online at sprawl.ihsp.mcgill.ca/PNAS2015. Downloadable data are also archived at dx.doi.org/10.5061/dryad.3k502.

ACKNOWLEDGMENTS. We thank John Ford, Drew Natuk, Parker Welch, Tabitha Fraser, and Chris Thomas for excellent research assistance and Danny Cullenward, Kevin Manaugh, and Navin Ramankutty for comments on an early draft. We also thank numerous counties (listed in *SI Appendix*) for providing parcel data. This work benefited greatly from the open source tools Pandas, Matplotlib, Git, and their dependencies. This work was partially funded by grants from the Canadian Social Sciences and Humanities Research Council, the Sustainable Prosperity Network, and the University of California Santa Cruz Committee on Research.

1. Siedentop S, Fina S (2012) Who sprawls most? Exploring the patterns of urban growth across 26 European countries. *Environment and Planning – Part A* 44(11):2765–2784.
2. Seto KC, Sánchez-Rodríguez R, Fragkias M (2010) The new geography of contemporary urbanization and the environment. *Annu Rev Environ Resour* 35(1):167–194.
3. Ewing R, Cervero R (2010) Travel and the built environment. *J Am Plann Assoc* 76(3): 265–294.
4. Seto K, et al. (2014) Human settlements, infrastructure and spatial planning. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ Press, Cambridge, UK).
5. Glaeser E (2011) *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier and Happier* (Macmillan, Basingstoke, UK).
6. Brueckner JK, Largey AG (2008) Social interaction and urban sprawl. *J Urban Econ* 64(1):18–34.
7. Fradkin PL (2005) *The Great Earthquake and Firestorms of 1906. How San Francisco Nearly Destroyed Itself* (Univ of California Press, Berkeley, CA).
8. Nelson AC (2013) *Reshaping Metropolitan America. Development Trends and Opportunities to 2030* (Island Press, Washington, DC).
9. Ramsey K (2012) *Residential Construction Trends in America's Metropolitan Regions* (United States Environmental Protection Agency, Washington, DC).
10. Badger E (2013) Have we reached peak sprawl? *Atlantic Cities*. Available at www.citylab.com/housing/2013/10/have-we-reached-peak-sprawl/7102/. Accessed May 19, 2015.
11. Hamidi S, Ewing R (2014) A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landsc Urban Plan* 128:72–82.
12. Parthasarathi P, Hochmair H, Levinson D (2015) Street network structure and household activity spaces. *Urban Stud* 52(6):1090–1112.
13. Knaap GJ, Song Y, Nedovic-Budic Z (2007) Measuring patterns of urban development: New intelligence for the war on sprawl. *Local Environ* 12(3):239–257.
14. Guo Z (2009) Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transp Res Part D Transp Environ* 14(5):343–352.
15. Clifton K, Ewing R, Knaap GJ, Song Y (2008) Quantitative analysis of urban form: A multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* 1(1):17–45.
16. Marshall WE, Piatkowski DP, Garrick NW (2014) Community design, street networks, and public health. *Journal of Transport & Health* 1(4):326–340.
17. Southworth M, Ben-Joseph E (2003) *Streets and the Shaping of Towns and Cities* (Island Press, Washington, DC), 2nd Ed.
18. Giacomin DJ, Levinson DM (2015) Road network circuity in metropolitan areas. *Environ Plann B Plann Des*, 10.1068/b130131p.
19. City of Dallas (1998) *City of Dallas Comprehensive Plan Vol. I: Goals and Policies* (City of Dallas, Dallas).
20. Handy S, Paterson RG, Butler K (2003) *Planning for Street Connectivity: Getting From Here to There* (APA Planning Advisory Service, Chicago).
21. Litman T (2014) *Roadway Connectivity in TDM Encyclopedia* (Victoria Transport Policy Institute, Victoria, Canada).
22. Millard-Ball A, Schipper L (2011) Are we reaching peak travel? Trends in passenger transport in eight industrialized countries. *Transp Rev* 31(3):357–378.

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.

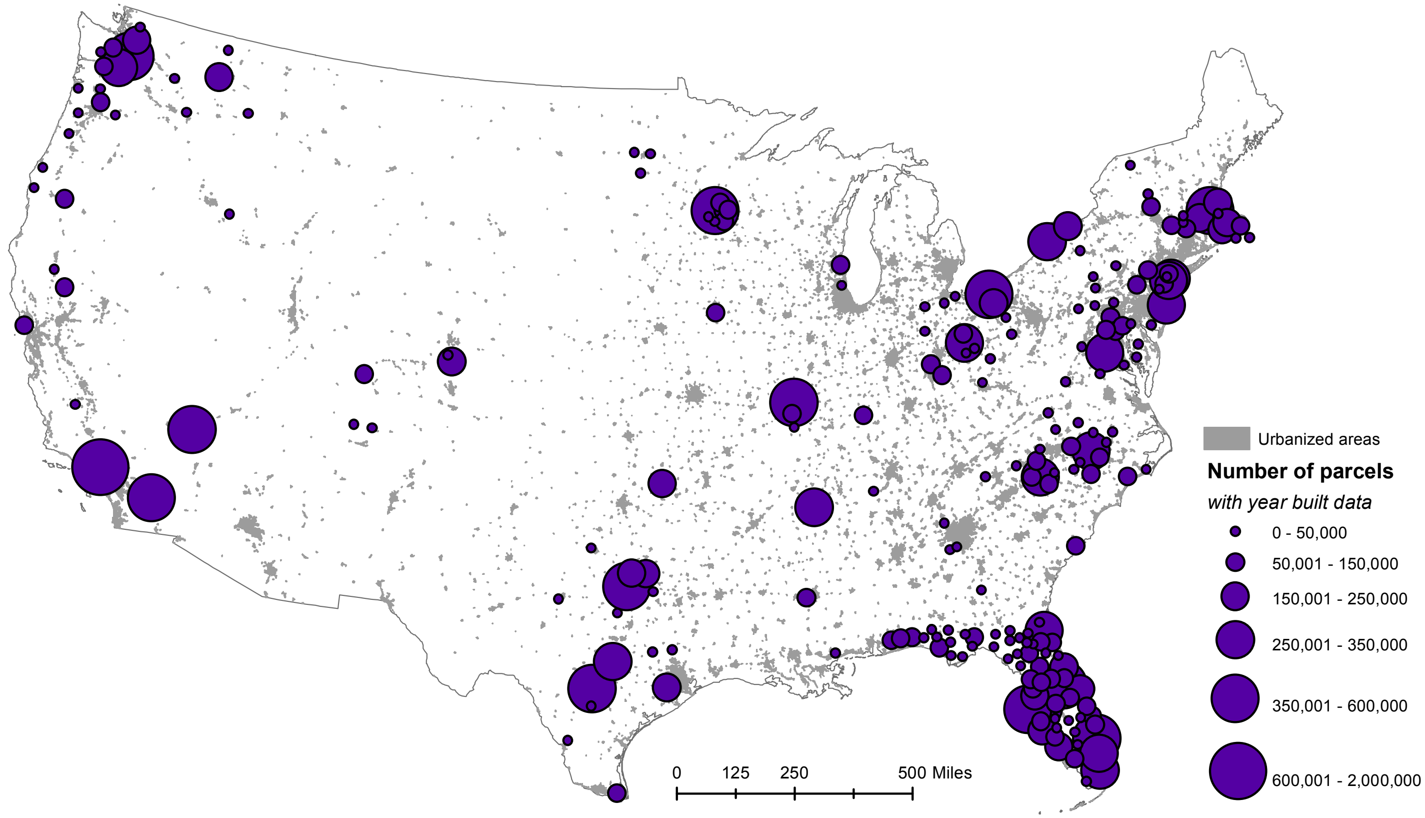

Figure S1: **Distribution of counties with parcel data.** We obtained year-built information for buildings in 226 counties with urban areas, broadly spanning the US and accounting for ∼33% of the urbanized area population.

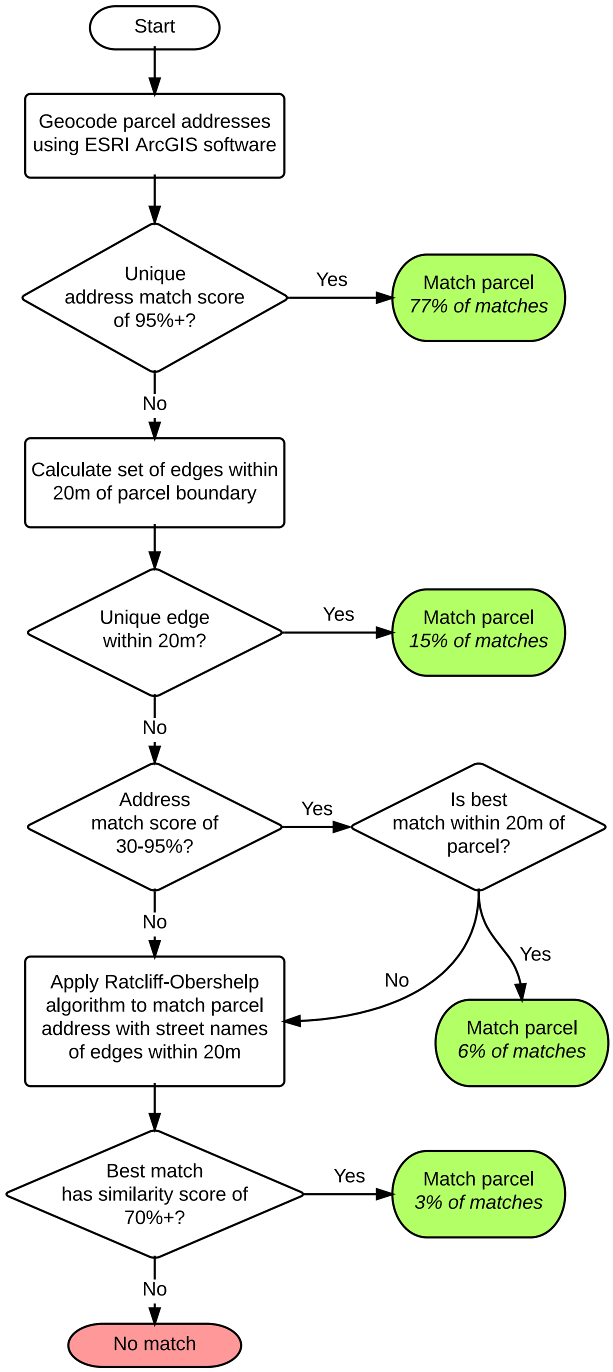

Figure S2: **Algorithm to match parcels to street edges.**

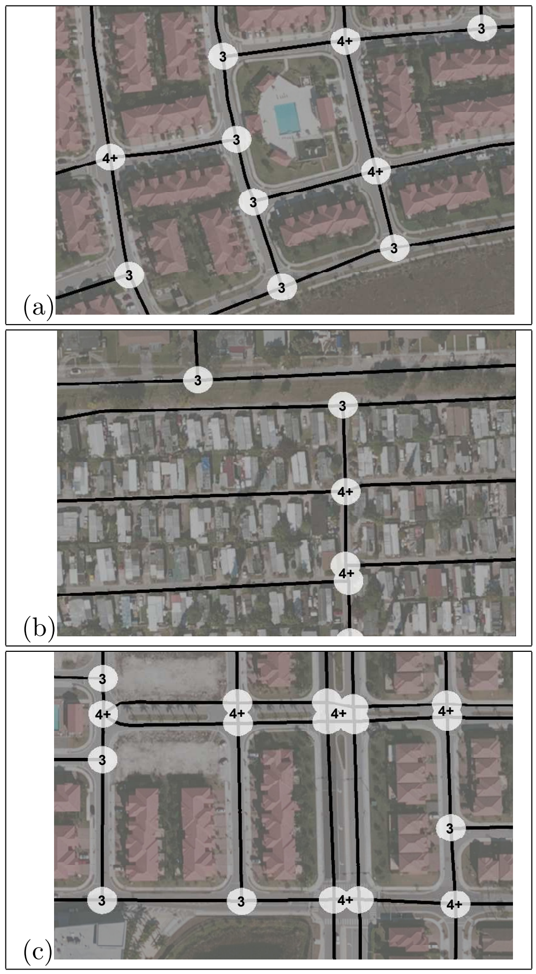

Figure S3: **Calculation of nodal degree.** Each geometric node is buffered (7.5m radius), and overlapping buffers are merged to create our dataset of nodes. In the simple case (a), calculated nodal degree is simply the number of connected edges at each geometric node. Where intersections are offset (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.

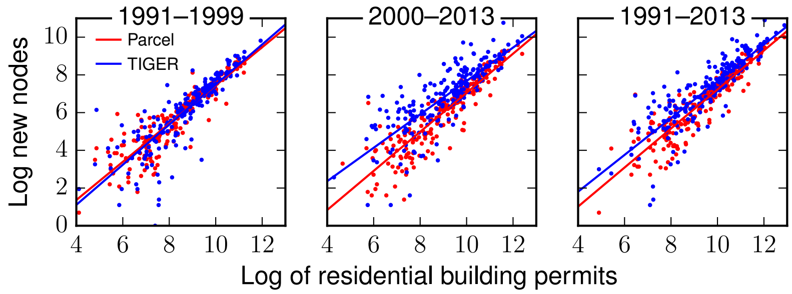

Figure S4: **Growth in nodes versus building permit issuance**. Comparison is for 1991–2013 in counties in our parcel-based series (N=224).

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.

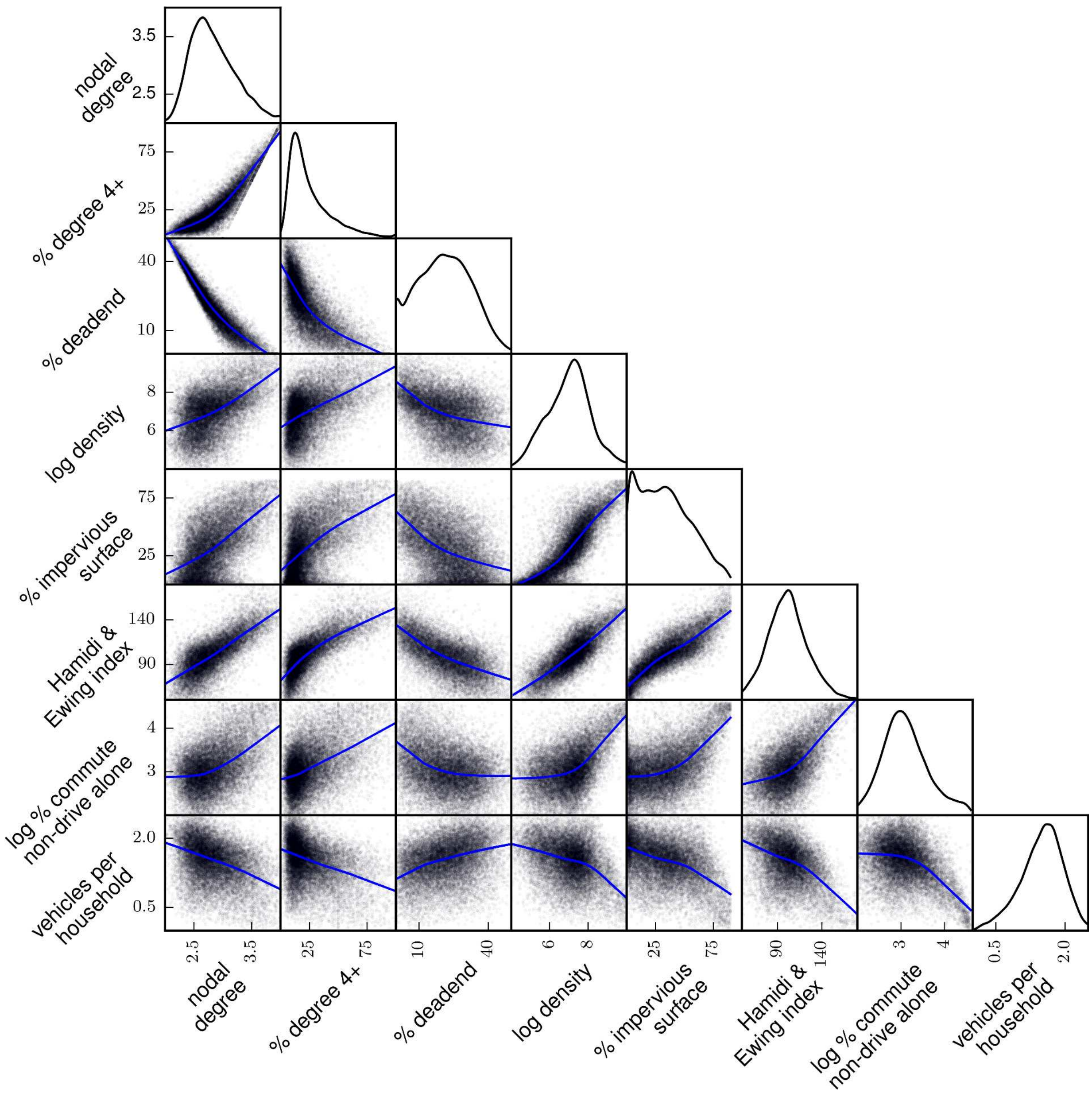

Figure S5: **Correlations between alternative measures of sprawl.** Log density, commute mode share and vehicle ownership are calculated based on the American Community Survey 2007-11. Impervious surface area is calculated based on the National Land Cover Database 2006 [14]. Hamidi & Ewing sprawl index 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.

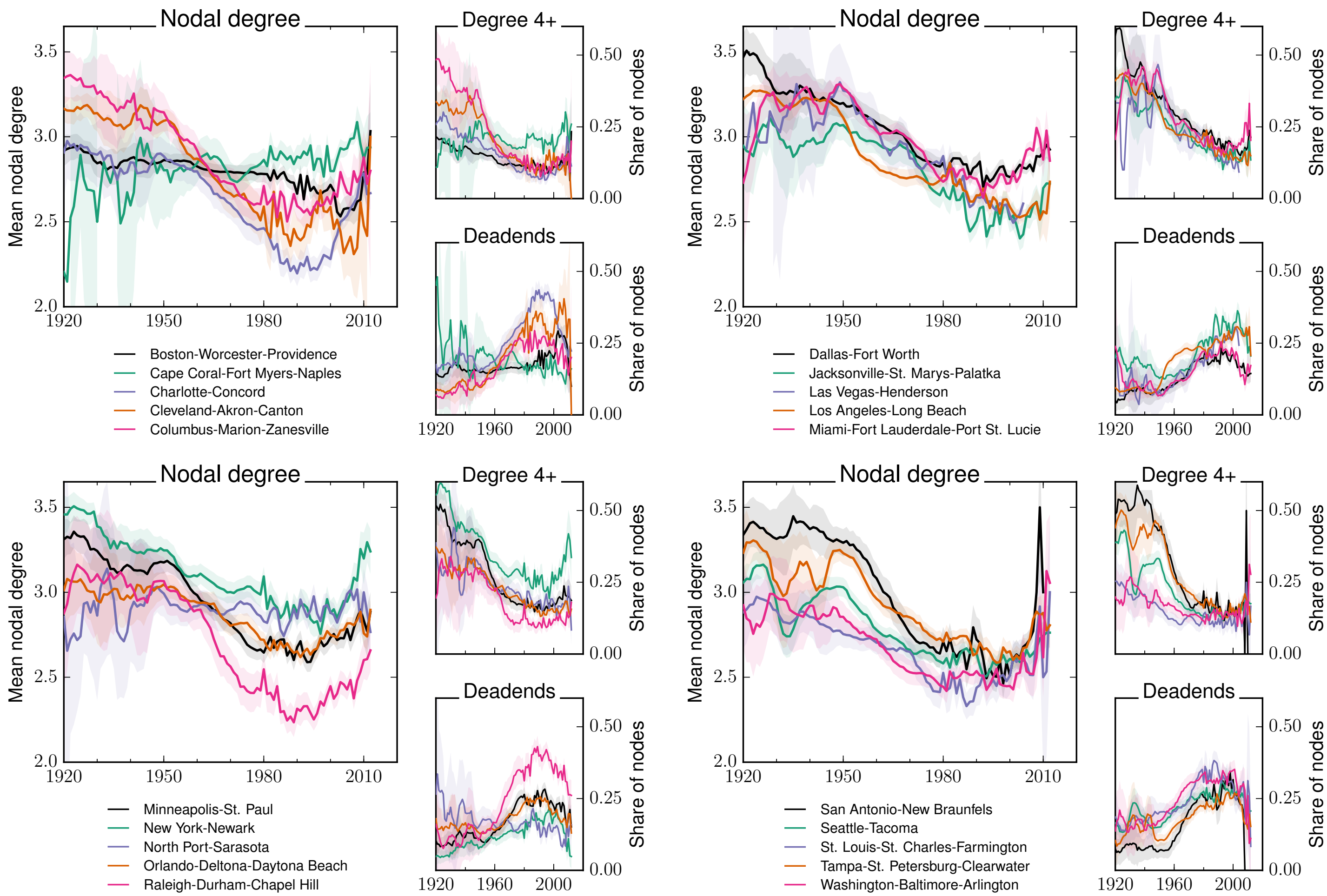

Figure S6: **Trends over time, metropolitan areas.** Figure 2 in the main text provides results for selected metropolitan regions. Here, we show results for the 20 largest metropolitan regions (Combined Statistical Areas as designated by the US Census Bureau) in our parcel-based dataset, as measured 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.

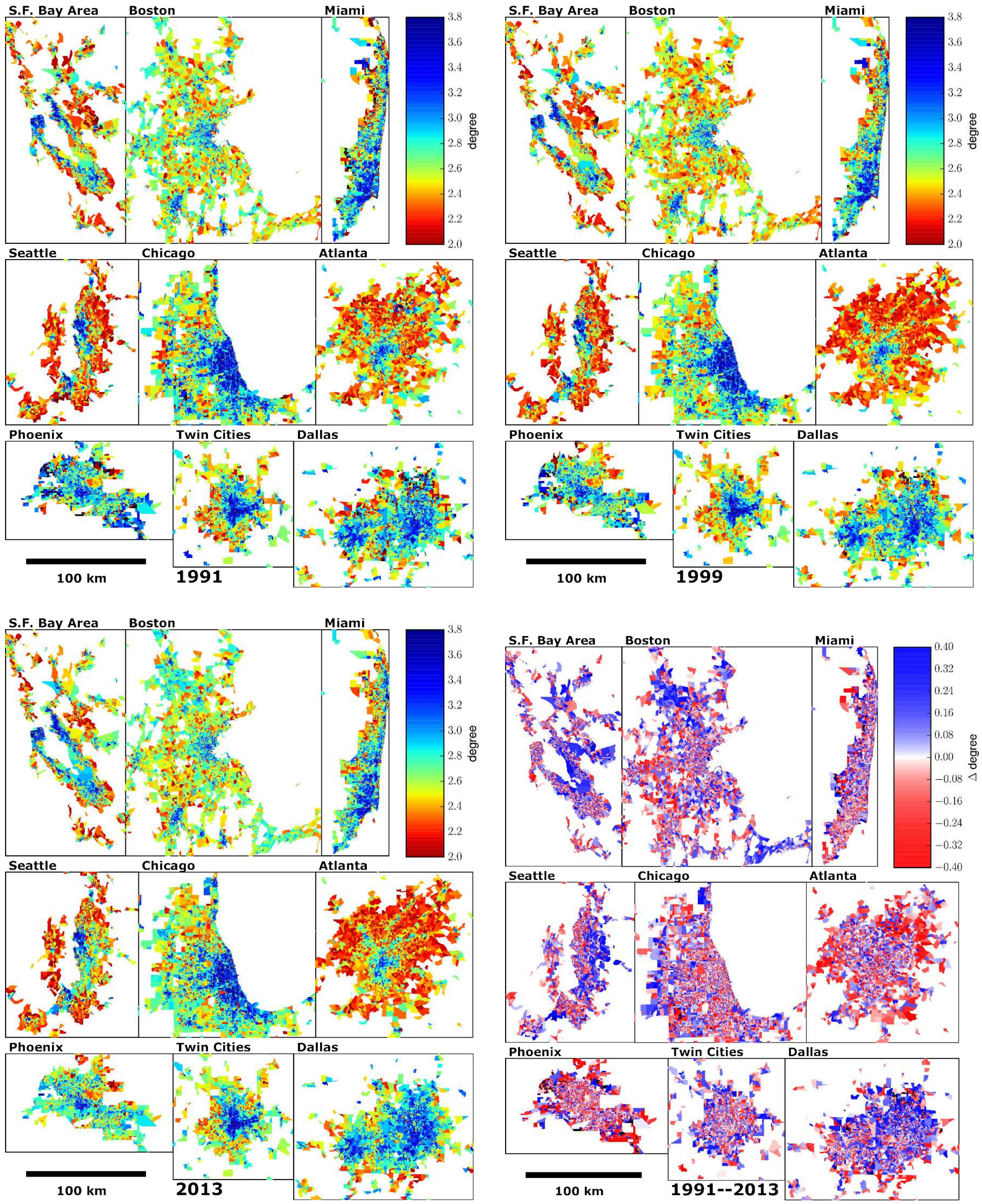

Figure S7: **Levels and changes over time in mean nodal degree in selected metropolitan areas**. Levels in mean nodal degree are shown at three points in time (top left, top right, and bottom left). The third plot (2013) is the one featured in Figure 4 of the main text. The bottom right panel shows the 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.

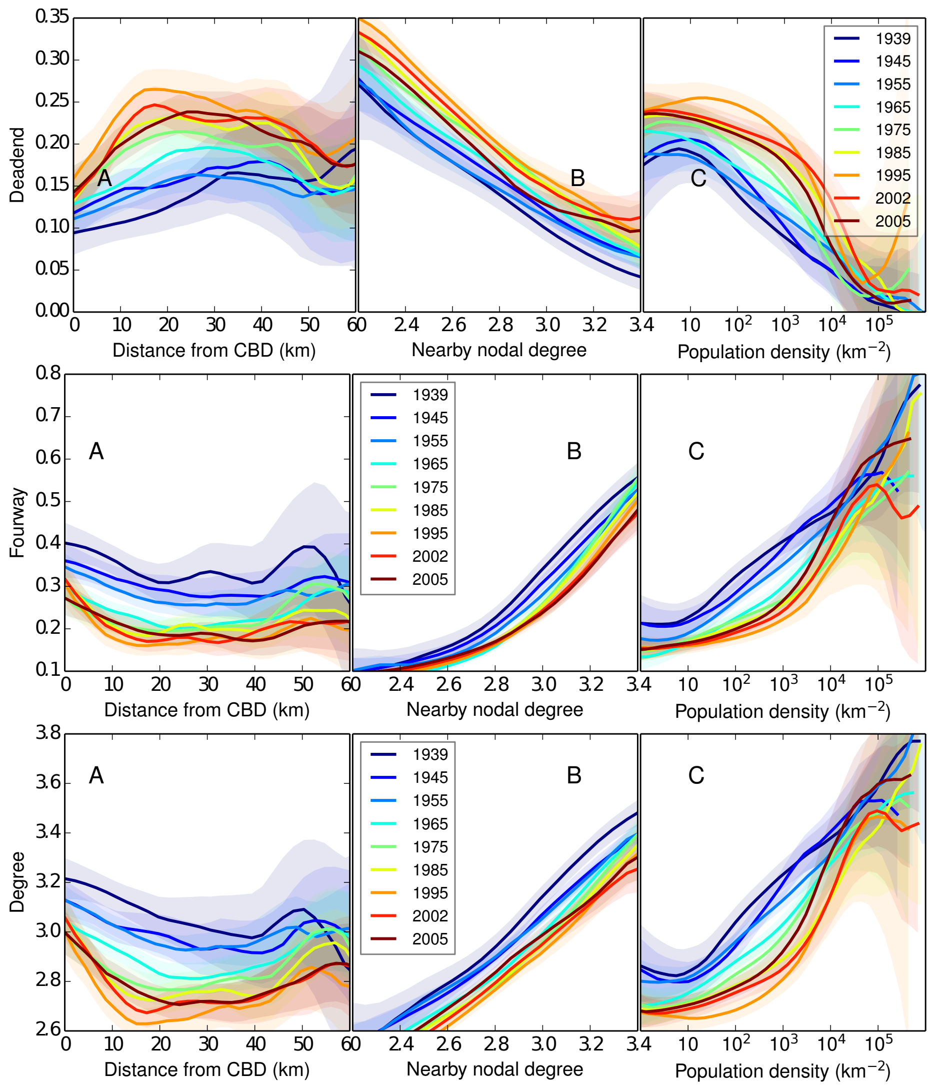

Figure S8: **Uniformity of shifts in sprawl.** Nonparametric estimates of the fraction of deadends (upper) and fraction of degree-four nodes (middle) and mean degree (lower, also in main text) as a function of (A) their distance from city center, (B) the mean nodal degree within 1 km, and (C) the local 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.

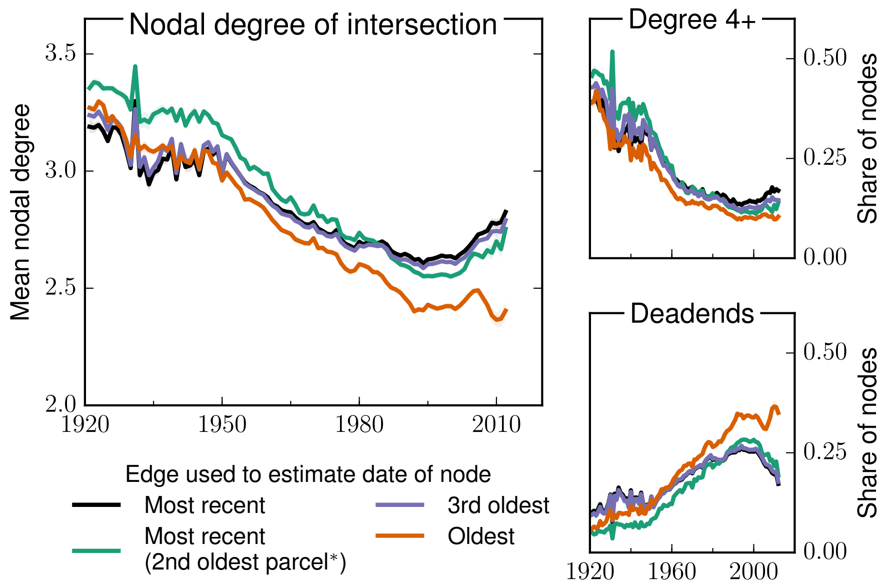

Figure S9: **Alternate methods to estimate the date of each node**. We assume that each edge was constructed at the time of the earliest parcel (building) on that edge, except as specified below. To estimate the year in which a node was built, we compare four methods. "Most recent" is our preferred measure, 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.