

Road Network Evolution in the Urban and Rural United States Since 1900

Keith Burghardt (✉ keithab@isi.edu)

USC Information Sciences Institute

Johannes Uhl

Institute of Behavioral Science, University of Colorado Boulder

Kristina Lerman

USC Information Sciences Institute

Stefan Leyk

Institute of Behavioral Science, University of Colorado Boulder

Research Article

Keywords: Road Network Evolution, Urban and Rural United States, human settlements mediating pollution

Posted Date: October 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-957212/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Computers, Environment and Urban Systems on July 1st, 2022. See the published version at <https://doi.org/10.1016/j.compenvurbsys.2022.101803>.

Abstract

We examine a key component of human settlements mediating pollution and congestion, as well as economic development: roads and their expansion in cities, towns, and villages. Our analysis of road networks in more than 850 US cities and rural counties since 1900 reveals significant variations in the structure of roads both within cities and across the conterminous US. Despite differences in the evolution of these networks, there are commonalities: newer roads tend to become less grid-like. These results persist across the rural-urban continuum and are therefore not just a product of urban growth. These findings illuminate the need for policies for urban and rural planning including the critical assessment of new development trends.

Background

Road networks provide the infrastructure for population growth and economic development. While previous studies have explored the evolution of urban road networks (1–5), the growth of suburban and rural road networks has been studied to a lesser degree, which challenges our understanding of the drivers of road evolution. To address this knowledge gap, we study road evolution across the rural-urban continuum of the conterminous US (CONUS) since 1900.

Researchers have studied how roads change over time using photography (6), satellite derived data (5, 7), census records and OpenStreetMap (3), census data, property tax records, and other geospatial data (4). Some have also used historical sources such as old maps (8–10), although these approaches are still in their infancy due to an often incomplete spatial and temporal coverage of catalogued maps and the difficulty of converting large-volumes of complex, low-quality images into analyzable data formats (11). Thus, existing approaches are limited in their temporal range and geographic coverage, do not account for regional variation when performing longitudinal analysis, and often fail to address scale effects, manifested in the modifiable areal unit problem (MAUP) (12, 13).

To address these challenges, we analyze road network data from the 2018 National Transportation Dataset (14), integrated with spatial data layers containing historical built-up areas and building densities since approximately 1900 (15, 16). We reconstructed historical road networks under plausible assumptions that roads were built at roughly the same time as the oldest nearby houses (see Methods). This integrated dataset enables us to study settlements and their changes through a road network lens at unprecedented temporal granularity and spatial resolution (17, 18). From these networks, holding over nine million nodes, and over fifteen million road segments, we extract several road network statistics, such as the mean degree (the number of roads at each intersection), road density (the kilometers of road per unit area), a local griddedness metric, and the orientation entropy of road segments (1), among others, to characterize and quantify key characteristics of urban development within spatial units of different granularity (e.g., metropolitan areas, counties, grid cells). Our analysis reveals a wide variance in these statistics, both within and between urban and rural areas. Moreover, these statistics vary markedly over time. However, this variance is far from random, with similar regions exhibiting similar statistical

trends. There are also common patterns in the growth of these networks, which points to common policies, including a reduction in the gridiron structure of newer roads, a structure associated with more walkable neighborhoods (1, 19). Notably, these trends persist along the rural-urban continuum; therefore, they are not strictly a product of urban growth. While the causes, drivers and implications of these trends are not fully understood, these findings offer new insights for rural and urban planners who aim to understand their neighborhood in context and develop appropriate policies, as well as modelers who aim to understand and predict development (20, 21).

Results

We carry out longitudinal and cross-sectional studies of the evolution of the US road networks since 1900, at spatial scales ranging from neighborhoods to the continental US. An example of the reconstructed road network is seen in Fig. 1 (more detailed examples can be seen in Supplementary Figs. S1 & S2). The Denver metropolitan area is seen as of 1900 (Fig. 1a), 1950 (Fig. 1b), and 2015 (Fig. 1c). The evolution of a subsection of the metropolitan area can be seen in Fig. 1d, with lighter colors denoting older roads, and darker colors representing newer ones. Data completeness is shown in Supplementary Figs. S3 & S4.

We begin our analysis on Office of Management and Budget-defined metropolitan statistical areas (MSAs) and micropolitan statistical areas (μ SAs), which allows results to be compared against previous work (2, 4). Collectively, these are known as Core Based Statistical Areas (CBSAs) (22). Figure 2 demonstrates how each statistic reveals interesting differences in road network structure between, as well as within, metropolitan areas (MSA and μ SA boundaries are shown in Supplementary Fig. S5). For example, azimuth variety (i.e., the variety of unique road orientation angles) is high (representing very irregularly oriented roads) in the Baltimore Washington area and low (regular roads) in the periphery of the Denver and Los Angeles metropolitan areas. Similarly, strong variations can be seen for the local griddedness metric, which is a spatial complement to the clustering coefficient often used in network analysis (2,23). It is defined as the number of city blocks (or more generally roads that form quadrilaterals) that meet at an intersection divided by the maximum number of city blocks that can meet at such an intersection (see Methods).

We observe low local griddedness (fewer square blocks) and low edge density (spaced out) suburbs, in mountainous regions west of Baltimore and Washington, DC, and Denver, as well as northern and eastern Los Angeles (see SI for more detail). Griddedness (and road density) is high, however, in the oldest parts of a city. Thus, our analysis reveals nuances in the level of griddedness within urban areas, as gridiron road structures appear in select parts of the city while other parts may have very different network characteristics. This is partly due to the fact that road networks in US cities have grown in very different ways. For example, across Los Angeles we observe local road networks that existed early in downtown Los Angeles or along the coastline that grew together over time, while the Washington DC region has comparatively old cities, and therefore roads were already widespread by 1900. The Denver area, in contrast, began in a tight-knit region around downtown, and moved outward progressively.

At the national level, Fig. 3 reveals broad trends in how road density varies across U.S. census regions ((24); henceforth referred to as regions). We find relatively low road density in the Midwest as well as the Northeast although the distribution in the Northeast shows a pronounced broad tail due in large part to the New York City MSA (Fig. 3a). The mean degree (i.e., mean number of roads per intersection) of networks, in contrast, is typically higher in the Midwest and lowest in the more mountainous regions (e.g., the Appalachian mountains) near the East coast (Fig. 3b). In agreement with these statistics, we find that orientation entropy, a proxy of the road network's regularity, is lowest (most regular) in the Midwest and highest (least regular) in the South and Appalachia. Complementing these observations, local griddedness and mean degree are highest in the Midwest and lowest in the South and mountainous regions in the West. These general results agree with previous work (3), where it was found that higher road gradients reduce the likelihood of gridiron road structure, by accident or design. Our temporal analysis, meanwhile, reveals that across 115 years some regions, such as the South and West, have seen great changes in their road networks, while the Northeast has been relatively stable due to limited additions of new roads in an already developed region (see Fig. 3c, Supplementary Fig. S6). In general, however, newer road networks tend to be less grid-like and less densely packed as they expanded into suburban areas (Fig. 2), similar to previous findings (3,4).

The evolution of road networks

Figure 4 analyzes the evolution of road networks within metropolitan areas based on network statistics extracted for each 1 × 1 kilometer grid cell, illustrating how networks have evolved since the first recorded building was constructed. For this analysis, we use several metrics including the ratio of dead end roads, the number of intersections per kilometer of road, nodes and length of road per area unit, and mean degree. We assess their correlation with the age of each grid cell, as shown in Fig. 4a (clusters for all of these metrics are seen in Supplementary Fig. S7). Broadly speaking, density- and griddedness-related metrics decrease over time, while azimuth variety shows mixed trends during the study period, and the dead end rate increases with road network age (correlations between these statistics and time are shown in Supplementary Fig. S8). These trends are similar for large cities (MSAs) and smaller cities (μ SAs) (Supplementary Fig. S9). Fig. 4b–c shows the spatial distributions of correlation coefficients between age and the two metrics azimuth variety and local griddedness, respectively, revealing strong spatial patterns. We try capturing this spatial variation by computing k-means clusters of temporal patterns in these statistics (25). We measure similarity between the time series of the CBSAs using the Dynamic Time Warping (DTW) distance metric (26). This metric yields large distances for time series that differ considerably in their trend, shape, and/or timing. We find the data series can be grouped well into just 3-4 clusters (indicated by the “elbow” in the DTW-based cluster inertia in Fig. 4d), a finding robust to how much low temporal and spatial coverage data is removed (see Supplementary Figs. S9, S10, & S11). Moreover, the other metrics shown in Fig. 4a tend to exhibit less pronounced spatial patterns in correlation and cluster analysis (Supplementary Figs. S7). We separately analyze MSAs and (μ SAs) to understand how their evolution is affected by the size of the urban area, as CBSAs of similar age can evolve very differently. These results reveal a Simpson's paradox (27) in that the trends in disaggregated

data differ from the overall trends shown in Supplementary Figs. S7, S8, & S10 and in previous work (3, 4). When mapping the computed clusters as shown in Fig. 4e–f, we find that the identified “types” of road network evolution at the city level follow strong spatial patterns. For example, we find that nearby cities in the Appalachian region or in the Northeast have similar trends in their azimuth variety. Similarity in griddedness trajectories for MSAs splits fairly cleanly between the East and the West. The temporal patterns for these clusters are shown in Fig. 4g–h, where azimuth variety grows fast in MSAs of clusters 1, 2, and 3, and decreases for cluster 4 (roughly covering Appalachia and the Northeast) where starting values were highest. In contrast, griddedness in MSAs decreases most in cluster 4 (West, Midwest) where values are highest and slower elsewhere (including the coastal regions).

While these univariate trends provide interesting insight, how do these metrics vary time jointly? The multivariate trajectories of CBSA-level network statistics over time are visualized in Fig. 5a, where we embed statistics for each city into two dimensions using UMAP (28). These embeddings are based on seven statistics computed for each CBSA and smoothed over time: the proportion of dead ends, mean degree, road distance per area, log of road distance, local griddedness and entropy (1), and proportion of intersections with four or more roads (details on data smoothing are seen in Supplementary Figs. S12 & S13). Changes in statistics are highlighted by radar charts computed for the Chicago, Washington, DC, Boston, and Denver MSAs. In Fig. 5c, we also show how these statistics vary across the UMAP projection, thus providing insight into the trends of individual cities. Our results demonstrate broad similarities but also considerable variation in city-level trends over time. Cities across but also within regions differ in their routes to their final statistics, yet again pointing to Simpson’s paradox in our data: trajectories, disaggregated to the level of a metropolitan region can differ, sometimes substantially, from any assumed overall trend, regionally or nationally. Nonetheless, we see some trends are consistent across cities, as shown in Fig. 5b, such as lower road density, fewer roads per intersection, and statistics consistent with less irongrid-like roads (although orientation entropy has recently started to decrease, implying more regular angles between intersections). Results are robust to data cleaning (Supplementary Fig. S14). These plots illustrate the heterogeneity in the evolution of cities that resulted in today’s urban areas of the US.

Figure 5b, meanwhile, reveals changes in road network statistics over time. There are general trends of increasing entropy and decreasing road density and griddedness across the US. We observe significant variance in the computed statistics across cities in early times, especially with road density and local griddedness. However, there are also notable commonalities, such as a tendency for newer regions to have lower griddedness and higher entropy (although entropy’s trend is non-linear). The trends for mean degree and mean griddedness reveal decreasingly grid-like networks over time (statistical significance of results are shown in Supplementary Fig. S11). While scholars have argued that grid structures enable efficient traffic flows and thus, may contribute to reduce emissions, congestion, and to increase the use of alternative, environmentally friendly transportation methods (1, 3, 19, 29, 30), more recently developed road networks appear to be less effective in that regard. But why? Some trends may be due to expanding urban areas towards non-flat topographies, such as in the mountains north of downtown Los Angeles or the Piedmont region of western Maryland (see Fig. 2), where grid-like road networks and high road

densities are not feasible. Residential development in these topographically more complex areas but also altered development patterns in periurban areas may help explain the decline in urban densities (31, 32), but future work needs to analyze these hypotheses in greater detail.

Differences in statistics between small and large cities (Supplementary Fig. S15) and regional network statistics (Supplementary Fig. S16), as well as trends based on most statistics, such as mean degree or dead-end rate, broadly agree with previous work (2–4, 7), confirmed by additional analysis in the SI (see Supplementary Figs. S17). However, in contrast with recent research (3, 4), we did not find a significant increase in mean degree or proportion of four-way intersections in the 21st century. Some of these results are likely data-dependent (4) but some observed differences may also be due to the MAUP (12, 13). Our large dataset allows us to analyze historical road networks, which are reconstructed from the age information of nearby settlements, at finer scales that do not depend on pre-defined boundaries at coarser resolution, such as a census tract.

Road networks across the rural-urban continuum

Finally, we analyzed and compared trends along the full rural-urban continuum (i.e., including counties outside of the CBSAs), using county-level rural-urban classes (33), stratified by regions (rural urban continuum values for each county are shown in Supplementary Fig. S3). While significant insight has been gleaned from analysis of urban road network growth (1–5, 34), Fig. 6 reveals its complement, road network growth in rural settlements. These results demonstrate that our findings are consistent and generalizable to rural settlements. Namely, the mean degree, road density, and straight road rate in developed areas within rural and urban settings are all decreasing from 1900 to 2015 across all regions (although there is less decrease in rural regions of the Northeast) (Figure 6a). Trends vary most for orientation entropy, which tends to increase in urban regions, but is stagnant and low in rural regions. This is likely because there is less incentive in rural regions to change road orientation over time, although one confounder in this metric is that entropy is lower when less data exists, which would be expected in rural areas. Comparing the road networks established in a fixed time period across the rural-urban continuum (Figure 6a), we observe high levels of persistence, indicating similar road construction trends in urban and rural places in a given time period. As a notable exception, road density in Midwestern settlements is higher in rural settlements than in urban settlements for portions of the road network established in the early 1900s, while this trend appears to invert in more recent time periods. Moreover, the differences between trends across the RUC per time period (Figure 6a) are mostly statistically significant, except for the entropy (Supplementary Fig. S18 & S19), and differences in temporal trends (Figure 6b) exhibit less statistical significance of regional differences for mean degree in peri-urban and rural settings (Supplementary Fig. S18 & S19), indicating high similarity in the degree of road networks in peri-urban and rural areas across regions. Trends are also consistent when split by MSA or μ SA, region, or year of city's maximum development (see relation between maximum development year and regions in Supplementary Fig. S20), with some minor differences, as shown in Supplementary Fig. S21. These trends over time and differences between regions are statistically significant, as shown in Supplementary Fig. S13.

Conclusions

We demonstrate how integration of large spatio-temporal datasets enables new detailed insights into long-term evolution of human settlements, through the lens of road networks across the rural-urban continuum in the US. We measure road network characteristics over time within varying units of analysis and differentiate resulting trajectories across geographical regions. This data-driven approach reveals previously hidden trends and regional patterns that fill important knowledge gaps in our understanding of how road networks have evolved, possible drivers of these changes, and what kind of differences we find in these networks across cities and regions. The continuous reduction in the proportion of gridiron roads in most cities is of particular importance as this reduction is associated with reduced walkability of neighborhoods (3,4). Notably, our findings reveal similar trends in rural regions which have been neglected in previous research. Our findings for urban areas broadly contrast with the New Urbanist school of thought (4), which promotes walkability of cities. Moreover, the observed reduction in roads that follow the school's goals, can be observed even in rural areas and suburban areas. This finding is somewhat unexpected due to the persistently low population density in rural settings and allows a more differentiated reflection of existing policies and concepts.

These insights could be used by researchers and policymakers for a number of applications in the future. First, it can help planners understand the trajectories of the urban and rural landscape and assess when and where infrastructure will need to be built. Next, the presented findings can offer a new understanding of the importance of various network characteristics over space and time and thus shed light on the various forms of development during different time periods and across regions. Finally, such insights could help us understand the (un)intended impacts of infrastructure development, both now and in the past, to inform future planning efforts. For example, it was long known that many highways were built at the expense of minority neighborhoods (35), but there is limited work quantifying the effect this, and other infrastructure projects, had on minority communities. Similarly, the development of road infrastructure can offer economic benefits that have yet to be fully quantified within urban areas (36) as well as rural settings. Given the high priority of infrastructure investment planning in the US, these insights are of particular importance and need to be considered an integral part of urban and rural planning.

While our results point to significant variation as well as commonalities in road network evolution, further analysis is needed to understand the global evolution of these networks (7), and their trends over extended periods of time. Moreover, the presented study focuses on local roads within developed areas. Thus, future work will also include highways that connect the local road networks. Moreover, these results only approximate the network existing at a given time period, as our models focus on network growth, and ignore network shrinkage (e.g., roads disappearing over time). Future work needs to therefore explore the historical network through, for example, automated analysis of historical maps (37).

Materials And Methods

Historical road networks modeling

We assume that the evolution of road networks is largely characterized by expansion over time, and, to a lesser degree, by densification. Changes in the geometric structure (e.g., layout, orientation) of road networks, or shrinkage are rare, and are assumed to be negligible in the case of the US during our study period. Thus, multi-temporal spatial data measuring the expansion of developed, or built-up land over time is commonly used to spatially constrain contemporary road networks to their assumed historical extents, under the assumption that the time period of earliest settlement roughly corresponds to the time period when nearby roads have been constructed. This strategy has been employed in related work analyzing road networks over time and at large spatial extents, using multi-temporal census enumeration units including rural-urban classifications (3), cadastral parcel data containing built-year information (4), multi-temporal remote-sensing derived urban extents (5), or width of residential streets over time (38). Other studies focusing on local scales have extracted multi-temporal road network data from historical maps using computer vision (e.g. (10)).

Herein, we use geospatial vector data from the United States Geological Survey (USGS) National Transportation Dataset (14), representing the US road network in approximately 2018. In order to model retrospective extents of built-up land, we use built-up areas from the Historical Settlement Data Compilation for the U.S. (HISDAC-US; (15, 16)), which are derived from parcel-level built-year information contained in Zillow's Transaction and Assessment Database (ZTRAX; (39)). These built-up areas (BUA) are available in 5-year intervals from 1810 to 2016, as a series of binary, gridded surfaces at a resolution of 250m ((40), Supplementary Fig. S2a). Likewise, historical estimates of the number of buildings (built-up property locations; BUPL, Supplementary Fig. S2b) per grid cell are available in the HISDAC-US repository (41). While HISDAC-US data coverage is sparse in some rural areas of the US, geographic coverage and temporal information is largely complete in urban regions (15). Moreover, the accuracy of the built-up extents layer increases over time (15, 16) reaching acceptable levels after 1900 (15).

Based on the gridded surfaces from the HISDAC-US we developed an approach to generate spatially generalized urban extents, consistent across different cities and over time. In a first step, we generated a built-up density surface in the starting year (i.e., 1910) within each 2015 CBSA boundary, using circular focal windows of radius $r = 1$ kilometer, containing the proportion of built-up area within the focal neighborhood. We then selected all grid cells with a focal built-up density greater than 5%. This method has previously been employed to discretize the rural-urban continuum into high density (urban) and lower density (periurban) strata (42) and shows high discriminative power between signals in remotely sensed spectral responses. For each CBSA and year, we then segmented the resulting contiguous groups ("patches") of urban grid cells and computed the sums of built-up area, building indoor area, and number of buildings per patch, and computed the percentile ranks of the patches within a CBSA according to the number of buildings they contain. We discarded small patches containing only a few buildings, likely representing scattered periurban settlements. To do so, we retained only patches that exceeded the 90th percentile in the first year when the density filtering yields at least one patch of built-up land (which may be later than 1910 for late-developing cities). This way, we ensure that urban areas are modelled based

on consistent criteria across space and time, and represented by smooth, contiguous, and largely gap-free areas (Supplementary Fig. S2c). We then clipped the NTD road vector data to the urban delineations in each year, yielding sub-networks that can be uniquely identified by the combination of CBSA, year, and patch identifier. Lastly, we identified the coordinates of the end points of the road vector lines that were introduced by the clipping, and recorded these locations for subsequent node analysis, since these nodes represent artificial cul-de-sacs introduced by the data processing and need to be excluded from subsequent statistical analysis (Supplementary Fig. S2d). In total, we analyzed 8.0 million nodes and 10.1 million edges within CBSAs.

CONUS-wide historical road network modeling

To derive trends of road network characteristics over time and across the rural-urban continuum, we used the NTD road network vector data (14, 33) and the first built-up year dataset (FBUY), available as a gridded dataset from the HISDAC-US data repository (15,16) (temporal and geographic coverage of these data are shown in Supplementary Figs. S3 & S4). We first identified grid cells developed within moving windows (time periods) of 40 years, in steps of 20 years, e.g., developed prior to 1900, 1880-1920, 1900-1940, 1920-1960, etc. For each county in the CONUS, we then extracted the road network vector objects within the areas corresponding to each 40-year development period and assigned an individual identifier to each contiguous group of developed grid cells (patches). We removed small, spatially isolated patches of under 0.31 square kilometers (corresponding to five 250 by 250 meter grid cells), as well as elongated patches of less than 500 meter width, likely representing settlements along highways and thus not relevant to characterize road networks in cities, towns, or places (see Supplementary Figs. S1 & S2). For the remaining patches, containing over 27 million road segments, we calculated a range of road network metrics, aggregated per county and time period. Based on the county-level rural-urban continuum code provided by the US Department of Agriculture (33), classifying each county into one of nine levels of “rurality”, we analyzed each of the network metrics in a bi-variate manner over time and across the rural-urban continuum, stratified by US census region (24) (see RUCC boundaries in Supplementary Fig. S5). In total, we analyzed 9.2 million nodes and 15.2 million edges across the rural-urban continuum.

Metropolitan-level historical road network modeling and statistical analysis

To construct CBSA-level analyses of the road networks, we group the patches constructed above into CBSA regions. If a patch is on the border of a CBSA region, we cut it off at the boundary, and edges that reach the boundary are removed; this is not common but is a reasonable way of defining patches associated with only one metropolitan or micropolitan area. At ten-year intervals between 1900 and 2010, as well as at 2015, we construct the road network topology using the Python library networkx, and remove nodes with degree two from the analysis, which we explain in more detail in the Supplementary Fig. S22. For each patch, we record the total road length, area, degree, proportion deadends and degree greater than

or equal to four, as well as local entropy and griddedness. These raw statistics are used to construct combined measures, e.g., the distance per unit area to be able to quantify the road length distance within all patch areas (which are a small proportion of the total CBSA area). Both CONUS-wide and CBSA level historical road networks were extracted using ESRI ArcPy (43), Safe Software Feature Manipulation Engine Desktop (44).

Grid-cell-level correlation analysis and time series clustering

Using the FBUY gridded surface from HISDAC-US (15, 16), we calculated the average settlement age within 1×1 kilometer grid cells located within the 2015 urban delineations derived from the density-based delineation method described above using GeoPandas (45) and SciPy (46) Python modules. The aggregation to 1×1 kilometer grid cells aims to avoid small sample sizes of road segments and intersections per grid cell, and thus, ensures the statistical support for the network statistics calculated per grid cell. We then identified all network nodes within a grid cell, as well as the centroids of all road segments (i.e., network edges) per grid cell. Based on the network statistics attributed to each node (e.g., nodal degree, local griddedness) and to each edge (e.g., azimuth, road segment length) we calculated road network statistics for each grid cell, such as average degree or total road length (see Fig. 2 in the main text). In total, we calculated seven grid cell-level network statistics (Fig. 4 in the main text and Supplementary Fig. S7). Note that due to potentially small sample sizes within grid cells, we replaced the orientation entropy by the variety of unique azimuth values per grid cell, calculated after discretizing the road segment azimuth into bins of 10 degrees. Based on the gridded surface indicating the average age per grid cell, and corresponding cell-level network statistics within each CBSA, we extracted cell-by-cell pairs of settlement age and road network statistics for each city. These vectors enable us to calculate correlations between age and network characteristics, for each city, considering the local, fine-grained variability of settlement age as it is associated to the characteristics of the road network. Moreover, we generated time series of each network characteristic for each city. In order to characterize the relative relationship between age and network characteristics, we discretized the age surface per CBSA into deciles. Thus, the resulting time series consist of the same number of observations (i.e., ten) and are independent from the absolute age of the cities. For each of the network characteristics, we conducted time series-based cluster analysis separately for MSAs and μ SA. We used the time-series k-means algorithm (TSK-means) (25) in conjunction with the Dynamic Time Warping (47) similarity metric to characterize the dissimilarity between time series, implemented in tslearn (48) Python module. In order to identify the optimum number of clusters k , we calculated the cluster inertia based on DTW similarity as a measure of separation between time series clusters for a range of k from 2 to 20 and used the commonly used elbow method (49) to identify the approximate number of clusters for each scenario. We normalized the cluster inertia of each clustering scenario into the range $[0, 1]$ in order to compare the cluster quality across the different network statistics (Fig. 4d in the main text).

Road Statistics

Griddedness has become critical to understanding traffic inefficiency and related problems in cities but measuring it has been difficult until recently. Namely, grid-like road networks appear to enhance walkability and lower relative vehicular travel in a city (3, 19). Griddedness (and related urban sprawl) metrics have been defined and implemented on several recent occasions (3, 5, 34). The methods to derive such metrics are sophisticated (3) and sometimes computationally expensive (5, 34). However, we aim for a simple, intersection-level measure to extend on previous work.

We develop the local griddedness metric, a spatial complement to the clustering coefficient often used in network analysis (2, 23). The local clustering coefficient of a node is defined as the proportion of triangles that exist whose vertex includes that node relative to the total number of possible triangles that could exist for a node of that degree. Unlike, e.g., social networks, road networks tend to have four-cycles, and more uniquely still, these cycles tend to be planar, meaning they all lie on a two-dimensional plane.

These constraints offer guidance to a unique spatial clustering coefficient, local griddedness, which is the proportion of four-cycles containing that vertex relative to the total number of planar four-cycles for a node with that degree. Degree one, two, and three nodes are common and special cases for intersections, however. If a node is the end of a dead-end road, we define the local griddedness to be zero. We do not analyze nodes of degree two, as we explain in more detail in the SI. Finally, it is unlikely for a three-road intersection to have three city blocks meet there, but more likely is that it ends in a T. The maximum number of city blocks is therefore defined as two and is otherwise k . While in most cases, local griddedness is between 0 and 1, we allow for rare instances in which, for example, degree-three nodes have a value up to $3/2$ (a “super gridded” node), in order for the T intersection to have a natural griddedness value of 1.0. This also means that roads that violate this planar assumption (e.g., those with bridges) may be greater than 1, but such instances are rare. Using this measure, any node can have its griddedness value rapidly calculated, allowing for extremely fine-grained analysis of road network spatial statistics.

In addition to this metric, we separately calculated azimuth variety and orientation entropy by binning the angles between road intersections into six-degree wide bins. Changing this width does not qualitatively change our findings but can change the absolute value of entropy (which can be as high as the log of the number of bins) as well as the azimuth variety.

Declarations

Acknowledgments:

We acknowledge access to the Zillow Transaction and Assessment Dataset (ZTRAX) through a data use agreement between the University of Colorado Boulder and Zillow Group, Inc. More information on accessing the data can be found at <http://www.zillow.com/ztrax>. The results and opinions are those of the authors and do not reflect the position of Zillow Group. Moreover, Safe Software, Inc., is acknowledged for providing a Feature Manipulation Engine (FME) Desktop license used for data

processing. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding:

Funding was provided by the National Science Foundation (grant 1924670) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (grant P2CHD066613) under SL.

Author contributions:

KB, JU, KL, and SL conceived the research, contributed to the investigation, and wrote the manuscript. KB and JU developed the methodology and visualization. KL and SL acquired funding, administered the project, and supervised the research.

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials. All code will be made available after acceptance on Gigantum.

References

1. G. Boeing, *Applied Network Science* 4, 67 (2019).
2. G. Boeing, *Environment and Planning B: Urban Analytics and City Science* 47, 590 (2020).
3. G. Boeing, *Journal of the American Planning Association* 87, 1 (2020). 21
4. C. Barrington-Leigh, A. Millard-Ball, *Proceedings of the National Academy of Sciences* 112, 8244 (2015).
5. C. Barrington-Leigh, A. Millard-Ball, *Proceedings of the National Academy of Sciences* 117, 1941 (2020).
6. E. G. Irwin, N. E. Bockstael, *Proceedings of the National Academy of Sciences* 104, 20672 (2007).
7. G. Boeing, *arXiv preprint arXiv:2009.09106* (2021).
8. D. Kaim, M. Szwagrzyk, K. Ostafin, *Data in Brief* 28, 104854 (2020).
9. S. Wang, *et al.*, *Sustainability* 11 (2019).

10. M. Saeedimoghaddam, Exploring the effectiveness of the urban growth boundaries in USA using the multifractal analysis of the road intersection points, a case study of Portland, Oregon, Ph.D. thesis, University of Cincinnati (2020).
11. B. Shbita, *et al.*, *The Semantic Web*, A. Harth, *et al.*, eds. (Springer International Publishing, Cham, 2020), pp. 409–426.
12. S. Openshaw, P. J. Taylor, *A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem* (Pion, London, 1979).
13. A. P. Masucci, E. Arcaute, E. Hatna, K. Stanilov, M. Batty, *Journal of The Royal Society Interface* 12, 20150763 (2015).
14. U.S. Geological Survey, National Geospatial Technical Operations Center, USGS National Transportation Dataset (NTD), <https://thor.f5.er.usgs.gov/ngtoc/metadata/waf/transportation/ntd/> (2018). Online; accessed 01 June 2020.
15. J. H. Uhl, *et al.*, *Earth System Science Data Discussions* pp. 1–43 (2021).
16. S. Leyk, J. H. Uhl, *Scientific data* 5, 180175 (2018).
17. S. Leyk, *et al.*, *Science Advances* 6 (2020).
18. J. H. Uhl, D. S. Connor, S. Leyk, A. E. Braswell, *Communications Earth & Environment* 2, 20 (2021).
19. R. Cervero, K. Kockelman, *Transportation Research Part D: Transport and Environment* 2, 199 (1997).
20. E. Strano, V. Nicosia, V. Latora, S. Porta, M. Barthelemy, *Scientific Reports* 2, 296 (2012).
21. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, G. B. West, *Proceedings of the National Academy of Sciences* 104, 7301 (2007).
22. Census Bureau, tl_2015_us_cbsa, <https://www2.census.gov/geo/tiger/TIGER2015/CBSA/> (2015). Online; accessed 01 June 2020.
23. D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998).
24. Census Bureau, cb_2018_us_region_500k, https://www2.census.gov/geo/tiger/GENZ2018/shp/cb_2018_us_region_500k.zip (2018). Online; accessed 01 June 2020.
25. X. Huang, *et al.*, *Information Sciences* 367-368, 1 (2016).
26. H. Sakoe, S. Chiba, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 43 (1978).
27. E. H. Simpson, *Journal of the Royal Statistical Society. Series B (Methodological)* 13, 238 (1951).
28. L. McInnes, J. Healy, N. Saul, L. Großberger, *Journal of Open Source Software* 3, 861 (2018).
29. A. Sharifi, *Building and Environment* 147, 171 (2019).
30. S. Gao, Y. Wang, Y. Gao, Y. Liu, *Environment and Planning B: Planning and Design* 40, 135 (2013).
31. S. Angel, J. Parent, D. L. Civco, A. Blei, The Persistent Decline in Urban Densities: Global and Historical Evidence of 'Sprawl', https://www.lincolninst.edu/publications/working_papers/persistent-decline-urban-densities (2017). Online; Accessed 01 June 2020.
32. G. Xu, *et al.*, *Landscape and Urban Planning* 183, 59 (2019).

33. USDA, Rural-Urban Continuum Codes, <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx> (2021). Online; accessed 01 June 2021.
34. C. Barrington-Leigh, A. Millard-Ball, *PLOS ONE* 14, 1 (2019).
35. D. Fitzpatrick, *Pittsburg Post-Gazette* (2000).
36. T. Jaworski, C. T. Kitchens, *The Review of Economics and Statistics* 101, 777–790 (2019).
37. J. H. Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, C. A. Knoblock, *IEEE Access* 8, 6978 (2020).
38. A. Millard-Ball, *Journal of the American Planning Association* 0, 1 (2021).
39. Zillow Inc., ZTRAX: Zillow Transaction and Assessment Dataset, <https://www.zillow.com/research/ztrax/> (2016). Online; accessed 01 January 2020.
40. J. H. Uhl, S. Leyk, Historical built-up areas (BUA) - gridded surfaces for the U.S. from 1810 to 2015, <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/J6CYUJ> (2020). Online; accessed 01 June 2020.
41. J. H. Uhl, S. Leyk, Historical built-up property locations (BUPL) - gridded surfaces for the U.S. from 1810 to 2015, <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/SJ213V> (2020). Online; accessed 01 June 2020.
42. S. Leyk, J. H. Uhl, D. Balk, B. Jones, *Remote sensing of environment* 204, 898 (2018).
43. ArcGIS Pro Python reference, <https://pro.arcgis.com/de/pro-app/latest/arcpy/main/arcgis-pro-arcpy-reference.htm> (2021). Online; accessed 01 June 2021.
44. FME Desktop, <https://www.safe.com/fme/fme-desktop/> (2021). Online; accessed 01 January 2021.
45. K. Jordahl, *et al.*, geopandas/geopandas: v0.8.1, <https://doi.org/10.5281/zenodo.3946761> (2020). Online; accessed 01 June 2020.
46. P. Virtanen, *et al.*, *Nature Methods* 17, 261 (2020).
47. M. Muller, *Dynamic Time Warping* (Springer, Berlin, Heidelberg, 2007).
48. R. Tavenard, *et al.*, *Journal of Machine Learning Research* 21, 1 (2020).
49. M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, B. D. Satoto, *IOP Conference Series: Materials Science and Engineering* 336, 012017 (2018).

Figures

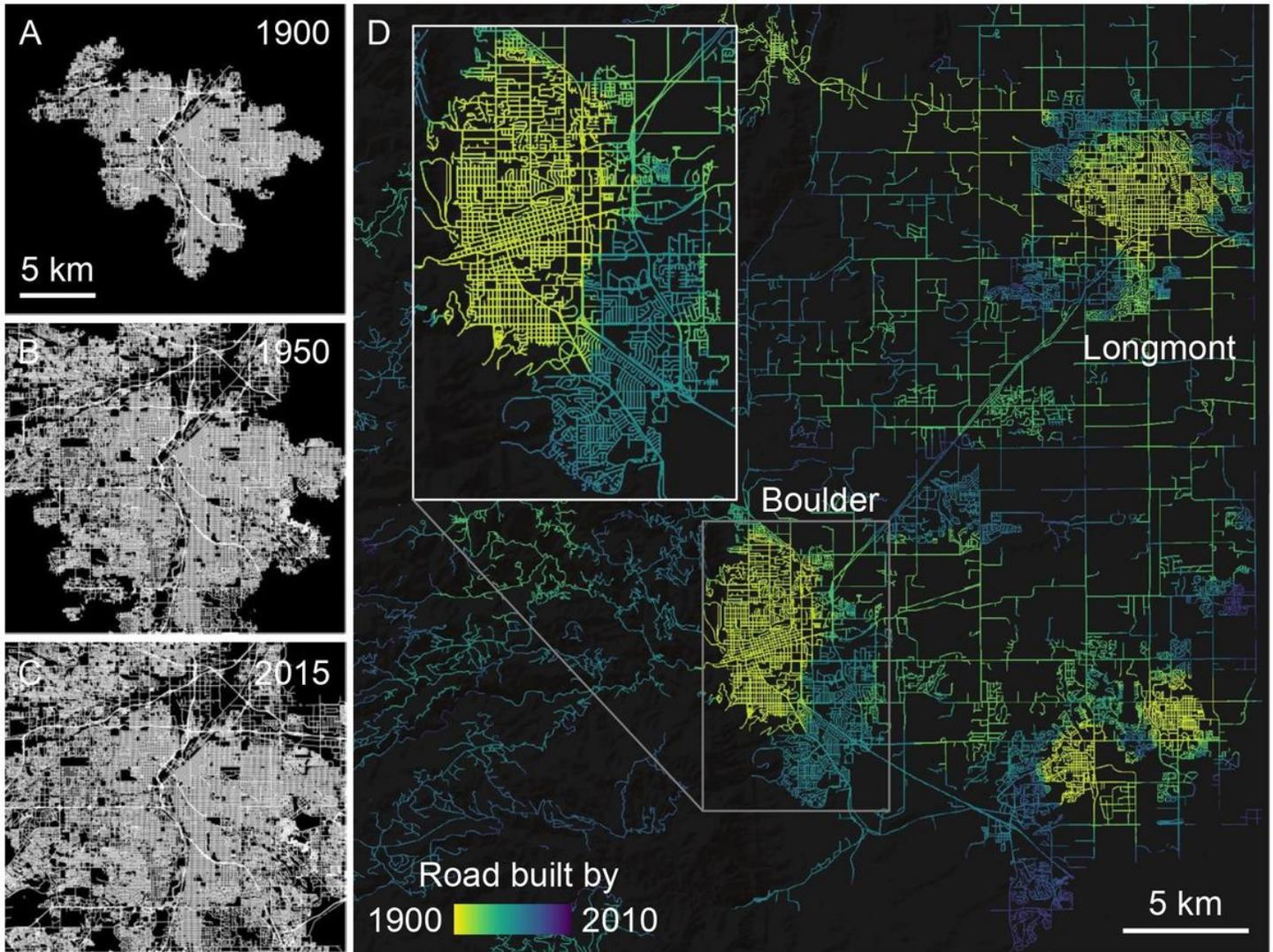


Figure 1

Growth of urban road networks over time. Denver metropolitan area in (a) 1900, (b) 1950, and (c) 2015; (d) Estimated road ages in a peri-urban setting near Boulder (Colorado) with street ages color-coded from light (1900) to dark (2015).

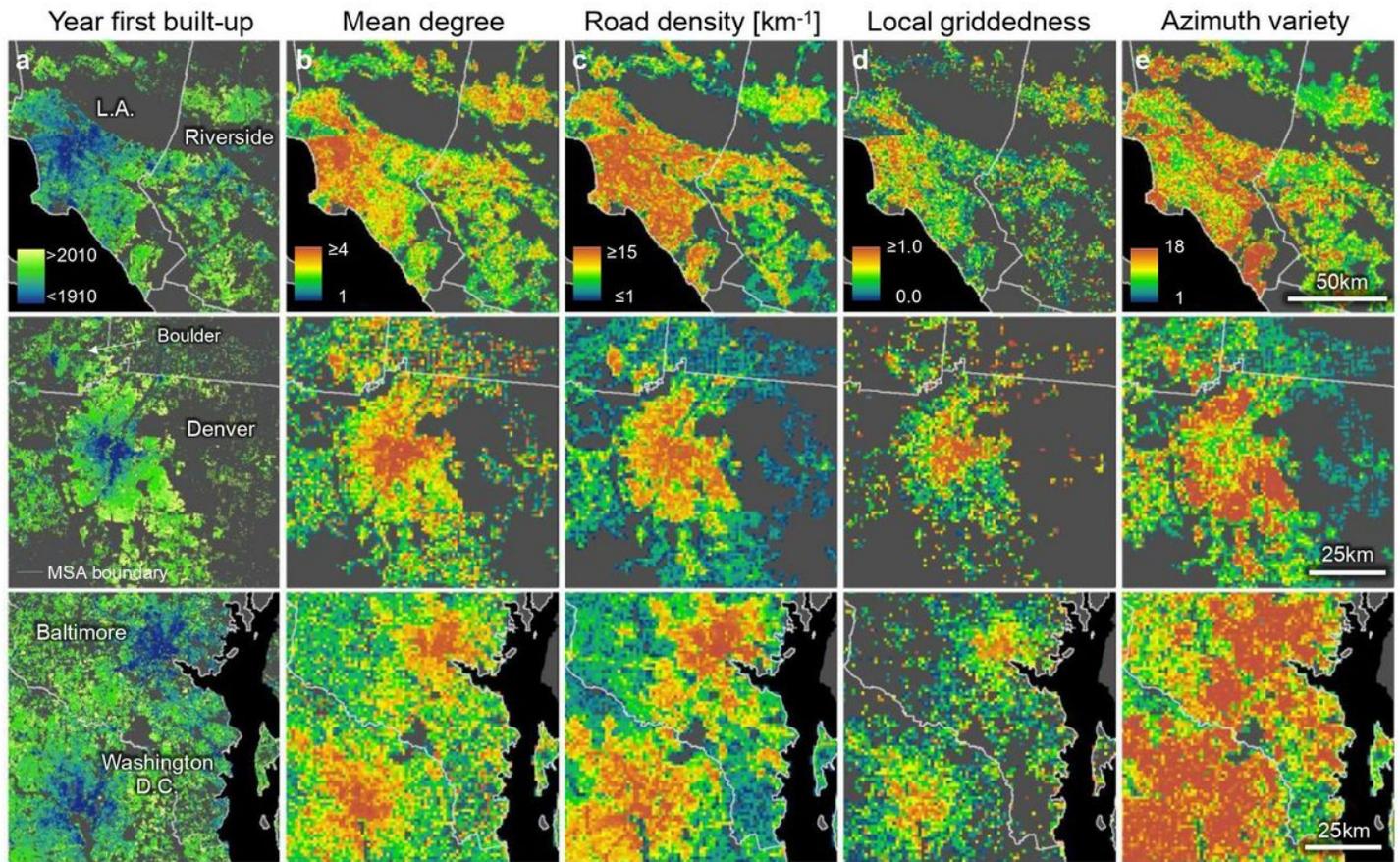


Figure 2

Gridded (1x1 kilometer) road network statistics within cities. Columns, left to right: inferred date of road construction, mean degree (number of roads at each intersection), road density (kilometers of road per square kilometer), local griddedness (degree to which an intersection is part of a grid topology), and azimuth variety (variety of road orientations, discretized to 10° bins). Rows represent from top to bottom: Los Angeles, Denver, and Baltimore-Washington metropolitan areas.

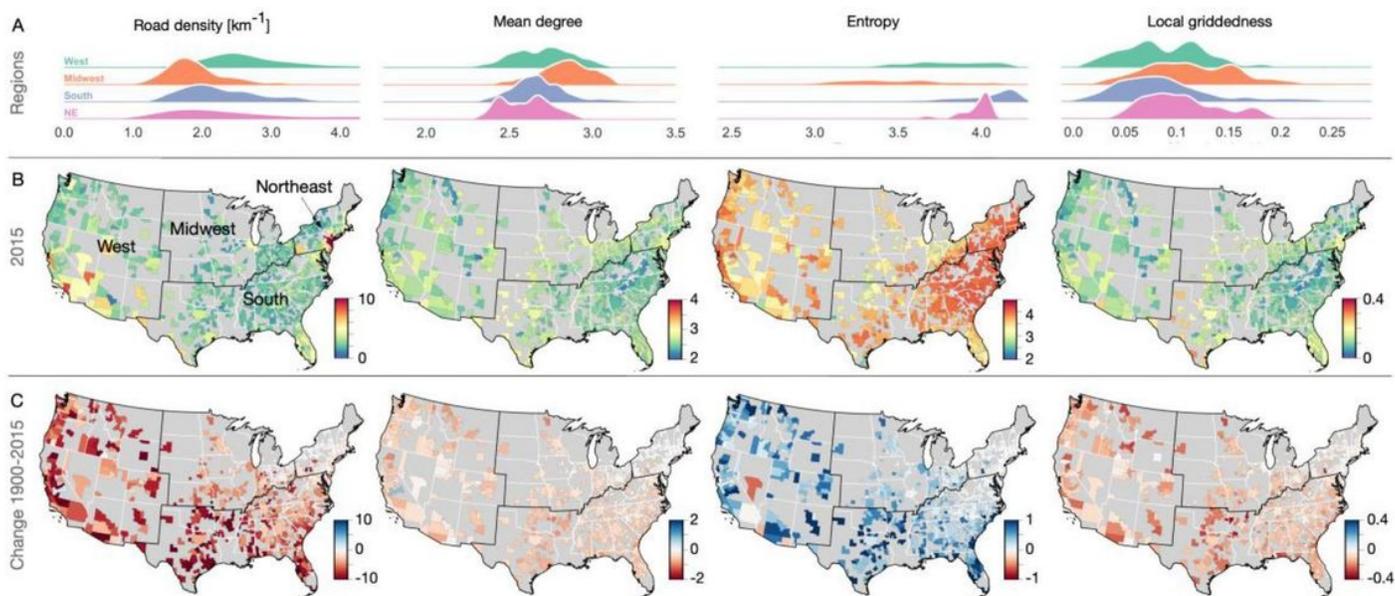


Figure 3

Variations in road network statistics across the US. (a) statistical distributions for road density, mean degree, orientation entropy [1], and local griddedness, split by region. (b) Road statistics for each CBSA. (c) Changes in road network statistics between 1900 and 2015.

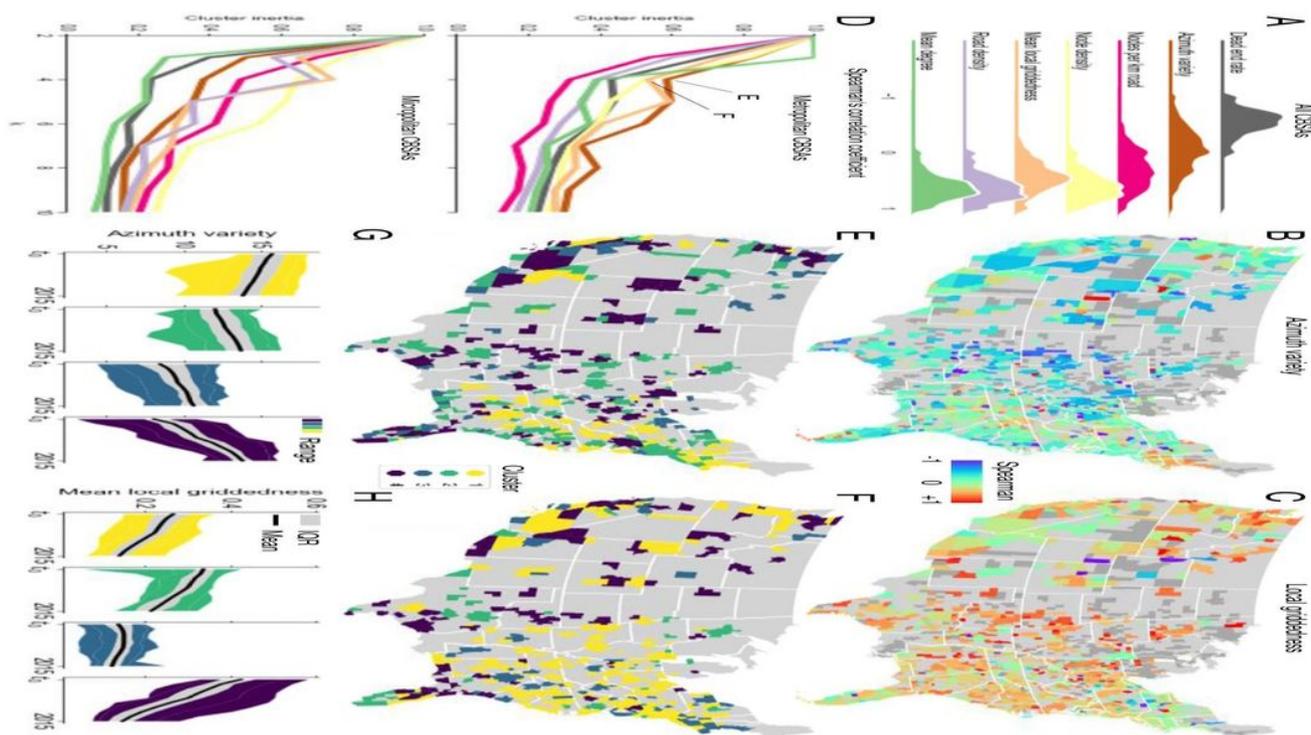


Figure 4

Grid-cell level correlation and time series clustering. (a) Distributions of CBSA-level Spearman correlation between network statistics and settlement age. (b–c) maps of Spearman correlation between age and azimuth variety and local griddedness, respectively, for CBSAs. (d) Dynamic time warping-based cluster inertia versus the number of clusters using time series k-means for CBSAs. Strong clustering occurs at an elbow in this plot, which is at roughly four clusters. (e–f) K-Means-based clusters with $k=4$ for (e) azimuth variety in MSAs and local griddedness in MSAs. (g–h) The time series of the clusters found in (e–f).

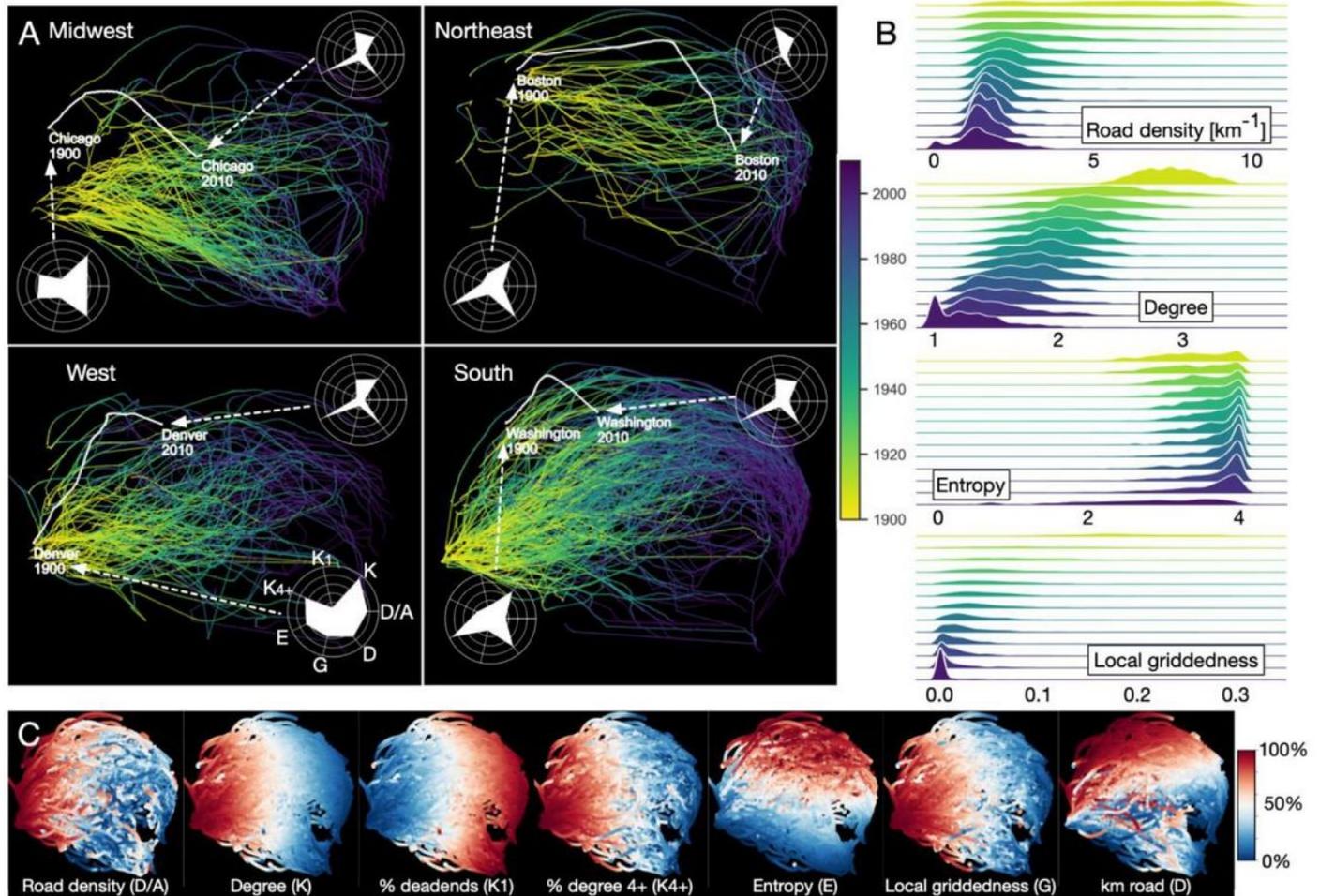


Figure 5

Road network statistics at the CBSA-level over time. (a) UMAP plot of cities embedded into seven road network statistics for each US region. Radar charts are road network statistics for example MSAs whose outer ring corresponds to the largest value of each network statistic. (b) Distributions of road density, mean degree, entropy, and mean griddedness per CBSA over time. (c) A closer inspection of the UMAP plots in (a) reveals network statistics vary strongly, such that cities in the upper left-hand corner of each plot have many roads built per year and have high entropy. Networks in the lower left-hand corner have higher local griddedness, road density, % degree 4+, and average degree, as well as lower % deadends.

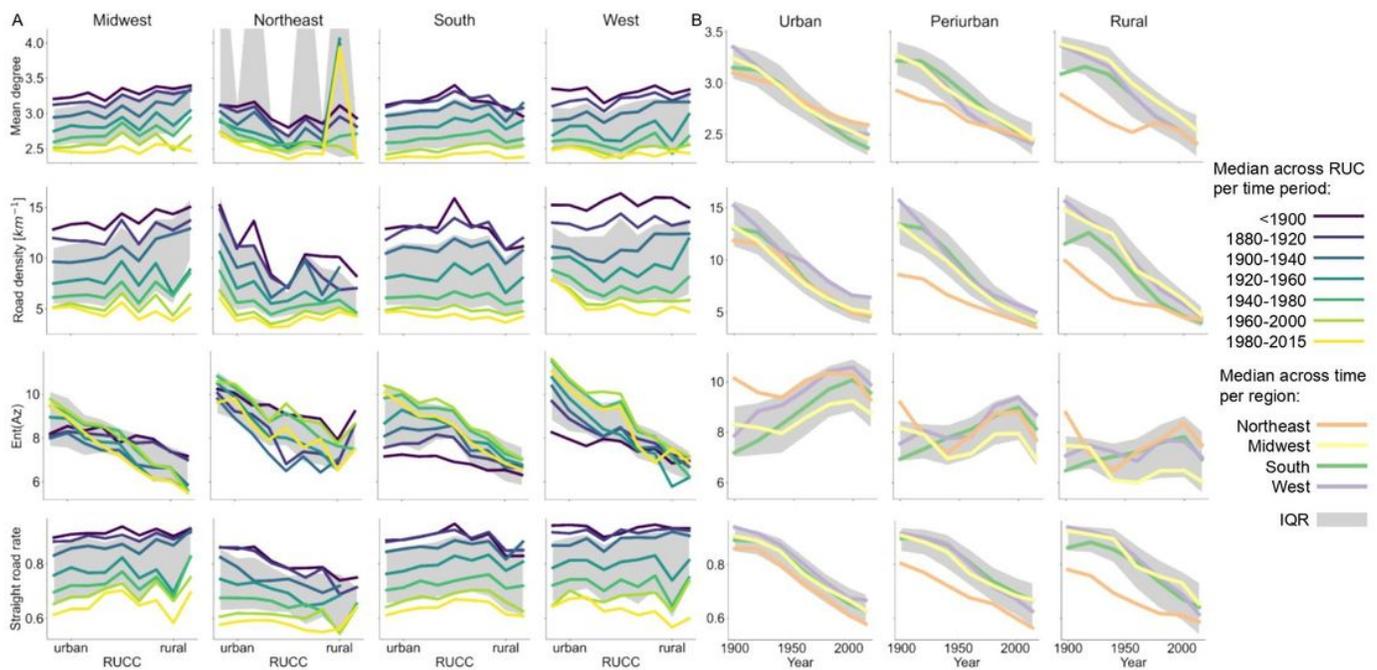


Figure 6

County-level network statistic trends across time, regions, and the rural-urban continuum (RUC). (a) Road network statistics for different time periods, as a function of the urban-rural gradient, using county-level rural-urban continuum codes (RUCCs), separately each US region. (b) Corresponding visualization of these decomposed trends over time, aggregated by region. Shown are the median values of county-level statistics within the strata defined by RUCC, time period, and region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NatureCommsupplementarymaterialstemplate2021.pdf](#)