

# A high-resolution global time series of street-network sprawl

Christopher Barrington-Leigh      Adam Millard-Ball  
McGill University                      UCLA

Published in  
*Environment and Planning B: Urban Analytics and City Science*, 2025.

DOI:10.1177/23998083241306829

Compiled: January 14, 2025

Latest version available online

## Abstract

Systems of street networks form a backbone for many aspects of human life and, once laid down, urban streets represent a nearly immutable influence on future urban form and concomitant travel, energy, and social outcomes. Moreover, as humanity is currently passing through its peak urbanization rate, decisions about how to design such networks at the local scale are being made faster than ever before. In this work, we quantify local street connectivity and provide a global, high-resolution time series of our Street Network Disconnectedness Index (SNDi) as an open data set. We derive a stylized version of the actual geographic road network from the 2023 vintage of OpenStreetMap by simplifying complex intersections, divided roads, and offset intersections. Using this functional representation of the network corrects systematic biases in derived properties of the network. We couple this simplified network with a newly-available time series of urbanization in order to compute SNDi and provide a dynamic analysis to the year 2019 and a cross-sectional analysis for 2023. We release our data as the raw network of edges and nodes and as aggregates to a 1 km grid, to countries, and to five subnational administrative levels. We also provide interactive visualizations at [sprawlmap.org](http://sprawlmap.org). Overall, our findings present a picture of rapidly worsening street network connectivity in many regions of the world.

**Acknowledgements:** We appreciate excellent research assistance from Ben Silverstein. We also thank Martino Pesaresi and Thomas Kemper at the European Commission Joint Research Centre for helping us interpret and utilize the GHSL data package. This work was supported by Social Science and Humanities Research Council of Canada Grant 435-2016-0531.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Measuring street-network sprawl</b>	<b>4</b>
<b>3</b>	<b>Simplifying the network</b>	<b>4</b>
3.1	Complex intersections . . . . .	5
3.2	Divided roads . . . . .	6
3.3	Other simplifications . . . . .	7
3.4	Simplification procedure . . . . .	7
<b>4</b>	<b>Computing connectivity measures</b>	<b>8</b>
4.1	Circuitry correction . . . . .	8
4.2	Urban region and time series identification . . . . .	8
4.3	Aggregation . . . . .	8
<b>5</b>	<b>Results</b>	<b>9</b>
<b>6</b>	<b>Data and code availability</b>	<b>13</b>
<b>7</b>	<b>Conclusion</b>	<b>13</b>
<b>A</b>	<b>Appendix</b>	<b>16</b>
A.1	Algorithm to simplify the street network . . . . .	16
A.1.1	Table of edges . . . . .	16
A.1.2	Drop redundant edges . . . . .	16
A.1.3	Identify clusters of nodes . . . . .	16
A.1.4	Anneal clusters . . . . .	16
A.1.5	Anneal nodes . . . . .	17
A.1.6	Drop parallel bicycle/pedestrian edges . . . . .	17
A.1.7	Classify divided roads . . . . .	18
A.2	Decomposition of changes . . . . .	18
A.3	Description of connectivity measures . . . . .	21
A.4	Definition of Street Network Disconnectedness Index (SNDi) . . . . .	23

# 1 Introduction

Street networks are a key dimension of urban form. The connectivity of streets shapes the travel choices and carbon footprints of a city’s residents (Ewing and Cervero, 2010; Barrington-Leigh and Millard-Ball, 2017), and possibly long-term densification opportunities as well. Less connected streets – which we call street-network sprawl – reduce accessibility by bus, walking and cycling, partly because their greater circuitry increases travel distances and makes public transportation service less feasible.

The importance of today’s decisions on street-network sprawl is amplified by lock-in and by the rate of change. Once laid down, street patterns rarely change, even after fires, wartime bombing, and earthquakes (Vazquez et al., 2023). And given that growth in the urban population is expected to peak by 2025 (United Nations, 2019), it likely that the 2020s will also represent the period of peak growth in urban street construction.

Our previous work (Barrington-Leigh and Millard-Ball, 2019, 2020) reported aggregate estimates of street-network sprawl for all countries and 200 selected cities. In this paper, we offer three major advances. First, we release for the first time the full data at the 1km level as a global grid, as well as the underlying street and path network as a graph of nodes and edges. In our earlier work and in this new dataset, we use raw data from OpenStreetMap (OSM) to compute 13 measures of connectivity, which we aggregate into a single index — the Street Network Disconnectedness Index (SNDi). Second, we offer a new algorithm for transforming the geometric representation of the street network into a simplified graph that is more appropriate for network analysis. Third, we update our cross-sectional data to use the October 2023 vintage (i.e., version) of OSM, and update the time series to examine shifts in street-network sprawl from 2015 to 2019, taking advantage of new longitudinal datasets on urban growth.

Our dataset complements the availability of global-scale population density (ORNL, 2011; Florczyk et al., 2019) and street connectivity datasets for specific countries (Boeing, 2020). Many analyses of urban form focus exclusively on population density, given the ready availability of data derived from national censuses and remotely sensed nighttime lights. Our street-network sprawl data provide a complement that reflects a separate and more enduring dimension of urban form. Our work is closest to that of Boeing (2021) who provides street connectivity indicators for each urban center in the world. We go beyond that to (i) provide a time series from 1975 to 2020; (ii) provide measures for a 1km grid that is not confined to urban center boundaries; (iii) include bicycle and pedestrian paths that are omitted from the Boeing dataset, but are crucial in the connectivity of streets in places such as Denmark (Barrington-Leigh and Millard-Ball, 2020); and (iv) simplify the street network to provide more meaningful measures of connectivity.

Our contributions in this paper are (i) producing a dataset for other researchers to use and build on, (ii) providing descriptive analysis of geographic patterns in street-network sprawl and trends over time, and (iii) identifying and addressing how to simplify a geometric road network for purposes of network and connectivity analysis. As we discuss below, simplification is important to avoid double counting complex intersections that are represented by multiple nodes, and roads with a median that are represented by multiple edges. As this is a “data” paper, we do not identify or test specific hypotheses or theoretical mechanisms, but our data can facilitate future research in the vein of regional and global-scale studies on the impact of street connectivity on urban history (Vazquez et al., 2023; Salazar Miranda, 2021), transportation choices (Marshall and Garrick, 2010; Hajrasouliha and Yin, 2015; Brenner et al., 2024), environmental outcomes (Rezaei and Millard-Ball, 2023), social relations, and more.

## 2 Measuring street-network sprawl

Street-network sprawl — our term for low street connectivity — is often quantified as the proportion of deadends or four-way intersections, average nodal degree, and/or intersection density (Marshall and Garrick, 2010; Ewing and Cervero, 2010). In North America, the empirical setting for most street connectivity research to date, these measures can distinguish between two of the most common street network configurations – a grid (where most nodes are degree-4) and subdivisions with cul-de-sac (where most nodes are degree-3 or degree-1). But nodal degree does not do justice to the connectivity of irregular networks typical of many Japanese, Middle Eastern, and European cities, where most nodes are degree-3 but pedestrian connectivity is clearly high.

Therefore, we quantify street-network sprawl using a wider range of attributes that can distinguish between more- and less-connected streets in a range of different urban design traditions. In addition to nodal degree and the fraction of deadends, we use measures of circuitry (the ratio between network distance and Euclidean distance) and sinuosity (the curviness of individual street edges). We also include several measures of the network function of each edge based on their graph theoretic properties. For example, we calculate the number and fraction of network “bridges” — streets that are the only connection between two different parts of the network, such as the single entrance to a gated community. Our 13 measures are described in Table A2 in the Appendix.

We collapse these 13 measures to a single index, the Street Network Disconnectedness index (SNDi) by using their first principal component (Table A3). Higher SNDi indicates more street-network sprawl, i.e. less connected streets. In our previous work, we validated SNDi using Google Street View imagery and national census data, and showed that SNDi corresponds to neighborhood walkability as well as individual decisions on car ownership and commute mode. The updated Principal Component Analysis (PCA) coefficients reported in Table A3 are similar to our previous estimates, indicating that our newly calculated SNDi has a similar interpretation.

Most of our analysis was carried out in a PostgreSQL/PostGIS database, with scripting and some network analysis functions undertaken in Python. We made use of moderate parallelization (56 processors) for many tasks, and of the  $\sim 0.7$  TB RAM available on our dedicated computation server. The entire planet took us about one month to process. The two most lengthy stages were the road network simplification ( $\sim 12$  days) and the calculations of circuitry and graph theoretic properties for each node ( $\sim 14$  days).

In general, our computational approach is similar to that described in our 2019 analysis. In the following sections, we briefly describe that approach, and focus in more detail on the changes we have made. In the Appendix, we provide a more extensive step-by-step discussion of our algorithm, and show the impact of the change to our algorithm and of updating the underlying OSM data from the 2019 to the 2023 vintage.

## 3 Simplifying the network

Maps that are geometrically accurate are not necessarily well-suited for network analysis and analyzing the connectivity of streets. Three examples are illustrative. In one case, a roundabout is functionally a single intersection, but geometrically, the roundabout may be depicted as a circular street, with each entrance (and possibly exit) accounting for a separate intersection. In another case, a slightly staggered intersection is functionally a degree-4 node, but geometrically might be represented as two degree-3 nodes, connected by a very short street. In a third case, a two-way street might be represented as two different one-way streets due to the presence of a median. In all

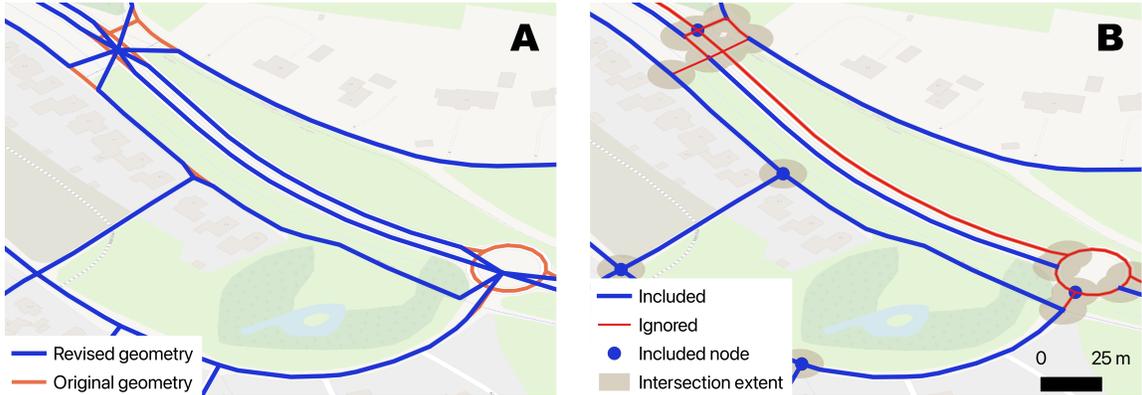


Figure 1: Treatment of complex intersections in 2019 (A) and 2022 (B) analysis. In Panel A, the intersection is collapsed to a single node. In Panel B, the original geometry is retained, but some edges are ignored in the aggregation. The extent of each complex intersection in Panel B is shown as the shaded area.

three cases, using the geometric rather than the functional representation will tend to bias upwards the number of nodes and also affect other properties of the network. Such biases will also inflate other measures often used in the urban planning literature, such as intersection density.

For this reason, we develop an algorithm to simplify the geometric representation of the street network provided by OSM. Our overall philosophy is to mitigate the dependence of our results on how the street network is represented in OSM — for example, the mapper may depict a staggered intersection with two nodes rather than a single node, map sidewalks as separate edges from the roadway, or represent a road with a median as two parallel edges, one for each direction. Note that this is conceptually separate from our process of annealing — i.e., removing nodes that are degree-2 (not intersections or deadends). While this annealing is called “simplification” by Boeing (2017), our simplification process goes beyond this.

We also simplified the network in our 2019 analysis, but here we improve on that approach in two ways related to (i) complex intersections and (ii) divided roads. We summarize our algorithm here and provide more detail in the appendix.

### 3.1 Complex intersections

Complex intersections include roundabouts, staggered intersections, and motorway interchanges that are functionally a single intersection, but are represented as multiple nodes in OpenStreetMap. In our 2019 analysis, we identified clusters where all nodes were within 20m of another node, and collapsed such clusters into single nodes at their centroids (Figure 1, panel A). This approach had computational advantages, but led to a slight overestimate of circuitry (all routes had to pass through the cluster centroid), and made the graphical representation non-intuitive.

In the present analysis, we adopt an approach that preserves all original nodes and edges (summarized in Table A2. ) We identify clusters of nodes where each node is within 20m (via the street network) of another node, and select a random node to represent that cluster (Figure 1, panel B). All other nodes are ignored at the later stage when we calculate aggregate connectivity

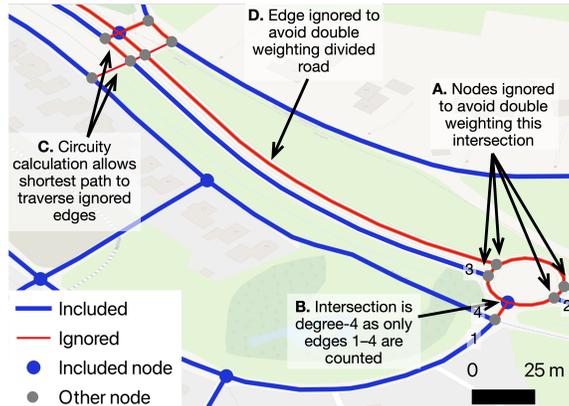


Figure 2: Rationale for excluding ignored nodes and edges from the aggregations (A, B, D), although the ignored edges are considered when calculating circuituity (C).

measures, in order to avoid double weighting these intersections (Figure 2A) and to avoid inflating the nodal degree of these intersections (Figure 2B). Moreover, edges  $< 20\text{m}$  in length are classified as “intra-cluster,” and are also ignored in the aggregation. However, these intra-cluster nodes and edges are retained for purposes of calculating circuituity — they can form part of the shortest path between two non-intra-cluster nodes (Figure 2C). In Figure 2, for example, the blue nodes would not even be connected without considering the intra-cluster and other ignored (red) edges.

### 3.2 Divided roads

In some cases, roads with a center median are depicted as two parallel edges in OSM, even if they are functionally the same road. This would effectively mean that these roads are double-weighted in our calculations of SNDi. To mitigate this issue, we identify sets of two or more parallel edges, defined as those that start and end in the same two clusters of nodes. Within each set, we first drop footpaths, bicycle paths, and similar edges where there is a parallel “car” edge; typically, these are footways or bicycle paths alongside a road. Of the remaining edges in each set, we identify those that have the same name (e.g. “Main Street”) and randomly select one edge to represent the set. The remaining edge or, rarely, edges are retained for purposes of calculating circuituity, but ignored for the aggregation (Figure 2D).

Overall, 8.0 million sets of parallel edges exist in our final dataset. Of the 17.7 million constituent edges, we tag 1.2 million as “ignore” because a parallel edge with the same OSM name exists.

As with many simplification algorithms, merging divided roads is a tradeoff between false positives (i.e., merging roads that are functionally separate) and false negatives (i.e., retaining two sides of a divided highway that are functionally a single road). It is also unclear where to set the dividing line — at what point do two parallel edges become functionally separate? Because we rely on heuristics and do not formally test the tradeoff, we err on the side of avoiding false positives by dropping only edges that meet both criteria (same OSM name and same start and end clusters of nodes). Thus, our approach improves on using the raw network, but does not address all cases of divided roads.

False negatives might arise where the name of an OSM way is missing or misspelled, or where the name differs between the two sides of a divided road. Spot inspections of a random sample of parallel edges suggests that where the OSM names differ, the edges are normally functionally separate (e.g., there are buildings between them) and using the OSM name is thus a useful discriminator. However, where the OSM name is missing (as is the case for 13.4 million parallel edges), more false negatives are likely to arise. The detailed algorithm (provided in the appendix) and our full code repository allow others to build on our work.

### 3.3 Other simplifications

As in our 2019 analysis, we drop duplicate geometries, and anneal degree-2 nodes, which by definition do not represent intersections. Their existence is an artefact of the import process or simply reflects the way data have been entered into OSM.

### 3.4 Simplification procedure

Because simplification can create further degree-2 nodes, we repeat certain steps of the process multiple times, as follows:

1. Create clusters: identify clusters of nodes where each node is within 20m of at least one other node in that cluster.
2. Within each cluster, randomly select one node to represent the cluster. Other intra-cluster nodes are retained but ignored in the subsequent aggregation, as are edges that link nodes in the same cluster.
3. Anneal degree-2 clusters: if a cluster only has two edges that link it to nodes outside the cluster, then compute the shortest path across that cluster. Retain that shortest path and remove other intra-cluster edges.
4. Anneal nodes: delete degree-2 nodes, i.e. any node that is neither a deadend nor an intersection.
5. Until no more changes are needed, repeat steps (1) through (4), as step (4) may have made some clusters smaller.
6. Drop footways and bicycle paths where they start and end in the same cluster as at least one other edge.
7. Anneal nodes, as step (6) may have created degree-2 nodes.
8. Identify and classify divided roads: where two or more edges start and end in the same cluster and have the same road name in OSM, randomly choose one of them to represent the road. Other edges are retained but ignored in the subsequent aggregation.
9. Anneal degree-2 clusters and anneal nodes as in steps (3) and (4) above, and repeat until no degree-2 nodes remain.

## 4 Computing connectivity measures

Detailed descriptions of the computation method for each connectivity measure from a network of nodes and edges are given in [Barrington-Leigh and Millard-Ball \(2019, Appendix A\)](#) and are unchanged. A summary table of the breadth and conceptual content of metrics is reproduced in [Table A1](#).

### 4.1 Circuitry correction

One change from our 2019 analysis relates to our circuitry measures, which we compute for every node  $i$  for multiple distance bands  $(d_1, d_2)$ . For each band  $(d_1, d_2)$ , e.g., 500m–1000m, we calculate the sum of the Euclidean distances from node  $i$  to every other node that lies between Euclidean distances  $d_1$  and  $d_2$ , and do the same for the network distances to every node in the same (Euclidean) distance band. We then calculate the log ratio of Euclidean to network distance. In our 2019 analysis, we simplified by excluding any node pairs greater than 3000m apart via the network. This underestimated circuitry, particularly for the higher distance bands. In the revised analysis, we relax this constraint to  $5d_2$  (e.g., 15km for the 2500m–3000m band). Where the network distance is greater than 15km or is not defined (as in an island with no road connection), we top-code it to 15km for computational reasons.

### 4.2 Urban region and time series identification

Our primary results are restricted to urbanized areas. In our 2019 analysis, we used a custom classification of urban areas based on country-specific density thresholds. Since then, the release of the Global Human Settlements Layer ([Pesaresi, 2023](#)) provides a consistent typology of urban settlements across countries, using the EUROSTAT “degree of urbanization” classification. Therefore, we now use the GHSL data (2023 release of GHS-SMOD) to identify urban areas. Specifically, we identify an edge or node as “urban” if it intersects a pixel classified by GHSL as “urban centre,” “dense urban cluster,” “semi-dense urban cluster,” or “suburban or peri-urban” (classes 30, 23, 22, and 21). The remaining classifications are various types of rural settlement.

We also use the GHSL data (specifically the 2023 GHS-BUILT-S raster) to develop a time series of street-network sprawl. Each GHS-BUILT pixel is classified as follows: land not built-up; built-up from 2005 to 2019 epochs; built-up from 1990 to 2004 epochs; built-up from 1975 to 1989 epochs; built-up before 1975 epoch. We calculate the built-up epoch based on when at least half of the ultimately-developed pixels in the 100m grid cell(s) that intersect each edge and node had been developed. For example, suppose that the intersecting grid cells have 20 pixels marked as developed by 1975, 60 by 1990, 70 by 2005, and 100 by 2020, with a further 50 pixels not built up in any of these epochs. Then, we would assign an epoch of 1975–89, as more than half of the built-up pixels (60/100) had been developed by 1990.

### 4.3 Aggregation

We aggregate to a 1km grid and to the administrative areas (countries and up to 5 subnational levels) demarcated in the Global Administrative Areas (GADM) dataset.

As in our 2019 analysis, we allow footpaths, bicycle paths, service roads, and similar edges to contribute to the connectivity of the network, but do not consider these nodes and edges in our aggregation, or as origins or destinations in the circuitry analysis. This exclusion helps to avoid

several potential biases. For example, the “service” tag is often used to identify private driveways, a practice that would inflate the fraction of dead ends. “Service” tags also represent other access roads that do not form part of the public street network, and aisles in parking lots, the internal connectivity of which has little relevance to urban form or travel behavior. The same is true for networks of walking paths in public parks, which can be represented in minimal or excruciating detail depending on the OSM contributor.

Even though these footpaths, bicycle paths, and service roads are *excluded* from the aggregation, they are *included* when computing graph properties and our measures of dendricity, nodal degree and circuitry. Inclusion of these edges markedly increases connectivity in places such as Denmark, where residential streets are often designed as deadends for cars but allow pedestrians and bicyclists to continue through. Intuitively, a pedestrian or cycle path improves our measure of connectivity indirectly, through its effect on nearby streets.

## 5 Results

The upper panel of Figure 3 shows SNDi at the country level for the entire stock of streets represented in OpenStreetMap. Countries with more connected streets (lower SNDi) are in blue, and those with higher SNDi in red. Connectivity is highest in much of South America (especially Argentina), continental Europe, North Africa, and parts of East Asia, especially Japan, South Korea, and Taiwan. At the other extreme, while the United States is the poster child for car-dependent suburban development, street connectivity is even lower (higher SNDi) in south and southeast Asia, particularly Thailand and Bangladesh.

The global differences in street connectivity are even more marked when considering streets constructed in the most recent of our four epochs, 2005-19 (Figure 3, lower panel). Southeast Asia, including Indonesia and the Philippines, is particularly notable for its high levels of SNDi in recent construction.

Within-country variation is also noticeable. In addition to the [downloadable data](#), our companion website at <https://sprawlmap.org> provides interactive maps of our results aggregated to five levels of sub-national administrative geographies and to a 1km grid.

Figure 4 shows trends in SNDi over the full range of our time series. In panels A and B, we have aggregated new development in urban areas across countries according to World Bank geographic and economic groupings. Many features are qualitatively consistent with our earlier work. However, updating the data to more recent years shows a generally steeper increase in street network sprawl in most regions and overall across Earth. Panel C shows trends aggregated to the country level for large countries, while panel D presents trends for large cities, but using a different time series which specifies urban development boundaries up to 2013 (Angel et al., 2012, 2016). While our previous work suggested that SNDi had plateaued in a number of locations and regions shown in Figure 4, the updated analysis provides instead a picture of increasingly disconnected development in the largest cities and generally in developing and middle income countries.

Even though the main feature of our new analysis is that the data are available at a highly disaggregated level, allowing for a variety of more detailed analysis, we provide one more example of high-level aggregation to extend the city-level picture. Figure 5 shows the average SNDi over the stock of streets in 2013 for a number of large cities, as well as that of new development in the 2000-2013 period. Patterns generally match the country-level picture of Figure 3.

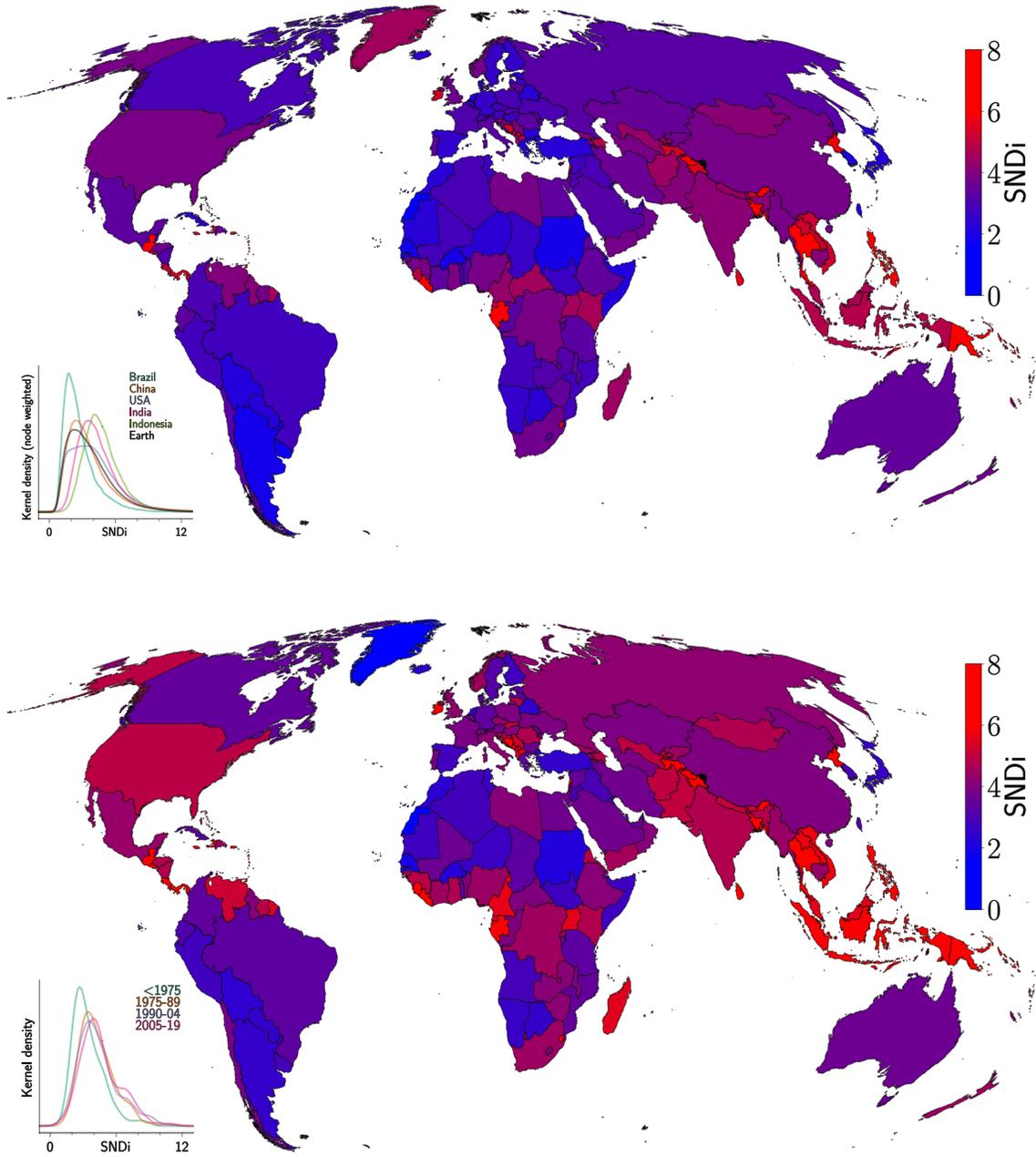


Figure 3: Upper panel: Street-network sprawl (SNDi) for the stock of streets, 2023. The inset shows the distribution (kernel density) of SNDi for 1km<sup>2</sup> grid cells in the five most populous countries. Lower panel: SNDi for streets developed in 2005–19. The inset shows how the distributions at the grid cell level (measured via kernel density) have shifted out over our four epochs, marking the shift towards less connected streets over time.

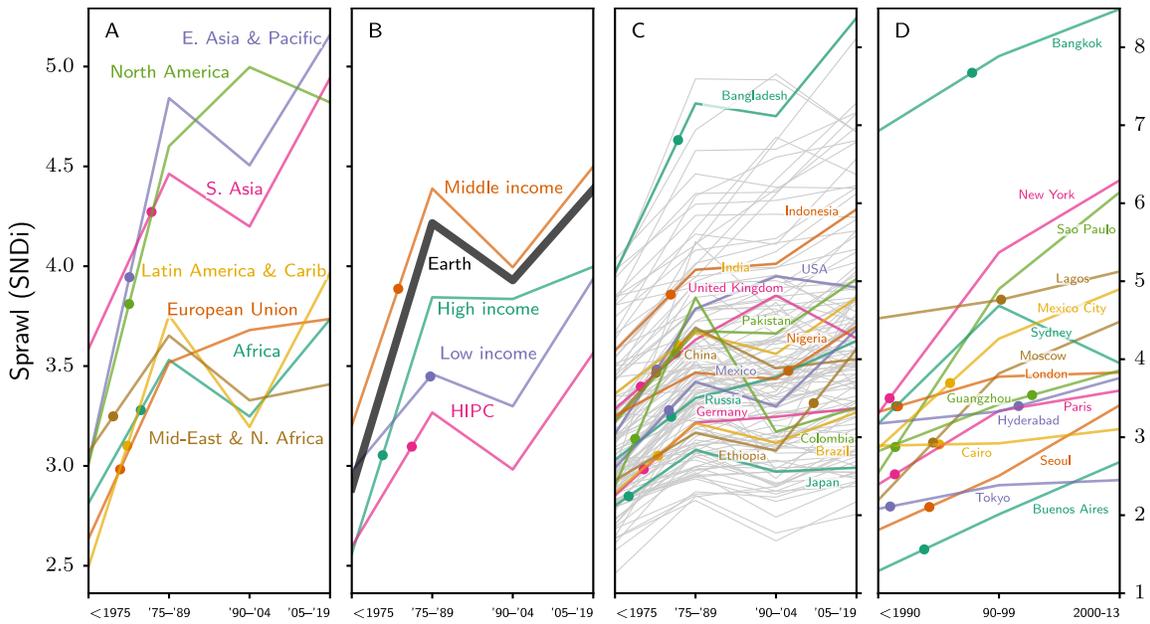


Figure 4: Trends in SNDi over time at the regional level (A), by the World Bank's country development classification (B), for a selection of large countries (C), and for a selection of large cities (D). HIPC is the World Bank's Heavily-Indebted Poor Country category. While trends show SNDi of new development during each time period, the solid circles show SNDi of the stock of all streets.

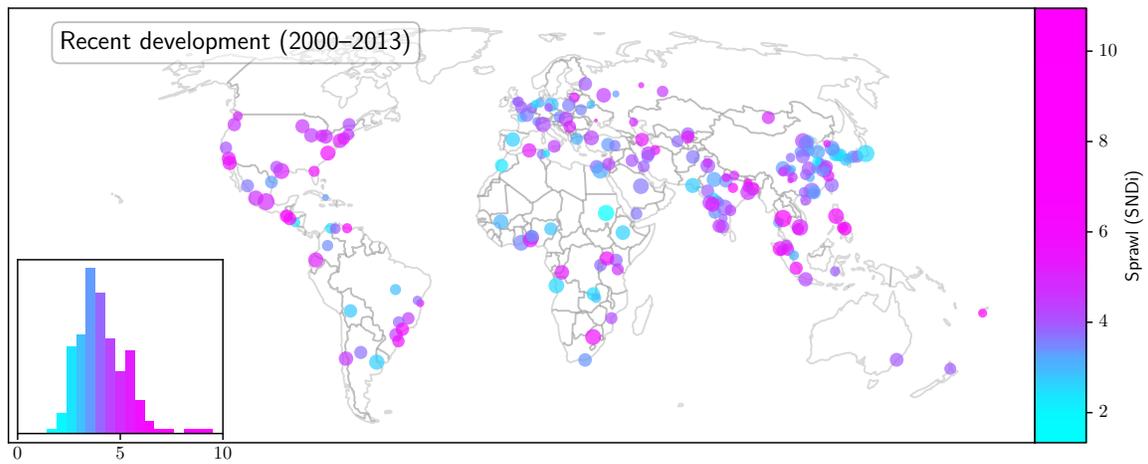
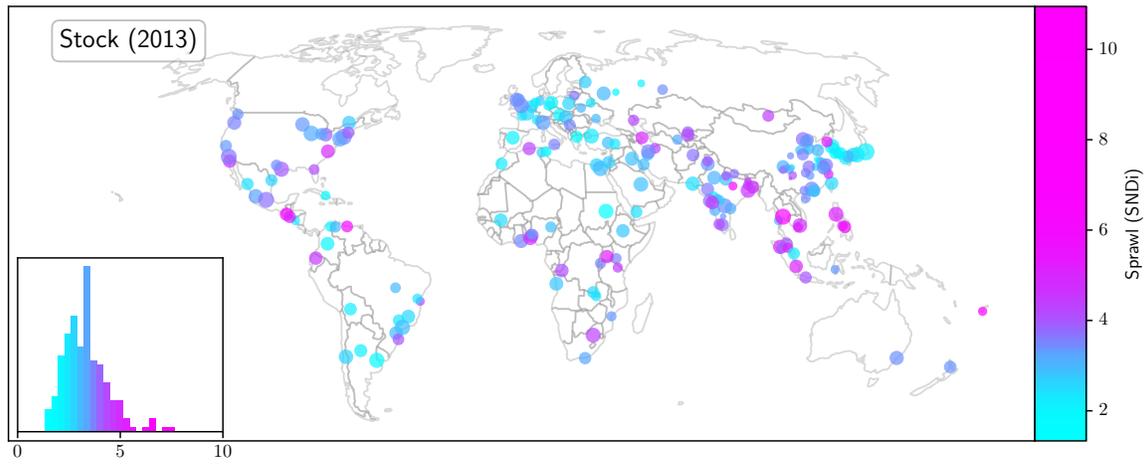


Figure 5: Distribution of SNDi for major cities in 2013 (top) and in recent development (bottom).

## 6 Data and code availability

Our analyses are available under a [CC BY 4.0 License](#) in three general geospatial formats:

1. as gridded data, specifying SNDi and its component metrics at 1km resolution,
2. as vector polygon data, aggregated to geopolitical boundaries at several spatial scales, using the Global Administrative Areas v4.1 dataset ([GADM, 2022](#)), from nodes and edges characterized as urban,
3. and as vector network data, specifying each node and edge and its characteristics.

Our provision of the underlying vector network data facilitates the creation of more detailed time series for specific locales. For instance, while our work relies on the GHSL in order to provide global coverage, data available for specific countries or metropolitan areas may offer higher temporal resolution (e.g. [Turner et al., 2023](#); [Barrington-Leigh and Millard-Ball, 2015](#)).

The data are hosted by OSF at <https://osf.io/c9hjy>.

We also provide our code as an open GitLab repository at [gitlab.com/cpbl/global-street-network-sprawl-sndi](https://gitlab.com/cpbl/global-street-network-sprawl-sndi). Given the hardware requirements and computation time, we suggest that our pre-prepared data products will be more valuable for most users. However, the code enables researchers to build on our algorithms and explore other extensions.

## 7 Conclusion

Our release of these analyses is intended to facilitate further exploration of patterns and trends in important characteristics of street networks worldwide. By releasing both spatially aggregated data and the underlying network layer, and both the SNDi and its component metrics, we hope to offer convenience and ease for moving beyond a simplistic and Americas-centric approach to quantifying street network connectivity. SNDi can serve as a point of comparison which has been validated across different historical development forms, and which detects stark and regionally-differentiated trends over recent decades. By providing a 45-year time series, we hope to facilitate studies of street connectivity that move beyond the descriptive to examine the causal role of streets in shaping economic, social, and environmental outcomes.

## References

- Angel S, Blei AM, Civco DL and Parent J (2012) *Atlas of urban expansion*. Cambridge, MA: Lincoln Institute of Land Policy. URL <http://atlasofurbanexpansion.org>.
- Angel S, Blei AM, Parent J, Lamson-Hall P, Sánchez NG, with Daniel L Civco, Lei RQ and Thom K (2016) *Atlas of Urban Expansion — 2016 Edition: Volume 1: Areas and Densities*. NYU Urban Expansion Program at New York University, UN-Habitat, and the Lincoln Institute of Land Policy. URL <http://www.lincolninst.edu/publications/other/atlas-urban-expansion-2016-edition>.
- Barrington-Leigh C and Millard-Ball A (2015) A century of sprawl in the United States. *Proceedings of the National Academy of Sciences* 112(27): 8244–8249.

- Barrington-Leigh C and Millard-Ball A (2017) More connected urban roads reduce US GHG emissions. *Environmental Research Letters* 12(4): 044008. DOI:10.1088/1748-9326/aa59ba. URL <https://doi.org/10.1088/1748-9326/aa59ba>.
- Barrington-Leigh C and Millard-Ball A (2019) A global assessment of street network sprawl. *PLoS ONE* 14(11): e0223078. DOI:10.1371/journal.pone.0223078. URL <https://doi.org/10.1371/journal.pone.0223078>.
- Barrington-Leigh C and Millard-Ball A (2020) Global trends toward urban street-network sprawl. *Proceedings of the National Academy of Sciences* DOI:10.1073/pnas.1905232116. URL <https://www.pnas.org/content/early/2020/01/13/1905232116>.
- Boeing G (2017) Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65(Supplement C): 126 – 139. DOI:<https://doi.org/10.1016/j.compenvurbsys.2017.05.004>. URL <http://www.sciencedirect.com/science/article/pii/S0198971516303970>.
- Boeing G (2020) A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science* 47(4): 590–608. DOI:10.1177/2399808318784595. URL <http://journals.sagepub.com/doi/10.1177/2399808318784595>.
- Boeing G (2021) Street Network Models and Indicators for Every Urban Area in the World. *Geographical Analysis* : gean.12281DOI:10.1111/gean.12281. URL <https://onlinelibrary.wiley.com/doi/10.1111/gean.12281>.
- Brenner AK, Haas W, Rudloff C, Lorenz F, Wieser G, Haberl H, Wiedenhofer D and Pichler M (2024) How experiments with superblocks in Vienna shape climate and health outcomes and interact with the urban planning regime. *Journal of Transport Geography* 116: 103862. DOI:<https://doi.org/10.1016/j.jtrangeo.2024.103862>. URL <https://www.sciencedirect.com/science/article/pii/S0966692324000711>.
- Ewing R and Cervero R (2010) Travel and the Built Environment. *Journal of the American Planning Association* 76(3): 265–294. DOI:10.1080/01944361003766766.
- Florczyk AJ, Corbane C, Ehrlich D, Freire S, Kemper T, Maffenini L, Melchiorri M, Pesaresi M, Politis P, Schiavina M, Sabo F and Zanchetta L (2019) GHSL Data Package 2019. DOI: 10.2760/290498. URL <https://ghsl.jrc.ec.europa.eu>.
- GADM (2022) Database of global administrative areas, v4.1. URL <http://gadm.org>.
- Hajrasouliha A and Yin L (2015) The impact of street network connectivity on pedestrian volume. *Urban Studies* 52(13): 2483–2497.
- Marshall W and Garrick N (2010) Effect of Street Network Design on Walking and Biking. *Transportation Research Record: Journal of the Transportation Research Board* 2198: 103–115. DOI: 10.3141/2198-12. URL <http://trrjournalonline.trb.org/doi/10.3141/2198-12>.
- ORNL (2011) LandScan High Resolution Global Population Data Set. URL <http://web.ornl.gov/sci/landscan/>.

- Pesaresi M (2023) GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). DOI:10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA. URL <http://data.europa.eu/89h/9f06f36f-4b11-47ec-abb0-4f8b7b1d72ea>.
- Rezaei N and Millard-Ball A (2023) Urban form and its impacts on air pollution and access to green space: A global analysis of 462 cities. *PLOS ONE* 18(1): e0278265. DOI:10.1371/journal.pone.0278265. URL <https://dx.plos.org/10.1371/journal.pone.0278265>.
- Salazar Miranda A (2021) The micro persistence of layouts and design: Quasi-experimental evidence from the united states housing corporation. *Regional Science and Urban Economics* : 103755DOI: <https://doi.org/10.1016/j.regsciurbeco.2021.103755>. URL <https://www.sciencedirect.com/science/article/pii/S0166046221001150>.
- Turner H, Lahoorpoor B and Levinson DM (2023) Creating a dataset of historic roads in sydney from scanned maps. *Scientific Data* 10(1): 683.
- United Nations (2019) World Urbanization Prospects: The 2018 Revision. Technical report, Department of Economic and Social Affairs, Population Division. URL <https://population.un.org/wup/Publications/>.
- Vazquez F, Millard-Ball A and Barrington-Leigh C (2023) Urban development & street-network sprawl in Tokyo. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* : 1–20DOI:10.1080/17549175.2023.2262698. URL <https://www.tandfonline.com/doi/full/10.1080/17549175.2023.2262698>.

## A Appendix

### A.1 Algorithm to simplify the street network

Figure A1 summarizes the algorithm that we use to simplify the street network, moving from the geometric representation provided by OSM to a functional representation that reflects the connectivity of the network. See the main text for more context and justification of why simplification is important. The key steps in Figure A1 are as follows, with select steps illustrated in Figure A2:

#### A.1.1 Table of edges

This is the database table produced by *osm2po*. Each row corresponds to one edge, and includes the edge geometry, the ids of the start and end nodes, and a code for the OSM tag such as *motorway*, *primary*, *footway*, etc. We tag footways, bicycle paths, service roads (which can be driveways, alleys, parking aisles, and similar edges) as “bicycle/pedestrian only.”

#### A.1.2 Drop redundant edges

We remove edges that will not affect the connectivity calculation. This step simplifies the network representation and speeds up subsequent analysis. Three distinct types of edges are dropped:

1. Isolated self loops. These are defined as where the start node is the same as the end node and where no other edges start or end at that node. (*Figure A2, Panel A*)
2. Edges that are tagged in OpenStreetMap as sidewalks or footways. By definition, these edges parallel or cross a road, and so do not affect connectivity.
3. Bicycle/pedestrian-only deadends. These do not affect connectivity as bicycle/pedestrian-only edges and nodes are excluded from the aggregation, and deadends cannot improve the connectivity of other edges. (*Panel B*)
4. Step 3 is repeated as deadends may be chained. (*Panel B*)

#### A.1.3 Identify clusters of nodes

A cluster is defined as a set of nodes where each node in the set is connected to at least one other node in the set by an edge of length  $\leq 20m$ . To avoid double weighting complex intersections, our connectivity metrics are based on clusters, not on their component nodes.

#### A.1.4 Anneal clusters

Annealing removes degree-2 clusters. Such clusters are an artefact of the *osm2po* import process, or of how the street network is represented in OSM. Degree-2 clusters are distinct from degree-2 nodes, which are handled in a subsequent step.

Our cluster annealing process has several distinct steps:

1. Drop deadends that start and end in the same cluster. (*Panel C*)
2. Drop isolated clusters, defined as clusters that are not connected to any other cluster.

3. Identify the edges that comprise the shortest path through the cluster. (*Panel D*)
4. Drop all other edges that are within the cluster (length  $\leq 20\text{m}$ , and starting and ending in the same cluster) (*Panel D*)

Degree-2 nodes will remain at this point, but these are dropped in the subsequent step (shown in *Panel E* and discussed in the next subsection).

#### A.1.5 Anneal nodes

Our process for annealing edges with degree-2 nodes is the same as in [Barrington-Leigh and Millard-Ball \(2019\)](#). The following text is reproduced from that source:

For each 2-degree node, we merge the two edges that meet at that node, and delete the node from the database. Our algorithm for accomplishing this is somewhat involved, because (1) in many cases, degree-2 nodes are not isolated, but exist in chains; (2) such chains have the possibility of being loops; and (3) edges are coded with a “start” and an “end” node, so that for any given degree-2 nodes, there are four possible relationships with the adjoining edges. Because of (1) and (2), one cannot safely process a long list of apparent degree-2 nodes with a single pass. We proceed in the following steps:

1. Delete all isolated loops constituted of exactly two parallel (in the graph theoretic sense) edges.
2. Select all edges with degree-2 nodes at *both* ends (“double-degree-2-edges”), build a network data structure from them, and find all connected components (using *networkx*, a network theory package implemented in Python). This gives all the chains consisting solely of degree-2 nodes.
3. These chains are annealed (nodes removed; edges joined) and re-inserted into the database. This replaces each original multi-edge double-degree-2-edge chain with a single double-degree-2-edge.
4. Repeatedly edit the edges/nodes database tables, annealing edges which have only one degree-2 node. This should properly reduce loops to simple self loop edges. When one edge of a pair has two degree-2 nodes, keep its edge id and drop the other’s.
5. Drop isolated self-loops, that is, edges which have a single degree-2 node that connects the start and end points, and are not connected to any other edges.
6. Drop parallel edges that have the same start and end node, and are not open to motor vehicles. Most of these parallel edges consist of sidewalks and walking paths that are represented separately from the street in the OSM database.
7. Drop island edges and nodes (where both nodes or both clusters are degree-1)

#### A.1.6 Drop parallel bicycle/pedestrian edges

We drop bicycle/pedestrian-only edges where there is a parallel edge (i.e., an edge with the same start node and the same end node) that is open to cars. These dropped edges do not affect connectivity because (i) bicycle/pedestrian-only edges are ignored in the aggregation, and (ii) almost-identical connectivity is provided by the parallel edge that is retained.

### **A.1.7 Classify divided roads**

We identify divided roads, where a road is represented by two or more parallel edges. We choose one of the edges to represent the road, and ignore the others in the aggregation. Typically, this situation arises when a road with a median is mapped in OSM as two separate edges, one for each direction of travel. The main text provides more discussion of this issue.

## **A.2 Decomposition of changes**

In most countries, the new algorithm changes our results only modestly, meaning that our 2019 interpretations of SNDi are qualitatively unchanged. Among large countries (Figure A3), the most marked changes are an increase in SNDi in Japan and Brazil, which is due mostly to the reweighted Principal Component Analysis (PCA) coefficients rather than changes in the underlying metrics, but these countries are still ranked as highly connected (low SNDi) in comparative terms. The addition of new data, moving from our 2019 analysis (based on an August 2018 OSM vintage) to the October 2023 vintage has little effect in most large countries, highlighting the stability of our composite SNDi measure over four years of additions by OpenStreetMap contributors.

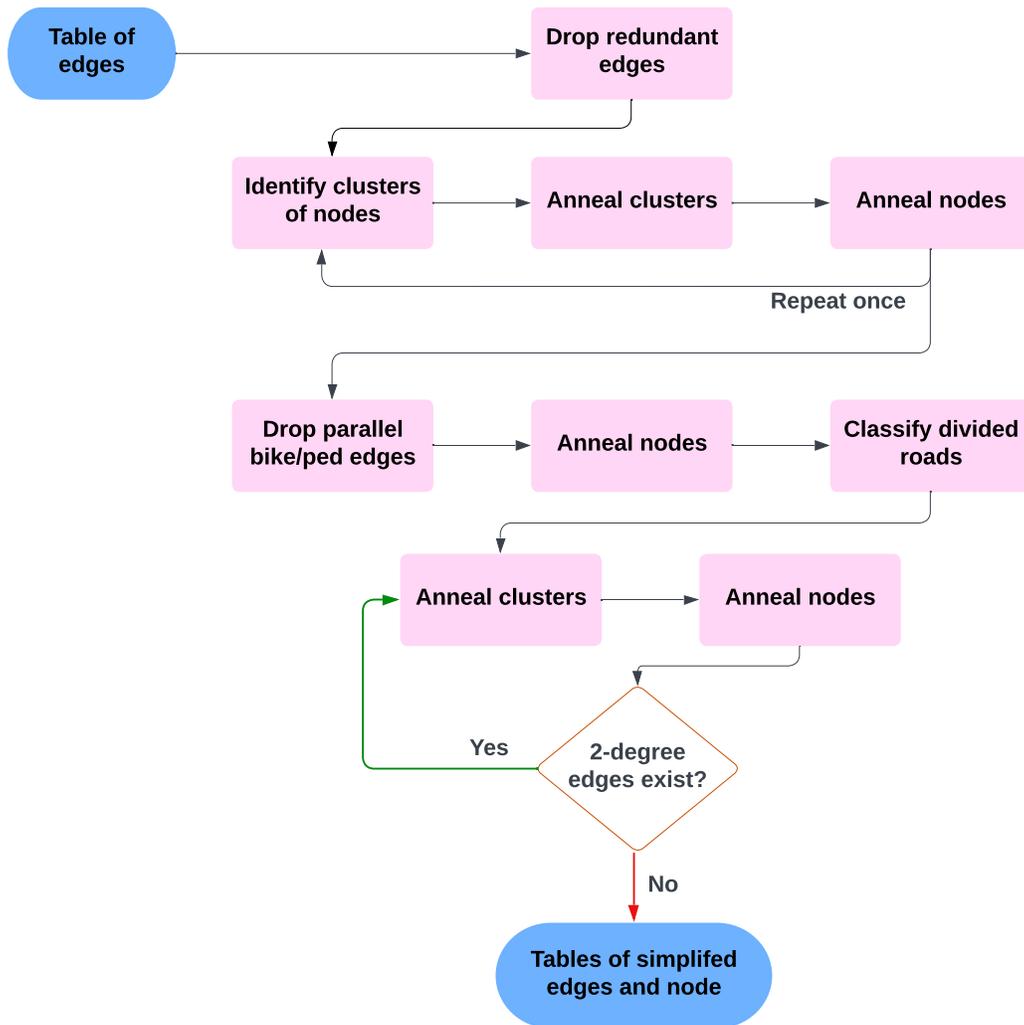


Figure A1: Summary of algorithm to simplify the road network

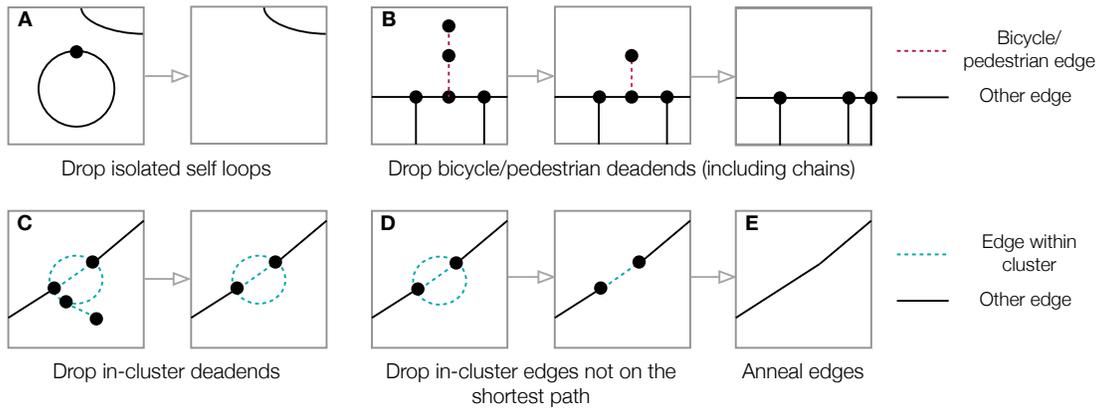


Figure A2: Examples of specific steps in the simplification process

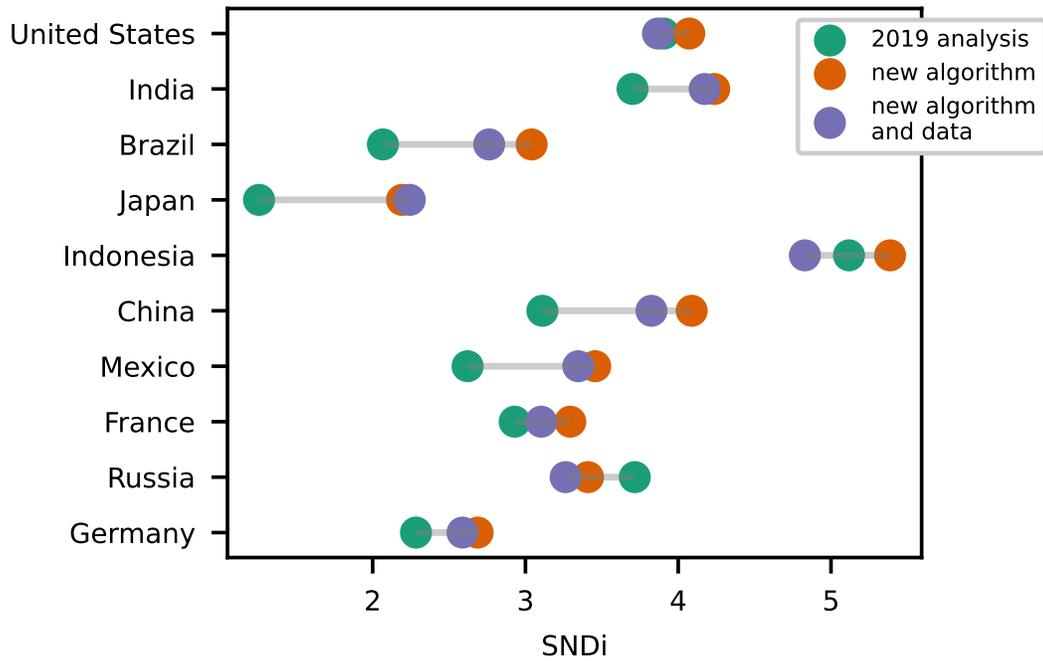


Figure A3: Comparison of 2019 and present analysis. The 10 largest countries (by number of nodes) are shown. The difference between the green and orange dots represents the impacts of our new algorithm. The difference between the orange and purple dot represents the impact of updated OSM data.

### A.3 Description of connectivity measures

	Concept	Core metric	Aggregated measure(s)	
Edges	Edge classification (dendricity)	dead-end	Fractions of edges; fraction of length	graph-theoretic
		self-loop		
		other bridge		
		part of a cycle		
Sinuosity		path length	Gross sinuosity, i.e. summed path length / summed end-to-end length	geographic
		end-to-end length		
Nodes	Nodal degree	dead-end	Mean degree; fraction of dead-ends and fraction of degree<4	graph-theoretic
		degree 3		
		degree>3		
	Node-node distance ratios (circuitry)	$r \leq 500$ m	Summed path length over summed end-to-end length for all proximate node pairs within radius $r$	geographic
		$500 \text{ m} < r \leq 1000$ m		
		$1000 \text{ m} < r \leq 1500$ m		
		$1500 \text{ m} < r \leq 2000$ m		
		$2000 \text{ m} < r \leq 2500$ m		
	Nodal density	Nodes within 500 m	Node-weighted mean area density	
		Nodes within 1000 m		
Nodes within 1500 m				
Nodes within 2000 m				
	Nodes within 2500 m			

Table A1: **Overview of core and aggregate connectivity measures.** Our database consists of computed values of the core metrics for each edge and node. When expressing aggregate metrics for administrative or city regions, we use the approach in the “Aggregation measure(s)” column. For our preferred results, we include only the edges and nodes we deem to be urban and that are open to motor vehicles, when aggregating. In order to align our metrics with increasing street-network sprawl, we use sparsity (inverse density) in place of density, and negative degree in place of mean degree.

Type	Node or edge	Definition	Considered when calculating nodal degree	Considered when calculating shortest paths	Considered as origins or destinations in circuitry calculations	Considered when calculating graph theoretic measures	Included in aggregation
Intra-cluster	node	Connected to an edge with length $\leq 20\text{m}$	N/A	Yes	One randomly chosen node represents each cluster	Yes	One randomly chosen node represents each cluster
Intra-cluster	edge	Link nodes within the same cluster, and are $\leq 40\text{m}$ in length	No	Yes	N/A	Yes	No
Divided road	edge	Edges that start and end in the same cluster, and have the same name in OSM	One randomly chosen edge represents each road	Yes	N/A	Yes	One randomly chosen edge represents each road
Bicycle and pedestrian	edge	Classified in OSM as pedestrian, path, track, cycleway, footway, steps, crossing, service	Yes	Yes	No	Yes	No
Bicycle and pedestrian	node	Would not exist if all bicycle and pedestrian edges were deleted, i.e. would be degree-0 or degree-2 after ignoring those edges	N/A	Yes	No	N/A	No
Core	both	All other edges and nodes	Yes	Yes	Yes	Yes	Yes

Table A2: We designate certain nodes and edges as intra-cluster, divided road, or bicycle and pedestrian. As discussed in the main text, this allows us to mitigate the dependence of our results on how the street network is represented in OSM, and to capture how bicycle and pedestrian paths contribute to connectivity.

#### A.4 Definition of Street Network Disconnectedness Index (SNDi)

Aggregated measure	PCA <sub>1</sub>	Mean	StdDev
Nodal degree (negative)	.29	-2.7	.41
Fraction deadends	.27	.23	.16
Log <sub>10</sub> Circuity (0-0.5km)	.24	.042	.098
Log <sub>10</sub> Circuity (0.5-1km)	.30	.18	.17
Log <sub>10</sub> Circuity (1-1.5km)	.32	.22	.17
Log <sub>10</sub> Circuity (1.5-2km)	.32	.037	.067
Log <sub>10</sub> Circuity (2-2.5km)	.31	.040	.036
Log <sub>10</sub> Circuity (2.5-3km)	.30	.22	.15
Fraction bridges (length)	.20	.22	.14
Fraction non-cycle (length)	.28	.21	.13
Fraction non-cycle (N edges)	.32	.20	.12
Fraction bridges (N edges)	.25	.19	.11
Log <sub>10</sub> Curviness	.16	.18	.10
Variance explained	48%		
Eigenvalue	6.3		

Table A3: Definition of SNDi. The first column shows the first principal component loadings. The second and third columns provide the global means and standard deviations. To calculate SNDi for a region, calculate the given aggregated measures from the nodes and edges in the region. Then subtract the Mean value, divide by the StdDev, multiply by the loadings, and take the sum. Finally, we add an arbitrary value of 4.64 to the result to get the published SNDi; this assures that observed values tend to be positive.