Supplementary information for ‘Network analysis of the Danish bicycle infrastructure: Bikeability across urban-rural divides’

Ane Rahbek Vierø[1] and Michael Szell[1,2]

[1]Networks, Data, and Society (NERDS), Data Science Section, IT University of Copenhagen, 2300 Copenhagen, Denmark
[2]Complexity Science Hub, Vienna, Austria

This is the supplementary information for the manuscript ‘Network analysis of the Danish bicycle infrastructure: Bikeability across urban-rural divides’. The document contains:

- Further details on the included OpenStreetMap (OSM) road and path network (Section 1).
- Illustration of updates to the OSM data set (Fig. S1).
- Illustration of method used for conflating OSM and GeoDanmark network data: Fig. S2
- Network density maps for the car and the total network (Fig. S3).
- Moran’s I values for test for spatial autocorrelation (Table S1).
- Maps of significant clusters of local spatial autocorrelation (Figures S4 to S7).
- Scatter plots for the relationship between LTS network densities and population densities (Fig. S8).
- Scatter plots for the relationship between network reach and population densities (Fig. S9).
- Histogram of the local component count (S10).
- Correlation heatmap for network metrics at the hex grid level (Figure S11).
- Cluster means for bikeability clusters (Table S2).
- Density distributions for cluster variables (Fig. S12).
- Overview of the effect on identified gaps and the number of disconnected components from the LTS gap threshold size (Figure S13).

1 Definition of input network

Below, we define the network included in the study in more detail and list the most important queries used in the network definition. For all code used in the network definition and classification, see the code repositories linked in the main paper.

- The network used in this study, as a starting point, includes all roads, paths, and bicycle tracks and lanes that have public access and are part of the public road network (i.e. excluding all OSM ways with access restrictions or OSM ways tagged with ‘highway’=‘service’ and ‘service’=‘driveway’).

- The *bikeable* subset of the total network includes all road categories where cycling is not prohibited. I.e. the bikeable network excludes OSM ways tagged with 'highway' = 'motorway'/'trunk'; 'motorroad' = 'yes'; or 'highway' = 'pedestrian'/'path'/'footway'/'bridleway', unless explicitly tagged as allowing cyclists. The bikeable network further excludes all OSM ways tagged as 'bicycle' = 'dismount'/'use_sidepath'/'no' and OSM ways tagged as having separate bicycle infrastructure ('cycleway'/'cycleway:left'/'cycleway:right'/'cycleway:both' = 'separate'/'use_sidepath'). Finally, the bikeable network excludes paths and bridleways where the surface is not conducive for everyday, utilitarian cycling (e.g. sand, clay, grass, woodchips, etc.), as well as all ways where 'highway' = 'track'.

- The *car* subset of the total network includes all OSM ways in the total network with a 'highway' type that by default allows for motorized traffic and is not explicitly tagged as 'motorcar'/'motor_vehicle' = 'no'.

- The bikeable and car network combination makes up the total final network used in the study, thus excluding all pedestrian-only infrastructure.

2 Figures & tables

Fig. S1: Separate mapping of road and bicycle track in OSM. White: Main road. Blue, dashed: Bicycle track running in parallel with a road but mapped separately. To avoid including the main road in the higher-stress bicycle network, a tag is added specifying that the bicycle tracks have been mapped separately.

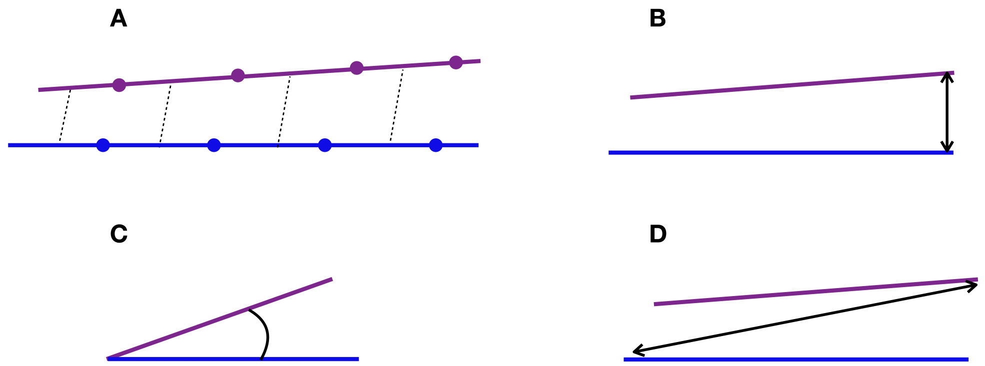

Fig. S2: Network conflation. The OSM network data is enriched with additional data on bicycle lanes and tracks using a feature matching method developed by (Vierø et al., 2024) with inspiration from (Koukoletsos et al., 2011). The feature matching algorithm first identifies potential matches for corresponding edge segments (A) based on the maximum distance between them (B) and then computes the best match based on the angle (C) and undirected Hausdorff-distance (D) (the maximum distance between any two points in the two corresponding line segments) between them.

Fig. S3: Network density for the car and the total network. A) Car network – absolute density. B) Car network – relative density. C) Total network – absolute density. Absolute network density values are in km/km^2, relative density values are in %.

Metric	Moran's I	p-value
Density - LTS 1	0.64	0.001
Density - LTS 2	0.63	0.001
Density - LTS 3	0.27	0.001
Density - LTS 4	0.28	0.001
Density - car	0.61	0.001
Density - total network	0.64	0.001
Local share of network - LTS 1	0.40	0.001
Local share of network - LTS 2	0.34	0.001
Local share of network - LTS 3	0.34	0.001
Local share of network - LTS 4	0.31	0.001
Local share of network - car	0.28	0.001
Components per km - LTS 1	0.00	0.044
Components per km - LTS≤2	0.00	0.306
Components per km - LTS≤3	0.00	0.267
Components per km - LTS≤4	0.00	0.367
Components per km - car	0.00	0.356
Components per km - total network	0.34	0.001
Largest component length - LTS 1	0.79	0.001
Largest component length - LTS≤2	0.81	0.001
Largest component length - LTS≤3	0.81	0.001
Largest component length - LTS≤4	0.93	0.001
Largest component length - car	0.80	0.001
Reach - 5 km - LTS 1	0.87	0.001
Reach - 5 km - LTS≤2	0.92	0.001
Reach - 5 km - LTS≤3	0.91	0.001
Reach - 5 km - LTS≤4	0.87	0.001
Reach - 5 km - car	0.84	0.001

Table S1: Global spatial autocorrelation: Moran's I. Spatial autocorrelation for results aggregated at the hex grid level. A case of maximum spatial clustering results in a Moran's I value of 1, while the opposite situation (dissimilar values are located next to each other) results in a Moran's I of a value of -1. A random spatial organization of values results in a Moran's I value around 0.

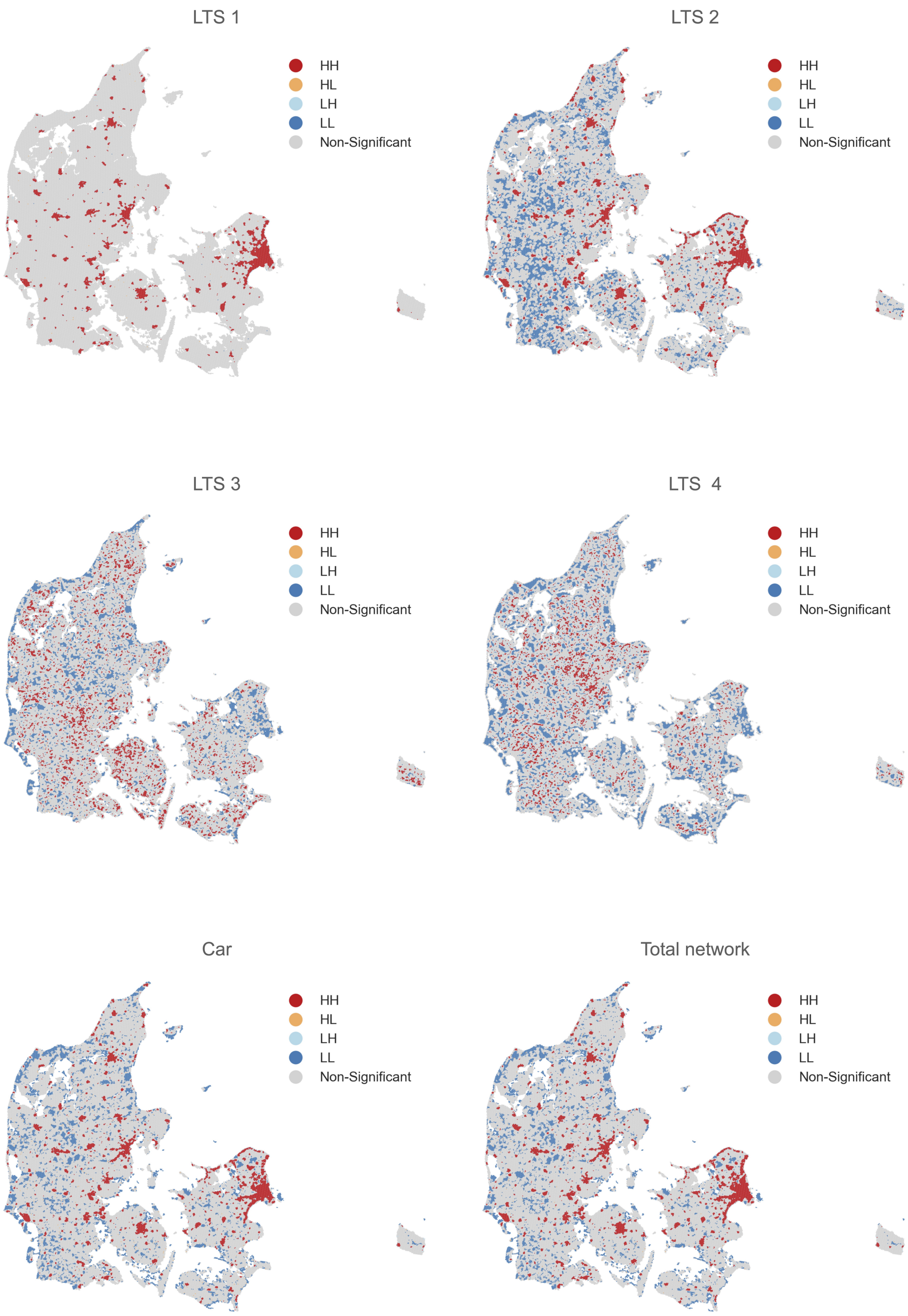

Fig. S4: Local indicators of spatial autocorrelation: Network density. Areas in red, 'High-High' (HH) represent clusters with a high network density of the given LTS (significant at $p<0.05$). Blue areas, 'Low-Low' (LL), indicate significant clusters of areas with a low density. 'High-Low' (HL) represents high LTS network densities surrounded by low densities, and 'Low-High' (LH) represents low densities surrounded by high densities. Grey: Areas where the spatial clustering of similar/dissimilar values is not statistically significant.

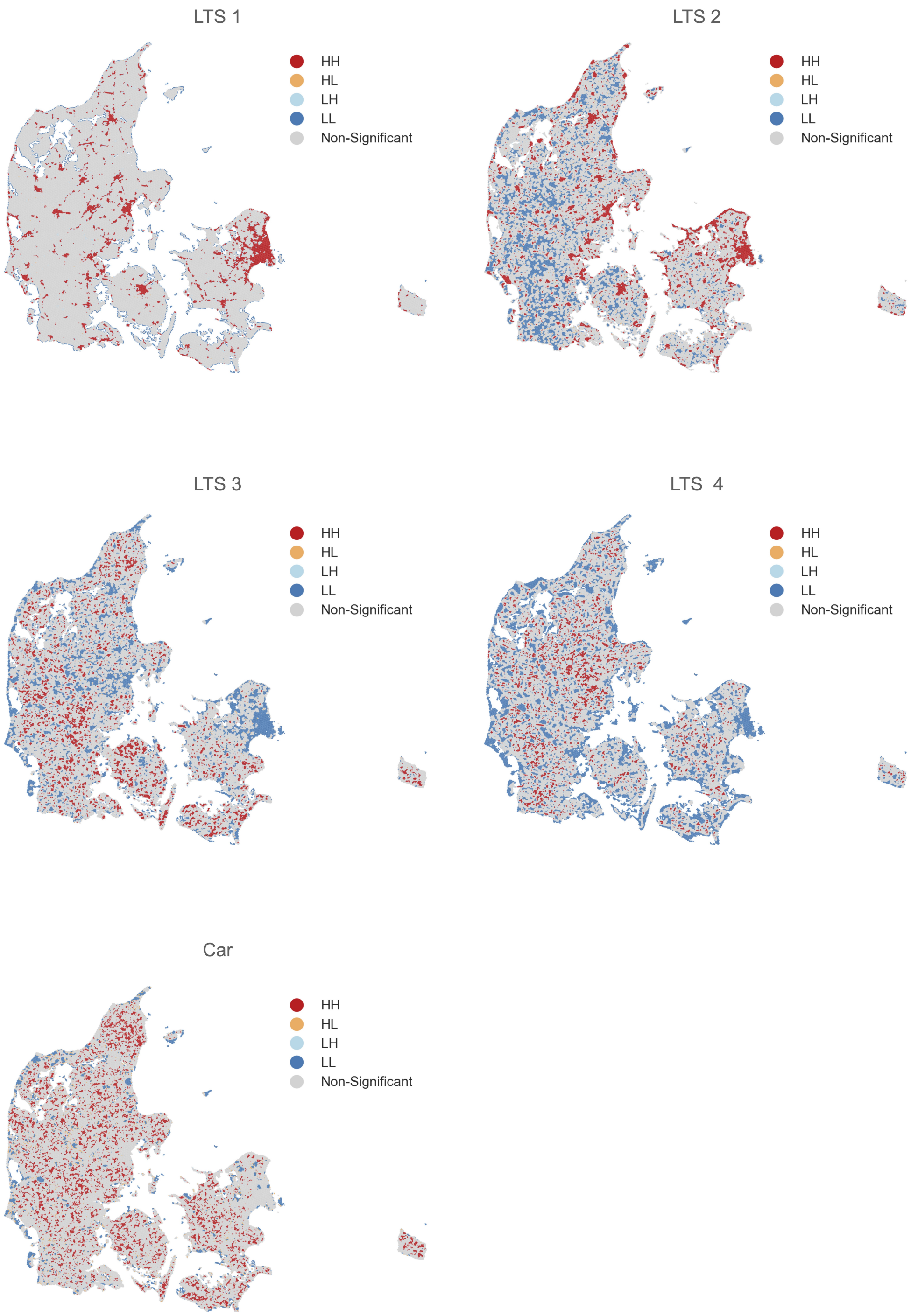

Fig. S5: Local indicators of spatial autocorrelation: Relative network density. Areas in red, 'High-High' (HH) represent clusters with a high local percentage of the given LTS level (significant at $p<0.05$). Blue areas, 'Low-Low' (LL), indicate significant clusters of areas with a low percentage. 'High-Low' (HL) represents a high local percentage surrounded by low local percentages, and 'Low-High' (LH) represents low local percentages surrounded by high percentage values. Grey: Areas where the spatial clustering of similar/dissimilar values is not statistically significant.

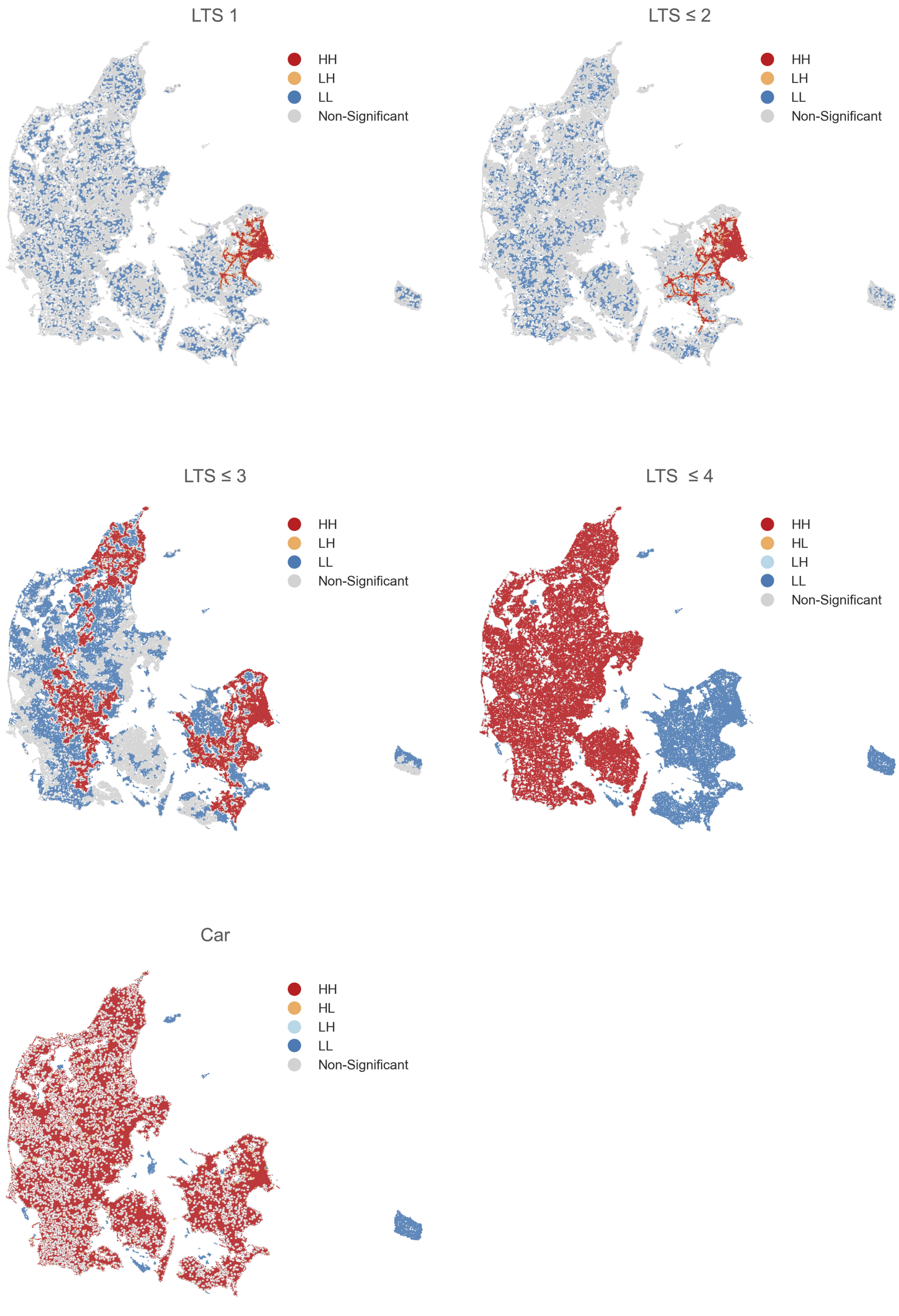

Fig. S6: Local indicators of spatial autocorrelation: Largest local component length. Areas in red, 'High-High' (HH) indicate clusters where the largest local component is large (significant at $p<0.05$). Blue areas, 'Low-Low' (LL), represent significant clusters of areas where the largest local component is small. 'High-Low' (HL) represents a large local component size surrounded by small largest local components, and 'Low-High' (LH) represents small largest local component sizes surrounded by large local components. Grey: Areas where the spatial clustering of similar/dissimilar values is not statistically significant.

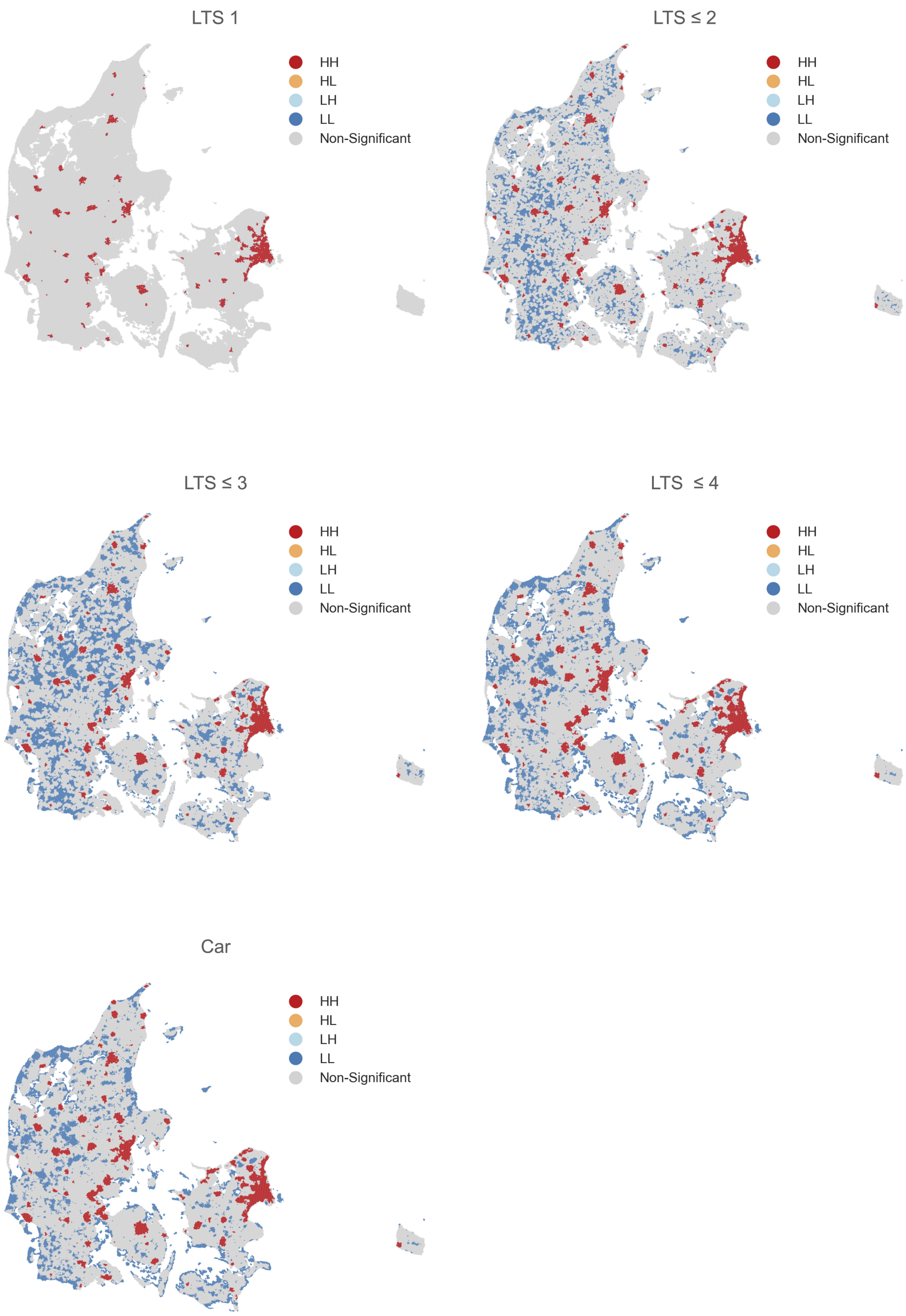

Fig. S7: Local indicators of spatial autocorrelation: Network reach. Areas in red, 'High-High' (HH) represent statistically significant clusters of high reach ($p<0.05$). Blue areas, 'Low-Low' (LL), indicate significant clusters of low reach. 'High-Low' (HL) represents high reach values surrounded by low reach values, and 'Low-High' (LH) represents low reach values surrounded by high reach values. Grey: Areas where the spatial clustering of similar/dissimilar values is not statistically significant.

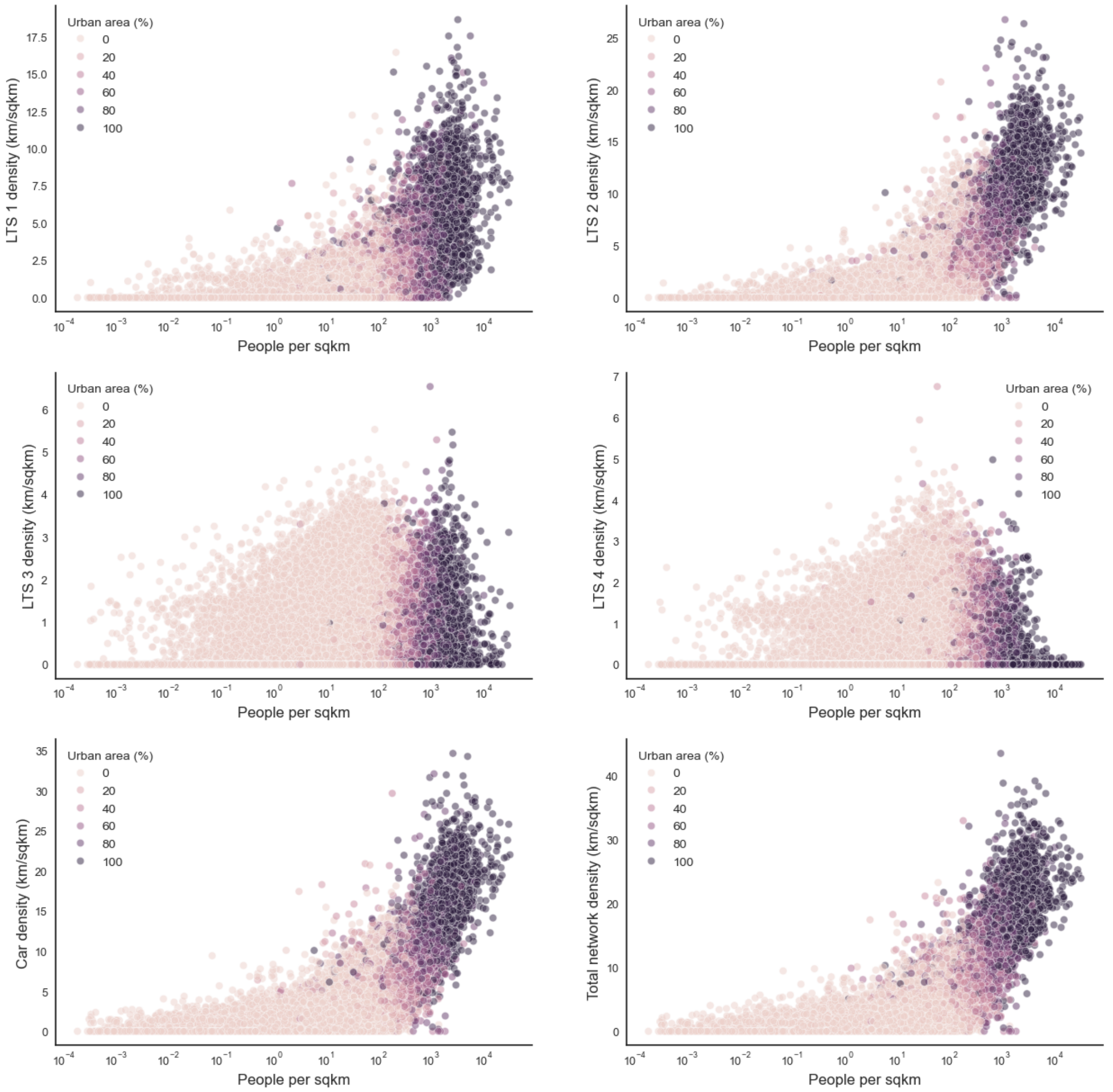

Fig. S8: Network density and population density. Color represents the percentage of the hex grid covered by an urban zone.

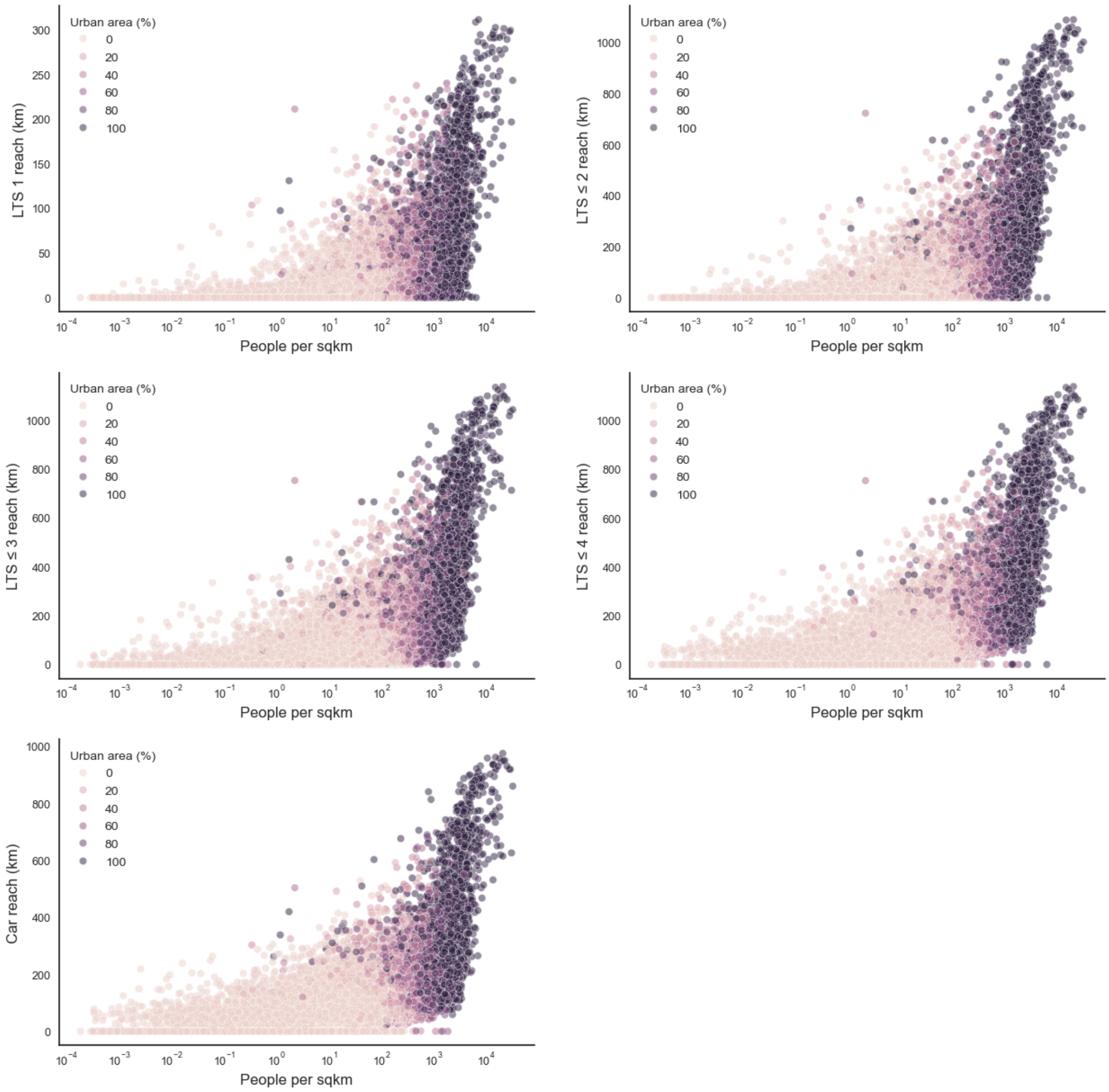

Fig. S9: Network reach (distance threshold = 5 km) and population density. Color represents the percentage of the hex grid covered by an urban zone.

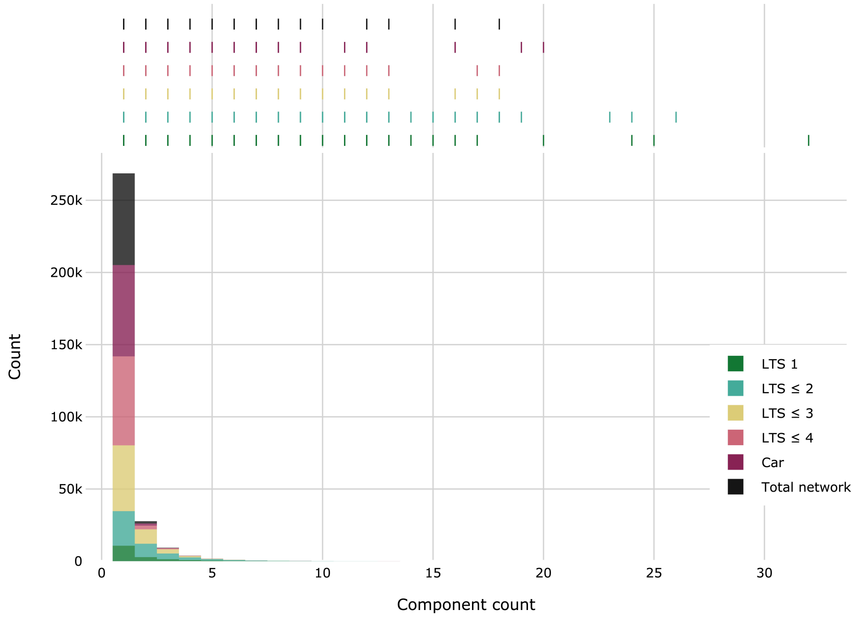

Fig. S10: Local component count distribution. Most hex grids have only a few components at any level of traffic stress, but a few locations have high component counts, particularly for LTS 1 and LTS$\leq$2.

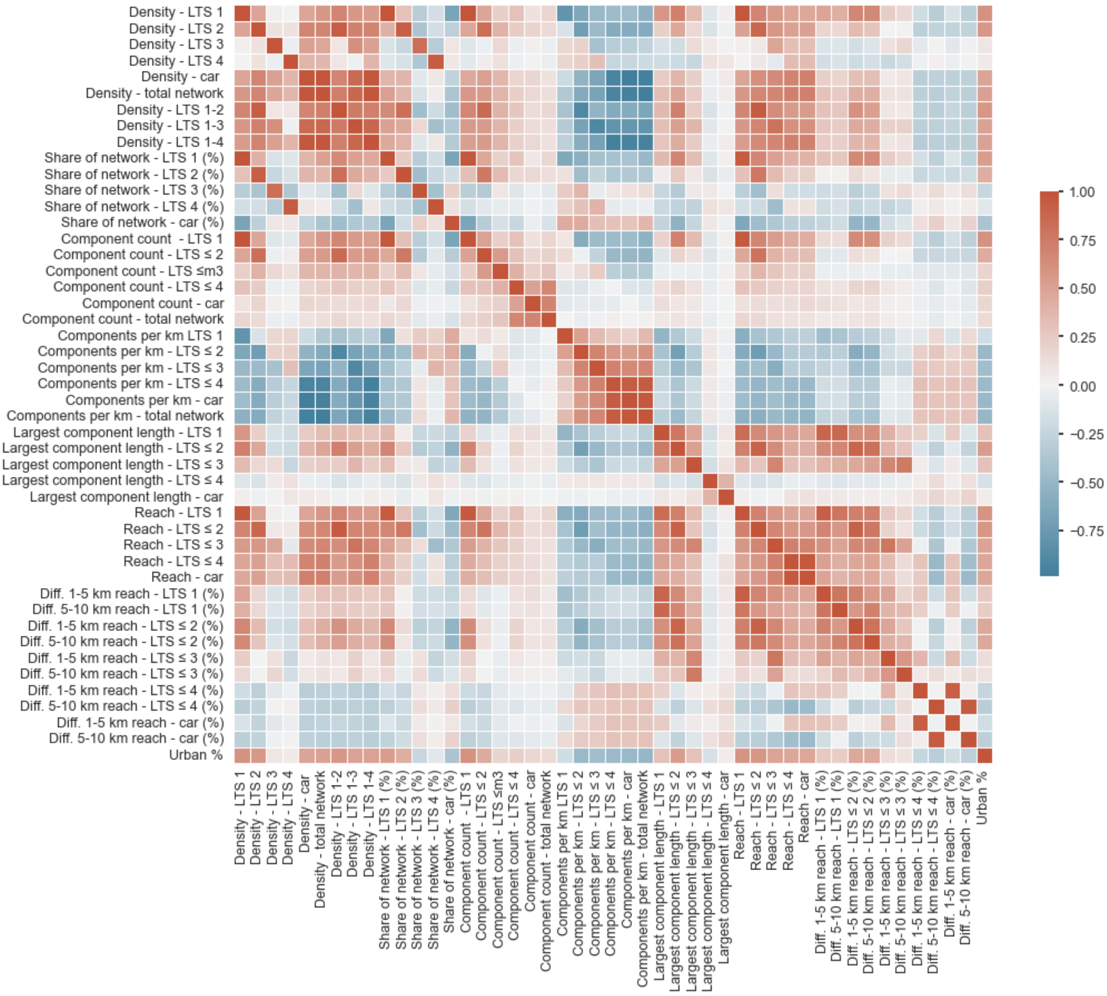

Fig. S11: Correlation heatmap. Correlation between network metrics at the hex grid level, using Spearman's rank correlation coefficient.

K-means variables	**1**	**2**	**3**	**4**	**5**
Density - LTS 1	0.08	1.96	1.87	4.29	6.55
Density - LTS 2	0.56	3.35	2.42	7.28	10.93
Density - LTS 3	0.70	0.67	0.64	0.52	0.36
Density - LTS 4	0.46	0.39	0.38	0.23	0.10
Density - car	1.83	5.60	4.75	10.56	16.18
Density - total network	1.89	7.02	6.15	13.65	19.76
Local share of network - car (%)	79.39	76.62	75.38	75.84	80.66
Reach - LTS 1	0.10	12.52	17.70	72.01	174.52
Reach - LTS≤2	4.47	56.65	60.07	275.90	671.88
Reach - LTS≤3	21.08	88.31	97.84	329.53	731.32
Reach - LTS≤4	60.11	124.97	136.98	356.25	740.81
Reach - car	58.04	114.43	120.81	288.31	601.90
Diff. 1-5 km reach - LTS 1 (%)	0.80	127.72	150.53	176.58	185.70
Diff. 5-10 km reach - LTS 1 (%)	0.00	18.58	92.52	83.20	109.69
Diff. 1-5 km reach - LTS≤2 (%)	17.01	152.21	165.59	182.73	187.93
Diff. 5-10 km reach - LTS≤2 (%)	3.91	31.39	124.16	76.23	104.83
Diff. 1-5 km reach - LTS≤3 (%)	91.56	167.93	175.13	183.34	187.80
Diff. 5-10 km reach - LTS≤3 (%)	44.40	69.54	120.98	82.35	104.86
Diff. 1-5 km reach - LTS≤4 (%)	139.81	177.54	182.27	184.15	187.87
Diff. 5-10 km reach - LTS≤4 (%)	96.86	104.82	127.02	92.64	106.18
Diff. 1-5 km reach - car (%)	140.54	175.60	179.67	185.03	188.85
Diff. 5-10 km reach - car (%)	97.76	105.05	125.41	98.80	111.02

Table S2: Cluster means.

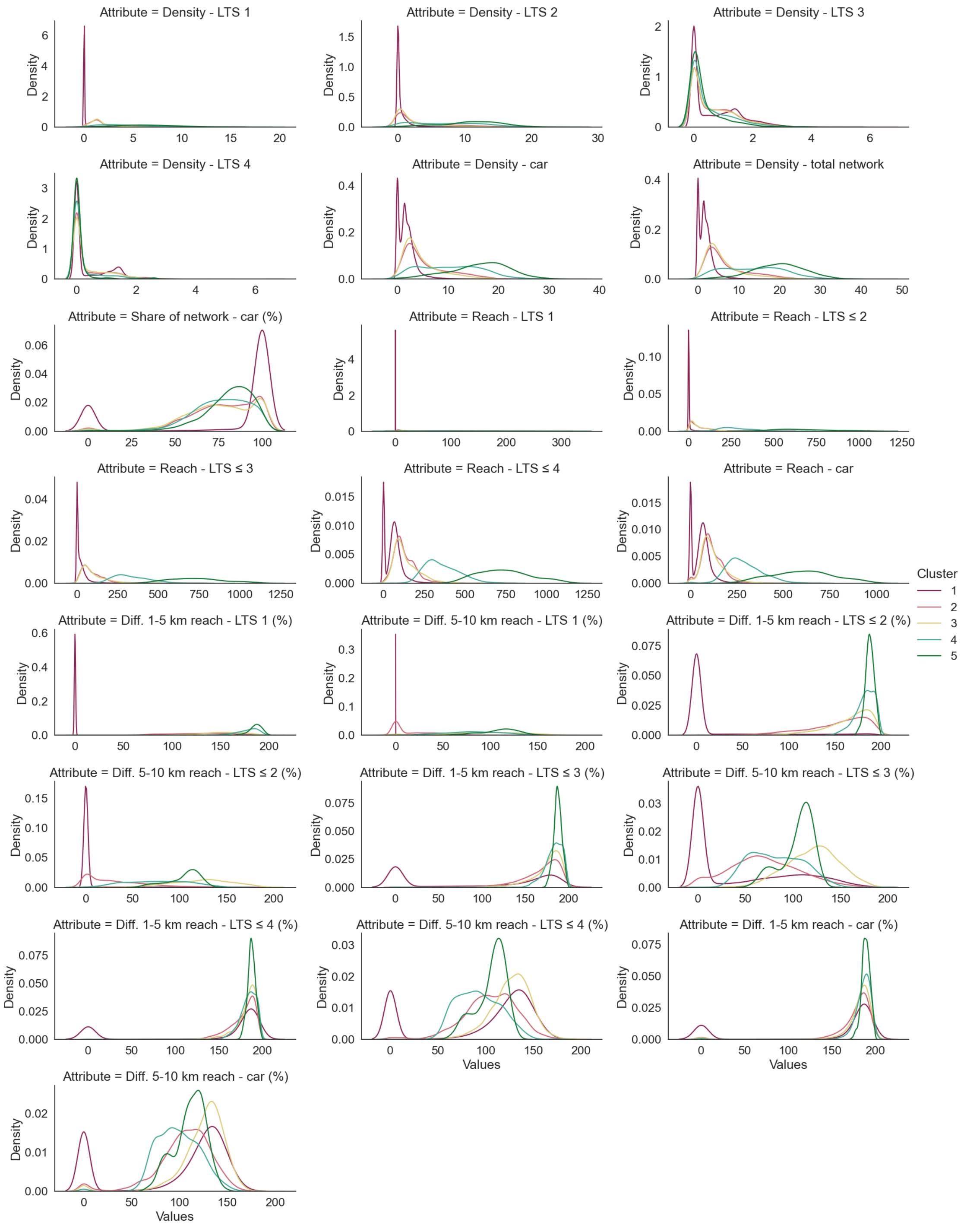

Fig. S12: Density distribution of cluster variables.

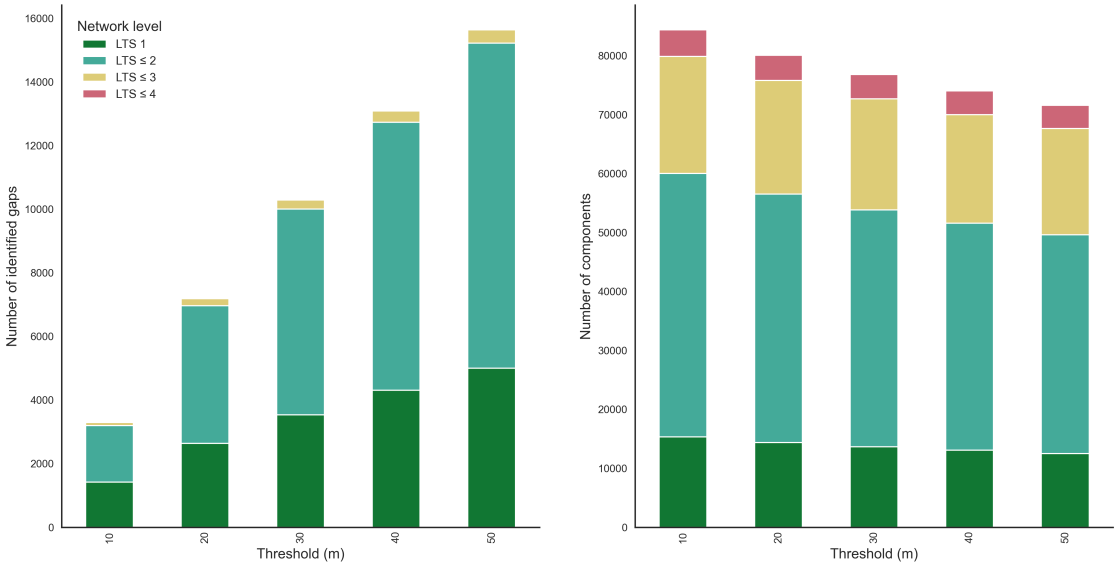

Fig. S13: Gap threshold size: Influence on closed gaps and disconnected component count. Number of identified gaps (left) and number of disconnected components (right) for each LTS level, depending on threshold size for closing gaps between network edges within the same LTS level.

References

Koukoletsos, T., Haklay, M. m., and Ellul, C. (2011). An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap.

Vierø, A. R., Vybornova, A., and Szell, M. (2024). BikeDNA: A tool for bicycle infrastructure data and network assessment. *Environment and Planning B: Urban Analytics and City Science*, 51(2):512–528.