Methodological Foundation of a Numerical Taxonomy of Urban Form: Supplementary material

Supplementary Material 1: Relational analytical framework

This research proposes and applies a relational framework of urban form for urban morphometrics.

Relational analytical framework (RF) of urban form is based on two concepts - topology and inclusiveness. The framework acknowledges that there are identifiable relations between all elements of urban form and their aggregations. As such, it accommodates all analytical aggregations into a singular framework, linking all potential measurable characters to the smallest element. Furthermore, it employs topological relations in the way it generates location-based aggregations of fundamental elements.

Unlike existing frameworks in literature, RF is analytical, not conceptual or structural. It does not try to propose a new theory of urban form; it has purely morphometric nature.

Within this research, RF is operationalised based on morphological tessellation.

The key principles of the tessellation-based relation framework are as follows.

1. Urban form is represented as building footprints, street networks and footprint-based morphological tessellation.
2. There is an identifiable relationship between buildings and street networks, buildings and street nodes and buildings and tessellation cells.
3. Morphometric characters are measured on scales defined by topological relations between elements.
    - Element itself
    - Element and its immediate neighbours
    - Element and its neighbours within n topological steps, either in a constrained or an unconstrained way.
4. Therefore, we can define subsets of RF as measurable entities of urban form based on fundamental elements and topological scales.
5. Subsets are overlapping, reusing each element within all relevant relations.

Since the relation between all elements is preserved throughout the process of their combination, we can always link values measured on one subset to another. For example, due to the fixed relation between building and street node, we can attach a node's degree value to a building as an element. The constrained topological relation can identify traditional area-based aggregations like block (as a combination of all tessellation cells which topological relation does not cross a street). As such, they allow us to combine both area-based and location-based aggregations while minimising MAUP for each of them.

Subsets of elements

Subsets are a combination of topological scales and fundamental elements. Overlap of morphometric characters derived from subsets, where each subset is representing a different structural unit, gives an overall characteristic of each duality building - cell, which can be later used for further analysis.

We can divide subsets into three topological scales: Small (or Single), Medium and Large.

Note that topological distance is possible to define within each layer (relations between buildings, relations between cells, relations between edges or nodes), but not as a combination of layers. The relation between building, its cell, its segment and its node is fixed and seen as a singular feature. That is why morphometric characters like covered area ratio of the cell are classified as a Small scale character.

Small/Single (S)
Small scale captures fundamental elements themselves (topological distance is 0 - itself). In the case of building and tessellation cell, it captures the individual character of each cell. In the case of street segment and node, it captures value for segment or node, which is then applied to each cell attached to it.

We have four subsets within small scale:

- building
- tessellation cell
- street segment
- street node

Figure S1: Diagrams illustrating the subsets on the small/single scale.

Medium (M)
The medium scale reflects topological distance 1. It captures individual character for each element derived from the relation to its adjacent elements.

- adjacent buildings
- neighbouring cells
- neighbouring segments
- linked nodes

Figure S2: Diagrams illustrating the subsets on the medium scale.

Large (L)
Large scale captures topological distance 2-n. In the case of cells, it captures individual character for each cell derived from the relation to cells within set topological distance. In the case of joined buildings and block, resulting measurable values are shared among all elements within such a structural unit. Block here is based on morphological tessellation and is defined as the contiguous portion of land comprised of cells which are normally bounded by streets or open space.

- joined buildings
- neighbouring cells of larger topological distance
- block (the maximum number of topological steps from element without the need to cross the street network)
- neighbouring segments of larger topological distance
- linked nodes of larger topological distance

Figure S3: Diagrams illustrating the subsets on the large scale.

The resulting combination of all subsets is overlapping, following, in principle, Alexander's (1966) schema of overlapping semi-lattice.

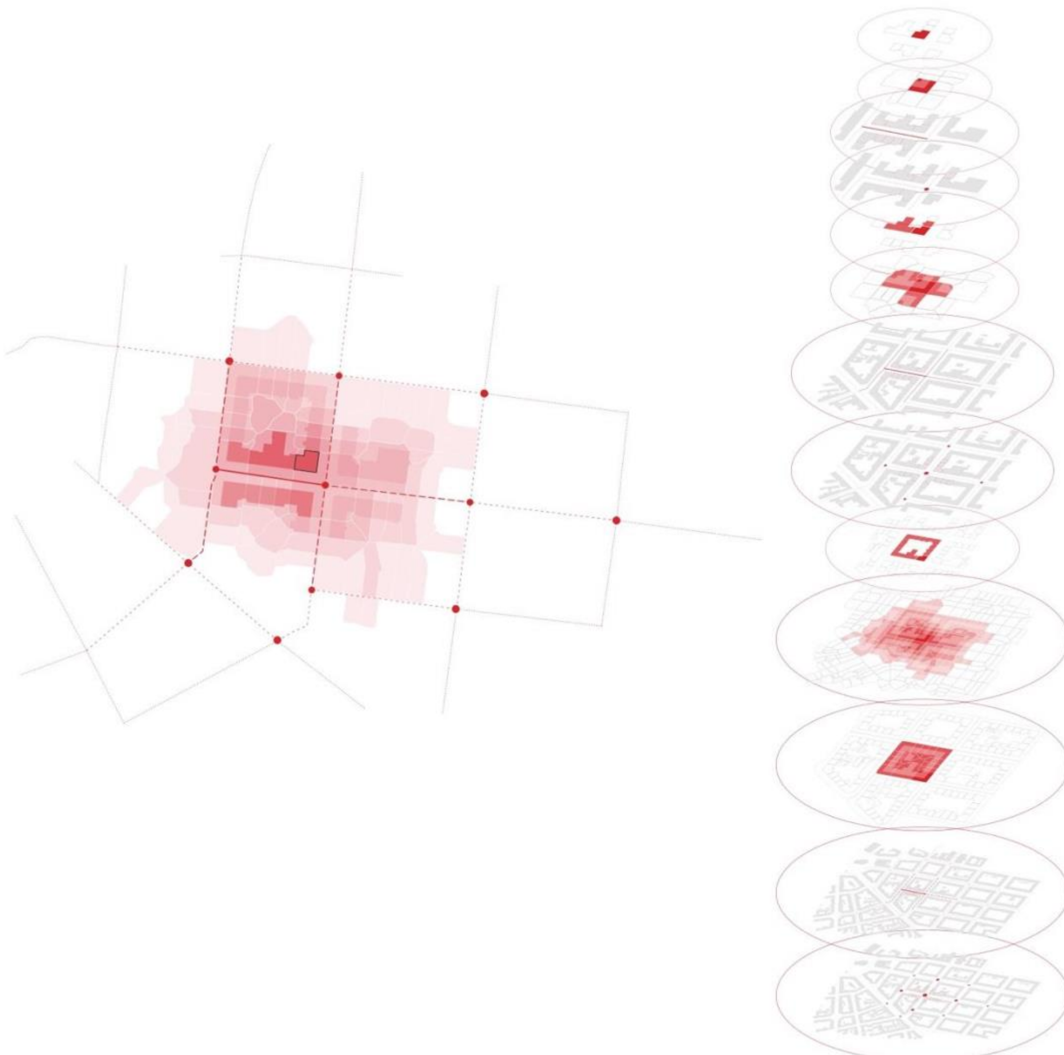

Figure S4: Diagrams illustrating the overlapping nature of the relational framework. The left diagram overlays all subsets on top of each other capturing the importance of each element for description of urban form around the indicated building. The darker the colour is, more times each element is used within various subsets. Diagram on the right shows all subsets aligned on top of each other describing the similar information while showing each subset directly.

Supplementary Material 2 Primary morphometric characters

Based on the principles described in Sneath and Sokal (1973), the following morphometric characters compose the final set of primary characters. For the implementation details, please refer to the original referred work and to the documentation and code of momepy, which contains Python-based implementation of each character.

index	element	context	category
area	building	building	dimension
height	building	building	dimension
volume	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
form factor	building	building	shape
volume to façade ratio	building	building	shape
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	shape
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension

circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
street alignment	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
floor area ratio	tessellation cell	tessellation cell	intensity
length	street segment	street segment	dimension
width	street profile	street segment	dimension
height	street profile	street segment	dimension
height to width ratio	street profile	street segment	shape
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity
height deviation	street profile	street segment	diversity
linearity	street segment	street segment	shape
area covered	street segment	street segment	dimension
buildings per meter	street segment	street segment	intensity
area covered	street node	street node	dimension
shared walls ratio	adjacent buildings	adjacent buildings	distribution
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
reached area	neighbouring segments	neighbouring segments	dimension
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
reached cells	neighbouring nodes	neighbouring nodes	intensity

reached area	neighbouring nodes	neighbouring nodes	dimension
number of courtyards	adjacent buildings	joined buildings	intensity
perimeter wall length	adjacent buildings	joined buildings	dimension
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
building adjacency	neighbouring buildings	cell queen neighbours 3	distribution
gross floor area ratio	neighbouring tessellation cells	cell queen neighbours 3	intensity
weighted reached blocks	neighbouring tessellation cells	cell queen neighbours 3	intensity
area	block	block	dimension
perimeter	block	block	dimension
circular compactness	block	block	shape
equivalent rectangular index	block	block	shape
compactness-weighted axis	block	block	shape
solar orientation	block	block	distribution
weighted neighbours	block	block	distribution
weighted cells	block	block	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
reached cells	street network	segment 3 steps	dimension
node density	street network	nodes 5 steps	intensity
reached cells	street network	nodes 3 steps	dimension
reached area	street network	nodes 3 steps	dimension
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
weighted node density	street network		intensity

local closeness centrality	street network	nodes 5 steps	connectivity
square clustering	street network	nodes within network	connectivity

Table S1: Table of primary morphometric characters. For detailed explanation, formulas and references, see the details below. Nomenclature follows the *Index of Element* model proposed by Fleischmann et al. (2020). Scale refers to the topological scale from which a character is derived, while context describes the actual set of elements used.

1. **Area of a building** is denoted as

$$a_{blg}$$

and defined as an area covered by a building footprint in m^2.

2. **Height of a building** is denoted as

$$h_{blg}$$

and defined as building height in m measured optimally as weighted mean height (in case of buildings with multiple parts of different height). It is a required input value not measured within the morphometric assessment itself.

3. **Volume of a building** is denoted as

$$v_{blg} = a_{blg} \times h_{blg}$$

and defined as building footprint multiplied by its height in m^3.

4. **Perimeter of a building** is denoted as

$$p_{blg}$$

and defined as the sum of lengths of the building exterior walls in m.

5. **Courtyard area of a building** is denoted as

$$a_{blg_c}$$

and defined as the sum of areas of interior holes in footprint polygons in m^2.

6. **Form factor of a building** is denoted as

$$FoF_{blg} = \frac{a_{blg}}{v_{blg}^{\frac{2}{3}}}.$$

It captures three-dimensional unitless shape characteristic of a building envelope unbiased by the building size (Bourdic et al., 2012).

7. **Volume to façade ratio of a building** is denoted as

$$VFR_{blg} = \frac{v_{blg}}{p_{blg} \times h_{blg}}.$$

It captures the aspect of the three-dimensional shape of a building envelope able to distinguish building types, as shown by Schirmer and Axhausen (2015). It can be seen as a proxy of volumetric compactness.

8. **Circular compactness of a building** is denoted as

$$CCo_{blg} = \frac{a_{blg}}{a_{blgC}}$$

where a_{blgC} is an area of minimal enclosing circle. It captures the relation of building footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle (Dibble et al., 2019).

9. **Corners of a building** is denoted as

$$Cor_{blg} = \sum_{i=1}^{n} c_{blg}$$

where c_{blg} is defined as a vertex of building exterior shape with an angle between adjacent line segments ≤ 170 degrees. It uses only external shape, courtyards are not included. Character is adapted from Steiniger et al. (2008) to exclude non-corner-like vertices.

10. **Squareness of a building** is denoted as

$$Squ_{blg} = \frac{\sum_{i=1}^{n} D_{c_{blg_i}}}{n}$$

where D is the deviation of angle of corner c_{blg_i} from 90 degrees and n is a number of corners.

11. **Equivalent rectangular index of a building** is denoted as

$$ERI_{blg} = \sqrt{\frac{a_{blg}}{a_{blgB}}} * \frac{p_{blgB}}{p_{blg}}$$

where a_{blgB} is an area of a minimal rotated bounding rectangle of a building (MBR) footprint and p_{blgB} its perimeter of MBR. It is a measure of shape complexity identified by Basaraner and Cetinkaya (2017) as the shape characters with the best performance.

12. **Elongation of a building** is denoted as

$$Elo_{blg} = \frac{l_{blgB}}{w_{blgB}}$$

where l_{blgB} is length of MBR and w_{blgB} is width of MBR. It captures the ratio of shorter to the longer dimension of MBR to indirectly capture the deviation of the shape from a square (Schirmer and Axhausen, 2015).

13. **Centroid - corner distance deviation of a building** is denoted as

$$CCD_{blg} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (ccd_i - \overline{ccd})^2}$$

where ccd_i is a distance between centroid and corner i and $\overline{ccd}$ is mean of all distances. It captures a variety of shape. As a corner is considered vertex with angle < 170º to reflect potential circularity of object and topological imprecision of building polygon.

14. **Centroid - corner mean distance of a building** is denoted as

$$CCM_{blg} = \frac{1}{n}\left(\sum_{i=1}^{n} ccd_i\right)$$

where ccd_i is a distance between centroid and corner i. It is a character measuring a dimension of the object dependent on its shape (Schirmer and Axhausen, 2015).

15. **Solar orientation of a building** is denoted as

$$Ori_{blg} = |o_{blgB} - 45|$$

where o_{blgB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions. There are multiple ways of capturing orientation of a polygon. As reported by Yan et al. (2007), Duchêne et al. (2003) assessed five different options (longest edge, weighted bisector, wall average, statistical weighting, bounding rectangle) and concluded a bounding rectangle as the most appropriate. Deviation from cardinal directions is used to avoid sudden changes between square-like objects.

16. **Street alignment of a building** is denoted as

$$SAl_{blg} = |Ori_{blg} - Ori_{edg}|$$

where Ori_{blg} is a solar orientation of the building and Ori_{edg} is a solar orientation of the street edge. It reflects the relationship between the building and its street, whether it is facing the street directly or indirectly (Schirmer and Axhausen, 2015).

17. **Cell alignment of a building** is denoted as

$$CAl_{blg} = |Ori_{blg} - Ori_{cell}|$$

where Ori_{cell} is a solar orientation of tessellation cell. It reflects the relationship between a building and its cell.

18. **Longest axis length of a tessellation cell** is denoted as

$$LAL_{cell} = d_{cellC}$$

where d_{cellC} is a diameter of the minimal circumscribed circle around the tessellation cell polygon. The axis itself does not have to be fully within the polygon. It could be seen as a proxy of plot depth for tessellation-based analysis.

19. **Area of a tessellation cell** is denoted as

$$a_{cell}$$

and defined as an area covered by a tessellation cell footprint in m^2.

20. **Circular compactness of a tessellation cell** is denoted as

$$CCo_{cell} = \frac{a_{cell}}{a_{cellC}}$$

where a_{cellC} is an area of minimal enclosing circle. It captures the relation of tessellation cell footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle.

21. **Equivalent rectangular index of a tessellation cell** is denoted as

$$ERI_{cell} = \sqrt{\frac{a_{cell}}{a_{cellB}}} * \frac{p_{cellB}}{p_{cell}}$$

where a_{cellB} is an area of the minimal rotated bounding rectangle of a tessellation cell (MBR) footprint and p_{cellB} its perimeter of MBR.

22. **Solar orientation of a tessellation cell** is denoted as

$$Ori_{cell} = |o_{cellB} - 45|$$

where o_{cellB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions.

23. **Street alignment of a building** is denoted as

$$SAl_{cell} = |Ori_{cell} - Ori_{edg}|$$

where Ori_{cell} is a solar orientation of tessellation cell and Ori_{edg} is a solar orientation of the street edge. It reflects the relationship between tessellation cell and its street, whether it is facing the street directly or indirectly.

24. **Coverage area ratio of a tessellation cell** is denoted as

$$CAR_{cell} = \frac{a_{blg}}{a_{cell}}$$

where a_{blg} is an area of a building and a_{cell} is an area of related tessellation cell (Schirmer and Axhausen, 2015). Coverage area ratio (CAR) is one of the commonly used characters capturing *intensity* of development. However, the definitions vary based on the spatial unit.

25. **Floor area ratio of a tessellation cell** is denoted as

$$FAR_{cell} = \frac{fa_{blg}}{a_{cell}}$$

where fa_{blg} is a floor area of a building and a_{cell} is an area of related tessellation cell. Floor area could be computed based on the number of levels or using an approximation based on building height.

26. **Length of a street segment** is denoted as

$$l_{edg}$$

and defined as a length of a *LineString* geometry in metres.

27. **Width of a street profile** is denoted as

$$w_{sp} = \frac{1}{n}\left(\sum_{i=1}^{n} w_i\right)$$

where w_i is width of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value. In the case of the open-ended street, 50 metres is used as a perception-based proximity limit (Araldi and Fusco, 2019).

28. **Height of a street profile** is denoted as

$$h_{sp} = \frac{1}{n}\left(\sum_{i=1}^{n} h_i\right)$$

where h_I is mean height of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value (Araldi and Fusco, 2019).

29. **Height to width ratio of a street profile** is denoted as

$$HWR_{sp} = \frac{1}{n}\left(\sum_{i=1}^{n} \frac{h_i}{w_i}\right)$$

where h_I is mean height of a street section i and w_i is the width of a street section i. The algorithm generates street sections every 3 meters alongside the street segment, and measures mean value (Araldi and Fusco, 2019).

30. **Openness of a street profile** is denoted as

$$Ope_{sp} = 1 - \frac{\sum hit}{2\sum sec}$$

where $\sum hit$ is a sum of section lines (left and right sides separately) intersecting buildings and $\sum sec$ total number of street sections. The algorithm generates street sections every 3 meters alongside the street segment.

31. **Width deviation of a street profile** is denoted as

$$wDev_{sp} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(w_i - w_{sp}\right)^2}$$

where w_i is width of a street section i and w_{sp} is mean width. The algorithm generates street sections every 3 meters alongside the street segment.

32. **Height deviation of a street profile** is denoted as

$$hDev_{sp} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(h_i - h_{sp}\right)^2}$$

where h_i is height of a street section i and h_{sp} is mean height. The algorithm generates street sections every 3 meters alongside the street segment.

33. **Linearity of a street segment** is denoted as

$$Lin_{edg} = \frac{l_{eucl}}{l_{edg}}$$

where l_{eucl} is Euclidean distance between endpoints of a street segment and l_{edg} is a street segment length. It captures the deviation of a segment shape from a straight line. It is adapted from (Araldi and Fusco, 2019).

34. **Area covered by a street segment** is denoted as

$$a_{edg} = \sum_{i=1}^{n} a_{cell_i}$$

where a_{cell_i} is an area of tessellation cell i belonging to the street segment. It captures the area which is likely served by each segment.

35. **Buildings per meter of a street segment** is denoted as

$$BpM_{edg} = \frac{\sum blg}{l_{edg}}$$

where $\sum blg$ is a number of buildings belonging to a street segment and l_{edg} is a length of a street segment. It reflects the granularity of development along each segment.

36. **Area covered by a street node** is denoted as

$$a_{node} = \sum_{i=1}^{n} a_{cell_i}$$

where a_{cell_i} is an area of tessellation cell i belonging to the street node. It captures the area which is likely served by each node.

37. **Shared walls ratio of adjacent buildings** is denoted as

$$SWR_{blg} = \frac{p_{blg_{shared}}}{p_{blg}}$$

where $p_{blg_{shared}}$ is a length of a perimeter shared with adjacent buildings and p_{blg} is a perimeter of a building. It captures the amount of wall space facing the open space (Hamaina et al., 2012).

38. **Alignment of neighbouring buildings** is denoted as

$$Ali_{blg} = \frac{1}{n}\sum_{i=1}^{n} \left|Ori_{blg} - Ori_{blg_i}\right|$$

where Ori_{blg} is the solar orientation of a building and Ori_{blg_i} is the solar orientation of building i on a neighbouring tessellation cell. It calculates the mean deviation of solar orientation of buildings on adjacent cells from a building. It is adapted from Hijazi et al. (2016).

39. **Mean distance to neighbouring buildings** is denoted as

$$NDi_{blg} = \frac{1}{n}\sum_{i=1}^{n} d_{blg,blg_i}$$

where d_{blg,blg_i} is a distance between building and building i on a neighbouring tessellation cell. It is adapted from Hijazi et al. (2016). It captures the average proximity to other buildings.

40. **Weighted neighbours of a tessellation cell** is denoted as

$$WNe_{cell} = \frac{\sum cell_n}{p_{cell}}$$

where $\sum cell_n$ is a number of cell neighbours and p_{cell} is a perimeter of a cell. It reflects granularity of morphological tessellation.

41. **Area covered by neighbouring cells** is denoted as

$$a_{cell_n} = \sum_{i=1}^{n} a_{cell_i}$$

where a_{cell_i} is area of tessellation cell i within topological distance 1. It captures the scale of morphological tessellation.

42. **Reached cells by neighbouring segments** is denoted as

$$RC_{edg_n} = \sum_{i=1}^{n} cells_{edg_i}$$

where $cells_{edg_i}$ is number of tessellation cells on segment i within topological distance 1. It captures accessible granularity.

43. **Reached area by neighbouring segments** is denoted as

$$a_{edg_n} = \sum_{i=1}^{n} a_{edg_i}$$

where a_{edg_i} is an area covered by a street segment i within topological distance 1. It captures an accessible area.

44. **Degree of a street node** is denoted as

$$deg_{node_i} = \sum_{j} \quad edg_{ij}$$

where edg_{ij} is an edge of a street network between node i and node j. It reflects the basic degree centrality.

45. **Mean distance to neighbouring nodes from a street node** is denoted as

$$MDi_{node} = \frac{1}{n}\sum_{i=1}^{n} \quad d_{node,node_i}$$

where $d_{node,node_i}$ is a distance between node and node i within topological distance 1. It captures the average proximity to other nodes.

46. **Reached cells by neighbouring nodes** is denoted as

$$RC_{node_n} = \sum_{i=1}^{n} \quad cells_{node_i}$$

where $cells_{node_i}$ is number of tessellation cells on node i within topological distance 1. It captures accessible granularity.

47. **Reached area by neighbouring nodes** is denoted as

$$a_{node_n} = \sum_{i=1}^{n} \quad a_{node_i}$$

where a_{node_i} is an area covered by a street node i within topological distance 1. It captures an accessible area.

48. **Number of courtyards of adjacent buildings** is denoted as

$$NCo_{blg_{adj}}$$

where $NCo_{blg_{adj}}$ is a number of interior rings of a polygon composed of footprints of adjacent buildings (Schirmer and Axhausen, 2015).

49. **Perimeter wall length of adjacent buildings** is denoted as

$$p_{blg_{adj}}$$

where $p_{blg_{adj}}$ is a length of an exterior ring of a polygon composed of footprints of adjacent buildings.

50. **Mean inter-building distance between neighbouring buildings** is denoted as

$$IBD_{blg} = \frac{1}{n}\sum_{i=1}^{n} \quad d_{blg,blg_i}$$

where d_{blg,blg_i} is a distance between building and building i on a tessellation cell within topological distance 3. It is adapted from Caruso et al. (2017). It captures the average proximity between buildings.

51. **Building adjacency of neighbouring buildings** is denoted as

$$BuA_{blg} = \frac{\sum blg_{adj}}{\sum blg}$$

where $\sum blg_{adj}$ is a number of joined built-up structures within topological distance three and $\sum blg$ is a number of buildings within topological distance 3. It is adapted from Vanderhaegen and Canters (2017).

52. **Gross floor area ratio of neighbouring tessellation cells** is denoted as

$$GFAR_{cell} = \frac{\sum_{i=1}^{n} \quad FAR_{cell_i}}{\sum_{i=1}^{n} \quad a_{cell_i}}$$

where FAR_{cell_i} is a floor area ratio of tessellation cell i and a_{cell_i} is an area of tessellation cell i within topological distance 3. Based on Dibble et al. (2019).

53. **Weighted reached blocks of neighbouring tessellation cells** is denoted as

$$WRB_{cell} = \frac{\sum blk}{\sum_{i=1}^{n} \quad a_{cell_i}}$$

where $\sum blk$ is a number of blocks within topological distance three and a_{cell_i} is an area of tessellation cell i within topological distance three.

54. **Area of a block** is denoted as

$$a_{blk}$$

and defined as an area covered by a block footprint in m^2.

55. **Perimeter of a block** is denoted as

$$p_{blk}$$

and defined as lengths of the block polygon exterior in m.

56. **Circular compactness of a block** is denoted as

$$CCo_{blk} = \frac{a_{blk}}{a_{blkC}}$$

where a_{blkC} is an area of minimal enclosing circle. It captures the relation of block footprint shape to its minimal enclosing circle, illustrating the similarity of shape and circle.

57. **Equivalent rectangular index of a block** is denoted as

$$ERI_{blk} = \sqrt{\frac{a_{blk}}{a_{blkB}}} * \frac{p_{blkB}}{p_{blk}}$$

where a_{blkB} is an area of the minimal rotated bounding rectangle of a block (MBR) footprint and p_{blkB} its perimeter of MBR.

58. **Compactness-weighted axis of a block** is denoted as

$$CWA_{blk} = d_{blkC} \times \left(\frac{4}{\pi} - \frac{16(a_{blk})}{p_{blk}^2}\right)$$

where d_{blkC} is a diameter of the minimal circumscribed circle around the block polygon, a_{blk} is an area of a block and p_{blk} is a perimeter of a block. It is a proxy of permeability of an area (Feliciotti, 2018).

59. **Solar orientation of a block** is denoted as

$$Ori_{blk} = |o_{blkB} - 45|$$

where o_{blkB} is an orientation of the longest axis of bounding rectangle in a range 0 - 45. It captures the deviation of orientation from cardinal directions.

60. **Weighted neighbours of a block** is denoted as

$$wN_{blk} = \frac{\sum blk_n}{p_{blk}}$$

where $\sum blk_n$ is a number of block neighbours and p_{blk} is a perimeter of a block. It reflects granularity of a mesh of blocks.

61. **Weighted cells of a block** is denoted as

$$wC_{blk} = \frac{\sum cell}{a_{blk}}$$

where $\sum cell$ is a number of cells composing a block and a_{blk} is an area of a block. It captures the granularity of each block.

62. **Local meshedness of a street network** is denoted as

$$Mes_{node} = \frac{e - v + 1}{2v - 5}$$

where e is a number of edges in a subgraph, and v is the number of nodes in a subgraph (Feliciotti, 2018). A subgraph is defined as a network within topological distance five around a node.

63. **Mean segment length of a street network** is denoted as

$$MSL_{edg} = \frac{1}{n}\sum_{i=1}^{n} \quad l_{edg_i}$$

where l_{edg_i} is a length of a street segment i within a topological distance 3 around a segment.

64. **Cul-de-sac length of a street network** is denoted as

$$CDL_{node} = \sum_{i=1}^{n} l_{edg_i}, if\ edg_i\ is\ cul - de - sac$$

where l_{edg_i} is a length of a street segment i within a topological distance 3 around a node.

65. **Reached cells by street network segments** is denoted as

$$RC_{edg} = \sum_{i=1}^{n} cells_{edg_i}$$

where $cells_{edg_i}$ is number of tessellation cells on segment i within topological distance 3. It captures accessible granularity.

66. **Node density of a street network** is denoted as

$$D_{node} = \frac{\sum node}{\sum_{i=1}^{n} l_{edg_i}}$$

where $\sum node$ is a number of nodes within a subgraph and l_{edg_i} is a length of a segment i within a subgraph. A subgraph is defined as a network within topological distance five around a node.

67. **Reached cells by street network nodes** is denoted as

$$RC_{node_{net}} = \sum_{i=1}^{n} cells_{node_i}$$

where $cells_{node_i}$ is number of tessellation cells on node i within topological distance 3. It captures accessible granularity.

68. **Reached area by street network nodes** is denoted as

$$a_{node_{net}} = \sum_{i=1}^{n} a_{node_i}$$

where a_{node_i} is an area covered by a street node i within topological distance 3. It captures an accessible area.

69. **Proportion of cul-de-sacs within a street network** is denoted as

$$pCD_{node} = \frac{\sum_{i=1}^{n} node_i, if\ deg_{node_i} = 1}{\sum_{i=1}^{n} node_i}$$

where $node_i$ is a node whiting topological distance five around a node. Adapted from Boeing (2017).

70. **Proportion of 3-way intersections within a street network** is denoted as

$$p3W_{node} = \frac{\sum_{i=1}^{n} node_i, if\ deg_{node_i} = 3}{\sum_{i=1}^{n} node_i}$$

where $node_i$ is a node whiting topological distance five around a node. Adapted from Boeing (2017).

71. **Proportion of 4-way intersections within a street network** is denoted as

$$p4W_{node} = \frac{\sum_{i=1}^{n} \quad node_i, if\ deg_{node_i} = 4}{\sum_{i=1}^{n} \quad node_i}$$

where $node_i$ is a node whiting topological distance five around a node. Adapted from Boeing (2017).

72. **Weighted node density of a street network** is denoted as

$$wD_{node} = \frac{\sum_{i=1}^{n} \quad deg_{node_i} - 1}{\sum_{i=1}^{n} \quad l_{edg_i}}$$

where deg_{node_i} is a degree of a node i within a subgraph and l_{edg_i} is a length of a segment i within a subgraph. A subgraph is defined as a network within topological distance five around a node.

73. **Local closeness centrality of a street network** is denoted as

$$lCC_{node} = \frac{n-1}{\sum_{v=1}^{n-1} \quad d(v,u)}$$

where $d(v,u)$ is the shortest-path distance between v and u, and n is the number of nodes within a subgraph. A subgraph is defined as a network within topological distance five around a node.

74. **Square clustering of a street network** is denoted as

$$sCl_{node} = \frac{\sum_{u=1}^{k_v} \quad \sum_{w=u+1}^{k_v} \quad q_v(u,w)}{\sum_{u=1}^{k_v} \quad \sum_{w=u+1}^{k_v} \quad [a_v(u,w) + q_v(u,w)]}$$

where $q_v(u,w)$ are the number of common neighbours of u and w other than v (ie squares), and $a_v(u,w) = (k_u - (1 + q_v(u,w) + \theta_{uv}))(k_w - (1 + q_v(u,w) + \theta_{uw}))$, where $\theta_{uw} = 1$ if u and w are connected and 0 otherwise (Lind et al., 2005).

Table below contains each character and its classification to scale following Fleischmann et al. (2020) and key used in additional figures across supplementary materials.

index	**element**	**grain**	**extent**	**id**
area	building	S	S	sdbAre
height	building	S	S	sdbHei
volume	building	S	S	sdbVol
perimeter	building	S	S	sdbPer
courtyard area	building	S	S	sdbCoA
form factor	building	S	S	ssbFoF
volume to façade ratio	building	S	S	ssbVFR

circular compactness	building	S	S	ssbCCo
corners	building	S	S	ssbCor
squareness	building	S	S	ssbSqu
equivalent rectangular index	building	S	S	ssbERI
elongation	building	S	S	ssbElo
centroid - corner distance deviation	building	S	S	ssbCCD
centroid - corner mean distance	building	S	S	ssbCCM
solar orientation	building	S	S	stbOri
street alignment	building	S	S	stbSAl
cell alignment	building	S	S	stbCeA
longest axis length	tessellation cell	S	S	sdcLAL
area	tessellation cell	S	S	sdcAre
circular compactness	tessellation cell	S	S	sscCCo
equivalent rectangular index	tessellation cell	S	S	sscERI
solar orientation	tessellation cell	S	S	stcOri
street alignment	tessellation cell	S	S	stcSAl
coverage area ratio	tessellation cell	S	S	sicCAR
floor area ratio	tessellation cell	S	S	sicFAR
length	street segment	S	S	sdsLen
width	street profile	S	S	sdsSPW
height	street profile	S	S	sdsSPH
height to width ratio	street profile	S	S	sdsSPR
openness	street profile	S	S	sdsSPO
width deviation	street profile	S	S	sdsSWD
height deviation	street profile	S	S	sdsSHD
linearity	street segment	S	S	sssLin
area covered	street segment	S	S	sdsAre
buildings per meter	street segment	S	S	sisBpM
area covered	street node	S	S	sddAre
shared walls ratio	adjacent buildings	S	S	mtbSWR
alignment	neighbouring buildings	S	S	mtbAli
mean distance	neighbouring buildings	S	S	mtbNDi
weighted neighbours	tessellation cell	S	S	mtcWNe
area covered	neighbouring cells	S	S	mdcAre
reached cells	neighbouring segments	S	S	misRea

reached area	neighbouring segments	S	S	mdsAre
degree	street node	S	S	mtdDeg
mean distance to neighbouring nodes	street node	S	S	mtdMDi
reached cells	neighbouring nodes	S	S	midRea
reached area	neighbouring nodes	S	S	midAre
number of courtyards	adjacent buildings	S	S	libNCo
perimeter wall length	adjacent buildings	S	S	ldbPWL
mean inter-building distance	neighbouring buildings	S	S	ltbIBD
building adjacency	neighbouring buildings	S	S	ltcBuA
gross floor area ratio	neighbouring tessellation cells	S	S	licGDe
weighted reached blocks	neighbouring tessellation cells	S	S	ltcWRB
area	block	S	S	ldkAre
perimeter	block	S	S	ldkPer
circular compactness	block	S	S	lskCCo
equivalent rectangular index	block	S	S	lskERI
compactness-weighted axis	block	S	S	lskCWA
solar orientation	block	S	S	ltkOri
weighted neighbours	block	S	S	ltkWNB
weighted cells	block	S	S	likWBB
local meshedness	street network	S	M	lcdMes
mean segment length	street network	S	S	ldsMSL
cul-de-sac length	street network	S	S	ldsCDL
reached cells	street network	S	S	ldsRea
node density	street network	S	M	lddNDe
reached cells	street network	S	S	lddRea
reached area	street network	S	S	lddARe
proportion of cul-de-sacs	street network	S	M	linPDE
proportion of 3-way intersections	street network	S	M	linP3W
proportion of 4-way intersections	street network	S	M	linP4W
weighted node density	street network	S	M	linWID
local closeness centrality	street network	S	M	lcnClo

square clustering	street network	S	L	xcnSCl

Table S2: Additional classification of primary morphometric characters.

Supplementary Material 3: Bayesian Information Criterion

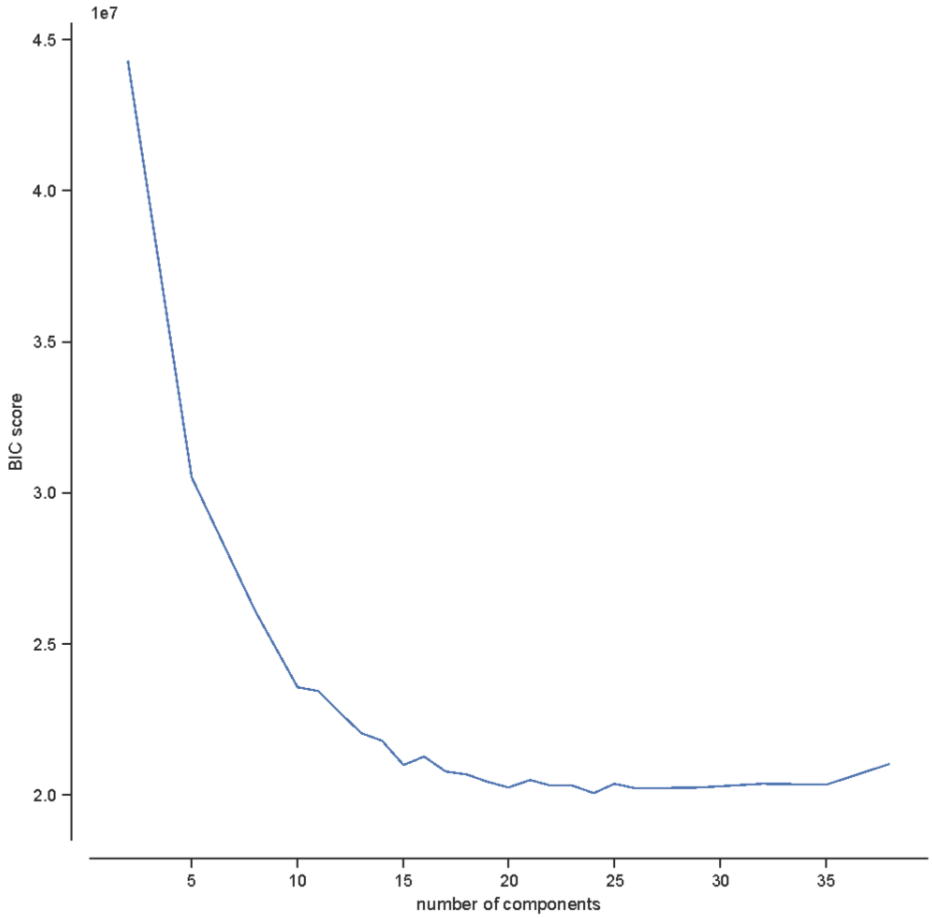

Figure S5: Bayesian Information Criterion score for the variable number of components in Prague case study. Shaded area reflects .95 confidence interval.

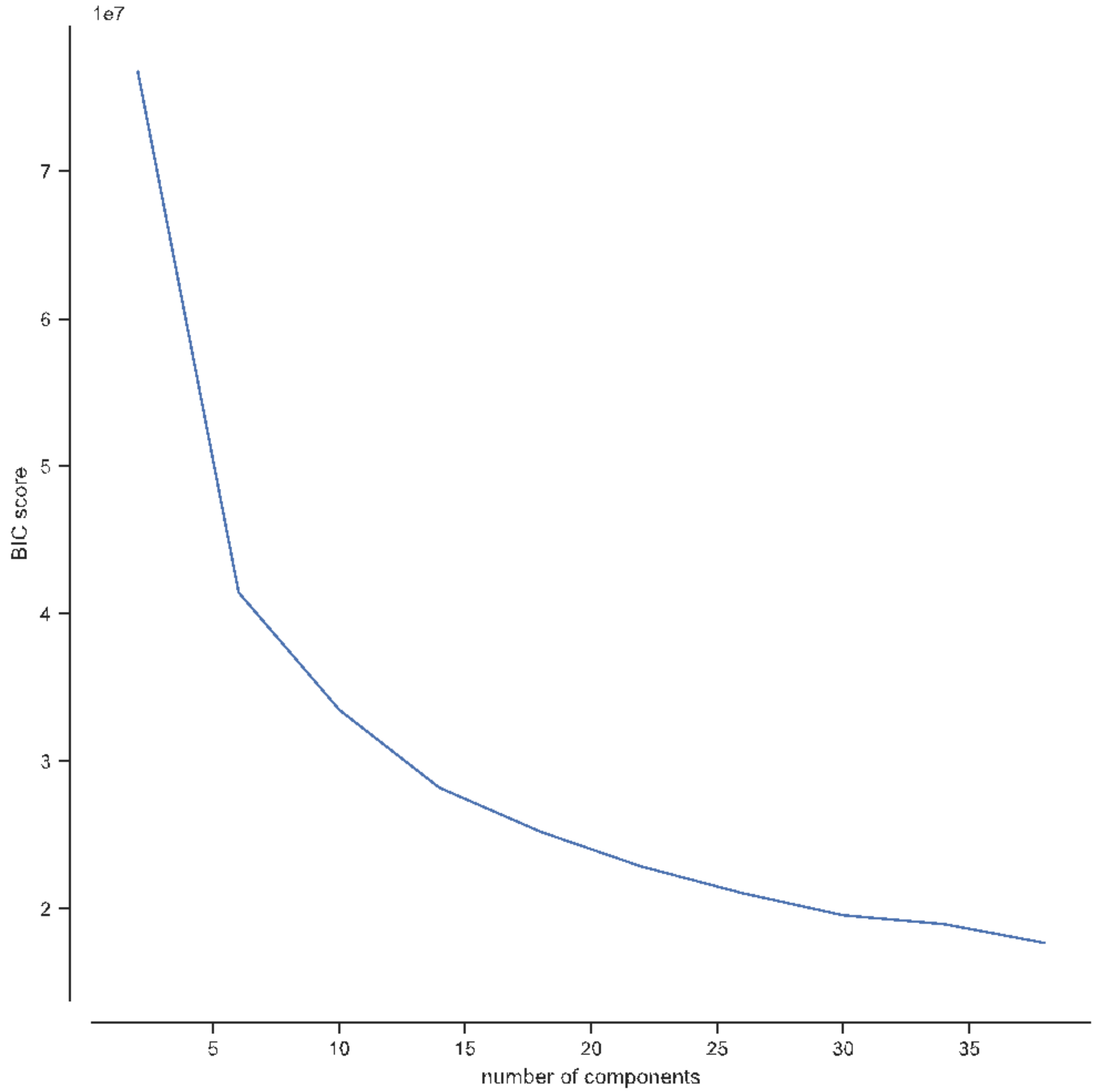

Figure S6: Bayesian Information Criterion score for the variable number of components in Amsterdam case study. Shaded area reflects .95 confidence interval, red line marks the first significant minimum.

Supplementary material 4: Full extent of presented maps illustrating spatial distribution of results of cluster analysis.

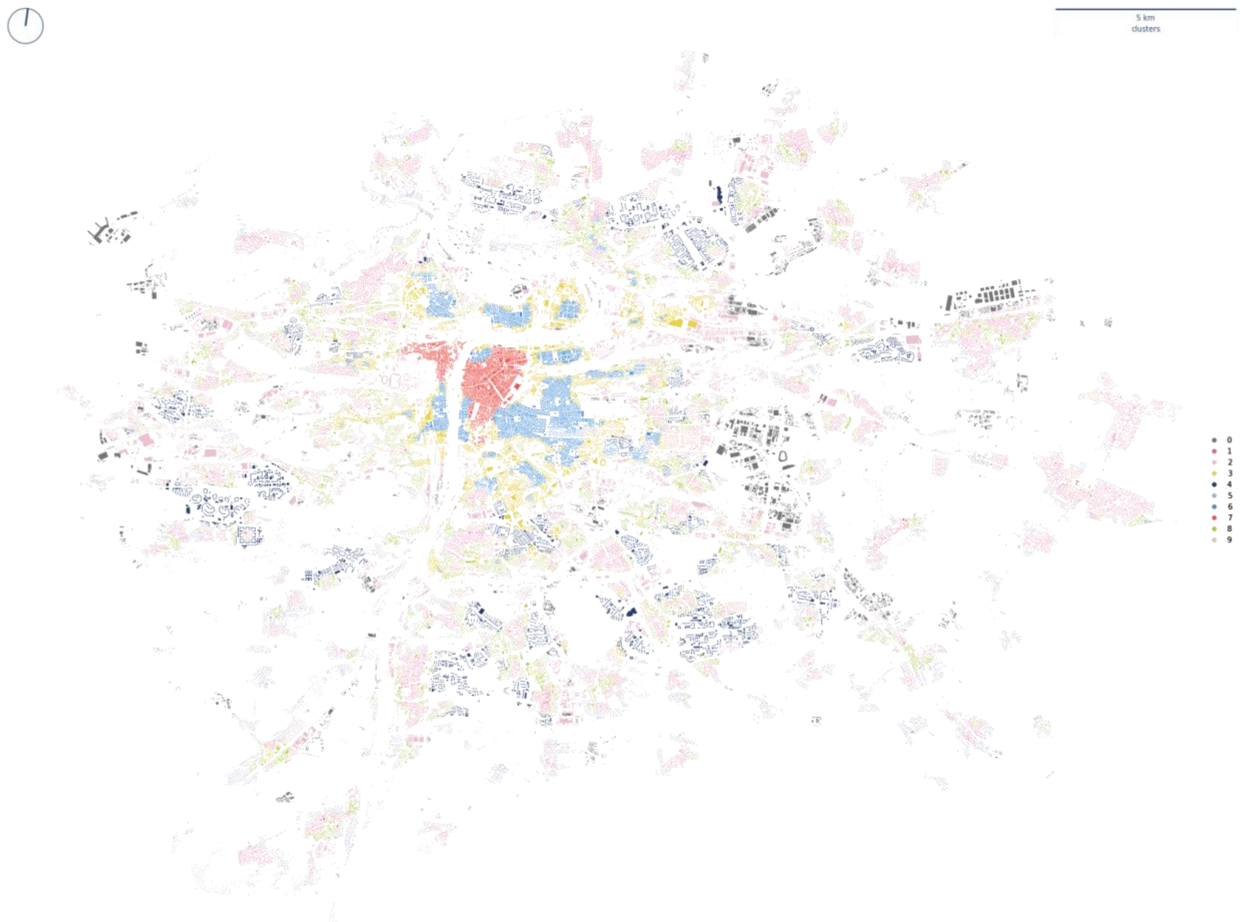

Figure S7: Spatial distribution of 10 detected clusters in Prague.

Figure S8: Spatial distribution of 10 detected clusters in Amsterdam.

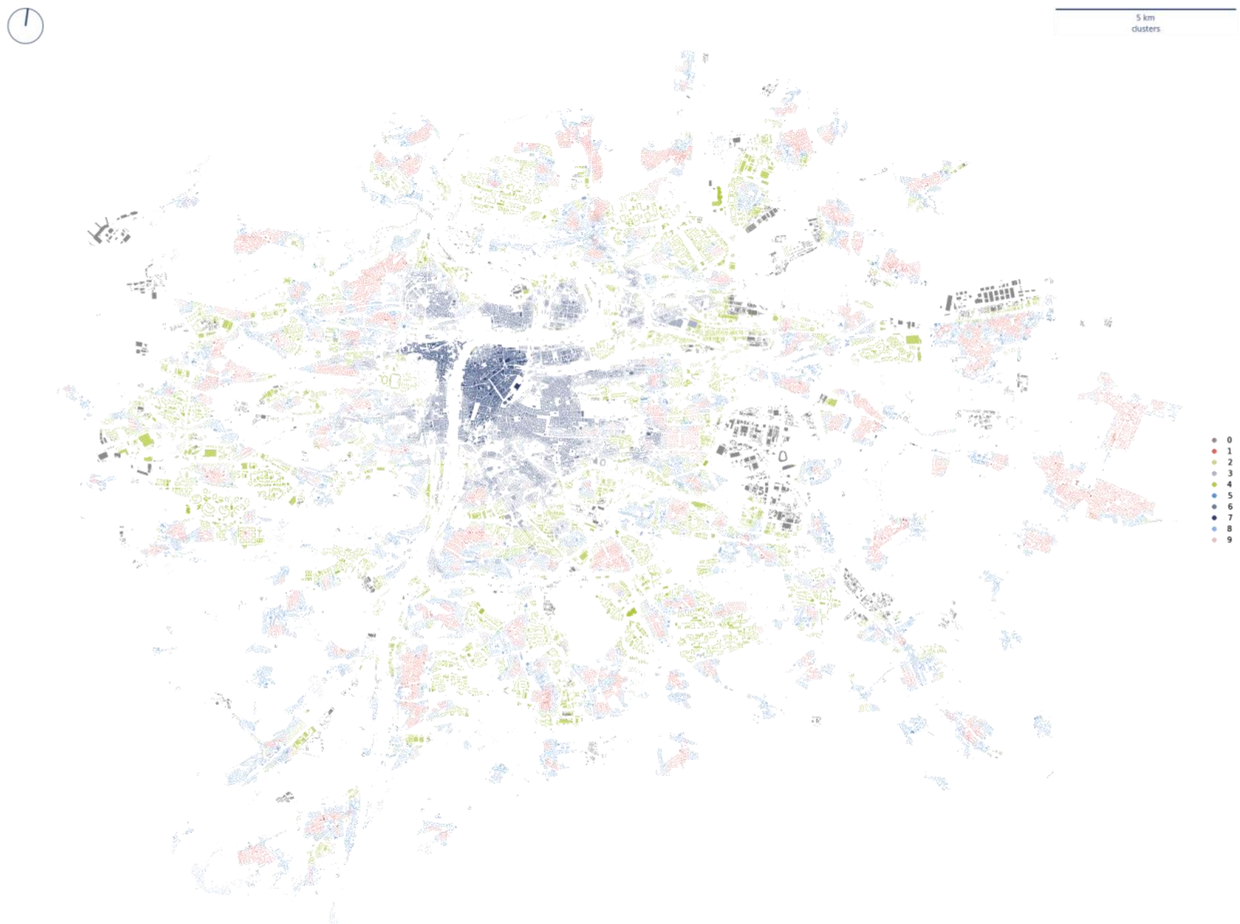

Figure S9: Spatial distribution of different branches of the combined dendrogram in Prague.

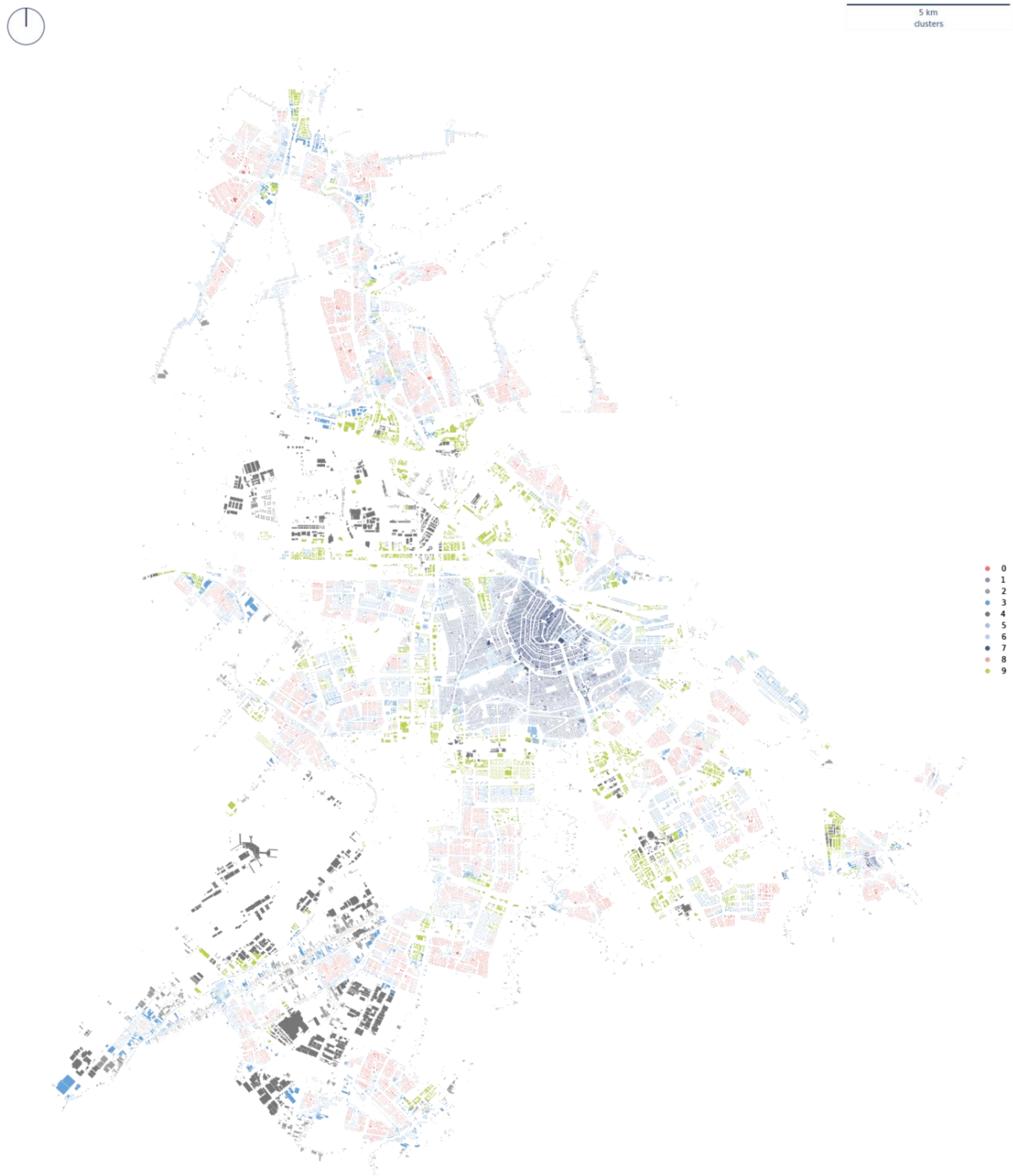

Figure S10: Spatial distribution of different branches of the combined dendrogram in Amsterdam.

Supplementary Material 5: Contingency tables

,

cluster	1840	1880	1920	1950	1970	1990	2012
0	349	85	263	1219	1506	1442	565
1	1138	513	3588	17095	4499	1265	1453
2	1407	621	1655	4537	3108	2530	1357
3	1392	1719	2658	2895	678	223	213
4	145	54	156	888	1993	6414	532
5	3442	568	1487	7677	3975	2084	3459
6	1413	2778	4109	2005	150	4	8
7	3177	110	73	49	0	0	1
8	2834	981	2661	9645	4259	2629	829
9	69	63	151	3764	1147	1573	1244

Table S3: Contingency table showing the counts of features per historical origin within individual clusters in Prague case study.

cluster	Multi-family housing	Single-family housing	Villas	Industry small	Industry large	other
0	112	617	3	322	1138	3497
1	437	27953	1164	3	0	33
2	3706	7238	203	972	789	2830
3	8472	577	136	93	26	626
4	9553	748	0	0	0	17
5	75	21590	147	50	22	1156
6	10070	231	153	0	0	34
7	2374	6	0	0	0	1057
8	4296	18110	1080	117	60	340
9	868	7015	79	0	0	120

Table S4: Contingency table showing the counts of features per predominant land use within individual clusters in Prague case study.

cluster	organic	perimeter block	village	garden city	modernism	production	services
0	0	17	377	213	39	3216	352
1	0	3	11384	16150	100	1	0
2	8	453	2937	2859	1394	2383	1085
3	192	6516	100	725	248	234	197
4	0	54	192	324	8782	17	49

5	0	0	13298	7824	40	33	61
6	604	8522	8	575	6	0	0
7	3281	49	0	0	0	3	78
8	0	263	6614	9900	2189	98	78
9	0	0	880	3176	1112	0	62

Table S5: Contingency table showing the counts of features per expert typology classes within individual clusters in Prague case study.

cluster	1800	1850	1900	1930	1945	1960	1975	1985	1995	2005	2020
0	2	6	25	653	757	5541	11488	10448	10153	3362	3327
1	314	0	5201	17479	5118	325	60	395	743	241	110
2	65	42	360	1794	914	1409	1949	1258	1280	1597	1230
3	59	27	303	2133	1072	1244	2189	1512	1906	1990	1452
4	2	0	62	32	27	81	267	288	420	477	361
5	927	24	2000	5825	2824	6583	3236	2564	3854	3662	3393
6	111	45	713	5116	2366	4643	8811	4463	5696	4171	3089
7	7153	98	1531	1828	692	145	213	362	722	386	125
8	31	24	371	7976	6716	11113	5369	1948	7652	2948	3739
9	127	25	359	658	322	1153	2453	1478	2082	2122	1698

Table S6: Contingency table showing the counts of features per historical origin within individual clusters in Amsterdam case study.

Case study	Data	Degrees of Freedom	N	χ^2	p-value	Cramér's V
Prague	Historical origin	54	140315	91599	< .001	0.331
Prague	Land use	45	140315	153672	< .001	0.468
Prague	Qualitative classification	54	119413	325351	< .001	0.674
Amsterdam	Historical origin	90	252385	218457	< .001	0.311

Table S7: Reported Chi-square and Cramér's V results for each tested dataset. All results indicate significant relationship as per Chi-square statistics and moderate to high association as per Cramér's V. V < .3 indicates low, .3 - .5 moderate, and > .5 high association.

Data and Code

The reproducible Python code is available in the form of Jupyter notebooks at <anonymised>.

The work is accompanied by an open-source Python package (available at <anonymised>) .

The morphological data (buildings, streets) for Prague case study were obtained from the city's open data portal (https://www.geoportalpraha.cz/en), while the validation layers were provided by Prague Institute of Planning and Development. The morphological data for Amsterdam are obtained from 3D BAG repository (Dukai, 2020) and Basisregistratie Grootschalige Topografie, BGT (http://data.nlextract.nl/)

- Dukai, B. (2020) '3D Registration of Buildings and Addresses (BAG) / 3D Basisregistratie Adressen en Gebouwen (BAG)', 4TU.ResearchData. doi: 10.4121/uuid:f1f9759d-024a-492a-b821-07014dd6131c.