Applicability of morphological tessellation and its topological derivatives in the quantitative analysis of urban form

Martin Fleischmann
University of Strathclyde
martin.fleischmann@strath.ac.uk

Ombretta Romice
University of Strathclyde
ombretta.r.romice@strath.ac.uk

Sergio Porta
University of Strathclyde
sergio.porta@strath.ac.uk

Keywords: morphological tessellation, plot, Voronoi diagram, contiguity, topology
Conference topic: Methods 1: embedding different approaches into the study of urban morphology

Abstract (max 200words)

Urban morphology is experiencing growth in quantitative analysis but faces a twofold problem concerning data used to represent urban form - their availability in the first place, and reliability in the second. Both problems concern several urban features, but are especially true in regard to the plot, whose GIS representation capturing a precise morphological meaning is scarcely available. This problem has challenged the authors of this paper to develop a different approach to spatial division. Using the morphological tessellation method, which entails partitioning of space starting from building footprints, utilising the principles of Voronoi tessellation, we explore the possibilities of such method in the definition of analytical units and test their reliability on the case study of Prague, Czechia.

Morphological tessellation and spatial units derived from this analysis are consistent in capturing built-up space uniformly, with no discontinuity. This allows us to follow the topology of morphological structure and determine analytical units based on their contiguity, capturing topological neighbours of variable distance. One of the side benefits of this approach is that block-like units, defined using morphological tessellation are also less dependent on the quality of street network data, whilst at the same time retain the information of variable street width. Empirical tests are presented to assess the reliability of the approach in identifying recurring patterns within urban form. Morphological tessellation partially substitutes the need to derive intricate plot forms from a combination of sources; the research shows furthermore a new approach to capturing spatial structure with reduced data demands.

Introduction

Quantitative urban morphology, or Urban Morphometrics studies urban form through the quantification of its spatial elements. Recently, it has entered the era of big data and powerful computing alongside other scientific fields. However, even though there is currently a variety of sources of data available for analytical purpose, this data does not always meet the standards urban morphologists would need. This is true of several urban form features, but plots are a particularly difficult case: essential to describe urban form, but very hard to be captured and represented rigorously and comprehensively.

To support quantitative analysis, urban morphometrics needs partitioning the built-up area consistently at both small and large scale. Historically, the most widely used spatial unit in urban morphology analysis is the plot, which however raises multiple concerns. First of all, there is no univocal definition of the term, which has been referred to land use (Conzen, 1960), accessibility (Porta and Romice, 2014), or ownership (Moudon, 1986). From an urban morphology perspective, even the purely physical definition proposed by Porta and Romice (2014)[1] shows problems, starting from the simple availability of the necessary information for the identification of the plots as GIS data. Most of the data portals and other GIS sources do not offer any plot layer, and if they do pots are derived from land ownership rather than legible spatial accessibility, while these two concepts often do not match on the ground (Kropf, 2018). Moreover, in sparse modernist urban tissues the reliability of existing plot information in GIS, no matter the definition, is questionable (Levy, 1999).

In urban spatial analysis, "partitioning space" means parcelling space into consistently defined measurable (spatial) units. A method that allows the consistent partition of space is a *morphological tessellation*. This method, which entails partitioning from building footprints, utilises the principles of Voronoi tessellation (Fleischmann et al., 2019). Morphological tessellation objectively defines spatial units named "morphological cells" through the influence that each building has on the space surrounding it (Hamaina et al., 2012). Thus, a morphological tessellation capturing the spatial configuration of urban form derived from building footprints is algorithmic, hence rigorous and objectively replicable. Moreover, it is fully comprehensive, in that it operates consistently across different types of urban form and its source information (building footprint) is widely available (Figure 1). Previous research demonstrates that the informational value of morphological tessellation is in fact similar to that of the plot, which makes the former a reliable alternative to the latter in urban morphometric analysis (Fleischmann et al., 2019).

[1] "A fenced portion of land that is entirely accessible from the public space" (Porta and Romice, 2014, p. 90)

Figure 1: Morphological tessellation of Podbaba neighbourhood in Prague, illustrating the behaviour of tessellation on various types of urban form. The structure of the tessellation mesh directly represents the spatial configuration of buildings in the area.

Since urban morphology analysis aims to capture patterns of urban form, it must describe single elements (buildings, plots, morphological cells) as well as their spatial configurations and relationships. To do so, larger analytical units must be identified. Generally, we can distinguish two approaches to do so: area-based and location-based (Berghauser Pont and Marcus, 2014). Area-based approaches divide space into preselected units, i.e. administrative boundaries (Gielen et al., 2017), abstract projected boundaries (typically grid) (Galster et al., 2001), or larger morphological structures such as a block or a Sanctuary Area (Dibble et al., 2017). However, such methods face two connected issues, together named "Modifiable Areal Unit Problem" (MAUP) (Openshaw, 1984): *scale issue* (how big the area of aggregation should be) and *aggregation issue* (where should we draw its boundaries). Area-based approaches are prone to both of them, particularly the latter: a change of the boundary, for example the voting district, might affect the analysis' results.

Location-based approaches generate analytical units independently for each source-element as a unique aggregation *around* it, typically at walking or driving distance, where distance is measured either along the street network (network distance) or an approximation of it (for example: as a crow flies). Therefore, the aggregated values are uniquely and consistently generated for each source-element (e.g. building), and the effect of arbitrary data aggregation is minimised, resolving MAUP's aggregation issue. For this reason, literature, including the research presented in this paper, prefers location-based analytical units, as their nature partially resolves MAUP. The scale problem part of MAUP is present also in location-based methods, and it is up to the methodology adopted on a case-by-case base to limit its effect to a minimum.

Currently, morphological literature relies on a few methods to define an aggregation of units in a location-based manner. The most straightforward is based on simple Euclidean (as-crow-flies) distance from the elements of analysis (typically radius of 400 metres around a building) (Schirmer and Axhausen, 2015). However, such an approach does not reflect the actual morphological situation on the ground. In some instances, such as in in traditional compact urban tissues, it can capture hundreds of buildings within 400 metres, but only a few in sparse modernist urban tissues, leading to fundamental differences in the amount of information captured, causing issues of comparability of such information.

Excluding effect of certain morphologies from the definition of aggregations overcomes method based on metric reach. Following street network or axial map of urban form (Berghauser Pont and Marcus, 2014; Marcus et al., 2017), it captures the area which is possible to reach within a set distance (mostly metric). As a location-based method, reach is useful because it reacts to unequal morphologies, but only through constrains that limit accessibility to space, rather than through a detection of a difference in urban form itself. The logic is based on the cognitive experience of cities but limited to accessible open spaces, excluding the intra-block relationships. It can generate situations of two buildings facing each other *across* the block (hence directly influencing each other) not being aggregated together, ignoring their relationship.

Both Euclidean distance and metric reach methods cannot capture the change in the granularity of urban tissues, hence effectively measuring different *information* in granular than sparse tissues. In the case of reach, the distance could be defined topologically as a number of steps on the network (represented by the graph) (Berghauser Pont and Marcus, 2014), allowing to recognize the change in the pattern of aggregations, but it still does not eliminate the issue of intra-block relationships. On top of that, network-based methods faces issues in data availability - street networks usually need significant adaptations before they can be used, as they are typically drawn for traffic purposes, not morphological ones; axial maps are scarcely available, and their generation needs a very specialised type of morphological knowledge limiting the applicability of such method.

The third method present in literature is K-nearest neighbour (KNN) analysis, which is also based on Euclidean distance but using it differently. It defines aggregation as a set number of nearest neighbours, defined via as-crow-flies distance. Whilst only scarcely used in urban morphology (Liqang et al, 2013), it has potential as such an approach might reflect changes in the granularity of urban tissues. However, due to the Euclidean definition of nearest neighbours, it cannot react to the detail of some spatial configurations (e.g., be able to detect linear patterns with natural boundaries between as features across the boundary might be closer than those within the pattern). Theoretically, KNN could be used together with reach analysis, joining both the ability to capture morphology represented by networks and scalability of KNN, but we are not aware of any research using this concept so far. However, it would still not resolve the issue of intra-block relationship.

To sum up, literature currently offers two general approaches to aggregation – area-based and location-based. The first one is prone to both parts of MAUP, for which reason it is preferred to use location-based methods. Three location-based methods used in literature are Euclidean distance, metric reach and KNN. However, to date none of these methods is able to capture the structure of urban form which, for its complexity and variance, requires both morphological detailed definition and sensitivity to intra-block relationships.

Thus, we present here a morphological tessellation, which is able to cover the built-up area as a continuum, thus allowing to identify all topological relations between morphological cells, including intra-block relationships, enabling to identify location-based aggregation sensitive to urban patterns.

Topology captures the information on adjacency of neighbouring elements (cells) - two cells are neighbouring if they share at least one point (so-called Queen contiguity) or one segment (so-called Rook contiguity). It defines the proximity of elements in terms of the number of steps needed to get from each element A to each element B. Topological relationships can be of two types - unconstrained, if not limited by any other element than tessellation itself, and constrained, if the step between two neighbours is impeded by constraint (a block is the maximum number of topological steps from element without the need to cross the street network, while the street network is the constraint in this case). Thus, we can define an aggregation around each element based on a number of topological steps (topological reach) on the morphological tessellation, where aggregation defined by n steps includes all morphological cells which we can reach within $x <= n$ steps.

The definition of aggregated analytical units via the topology of morphological tessellation can overcome issues of the three methods described above and provide a more consistent way to understand the relationship between adjacent elements of urban form (in the case of buildings or morphological cells). This paper is testing this hypothesis in the case of Prague, Czechia.

Methodology

The methodology of this research follows a twofold approach, analysing both small scale case studies and urban scale statistical data. Small scale case studies examine the difference between three methods extracted from literature (Euclidean distance, metric reach, K-nearest neighbour) and unconstrained topology of morphological tessellation in different types of urban form; large scale statistical analysis examines the parameters of these methods of aggregation across the whole of Prague.

Prague was selected as a case for its morphological richness of multiple historical layers still intact. The study area is limited by the administrative boundary of the city, but still extends the built-up area of the city by a large margin. The data used within this study were provided by Prague's Institute for Planning and Development (www.geoportalpraha.cz).

We compare how each of the tested methods aggregate tessellation cells (being smallest spatial unit) within two scales: one achieved by nine topological steps, the equivalent of approximately 400 metres used in morphological analysis to represent walking distance of 5 minutes; and one achieved by 4 topological steps, representing roughly 200 metres. The number of neighbours for KNN is then derived from mean number of neighbours captured by each of the topological distance, to keep the dimension comparable (Table 1).

Topology of MT	Euclidean	Metric reach	KNN
4 steps	200 metres	200 metres	70 neighbours
9 steps	400 metres	400 metres	320 neighbours

Table 1: Default topological distances and their equivalents. Values are derived from the summative analysis of topology-based aggregations defined around each morphological cell on Prague.

The aim of small-scale analysis is to understand how each of the four methods identifies and represent the same information across 5 types of urban form - medieval organically grown, 19th century compact perimeter blocks, 20th century mixed single and multi-family villas, 20th century modernist housing, and 20–21th century industrial estates. It should allow to ascertain how consistently each method distinguishes variations in form and morphological behaviours.

At the urban scale we conduct statistical analysis to compare the distribution of values derived from the whole of Prague (its administrative boundary). Statistical distribution of data across whole urban area describes the spread and variance of values, which can be used to assess the ability of each method to capture the intended information across types of urban tissue. To understand the different performance of each method, we compare distributions of two descriptive variables as a proxy for the performance assessment – number of neighbours and covered area.

The first variable is the number of neighbours captured. Neighbours represented by buildings and related tessellation cells capture most of the morphological information. For that reason, it is desirable to use method which will identify somewhat similar number of neighbours no matter the urban tissue to keep the similar essence and amount of information to maximise comparability of values, meaning that the distribution of such values should have relatively small standard deviation and be close to symmetrical distribution to have the similar positive and negative deviations from the mean.

The second variable should represent the concept of geographical extent of aggregation, as it bears the information of the scale of each type of urban tissue and therefore could describe the ability of each method to adapt to the scale. Amongst the possible measurable variables are mean distance to neighbours, maximum distance to neighbours and area covered by aggregation. Because they all represent scale and extent of aggregation of elements (buildings, tessellation cells), we use only area covered to represent them all as it is the most straightforward one and easy to understand. The statistical distribution of covered area should represent the adaptability of aggregation method. For that reason, ideal outcome should have high standard deviation and high range of values, meaning that many different options (levels of granularity of urban form) are all captured.

Analysis/Results

At the small scale, we studied how five urban types in Prague are represented by the four methods of aggregation. The Old Town, a tissue of a medieval origin which has grown organically, shows few differences between the four methods, with slightly larger footprints of aggregations defined by Euclidean distance and a morphological tessellation topology at both 200 (Figure 2a) and 400 (Figure 3a) metres distances. Numerically, the difference is clear, but for pattern-detection this difference is not substantial, suggesting that all methods are relatively equal in this tissue for both 200 and 400 metres (and equivalents). Differences might be explained by the high granularity of the tissue, with many elements on a relatively small area (the reason for KNN being the smallest) and complex configuration amongst them (buildings have many neighbours, expressed by more extensive topological-based aggregation).

]

Figure 2: Comparison of boundaries of aggregations defined by each of the tested method for 4 topological steps and equivalents (200 metres, 70 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice © IPR Praha, CC BY-SA 4.0

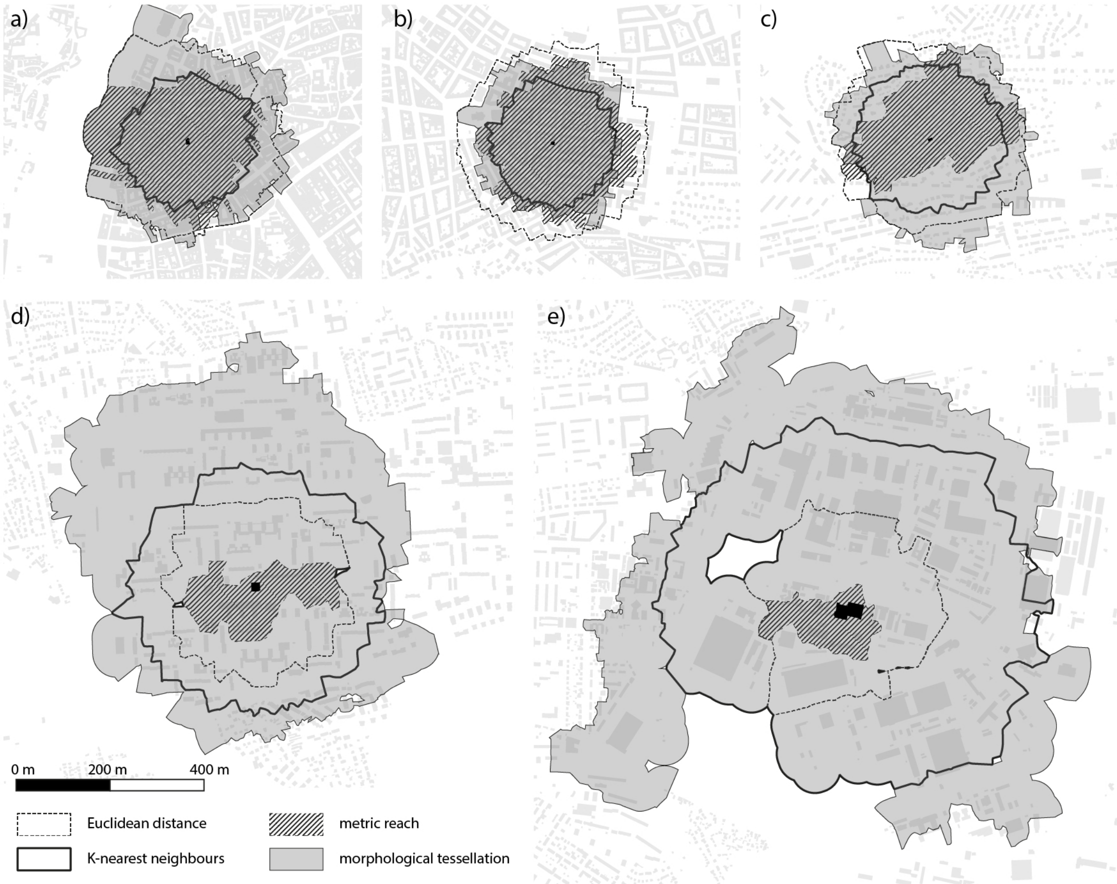

Figure 3: Comparison of boundaries of aggregations defined by each of the tested method for 9 topological steps and equivalents (400 metres, 320 neighbours). a) Old Town, b) Vinohrady, c) Hanspaulka, d) Jižní Město, e) Malešice © IPR Praha, CC BY-SA 4.0

In the second case, the urban tissue of 19th century of compact perimeter blocks in the Vinohrady neighbourhood, results almost matches the previous case. Due to the high granularity of this urban tissue, purely Euclidean distance-based area is the largest, while K-nearest neighbour the smallest (Figure 2b, Figure 3b).

Overall, in these two historic tissues, the difference is not significant to conclude that one method is better than the other - simple visual comparison shows that boundaries almost overlaps for both tested distances.

The first crucial differences are noticeable in Hanspaulka, an area of 20th century mixed single and multi-family villas, where the street network is becoming less dense and less connected than in the city centre, leading to the difference between the area captured by metric reach (smaller) and the other three methods, which almost overlap without any substantial distinction (Figure 2c, Figure 3c). This indicates that the street network plays a crucial role in the applicability of reach-based methods.

The twentieth-century modernist housing of Jižní Město is the first example of a post-WW2 urban tissue. The planning ideology behind it comes with the radical change of scale and a distinctive approach to streets and their connectivity. This change is reflected in how each of our tested methods captures the space: while the area defined by Euclidean distance remains mostly the same as in pre-WW2 tissues, the area captured by metric reach shrinks due to the convoluted street network. In theory, both topological and KNN definitions of aggregation should be able to capture the difference in scale and up to a certain level they do. However, KNN, even though being larger than metric-based methods, lacks the ability to deal with large pavilion-like buildings with many direct neighbours, unlike the topological definition which correctly reacts to the abrupt change of scale of the granularity of urban tissue and captures the relationship between high-rise buildings and their low-rise pavilion counterparts by acknowledging that they are neighbouring (Figure 2d, Figure 3d).

Whilst industrial type tissues are generally not the concern of urban morphology, as classified as specialist and treated differently than more ordinary fabric, they are nonetheless large, therefore important parts of our cities and as such deserve to be studied using the same approach as the more conventional ones. Their scale is radically different. Buildings are of the size of the traditional block or larger, the plot structure is mostly unorganised, and the street network is utilitarian only, following different principles than in residential or mixed-use parts of the city.These differences are captured through the application of our four methods. The network-based method is unreliable on this tissue, capturing the only a minor area around the building due to the major drop in a granularity and connectivity. The Euclidean distance of 200 or 400 metres, which seems to capture enough information in more granular urban tissues lacks the same capacity in this case. K-nearest neighbour analysis struggles to capture the peculiarities of this particular urban tissue, which is characterized by a large amount of additional built-up structures to main buildings, leading to the identification of smaller area that make a comparison with the other cases confused. The topological definition achieved by the morphological tessellation seems to tackle all issues of the other methods, whilst capturing a similar amount of information as it did in previous cases (Figure 2e, Figure 3e).

Overall, the differences between methods in defining aggregation are heavily dependent on the type of urban tissue analysed. More traditional (from a European perspective) urban tissues like medieval (Old Town) or perimeter blocks (Vinohrady) indicate that in these contexts the choice of the method is purely the matter of opinion and that the resulting value offered by the four methods is mostly similar. However, once we start focusing on post-WW2 development, we often observe a change of scale of urban patterns, which makes distance-based methods (Euclidean, metric reach) unable to react to such change. The information captured is consequently different in pre- and post-WW2 urban tissues, complicating the further comparability, whilst we seek similar and consistent data. If the urban patterns change their scale, the method of capturing such extent needs to be able to adapt to it. Our results of the small-scale analysis indicate that topology of morphological tessellation is the method able to fulfil this condition adequately.

Whilst small scale analysis illustrated the capacity of the four selected methods to provide stable information, it is only at an urban scale, through statistical analysis, that we can show a full overview of how the four methods perform.

As mentioned, neighbouring elements are bearing the primary information about urban patterns. For this reason, researchers aim to use methods capturing an equal number of neighbours across contexts. Such a method might be K-nearest neighbour, but due to the variety of urban configurations, a method needs a certain level of adaptability (which KNN with fixed number cannot provide). As Figure 4(a, b) shows, the statistical distribution of the number of neighbours captured is the most stable for topology of morphological tessellation, being almost perfect Gaussian distribution (the deviation in the number of neighbours is the same in both directions from the mean), with the smallest standard deviation (σ). The metric reach method to provides right-skewed distribution and Euclidean distance high deviation, which are both undesirable features in terms of stability of information.

Then, comparison of distributions of covered area aims to test the adaptability of each method. As mentioned, the changing scale of urban patterns means that the same level of information is spread to larger areas. Therefore, an ideal method should show high flexibility (the distribution of values should have large range and high standard deviation) in the area captured to fit all patterns possible. The results as shown in Figure 4(c, d) indicate that topology of morphological tessellation offers by a large margin the highest standard deviation out of tested methods, indicating that the change of the scale is captured successfully. Metric methods (Euclidean distance, metric reach) are the least flexible in this sense, while K-nearest neighbour might offer desired value alongside with morphological tessellation topology.

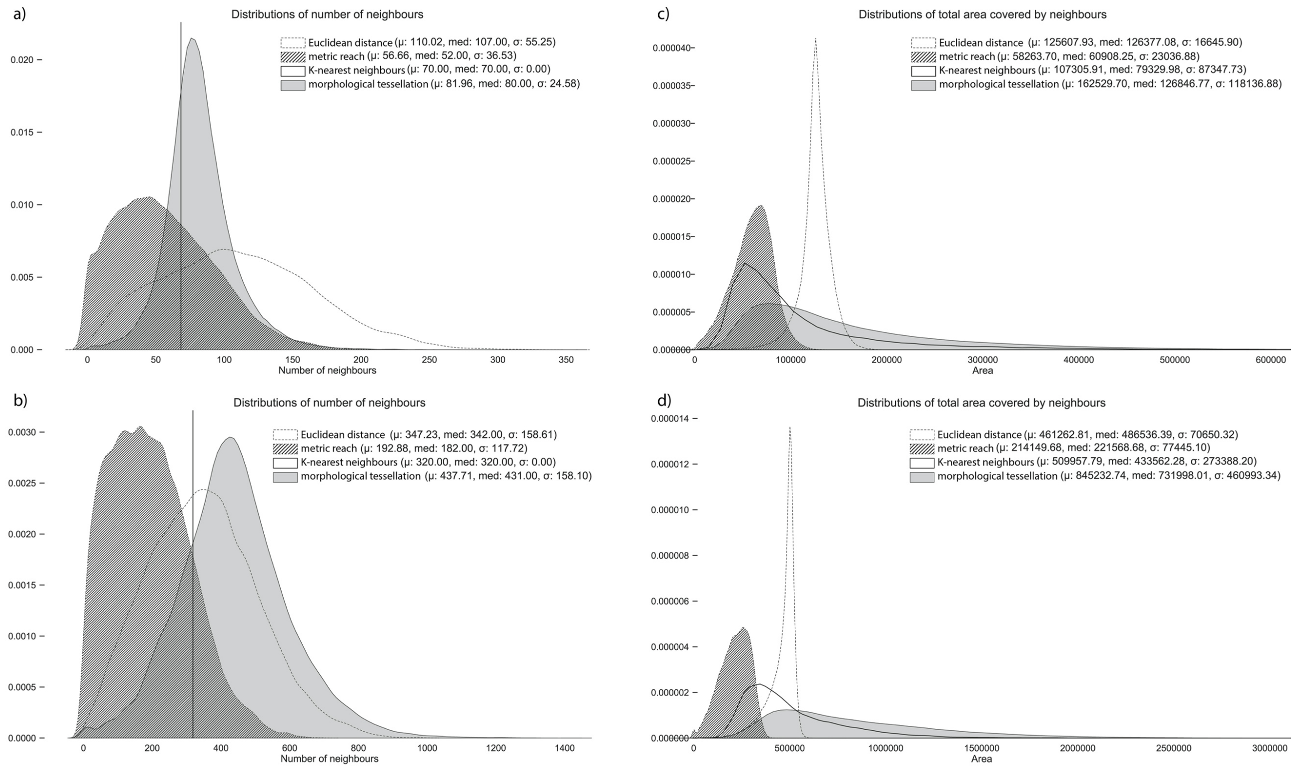

Figure 4: Statistical distributions of number of neighbours a) 4 steps and equivalents, b) 9 steps; and total covered area c) 4 steps and equivalents, d) 9 steps.

Even though there are differences between smaller and larger distances (4 steps / 9 steps and its equivalents), topologically defined aggregation seems to reflect desired outcomes (i.e. stability in number of elements captured and flexibility in metric values) in both statistical comparisons better than other tested methods. This finding is in line with the one we drew from small-scale case studies, indicating that topology of morphological tessellation is a valuable approach to be employed in morphological analysis.

Discussion/Conclusion

Existing location-based methods of aggregation of elements into larger analytical units all face some issues limiting their applicability and reliability. The alternative method presented is based on the topology of space as captured by the morphological tessellation. Such a method of partitioning space reflects the influence of each building on the space around it with the aim to overcome existing challenges and provide a context-sensitive method. Initial results of the twofold analysis of the topological ability of the morphological tessellation indicate that the type of urban tissue influences the outcome of morphological analysis, and that in the case of pre-WW2 traditional European-like urban tissues, all currently available methods of definition of aggregation are relevant and almost interchangeable. However, this is not the case with post-WW2 urban developments, as in them there has been a significant change in the scale of form's granularity. In these cases,

urban morphology needs to employ methods which are sensitive to the scale and configuration of urban form *and* at the same time can detect its granularity. The morphological tessellation and the topology derived from the analysis of its structure seem to be the most successful, sensitive method, suitable for general analysis.
All of the methods that have been tested partially solve one of the key issues identified in spatial analysis (MAUP), as data are aggregated independently for each element and there are no preselected boundaries in play.
No matter the results of presented analysis, methods extracted from literature have their role in morphological analysis. However, Euclidean definition and metric reach should be used in specific situations only, due to their abovementioned limitations. It is either in stable environments without abrupt changes of granularity or in a definition of larger-scale aggregations, where multiple urban tissues are included. In that case, the main benefits of the morphological tessellation — following the spatial configuration of urban patterns — is not so crucial and from certain scale does not even provide added value.
For the analysis done on the small scale (scale of urban tissue and smaller), Euclidean definition and metric reach do not provide a stable information, unlike KNN (which always captures the same number of elements of similar informational value, but not the same relationship) and the topology of morphological tessellation. Moreover, aggregation defined via the topology of morphological tessellation may be used even on the smallest scales of one or two steps as it will always capture intended comparable information based on the relationship of elements.
Currently, tools to generate a morphological tessellation and work with its topological relations are available under a MIT license as a part of open source Python package named "momepy – Urban Morphology Measuring Toolkit" (Fleischmann, 2019), which, together with its source code, is publicly available and is under active development.
Further research should focus on the question of the exact meaning and variation of topological distance and its definition for specific purposes. To date, the question of how many topological steps should be used for the analysis of urban form remains open. It is expected that it will vary depending on the scope of the research.
The morphological tessellation is a step towards achieving consistency in urban morphology in both definitions of the smallest spatial unit (Fleischmann et al., 2019) and analytical aggregation. The advantage of morphological tessellation is that it limits the data dependency as it is based on building footprints only and allows the elimination of subjectivity in the partitioning of space. Most importantly, it is context sensitive allowing the researcher to use the same method across different types of urban tissues whilst still get comparable information, much needed for reliable results of any statistical analysis.

References

Berghauser Pont M and Marcus LH (2014) Innovations in measuring density: From area and location density to accessible and perceived density. Nordisk Arkitekturforskning (2).
Conzen M (1960) Alnwick, Northumberland: a study in town-plan analysis. London: George Philip & Son.
Dibble J, Prelorendjos A, Romice O, et al. (2017) On the origin of spaces: Morphometric foundations of urban form evolution. Environment and Planning B: Urban Analytics and City Science 24(1): 239980831772507–24.
Fleischmann M, Romice O, Porta S (under review, 2019). Morphological tessellation as a way of partitioning space: improving consistency in urban morphology at the plot scale. Manuscript submitted for publication.
Fleischmann M (2019) momepy. Glasgow: Urban Design Studies Unit. Available from: https://github.com/martinfleis/momepy.
Galster GC, Hanson R, Ratcliffe MR, et al. (2001) Wrestling Sprawl to the Ground: Defining and measuring an elusive concept. Housing Policy Debate 12(4): 681–717.
Gielen E, Riutort-Mayol G, Palencia-Jiménez JS, et al. (2017) An urban sprawl index based on multivariate and Bayesian factor analysis with application at the municipality level in Valencia. Environment and Planning B: Urban Analytics and City Science.
Hamaina R, Leduc T and Moreau G (2012) Towards Urban Fabrics Characterization Based on Buildings Footprints. In: Lecture Notes in Geoinformation and Cartography, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 327–346. Available from: https://link.springer.com/chapter/10.1007/978–3–642–29063–3_18.
Issacharoff S (2002) Gerrymandering and political cartels. Harv. L. Rev., HeinOnline 116: 593.
Kropf K (2018) Plots, property and behaviour. Urban Morphology 22(1): 5–14.
Levy A (1999) Urban morphology and the problem of the modern urban fabric: some questions for research. Urban Morphology, INTERNATIONAL SEMINAR ON URBAN FORM 3: 79–85.
Liqiang Z, Hao D, Dong C, et al. (2013) A spatial cognition-based urban building clustering approach and its applications. International Journal of Geographical Information Science 27(4): 721–740.
Marcus LH, Berghauser Pont M and Bobkova E (2017) Cities as accessible densities and diversities. In:, pp. 1–13.
Moudon AV (1986) Built for change: neighborhood architecture in San Francisco. MIT Press
Openshaw S (1984) The Modifiable Areal Unit Problem.
Porta S and Romice O (2014) Plot-based urbanism: towards time-consciousness in place-making. In: Mäckler C and Sonne W (eds), Dortmunder Vorträge zur Stadtbaukunst [Dortmunder Lectures on Civic Art], Sulgen, DE: Niggli, pp. 82–111.
Schirmer PM and Axhausen KW (2015) A multiscale classification of urban morphology. Journal of Transport and Land Use 9(1): 101–130.