



2837

REPRINT

Geoffrey Caruso, Mohamed Hilal and Isabelle Thomas

Measuring urban forms from inter-building distances: combining MST graphs with a local index of spatial association

Landscape and Urban Planning, 163, 80-89, 2017

CORE
CENTER FOR OPERATIONS RESEARCH AND ECONOMETRICS

CORE
Voie du Roman Pays 34, L1.03.01
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
Email: immaq-library@uclouvain.be
https://uclouvain.be/en/research-institutes/immaq/core/reprints.html

Landscape and Urban Planning 163 (2017) 80–89

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Paper

Measuring urban forms from inter-building distances: Combining MST graphs with a Local Index of Spatial Association

Geoffrey Caruso[a,*], Mohamed Hilal[b], Isabelle Thomas[c]

[a] Institute of Geography and Spatial Planning, University of Luxembourg, Luxembourg
[b] CESAER – Centre d'Economie et Sociologie appliquées à l'Agriculture et aux Espaces Ruraux, UMR1041 INRA-AgroSup Dijon, France
[c] FRS-FNRS and CORE – Centre for Operations Research and Econometrics, Université catholique de Louvain, Belgium

HIGHLIGHTS

- Automatic method to characterise local urban patterns from inter-building distances.
- Combination of Minimum Spanning Tree graph and Local Index of Spatial Association.
- Preserve structural form of residential settlements.
- Non-aggregate complement to morphological and density based metrics.

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form
22 December 2016
Accepted 2 March 2017

Keywords:
Urban sprawl
Built-up morphology
Green space
Graph partitioning
Minimum Spanning Tree (MST)
Local Index of Spatial Association (LISA)

ABSTRACT

A new method is proposed for characterising local urban patterns at the scale of a large urban region. The approach overcomes the difficulties of surface-based representations of built-up morphologies and provides an efficient way to account for the proximity of built and non-built land. The strength of the approach resides in the direct use of the coordinates of each building, a very parsimonious input of external parameters, and a local spatial statistical perspective. The method consists in, first, constructing and clumping the Minimum Spanning Tree (MST) that links each building to its neighbours in order to obtain connected urban and suburban graph subsets, and second, measuring a Local Index of Spatial Association (LISA) on inter-building distances in order to characterise patterns. The method differentiates heterogeneous settlements from statistically significant homogeneous clusters with either long or short connections between buildings. The method is applied to the Southern periphery of Brussels (Belgium).
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In this article we propose a new approach to characterise local urbanisation forms. Our work is motivated by urban sprawl debates and the necessity to come up with planning tools that go well beyond simple density targets and therefore account for the local spatial arrangement of buildings. Our method is designed to avoid problems inherent to surface-based approaches and aggregation procedures that erase crucial information about how buildings and the green space in-between interact. We show how this can be achieved for large areas simply from knowing the precise location of buildings and the combination of existing graph theoric and spatial autocorrelation tools.

1.1. Urban sprawl and green space

Urban sprawl has been one of the hottest topic in urban planning for the last fifty years. Defined – by early discussants (Harvey & Clark, 1965, p. 2) – as a low density, scattered and heterogeneous mode of urban development at the periphery of cities, it is still a key issue that challenges environmental sustainability and bears important socio-economic costs (e.g. Ewing, 2008; Ewing & Hamidi, 2015). Planning practice today prescribes urban infill and compactness as ways to counteract sprawl. More specifically, the fragmentation aspect of sprawl is targeted because it is detrimental to the connectivity of wildlife habitats (e.g. Fischer & Lindenmayer, 2007) or because it pulls activities apart and brings them further away from main centres, thus increasing trip length and car-dependence (e.g. Ewing & Cervero, 2010; Handy, Boarnet, Ewing,

* Corresponding author at: Institute of Geography and Spatial Planning, University of Luxembourg, Maison des Sciences Humaines, Campus Belval. 11, porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg.
E-mail addresses: geoffrey.caruso@uni.lu (G. Caruso), mohamed.hilal@inra.fr (M. Hilal), isabelle.thomas@uclouvain.be (I. Thomas).

http://dx.doi.org/10.1016/j.landurbplan.2017.03.003
0169-2046/© 2017 Elsevier B.V. All rights reserved.

& Killingsworth, 2002) and eventually impacts human health, CO_2 emissions and energy use. Also the loss of open and agricultural land is targeted, because conversion is irreversible and is an important market failure when the resource value of open land is underestimated (Brueckner, 2000).

Densification policies, however, are not without unexpected impacts. Although it is still debated (see e.g. Ewing, Richardson, Bartholomew, Nelson, & Bae, 2014), it is now evident that making cities just more compact or dense is not a panacea and potentially holds adverse effects. For example, while sprawl increases commuting distances, compactness may lead to more congestion and time loss (e.g. Boarnet & Crane, 2001). In addition, compactness resulting from growth boundaries or green belts may lead to even longer commutes (e.g. Hall, 1973; Mace, Blanc, Gordon, & Scanlon, 2016) and the perception of density itself may lead to a decreasing attractiveness of centrality for households, especially where no compensatory green space are maintained in the center. In fact, while people enjoy social contacts and proximity to services and retail, they also enjoy the very close proximity to green space (e.g. Cavailhès et al., 2009; Geoghegan, 2002), which is a reason for emerging leapfrogging and sprawl forms (Caruso et al., 2011; Peeters et al., 2015). Moreover, while sprawl patterns increase total emissions because of longer commutes, they also reduce the exposure to trafic pollutants compared to more compact patterns (e.g. Schindler & Caruso, 2014). Besides these effects on households, compact cities also leave very little space for strong biodiversity links across the urban footprint. While the fragmentation of suburban and rural landscape is important to be treated, large and compact cities tend to become no man's land ('*halo*') for wildlife (Wade, Wickham, Zacarelli, & Riitters, 2009).

In this context, urban planning needs to come up with novel solutions that account for interactions between built-up forms and undeveloped land and should particularly recognise the ecological value of green links within cities and the value of very local green space for citizens. As part of identifying better policies, it is therefore important to measure urbanisation patterns with a high spatial granularity, including the relative local arrangement of buildings and open space. Despite much progress in urban form research, the vast majority of local authorities still evaluate and plan urban development in terms of density (of housing or population) per administrative unit, which is far too coarse to inform about local patterns, the interconnection of uses, or to analyse landscape changes (see e.g. Antrop, 2004). It typically hides information on how green space is positioned with respect to built-up land and how buildings relate to each other.

1.2. Measuring urban forms

There are many, diverse, and rather long traditions to analysing urbanisation forms across disciplines. The qualitative visual assessment of settlement shapes in terms of archetypes (star-shaped, street villages, etc.), and then the introduction of simple shape metrics in the 1960s (Boyce & Clark, 1964) can be seen as such a tradition in geography and planning. The last 20 years have seen advances along-with progress in GIS and concepts borrowed from landscape ecology (e.g. McGarigal & Marks, 1995; Turner & Gardner, 1991) or fractals (e.g. Batty & Longley, 1994; Frankhauser, 1998). Applied landscape metrics, using land use patches as base elements, and fractal metrics, using built-up areas or networks as input, are especially important to document the fragmented aspect of sprawl (Irwin & Bockstael, 2007) or the existence of non-built lacunes within the urbanised space (Thomas, Frankhauser, & Biernacki, 2008; Thomas, Frankhauser, & De Keersmaecker, 2007). Different traditions can be observed in urbanism, architecture and design studies, including qualitative urban morphology approaches to the analysis of town plans across time (Whitehand, 1992), the analysis of networks with space syntax (Hillier, 1996), space grammar design (March & Steadman, 1974), or spatial matrices (Berghauser Pont & Haupt, 2007), as well as their recent combinations (Griffiths, Jones, Vaughan, & Haklay, 2010; Stanilov, 2010; Ye & Van Nes, 2014).

Given the need to provide clarity in raging urban sprawl debates, Galster et al. (2001) conceptualised eight dimensions of urban sprawl and actually triggered a new literature, rather distinct from the previous set, with a focus on systematically addressing different facets of urban expansion forms comparatively across metropolitan areas. Arribas-Bel, Nijkamp, and Henk (2011), Frenkel and Ashkenazi (2008), Hamidi and Ewing (2014), Torrens and Alberti (2001), Wolman et al. (2005) are example belonging to this regional aggregate-multidimensionality line of research, and each propose deeper reviews of the field. Clifton, Ewing, Knaap, and Song (2008), Hasse and Lathrop (2003), Lowry and Lowry (2014), Song and Knaap (2004) also brought in multiple dimensions of urban forms but represent a field of research that stresses analysis at a more local scale (neighbourhoods, communities), where spatial configurations can be related to proximities, transportation systems and the accessibility of different activities, thus being more palatable to policymakers as argued by Song and Knaap (2004) and actually being closer to the urban morphology and urbanism approaches mentioned above.

We position our work at the intra-regional scale and contend that recent metrics are too aggregated, while those that use the most detailed information in the urban morphology field can hardly be made systematic across large areas given the need for diverse data (parcels, networks, building use, etc.). Most applications also need many exogenous parametric choices, such as anaysis windows, buffer size, contiguity type, thresholds, or range cut-offs. Our work is in line with Hasse and Lathrop (2003, p. 1022), searching for metrics that '*characterise urban growth at its atomic level, namely the housing unit*'. Conversely to Hasse and Lathrop (2003) however, who use buildings as a proxy for population and re-aggregate measures at patch level (thus following a landscape ecology methodology), we believe it is important to keep the housing unit scale, especially when the aim is to account for the local relationships between buildings and non-built uses. We also insist on parsimony in parameters and the sole use of a simple (and now almost ubiquitous) input data – building units – so that automation and comparisons are in reach.

Finally, we think that there is an intrinsic difficulty with using density-based approaches but also landscape metrics to measure forms because they rely on areal units. Homogeneity of land use over a certain surface is assumed whatever the resolution, while non-artificial land is usually interspersed between buildings and participate to the quality of a place for people or wildlife that an urban patch may have difficulties to represent. In that sense, landscape metrics fall in the same trap as density measures over a surface. Beyond spatial resolution issues, we argue that surface-based approaches, typically using density or aggregate measures over a series of units or land use patches, even if they can account for topologies and local forms (e.g. perimeter to area ratios or the share of a perimeter that is a contact between two land uses) handicap the description of urban patterns and particularly the spatial arrangement of buildings. Non-surfacic descriptors, based on graphs, are particularly appealing because they preserve structural information about location and proximities of buildings beyond the simpler contiguity aspect and in a manner that is less dependent on the quality of the spatial representation of buildings. Moreover, graphs can be handled in a very parsimonious manner, with few parametric inputs, hence applied systematically to large regions and comparatively across places. Urban and Keitt (2001) have addressed similar limitations and prospects in the context of landscape ecology and suggested increased use of graph-theoric approaches.

Just like using graphs instead of patches was an important perspective shift for landscape ecology (Foltête, Clauzel, & Vuidel, 2012; Foltête, Clauzel, Vuidel, & Tournant, 2012; Jaeger, 2000; Jaeger, Soukup, Madrinan, Schwick, & Kienast, 2011; Tannier, Foltête, & Girardet, 2012; Urban & Keitt, 2001), we argue that a similar shift is important to address urban sprawl forms from a residential perspective as well. In ecology, graphs contributed to further liaising landscape and individual processes with the promise of higher performance in conservation planning. We aim here at developing a graph-based measure of residential settlements from buildings and their local separation, therefore linking built-up and non-built parts of the landscape at the finest scale possible, eventually for a better consideration in planning.

We propose an innovative method that combines graph theory and local spatial autocorrelation analysis. The sole input of the method is the precise geographic location of all individual buildings. The method then consists of constructing a graph that links each building to its neighbours, subdividing the graph into subgraphs and then further into either homogeneous or heterogeneous local patterns with strong or loose connections between buildings. Those local patterns are identified in a statistically robust manner from a Local Index of Spatial Association (LISA) (Anselin, 1995) applied to the distance that separates each building within a subgraph.

2. Methods

Our methodology runs across a series of steps represented in Table 1 and described below.

Step 1. The first step is a simple GIS (Geographic Information System) pre-processing to retrieve centroids from polygons. Since our key variable of interest is the distance separating buildings (inter-building distance), transforming polygons to centroids leads to some approximation because of the shape and dimension of buildings vary. However, several studies have documented the effective simplification and reconstruction of surfaces by points (Amenta, Bern, & Kamvysselis, 1998; Pauly, Gross, & Kobbelt, 2002; Samozino, Alexa, Alliez, & Yvinec, 2006). In our case, we use a cloud of points as implicit ground surfaces of built-up areas. Our measurement error is small when comparing the distribution of inter-building distances with the distribution of surfaces across the study area. In our case study, the standard deviation in inter-building distances is 1.5 times longer than its mean (see results) while the standard deviation of building size is 30 m^2 for an average of 110 m^2, i.e. more than 3 times smaller. The approximation also seems homogeneous across the area and does not appear to impact our results. Both means and standard deviations do not show significant spatial variations across aggregated spatial units.
Step 2. Our second step consists in triangulating all centroids to obtain a Delaunay graph. A graph consists in a set of points (vertices) and arcs (edges or segments). The Delaunay graph is a subset of the complete graph (see Fig. 1 from a to b). It is undirected and planar and defines non-overlapping convex regions over the portion of space under study (Okabe, Boots, Sugihara, & Chiu, 1999). Most standard GIS systems have the ability to output such a tessellation. In particular the Delaunay graph can be output from Voronoi (or Thiessen) polygons generated from the original building points, by creating arcs between all vertices that share a Thiessen polygon boundary. The Delaunay graph and subsequent subgraphs have been computed here within MATLAB.
Steps 3 and 4. The third and fourth steps lead to the creation of a Minimum Spanning Tree (MST). The MST is a subset of the Delaunay graph (see Fig. 1 from b to c), where each building is linked to the neighbouring ones in such a way that the total cost

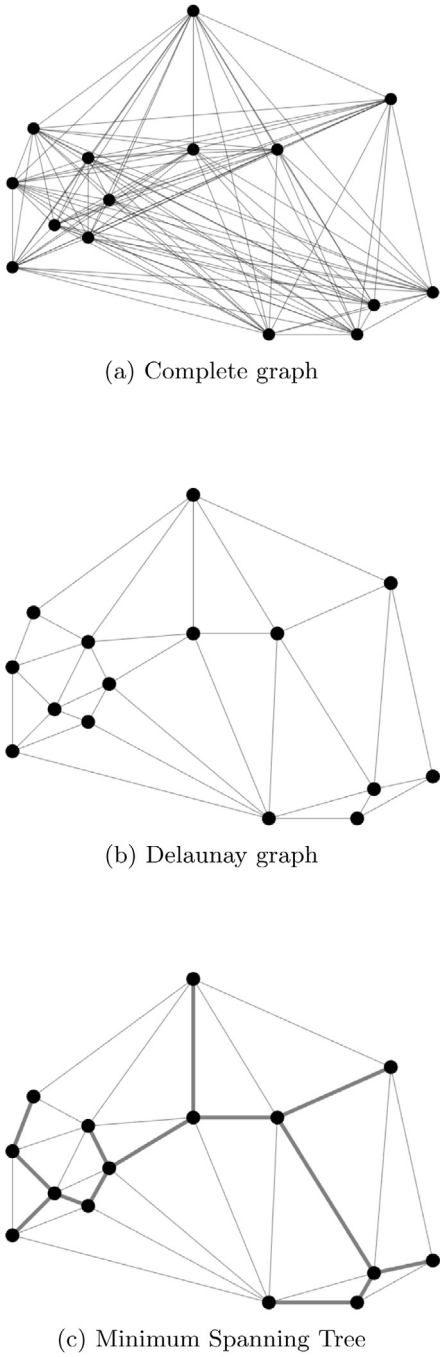


Fig. 1. Methodological steps 2–4: from Complete graph to Delaunay graph and the Minimum Spanning Tree (MST).

Table 1
Methodological steps.

Pre-processing	
Step 1	Polygon to centroid transform
Graph creation and segmentation	
Step 2	Computing Delaunay triangulation
Step 3	Weighting Delaunay graph edges with Euclidean distance
Step 4	Computing Minimum Spanning Tree
Step 5	Graph segmentation using distance cut-off (200 m)
Characterisation of local patterns	
Step 6	LISA analysis

of all edges is minimised. Minimum Spanning Trees (Graham & Hell, 1985) have long been of interest in many research fields (see Wu & Chao, 2004, for a comprehensive review of methods with applications), including applications in landscape and ecological connectivity (Fall, Fortin, Manseau, & O'Brien, 2007; Urban & Keitt, 2001). Connecting every building within a given area, in the cheapest possible way, is also very useful in urban network management since it represents a way of laying cables or pipes for transporting energy, waste, water, etc. MST graphs are particularly appealing because (i) they can be computed quickly and easily, (ii) they create a sparse subgraph that preserves structural information of the original graph, and (iii) they provide a simple way to identify clusters in sets of points. Indeed, deleting the long edges from a MST, like pruning a tree, returns connected components, or clusters, that keep their original (natural or backbone) structures.

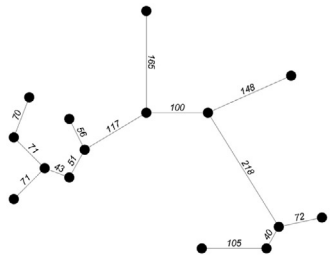
In our case the unitary cost considered to build the MST is simply the length of each segment, i.e. the inter-building distance. We weight every edge of the Delaunay with the Euclidean distance between the two vertices it connects (step 3) and then compute the MST (step 4) using Prim's algorithm. Note that using Prim or Kruskal algorithms for computing the MST does not actually require the Delaunay to be defined first (Okabe & Funamoto, 2000, p. 116). However the analysis of the Delaunay brings further information on distances and is also attractive because it is easily computed with most GIS software.

Step 5 The fifth action aims at subdividing the MST graph into subgraphs as shown in Fig. 2 from a to b. Since we use the MST graph rather than the Delaunay, applying a cut-off automatically leads to separate sub-graphs of close buildings.

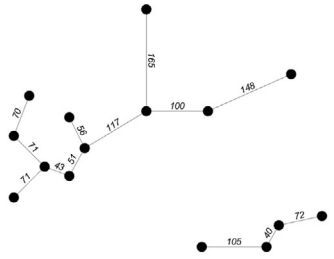
We use here an exogenous distance cut-off of 200 m. This is a value found in many studies, as early as Nordbeck (1971), with remote sensing or small area census-based approach to defining continuous urban settings see e.g. Donnay, Barnsley, and Longley (2003) p. 136, or Caruso (2001) for a review of definitions in Europe. Every segment longer than 200 m is removed at this stage.

Instead of a cut-off, there exist many segmentation algorithms to prune such a tree without any *ex-ante* knowledge, but directly from the data, using either local or global clustering criteria. The literature on these methods is quite huge with applications in many non geographical domains. The methods compare on efficiency, rapidity, capacity to identify cluster in non-homogeneous point patterns or complex images. This step 5 could easily be replaced for example by specific agglomerative or divisive algorithms using global optimisation criteria (e.g. minimising differences within groups, maximising inter cluster distance, or maximizing standard deviation reduction) similarly to Asano, Bhattacharya, Keil, and Yao (1988), Grygorash, Zhou, and Jorgensen (2006), Peter (2011)) or local information such as defining significantly larger segments among neighbours in line with the work pioneered by Zahn (1971). Our focus is slightly different here because we aim at characterising the within structure of subgraphs rather than identifying the subgraphs themselves in the most efficient manner. At this stage of research therefore we prefer a simple constant cut-off.

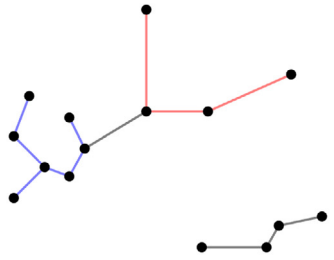
In urban contexts, the methods developed by Deng, Liu, Cheng, and Shi (2011), Estivill-Castro and Lee (2002), Okabe and Funamoto (2000) are appealing for identifying subgraphs. In particular we share similar interest with both Estivill-Castro and Lee (2002) and Deng et al. (2011) because they avoid using only a global argument to define clusters but focus on the local variation of segments' length. Both papers propose very good candidate algorithms to replace our simple cut-off. To avoid chaining and multiple bridge effects and thus a better identification of clusters, they start however with the Delaunay graph rather than the MST. While we can



(a) Minimum Spanning Tree (MST) and length of segments



(b) Subgraphs after 200m cut-off



(c) LISA map. Blue: homogeneous cluster of small inter-building distances (LL). Red: homogeneous cluster of large inter-building distances (HH). Grey: heterogeneous patterns)

Fig. 2. Methodological steps 5 and 6: Minimum Spanning Tree (MST) and LISA.

acknowledge this improves clustering (subgraphs identification), the MST is more in line with our goal to characterise internal structures at the local level, a representation of proximities that reduces computation time in the next stage (step 6), and mimics the linkage between buildings along streets rather than across backyards. This is a functional property of suburban or exurban settlements that cannot be ignored in planning. Some segments can be long but can still make sense for identifying the way people are able to travel from a group of houses to another, the absence of backyard to backyard functional connection (no path), or other barriers. The MST appears as a judicious choice compared to more complex set of tools such as SAINF (Okabe & Yoshikawa, 2003) or SANET (Okabe, Okunuki, & Shiode, 2006) that can explicitly account for infrastructures but need more exogenous parameters.

Step 6 This part of the methodology shares similarity with graph cutting methods that use a local perspective, i.e. a non global decision parameter to derive sub-groups. However, we consider local statistics here as a descriptor of pattern rather than for cutting.

We use the local Moran's *I* index (i.e. LISA), which is defined as the weighted product of the difference to the mean of the value of a variable at a certain observation and the same difference for all

other observations, with more weight given to the observations in close spatial proximity (Anselin, 1995). While it is a widespread way to identify clusters of similar values in space, with many applications related to socio-economic processes (Anselin & Rey, 2009), it has seldom been used to characterise spatial clusters of urban forms. Fan and Myint (2014) have applied local Moran and Getis indices to measure urban fragmentation, comparatively with landscape metrics, but used raster images, not vector objects. In addition, conversely to other Local Moran's I applications, we do not compute the index for a particular characteristic of a geographical object (building) but consider the link between two objects (the segment linking buildings) as an observation and the length of that link as the variable of interest. We therefore compare, for each segment of our MST subgraphs, the inter-building distance to neighbouring inter-building distances using the following:

$$I_i = \frac{\sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_j (z_i - \bar{z})^2} \quad (1)$$

where I_i is the LISA index (local Moran) for segment i within a given subgraph of the MST; z_i is the length of segment i; z_j is the length of every segment j other than i in the subgraph; $\bar{z}$ is the average length of all segments within the local subgraph; and w_{ij} is a pairwise weighting parameter indicating the proximity of i and j segments.

The spatial weight matrix w_{ij} is calculated independently of the MST segments, in other words all segments in the subgraph are considered for building the matrix irrespective of their connecting sequence in the MST. Two segments are simply considered as being neighbours ($w_{ij} = 1$) if the Euclidean distance between their centre falls within a distance cut-off of 200 m. $w_{ij} = 0$ otherwise. Tests have been undertaken with more complex proximity and contiguity matrices w_{ij} (available upon request) but were not more conclusive (see Peeters and Thomas (2009) for a wider discussion of network autocorrelation and determination of weights matrices). The significance of the LISA values is obtained after a Monte Carlo process (see Anselin, 1995), simulating the distribution of LISA, under null hypothesis, from repeated random sampling of data values. Pseudo p-values are calculated by comparing the observed value of the test statistic to the upper and the lower tails of the null distribution. Significance is reported at 0.05 level.

A significant positive value of I_i means that the segment i is surrounded by segments of similar length, i.e. that the inter-building distance is locally homogeneous. There are two such homogeneous patterns, as depicted in Fig. 2c: if the inter-building distance is large, nearby buildings are significantly further away from each other than expected in the subgraph and the segments coloured in red (a 'High' distance surrounded by 'High' distances, i.e. HH cluster to follow the usual LISA categorisation); if the inter-building distance is small, nearby buildings are significantly closer to each other than expected in the subgraph and the segments coloured in blue ('Low' distance surrounded by 'Low' distances, i.e. LL cluster).

It is very tempting to oppose the blue and red segments as one would oppose compact clusters to sprawl development. Our characterisation however is more precise. First, the red and and blue colours identify settings that are homogenous while sprawl in the literature may refer to heterogeneous settlements as well (Harvey & Clark, 1965). Second, the patterns are identified without a density constraint, and while blue settlements are indeed more internally connected and thus compact, they can also cover less space and have similar number of buildings than red settlements. We would rather talk about patterns with tight or loose links between buildings rather than high density or low density patterns, thus emphasizing the importance of the space in-between buildings.

A significant negative value of I_i means that segment i is much lower or higher than the surrounding ones, i.e. the segment is a spatial outlier relative to its neighbours (so-called 'High–Low' or 'Low–High' types). This case reflects a significantly heterogeneous cluster. If Moran's I is not significant, there is no specific cluster, which we also consider as an heterogeneous settlement (greyed in Fig. 2c).

3. Results and discussion

We apply our method to the Southern periphery of Brussels in Belgium. This is a highly dynamic commuting region – not comprising cities with high rise buildings – where post-war and more recent urbanisation has been loosely or more tightly embedded within a dispersed pattern of small towns and villages. Though generalisation from a single case study is always limited, this is an interesting case that offers some spatial heterogeneity (North/South and East/West) and has been studied earlier with fractals by Thomas et al. (2007).

Results are organised according to the three main methodological sections of Table 1.

3.1. Data input and pre-processing

The base material needed is a vector (cadastral) map of all residential buildings (Fig. 3). In this case, we use the data from the Plan de Localisation Informatique (PLI) developed by the regional administration (MRW-DGATLP, 2004). Unfortunately, this dataset includes polygons of the footprint of every building without any

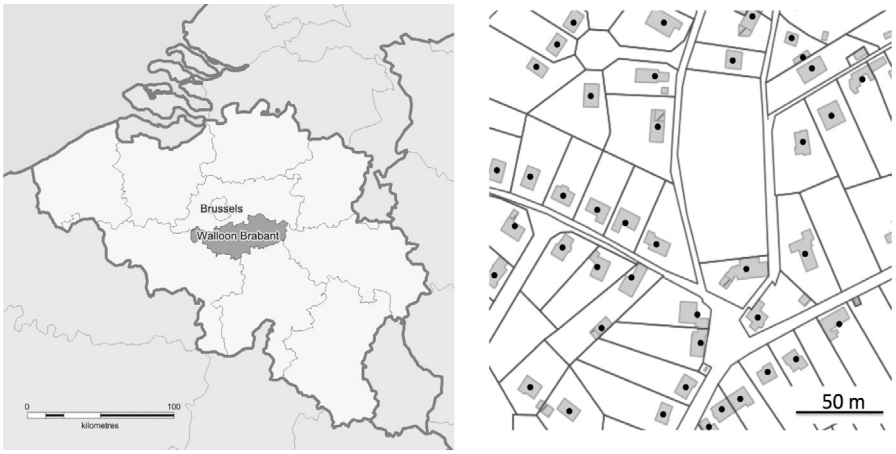
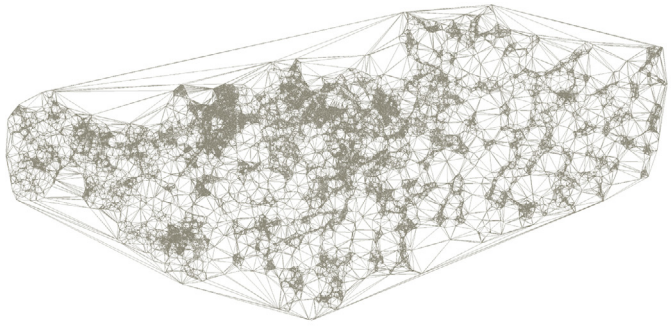


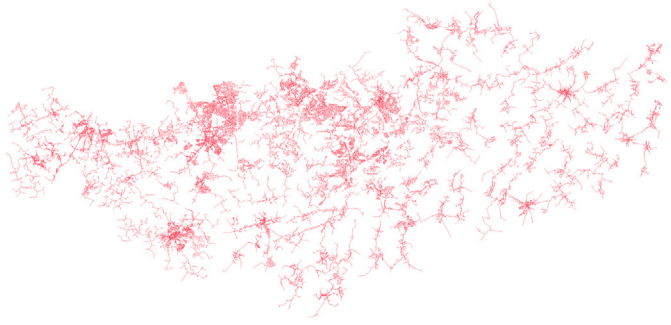
Fig. 3. Location of study area (left) and example of input data with plots, buildings and centroids (right).



(a) Centroids and limits of municipalities



(b) Delaunay graph



(c) Minimum Spanning Tree (MST)

Fig. 4. Steps leading to the MST.

reliable information about the use of buildings (e.g. residence, public service, retail, farm, etc.), nor about their occupants (number of inhabitants or jobs within a building). In the present case, the selection of 'residential' buildings is therefore made by excluding buildings larger than 5600 m^2, a cut-off chosen after photographic interpretation in order to remove stations and other large plants and infrastructures. After a qualitative cross-check, our results do not seem to suffer from this approximation because the number of very large buildings is marginal. In further applications however, it is desirable to select buildings from use information.

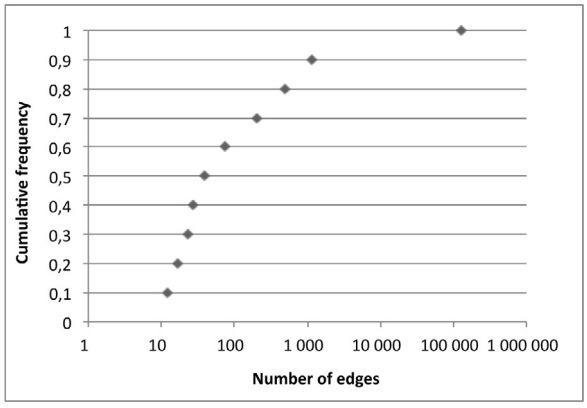
After cleaning for non-residential buildings and computing centroids, the study area hosts 202 531 points (Step 1). As shown from mapping all centroids (Fig. 4a) the buildings are located unevenly across the region, showing the urban influence of Brussels in the Northern part and a more rural context in the Eastern part.

3.2. Graph creation and segmentation

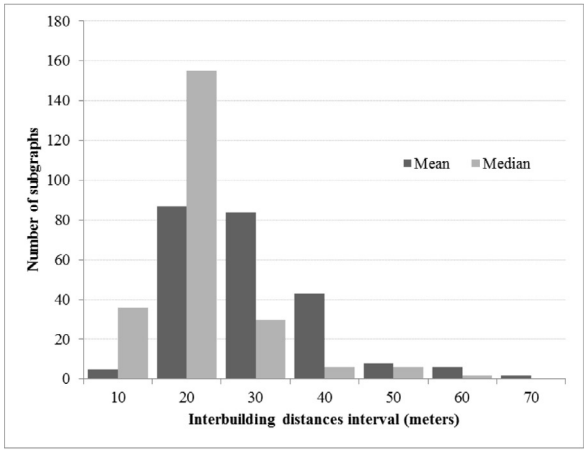
The second step leads to the Delaunay triangulation (Fig. 4b), comprising 607 564 segments with an average inter-building distance of 56 m and a median of 30 m. The standard deviation is 193 m, which is very large because the Delaunay connects villages and towns as well as all remote buildings and draw links over the borders of the study area. While the Delaunay could in principle be used as an input to the LISA analysis, we see from these statistics that it is the bearer of non-local information and border biases. Turning to the MST solves these problems and reduces considerably the number of segments to be characterised in later stages.

The MST is obtained (Step 4) after weighting each segment by its length (Step 3). The resulting MST (Fig. 4c) has – by construction – a number of segments equal to the number of buildings minus one, i.e. 202 530. The average inter-building distance is 20 m and the median is 15 m. The standard deviation is 29.6 m, which is a considerable reduction in the statistical variation compared to the Delaunay and shows the uneven distribution of points in the studied region.

The next stage (Step 5) consists in clumping the MST based on a given cut-off. As explained above we use a 200 m cut-off in line with the literature. We end up with 666 subgraphs, two thirds of which comprise less than 10 buildings. We remove those very small subgraphs because the LISA analysis would largely be non-significant given the limited number of units. Removing these subgraphs however do not significantly impact our analysis since only 1713 buildings are taken out, which is only 0.8% of the total number of buildings. The remaining subgraphs vary in size (number of segments) as illustrated in Fig. 5a. The logarithmic scale shows



(a) Cumulative frequency



(b) Distribution of inter-building distances

Fig. 5. Size of subgraphs.

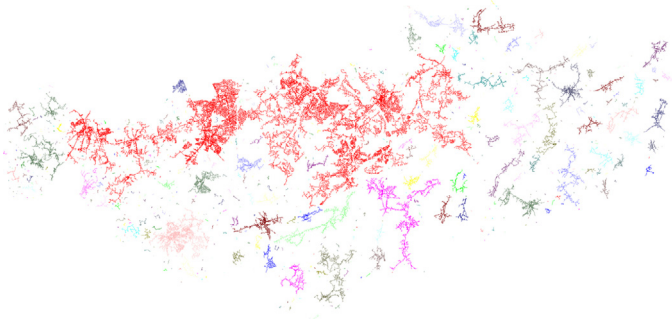


Fig. 6. Subgraphs obtained after a 200 m clump of the MST.

the existence of few very large subgraphs (top 10%) that disrupt the distribution.

The distance separating buildings within each subgraph varies around the mean identified earlier (20 m) (Fig. 5b) but the distribution being slightly skewed to the left. A large majority of subgraphs have a median within the range of 10–20 m. The distribution of the means is flatter than the medians and shows the existence of many subgraphs with larger inter-building distances, between 20 and 40 m, which might indicate the presence of a significant number of subgraphs with rather extensive low density morphologies. Comparatively, there are much less subgraphs with a smaller inter-building distance than the median, indicative of more compact settlements.

The subgraphs are mapped on Fig. 6 (except subgraphs with less than 10 buildings). Each subgraph is coloured differently on Fig. 6. The map visually reveals a variety of structures across the study area. We see that the graphs tend to be smaller and more elongated in the Eastern part. We can also particularly emphasize that the largest subgraph (in red) contains 112 410 segments, i.e. about half of the buildings in the study area. It is actually continuously connected to the urban agglomeration of Brussels city. We clearly see from the map of subgraphs that the contiguous urbanisation expands quite far to the South and West. This is in line with Thomas et al. (2007) who obtained a high fractal dimension for that area, i.e. a rather compact morphology, and with Van Hecke, Halleux, Decroly, and Merenne-Schoumaker (2009) who describe the area as a strong morphological urbanisation. Obtaining a single graph for the Northern area is a rejoinder to these studies but our results complements the picture by showing that the spatial expansion of the contiguous urbanisation goes further away and is no longer systematically related to main transportation axes. This indicates a strong continuation of ribbon development, typical of Belgium (De Decker, 2008; Thomas et al., 2007), and supports theoretical results for which there is no jump over non-developed land (no leapfrog) when transport costs have stronger effects than green amenity effects in the location decision of households (Caruso et al., 2015; Peeters et al., 2015).

3.3. Characterisation of local patterns (Step 6)

Before we analyse the Local Indicator of Spatial Association (LISA) per segment, it is worth reporting the global Moran's I for the entire set of segments. It is found to be significant and positive ($I = 0.193$, $p\text{-value} < 0.0001$) meaning that there is a positive spatial autocorrelation and a tendency to have similar local patterns next to each other.

Table 2 reports LISA categories for all segments, independent of their belonging to a particular subgraph. Half of the segments are located within heterogeneous settlements and a very large portion of them have a non significant local Moran's I (70 116 N.S.). The other half of segments belong to homogeneous settings where we find three times more cases within settlements where buildings are tightly connected, i.e. cluster of small inter-building distances (LL) than within settlements where buildings are loosely connected, i.e. cluster of large inter-building distances (HH).

Within those homogeneous but loosely connected settlements (HH), the inter-building distance is about 4 times longer than within the homogeneous but more compact environments (Table 2). The HH settlements are also more diverse as shown by a flatter and wider distribution (Fig. 7) with the standard deviation being 5 times larger (Table 2).

Fig. 8 shows the local pattern characterisation for the entire study area (8a) and zooms over a selection of three typical zones (8b–d).

Unsurprisingly, although the more tightly connected settlements (blue LL segments) represent the largest share of all segments (35%) they cover a rather limited portion of space compared to the homogeneous but loose settlements (HH). This is obviously the result of shorter inter-building distances, but the spatial distribution is far from being random and very well reflects our understanding of the geography and urbanisation history of the study area. The blue areas detected on the map by our method correspond well to suburban subcentres, small rural towns and early suburban developments, but also to some recently planned allotments, which was less expected.

The settlements with repeated long distances between buildings (red HH segments) represent only 13% of the total segments (buildings), they have a strong spatial imprint, particularly in the central part of the map, at the outskirts of the city of Brussels. Again the red segments identified correspond well to our understanding of the urbanisation process with sparser buildings and villas surrounded by large gardens (larger plots), a process started early in the 20th century by wealthier families looking for leisure homes in a picturesque countryside and continued with the increasing population and commuting to Brussels in a region long characterised by very weak planning restrictions.

Particularly interesting is our ability to discriminate former small towns and villages (blue) from the loosely connected built-up fabric (red) within the largest subgraph discussed above, which is usually considered as a unique bloc of continuous

Table 2
Characteristics of segments per local settlement types.

	Number of segments	Inter-building distance		
		Mean	Median	St. dev
Homogeneous local patterns				
HH	26 411 (13%)	38	31	23
LL	70 815 (35%)	9	8	4
Heterogeneous local patterns				
HL	17 902 (9%)	28	24	13
LH	14 875 (7%)	11	10	5
N.S.	70 116 (35%)	22	18	18
All	200 119 (100%)	19	15	17

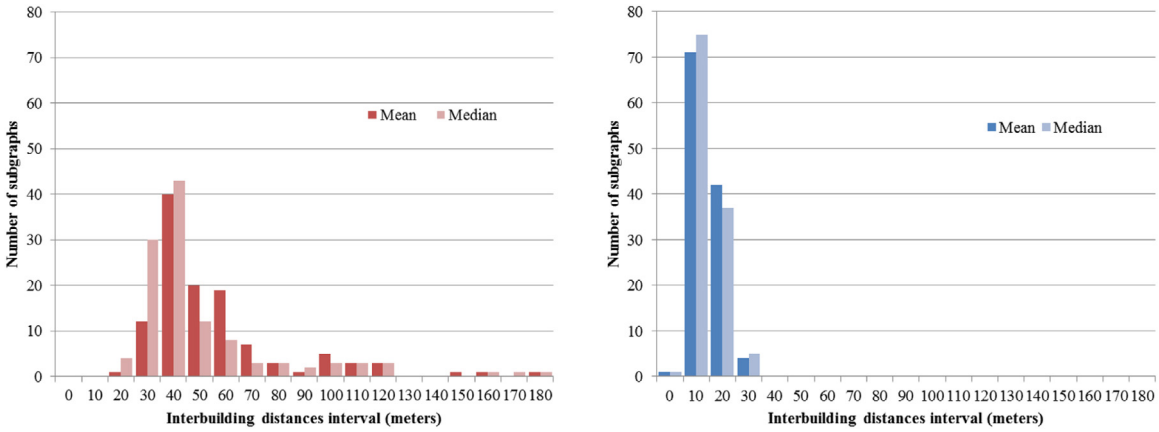


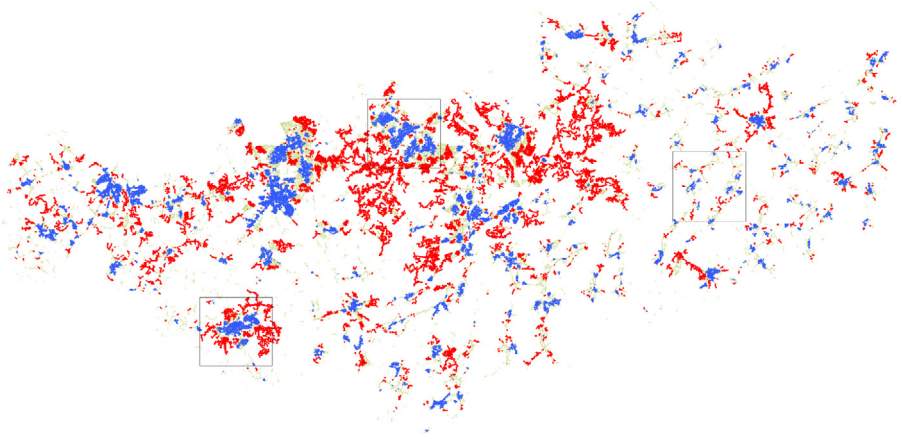
Fig. 7. Distribution of inter-building distances per local settlement types: HH (left) and LL (right).

urbanisation from Brussels (morphological agglomeration in the sense of Thomas, Cotteels, Jones, and Peeters (2012)). Our approach clearly adds up to standard methods that consider morphological continuity only. It actually combines continuity and density criteria, which is similar somehow to fractal analysis where morphologies are described simultaneously with density and lacunarity concepts. Conversely to fractals and density measures, however, we do not need to define an exogenous reference surface (window or spatial unit).

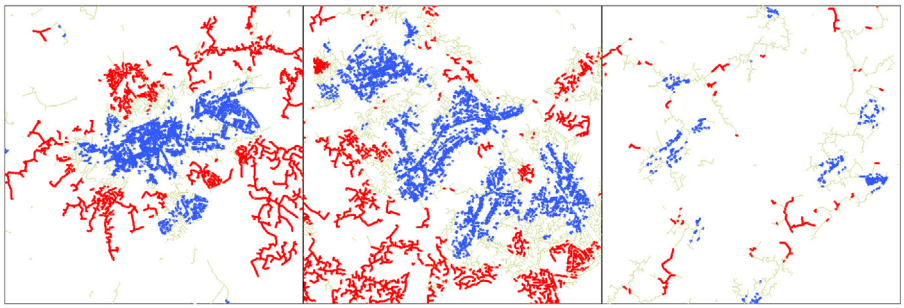
Beyond the centre-periphery urbanisation process from Brussels, a similar process is observed locally around many small centres (Fig. 8b and c). We can even identify (Fig. 8d) different colours and thus processes within linear villages, usually depicted under a single qualitative descriptor ('street-village', 'star-shaped settlement', or ribbon development). This emphasises the ability of our approach to locally isolate fuzzy zones at the border of small towns and villages, and hence different timing and facets of the periurbanisation process.

Finally, we see that heterogeneous patterns (LH, HL, N.S.) are often located between the red and blue areas, corresponding to interstitial morphological transitions or atypical mixed urbanisation forms.

Overall, our results and their spatial distribution confirm the ability of our automatic method to discriminate between different



(a) LISA map



(b) zoom South-West (c) zoom North-Centre (d) zoom East

Fig. 8. Blue: homogeneous cluster of small inter-building distances (LL). Red: homogeneous cluster of large inter-building distances (HH). Grey: heterogeneous patterns (HL, LH or N.S.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

kinds of urbanisation morphologies at a very local level from simple data and despite very few exogenous inputs.

4. Conclusion

We have developed an innovative combination of well known and simple methods inspired from graph theory and local spatial autocorrelation analysis in order to automatically identify different urbanisation morphologies directly from the geographic coordinates of buildings' centroids. This innovative combination is very parsimonious in parameters and results in an efficient, statistically robust, and sound identification of homogeneous and heterogeneous settlements in the Southern periphery of Brussels. Replicating the analysis to other cases would very valuable to assess whether the approach can be discriminant as well in other exurban contexts. At this stage, given our rather heterogeneous study case, the methodology is promising and demonstrates ability to isolate local patterns with fine granularity and capacity to discriminate centre-periphery structures at both regional and very local scales.

Being independent from any reference surface unit, the approach permits an identification and characterisation of groups of building while the positioning of buildings within their non-built environment and connection with other buildings are kept. Hence local settlements can be found with no *ex-ante* decision on their limit and their connection with the non-built environment. The structural form of residential settlements is preserved which we can easily relate to urbanisation history and processes, and later use to assess interactions between the built and non-built environments. Compared to other morphological approaches and density approaches based on administrative units or grids, this is a strong advantage for studying both ecological impacts of urbanisation patterns or for designing urban planning policies. One can for example think of using the mapped output for prioritising urban transport development within or near the more compact settlements but preserving ecological corridors and non-built land for recreational purpose.

Before we can use the methodology to design and recommend planning policies, priority should be given in the next stages to characterising the effects of inter-building distances and different local patterns identified on ecosystem services, on the ecology of species, or in terms of recreational and landscape value for people. In addition to progressing the descriptive analysis of urban forms and offering new measures for planning, there are also methodological improvements perspectives. In particular one can see limitations in the use of an exogenous cut-off, which we could relax in the future so it becomes endogenous, for example based on a local indicator of spatial association (LISA) at a higher scale.

Acknowledgements

This research has been carried out within the project TRUSTEE, funded under the RURAGRI ERA-NET, European Commission 7th Framework Research Programme (Contract 235175).

References

Amenta, N., Bern, M., & Kamvysselis, M. (1998). A new voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th annual conference on computer graphics and interactive techniques* (pp. 415–421).

Anselin, L. (1995). Local indicators of spatial association. *Geographical Analysis*, *27*(2), 93–115.

Anselin, L., & Rey, S. J. (2009, December). *Perspectives on spatial data analysis*. Springer Science & Business Media., google-Books-ID: sfTN7Kht9fEC.

Antrop, M. (2004). Landscape change and the urbanization process in Europe. *Landscape and Urban Planning*, *67*(1-4), 9–26.

Arribas-Bel, D., Nijkamp, P., & Henk, S. (2011). Multidimensional urban sprawl in Europe: A self-organizing map approach. *Computers, Environment and Urban Systems*, *35*, 263–275.

Asano, T., Bhattacharya, B., Keil, M., & Yao, F. (1988). Clustering algorithms based on minimum and maximum spanning trees. In *Proceedings of the 4th annual symposium on computational geometry* (pp. 252–257).

Batty, M., & Longley, P. A. (1994). *Fractal cities: A geometry of form and function*. Academic Press.

Berghauser Pont, M., & Haupt, P. (2007). The relationship between urban form and density. *Urban Morphology*, *11*(1), 62–65.

Boarnet, M. G., & Crane, R. (2001). Travel by design: The influence of urban form on travel. *Spatial Information Systems series*.

Boyce, R. R., & Clark, W. A. (1964). The concept of shape in geography. *Geographical Review*, *54*(4), 561–572.

Brueckner, J. K. (2000). Urban sprawl: Diagnosis and remedies. *International Regional Science Review*, *23*, 160–171.

Caruso, G. (2001). *Periurbanisation, the situation in Europe: A bibliographical note and survey of studies in the Netherlands, Belgium, Great Britain, Germany, Italy and the Nordic countries. Recueil des études et survey. DATAR (Direction de l'Aménagement du Territoire et de l'Action Régionale)*.

Caruso, G., Cavailhès, J., Peeters, D., Thomas, I., Frankhauser, P., & Vuidel, G. (2015). Greener and larger neighbourhoods make cities more sustainable! a 2d urban economics perspective. *Computers, Environment and Urban Systems*, *54*, 82–94.

Caruso, G., Vuidel, G., Cavailhès, J., Frankhauser, P., Peeters, D., & Thomas, I. (2011). Morphological similarities between DBM and a microeconomic model of sprawl. *Journal of Geographical Systems*, *13*(1), 31–48.

Cavailhès, J., Brossard, T., Foltête, J.-C., Hilal, M., Joly, D., Tourneux, F.-P., et al. (2009). GIS-based hedonic pricing of landscape. *Environmental and Resource Economics*, *44*(4), 571–590. http://dx.doi.org/10.1007/s10640-009-9302-8

Clifton, K., Ewing, R., Knaap, G.-J., & Song, Y. (2008, March). Quantitative analysis of urban form: A multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, *1*(1), 17–45. http://dx.doi.org/10.1080/17549170801903496

De Decker, P. (2008). Facets of housing and housing policies in Belgium. *Journal of Housing and the Built Environment*, *23*(3), 155–171.

Deng, M., Liu, Q., Cheng, T., & Shi, Y. (2011). An adaptive spatial clustering algorithm based on Delaunay triangulation. *Computers, Environment and Urban Systems*, *35*, 320–332.

Donnay, J., Barnsley, M., & Longley, P. (2003). *Remote sensing and urban analysis: GISDATA 9*. Taylor and Francis.

Estivill-Castro, V., & Lee, I. (2002). Argument free clustering for large spatial point-data sets. *Computers, Environment and Urban Systems*, *26*, 315–334.

Ewing, R., & Hamidi, S. (2015). Compactness versus sprawl: A review of recent evidence from the united states. *Journal of Planning Literature*, 1–20.

Ewing, R., Richardson, H. W., Bartholomew, K., Nelson, A. C., & Bae, C.-H. C. (2014). *Compactness vs sprawl revisited: Converging views*. CESifo Working Paper, No. 4571.

Ewing, R. H. (2008). Characteristics, causes, and effects of sprawl: A literature review. In J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, U. Simon, & C. ZumBrunnen (Eds.), *Urban Ecology* (pp. 519–535). US: Springer.

Ewing, R. H., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, *76*(3), 265–294.

Fall, A., Fortin, M.-J., Manseau, M., & O'Brien, D. (2007). Spatial graphs: Principles and applications for habitat connectivity. *Ecosystems*, *10*(3), 448–461. http://dx.doi.org/10.1007/s10021-007-9038-7

Fan, C., & Myint, S. (2014). A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning*, *121*, 117–128. http://www.sciencedirect.com/science/article/pii/S0169204613001953

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: A synthesis. *Global Ecology and Biogeography*, *16*(3), 265–280. http://dx.doi.org/10.1111/j.1466-8238.2007.00287.x

Foltête, J.-C., Clauzel, C., & Vuidel, G. (2012). A software tool dedicated to the modelling of landscape networks. *Environmental Modelling and Software*, *38*(0), 316–327. http://www.sciencedirect.com/science/article/pii/S1364815212001995

Foltête, J.-C., Clauzel, C., Vuidel, G., & Tournant, P. (2012). Integrating graph-based connectivity metrics into species distribution models. *Landscape Ecology*, *27*(4), 557–569.

Frankhauser, P. (1998). Fractal geometry of urban patterns and their morphogenesis. *Discrete Dynamics in Nature and Society*, *2*, 127–145.

Frenkel, A., & Ashkenazi, M. (2008). The integrated sprawl index: Measuring the urban landscape in Israel. *Annals of Regional Science*, *42*(99-121).

Galster, G. C., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling sprawl to the ground: Defining and measuring an elusive concept. *Housing Policy Debate*, *12*, 681–717.

Geoghegan, J. (2002). The value of open spaces in residential land use. *Land Use Policy*, *19*(1), 91–98.

Graham, R. L., & Hell, P. (1985). On the history of the minimum spanning tree problem. *Annals of the History of Computing*, *7*(1), 43–57.

Griffiths, S., Jones, C. E., Vaughan, L., & Haklay, M. (2010). The persistence of suburban centres in greater London: Combining conzenian and space syntax approaches. *Urban Morphology*, *14*, 85–99.

Grygorash, O., Zhou, Y., & Jorgensen, Z. (2006). Minimum spanning tree based clustering algorithms. In *Tools with Artificial Intelligence. ICTAI 06* (pp. 73–81).

Hall, P. (1973). *The containment of urban England*. London: PEP.

Hamidi, S., & Ewing, R. (2014). A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landscape and Urban Planning*,

128, 72–82. http://www.sciencedirect.com/science/article/pii/S0169204614001133

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*, *23*(August (2, Suppl. 1)), 64–73. http://www.sciencedirect.com/science/article/pii/S0749379702004750

Harvey, R. O., & Clark, W. A. (1965). The nature and economics of urban sprawl. *Land Economics*, *41*(1), 1–9. URL http://www.jstor.org/stable/3144884.

Hasse, J., & Lathrop, R. G. (2003). A housing-unit-level approach to characterizing residential sprawl. *Photogrammetric Engineering and Remote Sensing*, *69*(9), 1021–1030.

Hillier, B. (1996). *Space is the machine*. Cambridge: Cambridge University Press.

Irwin, E., & Bockstael, N. (2007). The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proceedings of the National Academy of Science of the United States of America*, *104*(52), 20672–20677.

Jaeger, J. A. (2000). Landscape division, splitting index, and effective mesh size: New measures of landscape fragmentation. *Landscape Ecology*, *15*(2), 115–130.

Jaeger, J. A., Soukup, T., Madrinan, L. F., Schwick, C., Kienast, F., et al. (2011). Landscape fragmentation in Europe. In *Tech. Rep. 2, Joint EEA-FOEN report*.

Lowry, J. H., & Lowry, M. B. (2014). Comparing spatial metrics that quantify urban form. *Computers, Environment and Urban Systems*, *44*, 59–67.

Mace, A., Blanc, F., Gordon, I., & Scanlon, K. (2016). *A 21st century metropolitan green belt*. Tech. rep., London School of Economics and Political Science.

March, L., & Steadman, P. (1974). *The geometry of environment: An introduction to spatial organization in design*. Cambridge, MA: MIT Press.

McGarigal, K., & Marks, B. J. (1995). *Spatial pattern analysis program for quantifying landscape structure*. Gen. Tech. Rep. PNW-GTR-351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Nordbeck, S. (1971). Urban allometric growth. *Geografiska Annaler Series B: Human Geography*, *53*, 54–67.

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. (1999). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Chichester: Wiley.

Okabe, A., & Funamoto, S. (2000). An exploratory method for detecting multi-level clumps in the distribution of points – a computational tool, VCM (variable clumping method). *Journal of Geographical Systems*, *2*, 111–120.

Okabe, A., Okunuki, K., & Shiode, S. (2006). The SANET toolbox: New methods for network spatial analysis. *Transactions in GIS*, *10*(4), 535–550.

Okabe, A., & Yoshikawa, T. (2003). Sainf: A toolbox for analyzing the effect of point-like, line-like and polygon-like infrastructural features on the distribution of point-like non-infrastructural features. *Journal of Geographical Systems*, *5*, 407–413.

Pauly, M., Gross, M., & Kobbelt, L. P. (2002). Efficient simplification of point-sampled surfaces. In *Proceedings of the conference on Visualization. IEEE Computer Society* (pp. 163–170). October.

Peeters, D., Caruso, G., Cavailhès, J., Thomas, I., Frankhauser, P., & Vuidel, G. (2015). Emergence of leapfrogging from residential choice with endogenous green space: Analytical results. *Journal of Regional Science*, *55*(3), 491–512.

Peeters, D., & Thomas, I. (2009). Network autocorrelation. *Geographical Analysis*, *41*(4), 436–443.

Peter, S. J. (2011). A novel dynamic minimum spanning tree based clustering method for image mining. *Journal of Discrete Mathematical Sciences and Cryptography*, *14*(5), 405–419.

Samozino, M., Alexa, M., Alliez, P., & Yvinec, M. (2006). Reconstruction with voronoi centered radial basis functions. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing (SGP)* (pp. 51–60).

Schindler, M., & Caruso, G. (2014). Urban compactness and the trade-off between air pollution emission and exposure: Lessons from a spatially explicit theoretical model. *Computers, Environment and Urban Systems*, *45*, 13–23.

Song, Y., & Knaap, G.-J. (2004). Measuring urban form is Portland winning the war on sprawl? *Journal of the American Planning Association*, *70*(2), 210–225.

Stanilov, K. (2010). Bridging the gap between urban morphology and urban modelling. *Urban Morphology*, *14*, 123–124.

Tannier, C., Foltête, J.-C., & Girardet, X. (2012). Assessing the capacity of different urban forms to preserve the connectivity of ecological habitats. *Landscape and Urban Planning*, *105*(1), 128–139.

Thomas, I., Cotteels, C., Jones, J., & Peeters, D. (2012). *Revisiting the extension of the Brussels urban agglomeration: New methods, new data. . . new results? e-BELGEO*. http://belgeo.revues.org/6074

Thomas, I., Frankhauser, P., & Biernacki, C. (2008). The morphology of built-up landscapes in Wallonia (Belgium): A classification using fractal indices. *Landscape and Urban Planning*, *84*(2), 99–115.

Thomas, I., Frankhauser, P., & De Keersmaecker, M.-L. (2007). Fractal dimension versus density of built-up surfaces in the periphery of Brussels. *Papers in Regional Science*, *86*(2), 287–308.

Torrens, P., & Alberti, M. (2001). *Measuring sprawl*. Working Paper 27, Centre for Advanced Spatial Analysis, University College of London.

Turner, M. G., & Gardner, R. H. (1991). *Quantitative methods in landscape ecology: An introduction*. Springer-Verlag.

Urban, D., & Keitt, T. (2001). Landscape connectivity: A graph-theoretic perspective. *Ecology*, *82*(5), 1205–1218.

Van Hecke, E., Halleux, J., Decroly, J., & Merenne-Schoumaker, B. (2009). *Noyaux d'habitats et regions urbaines dans une Belgique urbanisée*. Bruxelles: Monographie de l'enquête socio-economique générale 2001.

Wade, T. G., Wickham, J. D., Zacarelli, N., & Riitters, K. H. (2009). A multi-scale method of mapping urban influence. *Environmental Modelling and Software*, *24*(10), 1252–1256. http://www.sciencedirect.com/science/article/pii/S1364815209000826

Whitehand, J. W. R. (1992). Recent advances in urban morphology. *Urban Studies*, *29*, 617–634.

Wolman, H., Galster, G., Hanson, R., Ratcliffe, M., Furdell, K., & Sarzynski, A. (2005). The fundamental challenge in measuring sprawl: Which land should be considered? *The Professional Geographer*, *57*(1), 94–105.

Wu, B. Y., & Chao, K.-M. (2004). *Spanning trees and optimization problems*. Chapman and Hall/CRC. http://dx.doi.org/10.1201/9780203497289

Ye, Y., & Van Nes, A. (2014). Quantitative tools in urban morphology: Combining space syntax, spacematrix and mixed-use index in a GIS framework. *Urban Morphology*, *18*(2), 97–118.

Zahn, C. T. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, *100*(1), 68–86.