Spatial Signatures
Understanding (urban) spaces through form and function

July 2021

ABSTRACT: This paper presents the notion of spatial signatures as a characterisation of space based on form and function designed to understand urban environments. How we spatially arrange the building blocks that make up a city matters. On the one hand, it encodes many aspects of the phenomena that created such an arrangement in the first place. On the other, once in place, this arrangement of urban form and function underpins many outcomes, from economic productivity to environmental sustainability. Our approach unfolds in three main stages. First, we propose a new spatial unit –the Enclosed Tessellation (ET) cell– to delineate space in a way that is exhaustive and matches the underlying processes at which urban form and function operate. Second, we propose to attach a large variety of form and function-based characters to ET cells to describe each of these units. Third, to build spatial signatures, information on ET cells can be clustered using unsupervised learning techniques. This process results in a theory-informed, data-driven typology of space that follows form and function. We illustrate this approach by applying it to a sample of five very different cities scattered across the world. Our results demonstrate the ability to successfully differentiate different parts of a city that were built at different points in time and under different technological regimes, but also highlight broader comparisons about the nature of urban fabric in different regions. Our contribution resides in leveraging modern data, technology and methods to propose a detailed, consistent and scalable methodology that characterises urban form and function. The spatial signatures can be used across academic disciplines and by a variety of practitioners and policymakers supporting initiatives such as the Sustainable Development Goals.

Key words: Geographic Data Science, Urban Form, Urban Function

1. Introduction

How we spatially arrange the building blocks that make up a city matters. The map of many European cities tells the story of the different historical periods in which they were born, grew and, in some cases, contracted. American cities which "came of age" in the second half of the XXth Century would look very different had the automobile not been the defining technology of the time. And the stark contrasts between luxury developments and informal settlements that can be observed across many cities of the Global South are a reflection of the wide range of disparities and inequalities that those societies display. This encoding is also sticky. Although cities are constantly in flux, new innovations and waves of change rarely start from scratch. More commonly, they are added in a layered way. Over time, each phase, each change blends with the rest of the urban fabric to give a city its uniquely distinct pattern that defines it almost as a strand of DNA. The building blocks include the different elements of the built and natural environments of which cities are composed, but also the purpose they serve. Understanding the former thus requires us to consider urban *form*, while grasping the latter invites us to examine its *function*. Urban form and function are relevant for two main reasons. First, their fabric encodes the socio-economic history, technology and values of the society that has built them. Studying the nature and distribution of form and function in cities thus helps us better understand the societies that, over time, have shaped them. The second reason why urban form and function are important is that they are not only a history book recording the past, but also play an important role in defining the present and shaping the future. Once in place, their features and structure have direct implications for a wide range of outcomes, from productivity and job access to social inclusion and mobility, deprivation, service provision, energy consumption or carbon emissions, to name just a few.

However, the study of urban form and function is deeply fragmented. Work focusing on the form and function of cities is scattered across different academic disciplines and policy-making scenes. This is not necessarily a problem in itself since different backgrounds provide a richer picture. And there is much to be gained from a plurality of perspectives. It does however mean that the evidence available presents different interests as well as varying degrees of detail, consistency, and coverage. It is understandable that economists develop conceptualisations shaped around economic theories, while geographers do so, paying attention to spatial scales, and yet other disciplines bring different aspects to focus. Similarly, decision-makers interested in understanding aspects of urban form and function tend to see it through the lenses provided by the vantage point they occupy. Regional planners may try to obtain as much detail as possible for a relatively small geographical area; while supra-national organisations may prioritise scale and coverage at the expense of detail.

There is a clear need for detailed, consistent, and scalable evidence on urban form and function. Detailed and granular measurement that can be performed across large geographical extents in a comparable way can unlock insights that get lost when we can only observe certain regions of the picture developed for a large extent does not have sufficient detail. This is because many of the theoretical underpinnings of urban form and function that reflect its history and influence present and future outcomes tend to operate at fine scales but, to be able to observe meaningful differences, we need to consider many and different places. For example, the characters that define Medieval city centres in Europe quickly blur when the geographical unit considered is coarse. But, to be able to examine how these characters relate to different levels of walkability, or even of gas emissions, we require a large extent to set up meaningful comparisons.

Of detail, consistency and scale, the current research landscape described above provides, at best, any two. There is an abundance of detailed studies that consistently measure form and function, but these are in their majority confined to case studies with very limited geographical extents. On the other end of the scale spectrum, recent years have seen the appearance of work at a global scale that is consistent. However, their degree of detail tends to be hindered by data limitations. Finally, one could understand the multitude of detailed case studies in conjunction as a growing body of evidence that is able to reach a sizeable scale. But, in these case, the fragmentation discussed earlier often translates in a lack of consistency that prevents meaningful comparisons.

These limitations are beginning to be overcome thanks to recent advances on the data, technological and methodological fronts. New forms of data such as open cadastres, consumer datasets derived from modern business operations, or high resolution, public satellite imagery are greatly improving the descriptions we can build of cities (Arribas-Bel, 2014). Progressively, we are able to build denser and more up to date representations of urban environments at a cheaper cost. Technological developments such as the dramatic increase of computational power available to researchers, or improvements in computer algorithms and machine learning are lowering the entry barrier to advances that only a few years ago required a high degree of dedicated expertise to be able to benefit from. Perhaps more importantly, recent methodological contributions such as morphometrics (Dibble, 2016) or geographic data science (Singleton and Arribas-Bel, 2021) are paving the way to blend these advances with domain knowledge and urban theory, effectively enabling disciplines concerned with the form and function of cities to benefit from such developments.

In this paper, we introduce the spatial signatures as a characterisation of space based on form and function designed to understand urban environments. The spatial signatures are thus a delineation that divides geographical space based on its appearance (form) and how it is used (function). It is not a classification of space as much as a way of thinking around classifying space based on form and function. Hence, it is an ideal candidate for deployment on a wide range of data landscapes and geographical regions, as we illustrate in Section 4.

The spatial signatures hold important potential both for urban academics and policy-makers. From an academic point of view, they are relevant as a goal in themselves that allows us to better measure and study the spatial configuration of the building blocks that make up cities. But they also represent a platform on which other disciplines can build on to embed form and function on a variety of socio-environmental outcomes. For policy-makers, the spatial signatures provide a framework for detailed spatial understanding of the cities and territories their decisions affect. Because of the flexibility of the concept, they are useful both in the global north, where cities are constantly recast and retrofitted, as well as the global south, where most of the new urbanisation is currently taking place. In summary, the spatial signatures allow us to move forward in realising detailed, consistent, and scalable measurement of form and function in cities.

The remaining of the paper is structured as follows. Section 2 reviews existing literature on urban form and function, highlighting current gaps. Section 3 details our proposal of spatial signatures, including how we define conceptualise them, the spatial unit we develop to measure them –the enclosed tessellation cell–, and the embedding of form and function into such unit. In Section 4, we illustrate the flexibility of the spatial signatures by presenting an application to five rather different global cities. And we conclude in Section 5 with some reflection about the value and potential of our approach.

2. (Urban) form and function

2.1 *Form*

Urban form approaches environments from the perspective of their physical structure and appearance. Research studying urban form has a long tradition, dating back to the early XXth Century (Geddes, 1915, Trewartha, 1934). Urban morphology, subsequently, begun in the 1960s as an independent area of research. The field originated in parallel within geography (Conzen, 1960) and architecture (Muratori, 1959), reflecting its inherently multi-disciplinary nature, later reinforced by the inclusion of socio-economic elements, as in the work of Panerai et al. (1997). The original methods are predominantly qualitative, a tendency that persists today (Dibble, 2016). The first notable quantitative approaches date to the late 1980s and 1990s, reflecting advancements in computation and newly available data capturing the built environment. In this context, two strains of research have emerged. One focuses on cartographic (vector) representation of the urban environment, assessing its boundaries (Batty and Longley, 1987), street networks (Hillier, 1996, Porta et al., 2006) and other elements (Pivo, 1993). The second one is based on earth observation, exploiting remotely sensed data to capture change in the footprint of urban areas (Howarth and Boasson, 1983).

The current state of the art still retains this distinction between cartographic and remotely sensed approaches. A modern quantitative branch of urban morphology, or urban morphometrics, has emerged working predominantly discrete elements of urban form, and proposing an abundant selection of measurable characters that describe different aspects of form (Fleischmann et al., 2020b). As part of this trend, methods focusing on a single aspect (Porta et al., 2006) have been replaced by efforts to better reflect the complexity of urban form through the combination of multiple morphometric characters into a single model, often leading to data-driven typologies (Song and Knaap, 2007). This focus on classification is becoming more prominent, fueled by the possibilities afforded by new datasets increasingly available. Indeed, the literature is now able to produce typologies that start from small-scale studies focused on blocks and streets (Gil et al., 2012), and zoom out into larger areas with higher granularity (Schirmer and Axhausen, 2015, Araldi and Fusco, 2019, Bobkova et al., 2019, Dibble et al., 2019, Jochem et al., 2020).

Advances in remote sensing have also led to a range of classification frameworks based on various conceptualizations of the urban fabric. However, there is one significant difference between classification derived via morphometric characterization and the one based on remote sensing. Where the former is mostly unsupervised (Araldi and Fusco, 2019, Schirmer and Axhausen, 2015), exploiting the hidden structure in the data to develop organically the typology; the latter tends towards supervised techniques, relying on classes defined a priori (Pauleit and Duhme, 2000). Two emerging classification models used to inform these exercises are Local Climate Zones (Stewart and Oke, 2012), defining ten built-form types and seven land cover types, and used recently by Koc et al. (2017) or Taubenböck et al. (2020); and the Urban Structural Type, a generic typology based on the notion of internal homogeneity of types (Lehner and Blaschke, 2019).

2.2 *Function*

Urban function considers environments based on the activities that take place within them. The focus is thus not on what a space "looks like", as it is the case on urban form, but on "what it is used for". What activities occur within cities, how they are spatially configured, and how they relate to each other are key questions in this context. To the extent cities compress space and time to concentrate human activity of very diverse nature, the study of function is relevant to a variety of fields and is undertaken by a wider constituency of researchers. Disciplines as disparate as geography, economics or environmental sciences have contributed in their own way to our understanding of urban function. Furthermore, because function has direct implications for a wide range of social and environmental processes at different geographic scales, their study also falls within the realm of policy. Given the breadth of perspectives and goals, a complete overview of its contributions is beyond the scope of this paper. Instead, here we highlight what we consider the most relevant domains involved: environmental sciences, urban and public economics, urban and transport geography, planning, and sociology.

Environmental sciences have long considered urban function in the context of the broader interest on understanding the natural characteristics of the surface of the Earth. An area that has attracted much effort relates to the development of classifications of land cover and land use, the former describing the nature of surfaces while the latter focusing on how those surfaces are used. Several land cover classifications are available (e.g. CORINE, European Environment Agency, 1990, in Europe; the National Land Cover Database, Homer et al., 2012, in the US; or the Land Cover CCI, Defourny et al., 2012, globally), as well as some for land use (e.g. the Urban Atlas project, Copernicus Land Monitoring Service, 2021). While much of this research is not focused on urban environments, the urban remote sensing community (Weng and Quattrochi, 2018) is building a more explicit bridge between these approaches and the study of cities (e.g. Kuffer et al., 2016, Georganos et al., 2018, Jochem et al., 2018, Prasad, 2015, Stark et al., 2020).

A wide array of disciplines have developed more specific interests in urban function. Sustainability studies, for example, are interested in how function is configured within and across cities in so far as it relates to the level of emissions (Angel et al., 2018) or energy consumption (Silva et al., 2017). The social sciences have a long-standing interest on the spatial configuration of form because it affects several outcomes of prime interest. Depending on the nature of these outcomes, form is conceptualised in one or another way. Urban economics pays special attention to density of economic activity and, by extension, of population (Ahlfeldt and Pietrostefani, 2019, Duranton and Puga, 2020), since density is intimately related to theories of agglomeration, one of the intellectual pillars of the field. Public economics has paid attention the configuration of urban function to the extent that it determines the efficiency of certain public services provided by local governments (Carruthers and Ulfarsson, 2003, Hortas-Rico and Solé-Ollé, 2010). Sociologists and planners have also found that different spatial configurations of function over space is associated with different degrees of social mobility (Ewing et al., 2016) or socio-economic deprivation (Venerandi et al., 2018). More generally, transport researchers have built a robust body of knowledge linking urban function and its spatial distribution to travel behaviour (Boarnet et al., 2001), sustainability (Sevtuk and Amindarbari, 2020), or accessibility to jobs (Horner, 2004) and amenities (Diamond and Tolley, 2013), with clear implications for socio-economic disparities.

2.3 *Blending Form & Function*

Whilst much of the literature focuses either on form or function, the two are deeply interconnected. Function develops in the context provided by form; and, over time, form adapts and encodes function. However, there exists a few attempts to classify urban spaces considering both jointly. Bourdic et al. (2012) propose a comprehensive classification based on indicators ranging from form to biodiversity, culture and energy on a scale of individual cities. Several studies consider the link between form and land use (Song and Knaap, 2007, Song et al., 2013, Bourdic et al., 2012), with some authors even including land use a component of form characterisations (Dibble et al., 2019). At any rate, even when the two are combined, the scope of either, particularly function, is narrow rather than all-encompassing. For example, the geodemographic tradition (Harris et al., 2005, Webber and Burrows, 2018) studies populations based on where they live. Although this considers both, the focus is very much on the residential function, leaving aside others such as employment or amenities. Recent years have also seen work at the global scale connecting form and population density (Ewing et al., 2002, Zheng et al., 2014, OECD, 2018), many facilitated by the appearance of new datasets (e.g. Pesaresi et al., 2019, Sorichetta et al., 2015), alongside studies embedding accessibility and proximity to points of interests into their frameworks (Alexiou et al., 2016, Venerandi et al., 2019). Nevertheless, the body of research directly working with both form and function in a single framework is limited and tends to focus on particular functions.

3. Spatial Signatures

Despite the current sparsity of studies, we believe there are several benefits in considering form and function in tandem when trying to understand urban spaces. The two are deeply interconnected. This close correlation implies that outcomes observed across form tend to hold true for function, and viceversa. However, unique patterns emerge when particular types of form and function come together. We argue that it is only through the combination of form and function that cities are able to encode and reflect sophisticated aspects of human nature such as history, culture or technology. In these cases, considering only one or the other hinders rather than enables, as we risk missing important traits of the nature of a place. From a more empirical perspective, even when the two dimensions mostly overlap, there is value in considering them jointly. Some aspects of form and function are clear conceptually but challenging to measure. Broadening the pool of indices that can be deployed ensures better accuracy when characterising existing patterns on the ground. In this section, we detail our proposal to understand urban form and function through what we term "spatial signatures".

3.1 Definition

We propose the notion of *spatial signatures* as:

A characterisation of space based on form and function designed to understand urban environments

Spatial signatures provide exhaustive coverage for an area of interest by drawing organic boundaries that delineate portions of consistent morphological and functional characteristics. We will refer to a single *spatial signature* in two related but distinct ways: first, as one of the multiple classes that make up a wider typology of spatial signatures; and second, as a geographical instance of that class, a contiguous portion of territory that shares those morphological and functional traits. As such, spatial signatures can be seen as organically grown delineations that organise space into urban and rural, orderly and irregular, formal and informal. Laid out together, they can be used to explore urban extents, to parse through the complexity of their spatial structure, or to understand the evolution of cities. In bringing together both form and function, with a focus on the urban, spatial signatures provide a nexus between purely morphological characterisations and those entirely based on function. To the extent form and function are intrinsically connected, its combination leads to more robust portraits of the space that makes up cities. And, since the focus is on the urban, spatial signatures provide a complementary perspective to most land cover and use classifications, which historically pay more attention to the portion of space not occupied by cities.

The development of the spatial signatures approach carries several benefits to studies of cities and their footprint. The concept is data-driven but theoretically informed; granular but scalable; and flexible enough to be adapted to a wide variety of applied contexts, from data-rich to those with limited availability. Spatial signatures embed theoretical ideas about how cities are spatially arranged, how this configuration can be best conceptualised, and how it is perceived by humans into a data-driven framework that connects them to the vast amount of empirical information available representing the world. These theoretical underpinnings are sourced from a variety of disciplines, from architecture to environmental scienes, and thus are inherently interdisciplinary. The spatial signatures thus provide a shared vocabulary to bring together a variety of scholars and policy makers for whom form and function in cities is relevant, either as their object of study or as an input for their own domains of expertise.

Part of the flexibility of this approach stems from the fact it represents a way of thinking about form and function in cities as much as a set of techniques to parse through data. In the following two subsections, we cover the two core components of building spatial signatures: the delineation of atomic units that can be organically grown to delineate boundaries between signatures; and the approach to embed form and function into each of those units in a way that the aggregation is feasible.

3.2 Building blocks: the Enclosed Tesselation

This section proposes a novel and theoretically-informed delineation of space to support the development of spatial signatures. Since spatial signatures are conceptualised as highly granular in space, considering the ideal unit of analysis at which to measure them is of utmost importance. This step is worth spending enery and effort for two main reasons. First, if ignored, there is an important risk of incurring the modifiable areal unit problem (MAUP, Openshaw, 1981). The urban fabric is not a spatially smooth phenomenon; rather, it is lumpy, irregular and operates at very small scales. Choosing a spatial unit that does not closely match its distribution will subsume interesting variation and will hide features that are at the very heart of what we are trying to capture with spatial signatures. Second, and conversely, we see adopting a meaningful unit a step of analysis itself. Rather than selecting an imperfect but existing unit to try to characterise spatial signatures, delineating our own is an opportunity in itself to learn about the nature of urban tissue and better understand issues about distribution and composition within urban areas.

Let us first focus on what is required from an ideal unit of analysis for spatial signatures. We need a partition of space into sections of built *and* lived environment that can later be pieced together based on their characteristics. The result will feed into an organic delineation that captures variation in the appearance and character of urban fabric as it unfolds over space. To be more specific, a successful candidate for this task will need to fulfill at least three features: indivisibility, internal consistency, and exhaustivity. An ideal unit will need to be *indivisable* in the sense that if it were to be broken into smaller components, none of them would be enough to capture the notion of spatial signature. Similarly, every unit needs to be *internally consistent*: one and only one type of signature should be represented in each observation. Finally, the resulting delineation needs to be geographically *exhaustive*. In other words, it should assign every location within the area of interest (e.g. a region or a country) to one and only one class.

The existing literature does not appear to have a satisfying candidate to act as the building block of spatial signatures. Without attempting an exhaustive review, an endeavour beyond the scope of this article, the vast majority of existing approaches to delineate meaningful units of urban form and function fall within one of the following three categories. The first group relies on *administrative* units such as postcodes, census geographies or municipal boundaries (e.g. Taubenböck et al., 2020). These are practical as they usually are readily available. However, their partition of space is usually driven by different needs that rarely align with the measurement of spatial signatures, or indeed those of any morphological or functional urban process. Taubenböck et al. (2019) even argue that "administrative units obscure morphologic reality". An emerging body of work relies on granular, *uniform grids* as the main unit of analysis (e.g. Jochem et al., 2020). This choice is usually explicitly or implicitly motivated by the lack of a better, bespoke partitioning; the use of input data distributed in grids (e.g. satellite imagery); and the assumption that, with enough resolution, grids can be organically aggregated into units that match the processes of interest. A third approach followed mostly by the literature on urban

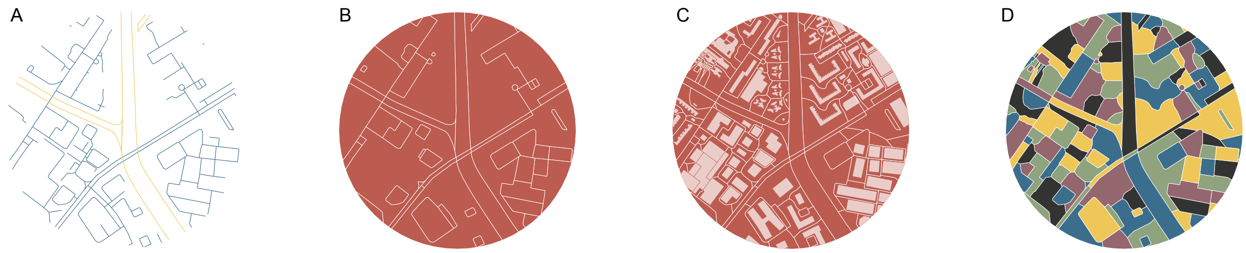

Figure 1: Diagram illustrating the sequential steps leading to the delineation of enclosed tessellaion. From a series of enclosing components, where blue are streets and yellow river banks (A), to enclosures (B), incorporation of buildings as anchors (C) to final tessellation cells (D).

morphology relies on the definition of morphometric units. These include street segments (Araldi and Fusco, 2019), plots (Bobkova et al., 2019), building footprints (Schirmer and Axhausen, 2015), or constructs such as the sactuary area (Mehaffy et al., 2010, Dibble et al., 2019). In all these cases, the choice is justified by the particular application in which it takes place. However none of these approaches meet the three characteristics we require for spatial signatures. Administrative boundaries are exhaustive but rarely indivisable or consistent when it comes to urban form, usually grouping very different types of fabric within a single area. Uniform grids are also exhaustive but, similarly to administrative definition, the arbitrariety of their delineation with respect to urban form usually leaves them divisable and internatlly inconsistent. Morphometric units are the most theoretically appealing since they are built to match the distribution of urban features and are usually granular enough to warrant internal consistency and indivisibility. Most of them are however not exhaustive as they are anchored to particular elements of the build environment, such as streets or building footprints, which do not provide full coverage. Plots would theoretically meet all characteristics but can be problematic due to their variable definition leading to different geometric representations (Kropf, 2018).

We propose the development of a new spatial unit that we term the *enclosed tessellation cell* (EC). An EC is defined as:

The portion of space that results from growing a morphological tesselation within an enclosure delineated by a series of natural or built barriers identified from the literature on urban form, function and perception.

Let us unpack this concept a bit further. ECs result from the combination of three sequential steps (Figure 1).

First, they rely on a set of enclosing components: features of the landscape that divide it in smaller, fully delimited portions. The list of what should be counted as enclosing is informed by theory and, as we will see below, may vary by context. But, as an illustration, it includes elements such as the street network, rivers and coastlines, or railways. Second, these enclosing features are integrated into a single set of boundaries that partition the geography into smaller areas. In some cases, they will be small, as with urban blocks in dense city centres; in others, they will be larger in size, as in rural sections with lower density of enclosing features. We call each of this fully delimited areas an enclosure. Third, enclosures are further subdivided using a morphological tesselation (Fleischmann et al., 2020a) that exhaustively partitions space based on a set of building footprints, which are used in this context as anchors to draw catchment polygons. This process generates geographical boundaries for a given area that result in a new spatial unit. This unit provides full geographical coverage without any overlap. Since the essence of the approach resides in growing a tessellation inside a set of enclosing features, we call the resulting areas "enclosed tesselation cells".

The enclosed tessellation (ET) intersects two perspectives of how space can be understood and organised. The first relies on the use of features that *delimit* the landscape and partition it into smaller, fully enclosed portions. These include the road and street networks, but also others such as railways or rivers. Each feature is conceptualised as a line that acts as a boundary, dividing space into what falls within each of its sides. A long tradition in the literature on urban perception relies on variations of these delimiters. Prominent early examples include the edges and paths highlighted by Lynch (1960) as two of the five core elements that define legibility and imageability of a city; as well as the later work inspired by this framework (e.g. Filomena et al., 2019).

The second perspective that ET integrate is a vision organised around *anchors*. In this view, space arises in-between a discrete set of relevant features. Unlike delimiters, these elements do not partition space per se, but instead act as origins to which the rest can be "attached". The choice of anchors might vary by context but, in this case, the literature on morphometrics has extensive evidence to support the use of buildings as the primary feature (Hamaina et al., 2012, Usui and Asami, 2013, Schirmer and Axhausen, 2015).

The combination of delimiters and anchors as the parsers of space make ECs an ideal spatial unit to study spatial signatures, one which meets the three requirements we outlined above. They are indivisable in that a single EC will contain no delimiters, at most a single anchor, and potentially none. They are also internally consistent because they are delineated as the area within the delimiters that contain at most one anchor. And finally ECs are exhaustive in that every location within the area of interest is assigned to one and only one EC, providing full geographical coverage without any overlap.

3.3 *Embedding form and function into spatial signatures*

This section covers how to develop spatial signatures once ET cells are delineated for an area of study. ET cells take the role of the structural unit. In themselves, they hold descriptive value in reflecting the configuration of the urban environment. They also operate as a container, into which other morphometric and functional characters can be embedded. The next stage to develop spatial signatures is to build form and function characters on top of ET cells. With this, we aim to describe both the intrinsic traits of each cell depending on its own geometry and nature, but to also include features of its immediate spatial context. To do this, we propose to collect a set of descriptors reflecting both form and function to capture the essence and definition of spatial signatures. This process will lead to a heterogeneous mix of morphometric characters, capturing patterns of physical, built-up environment; and functional characters, reflecting population, amenities, land use classification or historical importance. It is to be noted that every piece of information is considered within its spatial context. It is not the role of this section to provide a comprehensive list of all characters, morphometric and functional, that would need to be derived. Such list will depend on the specific context in which a spatial signatures classification is being developed including, for example, data availability or nature of the geographical area being considered. However, it should always aim to reflect the nature of the form and function of each place in as exhaustive a way as possible.

Collecting characters at the ET cell level is only half the task to develop spatial signatures. Given the granularity and multi-dimensionality of the information at this stage, we need to combine it in a way that retains its core characteristics but is easier to parse through. We propose a feasible aggregation of ET cells into spatial signatures using unsupervised learning. Again, it is not the role of this section to single out a technique, since many exist including K-Means, gaussian mixture models, or self-organizing maps (Kohonen, 1990), to name a few. We note there is no need to impose a spatial contiguity constraint as spatially contiguous clusters of cells in the same signature will emerge thanks to the inherent spatial autocorrelation of data derived from mutually overlapping *contexts*. These continuous groups of cells grouped in the same cluster is what we call instances of a spatial signature.

4. Illustration

The classification of form and function into spatial signatures is a conceptual framework and as such, can materialise in different ways depending on the particular implementation of a description of both form and function and the method of aggregation of enclosed cells into singatures. In this section, we present an illustration of the concept applied to five case studies reflecting different environments and heterogeneous input data. We choose this diversity to demonstrate the ability to adapt the classification to rather different situations. The five cities are displayed in Figure 2. The sample offers geographical variation covering Europe (Barcelona, Spain), North America (Houston, TX, United States), South America (Medellin, Colombia), Africa (Dar es Salaam, Tanzania) and South-east Asia (Singapore), coupled with cultural diversity, different planning paradigms involved in shaping the respective environments as well as varied historical and social contexts in which the selected cities were built and developed. At the same time, the selection brings a variety of input data covering both extremes in terms of quality (e.g., official mapping in Barcelona vs remote sensing in Houston), the richness of information on functional aspects of places (e.g., detailed data on the municipal level in Medellin vs global gridded datasets in Dar es Salaam) and scale (82,375 units in Barcelona vs 2,043,581 units in Houston). We present this variety to illustrate the flexibility of spatial signatures to accommodate varied inputs and adapt to a local specificity, while retaining the merit of the concept.

4.1 Method

The delineation of spatial signatures starts with the input data reflecting form and function of each place. We use enclosed tessellation, outlined in Section 3.2, as the core spatial unit. Therefore, the input data consists of building footprints and physical barriers –delimiters– denoting streets, railways, and water bodies. Using these delimiters, we first identify the geometry of enclosures, which we combine with building footprints to grow ET cells. The resulting set allows for a

Figure 2: Selection of case studies covering geographical variation, cultural diversity, different planning paradigms involved in shaping the respective environments as well as varied historical and social contexts in which the selected cities were built.

comprehensive morphometric analysis composed of characters capturing individual aspects of form, and contextualisation, following the model proposed by Fleischmann et al., 2021. In the latter, we include the distribution of each character within the neighbouring context of each tessellation cell. Function is captured as a heterogeneous set of characteristics reflecting features from population density to location of amenities. All aspects are linked to ET cells using the most appropriate method for each data input (e.g., areal interpolation, network accessibility). The complete list of used characters relfecting both form and function, as well as implementation details are available in Appendix A.

Spatial signatures are then identified using cluster analysis on the form-function characteristics attached to ET cells. The combined data reflecting both form and function are therefore standardised and clustered using K-Means. Since the number of classes is not known a priori, we use clustergrams (Schonlau, 2002) to understand the behaviour of different solutions and select the optimal number of groups. Appendix A contains details on the clustergrams used. The final clustering is run with 1000 initialisations to ensure stability of results. Once clustered, we aggregate contiguous ET cells assigned into the same cluster to create instances of spatial signatures.

4.2 Results

Figures 3-4 illustrate the resulting spatial signatures in the respective case studies. The geometries reflect the spatial extent of individual signatures derived from the enclosed tessellation with colour coding reflecting the type of a signature, i.e. the initial cluster. Two areas within the same type are expected to share form and function characteristics, being more similar to each other than to the other classes. Note that the similarity of different colours does not encode similarity of signatures.

The granularity of classification, that is the number of clases identified through the clustergram, ranges from nine (Houston) to 19 (Medellin) signature types. However, it is interesting that

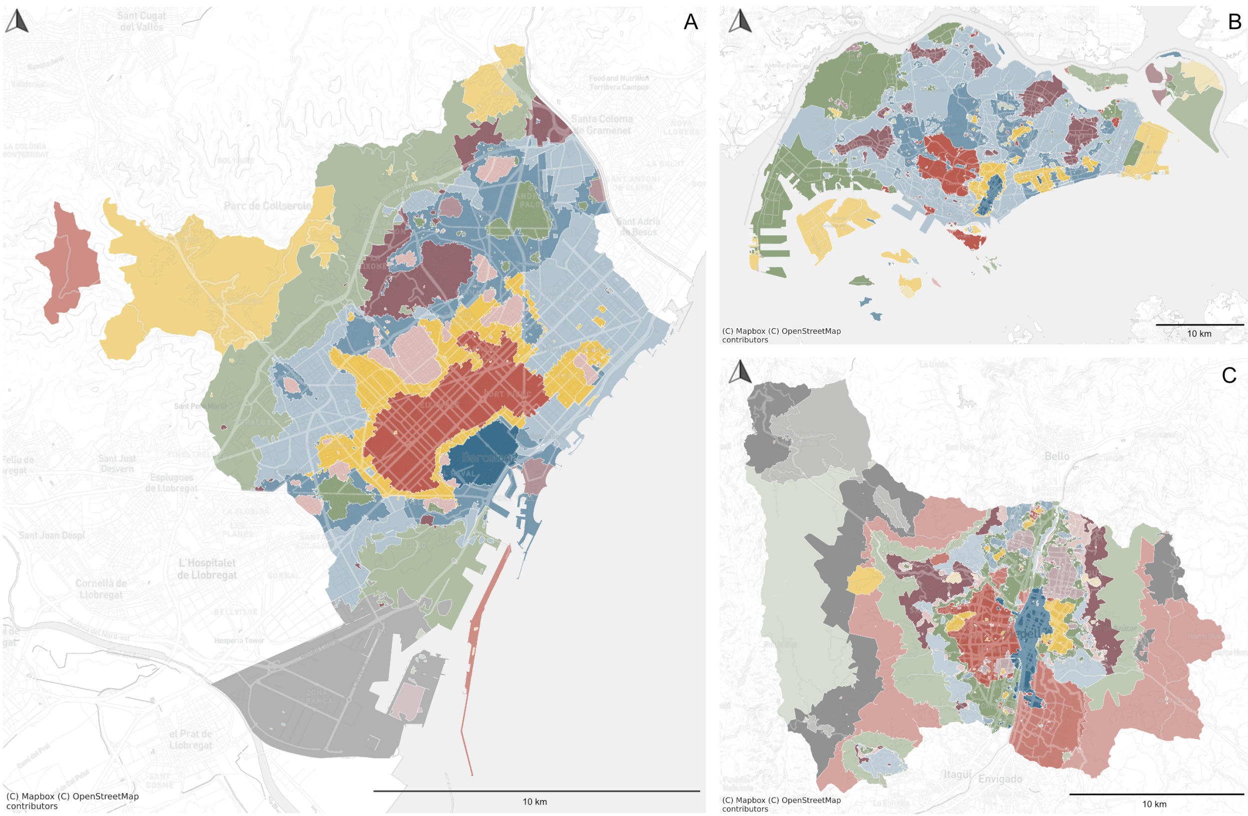

Figure 3: Resulting spatial signatures in the case of Barcelona (A), Singapore (B) and Medellin (C). Colours are used to distinguish between types within a single case. Note scale differs across maps given the different extent covered by each urban area.

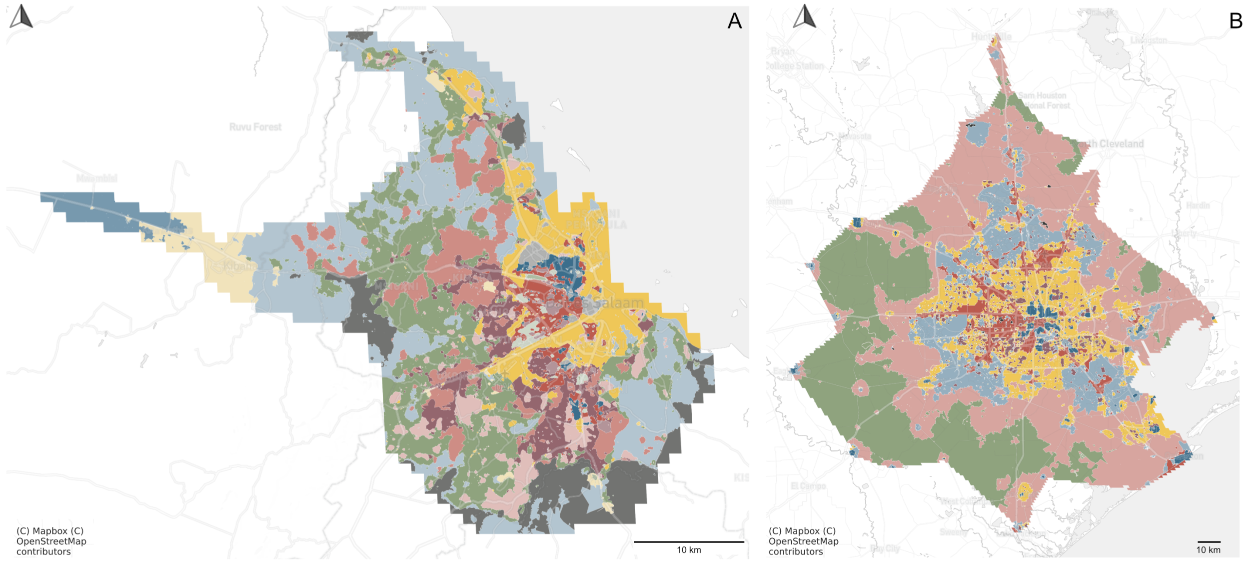

Figure 4: Resulting spatial signatures in the case of Dar es Salaam (A) and Houston (B). Colours are used to distinguish between types within a single case.

the actual number is not dependent on the size of each city but rather on the heterogeneity in form and function of each place. This phenomenon is best illustrated on the comparison between Houston and Barcelona, the largest (2 million cells) and the smallest (80 thousand cells) case. Houston, archetype of North American, post WW-II sprawling urban fabric shows considerably smaller diversity of spatial patterns (nine spatial signature types) than Barcelona (16 types). The distribution of cluster sizes follows the same unequal pattern across all cases. The most extensive types contain between 15% and 28% of all observations, gradually decreasing towards a small number of outlier clusters containing less than one per cent of all observations within each sample. All illustrations include both extremes on the urbanisation continuum, with delineated central districts on the one hand and non-urban countryside signatures on the other. We interpret this transition as a gradual move from more urban signatures to less so. The only exception where this pattern is not as clear (but still present) is Singapore, whose geographical extent is limited to the the main island and thus does not allow for full transition.

Barcelona is known for its extension grid ("Eixample"), which is captured as a unique signature (red). However, this development is historically an infill between the city's medieval core and smaller preexisting peripheric settlements. Both core and periphery are reflected in the typology of signatures, which embody these historical origins. The spatial transition between historical organic fabric and the heavily planned Eixample is reflected as another signature (yellow), stitching together different patterns into a coherent city.

The spatial distribution of signatures in Medellin tells the story of its intricate topography, even though the input data do not contain any explicit information on it. The city lies in a valley surrounded by steep slopes. While the central parts are found on the relatively flat terrain allowing paradigmatic planning and regularity of form and function, hillsides become more vernacular leading to a sharp urban edge where topography limits further development.

Signatures in Dar es Salaam reflect changes in the degree of formality in development, with formal areas (yellow) distributed across the central parts of the city, in the vicinity of the coastline. The transition between different degrees of formality is not always gradual as most informal parts of the city (light green and dark blue) appear as infills of the space not occupied by more planned neighbourhoods.

The character of spatial signatures in Houston follows two primary principles. One forms the spine of activity spreading from the city centre radially to the suburbs. The other fills the areas in-between the former, suggesting the decline of compact, walkable urban blocks into the convoluted, dendritic street network patterns of modern suburbs. The change in these predominantly residential signatures is gradual and reflects the waves of development the city has experience during the post-war period.

A similar situation can be found in Singapore, where different types of signatures can be linked to the period when they were developed. Contrary to previous cases, this evolution and, consequently, spatial signatures followed a radial pattern. This process is not entirely contiguous, and thus appear major infills built in the last 50 years.

5. Conclusions

This paper proposes the notion of spatial signatures as a characterisation of form and function designed to understand urban spaces. As such, spatial signatures have the potential to provide unique insight into the ways human populations create and inhabit cities. Developing spatial signatures begins with a partition of space that is theoretically aligned with their purpose. To this end, we propose the enclosed tessellation. With appropriate spatial units at hand, we show how form and function can be quantitatively built in. The illustration in the the previous section demonstrates how the resulting patchwork of signatures reflects a broad range of aspects stretching from topography to design, through history into the current use of space, each influencing in its own way the nature of the urban fabric.

The flexibility of our approach is evidenced in the variety of scales and shapes picked up in the analysis, and stems in great part from the choice of the spatial unit we propose –the ET cell. Unlike other common units, such as uniform grids or administrative boundaries, ET cells provide full coverage while adapting to varying conditions depending on both the urban and data relities of the application at hand. Together with the use of the Enclosed Tessellation, the main contribution of spatial signatures resides in the combination of form and function into a single classification of space. We believe this approach results in robust characterisations and is able to provide more insights than the sum of those from each aspect alone. In this respect, spatial signatures bridge purely morphological approaches based on concepts like the morphological region (Oliveira and Yaygin, 2020), Local Climate Zone (Stewart and Oke, 2012) or Urban Structural Type (Lehner and Blaschke, 2019), with functional approaches such as land use/land cover classifications (Georganos et al., 2018) or mobility and population (Gale et al., 2016). In doing so, they provide a complimentary view that adds a new perspective rather than replaces existing classifications.

Rather than a particular technique or a rigid application, the Spatial Signatures provide a *way of thinking* about building detailed, scalable and consistent characterisations of form and function in cities. It is a way of conceptualising built (and non-built) environment. In this context, the outputs from different regions or countries can be understood as different manifestations of similar concepts. Specific regions may present specific characteristics; and data landscapes vary significantly across the globe, as our illustration shows. In both cases, the spatial signatures can highlight and adapt to these cicumstances, retaining the conceptual framework.

Such flexibility and intellectual malleability makes the spatial signatures an excellent candidate to become a platform that brings together different disciplines interested in cities, their form and how activity is distributed within them. We envision this approach, and its outputs from particular classifications, as a useful input to integrate urban form and function in research across disciplines such as geography, planning, economics or sociology. Similarly, since they operationalise conceptual ideas about how our current cities can adapt to the main challenges of the century, spatial signatures can play an important role in tracking progress on initiatives such as the UN's Sustainable Development Goals. Given the rapid urbanisation in the Global South, and the constant retrofitting of cities in the Global North, developing consistent frameworks to characterise cities and track their evolution has never been more important. We hope the present paper contributes in this direction and can be the seed of further discussion and progress on these challenges.

References

Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.

Alexiou, A., Singleton, A., and Longley, P. A. (2016). A Classification of Multidimensional Open Data for Urban Morphology. *Built Environment*, 42(3):382–395.

Angel, S., Franco, S. A., Liu, Y., and Blei, A. M. (2018). The shape compactness of urban footprints. *Progress in Planning*.

Araldi, A. and Fusco, G. (2019). From the street to the metropolitan region: Pedestrian perspective in urban fabric analysis:. *Environment and Planning B: Urban Analytics and City Science*, 46(7):1243–1263.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49:45–53.

Batty, M. and Longley, P. A. (1987). Fractal-based description of urban form. *Environment and Planning B: Planning and Design*, 14(1961):123–134.

Boarnet, M. G., Crane, R., et al. (2001). *Travel by design: The influence of urban form on travel*. Oxford University Press on Demand.

Bobkova, E., Berghauser Pont, M., and Marcus, L. (2019). Towards analytical typologies of plot systems: Quantitative profile of five European cities. *Environment and Planning B: Urban Analytics and City Science*, page 239980831988090.

Bourdic, L., Salat, S., and Nowacki, C. (2012). Assessing cities: A new system of cross-scale spatial indicators. *Building Research & Information*, 40(5):592–605.

Carruthers, J. I. and Ulfarsson, G. F. (2003). Urban sprawl and the cost of public services. *Environment and Planning B: Planning and Design*, 30(4):503–522.

Conzen, M. R. G. (1960). Alnwick, northumberland: a study in town-plan analysis. *Transactions and Papers (Institute of British Geographers)*, (27):iii–122.

Copernicus Land Monitoring Service (2021). Urban atlas.

Defourny, P., Kirches, G., Brockmann, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., and Santoro, M. (2012). Land cover cci. *Product User Guide Version*, 2.

Diamond, D. B. and Tolley, G. S. (2013). *The economics of urban amenities*. Elsevier.

Dibble, J., Prelorendjos, A., Romice, O., Zanella, M., Strano, E., Pagel, M., and Porta, S. (2019). On the origin of spaces: Morphometric foundations of urban form evolution. *Environment and Planning B: Urban Analytics and City Science*, 46(4):707–730.

Dibble, J. L. (2016). *Urban morphometrics: towards a quantitative science of urban form*. PhD thesis, University of Strathclyde.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.

European Environment Agency (1990). CORINE Land Cover. pages 1–163.

Ewing, R., Hamidi, S., Grace, J. B., and Wei, Y. D. (2016). Does urban sprawl hold down upward mobility? *Landscape and Urban Planning*, 148:80–88.

Ewing, R. H., Pendall, R., and Chen, D. D. (2002). *Measuring sprawl and its impact*, volume 1. Smart Growth America Washington, DC.

Filomena, G., Verstegen, J. A., and Manley, E. (2019). A computational approach to 'The Image of the City'. *Cities*, 89:14–25.

Fleischmann, M., Feliciotti, A., Romice, O., and Porta, S. (2020a). Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems*, 80:101441.

Fleischmann, M., Feliciotti, A., Romice, O., and Porta, S. (2021). Methodological foundation of a numerical taxonomy of urban form.

Fleischmann, M., Romice, O., and Porta, S. (2020b). Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environment and Planning B: Urban Analytics and City Science*, page 2399808320910444.

Gale, C. G., Singleton, A. D., Bates, A. G., and Longley, P. A. (2016). Creating the 2011 area classification for output areas (2011 oac). *Journal of Spatial Information Science*, 2016(12):1–27.

Geddes, P. (1915). *Cities in evolution: an introduction to the town planning movement and to the study of civics*. London, Williams.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., and Wolff, E. (2018). Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE geoscience and remote sensing letters*, 15(4):607–611.

Gil, J., Montenegro, N., Beirão, J. N., and Duarte, J. P. (2012). On the Discovery of Urban Typologies: Data Mining the Multi-dimensional Character of Neighbourhoods. *Urban Morphology*, 16(1):27–40.

Hamaina, R., Leduc, T., and Moreau, G. (2012). Towards Urban Fabrics Characterization Based on Buildings Footprints. In *Bridging the Geographic Information Sciences*, volume 2, pages 327–346. Springer, Berlin, Heidelberg, Berlin, Heidelberg.

Harris, R., Sleight, P., and Webber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting*, volume 8. John Wiley & Sons.

Hillier, B. (1996). *Space is the machine : a configurational theory of architecture*. Cambridge University Press, Cambridge.

Homer, C. H., Fry, J. A., and Barnes, C. A. (2012). The national land cover database. *US Geological Survey Fact Sheet*, 3020(4):1–4.

Horner, M. W. (2004). Exploring metropolitan accessibility and urban structure. *Urban Geography*, 25(3):264–284.

Hortas-Rico, M. and Solé-Ollé, A. (2010). Does urban sprawl increase the costs of providing local public services? evidence from spanish municipalities. *Urban studies*, 47(7):1513–1540.

Howarth, P. J. and Boasson, E. (1983). Landsat digital enhancements for change detection in urban environments. *Remote sensing of environment*, 13(2):149–160.

Jochem, W. C., Bird, T. J., and Tatem, A. J. (2018). Identifying residential neighbourhood types from settlement points in a machine learning approach. *Computers, Environment and Urban Systems*, 69:104 – 113.

Jochem, W. C., Leasure, D. R., Pannell, O., Chamberlain, H. R., Jones, P., and Tatem, A. J. (2020). Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science*, page 239980832092120.

Koc, C. B., Osmond, P., Peters, A., and Irger, M. (2017). Mapping local climate zones for urban morphology classification based on airborne remote sensing data. In *2017 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4. IEEE.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.

Kropf, K. (2018). Plots, property and behaviour. *Urban Morphology*, 22(1):1–10.

Kuffer, M., Pfeffer, K., and Sliuzas, R. (2016). Slums from space-15 years of slum mapping using remote sensing. *Remote Sensing*, 8(6):455.

Lehner, A. and Blaschke, T. (2019). A Generic Classification Scheme for Urban Structure Types. *Remote Sensing*, 11(2):173.

Lynch, K. (1960). *The Image of the City*, volume 11. MIT press.

Mehaffy, M., Porta, S., Rofe, Y., and Salingaros, N. (2010). Urban nuclei and the geometry of streets: The 'emergent neighborhoods' model. *Urban Design International*, 15(1):22–46.

Muratori, S. (1959). Studi per una operante storia urbana di venezia. *Palladio*, 1959:1–113.

OECD (2018). *Rethinking Urban Sprawl*.

Oliveira, V. and Yaygin, M. A. (2020). The concept of the morphological region: Developments and prospects. *Urban Morphology*, 24(1):18.

Openshaw, S. (1981). The modifiable areal unit problem. *Quantitative geography: A British view*, pages 60–69.

Panerai, P., Castex, J., and Depaule, J.-C. (1997). *Formes urbaines: de l'îlot à la barre*. Editions Parentheses.

Pauleit, S. and Duhme, F. (2000). Assessing the environmental performance of land cover types for urban planning. *Landscape and urban planning*, 52(1):1–20.

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., and Maffenini, L. (2019). Ghs settlement grid, updated and refined regio model 2014 in application to ghs-built r2018a and ghs-pop r2019a, multitemporal (1975-1990-2000-2015) r2019a. *European Commission, Joint Research Centre (JRC)*.

Pivo, G. (1993). A taxonomy of suburban office clusters: the case of toronto. *Urban Studies*, 30(1):31–49.

Porta, S., Crucitti, P., and Latora, V. (2006). The network analysis of urban streets: A primal approach. *Environment and Planning B: Planning and Design*, 33(5):705–725.

Prasad, S. (2015). *Remotely sensed data characterization, classification, and accuracies*, volume 1.

Schirmer, P. M. and Axhausen, K. W. (2015). A multiscale classification of urban morphology. *Journal of Transport and Land Use*, 9(1):101–130.

Schonlau, M. (2002). The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses. *The Stata Journal*, 2(4):391–402.

Sevtuk, A. and Amindarbari, R. (2020). Does metropolitan form affect transportation sustainability? evidence from us metropolitan areas. *Environment and Planning B: Urban Analytics and City Science*, page 2399808320971310.

Silva, M., Oliveira, V., and Leal, V. (2017). Urban form and energy demand: A review of energy-relevant urban attributes. *Journal of Planning Literature*, 32(4):346–365.

Singleton, A. and Arribas-Bel, D. (2021). Geographic data science. *Geographical Analysis*, 53(1):61–75.

Song, Y. and Knaap, G.-J. (2007). Quantitative Classification of Neighbourhoods: The Neighbourhoods of New Single-family Homes in the Portland Metropolitan Area. *Journal of Urban Design*, 12(1):1–24.

Song, Y., Popkin, B., and Gordon-Larsen, P. (2013). A national-level analysis of neighborhood form metrics. *Landscape and Urban Planning*, 116:73–85.

Sorichetta, A., Hornby, G. M., Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data*, 2(1):150045.

Stark, T., Wurm, M., Zhu, X. X., and Taubenböck, H. (2020). Satellite-based mapping of urban poverty with transfer-learned slum morphologies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5251–5263.

Stewart, I. D. and Oke, T. R. (2012). Local Climate Zones for Urban Temperature Studies. *Bulletin of the American Meteorological Society*, 93(12):1879–1900.

Taubenböck, H., Debray, H., Qiu, C., Schmitt, M., Wang, Y., and Zhu, X. (2020). Seven city types representing morphologic configurations of cities across the globe. *Cities*, 105:102814.

Taubenböck, H., Weigand, M., Esch, T., Staab, J., Wurm, M., Mast, J., and Dech, S. (2019). A new ranking of the world's largest cities—do administrative units obscure morphological realities? *Remote Sensing of Environment*, 232:111353.

Trewartha, G. T. (1934). Japanese cities distribution and morphology. *Geographical Review*, 24(3):404–417.

Usui, H. and Asami, Y. (2013). Estimation of mean lot depth and its accuracy. *Journal of City Planning Institute of Japan*, 48(3):357–362.

Venerandi, A., Fusco, G., Tettamanzi, A., and Emsellem, D. (2019). A machine learning approach to study the relationship between features of the urban environment and street value. *Urban Science*, 3(3):100.

Venerandi, A., Quattrone, G., and Capra, L. (2018). A scalable method to quantify the relationship between urban form and socio-economic indexes. *EPJ Data Science*, 7:1–21.

Webber, R. and Burrows, R. (2018). *The Predictive Postcode: The Geodemographic Classification of British Society*. Sage.

Weng, Q. and Quattrochi, D. A. (2018). *Urban remote sensing*. CRC press.

Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):1–55.

Appendix A. Technical appendix

A. The complete lists of used characters relfecting both form and function across case studies

Table 1: The complete list of form characters used in the Barcelona case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

index	element	context	category
area	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	dimension
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension
circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
street alignment	tessellation cell	tessellation cell	distribution
length	street segment	street segment	dimension
width	street profile	street segment	dimension
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity
linearity	street segment	street segment	shape
area covered	street segment	street segment	dimension
buildings per meter	street segment	street segment	intensity
area covered	street node	street node	dimension
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension

Continued on next page

index	element	context	category
reached area	neighbouring segments	neighbouring segments	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
shared walls ratio	adjacent buildings	adjacent buildings	distribution
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
weighted reached enclosures	neighbouring tessellation cells	cell queen neighbours 3	intensity
area	enclosure	enclosure	dimension
perimeter	enclosure	enclosure	dimension
circular compactness	enclosure	enclosure	shape
equivalent rectangular index	enclosure	enclosure	shape
compactness-weighted axis	enclosure	enclosure	shape
solar orientation	enclosure	enclosure	distribution
weighted neighbours	enclosure	enclosure	distribution
weighted cells	enclosure	enclosure	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
node density	street network	nodes 5 steps	intensity
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
degree weighted node density	street network	nodes 5 steps	intensity
local closeness centrality	street network	nodes 5 steps	connectivity
perimeter wall length	adjacent buildings	joined buildings	dimension
number of courtyards	adjacent buildings	joined buildings	intensity
square clustering	street network	street network	connectivity

Table 2: The complete list of function characters and transfer methods used in the Barcelona case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

character	input spatial unit	transfer method
population	block	Building-based Dasymetric mapping
number of car parks	block	Dasymetric mapping
number of other items that are not premises	block	Dasymetric mapping
land use	parcel	Spatial join (centroid)
number of dwellings	building	Attribute join

Continued on next page

character	input spatial unit	transfer method
current use	building	Attribute join
age	building	Attribute join
heritage	points	Accessibility -# within 15min
heritage	polygons	Spatial join
culture (cinemas, museums, libraries, theaters)	points	Accessibility - distance to nearest / # within 15min
parks	points	Accessibility - distance to nearest / # within 15min
economic census	poitnts	Accessibility - distance to nearest / # within 15min
restaurants	point	Accessibility - distance to nearest / # within 15min
trees	points	Spatial join (count)
NDVI	raster 1m	Zonal stats

Table 3: The complete list of form characters used in the BarceMedellinlona case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

index	element	context	category
area	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	dimension
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension
circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
street alignment	tessellation cell	tessellation cell	distribution
length	street segment	street segment	dimension
width	street profile	street segment	dimension
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity

Continued on next page

index	element	context	category
linearity	street segment	street segment	shape
area covered	street segment	street segment	dimension
buildings per meter	street segment	street segment	intensity
area covered	street node	street node	dimension
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension
reached area	neighbouring segments	neighbouring segments	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
shared walls ratio	adjacent buildings	adjacent buildings	distribution
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
weighted reached enclosures	neighbouring tessellation cells	cell queen neighbours 3	intensity
area	enclosure	enclosure	dimension
perimeter	enclosure	enclosure	dimension
circular compactness	enclosure	enclosure	shape
equivalent rectangular index	enclosure	enclosure	shape
compactness-weighted axis	enclosure	enclosure	shape
solar orientation	enclosure	enclosure	distribution
weighted neighbours	enclosure	enclosure	distribution
weighted cells	enclosure	enclosure	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
node density	street network	nodes 5 steps	intensity
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
degree weighted node density	street network	nodes 5 steps	intensity
local closeness centrality	street network	nodes 5 steps	connectivity
perimeter wall length	adjacent buildings	joined buildings	dimension
number of courtyards	adjacent buildings	joined buildings	intensity
square clustering	street network	street network	connectivity

Table 4: The complete list of function characters and transfer methods used in the Medellin case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

character	input spatial unit	transfer method
trees	points	Spatial join (count)
parks	polygons	Distance to nearest
heritage large areas	polygons	Spatial join (boolean)
land use	polygons	tobler
pois	points	Accessibility - distance to nearest / # within 15min
public spaces	polygons	Accessibility - area within radius
population	raster	Zonal stats
NDVI	raster	Zonal stats

Table 5: The complete list of form characters used in the Dar es Salaam case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

index	element	context	category
area	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	dimension
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension
circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
length	street segment	street segment	dimension
width	street profile	street segment	dimension
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity
linearity	street segment	street segment	shape

Continued on next page

index	element	context	category
area covered	street segment	street segment	dimension
buildings per meter	street segment	street segment	intensity
area covered	street node	street node	dimension
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
reached area	neighbouring segments	neighbouring segments	dimension
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
weighted reached enclosures	neighbouring tessellation cells	cell queen neighbours 3	intensity
reached neighbors	neighbouring tessellation cells	cell queen neighbours 3	intensity
reached area	neighbouring tessellation cells	cell queen neighbours 3	dimension
area	enclosure	enclosure	dimension
perimeter	enclosure	enclosure	dimension
circular compactness	enclosure	enclosure	shape
equivalent rectangular index	enclosure	enclosure	shape
compactness-weighted axis	enclosure	enclosure	shape
solar orientation	enclosure	enclosure	distribution
weighted neighbours	enclosure	enclosure	distribution
weighted cells	enclosure	enclosure	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
reached area	street network	segment 3 steps	dimension
reached cells	street network	segment 3 steps	intensity
node density	street network	nodes 5 steps	intensity
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
degree weighted node density	street network	nodes 5 steps	intensity

Continued on next page

index	element	context	category
local closeness centrality	street network	nodes 5 steps	connectivity
square clustering	street network	street network	connectivity

Table 6: The complete list of function characters and transfer methods used in the Dar es Salaam case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

character	input spatial unit	transfer method
population	raster 100m	Zonal stats
NDVI	raster 10m	Zonal stats
land cover	raster	Zonal stats
night lights	raster	Zonal stats

Table 7: The complete list of form characters used in the Houston case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

index	element	context	category
area	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	dimension
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension
circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
length	street segment	street segment	dimension
width	street profile	street segment	dimension
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity
linearity	street segment	street segment	shape
area covered	street segment	street segment	dimension

Continued on next page

index	element	context	category
buildings per meter	street segment	street segment	intensity
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
reached area	neighbouring segments	neighbouring segments	dimension
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
weighted reached enclosures	neighbouring tessellation cells	cell queen neighbours 3	intensity
reached neighbors	neighbouring tessellation cells	cell queen neighbours 3	intensity
reached area	neighbouring tessellation cells	cell queen neighbours 3	dimension
area	enclosure	enclosure	dimension
perimeter	enclosure	enclosure	dimension
circular compactness	enclosure	enclosure	shape
equivalent rectangular index	enclosure	enclosure	shape
compactness-weighted axis	enclosure	enclosure	shape
solar orientation	enclosure	enclosure	distribution
weighted neighbours	enclosure	enclosure	distribution
weighted cells	enclosure	enclosure	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
reached area	street network	segment 3 steps	dimension
reached cells	street network	segment 3 steps	intensity
node density	street network	nodes 5 steps	intensity
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
degree weighted node density	street network	nodes 5 steps	intensity
local closeness centrality	street network	nodes 5 steps	connectivity
square clustering	street network	street network	connectivity

Table 8: The complete list of function characters and transfer methods used in the Houston case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

character	input spatial unit	transfer method
population	raster 100m	Zonal stats
NDVI	raster 10m	Zonal stats
land cover	raster	Zonal stats
night lights	raster	Zonal stats
employment	census block	Dasymetric interpolation
historical sites	point	Accessibility

Table 9: The complete list of form characters used in the Singapore case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

index	element	context	category
area	building	building	dimension
perimeter	building	building	dimension
courtyard area	building	building	dimension
circular compactness	building	building	shape
corners	building	building	shape
squareness	building	building	shape
equivalent rectangular index	building	building	shape
elongation	building	building	shape
centroid - corner distance deviation	building	building	shape
centroid - corner mean distance	building	building	dimension
solar orientation	building	building	distribution
street alignment	building	building	distribution
cell alignment	building	building	distribution
longest axis length	tessellation cell	tessellation cell	dimension
area	tessellation cell	tessellation cell	dimension
circular compactness	tessellation cell	tessellation cell	shape
equivalent rectangular index	tessellation cell	tessellation cell	shape
solar orientation	tessellation cell	tessellation cell	distribution
coverage area ratio	tessellation cell	tessellation cell	intensity
street alignment	tessellation cell	tessellation cell	distribution
length	street segment	street segment	dimension
width	street profile	street segment	dimension
openness	street profile	street segment	distribution
width deviation	street profile	street segment	diversity
linearity	street segment	street segment	shape
area covered	street segment	street segment	dimension
buildings per meter	street segment	street segment	intensity
area covered	street node	street node	dimension

Continued on next page

index	element	context	category
alignment	neighbouring buildings	neighbouring cells (queen)	distribution
mean distance	neighbouring buildings	neighbouring cells (queen)	distribution
weighted neighbours	tessellation cell	neighbouring cells (queen)	distribution
area covered	neighbouring cells	neighbouring cells (queen)	dimension
reached area	neighbouring segments	neighbouring segments	dimension
reached cells	neighbouring segments	neighbouring segments	intensity
degree	street node	neighbouring nodes	distribution
mean distance to neighbouring nodes	street node	neighbouring nodes	dimension
shared walls ratio	adjacent buildings	adjacent buildings	distribution
mean inter-building distance	neighbouring buildings	cell queen neighbours 3	distribution
weighted reached enclosures	neighbouring tessellation cells	cell queen neighbours 3	intensity
area	enclosure	enclosure	dimension
perimeter	enclosure	enclosure	dimension
circular compactness	enclosure	enclosure	shape
equivalent rectangular index	enclosure	enclosure	shape
compactness-weighted axis	enclosure	enclosure	shape
solar orientation	enclosure	enclosure	distribution
weighted neighbours	enclosure	enclosure	distribution
weighted cells	enclosure	enclosure	intensity
local meshedness	street network	nodes 5 steps	connectivity
mean segment length	street network	segment 3 steps	dimension
cul-de-sac length	street network	nodes 3 steps	dimension
node density	street network	nodes 5 steps	intensity
proportion of cul-de-sacs	street network	nodes 5 steps	connectivity
proportion of 3-way intersections	street network	nodes 5 steps	connectivity
proportion of 4-way intersections	street network	nodes 5 steps	connectivity
degree weighted node density	street network	nodes 5 steps	intensity
local closeness centrality	street network	nodes 5 steps	connectivity
perimeter wall length	adjacent buildings	joined buildings	dimension
number of courtyards	adjacent buildings	joined buildings	intensity
square clustering	street network	street network	connectivity

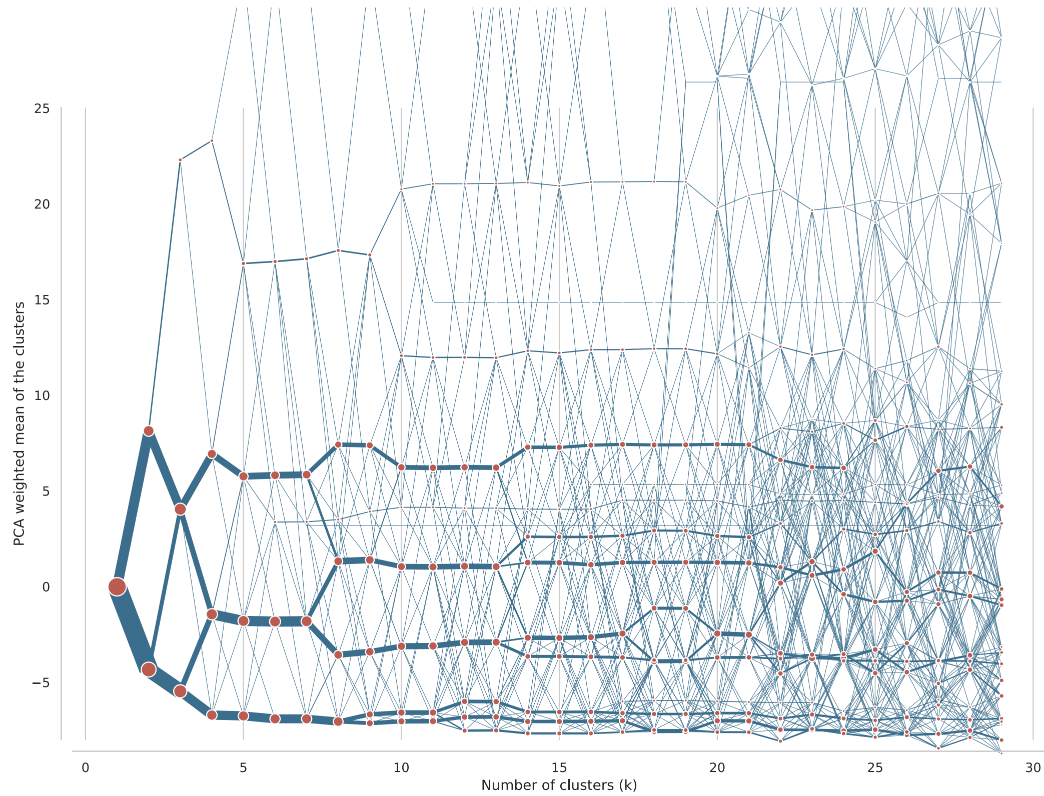

Figure 5: Clustergram (truncated along the vertical axis) illustrating the behaviour of dataset in different clustering options. The optimal number of clusters derived using the clustergram is 16.

Table 10: The complete list of function characters and transfer methods used in the Singapore case study. The implementation details are available in Jupyter notebooks available at ¡anonymised for peer-review¿.

character	input spatial unit	transfer method
NDVI	raster	Zonal stats
population	raster	Zonal stats
night lights	raster	Zonal stats
eating est	point	Accessibility
supermarkets	point	Accessibility
land use	polygon	Areal interpolation
parks	point	Accessibility
monuments	point	Accessibility

B. Clustergrams

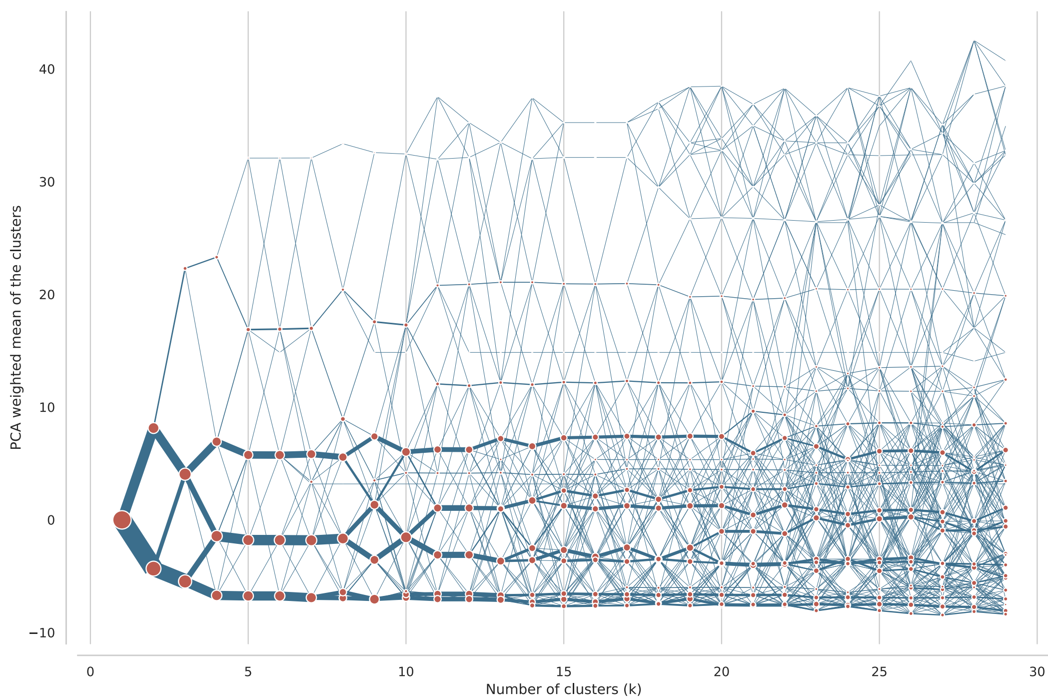

Figure 6: Clustergram (truncated along the vertical axis) illustrating the behaviour of dataset in different clustering options. The optimal number of clusters derived using the clustergram is 19.

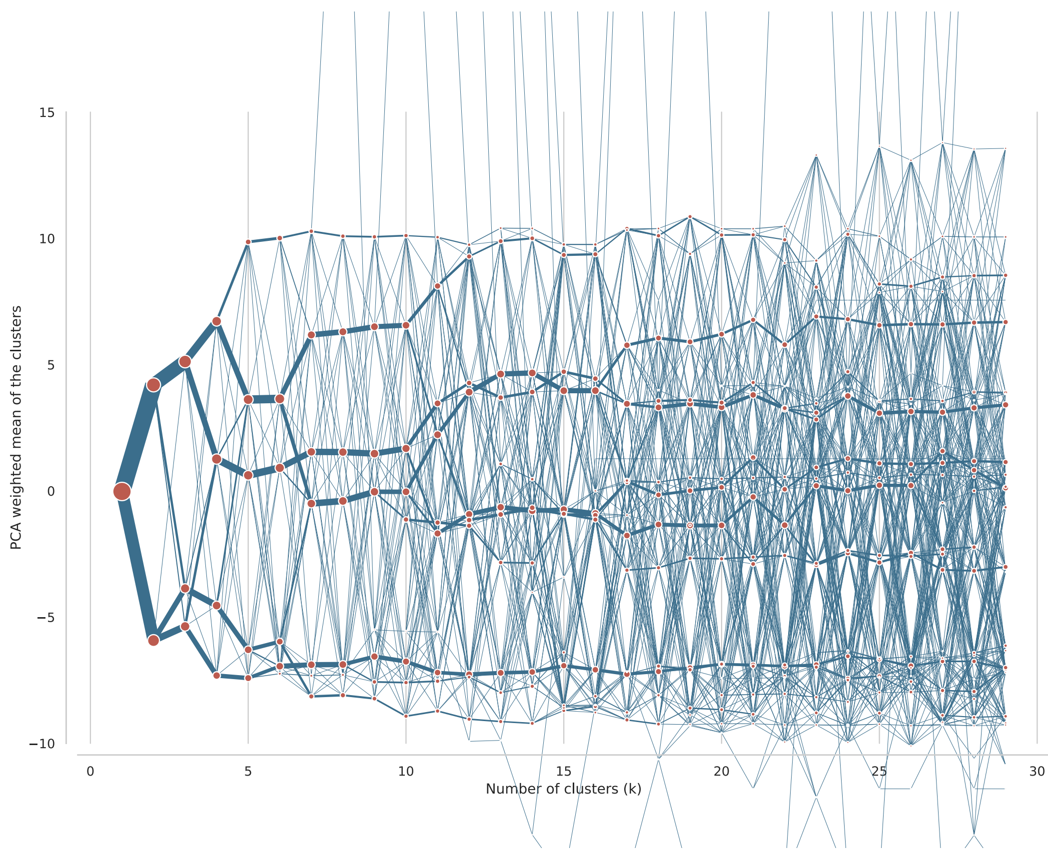

Figure 7: Clustergram (truncated along the vertical axis) illustrating the behaviour of dataset in different clustering options. The optimal number of clusters derived using the clustergram is 17.

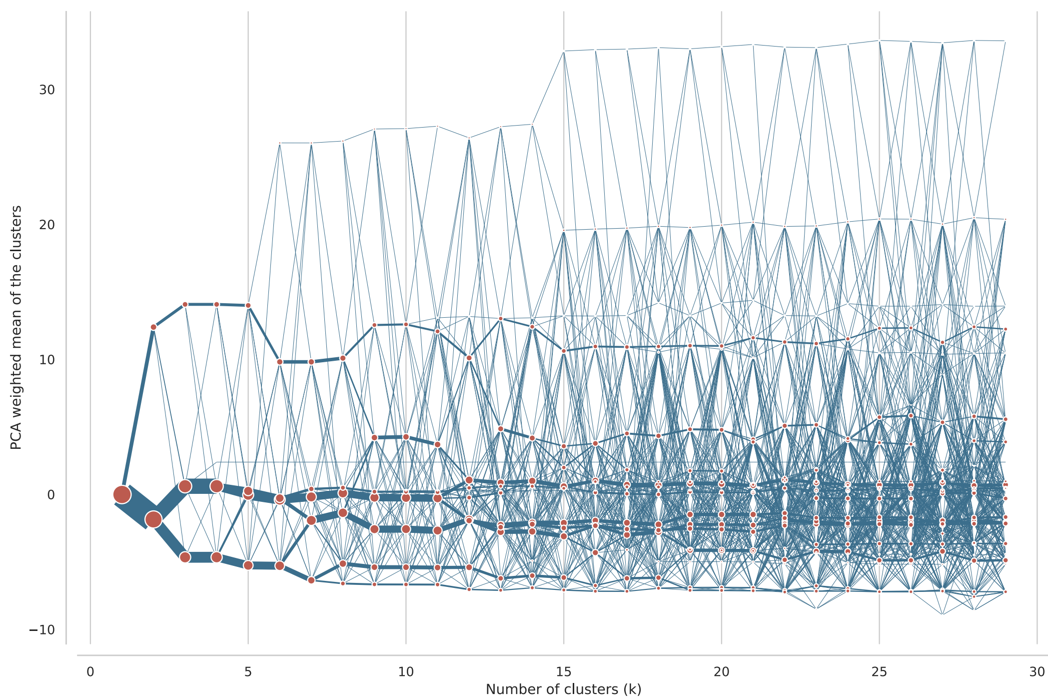

Figure 8: Clustergram (truncated along the vertical axis) illustrating the behaviour of dataset in different clustering options. The optimal number of clusters derived using the clustergram is 9.

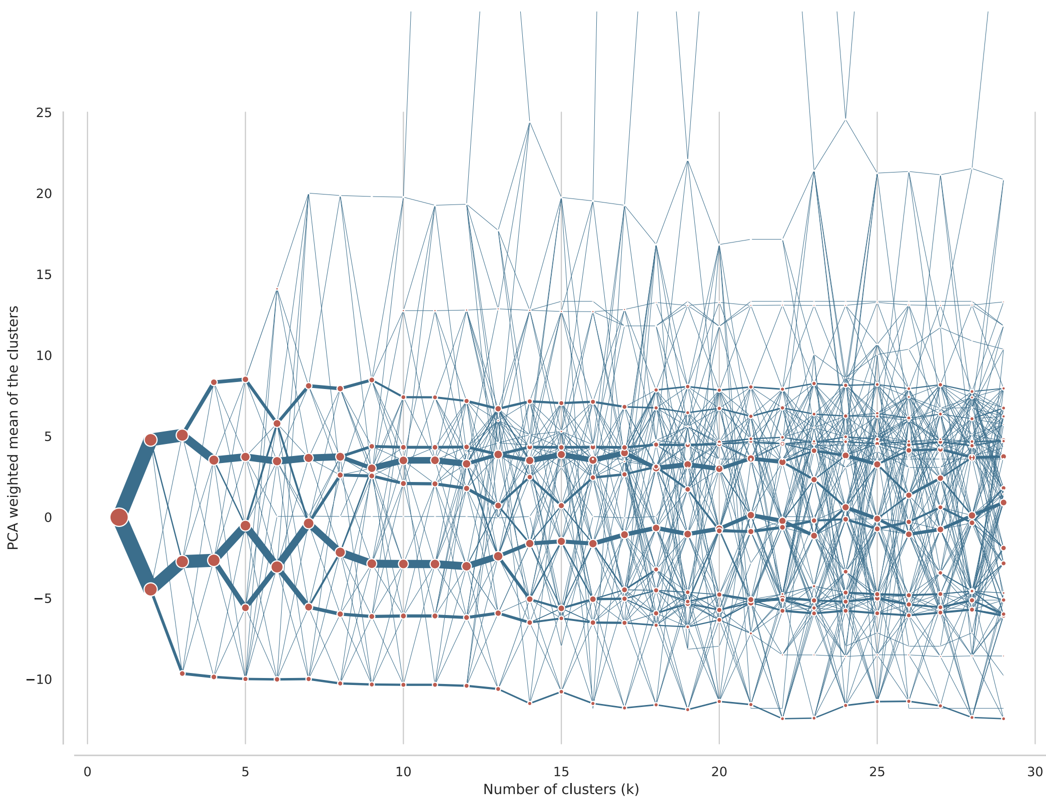

Figure 9: Clustergram (truncated along the vertical axis) illustrating the behaviour of dataset in different clustering options. The optimal number of clusters derived using the clustergram is 16.