Proceedings of the Ninth International Space Syntax Symposium
Edited by Y O Kim, H T Park and K W Seo, Seoul: Sejong University, 2013

UNSUPERVISED CLASSIFICATION OF EVOLVING METROPOLITAN STREET PATTERNS

046

Miguel Serra
CITTA, University of Porto
e-mail: alvim.serra@fe.up.pt

Jorge Gil
Delft University of Technology
e-mail: j.a.lopesgil@tudelft.nl

Paulo Pinho
CITTA, University of Porto
e-mail: pcpinho@fe.up.pt

Abstract

The definition of urban typomorphologies is fundamental both for the description as well as for the prescription of urban form. However, urban typomorphologies are traditionally defined through time-consuming analytical procedures, which are based mainly on the personal knowledge and ability of the analyst and on ad-hoc generalizations. Such generalizations are also usually context-dependent and apparently not fit to deal with contemporary metropolitan and suburban forms, whose emergent and very different morphologies have until now eluded stringent classifications. In this paper we explore the possibility of defining urban typomorphologies through the use of non-supervised classification techniques, focusing on urban street patterns. We illustrate the method through the analysis of the expansions of the street network of Oporto's metropolitan area, over the last sixty years. We have modelled such network using space syntax's axial mapping technique on four different historical moments (2011, 1998, 1977, and 1951). By comparing these four axial models and by cross-checking them with old cartographies and current orthophotomaps, we were able to isolate and to identify each and every urban development intervention entailing the creation of new streets (as represented by their axial lines), occurring within the metropolitan network along the study's time-span. Using this dataset of street-network interventions (amounting to 4208 objects), we quantified a certain number of geometrical and topological attributes describing each intervention's street-layout morphology. This database was then explored by non-hierarchical classification techniques and a number of significant clusters were extracted. We conclude by showing that such clusters correspond to clearly different morphotypes of street-layouts, with different distributions over time. Based on these morphotypes we propose a new taxonomy of metropolitan street network expansions. We suggest that their identification can provide a solid basis to support the formulation and implementation of planning and urban design policies.

Keywords: *unsupervised, classification, typomorphology, metropolitan*

Theme: *Modelling and Methodological Developments*

1. Introduction

To classify is to understand. When one divides a subject of study into meaningful categories, one may grasp order underneath the large diversity of individual phenomena. In nature very few things occur in small, sharp and indisputable variability; facts and objects must be arranged in an orderly fashion before their unifying principles can be discovered. That is why classification is one of the fundamental concerns of science.

Likewise, in urban and architectural morphological studies, there is a long tradition of classification of existing built forms, known as typomorphology. Typomorphology is the study of urban form derived from typical spaces and structures (Moudon 1994). It presumes the existence of certain morpho-structural similarities among the variety of existing built forms, which allow for their grouping into intrinsically distinct categories, called *types*. The type is an elemental object that embodies a morphological idea and has little or nothing to do with function, even if it is sometimes wrongly equated with it. Indeed, buildings and cities are largely independent of their use at one certain moment in time and can shelter many different functions throughout history, while keeping their form generally intact (Rossi 1966).

Urban typomorphological studies are important for two main reasons. Firstly, they constitute a way of understanding the shape of cities and its evolution. They try to unveil the fundamental elements that construct the city at each moment in history, their permanence, disuse or mutation through time, and the fundamental physical and spatial relations that they establish between themselves. Secondly, typomorphological studies can also contribute to inform urban planning and design, whether as preliminary analytical procedures of the design and planning processes or as prescription instruments for the implementation of urban plans and policies concerning the physical and spatial structure of the city (Moudon 1994). These instrumental and prescriptive dimensions of typomorphological studies, even if traditionally relatively marginal (Hall 2008), have gained particular importance recently, leveraged by concerns on the preservation of the character of historical urban areas (Kropf 2011), by the enthusiastic adoption in the U.S. and in the U.K. of form-based codes (Carmona, Marshall et al. 2006) and by the increasingly inescapable agenda on urban sustainability (Jabareen 2006).

However, urban typomorphological classification has never attained the degree of rigour and systematization of other disciplines, as biological taxonomy or structural linguistics. There are obvious reasons for this, since urban morphology is a recent field of scientific enquiry with a restricted research community, not comparable to those of other social sciences let alone those of the natural sciences. One may also argue that the problem lays in the fact that the objects of urban typomorphology are more difficult to define than, say, those of biological taxonomy (e.g. the difference between a plant and an animal) or of geology (e.g. the similitude in the constitution of two types of minerals); or, still, that urban form is perhaps an intrinsically ill-defined subject. But these seem only poor excuses not to push forward urban typomorphology into the realm of consistent classification, as it has happened in other disciplines which, in their beginnings, were eventually loosely defined and not totally objective. In fact, rather than being created by any insurmountable variability of the subject, the problem seems rooted instead on uncertainties regarding which morphological attributes to consider, on the definition of the desirable degree of resolution of the classification, on the classification criteria themselves or even on the lack of consistency of the adopted terminology of types (Marshall 2005).

A recent paper (Gil et al. 2012) has collated some of the shortcomings of traditional typomorphological analysis reported in the literature. These include the laborious and time-consuming nature of traditional analytical procedures (thus limited to small samples and to a few morphological dimensions), their relative opacity and subjectivity (relying mainly on the personal knowledge and ability of the analyst), their strong cultural and geographical context-dependency (thus not obviously applicable in other urban settings) and their subsequent questionable reproducibility or generalization. Regarding the recent instrumental use of typomorphology in form-based codes and urban preservation, Samuels (2008) makes a stringent critique of the shallowness of the main-stream and public agencies approaches, which seem to emphasize only superficial and ephemeral architectural features and show little awareness of the importance of other deeper and less visual aspects of urban form, such as the structure of the public space system. Other authors have yet noted the seeming inadequacy (or, at least, the lack of application) of traditional typomorphological analysis in contemporary suburban and metropolitan contexts (Levy 1999; Stanilov 2004) and even its irrelevance when faced with the explosive urbanization patterns of developing countries (Shane 2011). In fact, typomorphological analysis has been almost entirely dedicated to traditional or historical urban contexts, with only a few incursions into the contemporary urban periphery (Southworth and Owens 1993; Case-Sheer 2001; Stanilov and Case-Sheer 2004) or on the metropolitan scale (Stanilov 2002). The emergent and very different morphology of these new urban territories and scales has until now eluded stringent typomorphological classifications.

In this paper we explore, develop and exemplify the possibility of defining urban typologies through the use of unsupervised classification techniques, a method proposed in a previous paper (Gil et al. 2012) that we believe is capable of coping with the abovementioned difficulties. The method has several advantages over traditional typomorphological analysis. Firstly, it is consistent, objective and reproducible, being based on well-established statistical algorithms; secondly, because it uses computerized data analysis, it can handle a very large number of urban form dimensions, both quantitative and qualitative; and thirdly, it produces results which are derived only from the data itself and not from pre-defined types, therefore revealing the intrinsic nature of local types as well as allowing for the discovery of others, previously unknown.

We illustrate the method through the analysis of the street patterns formed by the expansions of the public space network of Oporto's metropolitan area[1] over the last sixty years, extracting a number of significant clusters. We conclude by showing that these clusters correspond to clearly differentiated morpho-types, with different distributions over time, and by suggesting that their identification can provide a solid basis to support the formulation and implementation of planning and urban design policies.

2. Methodology

Nowadays, across a wide variety of fields, data are being collected and accumulated at dramatic pace, producing massive databases that make impracticable the traditional procedures of manual extraction of patterns. Rather, new automatic or semi-automatic methods are adopted, based on exploratory statistical techniques (as cluster and dimension reduction analysis), which are known as *unsupervised classification* methods. In many disciplines such classification techniques have produced huge advances and have become unavoidable. It is our belief that this could be also the case with urban typomorphology. The fact of traditional typomorphological analysis being overwhelmingly supported solely by visual observation methods is perhaps its greatest frailty. Even if we humans are particularly good at finding spatial relations in seemingly disordered patterns (something for long recognized and studied by Gestalt psychology) we are also very prone to see order where there is none (a common

[1] Oporto is the second largest Portuguese city and the core of largest metropolitan area of the peninsular northwest, with a population of approximately 2.300.000 inhabitants

phenomenon known as apophenia, or the tendency to see patterns in random data). Moreover, there are certain types of morphological order (e.g. configuration, in the syntactic sense of the term) that are simply not visually accessible. A systematic, consistent and robust classification system should not be based solely on visual perceptions (even is these may prove necessary and useful), but rather in well-defined and non-subjective classification criteria. In any classification exercise the two key issues are the choosing of the relevant attributes and the definition of the most parsimonious number of classes that are able to describe the variability of the phenomena under study. As we will see, the latter issue can be tackled by the proposed method, but the former deserves some preliminary discussion.

One of the main findings of urban morphology is the scale of resistance to change of urban form components (Case-Sheer 2001; Whitehand 2001). It is well known that the architectural superstructure changes quite often, that the structure of land-tenure is more resilient to change and that the open space system is the most perennial of all urban form components: once laid out it endures for very long periods of time, ultimately for millennia. Furthermore, the study of urban spatial networks (of which space syntax is perhaps the most fertile manifestation) has shown that the long-sought relations between urban form and urban functioning were not to be found on the superficial looks of buildings, but in the non-visual, deep topological structures of urban street networks (Hillier 1996; Hillier 2009). Thus, because of their structural role on the definition of future urban form and because of their impact on urban functioning, street systems ought to be seen as preferential targets of morphological study.

Since streets are the most perennial and determinant of all elements of urban form, studying street patterns in order to define urban typomorphologies seems reasonable. However, pure typomorphological studies of street patterns and layouts, or at least taking street systems as main research objects, are surprisingly rare. The patterns composed by streets are frequently conceived as a subsidiary part of urban tissues, or just as the negative of built forms, in figure-ground representations. They are acknowledged, but seldom analyzed and even more rarely quantified. We refer here to studies emanating from the process-typological and historical-geographical approaches to urban form, using Karl Kropf's (2009) classification, which are the traditional sources of urban typomorphological studies. However, even within the space syntax field (or configurational approach, according to Kropf), which is wholly dedicated to the study of street systems, typological classifications of street patterns are extremely scarce.

This raises again the question of what attributes to consider. In contrast with built forms, typomorphologies of street systems normally do not go beyond very general pattern designations. There is, however, one remarkable exception: Stephen Marshall's (2005) book "Streets & Patterns". Marshall establishes a basic distinction between *compositional* and *configurational* descriptors of street patterns. "Composition refers to absolute geometric layout, as represented in a scale plan, featuring absolute position, lengths, areas, [etc]. Configuration refers to topology, as represented on [a graph], featuring links and nodes, their ordering, […] adjacency and connectivity." Op. Cit, p. 86. This author shows how existing classifications are in fact mixtures and permutations of characteristics falling into these two basic categories.

These may be seen, in fact, as the two basic quantitative descriptors of form: geometry and topology. Geometry describes the absolute composition of forms, ultimately in such a complete way that only geometrically equal shapes (i.e. whose superposition is perfect at all points) may be said to be similar. Geometry deals with sizes, lengths, widths and heights, relative positions (distances) and angles; it aims at describing integrally the metrical attributes of a given shape or space. Topology, on its part, deals only with the connectedness, continuity and boundary of forms, while ignoring their sizes, areas or volumes, specific directional orientations, or any type of measurable distance within them; topological properties are the deepest and most invariant morphological attributes of a shape or a space. It is clear that these two types of descriptions are saying quite different things, but that they may be both relevant in characterizing urban street patterns. One may measure thoroughly all the geometric properties of a given street layout and represent them accurately on a plan. However, such representation, even if depicting the exact shape of the layout, says little about its topology, e.g. about the number of possible routes between each of its streets or about the degree of connectivity or continuity of those streets. These are aspects that are undoubtedly different between a grid-like street pattern and a tree-like street pattern, with many culs-de-sacs. Conversely, we may know exactly the average connectivity of a given street layout and the exact number and type of its junctions, but this tells us almost nothing about its general geometric shape or about the total length of its streets. Thus, both geometrical and topological (or compositional and configurational) descriptions seem necessary, in order to fully discriminate relevant morphological aspects between potentially different types of street patterns.

One last methodological decision of the study was to analyse all interventions together, independently of their period of construction. Only after their simultaneous classification would they be analysed by time period. This guarantees a strong degree of consistency in our classification while allowing checking for variations in the distribution of types over time, revealing potential temporal shifts in morphological trends.

Finally, and after some preliminary screening of the data (see section 3), we performed an unsupervised classification method on this database, namely the k-means algorithm. This is an exploratory multivariate analysis technique, which allows to group observations into homogeneous groups derived only from the data itself, thus defining urban morphological types. It is an iterative process that seeks to minimize distances between observations in a cluster, based on similarity of measurements between them. The point corresponding to (or nearest to) the final centroid of each cluster can be seen as the archetype of that cluster. The definition of the number of clusters to be found is a previous condition of this method. But it can be used in an exploratory way, by running the algorithm several times for an increasing number of clusters, and then plotting the sum of the squares of the distances of each point to the centroid it belongs against the number of clusters. This produces a graph (or 'scree plot') of a continuously descending curve, usually with a varying slope. One needs to look for sudden variations of slope, namely the first point where the curve's slope becomes suddenly less. This means that beyond that number of clusters the decrease in the distance to the clusters centroids is no longer explaining a significant amount of variance in the data, and therefore the ideal number of clusters was attained.

3. Case study

For the reasons outlined in the previous section, we have elected the open space system as the main object of study, constructing a diachronic axial model of Oporto's metropolitan region, covering an area of 1600 Km^2 contained in a circle of 25 Km radius, centred in the centroid of Oporto's municipality limits. Within this study area are comprised 13 municipalities of Oporto's metropolitan region and 68135 axial lines with a total length of 9649 Km. We have chosen to represent the street network as an axial map due to reasons connected with a broader research of which this work is part. But, for all intents and purposes, the proposed method is also applicable to other kinds of street network representation, such as common GIS road-centre lines. This axial map was drawn in a GIS platform over orthophotomaps[2] dated from 2011. Subsequently, that axial map was edited over increasingly older cartographies[3], dating from

[2] We used Bing Maps Aerial, a free web mapping service from Microsoft, built-in in ESRI's ArcGIS 10 and offering worldwide orthographic aerial and satellite imagery

[3] Military cartographies from the Portuguese Military Geographic Institute (IgeoE) at the 1:25000 scale

1951, 1977 and 1998, in order to recreate the metropolitan street network on those periods. On a next step, on each of the temporal versions of the axial map, the individual grid interventions that appeared over time were identified through the simultaneous visualization of the current orthophotomaps over the sub-set of all new lines in each period. For each grid intervention an ID attribute was recorded as well as its period of construction, amounting to a total of 4208 occurrences along the study's time span. It is important to note that these 4208 interventions do not correspond to building construction, but to road construction; evidently, many more buildings appeared during that time, not only along new streets but also along existing ones.

With all street interventions isolated in this manner, we proceed to characterize and classify them regarding their morphologic constitution. Accordingly, we have defined an initial set of 20 variables, describing the composition and the configuration of our 4208 grid interventions. Some of these variables were computed directly on the grid interventions' dataset using simple GIS algorithms; others are ratios and combinations of the previous (see Figure 3). After a preliminary co-variance analysis this set of variables was reduced to a total of 11 variables, due to strong co-linearity between some of them. Correlations were much stronger between the variables reflecting the composition of the grid interventions than between those reflecting their configuration, with road length being the most discriminant one. Some significative correlations were also found among configurational attributes, but these could be explained by the very nature of street layouts. Indeed, it seems unreasonable to ignore the count of internal and external blocks, as well as the count of street junctions by their type, just because these properties are bounded to occur simultaneously. In fact, their correlations are explainable by that fact and not because they truly represent similar things. The final group of 11 variables is described in the following paragraphs.

When an intervention is made to the metropolitan spatial network through the construction of a new set of streets, two types of characteristics are worth noting: the internal (or intrinsic) morphologic characteristics of the intervention (i.e. those that are defined solely by the set of new streets) and the external (or extrinsic) morphological characteristics of the intervention (i.e. those that are defined by the new streets and by the existing ones, with which they relate or connect to). In Figure 1, the hypothetical existing grid is represented by thick, black dashed lines; and the newly built streets of the two interventions, by thin black lines (representing axial lines) and by thick light-grey segments (representing axial segments). Thus, we will be looking at compositional and configurational attributes of each intervention, both at the internal level (i.e. concerning only the intervention itself) and at the external level (i.e. concerning the relations that each interventions establishes with the existing grid).

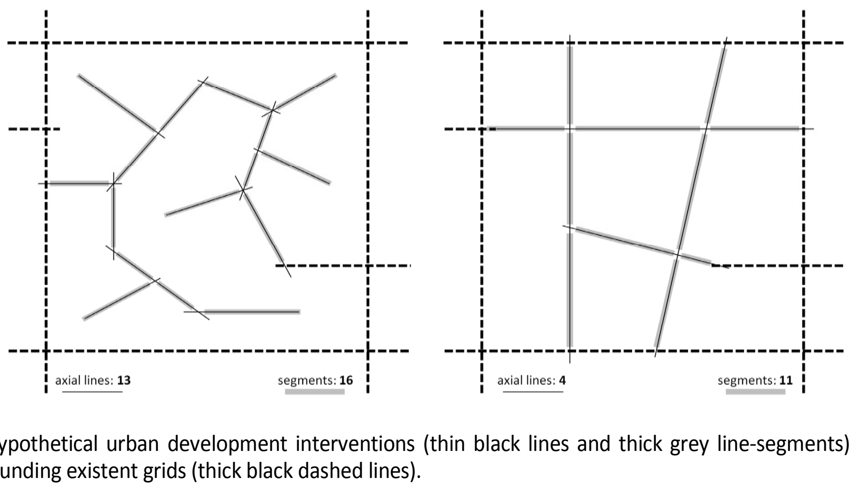

Figure 1: Two hypothetical urban development interventions (thin black lines and thick grey line-segments) and their immediate surrounding existent grids (thick black dashed lines).

By looking at Figure 1, it is possible to see that the simple count of the number of axial lines and segments making up each intervention is already a basic description of its morphology. The interventions of Figure 1 have approximately equal sizes (the one on the left is 8% larger), as defined by their road-lengths (the intervention on the left has a length of 35.2 and the one on the right 32.5; values in arbitrary but equal units). However, the intervention on the left has much more axial lines than the one on the right (13 and 4 respectively, thus more than the triple); whereas the difference in the number of segments is much smaller (16 and 11 respectively, a difference of approximately one third).

This happens because the intervention on the left is much more sinuous and curvilinear than the one on the right and has a greater number of internal street junctions (6), while the one on the right has slightly fewer internal junctions (4) but is much more linear and grid-like. However, the number of street-segments between junctions in both cases is similar. We see then, that the simple relation between the number of axial lines and segments composing each intervention is already providing some basic information about their geometry or composition. Another compositional variable immediately accountable for in Figure 1 is the total road-length of each intervention, which is directly related with their sizes (which are approximately equal, in this case).

Figure 2 shows the same two hypothetical interventions, but now with all the proposed morphological variables accounted for, and with the properties they measure graphically depicted. We have 5 compositional variables, namely: the total road-length of the intervention (“road_len”, in Figure 1); the number of axial lines (“ax_count”); the number of axial segments (“seg_count”); the ratio between the number of segments and axial lines (“seg_ax”); and the compactness of the internal blocks (if present) created by the intervention (“iCyc_APR”). All these compositional variables represent internal morphological properties.

In addition, we have 6 types of configurational attributes, divided into 9 independent variables, namely: the number of internal blocks created by the intervention (“int_Cyc”); the number of new sub- divisions of existing blocks created by the intervention (“ext_Cyc”); the ratio between the number of axial lines and the number of junctions (or links) the intervention has with the existing grid (“ax_ExtL”); the number of culs-de-sac, or dead-ends (“d_End”); the type and number of internal junctions (“int_T”, “int_X”); and the type and number of external junctions (“ext_T”, “ext_X”, “ext_I”). Thus, some configurational variables represent internal morphological properties, while others are aimed at the external relations that each intervention establishes with the existing grid. Figure 3 summarizes all the variables, organizing them by their compositional or configurational nature and by the type of property (internal or external) which they describe.

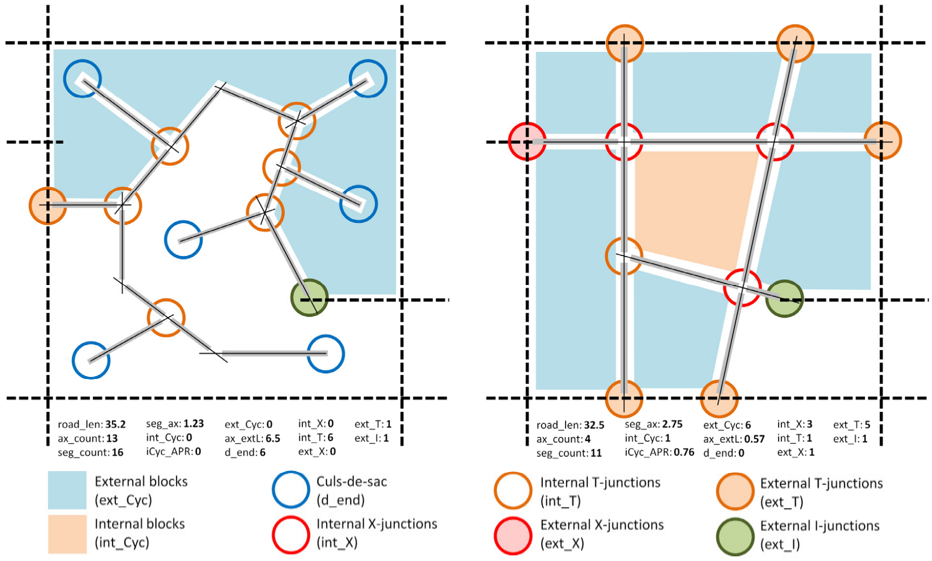

Figure 2: Graphical and numerical illustration of the compositional and configurational variables.

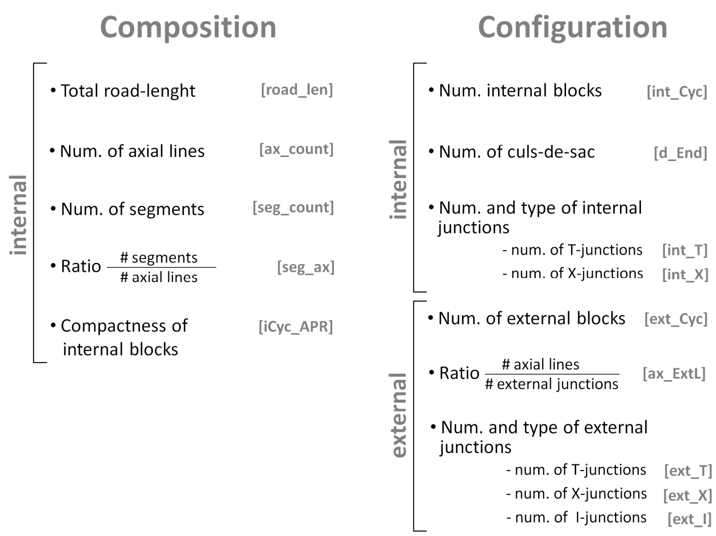

Figure 3: Summary of compositional and configurational variables.

4. Results

All these variables showed strongly right-skewed distributions with long tails, so all have been normalized through a logarithmic transformation. A simple outlier analysis (with the common criterion of identifying outliers above 2.55 or below -2.55 in the range of each variable's z-score) was performed in order to find extreme values, which are known to bias clustering procedures. No variables with severe outliers were found, except total road length and only regarding positive outliers. Additionally, this simple procedure was able to identify the entire network of metropolitan highways, which are a very different kind of grid interventions when compared to the myriad of incremental, non-planned and market-driven interventions constituting the bulk of the observed urban development. Large road infrastructures are products of centrally-planed decisions and are built in a non-incremental way (i.e. all of a sudden, in urban time terms). They should not be considered for the clustering procedure in any case, not only for technical reasons (i.e. statistical consistency) but also for conceptual reasons. Metropolitan growth in the Oporto region is fundamentally spontaneous and emergent, produced by private agents acting at the micro-scale, in a context of strong political and institutional fragmentation. Our aim is to find structural regularities among that emergent segment of urban growth in order to begin to understand the true morphological nature of contemporary city building.

After this first fundamental division of the grid interventions into incremental/non-planed and sudden/planned, the analysis of the variables' distributions suggested still one necessary and pertinent sub-division. The number of internal blocks showed a bi-modal distribution, with a very high peak at 0, followed by a much lower, quasi-normal distribution of values, meaning that there was a fundamental divide between not-outlying grid interventions: those with internal blocks and those without, hereafter called *cellular* and *linear*. Together with the overwhelming majority of linear interventions (87%), this divide constitutes an interesting first finding. The difference can be partially explained by the smaller size of linear interventions (mean road length 424m) when compared with the cellular type (mean 1142m), hence entailing a minor capital investment. But the differential is so great that one can say that internal block creation is indeed an exception, at least in the studied case. Figure 4 shows the geographical distributions of these three fundamental types of interventions, outliers, cellular and linear.

Figure 4: Geographical distributions of the database of street interventions of Oporto metropolitan region. From top to bottom and left to right: the full set, outlier, cellular and linear interventions.

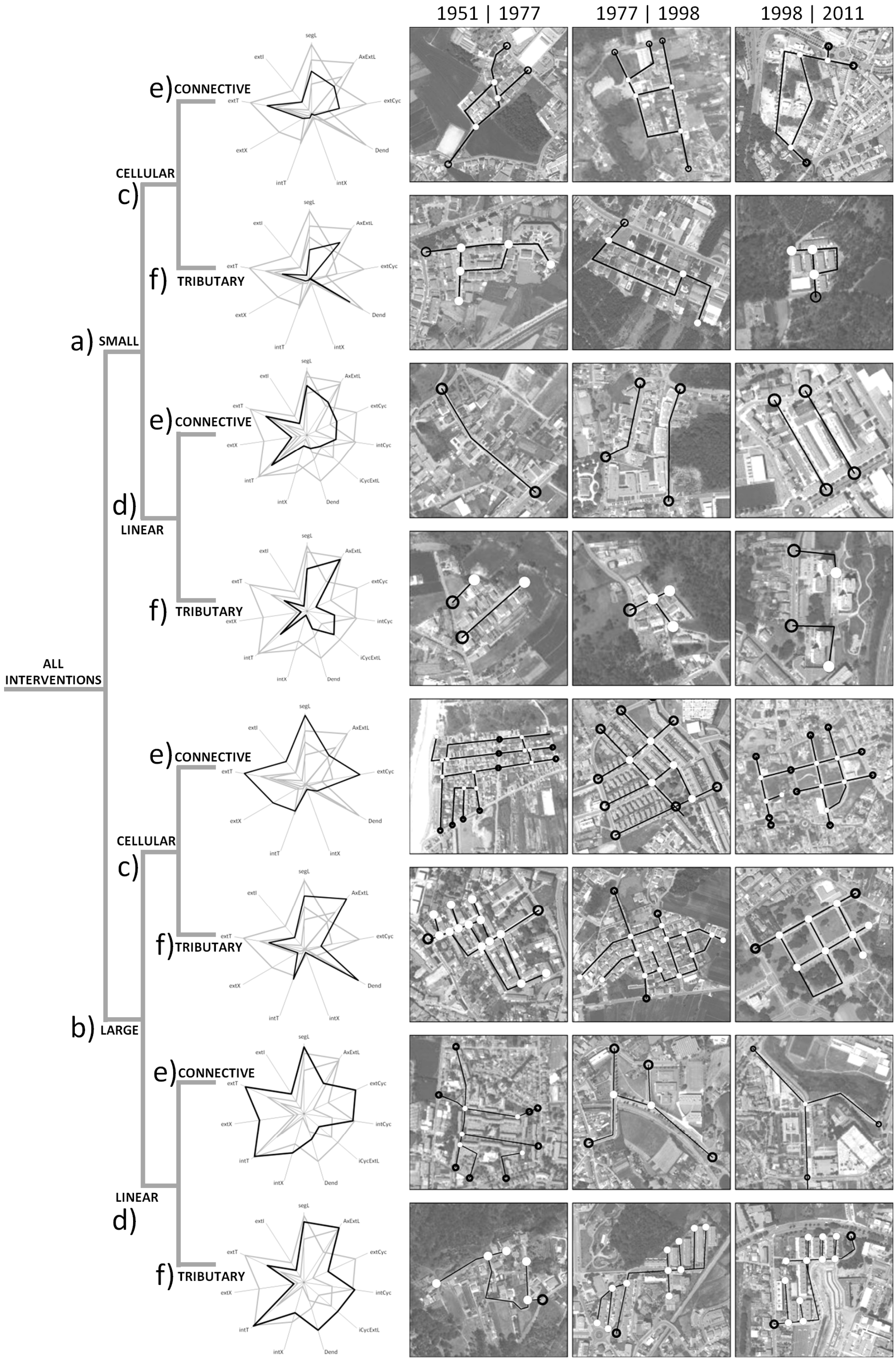

Figure 5: The proposed taxonomic nomenclature and the 8 morphotypes of street patterns identified. Each type is represented by its profile (i.e. the radar-plot depicting its mean values on each morphological dimension) and by its archetypes (i.e. the cases closer to each cluster's centroid) from each period of analysis.