Unsupervised Classification of Evolving Metropolitan Street Patterns

Miguel Serra
CITTA - Research Centre for the Territory, Transports and Environment. Faculty of Engineering
University of Porto
alvim.serra@fe.up.pt

Jorge Gil
Department of Urbanism, Faculty of Architecture, Delft University of Technology
j.a.lopesgil@tudelft.nl

Paulo Pinho
CITTA - Research Centre for the Territory, Transports and Environment. Faculty of Engineering
University of Porto
pcpinho@fe.up.pt

Keywords: unsupervised, classification, typomorphology, metropolitan

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 through 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. But there are also less obvious reasons. The problem seems rooted in the very definition of which morphological attributes to consider, in the degree of resolution of the desired classification, in the classification criteria themselves or even in the lack of consistency regarding the adopted terminology of types (Marshall 2005).

A recent paper by the second author (Gil, Montenegro 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, Montenegro 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 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 are adopted, based on exploratory statistical techniques (as cluster and dimension reduction analysis), which are known as unsupervised learning 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 main difficulty in any classification exercise is 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). Therefore, because of its structural role on the definition of future urban form and of its impact on urban functioning, the open space system ought to be seen as a preferential target of morphological study.

For these reasons, we have elected the open space system as our main object of study, constructing a diachronic axial model of Oporto’s metropolitan region, covering an area of 1600 Km² 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 dated from 2008. Subsequently, that axial map was edited over increasingly older cartographies, dating from 1950, 1975 and 2000, 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 intended to characterize and classify them regarding their morphologic constitution. This raises again the question of what attributes to consider. In contrast with built forms, typomorphologies of street systems are very scarce and normally don’t 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. 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. After a preliminary co-variance analysis, this set of variables was reduced to a total of 11 variables (see section 3).

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. 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 significative amount of variance in the data, and therefore the ideal number of clusters was attained.

### 3. Results

As mentioned before, the group of the initial twenty variables had strong correlations between them, showing that it could usefully be reduced. 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. The final group of 11 variables is the following:

- Total length of new streets (in meters) of each intervention [segL];
- Number of sub-divisions of existing blocks created by each intervention [extCyc];
- Number of new blocks created within each intervention (i.e. internal blocks) [intCyc];
- Proportion of new axial lines to the number of links with the existing grid (i.e. external links) each intervention has \([\text{AxExtL}]\);
- Proportion of internal blocks to the number of external links \([\text{AxExtL}]\);
- Number of culs-de-sac of each intervention \([\text{Dend}]\);
- Number of internal X-junctions (i.e. with four incident streets) \([\text{intX}]\);
- Number of internal T-junctions (i.e. with three incident streets) \([\text{intT}]\);
- Number of external X-junctions \([\text{extX}]\);
- Number of external T-junctions \([\text{extT}]\);
- Number of external i-junctions (i.e. external links that are simple continuations of the existing grid, with just one incident street) \([\text{extI}]\).

All these variables showed strongly right-skewed distributions with long tails, so all have been normalized by a logarithmic transformation. A simple outlier analysis (with the common criterion of identifying outliers above 3.5 or below -3.5 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 and motorways, 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-planned 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-planned 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 \textit{cellular} and \textit{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 really an exception, at least in the studied case. Figure 1 shows the geographical distributions of these three fundamental types of interventions, outliers, cellular and linear.

The cellular and linear sub-sets were finally explored separately using k-mean clustering. The study of the best number of clusters to extract suggested four clusters for each set. The composition of such clusters can be understood by their profile. These profiles (in our case, produced using the mean of each variable's values standardized to a [0,1] interval) show the cluster's composition plotted as polar graphs (Figure 2). Besides the basic separation into cellular and linear interventions (Figure 2, c and d), one can easily state the different morphological nature of each cluster. First, they differ in size (as expressed by \([\text{segL}]\) variable), with two clusters (both in the cellular and linear groups) scoring low in \([\text{segL}]\) (Figure 2, a) and other two scoring higher (Figure 2, b). Second, they differ in their external connectivity, i.e. the relations they establish with the existing grid, as expressed by the variables accounting for external junctions \([\text{extT}, \text{extX}, \text{extI}]\), number of subdivisions of existing blocks \([\text{extCyc}]\) and the ratios between the number of external links and number of axial lines or the number of internal blocks \([\text{AxExtL}, \text{iCycExtL}]\) (these last two attributes revealing higher external connectivity through lower values). Also, they differ in internal connectivity's terms, as expressed by the variables accounting for internal junctions \([\text{intT}, \text{intX}]\), by the number of internal blocks \([\text{iCyc}]\) (here only in the cellular case) and by the number of culs-de-sac, \([\text{Dend}]\). Hence, in terms of external and internal connectivities, we can indentify four well-connected clusters (Figure 2, e) and four others ill-connected (Figure 2, f), both in the cellular and in the linear groups. In Figure 2, we show all the clusters profiles together with an image of the corresponding archetype of each cluster for each period of analysis making clear their typomorphological similarity, despite of the evident differences - yet superficial, we believe - of the built forms.
The consistency of these results allows us to propose names for these morphotypes. Following the long tradition of biological taxonomy, in which each part of a species’ binomial Latin nomenclature indicates the attributes that make it unique, we propose a composite trinomial nomenclature based on the size, on the topological property of ciclicity (i.e. the presence or absence of internal blocks) and on the external and internal connectivities of the interventions. The nomenclature results from the eight possible combinations of the following hierarchy: Small / Large → Cellular / Linear → Connectives / Tributaries. These are the names given to the types in Figure 2. Following the hierarchy, one can find the name of each type (e.g., “small linear connectives” or “large cellular tributaries”).

The last step of our analysis, the inspection of the distribution of the discovered types through time, showed also interesting results. Figure 3 shows that the pace of grid growth (which can be seen as a proxy of urban growth) increases significantly from the second to the third period of analysis (1975-2000), corresponding to the construction boom that Portuguese cities suffered at the time. The overwhelming majority of linear expansions is also evident. Moreover, during the same period, there is a dramatic increase in the creation of outlier interventions (the long, planned, heavy-duty road infrastructures), which is a clear sign of the coalescence of Oporto region as a metropolitan area at that time.

But the graphs showing the frequency of the identified types in the several periods of analysis are particularly revealing. They show clear variations on the incidence of each type at each period of analysis. It becomes clear, against what one would think intuitively, that the morphological composition of metropolitan urban growth is definitively not always the same through time, with startling variations. There is a sharp increase in the production of small cellular tributaries during the 1975-2000 period (when this type clear surpasses the production of small cellular connectives), followed by an even more sharp decrease in the next period. Also, it is evident a strong increase in the production of large linear connectives together with an equally strong decrease of small linear tributaries during the 2000-2008 period.
Figure 2 - The proposed hierarchical nomenclature and the 8 morphotypes of street patterns identified.
Figure 3 - From top to bottom and left to right: production of cellular and linear cellular and linear types by period (values are estimated number of interventions per year in each period); production of outlier interventions by period (% of total length built along the study's time span); frequency of cellular interventions by period and frequency of linear interventions by period (values of % of total number of interventions of each type in each period).

4. Discussion

The method explored in this paper offers an objective classification of a large sample of contemporary metropolitan interventions that enables us to start describing and understanding the morphological composition of an extremely diverse urban territory. From this understanding, one can begin to establish new criteria for the spatial planning of the metropolitan area. The aim is to define a vision for the territory and its general development, based on objective measurements of the specific morphological reality. The method presented here points a path in that direction.

The unique morphological, geographic, social and climatic characteristics of each metropolitan territory, but also the socio-economic cultures of the stakeholders involved in the planning process, make it challenging to implement foreign urban intervention models that would require a tabula rasa approach and a profound change in attitudes. How to produce urban planning and design guidelines that are sensitive to the existing context?

One approach is to draw assessment criteria from existing local regulations and best practice guidance and to test the urban types herein identified against those. This way one can identify the best performing type(s), or measure the degree to which each type deviates from predefined goals. Because assessment criteria rarely relate directly to urban morphological descriptions, not least due to the lack of consistency of such descriptions, there is a danger of becoming trapped in a ‘business as usual’ situation in terms of the definition of a desirable urban morphological structure. For lack of explicit urban design and morphological guidance, the least problematic type, and not the best possible solution, becomes best practice and progress towards the overall goal is very slow.

The recommended approach is to select in advance reference cases of assumed quality and describe them using the 11 morphological parameters used here. From those descriptions one can extract a set of target values, or value ranges, for each parameter. The types identified and specific to the territory can then be assessed against the set of equivalent morphological target values, thus obtaining a detailed diagnostic for each type. This offers material to create recommendations and best practice guidance built on the existing local reality that is more explicit in urban morphological terms and eventually more challenging in terms of goals.
5. Conclusion

In this paper we have explored a method of unsupervised learning to identify urban morphological types in the Oporto metropolitan region, and analysed their occurrence through a period of 60 years. The interventions made to the street network during that period represent non-traditional metropolitan and suburban street patterns. Using this method we were able to develop a consistent classification, clearly distinguishing typological regularities and consistencies in an otherwise very complex and diverse morphological reality. We were also able to identify the predominance of certain types in specific time periods. These findings support an objective understanding of the evolution of the metropolitan region outside more consolidated and traditional urban centres. And in turn are a first step towards the development of design guidance and a typomorphology that is relevant to the contemporary practice.

6. References


---

i 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.

ii The axial map is a space syntax modelling technique that represents the entire public space grid as a network of straight lines describing its basic geometry. See (Hillier 1996) for details on axial mapping.

iii 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.

iv Military cartographies from the Portuguese Military Geographic Institute (IgeoE) at the 1:25000 scale.

v But we do not go as far as to suggest such a nomenclature in Latin.