

#85

MULTIDIMENSIONAL STUDY OF URBAN SQUARES THROUGH PERIMETRAL ANALYSIS:

Three Portuguese case studies

JOÃO V. LOPES

ISCTE – Instituto Universitário de Lisboa
jvls@iscte.pt

ALEXANDRA PAIO

ISCTE – Instituto Universitário de Lisboa
alexandra.paio@iscte.pt

JOSÉ NUNO BEIRÃO

Faculdade de Arquitectura da Universidade de Lisboa
jnb@fa.ulisboa.pt

LUÍS NUNES

ISCTE – Instituto Universitário de Lisboa
luis.nunes@iscte.pt

CAMILLA PEZZICA

Università di Pisa
camilla.pezzica@gmail.com

ABSTRACT

This paper addresses one of the most symbolically and socially meaningful elements of the public open space: the urban square (Portuguese: praça). Besides their urban centrality, these spaces' potential for liveliness depends on multiple factors and their identity as a place may only be grasped by formal methods able to embrace their latent complexity and address the multi-scale and multivariate correlations of factors that defy human cognitive capabilities. This paper presents a synchronic multidimensional analysis of three Portuguese historic squares: Largo da Oliveira, Praça de São Tiago (Guimarães) and Praça do Giraldo (Évora), representative of the national historic heritage.

KEYWORDS

Public Open Space, Urban Squares Morphology, Public-Private Interface, Syntactic Analysis, Multidimensional Analysis

1. INTRODUCTION

The word *square* by which we translate the Latin – *platea* – derived terms (piazza, plaza, praça, piață) is used to identify an urban space of exceptional character in the morphology of the public open space. Although usually denoting a designed formal space, in this context we use the term in a broader sense including the informal formations that share a set of particular morphologic and cultural semantic-symbolic values, rooted in history, that make them more than mere

urban voids, pedestrian paths or focal points. In Portuguese, this concept is embodied in other toponymies like *largo*, *terreiro* or *rossio*.

The liveliness and the success of interventions in these spaces is defined not only by their centrality but also by local features like the permeability of their defining planes, besides notions of comfort, security, use and appropriation *affordances* of a micro urban morphology (Habraken and Teicher, 1998; Gehl, 2011). To address this complexity, a methodology is presented which, collecting contributions from the disciplines of urban morphology and site analysis, resorts to multivariate statistical analysis and inductive patterns search techniques in large datasets by data mining.

The notion of interface, as the potential of a boundary to work as a threshold, is at the very basis of the configurational and morphological analysis, as well as at the basis of the design of the built environment itself (Palaiologou et al., 2016). By discretizing the boundary space of squares, we delve in its potential as interface between architecture and public open space by observing how an a priori set of heterogeneous attributes manifest themselves along their perimeters. The selection of attributes takes an agnostic attitude melding analytical **themes** (morphologic, environmental, syntactic and functional), **scales** (global and local) and **standpoints** (cognitive-unbounded and perceptive-bounded). This strictly data-driven approach is methodologically important in the sense that the proposed inductive data mining methodology is eminently theory agnostic, local (conclusions belong to the data set analysed), and resorts to case-based reasoning (Ahu et al., 2011).

The two cities where the case studies are located (Guimarães and Évora) are paradigmatic examples of the north-south and coast-interior dichotomies of the Portuguese mainland territory, entrenched between *Atlantic and Mediterranean cultures* (Ribeiro, 2011). Nowadays, tourism and the location of university centres are crucial factors for development of these two medium-sized cities, fostered by UNESCO classification of their historical centres. The three case studies are squares of essentially informal genesis, located in the cities' consolidated historical fabric within well-preserved medieval walls. Nowadays they are carefully preserved symbolic and leisure spaces, located in mainly pedestrianized neighbourhoods where car circulation is not disruptive of their tourist and civic event activity. Besides the considerable difference in area between Praça do Giraldo (Évora) and the squares in Guimarães (Largo da Oliveira and Praça de São Tiago), their historic function was distinct. The Praça do Giraldo (Figure 1) is an example of an ancient *rossio*: a multifunctional space contiguous to the external side of a city gate (in this case of the disappeared roman city wall), and considered a hallmark of the Portuguese medieval urbanism. These spaces were focal irradiation points for subsequent urban expansions



Figure 1 - From the top to down: Évora city center, Praça do Giraldo planimetry and photos, Guimarães city center. Praça de São Tiago (b.) and Largo da Oliveira (c.) planimetry and photos. a. and d. in Silva (2009).

(Coelho, 2008). The Largo da Oliveira has a more formal and symbolic character, functioning as the forecourt of the church that dominates and gives its name to the place, while Praça de São Tiago is a classic medieval market square (previously named Praça do Peixe - Fishmarket Square) surrounded mainly by typical houses. Together, these two squares, form an interesting system connected through the archway of the open floorplan of the elevated medieval city hall building. In the presented investigation, they are studied as separated spaces, where each one is part of the context of the other (Figure 1).

2. OBJECTIVES

The main objectives of the presented investigation are: (i) to describe and classify squares' perimetral space, through a generalizable non-supervised analysis method, which naturally groups sampling points around their perimeter, taking into account a set of heterogeneous attributes simultaneously; (ii) to determine the attributes that better explain functional distribution and morphological diversity along the square perimeter; (iii) to assess how these correlations of factors vary from case study to case study; and finally (iv) to observe how they relate with site surveys and empirical knowledge of those spaces, from the perspective of urban morphology and human behaviour (Gehl, 2011).

3. BACKGROUND AND MOTIVATION

While theory integration between space syntax and urban morphology is mainly carried out at the urban macro scale, e.g. Place Syntax (Stahle et al., 2005), most of its successful and well known practical applications lay at the intermediate scale, between architecture and the city (e.g. Trafalgar square). This follows the schism between planning and design, hindering the development of relations with the typo-morphologic and architectonic stands on the city micro morphologic analysis, that could promote a more sustainable evidence-based urban design.

As pointed out by Palaiologou et al. (2016), the primacy of movement in space syntax theory as the main space usage, rooted in the axial analysis and concepts like *natural movement* (Hillier et al., 1993) or *movement economy* (Hillier, 1996), came at the price of privileging the generic city *over the complex and culturally specific street-building interface*, the lived space in which *individual action becomes social practice*. Problematising this idea of the street-building interface and claiming for a *recalibration of scales*, the authors make a stand on the role of *constitutedness* to shape its own field of encounters, at the threshold of architecture and urban scales, and make a specific contribution to the virtual community concept (Hillier, 1996). One could say that the regularity, and the character of the interface porosity, culturally shape the dynamics of the public space.

The syntactic exploration of the relations between street interface and liveliness has been carried out by authors like van Nees (2007) and Beirão and Koltsova (2015), driven by space syntax seminal ideas like the formalization of the interface relation between private and public realms (enclosed – *carrier space*) and the *ideographic language* of spatial arrangements, the morphogenetic model of *elementary cell* aggregation that gives rise to the *beady ring* settlement structure, or the *constitutedness* of space and the interface map (Hillier and Hanson, 1984). At this micro analysis scale the linking between syntax, morphology, typology and architecture becomes natural and the (tridimensional) form and specific aspects gain prominence over the generic and location ones.

The relational or configurational nature of urban space syntax, as the study of the field of possibilities over a discretized system, has the partitioning of the continuum of the public open space at the heart of its theory foundations. Lacking the formal clarity of the axis, the convex space or the isovist (Benedikt, 1979), the translation of urban morphology fundamental elements, like "square", to a formal language is an *ill-defined problem*, even if stripped of all its semantic values. However, perceptive and cognitive explorations may cast some light on this *threshold* problem (e.g. Lynch, 1960; Thiel, 1961; Turner, 2003).

Accepting the square as an urban element per se authors like Campos (1997) and Campos and Golka (2005) investigate the relationship between patterns of use, network configuration and visibility by studying the penetration of axial lines and the effect of visual fields on the space of London squares. Cutini (2003) studies Tuscan historic squares (*piazzas*), focusing on the relationship between centrality, potential of enclosure, and extension, proposing a new compound visual graph analysis (VGA) index that depicts the hierarchy and performance potential of convex spaces to work as squares.

The classic methods of urban morphologic and syntactic analysis, usually limit the number of variables, to respond to human cognitive and perceptive limitations, restricting the simultaneous expression of features that give spaces their uniqueness. Serra, Gil and Pinho (2013) compile the shortcomings of traditional typomorphological approaches, namely their time-consuming methods, which also restrain the quantity of examples and dimensions considered, and their theoretical partiality.

The identified deficiencies can be handled using new computational methods that allow for multidimensional analysis and typological classifications based on data mining. This interdisciplinary subfield of computer science can be understood as the practical application of machine learning, itself a subfield of Artificial Intelligence (AI) dealing with automatic learning from data (Witten and Frank, 2005). The main objectives of data mining, also known as knowledge discovery in databases (KDD) (Han and Kamber, 2001), are knowledge extraction, prediction and hypothesis generation from data, by favouring an inductive approach. This approach is in opposition to confirmatory techniques, which require *a priori* hypotheses formulation, and restrict hidden information discovery (Miller and Han, 2001). The automatic learning may be divided in three methods: (i) unsupervised, when there is no *a priori* labelled data (e.g. used in segmentation, clustering and dimensionality reduction); (ii) supervised, when there is a priori labelled data (e.g. used in predictive models like regression, classification and rule induction); and semi-supervised, when data is partially labelled. Nowadays these techniques are widely applied in many fields of science, engineering and business. When spatial data is involved, like in geographic information systems and geoscience, it is designated by spatial data mining (Demsar, 2006).

Within urban morphological studies, data mining supports analyses at different scales: from Laskari et al. (2008) study on urban identity through quantifiable attributes on blocks' shape at district level, to street patterns in metropolitan areas. Gil et al. (2012), in an unsupervised classification of the urban fabric of two neighbourhoods of Lisbon, focusing on street and block elements, mention the possible integration of these techniques in design. Ahu et al. (2011) explore the potential of supervised learning as a methodology of knowledge discovery in micro urban feature analysis on the historical fabric of a neighbourhood in Istanbul. Chazar and Beirão (2013) point the potential of extending the methodology to deal with non-physical qualities, leading to a better understanding of the public open space morphology. Multidimensional analysis of the latter is rarer: Laskari (2014) analyses the blocks' residual void space in a neighbourhood of Athens, through a set of 13 properties and by different clustering methods. The work of Hanna (2009) on the principal component analysis of graph spectra and self-organizing maps (SOM), and of Al-Sayed (2013) with supervised artificial neural networks (ANN), analysing both design process and the configuration of urban grids, are examples of the application of related techniques in the field of syntactic analysis.

4. DATASETS AND METHODS

Resorting to a previous investigation of the authors on a general method for the multiscale and multivariate analysis/classification of public open space (Lopes et al., 2015), it was chosen a subset of the proposed indicators, that seemed significant to the current study.

#	Att. Name	Unit	Type	Theme	Att. Description
1	SpID	-	Ordinal	ID	Sp Point ID
2	x	m	Real	Geo location	Sp point localization x value
3	y	m	Real	Geo location	Sp point localization y value
4	z	m	Real	Geo location	Sp point localization z value
5	PerimCon	-	Integer	Connectivity	Sp point connectivity
6	PerimCluster	-	Real	Connectivity	Sp points Clustering coefficient
7	IsovArea	m ²	Real	Isovist	Area of the Sp point 2D isovist
8	IsovPerim	m	Real	Isovist	Perimeter of the Sp point 2D isovist
9	IsovMaxRad	m	Real	Isovist	Maximal radial of the Sp point 2D isovist
10	IsovComp	-	Real	Isovist	Compactness of the Sp point 2D isovist
11	SpIsovOcl	m	Real	Isovist	Occlusivity of the Sp point 2D isovist
12	VGAInt	-	Real	VGA	Sp Point Visual IntegrationHH Rn VGA value
13	VGAControl	-	Real	VGA	Sp Point Visual Control VGA value
14	VGAControlla	-	Real	VGA	Sp Point Visual Controllability VGA value
15	VGACluster	-	Real	VGA	Sp Point Clustering Coefficient VGA value
16	Fah	m	Real	Facade	Facade height at the Sp point
17	VFahMax	m	Real	Facade	Maximum facade height of the visible Sp points
18	VFahSD	-	Real	Facade	Visible fcade height standard deviation
19	VFfaArea	m ²	Real	Facade	Visible facade area
20	FaOr	deg	Real	Facade	Facade orientation (y=North=0 degrees)
21	Max3DDist	m	Real	Facade	Maximal distance to sampled facades (eye h=1,5m)
22	VSkyF	%	Real	Environment	Sample Point visible Sky Factor
23	Solar	-	Integer	Environment	Number of Sp point annual insolation hours
24	VPpQty	-	Integer	Constitutedness	Quantity of visible Pp points
25	MeanDist	m	Real	Constitutedness	Distance to the nearest Pp point
26	MinDist	m	Real	Constitutedness	Distance to the nearest Pp point
27	NearUse	-	Nominal	Use	Land use nearest visible Pp point
28	VMonum	%	Real	Use Visibility	Percentage of visible Pp monumental use
29	VPublicServ	%	Real	Use Visibility	Percentage of visible Pp public service use
30	VHotelFood	%	Real	Use Visibility	Percentage of visible Pp hotel and catering use
31	VComerc	%	Real	Use Visibility	Percentage of visible Pp commercial use
32	VResid	%	Real	Use Visibility	Percentage of visible Pp residential use
33	VOther	%	Real	Use Visibility	Percentage of visible Pp other uses
34	BVol	m ³	Real	Urban Index	Built volume at Sp point location

Table 1 - Attribute list, metadata and description.

They comprise 34 attributes, divided in ten themes, representative of morphological (space shape/built form), syntactic and visibility (connectivity/VGA/isovists), environmental (topography/solar exposure) and functional (permeability/uses) traits (Table 1). Attributes can be global or local, and either the expression of a property at a sampling location or aggregations of values as presented to visibility at that point. Consequently, besides isovists, it is taken another approach on visibility (akin to VGA) as a discrete connectivity graph of the sampling points. Recording the data structure of that graph one can more easily query on connected sampling points' attributes and aggregate their values in a meaningful manner that may represent the way they express themselves to visibility.

The method implies the definition over the square's models of (i) a square working perimeter and (ii) two sets of points:

(i) The square working perimeter

As a generic rule the perimeter of the square is the minimal polygon (not necessarily convex) enclosed by its built frontage edges and defined by the horizontal and vertical projections of their main façades. As this is an ambiguous description, some assumptions are made clear: existing public ground floor galleries are considered but do not define the perimeter; accessible fenced public spaces that are adjacent to the square are considered part of the square; built façades that partially face the square are considered as an indivisible unit belonging to the square. The proposed perimeters are working perimeters and other definitions may be used, but an effort is made to keep consistency across the case studies.

(ii) Two sets of points

a) Physical permeability points (Pp)

Points located at the centre of ground floor openings (the interface between public and private) in and around the square (with a buffer of approximately 50 meters). These points record the use they serve. Other building openings for lighting, ventilation or views are not considered in this study.

b) Sampling points (Sp)

Points evenly distanced 1,0 m, approximating human scale, numbered and ordered along the squares' defined perimeter. They work as a neutral referential from which tables are built. The altimetry of these points will depend on the attribute being recorded.

The method implies an initial global configurational analysis of the expanded urban system where the squares are embedded. In the presented case studies, constructed upon Guimarães and Évora municipal georeferenced official cartography, an axial and segment analysis at the urban perimeter scale, and a VGA analysis at their historical centres were conducted (Figures 2 and 3 respectively). The importance of the three squares in their urban system can be depicted on the syntactic maps, and made clearer by mapping Cutini's *interaction index k* (Cutini, 2003) of the set of squares of those central urban areas (Figure 3).

Subsequently, the focus is directed to urban micro-morphology analysis and local properties, as they are expressed and recorded along the perimeter of the square (Figure 4), using the methodology described above, and in the use of data mining techniques for the visualization and exploration of the datasets so produced.

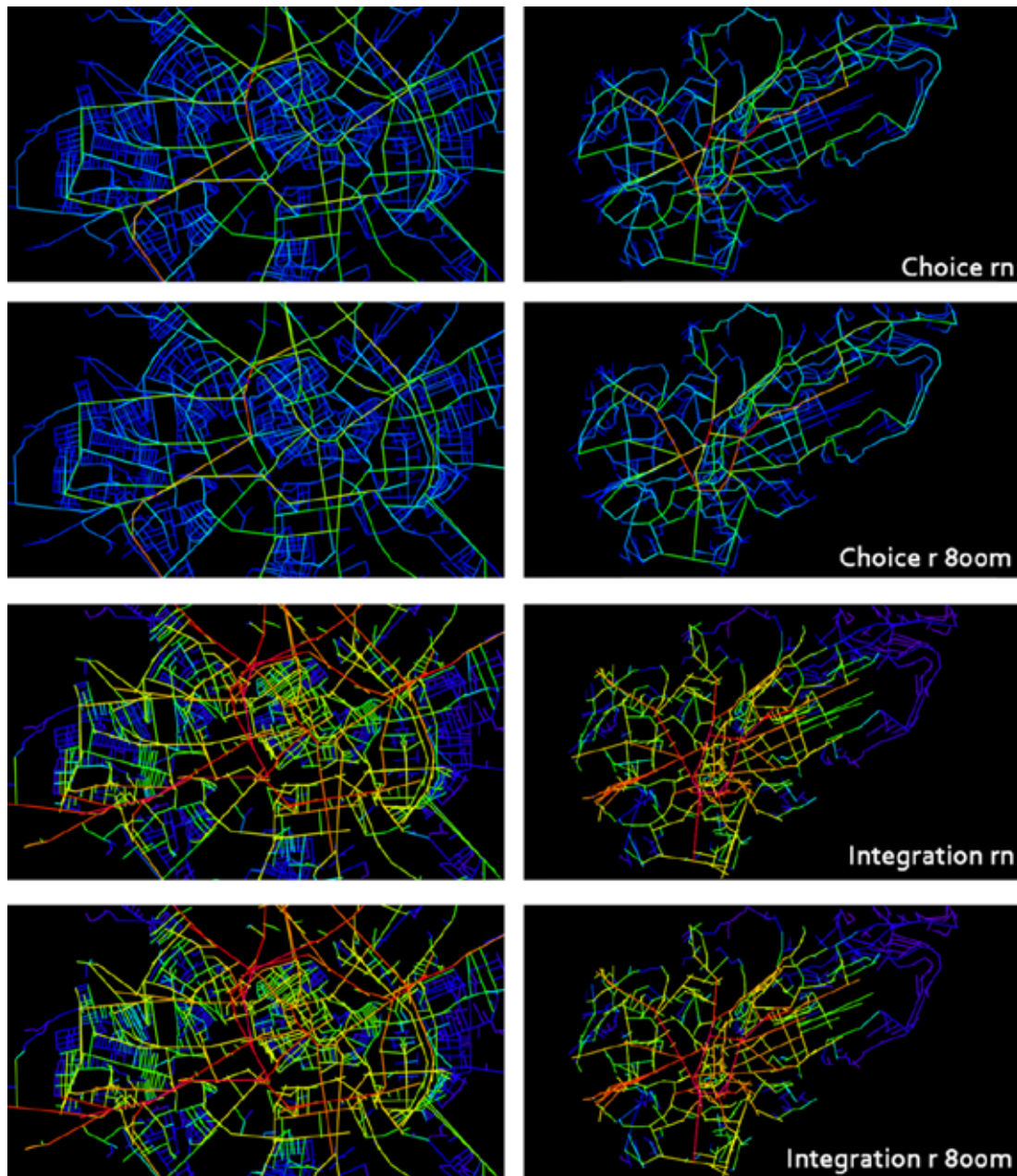


Figure 2 - Évora (left) and Guimarães (right) segment analysis.

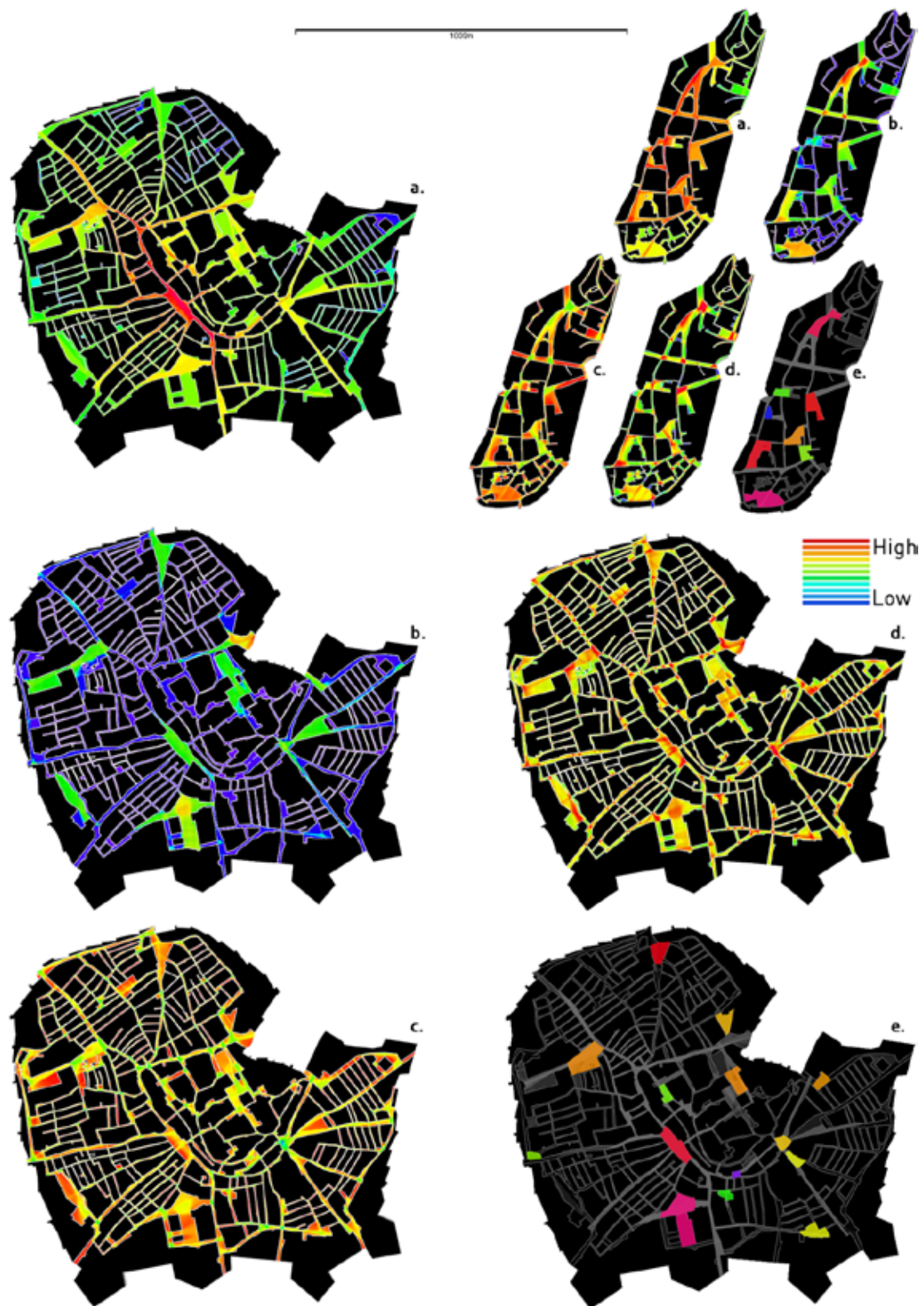


Figure 3 - Évora and Guimarães VGA analysis (1m grid resolution): a. Integration HH, b. Connectivity, c. Clustering Coefficient, d. Control, e. Interaction Index (Cutini, 2003).

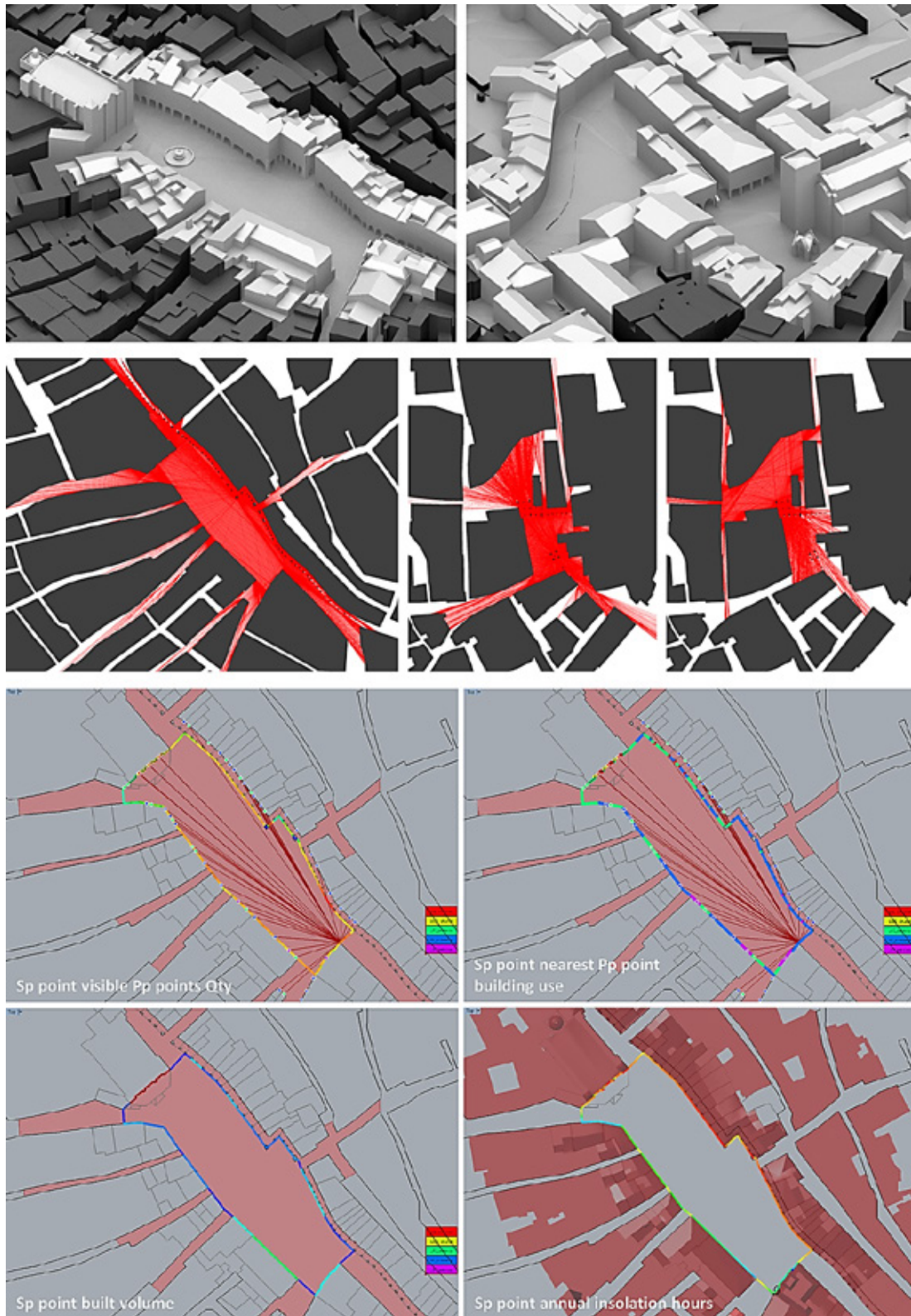


Figure 4 - From the top: 3D mass models; overlapping perimetral isovists (from the left: Pr. Giraldo, L g. Oliveira; Pr. São Tiago); mapping of uses and attributes (Pr. Giraldo).

The first step is to get a sense of the data and validate data quality. Besides data exploratory techniques and descriptive statistics, the mapping of the information is essential for checking out errors and for dealing with outliers. This pre-processing phase includes the normalization of the numeric attributes as identical weight is given to all of them.

Following the objectives, the data is modelled by exploring mainly unsupervised learning techniques: (i) dimensionality reduction, by principal components analysis (PCA), and (ii) partitioning clustering.

PCA allows the determination of a smaller set of artificial variables (the components, formed by a linear weighted composition of the original attributes) that summarize the original data with minimized loss of information. There are as many components as attributes, ordered by their contribution to variance. The first few ones normally attend for the larger amount, allowing to work with fewer dimensions and to produce clearer representations in 2D plots. By analysing the composition of those components, one can filter the most important attributes for further analysis (Witten and Frank, 2005).

Clustering is an unsupervised classification process that assigns objects to groups (clusters), so that the objects of each group are more similar to one another than with the objects of other groups. There are many clustering algorithms; here it is used K-means with the Euclidean distance function, used to express dissimilarity, whose results are easier to interpret. Clustering discovers natural groups of objects or variables, identifies extreme and archetypal examples (based on the distance to the cluster centre: the prototype) and suggests interesting hypotheses about relationships (Witten and Frank, 2005). K-means is here implemented by a derivative algorithm, X-Means, that determines the correct number of clusters based on a Bayesian Information Criteria (BIC) partition heuristic, and balancing the trade-off between precision and model complexity (Pelleg and Moore, 2000).

To assess strong predictors on the location of uses along the square perimeter, Predictor Screening (SAS, 2009) is used; this is a supervised learning method of screening many attributes for their ability to predict an outcome, which is based on random decision trees (bootstrap forest partitioning).

The workflow (Figure 5) implies the functional surveying and mapping of Pp points in the vicinity of the sites, and the creation of 2D and 3D (topographic and volumetric) CAD models suitable for the scale and analysis at hand. The large scale syntactic analyses are produced in DepthmapX (Varoudis, UCL), and the morphologic local analyses using algorithm design software (Rhino/Grasshopper and its plugins Ladybug/Honeybee for environmental analysis) as a mean of gaining control over the analytical tools (Figure 4). For each attribute, data is collected by a specific script and converted in a vector on a multidimensional table. Here simple spreadsheets are used since the case studies are few, but the ideal workflow would include data storage in a central spatial object-relational database (like PostgreSQL extended with PostGIS). For exploratory data analysis and data mining we resorted to the student versions of Rapidminer (RapidMiner Inc.) and JMP (SAS Inc.) and, for visualization and mapping, to QGIS and Rhino/Grasshopper. Data visualization is essential for error checking and for the critical interpretation of the results.

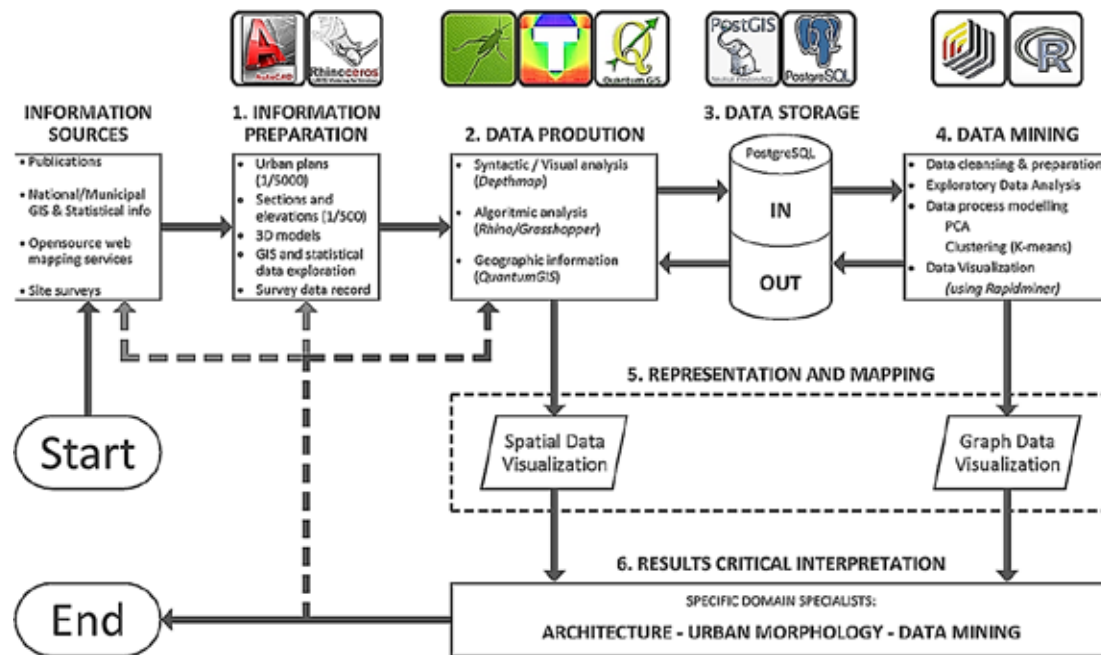


Figure 5 - Workflow diagram illustrating the interactive nature of the ideal process.

5. RESULTS

The statistical summary in Table 2 (left side) shows aggregated values of the raw data. Some blocks of values appear: Praça do Giraldo concentrates the generality of the highest mean values, as it is the larger space and data is unscaled, and Largo da Oliveira has the lowest attribute mean values related with isovists area, connectivity and VGA, while Praça de São Tiago has the lowest attributes related with isovists perimeter and façades. This last case study, although with high standard deviation, has the highest mean and maximum values of isovist compactness and visible Pp points. These attributes together with the environmental ones, make these Guimarães case studies diverge. The distinct functional character of the case studies is also depicted in the mean values of attributes related with proximity and visibility to Pp points: Praça do Giraldo with commerce, Largo da Oliveira with leisure and monumental buildings and Praça de São Tiago with residential use. Maximum and minimum values are related to squares' size, the detail of the defined perimeters and the type of attributes (NearUse appears as a 0 or 1 binominal attribute).

The pairwise correlation analysis on Table 2 (right side) shows the pairs of attributes correlated with an absolute value above 0.7, which are mainly the known cases of correlations between VGA, isovists and connectivity, but cross relations with non-syntactic attributes can also be depicted. The negative correlation on the location of residential/leisure and public service uses is also highlighted. Two pairs of attributes have values above 0.95, across all the case studies, so a selection is made on redundant attributes (IsovPerim and VFaArea are dropped off the analysis).

The data presents some outliers but, as they don't are the result of measurement errors but of idiosyncratic aspect of the sites, the detail of analysis and of the selected perimeter, they are not discarded.

In parallel with clustering, a PCA analysis is conducted, revealing that, in all squares, for an explanation of 95% of the variance of the data about 15 components are necessary, which means a general weak correlation. The first two components explain an average value of 36% of variance and are used in the scatter plot visualization of the clustering results (Figures 6 and 8).

Attribute	Mean			Std Dev			Minimum			Maximum				
	PrGirald	LgOliveira	PrSãoTiago	PrGirald	LgOliveira	PrSãoTiago	PrGirald	LgOliveira	PrSãoTiago	PrGirald	LgOliveira	PrSãoTiago		r
1 PerimCon	222.6	95.8	119.7	64.2	34.5	43.1	42.0	20.0	35.0	293.0	155.0	195.0	PerimCon	0.59
2 PerimCluster	0.8	0.7	0.7	0.1	0.1	0.1	0.3	0.3	0.2	0.9	1.0	1.0	IsovPerim	0.59
3 IsovArea	4420.3	1541.2	1843.3	1091.9	751.0	656.0	216.9	241.1	0.0	5989.2	4053.4	2745.7	IsovArea	0.52
4 IsovPerim	862.5	705.2	473.5	250.8	300.5	214.7	169.6	124.7	2.5	1776.4	1701.8	1107.1	IsovArea	0.51
5 IsovMaxRad	157.4	108.0	92.7	46.7	65.3	25.2	38.5	38.8	19.8	336.2	363.1	142.8	IsovMaxRad	0.78
6 IsovComp	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.3	0.6	IsovArea	0.78
7 IsovOcl	554.2	543.9	289.5	209.5	266.8	196.5	126.7	60.8	2.5	1426.1	1518.0	938.9	IsovPerim	0.78
8 VGAInt	6.4	3.7	3.7	0.6	0.3	0.3	4.8	3.0	3.1	7.5	4.2	4.5	IsovArea	0.78
9 VGAControl	1.1	0.9	1.0	0.3	0.3	0.3	0.3	0.3	0.3	2.0	1.5	1.9	IsovPerim	0.78
10 VGAControlla	0.4	0.2	0.3	0.1	0.1	0.1	0.2	0.1	0.1	0.4	0.4	0.4	IsovArea	0.78
11 VGACluster	0.8	0.7	0.8	0.1	0.1	0.1	0.7	0.5	0.5	1.0	1.0	1.0	IsovPerim	0.78
12 FaH	11.5	9.8	7.3	5.8	5.1	3.4	0.0	0.0	0.0	22.2	22.5	13.5	IsovArea	0.78
13 VFahMax	21.4	20.0	12.3	1.9	4.6	1.4	14.1	9.8	9.0	22.2	22.5	13.5	IsovPerim	0.78
14 VFahSD	5.7	4.5	3.1	0.5	0.9	0.9	4.2	1.4	0.5	7.1	6.9	5.2	IsovArea	0.78
15 VFahArea	2585.8	929.8	908.4	752.4	376.1	339.8	506.0	157.5	233.4	3492.6	1610.1	1553.8	IsovPerim	0.78
16 FaOr	197.0	164.5	148.4	101.0	99.4	105.8	1.6	4.2	0.6	350.0	313.1	359.1	IsovArea	0.78
17 Max3DDist	100.3	49.8	61.1	23.1	11.4	20.3	35.8	16.0	17.6	131.1	68.9	98.0	IsovPerim	0.78
18 VSkylF	33.1	24.8	30.7	5.8	6.6	7.0	15.3	3.6	0.0	48.8	34.7	42.1	IsovArea	0.78
19 Solar	1463.2	959.6	1146.2	676.7	528.7	811.6	0.0	0.0	0.0	2808.0	2085.0	2429.0	IsovPerim	0.78
20 VPPQty	35.7	34.6	45.7	9.3	16.3	15.3	2.0	2.0	12.0	62.0	71.0	76.0	IsovArea	0.78
21 MeanDist	54.4	34.6	35.3	11.0	5.8	9.2	20.2	16.1	12.3	84.4	55.5	60.5	IsovPerim	0.78
22 MinDist	7.5	4.9	2.5	10.2	5.5	4.0	0.0	0.0	0.0	41.7	40.3	33.4	IsovArea	0.78
23 Near=Monum	0.1	0.2	0.0	0.3	0.4	0.0	0.0	0.0	0.0	1.0	1.0	0.0	IsovPerim	0.78
24 Near=PublicServ	0.1	0.1	0.1	0.3	0.2	0.3	0.0	0.0	0.0	1.0	1.0	1.0	IsovArea	0.78
25 Near=HotelFood	0.1	0.3	0.3	0.3	0.5	0.5	0.0	0.0	0.0	1.0	1.0	1.0	IsovPerim	0.78
26 Near=Commerc	0.4	0.1	0.1	0.5	0.3	0.3	0.0	0.0	0.0	1.0	1.0	1.0	IsovArea	0.78
27 Near=Resid	0.2	0.3	0.5	0.4	0.5	0.5	0.0	0.0	0.0	1.0	1.0	1.0	IsovPerim	0.78
28 Near=Other	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	1.0	1.0	1.0	IsovArea	0.78
29 VMonum	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.5	0.0	IsovPerim	0.78
30 VPublicServ	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.3	0.1	0.4	IsovArea	0.78
31 VHotelFood	0.1	0.3	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.3	0.8	0.4	IsovPerim	0.78
32 VCommerc	0.4	0.1	0.1	0.1	0.1	0.0	0.2	0.0	0.0	0.6	0.6	0.2	IsovArea	0.78
33 VResid	0.3	0.5	0.6	0.1	0.1	0.1	0.1	0.2	0.0	0.6	0.9	0.7	IsovPerim	0.78
34 VOther	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.1	IsovArea	0.78
35 BVol	3465.0	3962.2	710.5	4665.0	4636.5	672.9	0.0	0.0	0.0	17726.1	11749.5	4622.3	IsovPerim	0.78

Table 2 - Left side: statistical summary of non-spatial attributes (note: nominal attribute NearUse is disaggregated); Pr.Girald: N= 352; Lg.Oliveira: N= 249; Pr.São Tiago: N= 280. Right side: absolute values of correlation $|r| \geq 0.7$; in bold: discarded attributes; in underlined italics: attributes not directly related to isovists or VGA; in box: pairs that appear in all squares.

The clustering process of the Sp points is made in the non-positional attributes (x, y, z), normalized using z transformation (or statistical transformation, where the final data has a normal distribution with zero mean and unit variance) and by the X-Means algorithm. Three clustering experiments are performed: first, a local-individual approach to each square Sp points dataset separately; second, a clustering of the clusters' prototypes of all the squares; and third a global-joined analysis of all the Sp point datasets from the three squares simultaneously.

The characterization of the clusters (the most distinctive attributes of their class) is made by the comparison of the mean value of the attributes of each cluster (its prototype or archetype), with the mean of all the examples in the dataset, which, with the applied normalization, is zero by definition. So, high and low values indicate values significantly above and below the mean values of the whole dataset (in standard deviations). Figure 6 shows a table of the set of attributes, by square, that include the top five ranking of their normalized absolute value, and the plotting of examples in the first two principal components; and Figure 7 shows their mapping on site. The numbering of the clusters has no special meaning and they cannot be compared across case studies by their designation.

Praça do Giraldo. For this square, there are two major clusters (C1 and C2) whose values concentrate around the global mean (the origin of the coordinates in the scatterplots). Cluster C1 collects all the highest values of solar exposure, VGA control and minimal distance to Pp points. Façade orientation takes low values (good solar south-west exposition), with contrary signal to cluster C2, which concentrates all the highest values of perimetral connectivity, isovist

occlusivity and VGA clustering coefficient, meaning higher degree of convexity and jaggy visual fields simultaneously (Sp points face the northern gallery). Clusters C₃ and C₄ gather the extreme examples with values way off below the median regarding connectivity, isovists and VGA, especially isovist area, global VGA integration and controllability as their Sp points are located in a segregated area of transition between street and square, with the highest values of visibility to residential use. Moreover, C₃ differs from C₄ by the inverse relation in the attributes dominated by the nearby church: VMonum and VFahSD, which records the skyline variability.

Largo da Oliveira. In this case study the attributes related with isovists, VGA, proximity to commerce and visibility to monumental buildings separate clusters C₁ and C₂ from C₃ and C₄. Extreme high values of the last two features, as well as of visual axis, isovist area and occlusivity, with low value of visible Pp points and related perimetral connectivity dominate cluster C₁. The latter are attributes that disentangle C₂ from the other clusters, which collect all the high values in the syntactic measures, corresponding to the more central stable space. Clusters C₃ and C₄ locate mainly in the East side of the square, due more to morphologic and functional aspects than to the value of façade orientation (like in Évora). The proximity to public services and the built volume of the surroundings, the old convent and church, dominate C₃ and distinguish it from C₄, what is reinforced by the opposite values of visibility to restaurants and cafés doorways, located across the square, and the skyline silhouette.

Praça de São Tiago. This square has two major clusters (C₃ and C₄) that are the closest to the global mean of this square. C₃ collects all the high values of isovists and VGA except the ones related with clustering coefficient (perimetral and extended VGA), due to the concave shape of its urban frontage, as well as solar exposure, along with positive values of visibility to public and touristic services, distinctive traits to C₄. The latter collects all the high values of visibility to high façades, with low values of VGA integration.

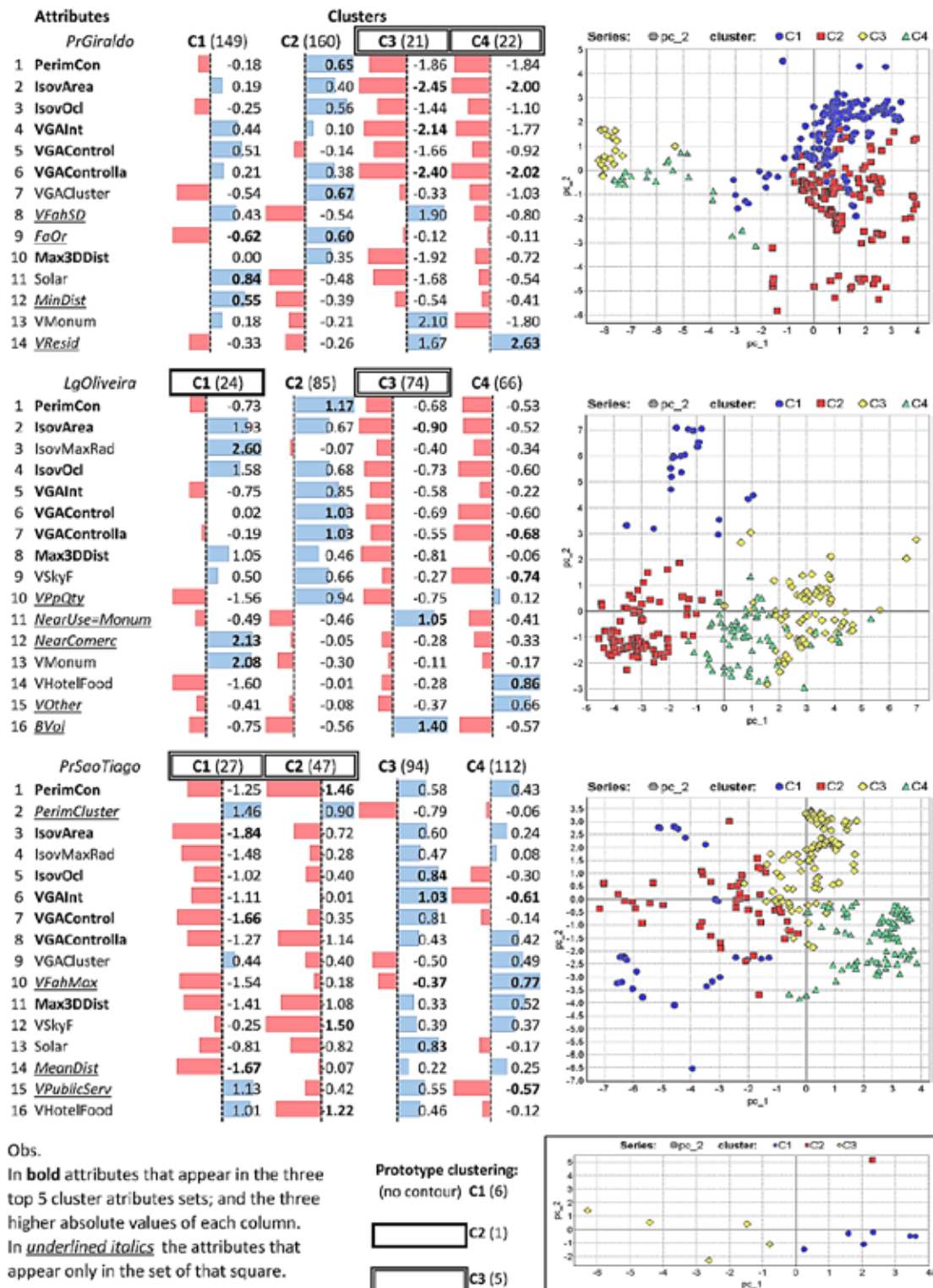


Figure 6 - Clustering on each square's Sp points and on the prototypes of all the squares (bottom right). In parentheses: the number of examples.

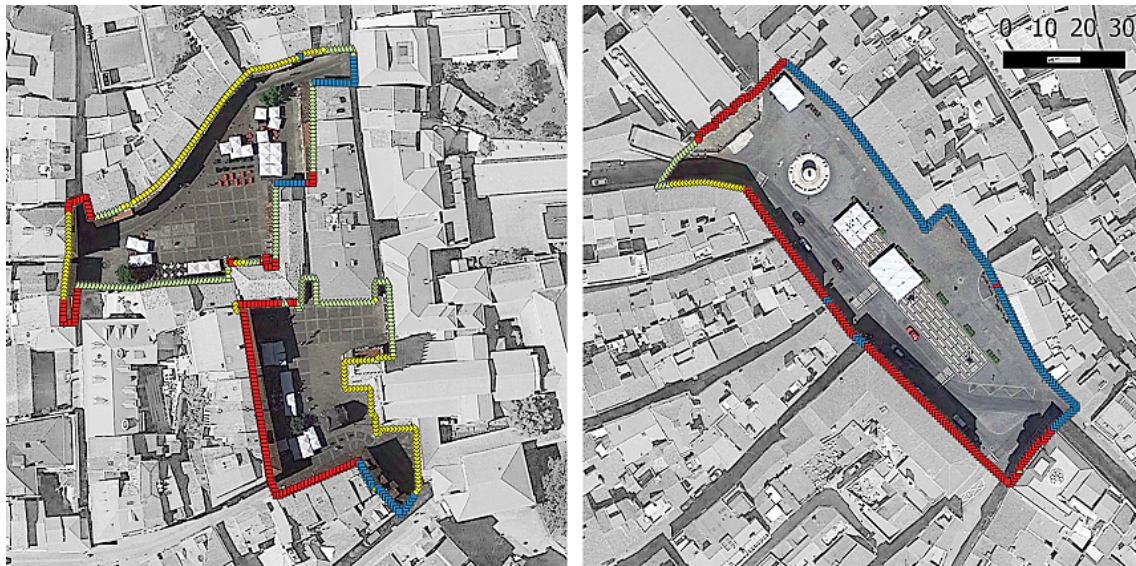


Figure 7 - Mapping of the clusters on site (representations are local to each square and follow Figure 6 legends).

In the second clustering experiment on the cluster prototypes of all the squares three clusters are highlighted (Figure 6, bottom left). One of the clusters is an outlier with only one example: cluster C1 of Largo da Oliveira. This comes from the extreme positive values of isovist attributes, exclusive proximity to commerce and visibility to the church that diverges highly from the mean values of this dataset. It's located in a vestibular space to the square and a museum entrance with low visibility to the central space but through which penetrate long axial lines from south and across to the nearby Praça de São Tiago.

The third clustering experiment on the joined datasets of Sp points of the three squares also produces four clusters, individualizing Praça do Giraldo main central space, with the highest mean of Sp's VGA integration, isovist area and visibility to commerce (Figure 8, top). The outlier cluster of Largo da Oliveira (C2 in this analysis) keeps its independence and Guimarães case studies share clusters C3 and C4 which separate the more segregated and irregular spaces from the main urban frontages. The key features that individualize those clusters are the opposite signs of VGA controllability, sky factor and the quantity of visible Pp points. It's also possible to observe in the plot of Figure 8 (top left) that Guimarães squares locate in the negative half plane of the PC1 axis, and are well separable from the main cluster of Évora.

In the exploratory experiment using Screen Predictors (SAS, 2009), we analysed the joined dataset of all the Sp points in the discovery of strong predictors on the distribution of uses. Discarding the unidentified uses (Near=Other and VOther) it is defined as our response variable the NearUse nominal attribute and all the others as explanatory ones (Figure 8, bottom). The results are not surprising for Near=Monum (mainly churches in our case studies) as the main predictor is its characteristic extraordinary built volume; public buildings and commercial use have direct configurational related predictors, perimetral connectivity and VGA integration, respectively, corroborating long established space syntax findings; touristic/restaurant activity as well residential use have visibility as their strong predictors. The former has its own visibility as its main predictor, pointing to a clustering phenomenon on its location, and the latter has visibility to commercial use.

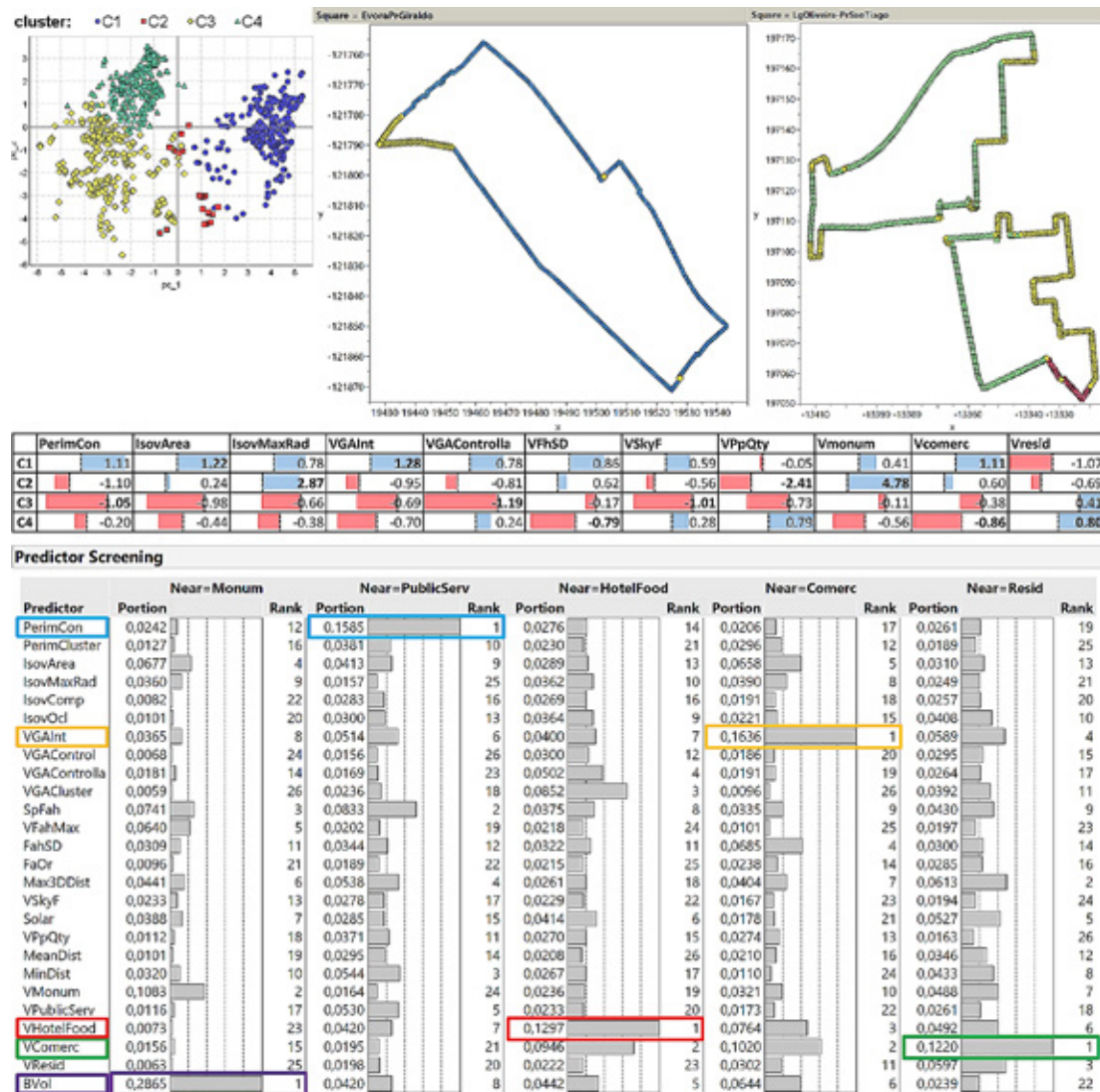


Figure 8 - Analysis on the joined datasets. Top: X-means clustering (the plotted squares are not at the same scale); bottom: predictor screening on the use location variables.

6. SUMMARY OF RESULTS AND DISCUSSION

In a critical summary of the results we can assert:

1. Main correlations exist between visibility-configurational related attributes (connectivity, visibility and VGA), and this study shows that their relation to morphologic or environmental ones (e.g. façade area or sky visibility) is highly local.
2. Given the heterogeneity of the attributes, general correlation is low and sharp clusters do not predominate, however their number (four), even if automatically determined, is constant across the analysis of the Sp point datasets (individually and jointly).
3. Cluster description by prototype attribute values is heterogeneous but cross relations are highlighted, capturing some site idiosyncrasies.
4. Key discriminating attributes for clustering (appearing in all case studies prototypes definitions) are mainly, directly or indirectly, configurational.
5. Prototype clustering shows an outlier where the functional features are fundamental for its description, and not only physical or configurational aspects.

6. The joined dataset clustering analysis, separates Évora example from Guimarães' ones, even if no spatial data is used. Furthermore, it shows some capacity for generalization as clusters appear across squares, but only extremely divergent clusters remain from the local-individual to the global-joined analysis.
7. Attributes related with physical proximity to Pp points (Use attributes) have low expression in clustering. This should be further investigated since in an experiment using a normalization by range [0,1] they gain a greater predominance.
8. Within our joined dataset universe, strong predictors analysis on use location corroborates the predominance of configurational related attributes as main predictors. An interesting finding is the visual clustering phenomena of touristic/restaurant activity location, as its strongest location predictor is its own visibility.

The empirical analysis of the relation of people's appropriation of space with the spaces defined by the clustering of Sp points seems, for instance, to corroborate the preference for highly clustered convex spaces in the location of terraces (Figure 7, e.g. Praça de São Tiago). However, these spaces are highly conditioned by the potential of transformation of built typologies, mainly their ground floor, and by peripheral car circulation which tends to segregate the more domestic sides of the squares.

7. CONCLUSIONS

Going beyond a perimetral syntactic analysis, as possible in Depthmap, or a shape analysis by perimetral connectivity, like in Psarra and Grajewski (2001), the present study illustrates a specific method for the analysis of the urban square, which tries to capture its latent complexity. Taking a stand on the micro urban morphologic analysis and the individuality and potential of the perimetral space to work both as interface between public - private and architecture - urban space, this space is synchronically and multidimensionally analysed in three Portuguese case studies resorting to data mining techniques. These seem to be most appropriate to support a data-driven approach that tries to blend standpoints, scales and themes of analysis, reasserting the centrality of the study object, the urban space/square, in the realm of the varied field of the urban studies.

Although our approach is eminently analytical, its application to design can be envisaged in a strategic level. In fact, it can help in promoting a more evidence-based design and to gap the barriers between analysis, decision and project, resorting to case-based reasoning, which is inherent to the presented methodologies.

We should stress that the analysis is highly sensitive to the *a priori* definition of the perimeter and to the data normalization methods. In future work these issues should be addressed, as well as clustering quality measures and the analysis of the frontier between clusters (both in urban and data spaces).

The basic concepts in which we delve are, beyond shape and proximity, visibility and permeability, which in the space of the square have a particular character. While in the channel space of the street the intervisibility of façade and permeability are defined mainly by bilateral frontality, in the square that affinity is augmented by the presence of third elements in lateral relation. This kind of relation mediated by a third element is exactly at the root of configurational analysis and complexity itself (Poincaré's *three-body problem*).

New tools (concrete and conceptual) impelled the space syntax project in the 1980's, namely graph and network theories, and backed up its structuralist approach (Hillier, 1996). As the theory fundamentals get stabilized, its boundaries get more convulsed, and amidst a data driven society a new set of tools seem necessary to support our cognitive limitations and open new frontiers. Data mining and machine learning are therefore proposed here as methods, not only of systematic analysis on urban data and hypothesis testing, but also hypothesis creation and speculative devices.

ACKNOWLEDGMENTS

The first author is supported by FCT grant SFRH/BD/95148/2013. The author acknowledges the support of Guimarães and Évora municipalities for the accessibility to official cartography.

REFERENCES

- Ahu, S., Cagdas, G., Sariyildiz, S. (2011), 'A Multi-Dimensional Exploration of Urban Attributes by Data Mining', *CAAD Futures 2011: Designing Together*, ULg.
- Al-Sayed, K. (2013), 'SYNTHETIC SPACE SYNTAX: A generative and supervised learning approach in urban design', *Proceedings of the Ninth International Space Syntax Symposium*, Seoul: Sejong University.
- Benedikt, M (1979), 'To take hold of space: isovists and isovist fields', *Environment and Planning B: Planning and Design*, 6(1), 47–65.
- Beirão, JN. and Koltsova, A. (2015), 'The Effects of Territorial Depth on the Liveliness of Streets', *Nexus Network Journal*, 17(1), pp.73–102.
- Campos, M. (1997), 'Strategic spaces: patterns of use in public squares of the city of London', *Space Syntax - First International Symposium Proceedings*, London, p. 26.1-26.11.
- Campos, A., Golka, T. (2005), 'Public Spaces Revised: of the relationship between Patterns of Stationary Activities and Visual fields', *Proceedings 5th International Space Syntax Symposium*, Delft.
- Chaszar, A. and Beirão, J. (2013), 'Feature Recognition and Clustering for Urban Modelling-Exploration and Analysis in GIS and CAD', *Proceedings of the 18th CAADRIA Conference*, Singapore, p. 601-610.
- Coelho, C.D. (2008), *A praça em Portugal. Inventário de espaço público*, Lisbon: DGOTDU.
- Cutini, V. (2003), 'Lines and squares: towards a configurational approach to the morphology of open spaces', *4th International Space Syntax Symposium*, London.
- Demser, U. (2006), *Data Mining of Geospatial Data: Combining Visual and Automatic Methods*, PhD Thesis.KTH.
- Gehl, J. (2011), *Life Between Buildings. Using Public Space*, London: Island Press.
- Gil J., Beirão J., Montenegro N., Duarte J. (2012), 'On the discovery of urban typologies: Data mining the many dimensions of urban form.' In *Urban Morphology*, vol. 16 (1), p.27–40.
- Habraken, N. J., Teicher, J. (1998), *The structure of the ordinary: form and control in the built environment*, Cambridge, MA: MIT Press.
- Han, J., Kamber, M. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann.
- Hanna, S. (2009), 'Spectral comparison of large urban graphs', *Proceedings of the 7th International Space Syntax Symposium*, Stockholm.
- Hillier, B. and Hanson, J. (1984), *The Social Logic of Space*, Cambridge: Cambridge University Press.
- Hillier, B. (1996), *Space is the machine: a configurational theory of architecture*, Cambridge: Cambridge University Press.
- Hillier B., Penn A., Hanson J., Grajewski T., Wu J. (1993), 'Natural Movement: or, Configuration and Attraction in Urban Pedestrian Movement', *Environment and Planning B: Planning and Design*, 20.
- Laskari A., Hanna S., Derix C. (2008), 'Urban Identity Through Quantifiable Spatial Attributes', *Design Computing and Cognition*, Vol. 08, p.615–634.
- Laskari, A. (2014), 'Multidimensional Comparative Analysis for the Classification of Residual Urban Voids', *Proceedings of the 32nd eCAADe Conference*, Newcastle upon Tyne, p. 283-292.
- Lopes, J., Paio, A., Beirão, J.N., Pinho, E.M., Nunes, L. (2015), 'Multidimensional Analysis of Public Open Spaces - Urban Morphology, Parametric Modelling and Data Mining', *Proceedings of the 33rd eCAADe Conference*, Vol 1, p. 351–360.
- Lynch, K. (1960), *The Image of the City*, Cambridge Massachussetts: MIT press.
- Miller, H.J. and Han, J. (2001), 'Geographic data mining and knowledge discovery, An overview'. In Miller, H.J. and Han, J. (eds), *Geographic data mining and knowledge discovery*, Taylor & Francis.

- Palaiologou, G., Griffiths S., Vaughan L. (2016), 'Reclaiming the virtual community for spatial cultures: Functional generality and cultural specificity at the interface of building and street', *The Journal of Space Syntax*, Vol. 7 (1), p. 25-54.
- Pelleg, D., Moore, A. (2000), 'X-means: Extending K-means with efficient estimation of the number of clusters', *Proceeding of the 17th International Conference on Machine Learning*, p.727-734.
- Psarra, S., Grajewski, T. (2001), 'Describing shape and shape complexity using local properties', *3th International Space Syntax Symposium*, Atlanta.
- Ribeiro, O. (2011 [1945]), *Portugal, o Mediterrâneo e o Atlântico*, Estudo Geográfico, Lisboa: Letra Livre.
- SAS Institute Inc. (2009), *JMP® 9 Modeling and Multivariate Methods*, Cary, NC: SAS Institute Inc.
- Serra, M., Gil, J., Pinho, P. (2013), 'Unsupervised Classification of Evolving Metropolitan Street Patterns', *Ninth International Space Syntax Symposium*, p. 045:1-046-10.
- Silva, M. (2009), *Forma e circunstância: a praça na cidade portuguesa contemporânea*, Unpublished Phd thesis. FAUP-UP, Porto.
- Stähle, A., Marcus, L., Karlström, A. (2005), 'Place Syntax - Geographic Accessibility with Axial Lines in GIS', *Proceedings 5th International Space Syntax Symposium*, Delft, pp. 131-144.
- Thiel, P. (1961), 'A sequence experience notation for architectural and urban space', *Town Planning Review*, 32, pp.33-52.
- Turner, A. (2003), 'Analysing the visual dynamics of spatial morphology', *Environment and Planning B: Planning and Design*, 30(5), pp.657-676.
- van Nes, A., López, M. J. J. (2007), 'Micro scale spatial relationships in urban studies: The relationship between private and public space and its impact on street life'. In Kubat, A. S., Ertekin, Ö., Güney, Y. I. and Eyübolou, E. (eds.), *Proceedings of the Sixth International Space Syntax Symposium*, Istanbul: I.T.U. Faculty of Architecture, p.023.1-12.
- Witten I., Frank E. (2005), *Data mining: practical machine learning tools and techniques*, Boston: Morgan Kaufman.