# 94

# MULTIDIMENSIONAL STUDY OF URBAN SQUARES THROUGH PERIMETRAL ANALYSIS:

## Three Portuguese case studies

_____

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 that embrace that latent complexity and address the multi-scale and multivariate correlations of factors that defy human cognitive capabilities. This paper will present a synchronic multidimensional analysis of three Portuguese historic squares: Praça da Oliveira, Praça de Santiago (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 designating a designed formal space, in this context we use the term in a broader sense including 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 covers toponymies like _largo_, _terreiro_ or _rossio_.

The liveliness and the success of interventions in these spaces is defined not only by its centrality but by local features like the permeability of its defining planes, on notions of comfort, security, use and appropriation affordances of a micro urban morphology

(Habraken&Teicher, 1998). To address this complexity, we present a methodology that 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 of the design of the built environment itself (Hillier&Hanson, 1984). Discretizing the boundary space of the squares we delve in its potential as interface between architecture and public open space observing as an *a priori* set of heterogeneous attributes manifest themselves along its perimeter. 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 inductive data mining methodology proposed 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 territory, entrenched between *Atlantic and Mediterranean cultures* (Ribeiro, 2011). At the present time tourism and the location of university centres are crucial factors for development of these two medium-sized cities, compelled 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, in mainly pedestrianized neighbours where car circulation is not disruptive of their tourist and civic event activity. Besides a considerable difference in area between Praça do Giraldo (Évora) and the examples from Guimarães (Largo da Oliveira and Praça de São Tiago), their historic function was distinct. The Praça do Giraldo (fig.1) is an example of an ancient *rossio* a multifunctional space contiguous to the external side of a city gate (in this case in the disappeared roman city wall), and considered a hallmark of the Portuguese medieval urbanism. These spaces were focal points of irradiation for subsequent urban expansions (Coelho, 2008).

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

The Largo da Oliveira as a more formal and symbolic character functioning as the forecourt of the church that dominates and gives its name to the place, and Praça de São Tiago is a classic medieval market square (previously named Praça do Peixe - Fishmarket Square) surrounded mainly by typical habitational houses. These form an interesting system as they are connected through the archway of the open floorplan of the elevated medieval building of the city hall. Here administrative power is literally between the religious and the economic one. In the present investigation they are studied as separated spaces, where each one is part of the context of the other (fig. 1).

## 2. OBJECTIVES

The main objectives of the present investigation are: (i) to describe and classify the squares' perimetral space through a non-supervised method that naturally groups sampling points around their perimeter taking in 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 on those spaces, from the perspective of urban morphology and human behaviour (Gehl, 2011).

## 3. BACKGROUND AND MOTIVATION

Meanwhile theory integration between space syntax and urban morphology is mainly carried out at the urban macro scale, like in 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 and hinders the development of relations with typo-morphologic and architectonic stands on the micro morphologic analysis of the city form 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*. Problematizing this idea of 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).

The syntactic exploration of the relations between street interface and liveliness has been carried out by authors like van Nees (2007) and Beirão&Koltsova (2015), driven by space syntax seminal ideas like the formalization of the interface relation between private and public realms and the morphogenetic model of the aggregation that gives rise to the *beady ring* structure, or the constitutedness of space and the interface map (Hillier&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. Totalizing generalization gives way to explanatory power on culturally ingrained practices.

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. Although perceptive and cognitive explorations my cast some light on this *threshold* problem (Turner, 2003; Thiel, 1961; Lynch, 1960). At this micro level, the type of interface between public open space and private space becomes important, culturally specific and culturally determinant of the public space use. One could say that the regularity and the character of the interface porosity culturally shape the dynamics of the public space.

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 a square.

The classic methods of urban morphologic and syntactic analysis, usually limited to the analysis of single or pairs of variables, so as to respond to human cognitive and perceptive limitations, restrict the simultaneous expression of features that give spaces their

uniqueness and characterize them as *places*. Serra, Gil and Pinho (2013) compile the shortcomings of traditional typomorphological approaches, namely their time-consuming methods restraining the amount of examples and dimensions considered, and theoretical partiality. Some recent formal approaches try to escape these limitations by relating attributes and classifying examples through clustering in multivariate tri or quadrangular graphs (Berghauser & Haupt, 2010).

The identified deficiencies can be handled by the use of 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&Frank, 2005). Data mining, also known as knowledge discovery in databases (KDD) (Han&Kamber, 2001) main objectives are knowledge extraction, prediction and hypothesis generation from data, favouring an inductive (bottom-up) approach. This approach is in opposition with confirmatory techniques that require a priori hypotheses formulation that restrict hidden information discovery (Miller&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 relatively easier to implement by computational methods and are widely applied to many fields of science, engineering and business. When spatial data his involved, like in geographic information systems and geoscience, it is designated by spatial data mining (Demsar, 2006).

Within urban morphological studies, they support 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) explores 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-formal spatial qualities, leading to a better understanding of the public open space morphology.

Although multidimensional morphologic analysis of the urban void are 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 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 previous investigation on a general methodology for the multiscale and multivariate analysis and classification of public open space (Lopes et al, 2015), the first step was to choose a subset of the proposed indicators that seemed significant in the scope of the present analysis. The attributes were divided in nine main themes: Geolocation, Connectivity, Isovists, Visual graph analysis, Façade attributes, Environmental features, Constitutedness, Use, and Urban indexes. They comprise 34 attributes representative of morphological (space shape/built form), syntactic and visibility (connectivity/VGA/isovists), environmental (topography/solar exposure) and functional (permeability/uses) traits (Tab. 1). Attributes can be global or local and the expression of a property at a sampling location or aggregations of values as presented to visibility at that point.

| # | 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 | m2 | 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 | VFaArea | m2 | 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 | m3 | Real | Urban Index | Built volume at Sp point location |

Tab. 1 : Attribute list metadata and description

Consequently besides isovists we have taken another approach on visibility (akin to the VGA method) as a discrete connectivity graph of the sampling points. Recording the data structure of that graph we can more easily query on connected sampling points' attributes and aggregate them in a meaningful way that may represent the way they present themselves to visibility. Another practical advantage is that some attributes may be extrapolated and not analytically or geometrically calculated.

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 a whole unit belonging to the square. The proposed perimeters are working perimeters and other definitions may be used, but an effort was 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 will 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 (Fig.2), and a VGA analysis at their historical centres were conducted (Fig. 3). 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 (Fig. 3).
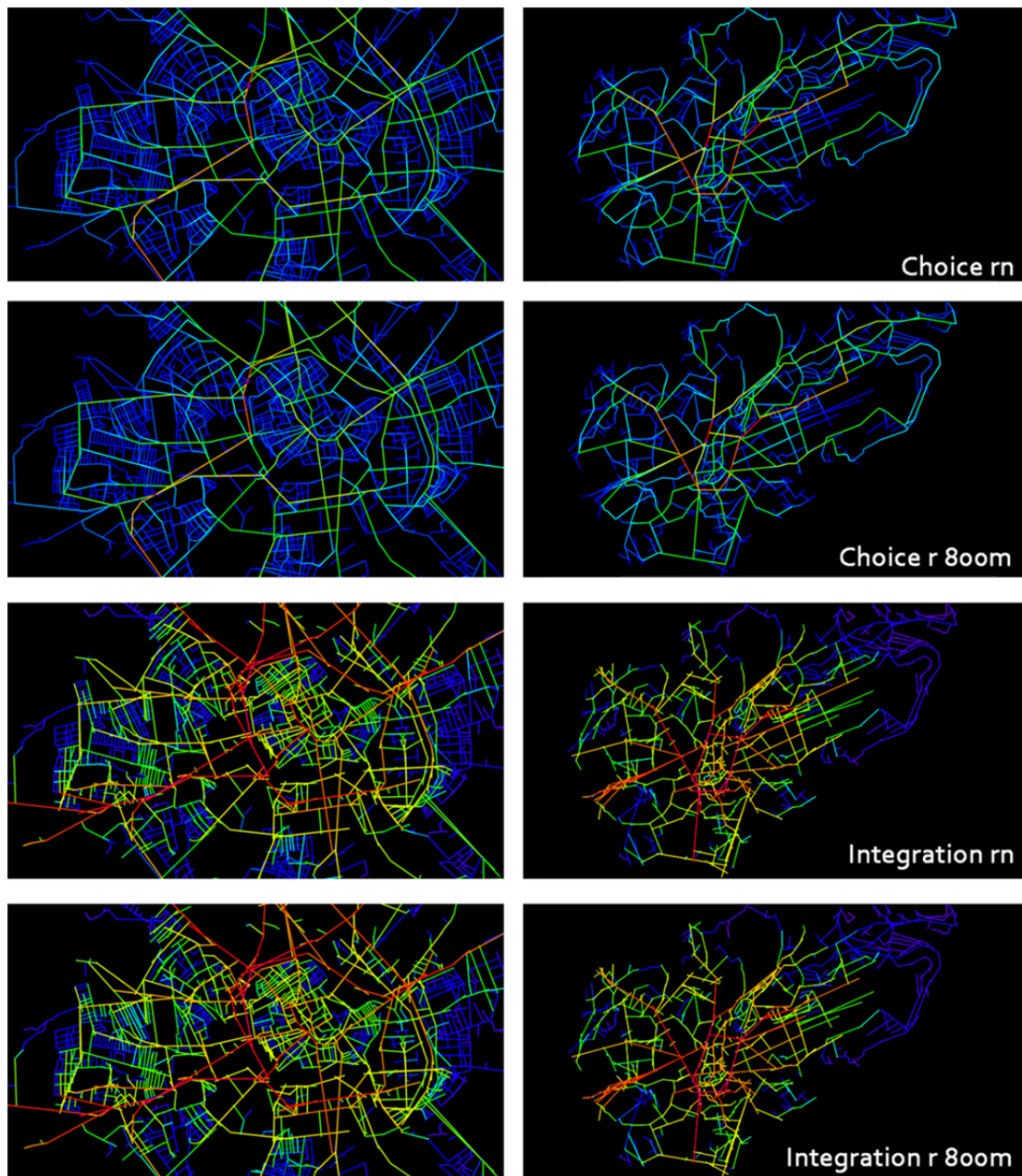
Fig. 2: Évora and Guimarães segment analysis.
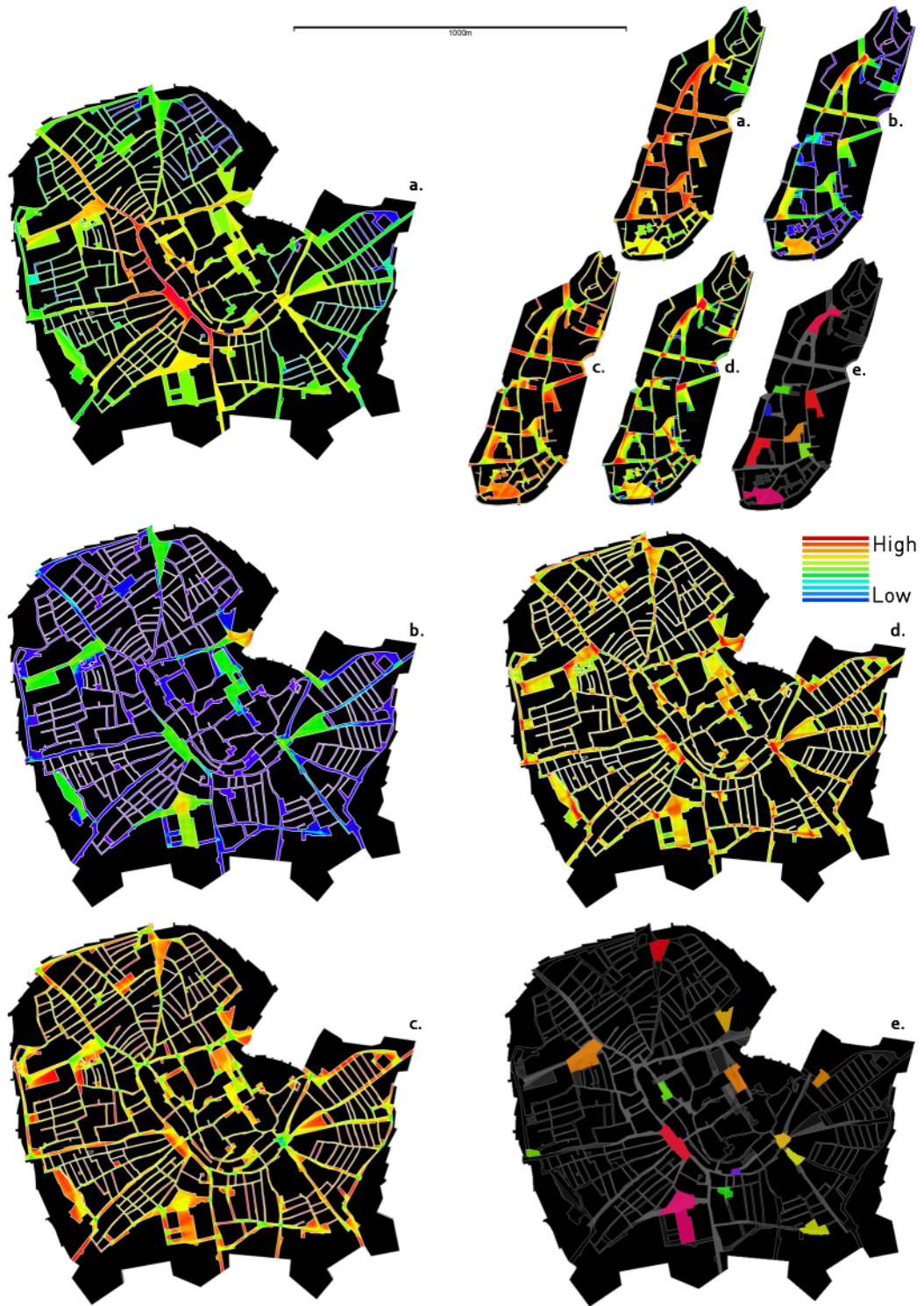
Fig. 3: Évora and Guimarães VGA analysis (1m grid resolution): a. Integration HH, b. Connectivity, c. Clustering Coefficient, d. Control, e. Interaction Index (Cutini, 2003).

Then, the focus is to put on urban micro-morphology analysis and local properties, as they are expressed and recorded along the perimeter of the square, using the methodology described above, and in the use of data mining techniques for the visualization and

exploration of the datasets so produced.
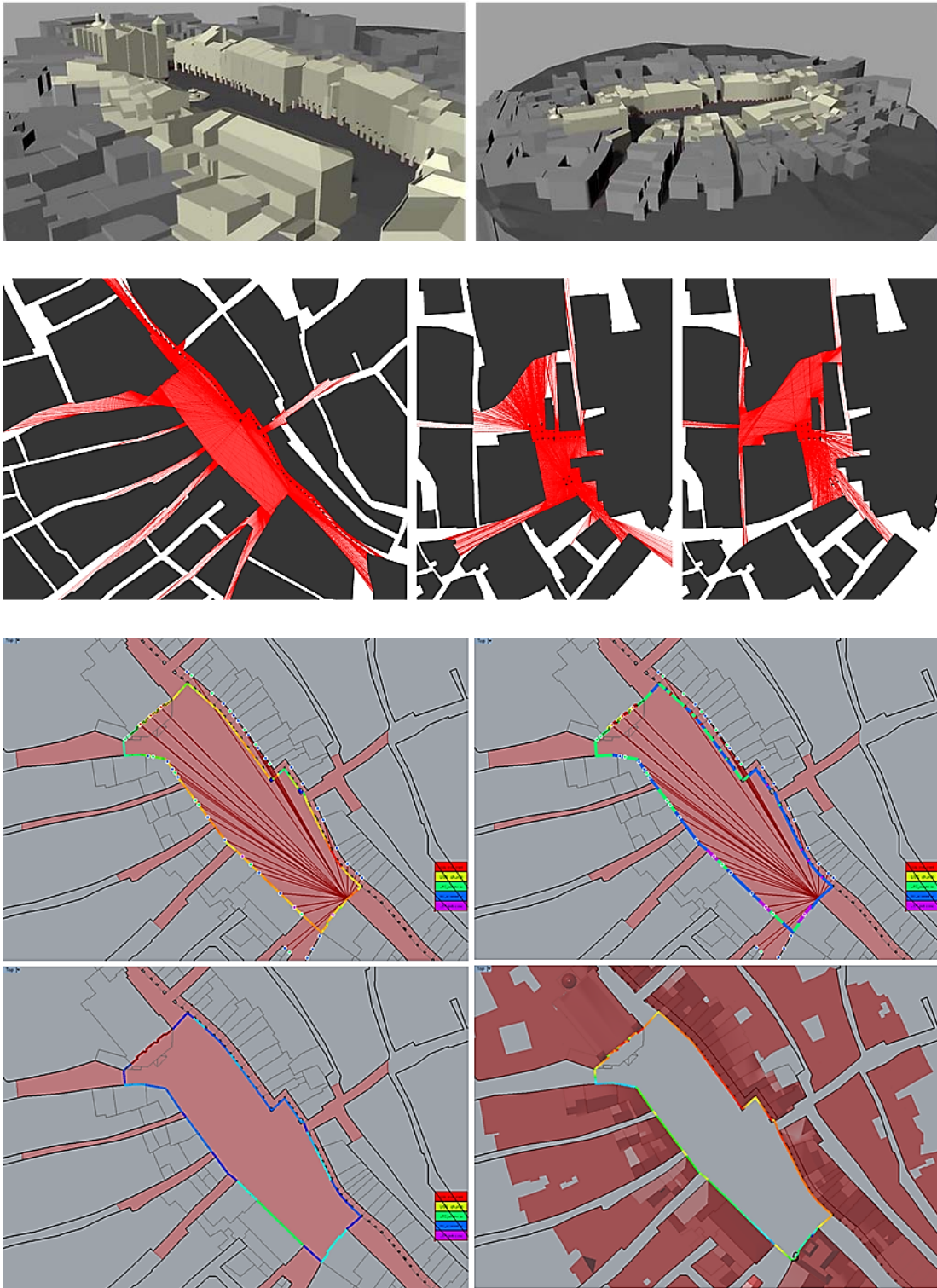


Fig.4: From the top: 3D model of Pr. Giraldo; overlapping isovists (from the left: Pr. Giraldo, L g. Oliveira; Pr. São Tiago); mapping of uses and attributes (Pr. Giraldo).

The first step was to get a sense on the data and validate its quality. Besides data exploratory techniques and descriptive statistics, the mapping of the information is

essential for checking out errors and dealing with outliers. This pre-processing phase includes the normalization of the numeric attributes as identical weight was given to all of them.

Following our objectives the data was 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 normally attend for the larger amount, allowing working in a lower dimension and clearer representations in 2D plots. Analysing the composition of those components one can filter the most important attributes for further analysis (Witten&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; we are using k-means with the Euclidean distance function (used to express dissimilarity), whose results are easier to interpret. It discovers natural groups of objects or variables, identifies extreme and archetypal examples and suggests interesting hypotheses about relationships (Witten&Frank, 2005). We use 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, we used Predictor Screening (SAS, 2009), a supervised learning method of screening many predictors for their ability to predict an outcome based on random decision trees (bootstrap forest partitioning).

The workflow (Fig. 5) implied the functional surveying and mapping of Pp points in the neighbour 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 were produced in DepthmapX (UCL and Tasos Varoudis), and the morphologic local ones using algorithm design software (Rhino/Grasshopper and its plugins Ladybug/Honeybee for

environmental analysis) as a mean of gaining control over the analytical tools (Fig. 4). Data was collected by the different scripts and converted in a multidimensional vector on a table. We used simple spreadsheets as our case studies are few, but the ideal workflow includes the data storage in a central spatial object-relational database (like PostgreSQL extended with PostGIS). For the mining and plotting of the data we resorted to student versions of Rapidminer (RapidMiner, Inc) and JMP (SAS Institute Inc.) and for visualization and mapping to QGIS and Rhino/Grasshopper. Data visualization is essential for error checking and to the critical interpretation of the results.



Fig. 5: Workflow diagram illustrating the interactive nature of the ideal process.

## 5. RESULTS

The statistical summary in Tab. 2 shows aggregated values of the raw data. Some blocks of values appear with Praça do Giraldo concentrating the generality of the highest mean values, as it is the larger space and data is unscaled, and Largo da Oliveira the lowest attribute mean values related with isovists area, connectivity and VGA, while Praça de São Tiago the ones 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 and environmental attributes 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

recorded uses: 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 with size of the squares, the detail of the defined perimeter and the type of attribute (NearUse appears as a 0 or 1 binominal attribute).

The correlation analysis (Tab. 2, left) shows two pairs of attributes with value above 0.95 and a selection was made on them (IsovPerim and VFaArea were dropped in the analysis). That table shows the pairs of attributes correlated above 0.7, and these are mainly the known cases of correlations between VGA, isovists and connectivity, and cross relation with non-syntactic attributes can be depicted. The negative correlation on the location of residential and leisure and public services uses is also highlighted. The data presents some outliers but as they don't come from measure errors but from idiosyncratic aspect of the sites, the detail of analysis and the selected perimeter they were not discarded.

In parallel with clustering a PCA analysis was 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.

| Attribute | Mean | | | Std Dev | | | Minimum | | | Maximum | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PrGiraldo | LgOliveira | PrSaoTiago | PrGiraldo | LgOliveira | PrSaoTiago | PrGiraldo | LgOliveira | PrSaoTiago | PrGiraldo | LgOliveira | PrSaoTiago |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 15 VFaArea | 2585.8 | 929.8 | 908.4 | 752.4 | 376.1 | 339.8 | 506.0 | 157.9 | 233.4 | 3492.6 | 1610.1 | 1553.8 |
| 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 |
| 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 |
| 18 VSkyF | 33.1 | 24.8 | 30.7 | 5.8 | 6.6 | 7.0 | 15.3 | 3.6 | 0.0 | 48.8 | 34.7 | 42.1 |
| 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 |
| 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 |
| 21 MeanDist | 54.4 | 34.6 | 36.3 | 11.0 | 5.8 | 9.2 | 20.2 | 16.1 | 12.3 | 84.4 | 55.5 | 60.5 |
| 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 |
| 23 NearUse=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 |
| 24 NearUse=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 |
| 25 NearUse=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 |
| 26 NearUse=Comerc | 0.4 | 0.1 | 0.1 | 0.5 | 0.3 | 0.3 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |
| 27 NearUse=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 |
| 28 NearUse=Other | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 1.0 |
| 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 |
| 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 |
| 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 |
| 32 VComerc | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.2 | 0.0 | 0.0 | 0.6 | 0.6 | 0.2 |
| 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 |
| 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 |
| 35 BVol | 3465.0 | 3962.2 | 710.5 | 4666.0 | 4636.5 | 672.9 | 0.0 | 0.0 | 0.0 | 17726.1 | 11749.5 | 4622.3 |

| PrGiraldo | | r |
|---|---|---|
| PerimCon | VfaArea | 0.99 |
| IsovPerim | IsovOcl | 0.95 |
| IsovArea | VGAControlla | 0.82 |
| IsovArea | VPpQty | 0.81 |
| IsovMaxRad | Max3DDist | 0.78 |
| PerimCon | IsovArea | 0.78 |
| IsovArea | VGAInt | 0.78 |
| VGAInt | VGAControlla | 0.78 |
| PerimCon | VGAControlla | 0.75 |
| VGAInt | VGAControl | 0.75 |
| IsovArea | VfaArea | 0.74 |
| IsovArea | IsovPerim | 0.73 |
| VGAInt | VPpQty | 0.73 |
| VSkyF | VPpQty | 0.71 |
| Max3DDist | MeanDist | 0.71 |
| VGAControlla | VfaArea | 0.71 |
| VGAControlla | VPpQty | 0.71 |
| VHotelFood | VResid | -0.71 |
| IsovPerim | IsovComp | -0.88 |
| IsovComp | IsovOcl | -0.89 |
| **LgOliveira** | | |
| IsovPerim | IsovOcl | 0.99 |
| PerimCon | VfaArea | 0.98 |
| IsovArea | IsovPerim | 0.80 |
| PerimCon | VGAControlla | 0.77 |
| IsovArea | IsovOcl | 0.77 |
| VGAControlla | VfaArea | 0.76 |
| PerimCon | VGAControl | 0.76 |
| VGAControl | VGAControlla | 0.75 |
| VGAInt | VfaArea | 0.73 |
| VfaArea | VPpQty | 0.73 |
| PerimCon | VPpQty | 0.73 |
| VGAControl | VfaArea | 0.72 |
| PerimCon | VGAInt | 0.72 |
| PerimCon | VSkyF | 0.70 |
| VGAControl | VGACluster | -0.77 |
| **PrSaoTiago** | | |
| PerimCon | VfaArea | 0.99 |
| IsovPerim | IsovOcl | 0.96 |
| VGAControlla | VfaArea | 0.87 |
| PerimCon | VGAControlla | 0.86 |
| IsovArea | MeanDist | 0.76 |
| PerimCon | Max3DDist | 0.72 |
| VfaArea | VSkyF | 0.71 |
| PerimCon | VSkyF | 0.71 |
| PerimCon | IsovArea | 0.70 |
| VfaArea | Max3DDist | 0.70 |
| IsovPerim | IsovComp | -0.71 |
| VPublicServ | VResid | -0.80 |

Tab. 2 : Statistical summary (Pr.Giraldo: N= 352; Lg.Oliveira: N= 249; Pr.São Tiago: N= 280). Left absolute values of correlation (r) over 0,7. In **bold**: discarded attributes; _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 was 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.

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

**Praça do Giraldo.** In this square two clusters (C1 and C2) domain the results and their values concentrate around the global mean. 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 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 from 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 is dominated by two 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.

**Attributes** — **Clusters**

### PrGiraldo

| | Attribute | C1 (149) | C2 (160) | C3 (21) | C4 (22) |
|---|---|---|---|---|---|
| 1 | PerimCon | -0.18 | 0.65 | -1.86 | -1.84 |
| 2 | IsovArea | 0.19 | 0.40 | -2.45 | -2.00 |
| 3 | IsovOcl | -0.25 | 0.56 | -1.44 | -1.10 |
| 4 | VGAInt | 0.44 | 0.10 | -2.14 | -1.77 |
| 5 | VGAControl | 0.51 | -0.14 | -1.66 | -0.92 |
| 6 | VGAControlIa | 0.21 | 0.38 | -2.40 | -2.02 |
| 7 | VGACluster | -0.54 | 0.67 | -0.33 | -1.03 |
| 8 | VFahSD | 0.43 | -0.54 | 1.90 | -0.80 |
| 9 | FaOr | -0.62 | 0.60 | -0.12 | -0.11 |
| 10 | Max3DDist | 0.00 | 0.35 | -1.92 | -0.72 |
| 11 | Solar | 0.84 | -0.48 | -1.68 | -0.54 |
| 12 | MinDist | 0.55 | -0.39 | -0.54 | -0.41 |
| 13 | VMonum | 0.18 | -0.21 | 2.10 | -1.80 |
| 14 | VResid | -0.33 | -0.26 | 1.67 | 2.63 |

### LgOliveira

| | Attribute | C1 (24) | C2 (87) | C3 (53) | C4 (85) |
|---|---|---|---|---|---|
| 1 | PerimCon | -0.73 | 1.17 | -0.68 | -0.53 |
| 2 | IsovArea | 1.93 | 0.67 | -0.90 | -0.52 |
| 3 | IsovMaxRad | 2.60 | -0.07 | -0.40 | -0.34 |
| 4 | IsovOcl | 1.58 | 0.68 | -0.73 | -0.60 |
| 5 | VGAInt | -0.75 | 0.85 | -0.58 | -0.22 |
| 6 | VGAControl | 0.02 | 1.03 | -0.69 | -0.60 |
| 7 | VGAControlIa | -0.19 | 1.03 | -0.55 | -0.68 |
| 8 | Max3DDist | 1.05 | 0.46 | -0.81 | -0.06 |
| 9 | VSkyF | 0.50 | 0.66 | -0.27 | -0.74 |
| 10 | VPpQty | -1.56 | 0.94 | -0.75 | 0.12 |
| 11 | NearUse=Monum | -0.49 | -0.46 | 1.05 | -0.41 |
| 12 | NearComerc | 2.13 | -0.05 | -0.28 | -0.33 |
| 13 | VMonum | 2.08 | -0.30 | -0.11 | -0.17 |
| 14 | VHotelFood | -1.60 | -0.01 | -0.28 | 0.86 |
| 15 | VOther | -0.41 | -0.08 | -0.37 | 0.66 |
| 16 | BVol | -0.75 | -0.56 | 1.40 | -0.57 |

### PrSaoTiago

| | Attribute | C1 (41) | C2 (36) | C3 (109) | C4 (94) |
|---|---|---|---|---|---|
| 1 | PerimCon | -1.25 | 1.46 | 0.58 | 0.43 |
| 2 | PerimCluster | 1.46 | 0.90 | -0.79 | -0.06 |
| 3 | IsovArea | -1.84 | 0.72 | 0.60 | 0.24 |
| 4 | IsovMaxRad | -1.48 | 0.28 | 0.47 | 0.08 |
| 5 | IsovOcl | -1.02 | 0.40 | 0.84 | -0.30 |
| 6 | VGAInt | -1.11 | 0.01 | 1.03 | -0.61 |
| 7 | VGAControl | -1.66 | 0.35 | 0.81 | -0.14 |
| 8 | VGAControlIa | -1.27 | 1.14 | 0.43 | 0.42 |
| 9 | VGACluster | 0.44 | 0.40 | -0.50 | 0.49 |
| 10 | VFahMax | -1.54 | 0.18 | -0.37 | 0.77 |
| 11 | Max3DDist | -1.41 | 1.08 | 0.33 | 0.52 |
| 12 | VSkyF | -0.25 | 1.50 | 0.39 | 0.37 |
| 13 | Solar | -0.81 | 0.82 | 0.83 | -0.17 |
| 14 | MeanDist | -1.67 | 0.07 | 0.22 | 0.25 |
| 15 | VPublicServ | 1.13 | 0.42 | 0.55 | -0.57 |
| 16 | VHotelFood | 1.01 | 1.22 | 0.46 | -0.12 |

Obs.

In **bold** attributes that appear in the three top 5 cluster atributes sets; and the three higher absolute values of each column. In *underlined italics* the attributes that appear only in the set of that square.

Prototype clustering:
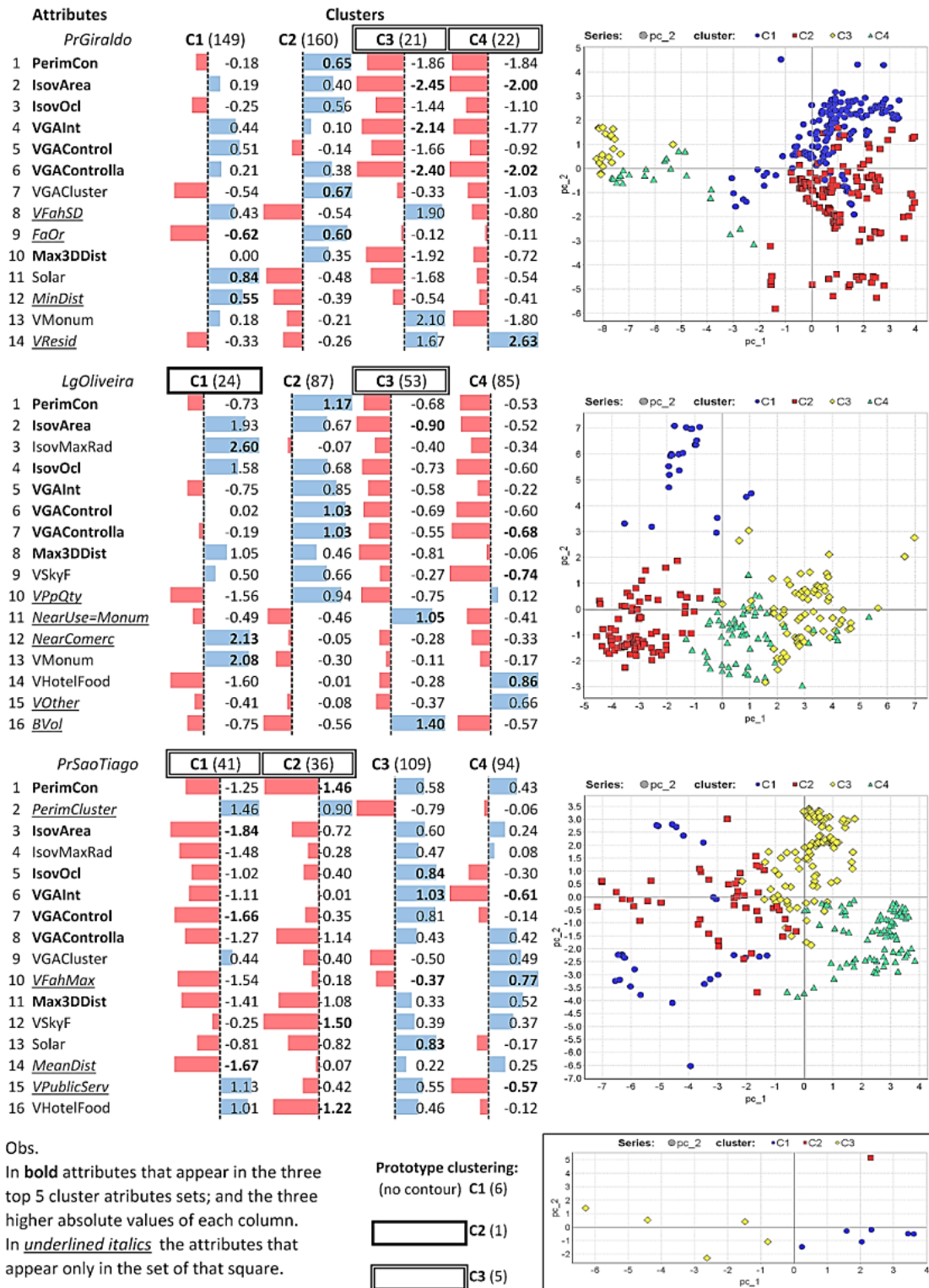(no contour) C1 (6)

C2 (1)

C3 (5)

Fig.6: Clustering on each square's Sp points and on the prototypes of all the squares (bottom right). In parentheses the number of examples.
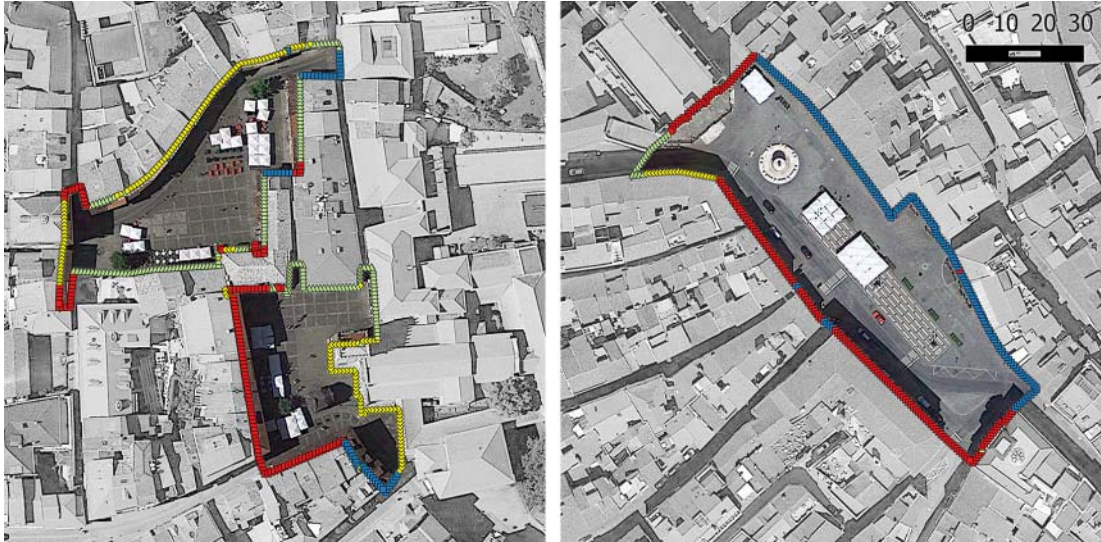
Fig. 7: Mapping of the clusters on site (representations are local to each square and fallow Fig. 6 legends).

Performing another X-means clustering on the cluster prototypes of all the squares three clusters are highlighted (Fig. 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 of the all 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 clustering in the total set of the 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 (Fig. 8, above). 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 Fig. 8 that Guimarães squares locate in the negative half plane of the data PC1, and are well separable from the main cluster of Évora.

|  | PerimCon | IsovArea | IsovMaxRad | VGAInt | VGAControlla | VFhSD | VSkyF | VPpQty | Vmonum | Vcomerc | Vresid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1.11 | 1.22 | 0.78 | 1.28 | 0.78 | 0.86 | 0.59 | -0.05 | 0.41 | 1.11 | -1.07 |
| C2 | -1.10 | 0.24 | 2.87 | -0.95 | -0.81 | 0.62 | -0.56 | -2.41 | 4.78 | 0.60 | -0.69 |
| C3 | -1.05 | -0.98 | -0.66 | -0.69 | -1.19 | -0.17 | -1.01 | -0.73 | 0.11 | -0.38 | 0.41 |
| C4 | -0.20 | -0.44 | -0.38 | -0.70 | 0.24 | -0.79 | 0.28 | 0.79 | -0.56 | -0.86 | 0.80 |

### Predictor Screening

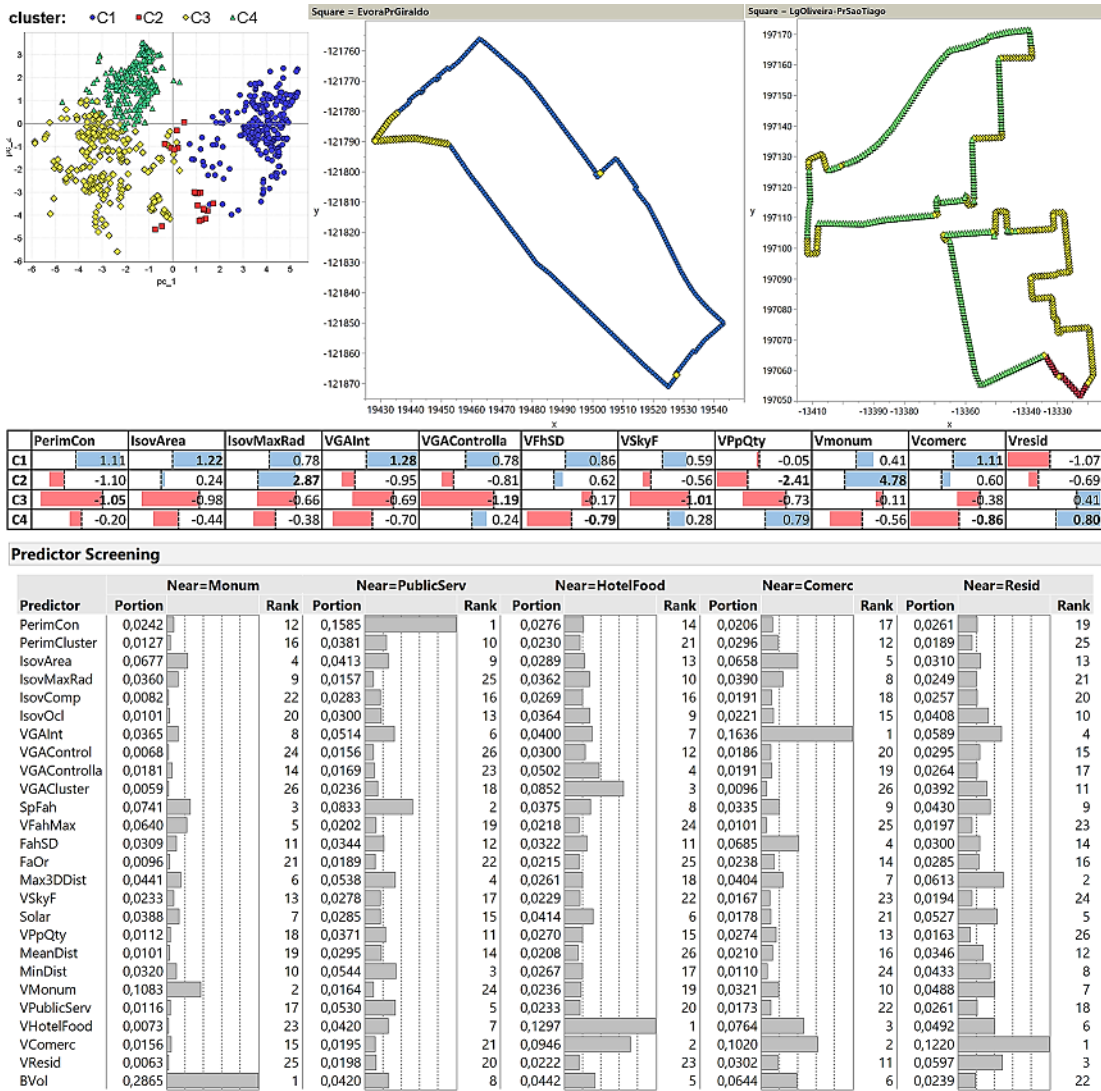| Predictor | Near=Monum Portion | Rank | Near=PublicServ Portion | Rank | Near=HotelFood Portion | Rank | Near=Comerc Portion | Rank | Near=Resid Portion | Rank |
|---|---|---|---|---|---|---|---|---|---|---|
| PerimCon | 0,0242 | 12 | 0,1585 | 1 | 0,0276 | 14 | 0,0206 | 17 | 0,0261 | 19 |
| PerimCluster | 0,0127 | 16 | 0,0381 | 10 | 0,0230 | 21 | 0,0296 | 12 | 0,0189 | 25 |
| IsovArea | 0,0677 | 4 | 0,0413 | 9 | 0,0289 | 13 | 0,0658 | 5 | 0,0310 | 13 |
| IsovMaxRad | 0,0360 | 9 | 0,0157 | 25 | 0,0362 | 10 | 0,0390 | 8 | 0,0249 | 21 |
| IsovComp | 0,0082 | 22 | 0,0283 | 16 | 0,0269 | 16 | 0,0191 | 18 | 0,0257 | 20 |
| IsovOcl | 0,0101 | 20 | 0,0300 | 13 | 0,0364 | 9 | 0,0221 | 15 | 0,0408 | 10 |
| VGAInt | 0,0365 | 8 | 0,0514 | 6 | 0,0400 | 7 | 0,1636 | 1 | 0,0589 | 4 |
| VGAControl | 0,0068 | 24 | 0,0156 | 26 | 0,0300 | 12 | 0,0186 | 20 | 0,0295 | 15 |
| VGAControlla | 0,0181 | 14 | 0,0169 | 23 | 0,0502 | 4 | 0,0191 | 19 | 0,0264 | 17 |
| VGACluster | 0,0059 | 26 | 0,0236 | 18 | 0,0852 | 3 | 0,0096 | 26 | 0,0392 | 11 |
| SpFah | 0,0741 | 3 | 0,0833 | 2 | 0,0375 | 8 | 0,0335 | 9 | 0,0430 | 9 |
| VFahMax | 0,0640 | 5 | 0,0202 | 19 | 0,0218 | 24 | 0,0101 | 25 | 0,0197 | 23 |
| FahSD | 0,0309 | 11 | 0,0344 | 12 | 0,0322 | 11 | 0,0685 | 4 | 0,0300 | 14 |
| FaOr | 0,0096 | 21 | 0,0189 | 22 | 0,0215 | 25 | 0,0238 | 14 | 0,0285 | 16 |
| Max3DDist | 0,0441 | 6 | 0,0538 | 4 | 0,0261 | 18 | 0,0404 | 7 | 0,0613 | 2 |
| VSkyF | 0,0233 | 13 | 0,0278 | 17 | 0,0229 | 22 | 0,0167 | 23 | 0,0194 | 24 |
| Solar | 0,0388 | 7 | 0,0285 | 15 | 0,0414 | 6 | 0,0178 | 21 | 0,0527 | 5 |
| VPpQty | 0,0112 | 18 | 0,0371 | 11 | 0,0270 | 15 | 0,0274 | 13 | 0,0163 | 26 |
| MeanDist | 0,0101 | 19 | 0,0295 | 14 | 0,0208 | 26 | 0,0210 | 16 | 0,0346 | 12 |
| MinDist | 0,0320 | 10 | 0,0544 | 3 | 0,0267 | 17 | 0,0110 | 24 | 0,0433 | 8 |
| VMonum | 0,1083 | 2 | 0,0164 | 24 | 0,0236 | 19 | 0,0321 | 10 | 0,0488 | 7 |
| VPublicServ | 0,0116 | 17 | 0,0530 | 5 | 0,0233 | 20 | 0,0173 | 22 | 0,0261 | 18 |
| VHotelFood | 0,0073 | 23 | 0,0420 | 7 | 0,1297 | 1 | 0,0764 | 3 | 0,0492 | 6 |
| VComerc | 0,0156 | 15 | 0,0195 | 21 | 0,0946 | 2 | 0,1020 | 2 | 0,1220 | 1 |
| VResid | 0,0063 | 25 | 0,0198 | 20 | 0,0222 | 23 | 0,0302 | 11 | 0,0597 | 3 |
| BVol | 0,2865 | 1 | 0,0420 | 8 | 0,0442 | 5 | 0,0644 | 6 | 0,0239 | 22 |

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

In another exploratory experiment using Screen Predictors, we analysed the normalized dataset of all the points in the discovery of strong predictors on the distribution of uses. Discarding the unidentified uses (Near=Other and VOther) we defined as our response variable the NearUse and all the others as explanatory ones (Fig. 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, corroboration 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.

## 6. RESUME OF RESULTS AND DISCUSSION

In a critical resume of the results we can assert:

1. Main correlations exist between visibility configurational related attributes (connectivity, visibility and VGA), and this studies shows that its 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 cluster are not predominate, however their number (four), if even automatically determined, is constant across the analysis of Sp point datasets.

3. Cluster description by prototype attribute values is heterogeneous but cross relation are highlighted, capturing some of the site idiosyncrasies.

4. Main discriminating attributes for clustering (appearing in all case studies prototypes definitions) are mainly direct 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. General analysis also produced four clusters, separating Évora example from Guimarães ones, even if no spatial data was used. It shows some capacity for generalization as clusters appear across squares, but only extremely divergent cluster remain from the local to the global analysis.

7. Attributes related with physical proximity to Pp points (Use attributes) have low expression in clustering. This has to be investigated deeper as in an experiment using a normalization by range [0,1] they gain a greater predominance.

8. Within our dataset universe, strong predictors analysis on use location corroborates the predominance of configurational related attributes as main predictors. An interesting finding is the clustering phenomena of touristic and leisure activity that tends to cluster as its strong predictor is its on visibility.

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

## 7. CONCLUSIONS

Going beyond a perimetral syntactic analysis, as it can be made 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 sincroniquelly and multidimensionally analysed in three Portuguese case studies resorting to data mining techniques. These seem 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.

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*).

We should stress that the analysis is highly sensitive to the *a priori* definition of the perimeter and to the method of data normalization. In future work these should be addressed as well clustering quality measures and the analysis of the frontier between clusters.

Although our approach is eminently analytical its application to design  can be envisaged in a strategic level. 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 that is inherent to the presented methodologies.

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 in the midst of 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 here proposed as methods not only of systematic analysis on urban data and hypothesis testing but also hypothesis creation and speculative devices.

## ACKNOWLEDGMENTS

## REFERENCES

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.

Campos, A. (1997), 'Strategic Spaces: Patterns of Use in Public Squares in the City of London', *Proceedings 1st International Space Syntax Symposium*, London, 26.1-26.11

Campos, A. Golka, T. (2005), 'Public Spaces Revised: A Study of the relationship between Patterns of Stationary Activities and Visual fields', *Proceedings 5th International Space Syntax Symposium*, Delft

Ahu, S., Cagdas, G., Sariyldiz, S. (2011), 'A Multi-Dimensional Exploration of Urban Attributes by Data Mining', *CAAD Futures 2011 : Designing Together*, ULg.

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. & Koltsova, A. (2015), 'The Effects of Territorial Depth on the Liveliness of Streets', *Nexus Network Journal*, 17(1), pp.73–102.

Berghauser Pont, M. & Haupt, P. (2010), *Spacematrix: Space, Density and Urban Form*, Rotterdam: NAi Pub.

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. and 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.

Demsar, 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.

Hanna, S. (2009), 'Spectral comparison of large urban graphs', *Proceedings of the 7th International Space Syntax Symposium*, Stockholm.

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

Hillier, B. & Hanson, J. (1984), *The Social Logic of Space*, 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. & Han, J. (2001), 'Geographic data mining and knowledge discovery, An overview'. In Miller, H.J. & 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. and 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 and 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'. In *Ninth International Space Syntax Symposium*p, p. 045:1-046-10.

Silva, M. (2009), *Forma e circunstância: a praça na cidade portuguesa contemporânea*, Unpublished Phd thesis. FAUP-UP, Porto.

Ståhle, 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. and 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.