Article

Classification of Commercial Structures: A Cross Sectional Analysis in Lisbon, 1995-2010

Rui Colaço[1,*] and João de Abreu e Silva[1]

[1] CERIS – Civil Engineering Research and Innovation for Sustainability. Departamento de Engenharia Civil, Arquitectura e Georrecursos (DECivil), Instituto Superior Técnico, Universidade de Lisboa. Av. Rovisco Pais, 1049-001, Lisboa, Portugal. ruimccolaco@tecnico.ulisboa.pt (R.C.), jabreu@tecnico.ulisboa.pt (J.A.S.).

* Correspondence: ruimccolaco@tecnico.ulisboa.pt

Abstract: Commercial classification is essential to spatially describe commerce. Although most of the proposed classifications have been classic functional analyses, rooted on qualitative studies, some recent classifications efforts are found to rely heavily on more quantitative methods, due eventually to increasing data availability and technological advancements. In this paper, a classification is proposed, using k-means clustering and a minimal set of variables (density, diversity and clustering) to derive the commercial structure of Lisbon. The classification is implemented for 1995, 2002 and 2010, using a 150m-sided square grid. The cross-sectional analysis of the results shows the rise of shopping-malls against city centre decline and gentrification, along with changes in cluster composition considering 9 different commercial categories (6 categories of retail, and restaurants, cafes and bars). These findings are in line with literature, thus supporting the obtained classification. Since the classification can be used to accurately describe the commercial structure of the city in different time periods, it is implied that it may also be generalized to different cities. Furthermore, the potential use of cluster membership in retail location models, which is an advantage of the proposed classification, could help strengthen the relationship between location modelling and commercial classification, thus reinforcing the role of commercial studies in urban planning and policymaking.

Keywords: commercial studies; commercial classification; retail location model; cluster analysis.

1. Introduction

Ever since the 1960's, retail has been considered a driver of sustainable communities. "Not only do public characters spread the news and learn the news at retail, so to speak. They connect with each other and thus spread word wholesale, in effect" [1] (p. 70). The communal high-street where "neighbours bump into each other" [2] (p. 1) promotes social interaction and adds to the centres multi-functionality. This generates urbanization economies from which innovation stems [3], stimulates local employment, encourages walking instead of driving [4], and increases safety and the feeling of belonging to a place, since at any given moment of the day, one can find people in the street [5].

This image of the communal high-street, though, has become increasingly disconnected from reality. Griffiths *et al.* [2] suggest this has been especially noticeable in the last three decades, but some authors propose that the traditional model of retail location, concentrated on city centres, has actually been threatened since the 1960's. Jacobs' [1] work fell into the public eye, after all, because it defended that maintaining the richness of the urban system, which included retail, was more interesting than implementing Robert Moses' "slum clearance" projects. And while urban sprawl became the first and main threat to the high-street, other factors, associated with social and demographic changes, such as an increasing purchase power and women entering the work force, also contributed to that phenomenon [6,7]. New ways of doing business (shopping malls, big box retail) would also add to the decline of the high-street during the 1960's and 1970's; and this situation, reinforced by the spread of neoliberal principles in the 1980's, would lead to the decline of city centres in general, especially in Europe [8,9]. And although extensive funding and different management models have been implemented since to promote town centre revival [10–12], the high-street continues to suffer from intense competition from other business models. In recent times, one of the most disruptive forces against the high-street is online shopping. In 2009, it was already considered to have the potential to replace in-store shopping [13]; in 2019, online purchases represented 14.1% of the total retail worldwide sales [14].

Considering that in Europe 74.5% of the population already lives in urban settlements [15], evaluating the past and current location patterns of commercial activity is essential in understanding the impact of future developments on city centre dynamics. This evaluation has conventionally been done in two ways: one from a more qualitative perspective, based on systematic observation of local case studies, and another from a more quantitative perspective, based on location modelling; or as some authors have mentioned, from the perspective of humanities and social philosophy and from the perspective of sciences [16] – and seldom communicating with each other.

Hence, retail classification studies, beginning with Proudfoot [17], propose that location phenomena is to be interpreted and presented based on data and observation, without relying on mathematical abstractions. This line of research has evolved so that retail structures can be described with great detail [18–21], which is essential for transmitting results. The drawback is that generalization becomes extremely hard. For example, Guy [21] considers that at least 10 dimensions can be used in retail classification (e.g. size of store, development history, development type), each with several possible categories. In a classification system as detailed, no two cities will be the same. Although that is, of course, true, if a researcher intends to find regularities to explain store location, it will be difficult to collect sufficient empirical data, because each case will be a specific case.

From the more quantitative perspective comes a somehow different way to analyse the processes leading to agglomeration, from which different commercial structures emerge. Hence, competitive clustering accounts for the clustering of activities of the same type [22–24]; urbanization economies and multipurpose shopping explain the clustering of activities of any type [3,25]; and increasing returns to scale [26,27] support the (ever increasing) density of centres. Density, diversity and clustering can be translated into mathematical expressions that can be generalized to different centres. And if this perspective is considered "somehow different" from the more qualitative one – but not entirely different – is because both perspectives are ultimately different ways to tackle the same questions. Acknowledging that the Central Business District (CBD) has "a marked concentration of shopping goods stores (….) which are located within that focal area of intra-city transportation collectively most accessible to the entire city population" as did Proudfoot [17] (p. 428) is not different from addressing the advantages in agglomeration economies from a strict economic perspective [26,28].

What is found is that this lack of communication between different perspectives leads to a gap in retail studies. There is no classification that does not rely specifically on form or function of local case studies, and at the same time, is sufficient to describe a commercial structure, in a manner that allows for its subsequent use in location models. This usage precludes the inclusion in the classification of market area characteristics [29–31], accessibility [32,33] and the presence of amenities [34], since these are all common explanatory variables used on location models.

Therefore, the purpose of this paper is to devise a self-contained quantitative classification of commercial structures, which can also be interpreted along the lines of more qualitative studies. By relating different analytical perspectives (social and economic, qualitative and quantitative), this classification will also help strengthen the relationship between location modelling and commercial classification, thus reinforcing the role of commercial studies in urban planning and policymaking. The classification is obtained in a bottom-up perspective [35,36] using k-means cluster analysis and a minimal set of variables: density, diversity and clustering. The use of this set of variables allows for cluster membership to be used in location models, as it is independent of common explanatory variables used in these models.

The classification is implemented for three different periods, corresponding to the years of 1995, 2002 and 2010. These three years were chosen because they correspond to retail censuses collecting business establishments disaggregated at point level [37] that are also close to the national housing and population censuses (1991, 2001 and 2011), since census data may latter be used to interpret changes in commercial clusters. Comparing the structures in a cross-sectional manner will help validating the classification, since the results are expected to be validated by literature and anecdotal evidence on commercial changes occurring between 1995 and 2010.

The paper is structured as follows: Section 2 presents a literature review, from classic functional analysis to classifications based on state-of-the-art methods, including references to novel approaches to retail location that have allowed for the latter to be devised. Section 3 presents the data and methodology, along with the case study, Lisbon. The results are presented and discussed on Section 4. The conclusions are presented on Section 5, along with possible lines of research that may be pursued in further developments of the present analysis.

2. Literature Review

The first systematic description of the retail structure of a city can eventually be attributed to Proudfoot [17]. Based on the analysis of 9 cities in the USA, and considering "class of commodities sold; special concentration or dispersion of outlets; and character of customer tributary areas" [17] (p. 425), the author found that most cities presented a similar structure, consisting of 1) a Central Business District (CBD), 2) an outlying business centre (or centres), 3) a principal business thoroughfare, 4) neighbourhood business streets and 5) isolated store clusters (isolated stores are also mentioned). The CBD concentrated most of the retail, both in quantity and diversity, with the outlying business districts being smaller versions ("miniatures") of it. The principal business thoroughfare was the main artery leading to the CBD, relying on through traffic, and the visibility that came with it. Neighbourhood business streets and isolated store clusters catered for local customers, and were located at walking distance in residential areas, and so were isolated stores. This classification, though rooted in urban social studies, holds against economic theory: acknowledging the advantages in the concentration of activities is, ultimately the same as addressing the advantages in agglomeration economies.

The classification methodology of Berry [18,38] proposes that, in any city, one can find a set of nested nucleations, down from the metropolitan CBD to small convenience centres, eventually connected by ribbon structures (developed around streets, or highway oriented). These ribbons draw patronage from through traffic. Specialized areas (like entertainment districts) exist where the type of business demands for a specific location that doesn't fit the proposed hierarchical structure The influence of CPT is evident on the classification system, which is eventually explained by Berry's previous work [39,40]. The possibility of mapping the results creates a direct relationship with urban planning, and retail classification has been used for that purpose since at least Berry [18] for Chicago's community renewal plan.

Davies and Bennison [19] suggest that there are essentially three conformations of retail structures: nucleated facilities, ribbon developments and specialised functional areas. They need not to be "mutually exclusive or spatially discrete" [19] (p.271), since they occur as a consequence of their intended patronage, and thus may be superimposed. Nucleated facilities are central to a body of consumers (general accessibility); ribbon developments rely on through traffic (arterial accessibility) and specialised functional areas occupy areas of prestige or, in general, possessing some distinct attribute (special accessibility). The resulting classification system is not clearly hierarchical.

A post-hierarchical classification is assumed by Brown [20], based on form and function. Function is divided into general, specialist and ancillary, while form is separated into cluster (unplanned), cluster (planned), linear and isolated. The lack of an evident hierarchy is relevant since the imposition of hierarchy had been criticized [22,30], considering for example that new business models may develop beyond a strict hierarchical system. Thus, the post-hierarchical classification is designed so it can fit a wider range of scenarios.

Guy [21] presents a comprehensive analysis of retail classification methods, and the variables used to classify different centres. Even though no classification is proposed, this very thorough review is relevant for anyone who wishes to build upon the existing methods. Several dimensions used in retail classification are synthesized by the author, such as type of goods sold, store size, store ownership, catchment area, physical form, development history, development type, function and location. The use of hierarchical and non-hierarchical methods is also discussed, but ultimately, the author does not recommend any classification: these should be made from "a compromise between comprehensiveness and simplicity (…) too many categories make analysis and interpretation difficult; too few can obscure essential insights." [21] (p. 263).

After 2000, the unprecedented availability of data, along with technological advancements, allowed for innovative approaches in identifying centralities and hence, commercial centres. The goal is not necessarily to contribute to commercial classification, but the methods and variables that result from these analyses are useful for that purpose. Thurstain-Goodwin and Unwin [41] propose the use of density surfaces to visualize morphological and functional dimensions of urban centres (such as the presence of retail), allowing to relate them more easily. Density surfaces are also used to assess the multifunctional character of cities, and its diversity, considering land-use and hence the presence of economic activities [42]. Retail has also been integrated with graph-like measures of accessibility or space syntax [4,33,43] and related with street integration using percolation theory [44]. Big data is being progressively used to interpret retail location [45].

Considering specifically commercial structures, a renewed interest in classification studies, relying on strong quantitative analysis, namely by implementing different types of clusters analysis, is demonstrated by several authors. Hence, Araldi and Fusco [35] propose a "Retail Fabric Assessment" (RFA) in the French Riviera, Carpio-Pineda and Gutierrez [36] search for a "metropolitan geography of commercial spaces" in Madrid and Saraiva *et al.* [46] devise an analysis of "multi-diversity clusters of commercial activities" in Porto. The variables used to derive the clusters both in the French Riviera and in Porto are in line with classic retail classification studies, though the analyses take advantage of innovative methods, such as fractal counting in the former, and kernel density estimation in the later. In Madrid, the variables are derived from social network data, in line with the abovementioned research [44,45].

Finally, we are faced with the issue of classifying establishments into different categories to estimate diversity. Nelson [47] divides goods into search and experience goods, and while 2 categories are sufficient to compare, for example, online shopping adoption [48,49], they are not enough to estimate diversity. Jensen [50] reaches 5 categories, matching closely the U.S. Department of Labour Standard Industrial Classification (SIC) by using an adaptation of Potts' algorithm. Sevtsuk [51] uses the three-digit NAICS (North American Industry Classification System) system, which results into 13 individual categories of establishments. Saraiva *et al.* [46] use 5 categories, based on frequency of the purchase and the proximity of the location. The common ground is the purpose of the research: do we have a reason to separate search from experience goods? Do we want to create our own classification algorithm? Do we expect the frequency of the purchase to impact on the cluster composition? The answer to all of these is no. We want to approach the data with the minimum of preconceived ideas, expecting only that density, diversity and clustering will be sufficient to separate commercial areas that are structurally different. Our approach is hence similar to that of Sevtsuk [51]: considering the classification used by the Lisbon City Council [52], which is itself an adaptation of the official Portuguese Classification for Economic Activities ("CAE"), we group commercial activity in 9 categories, that are similar to the three-digit NAICS used by the author [51]. This process is explained in the following section.

3. Materials and Methods

3.1. Case Study

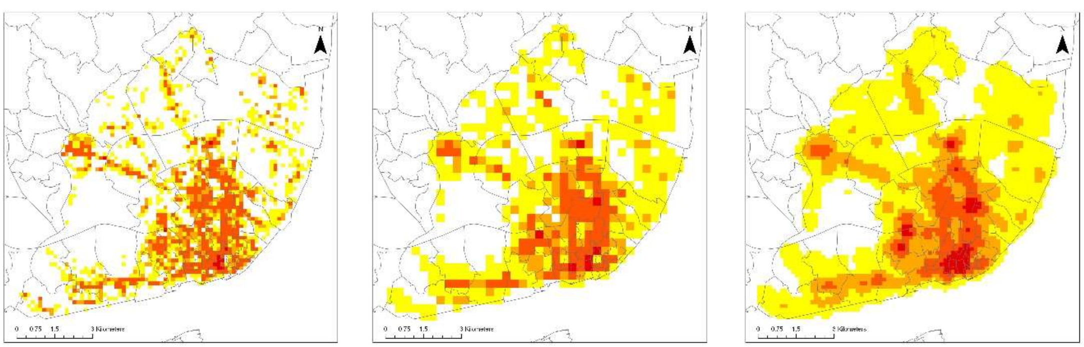
The case study is Lisbon, the capital and largest city in Portugal. Lisbon occupies an area of 100.05 km2 [53], with a population of 662.572 inhabitants in 1991 and 547.733 in 2011 [54]. It lies strategically on the river Tagus estuary, with its historic districts being located adjacently to the river which, for centuries, was the main support for goods transport. Commercial activity has been present at what is now the historical CBD since the 15th century, when the Portuguese Maritime Exploration gave Lisbon a global mercantile dimension [55,56]. Lisbon´s geographic location has since made it a desirable business and services location, ranking high amongst other European cities [57]. As expected, it presents a rich retail structure, which has been analysed by other authors: notably, the first classification of the commercial structure, using Berry's terminology [18,38], is included in a study of the functional structure of the city by Gaspar [56]. This work has subsequently been extended and updated [58,59]. Brown's non-hierarchical classification has also been applied to Lisbon [60]. Since these analyses have been based on classic commercial classification methods, they have relied heavily on data and observation, eventually focusing also on individuals' behaviours towards shopping. The stronger quantitative component that was found in more recent classification studies [35,36,46] is, in general, absent. Nevertheless, similar spatial patterns could be found between the functional analysis of Lisbon in 1994 [58] and the results of our analysis in 1995, which demonstrates, to some extent, that it is relevant to devise a classification that draws from both perspectives.

3.2. Data and Variables

For the following analyses, we used a geo-referenced database of establishments containing 14.046 commercial locations (establishments) for 1995, 16.378 locations for 2002 and 16.005 locations for 2010 [37]. For the purpose of the present research, we grouped commercial establishments (retail, and restaurants, cafes and bars) in 9 categories, based on the Lisbon City Council's classification [52], which is itself an adaptation of the Portuguese Classification of Economic Activities ("CAE"). The categories include occasional and frequently purchased goods, experience and search goods. The analysis of what may be found in each cluster is made *after* the cluster analysis, with the expectation that it, too, will help validate the obtained commercial structure.

Commercial activity was therefore divided into the following categories: Foodstuffs (supermarkets, bakeries, groceries and similar establishments); Personal Use Items (mainly clothing, clothing accessories and shoes); Household Articles (mainly furniture, home appliances, and home decoration items); Health and Hygiene (essentially pharmacies and optical shops, perfumes and cosmetics); Leisure items (sporting goods, bookshops, music stores, etc.); Other items (all other items non included in the remaining categories); Restaurants and similar establishments; Cafes and similar establishments; Bars and similar establishments.

Having access to data at point level provides several possibilities for estimating the variables to be used in the cluster analysis. The Modifiable Areal Unit Problem (MAUP) [61], though, points to the potential bias that is introduced in a statistical analysis when using areal units that are larger than the individual observation. The solution would be to use individual observations, which is mostly unfeasible: Araldi and Fusco (2019) use indicators that are based on reporting all data to the commercial establishments (points), but by using several radii. And the radii, too, imply areal limits. Other authors dealing with the issue of using very disaggregated information (and specifically, pertaining commercial location) have found their way around the problem by creating density surfaces [41,42,44]; Carpio-Pinedo and Gutiérrez [36] (p. 5) used 600m x 600m cells and consider that “the use of equal cells solves the Modifiable Areal Unit Problem”. All methods, though, require the definition of thresholds: radii to report data to points, or to create density surfaces; or the size of the cells. All these are areal limits. The problem is eventually unsolvable, but the point of Openshaw [61] was primarily for researchers to address the MAUP. Hence, a preliminary analysis was implemented considering both a 325m-sided square grid with an area of 105.625m2 (which is the average area of a census block in the Metropolitan Area of Lisbon considering 1991, 2001 and 2011 census blocks) and a 150m-sided square grid with an area of 22.500m2 (the average area of a block in the city of Lisbon)). A second approach was devised by using the Point Density Estimation tool from ArcGIS ArcMap® (which “calculates a magnitude-per-unit area from point features that fall within a neighbourhood around each cell” [62]) to produce a density surface considering a 300m radius and the same 150m-sided square grid. Though the results found using the 325m-sided square grid or the density surface were intuitive and, to a certain extent, easier to explain (ultimately, it would be simpler to explain the commercial structure considering data aggregated into parishes than disaggregated into blocks), the conclusion is that using one of those approaches would result in losing a great deal of information – which is the advantage of having it at point level. Therefore, we find that the grid of 22.500m2 square polygons is the best suited for our analysis, since it relates commercial points with blocks, which can then be described at the scale of the city. All information was associated with this grid, after which it became possible to calculate the density, diversity and clustering indicators.



(a) **(b)** **(c)**

Figure 1. Total Density of Establishments, 1995, using: (**a**) a 150m-sided square grid; (**b**) a 325m-sided square grid; (**c**) the Point Density tool of ArcGISPro® considering a 300m radius.

Density is simply the number of points per unit, as presented in Equation 1.

$$DS_i = \frac{NCi}{Ai} \tag{1}$$

where D_i is the density of a cell i, NR_i is the number of commercial establishments found at that cell, and A_i is the area of the cell. Since all cells have the same area, the interpretation will be the same considering simply consider DS = NC. This allows for a straightforward interpretation of density: instead of having (e.g.) 0.022 units/m2, or 22.2 units per 1.000m2, we will have 50 units per block (etc.). Although the initial intention was to use density of any commercial establishment, we decided to separate commercial establishments into those located in the high-street and those located in shopping malls, for a better interpretation of results. This is not different from using a variable such as area of a commercial lot, or stores-per-building [36] to separate street-level establishments from shopping malls, and is more straightforward. Density is thus represented by two variables: DSS and DSM, which are obtained in the same manner presented in Equation 1, but with DSS representing the density of commercial establishments at street level (high-street), and DSM representing the density at shopping malls.

As for Diversity, Batty *et al.* (2004) (p. 325) point that the simplest measure of diversity is "a direct or raw count of the number of activities", as presented in Equation 2 (here termed DVT, for "diversity-variety of types").

$$DVT_i = \sum_k b(i,k) \tag{2}$$

where k represents the different categories of commercial establishments. This Diversity indicator is the same as the Variety indicator used by Araldi and Fusco [35], and is equal to 0 where no type of commercial activity exists, and K where all categories are present. The indicator is intended to provide an immediate "snapshot" of the city in terms of diversity. What is being admitted is that a block with (e.g.) 9 categories and 9 stores, or a block with 9 categories and 90 stores share a degree of similarity – which is true, and relevant for a cluster analysis. This indicator alone, though, is insufficient to characterize diversity. Using the same example, if there are 9 categories and 90 stores, but in one block each type occupies 10 stores, and on another, one type occupies 82, then this dimension of (dis)similarity should also be assessed. For this effect, numerous diversity indicators have been proposed. In a very extensive revision of diversity indices, Jost [63] (p. 364) concludes that "most nonparametric diversity indices used in the sciences are monotonic functions of

$$\sum_{i=1}^{S} p_i^q \tag{3}$$

or limits of such functions as q approaches unity" (in our case, p is the proportion of each type s at each place i, and S is the total number of categories), while also mentioning the potential misinterpretations that arise from using overly complex indices to assess diversity. An example is given with the Shannon-Wiener index, which gives the uncertainty in the outcome of a sampling process. Though it can be used as a measure of diversity, the interpretation is not straightforward, and especially not as straightforward as it may be implied in some analyses. Therefore, we use the simplest indicator that can be derived from this function: the Simpson index [64], which measures the degree of concentration of individuals classified in different categories, with the notation being presented in Equation 4 (here termed DVC, for "diversity-concentration"):

$$DVCi = \sum_{i=1}^{K} p_i^2 \tag{4}$$

where p is the proportion of each type k at each place i, and K is the total number of categories. This index is usually considered a measure of concentration and equals 1 if only one type of establishment is present at block-level; values closer to 0 indicate a competitive use of the available stores by each of the K categories. For Diversity, we don´t distinguish between high-street and shopping mall: what this means is that we consider that a shopping mall with k categories equally distributed is ultimately not different from a high-street cluster of similar diversity: it may attract different customers, but not because of diversity.

For Clustering, we use the concept of spatial lag, calculated as a variable to address spatial effects [65,66]. The use of clustering is, to some extent, another way of dealing with the MAUP: in econometric models used to address spatial phenomena, it is often found that the dependent variable y in place i is affected by the independent variables in both place i and j. The explanation is that the outcome of y in i is not confined to what takes place in *i*. Addressing this issue is especially important when dealing with the MAUP, since it implicates that the limit of the unit of analysis is interfering with the results. In our case, a block may be coincident with a commercial cluster; but eventually, some clusters may include several blocks. Since we want to use a unit of the same size in the entire city, one way to address this issue is by creating a variable that is simply the mean value of the density of the adjacent cells. This implies that if two blocks are equal in density and diversity, but one has a much lower lag value (clustering), then this block is isolated (or integrated in a less dense structure), while the one with the higher lag value is integrated into a larger (or denser) structure. This variable will separate them – and the limit of the unit will be better fitted to the commercial structure. Therefore, clustering provides structure to the data, while also addressing the MAUP.

Formally, Clustering is obtained in two steps: first, a spatial weights matrix W is created as considering the weights w defined in 5.

$$w_{ij} = \begin{cases} 1 \; if \; d_{ij} \; \leq d^{*} \; for \; i \neq j \\ 0 \; otherwise \end{cases} \tag{5}$$

where d_{ij} is the distance between two cells, and d^* is the threshold distance – the distance at which cell j is considered to be independent of i. If this distance is calculated in cells, then if d is equal to (e.g.) 1, what is being assumed is that only the adjacent cells (the "first order neighbours") will be considered as being part of the same cluster. This would imply a cluster of 9 blocks using the 150mx150m grid. In order to ensure that even the largest clusters are addressed, second order neighbours are also considered (another reason for doing this is that we want our cluster analysis to be balanced, with 2 variables for each of the three aspects being dealt with; a cluster analysis of different aspects (density, diversity, clustering) where one is under or overrepresented might bias the outcome). Therefore, the Clustering variables are obtained according to the notation of Equation 6.

$$CL_d = \frac{\sum_j w_{ij} * Dens_j}{\sum_j Dens_j} \tag{6}$$

with CL_1 being the result of Equation 6 when d=1, and CL_2 the result when d=2. Only the density of high-street retail is considered for this variable: if a shopping mall is clustered with high-street retail, then it will be considered integrated in the general commercial structure (for simplicity, CL1 and CL2 will be used as the notations of CL_1 and CL_2).

Table 1 presents the summary of the variables considered in the cluster analysis.

Table 1. Summary of the Variables

Variable	Description		Mean (SD)		
			1995	**2002**	**2010**
DSS	Density of Commercial Establishments - High Street	number of establishments per block	8,79 (11,395)	8,69 (11,524)	7,87 (10,692)
DSM	Density of Commercial Establishments - Shopping Malls	number of establishments per block	1,10 (7,802)	1,62 (12,102)	1,43 (10,905)
DVT	Diversity (Variety)	number of categories per block	3,99 (2,571)	3,76 (2,504)	3,50 (2,410)
DVC	Diversity (Concentration)	Simpson index	0,45 (0,318)	0,46 (0,314)	0,48 (0,315)
CL1	Clustering (1st order)	average number of establishments in 1st order neighboring blocks	8,33 (8,618)	8,62 (9,387)	7,84 (9,015)
CL2	Clustering (2nd order)	average number of	7,45	7,63	6,92

	establishments in 2nd order neighboring blocks	(7,307)	(7,739)	(7,508)
N		1420	1492	1721

Note: The value in parenthesis is the standard deviation. N is the number of cells presenting commercial activity.

3.3. Methodology vou aqui

A straightforward definition of cluster analysis is proposed by Anselin [67]: "the goal of clustering methods is to achieve compact groups of similar observations that are separated as much as possible from the other groups". In other words, and considering any clustering technique, the purpose will be to maximize intra-cluster similarity and minimize between-cluster similarity, so that the entities contained in a cluster k are very homogenous, with the clusters being as heterogeneous as possible between themselves [67,68].

Xu and Tian [69] present a comprehensive survey of 19 categories of clustering techniques, from traditional to modern methods. The advantages and disadvantages of each method are presented and discussed. All modern methods present estimation time as a disadvantage, as a result of the increasing complexity of the algorithms. Also, being relatively new, the number of cases in which they have been applied may not be sufficient to grasp other disadvantages. The main advantage of recent methods is that they intend to ensure that a global optimal solution is attained, while traditional methods, such as partitioning and hierarchical methods, may produce only a local optimum solution. Anselin [67] that proposes that this disadvantage may be minimized by performing a sensitivity analysis to assess the robustness of the result. Another advantage of partitioning and hierarchical methods is that they have been extensively studied, allowing for common mistakes to be avoided and sounder results to be obtained, even if a global optimum cannot be assured.

For cluster analysis, using either hierarchical or partitioning methods, GeoDa [70] was considered to be the best software to implement our analysis, since it contains an extensive set of statistical tools for spatial analysis, as well as a graphic interface for visualizing maps. Since k-means clustering requires for the number of clusters to be defined before the algorithm is implemented, hierarchical methods were used in an exploratory way to set the initial number of seeds. Most of the solutions implied a structure of 5-6 clusters, and hence, 5-6 seeds to be used in the k-means.

The partitioning algorithm in k-means uses the squared Euclidean distance as the measure of dissimilarity, with the overall objective of allocating "each observation i to a cluster h out of the k clusters so as to minimize the within-cluster similarity W over all k clusters" [67], with the notation being presented in Equation 7.

$$minW = \sum_{h=1}^{k} \sum_{i \in h} \sum_{j \in h} \left\| x_i - x_j \right\|^2 \quad (7)$$

where x_i and x_j are p-dimensional vectors representing the p variables that describe the observations. The allocation process is based on iterative relocation that ensures that at each step, "the total across clusters of the within-cluster sums of squared errors (from the respective cluster means) is lowered." [67]. The algorithm stops when no improvement is possible. The first set of k cluster centres is obtained "by sampling k distinct observations ("seeds") from the full set of n observations" [67]. This sampling can be random, but in GeoDa, it is also possible to implement a second sampling procedure, referred to as k-means++. The concept behind k-means++ is squared distance weighting [71], which is implemented considering random sampling in the first iteration only; after that, each seed is located farther away from the original seed, "with the aim of providing better coverage of the sample" [67]. In both procedures, it is also possible to specify the number of rounds that take place before that first iteration. Thus, and for a sensitivity analysis, both sampling procedures, and different numbers of initial rounds, were implemented and compared. The ratio of between sum of squares to total sum of squares

(BSS/TSS) can also be compared for guidance, but if the structure is "stable" (small or no changes in cluster members), it can be assumed that the final solution has been attained [67].

Finally, and since the objective function is sensitive to the scale of the variables, all variables are standardized (z-standardization); to simplify interpretation, the variables will be presented in their original scale in the output tables.

4. Discussion

The cluster analysis was performed firstly on the 1995 dataset, using k-means with 5 and 6 seeds. The latter provided a more intuitive description of the commercial structure of Lisbon, considering previous studies [56,58] and retail classification theory, in general. Thus, we find that those 6 clusters correspond to 1) the CBD, or cells (blocks) presenting the same characteristics; 2) the "Expansion area of the centre" (a term used by Cachinho in 1994 [58] to describe a functional area that roughly corresponds to cluster 2 in 1995; it should be noted that naming this area "expansion area" is more evident if we consider that Lisbon is the centre of a metropolitan area, to which the CBD *and* its "expansion area" are central), which contains both nucleated and ribbon structures, superimposed or not, but equal in density, diversity and clustering; and 3) "Local Store Clusters" and 4) "Isolated Stores", not dissimilar from the ones of Proudfoot [17], and thus catering for local customers. Clusters 5 and 6 correspond to "new" business models: cells classified as 5 present commercial activity in shopping malls (DSM), but are integrated in the urban fabric; they are, to some extent, different miniatures of the centre, relying on a different business model – and thus termed "Integrated Shopping Malls". Cluster 6 contains just one observation, corresponding to the largest shopping mall in Lisbon in 1995, with more than 200 stores, and close to a motorway interchange. Cluster 6 is representative of a period described by Fernandes and Chamusca [8] as the "retail revolution post-modernity", characterized by the emergence of new retail concepts (namely shopping malls and big box retail), with a focus on car oriented accessibility, from which a centre-periphery competition started to emerge. This "revolution" started as early as 1969 in France, reaching Portugal in 1985, precisely with the opening of the shopping mall found in cluster 6 ("Amoreiras Shopping Center"). Cluster 6 thus corresponds to "Large Shopping Malls".

These conclusions are supported by the data presented in Table 2. The number of categories (DVT) is higher at the CBD, at Integrated Shopping Malls and at Large Shopping Malls. The density of high-street commercial activity (DSS) is similar in the Expansion Area of the centre and in the Integrated Shopping Malls, supporting the assertion that the latter structures are integrated in the urban fabric, revealing a relationship which is not necessarily "predatory" towards the high-street.

Expected validation comes also from analysing the original dataset. Table 3 presents the mean number of stores in each cluster cell for each of the 9 categories of commercial activity. The results support the proposed classification: in 1995, the mean number of Foodstuffs in the CBD equals 4.62, while in the Expansion Area of the centre this number is 2.56, and in the Local Store Clusters is 1.35. In comparison, Personal Use Items is 15.94 in the CBD, 2.59 in the Expansion Area and 0.55 in Local Store Clusters. Comparison goods tend to cluster, while daily purchase goods do not need to [4,17,46]. Local Store Clusters, which present an average of 3.67 categories (DVT), are essentially composed of daily purchased goods: Foodstuffs, Restaurants, Others and Cafes, as could be expected, and confirmed by anecdotal evidence. These 4 categories also present the highest mean values in the cells classified as Isolated Stores; the results of the cluster analysis are robust, even when using a classification that did not specifically consider purchase frequency. As for the new business models, the mean values are similar for most categories on the CBD and on the Integrated Shopping Malls, except notably for Restaurants, which present a lower mean value in the latter, since the Integrated Shopping Malls did not rely, in general, on large food courts. As for the Large Shopping

Malls, the density of stores presented is much higher than that of the Integrated Shopping Malls, and even of the CBD, especially considering Personal Use Items and Leisure Items.

Finally, a sensitivity analysis was performed. The process was initiated considering both k-means++ (the default in GeoDa) and a random sampling procedure, and 150 (default) and 1000 initial rounds. The ratio of between to total sum of squares did not change significantly, nor did the cluster members. The commercial structure of Lisbon in 1995, presented in Figure 2, is thus derived from the default k-means++ initiation, with 150 rounds.

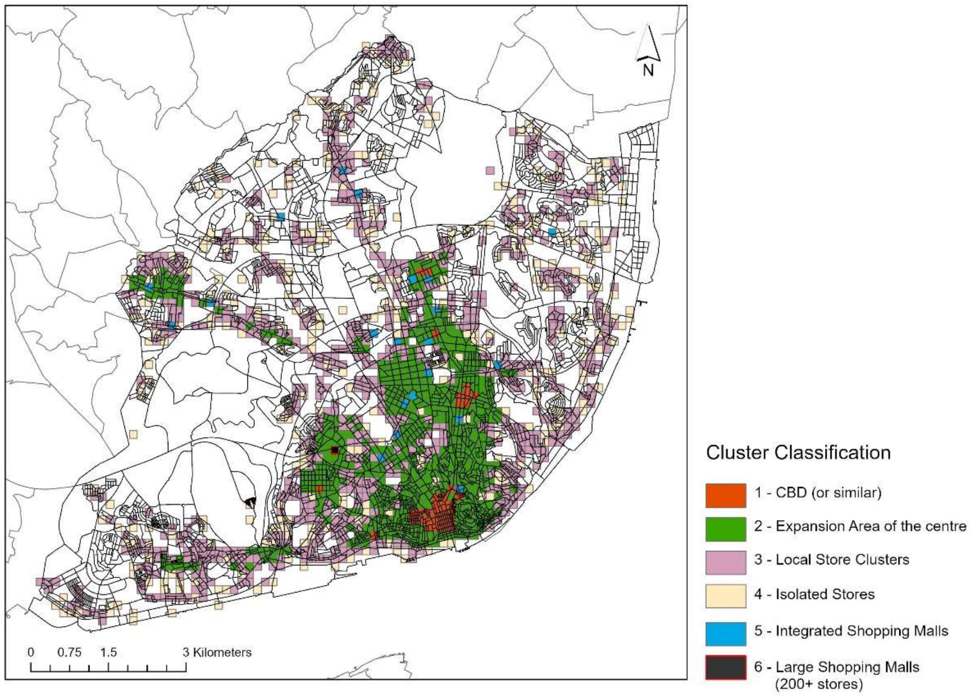


Figure 2. Cluster Classification – 1995.

The results for 2002 are similar to those of 1995. A better interpretation comes eventually from noticing the changes in the number of cells in each cluster (Table 2). The CBD decreased from 32 to 28 cells, and the Large Shopping Malls increased by 2 cells. A straightforward interpretation is that, in this period, these shopping malls competed directly with the CBD. Only the core (historic) CBD resisted, along with the densest, most diverse, and most central blocks of the remaining centres. But apart from the CBD, only the Integrated Shopping Malls decreased in number of cells (from 21 in 1995 to 15 in 2002), and the densities of both high-street (DSS) and establishments in shopping malls (DSM) increased, in general. This was therefore a period when commercial activity was growing, due partly to the expansion of the city north and northwest; but the Large Shopping Malls had started to strongly compete with the CBD, thus threatening city centre vitality. This would lead to tightening measures in urban retail policy in the 1990's: with the decline of independent retail in the city centres, the pro-developer attitude that had been adopted in the previous decades had to be replaced by a pro-town centre attitude by the local government [8]. In Portugal, this resulted in the implementation of Special Projects of Commercial Urbanism to engage retailers, retailers' trade associations and municipalities in promoting high-street retail [7]. In Lisbon, approximately 14 million euros were made available for the modernization of stores, public space interventions and promotional actions (among other more specific goals) between 1998 and 2007 [7]. The number of participants, though, is insufficient to establish any direct correlation with the changes that are visible in our analysis (190 establishments, located in different areas of the city [7]).

As for cluster composition, the diversity remained essentially the same (DVT), but the activity became more concentrated (DVC). The mean number of Foodstuffs decreased, but most of the other categories increased, especially Personal Use Items and Restaurants. The former increased so much that it became more likely to find a clothing store than a cafe, even at a Local Store Cluster; and the latter presented the highest mean value within the Isolated Stores, surpassing even Foodstuffs. Shopping (not for daily necessities, but eventually as a leisure activity) and eating out became so common that an isolated restaurant could still expect to find clients, and so could a clothing store located on a local store cluster. This can be attributed to a change in consumers' habits and the growing symbolism of shopping: from acquiring a good, to performing the act of shopping [36,72].

Here, too, a sensitivity analysis was performed with no significant changes to the structure. The final solution results from a k-means++ initiation, with 150 rounds, and is presented in Figure 3.

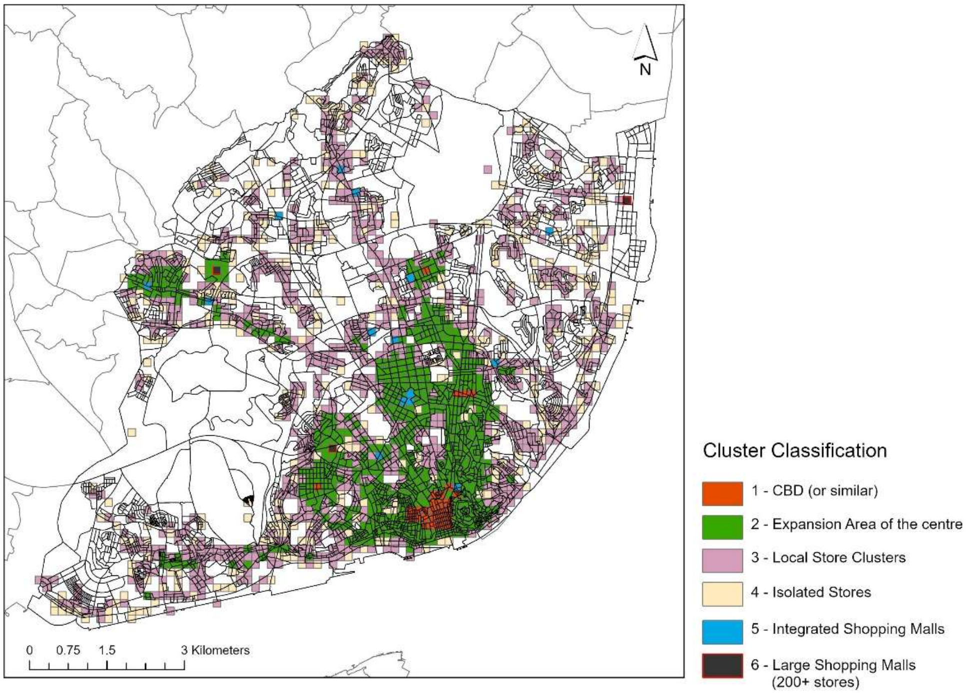


Figure 3. Cluster Classification – 2002.

For 2010, a 6-cluster structure did not emerge from the data. The resulting structure could not be related with the ones of 1995 and 2002. But a structure with 5 clusters was found to be related with those. The explanation is that between 2002 and 2010, the Integrated Shopping Malls became indistinguishable from the Expansion Area of the centre and the Local Store Clusters: with the closing of older shopping malls (pre-90's), the remaining shopping malls became too similar to those clusters to be separated into a different structure. The Expansion Area of the centre and the Local Store Clusters thus present a higher mean value for the density of stores in shopping malls (DSM) in 2010 when compared to 2002. The Local Store Clusters increased, pertaining number of cells, and so did the Isolated Stores, due to further urban expansion and consolidation of the city north and northwest. The CBD also increased (from 28 to 32 cells) but became less dense (DSS) and less diverse (more monofunctional) (DVS) in 2010. Looking at Table 3, most of the categories in the CBD decreased pertaining mean number of units. The exception is found in restaurants, cafes and bars, all of which increased in the CBD when compared to 2002. This had been noted already in 2001 [59]: the CBD was shifting north up the main avenues leading to it. A large part of businesses and services had left the CBD during the 1990's. Residents had started leaving it even before. The town center revival measures were clearly insufficient to address these issues, in Lisbon as in different other countries [10,11]. The combination of empty houses with the displacement of businesses and services, and the concentration of restaurants and bars, would make these areas fertile ground for the substitution of residents by tourists on the following decade: in Lisbon, as in Barcelona and New Orleans [73].

After a sensitivity analysis, we arrive at the result presented in Figure 3. The Integrated Shopping Malls were removed and hence, the clusters are numbered from 1 to 4, and then 6, to allow for a coherent comparison between all information being presented.

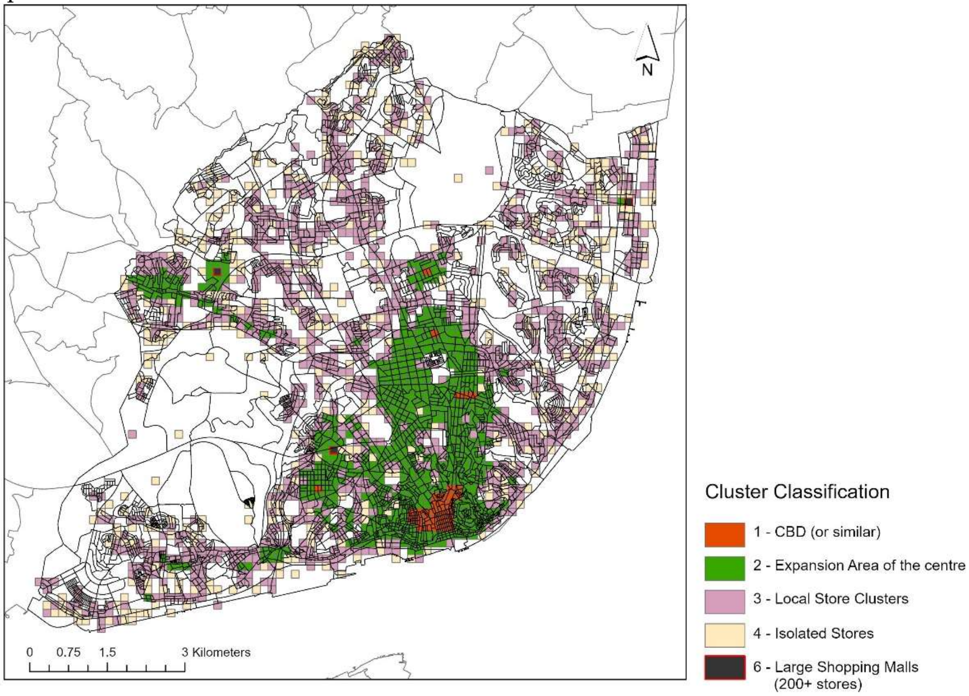


Figure 3. Cluster Classification – 2010.

The 2010 structure is not dissimilar from the one arrived at by Alho and de Abreu Silva [74] considering the Logistics Profiles concept [75]. The "CBD and the Expansion Area of the centre are related with the "CBD/Business centres/Large commercial" found by the authors, with an overlap of the Expansion Area of the centre and the Local Store Clusters with some of the "Residential areas with local trade". Considering that for 1995, a relationship could be established with the 1994 analysis of Cachinho [58], and though assuming the differences that are related with the scale of disaggregation and the purpose of each specific research project, the present analysis is, to some extent, validated by previous research. Thus, and considering our specific research purpose, we find that the method and the minimal set of variables are appropriate for commercial classification, since they allow to describe commercial structures in detail and in different periods in time. This suggests that the classification will be appropriate to describe also different spaces, which would allow for its generalization to other cities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2021

Table 2. Commercial Variables in 1995, 2002 and 2010 –Mean Value considering Cluster Membership

Cluster	1 – CBD			2 – Expansion Area of the centre			3 – Local Store Clusters			4 – Isolated Stores			5 – Integrated Shopping Malls		6 – Large Shopping Malls		
Year	1995	2002	2010	1995	2002	2010	1995	2002	2010	1995	2002	2010	1995	2002	1995	2002	2010
DSS	53,21 (24,030)	59,14 (28,341)	57,16 (22,636)	15,75 (9,651)	16,25 (9,910)	15,22 (9,715)	5,57 (3,975)	5,59 (3,936)	5,32 (3,868)	1,21 (0,830)	1,22 (0,816)	1,20 (0,804)	11,19 (13,578)	7,33 (11,998)	6,00	0,33 (0,577)	0,67 (1,155)
DSM	1,35 (3,773)	3,29 (8,210)	5,38 (16,070)	0,67 (2,357)	1,05 (3,634)	2,14 (8,252)	0,33 (1,597)	0,54 (2,336)	0,77 (3,895)	0,09 (0,536)	0,08 (0,509)	0,09 (0,416)	37,00 (12,377)	53,93 (17,698)	227,00	223,67 (77,261)	218,00 (73,980)
DVT	8,26 (0,751)	7,96 (0,793)	8,00 (0,842)	6,23 (1,937)	6,11 (2,030)	5,76 (2,166)	3,67 (1,695)	3,51 (1,667)	3,38 (1,621)	1,03 (0,172)	0,97 (0,224)	0,94 (0,250)	7,95 (0,498)	3,20 (2,731)	8,00	0,33 (0,577)	0,67 (1,155)
DVS	0,20 (0,045)	0,21 (0,047)	0,22 (0,056)	0,23 (0,101)	0,22 (0,094)	0,23 (0,107)	0,34 (0,135)	0,35 (0,134)	0,36 (0,129)	0,99 (0,056)	0,99 (0,041)	1,00 (0,025)	0,20 (0,031)	0,21 (0,027)	0,20	0,23 (0,021)	0,25 (0,003)
CL1	39,59 (14,393)	46,27 (17,003)	45,05 (14,005)	14,86 (6,494)	15,84 (7,818)	15,76 (7,791)	4,97 (3,703)	4,99 (3,827)	4,47 (3,259)	3,00 (3,486)	3,51 (4,127)	3,11 (4,024)	9,90 (7,434)	11,74 (10,661)	3,75	3,83 (1,688)	5,71 (1,041)
CL2	25,81 (11,800)	31,55 (12,410)	30,67 (12,687)	13,43 (6,493)	13,54 (7,170)	13,85 (7,079)	4,48 (4,096)	4,59 (4,254)	4,06 (3,586)	3,51 (4,199)	4,24 (5,208)	3,29 (4,126)	7,67 (5,043)	7,84 (5,390)	7,88	3,25 (4,732)	3,60 (3,737)
N	34	28	32	418	430	449	618	680	824	328	336	413	21	15	1	3	3

Note: The value in parenthesis is the standard deviation.

Table 3. Commercial Establishments in 1995, 2002 and 2010 - Mean Number of Units considering Cluster Membership

Cluster	1 – CBD			2 – Expansion Area of the centre			3 – Local Store Clusters			4 – Isolated Stores			5 – Integrated Shopping Malls		6 – Large Shopping Malls		
Year	1995	2002	2010	1995	2002	2010	1995	2002	2010	1995	2002	2010	1995	2002	1995	2002	2010
Foodstuffs	4,62	3,79	4,00	2,56	2,43	1,91	1,35	1,27	0,94	0,35	0,26	0,21	3,43	3,07	8,00	5,67	6,67
	(3,601)	(2,936)	(2,874)	(2,408)	(2,432)	(1,975)	(1,485)	(1,426)	(1,213)	(0,767)	(0,693)	(0,515)	(2,271)	(2,219)		(3,055)	(3,055)
Personal Use	15,94	18,96	18,19	2,59	2,76	2,85	0,55	0,53	0,50	0,02	0,03	0,01	16,05	21,47	82,0	89,67	96,33
	(11,714)	(13,956)	(12,923)	(3,215)	(3,551)	(4,276)	(0,966)	(1,027)	(1,391)	(0,173)	(0,312)	(0,120)	(6,344)	(7,818)	0	(25,541)	(32,316)
Household	5,38	5,43	3,91	1,74	1,91	1,47	0,44	0,45	0,44	0,07	0,07	0,05	5,19	5,93	30,0	26,67	16,33
	(3,782)	(4,086)	(3,392)	(1,881)	(2,103)	(1,800)	(0,856)	(0,820)	(0,872)	(0,261)	(0,253)	(0,265)	(2,732)	(1,831)	0	(12,220)	(7,767)
Health and Hygiene	2,50	3,07	3,16	0,72	0,78	0,85	0,21	0,20	0,26	0,03	0,03	0,02	2,67	2,80	6,00	9,67	11,67
	(2,874)	(3,208)	(2,665)	(0,866)	(0,932)	(1,110)	(0,459)	(0,458)	(0,581)	(0,172)	(0,162)	(0,146)	(1,494)	(1,971)		(2,887)	(1,155)
Leisure	6,00	7,61	6,63	1,74	1,87	1,84	0,59	0,61	0,64	0,11	0,10	0,08	7,76	9,80	39,0	25,00	19,67
	(4.000)	(5,412)	(4,164)	(1,600)	(1,677)	(1,696)	(0,950)	(0,986)	(1,058)	(0,473)	(0,432)	(0,414)	(4,024)	(4,539)	0	(12,490)	(8,327)
Other	7,47	8,57	9,72	2,56	2,62	2,67	0,91	0,96	0,87	0,26	0,29	0,20	6,19	7,40	29,0	25,67	25,00
	(4,794)	(5,131)	(7,574)	(2,236)	(2,430)	(2,498)	(1,154)	(1,225)	(1,192)	(0,608)	(0,601)	(0,559)	(3,737)	(7,395)	0	(12,342)	(10,583)
Restaurants	8,12	10,11	10,94	2,61	2,84	3,42	0,94	1,06	1,25	0,27	0,33	0,47	3,52	6,33	28,0	27,67	29,67
	(5,431)	(6,238)	(7,611)	(2,404)	(2,601)	(3,157)	(1,228)	(1,232)	(1,567)	(0,674)	(0,721)	(0,852)	(1,990)	(4,701)	0	(4,509)	(5,686)
Cafes	3,18	3,82	4,19	1,50	1,70	2,00	0,72	0,86	1,08	0,16	0,17	0,24	3,14	4,20	11,0	13,67	13,33
	(1,977)	(2,212)	(2,608)	(1,266)	(1,348)	(1,590)	(0,887)	(0,975)	(1,102)	(0,395)	(0,426)	(0,524)	(1,526)	(2,908)	0	(6,429)	(6,028)
Bars	1,35	1,07	1,81	0,42	0,39	0,33	0,21	0,18	0,11	0,04	0,02	0,00	0,24	0,27	0,00	0,33	0,00
	(1,276)	(1,538)	(4,231)	(0,921)	(0,898)	(0,895)	(0,447)	(0,432)	(0,451)	(0,202)	(0,153)	(0,070)	(0,700)	(0,799)		(0,577)	(0,000)

Note: The value in parenthesis is the standard deviation.

5. Conclusions

This paper sets at contributing to the classification of retail structures, considering a bottom-up approach, in line with recent research [35,36,46] on which cluster analysis was used to derive commercial classification. Cluster analysis presents the advantage of consistency: once the variables are determined, the process can be generalized, allowing for systematic comparisons. We find also that implementing a method that can be accepted by researchers coming from different backgrounds will help stimulate communication between different fields of research. This is achieved by devising an analysis rooted in social analyses of functional structures [17–19], while using variables that are based on classic concepts of spatial economy: density, diversity and clustering [22,26,76]. Using a minimal set of variables makes also for a parsimonious model and allows for the process to be reproduced with relative ease. This is seen as a strength of the proposed classification since it allows for researchers to adapt it to specific research purposes. In the present case, we compare already 3 different time periods, with the results finding direct relation, to some extent, with previous research on different subjects [58,74]. Considering this, we find that the proposed classification may eventually be generalized to other cities, and eventually adapted for other specific purposes. That the set of variables is parsimonious, simple and transparent will help to explain the results to policymakers and the public in general, which is essential for gaining their trust [77]. Hence, the classification has also the potential to be adopted by public sector practitioners, enabling the collection of empirical evidence by different entities, which could be used to improve the proposed methodology.

From the implementation of the proposed classification, it was possible to obtain and compare the commercial structure of Lisbon in 1995, 2002 and 2010. The results provided validation to the classification, not only from the relation with previous research [58,74] but also from the conclusions, that are supported by literature. The first conclusion is that a hierarchical structure is still present and, to some extent, according to CPT. Although shopping malls have been competing with the CBD for decades (especially after 1985, in Lisbon), it is reasonable to admit that the CPT-like structure will still be around for some time, as is confirmed by the work of other researchers [4,36,78]. What apparently changes is the cluster composition. Already in 2010, the CBD was presenting some specialization in commercial activities that can easily be related with tourism: restaurants, cafes and bars. This has since turned into a gentrification problem, as has happened in other cities [73,79] and shows the importance of a timely analysis in guiding policymaking. A second conclusion is that some shopping malls can eventually be integrated in the commercial structure of a city without displaying a predatory behaviour towards the high-street. This will need to be further explored. It is possible, for example, that the consumers take advantage of the food courts and parking facilities of the shopping malls, while at the same time shopping at the neighbouring commercial districts. This is relevant for policymaking, since it would imply that the possibility of parking a car has a positive effect on commercial patronage, which is not restricted to shopping malls, but also to the commercial districts where they are located. Hence, the implementation of car parking restrictions should eventually consider the result of this analysis: while it is necessary to limit car access to cities' centres, this may impact the high-street on a negative way. Other conclusions come from analysing the changes on commercial composition in different clusters. For example, a decreasing number of foodstuffs in local store clusters may lead to an increase in frequent shopping trips that cannot be made on foot. Finally, a decrease in diversity in any cluster will ultimately lead to a decrease in urbanity, considering what we came to expect from living in a city, being that social interaction, access to goods and services, or the feeling of belonging to a place.

As for the potential for modelling, using a minimal set of variables that excludes both specific characteristics of commercial activity (size of stores, development history, development type, etc.) and common explanatory variables in location models (market area characteristics, accessibility, etc.) is advantageous, since it allows for the classification itself (cluster membership) to be used in modelling. With cluster membership being independent from most explanatory variables (which solves potential issues of collinearity), while the clustering variables address the issue of spatial lag, the classification has the potential to be used as a dependent variable in location models aiming to explain the factors driving its changes, or as in independent variable in location models aiming to explain changes in commercial location.

Author Contributions: Conceptualization, R.C. and J.A.S.; methodology, R.C. and J.A.S.; software, R.C.; validation, R.C. and J.A.S; formal analysis R.C.; investigation, R.C.; resources, R.C. and J.A.S.; data curation, R.C.; writing—original draft preparation, R.C.; writing—review and editing, R.C. and J.A.S.; visualization, R.C.; supervision, J.A.S.; project administration, R.C. and J.A.S.; funding acquisition, R.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Fundação para a Ciência e a Tecnologia, through Rui Colaço's PhD grant (SFRH/BD/136003/2018).

Data Availability Statement: Publicly available datasets were used in this study. The data can

be found here: https://geodados-cml.hub.arcgis.com/.

Acknowledgments: The authors would like to thank Câmara Municipal de Lisboa for providing information on the data available at their website.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jacobs, J. *The Death and Life of Great American Cities*; **1961**. https://doi.org/10.1002/9781119084679.ch4.
2. Griffiths, S.; Vaughan, L.; Haklay, M. M.; Emma Jones, C. The Sustainable Suburban High Street: A Review of Themes and Approaches. *Geogr. Compass* **2008**, *2* (4), 1155–1188. https://doi.org/10.1111/j.1749-8198.2008.00117.x.
3. Jacobs, J. *The Economies of Cities*; Random House: New York, **1969**.
4. Sevtsuk, A. Location and Agglomeration: The Distribution of Retail and Food Businesses in Dense Urban Environments. *J. Plan. Educ. Res.* **2014**, *34* (4), 374–393. https://doi.org/10.1177/0739456X14550401.
5. Salgueiro, T.; Erkip, F. Retail Planning and Urban Resilience - An Introduction to the Special Issue. *Cities* **2014**, *36*, 107–111. https://doi.org/10.1016/j.cities.2013.01.007.
6. Fernandes, J. R.; Cachinho, H.; Ribeiro, C. *Comércio Tradicional Em Contexto Urbano: Dinâmicas de Modernização e Políticas Públicas*; Porto, **2000**.
7. Guimarães, P. O Planeamento Comercial Em Portugal : Os Projectos Especiais de Urbanismo Comercial, Universidade de Lisboa, **2015**.
8. Fernandes, J. R.; Chamusca, P. Urban Policies, Planning and Retail Resilience. *Cities* **2014**, *36*, 170–177. https://doi.org/10.1016/j.cities.2012.11.006.
9. Jacques, T. The State, Small Shops and Hypermarkets: A Public Policy for Retail, France, 1945–1973. *Bus. Hist.* **2018**, *60* (7), 1026–1048. https://doi.org/10.1080/00076791.2017.1413092.
10. Warnaby, G.; Alexander, A.; Medway, D. Town Centre Management in the UK: A Review, Synthesis and Research Agenda. *Int. Rev. Retail. Distrib. Consum. Res.* **1998**, *21* (1), 15–31. https://doi.org/10.1080/095939698342850.
11. Cook, I. R. Private Sector Involvement in Urban Governance: The Case of Business Improvement Districts and Town Centre Management Partnerships in England. *Geoforum* **2009**, *40* (5), 930–940. https://doi.org/10.1016/j.geoforum.2009.07.003.
12. Findlay, A.; Sparks, L. Town Centre First, Second, or Not at All? *T. Ctry. Plan.* **2014**, 158–160.
13. Cao, X. E-Shopping, Spatial Attributes, and Personal Travel. *Transp. Res. Rec. J. Transp. Res. Board* **2009**, *2135* (1), 160–169. https://doi.org/10.3141/2135-19.
14. Oberlo. Global Ecommerce Sales from 2017 to 2023 https://www.oberlo.com/statistics/global-ecommerce-sales (accessed Jul

14, 2020).
15. United Nations. *World Urbanization Prospects - The 2018 Revision*; New York, **2018**; Vol. 12. https://doi.org/10.4054/demres.2005.12.9.
16. Portugali, J.; Meyer, H.; Stolk, E.; Tan, E. *Complexity Theories of Cities Have Come of Age*; Portugali, J., Meyer, H., Stolk, E., Tan, E., Eds.; Springer: Berlin Heidelberg, **2012**.
17. Proudfoot, M. J. City Retail Structure. *Econ. Geogr.* **1937**, *13* (4), 425–428. https://doi.org/10.1126/science.11.277.620.
18. Berry, B. J. L. *Commercial Structure and Commercial Blight*; University of Chicago Press: Chicago, **1963**.
19. Davies, R. L.; Bennison, D. J. Retailing in the City Centre: The Characters of Shopping Streets. *Tijdschr. voor Econ. en Soc. Geogr.* **1978**, *69* (5), 270–285. https://doi.org/10.1111/j.1467-9663.1978.tb00864.x.
20. Brown, S. Retail Location: The Post Hierarchical Challenge. *Int. Rev. Retail. Distrib. Consum. Res.* **1991**, *1* (3), 367–381. https://doi.org/10.1080/09593969100000034.
21. Guy, C. M. Classifications Centres : Some Methodological Issues. *GeoJournal* **1998**, *45* (1), 255–264. https://doi.org/10.1023/A:1006960414161.
22. Hotelling, H. Hotelling - Stability in Competition. *The Econometric Journal*. 1929, pp 41–57.
23. Eaton, B. C.; Lipsey, R. G. The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition. *Rev. Econ. Stud.* **1975**, *42* (1), 27. https://doi.org/10.2307/2296817.
24. D'Aspremont, C.; Gabszewicz, J. J.; Thisse, J.-F. On Hotelling's "Stability in Competition." *Econometrica* **1979**, *47* (5), 1145–1150.
25. McLafferty, S. L.; Ghosh, A. Multipurpose Shopping and the Location of Retail Firms. *Geogr. Anal.* **1986**, *18* (3), 215–226. https://doi.org/10.1111/j.1538-4632.1986.tb00094.x.
26. Krugman, P. Increasing Returns and Economic Geography. *J. Polit. Econ.* **1991**, *99* (3), 483. https://doi.org/10.1086/261763.
27. Fujita, M.; Krugman, P.; Mori, T. On the Evolution of Hierarchical Urban Systems. *Eur. Econ. Rev.* **1999**, *43* (2), 209–251. https://doi.org/10.1016/S0014-2921(98)00066-X.
28. Weber, A. *Theory of the Location of Industries*; University of Chicago Press: Chicago, **1909**.
29. Dijk, J. Van; Pellenbarg, P. H. Firm Relocation Decisions in The Netherlands. **2000**, *219*, 191–219.
30. Elgar, I.; Miller, E. J. How Office Firms Conduct Their Location Search Process?: An Analysis of a Survey from the Greater Toronto Area. *Int. Reg. Sci. Rev.* **2010**, *33* (1), 60–85. https://doi.org/10.1177/0160017609331398.
31. Bodenmann, B.; Axhausen, K. *Synthesis Report on the State of the Art on Firmographics*; **2010**.
32. Waddell, P.; Ulfarsson, G. F. Accessibility and Agglomeration : Discrete-Choice Models of Employment Location by Industry Sector. *82nd Annu. Meet. Transp. Res. Board Washingt. DC* **2003**, No. 206.
33. Ravulaparthy, S. K.; Goulias, K. G. Characterizing the Composition of Economic Activities in Central Locations. *Transp. Res. Rec. J. Transp. Res. Board* **2014**, *2430* (1), 95–104.
34. Glaeser, E.; Kolko, J.; Saiz, A. The Consumer City. *J. Econ. Geogr.* **2001**, *1*, 27–50. https://doi.org/10.4337/9781784717445.00009.
35. Araldi, A.; Fusco, G. Retail Fabric Assessment: Describing Retail Patterns within Urban Space. *Cities* **2019**, *85* (September 2018), 51–62. https://doi.org/10.1016/j.cities.2018.11.025.
36. Carpio-Pinedo, J.; Gutiérrez, J. Consumption and Symbolic Capital in the Metropolitan Space: Integrating 'Old' Retail Data Sources with Social Big Data. *Cities* **2020**, *106* (October 2019), 102859. https://doi.org/10.1016/j.cities.2020.102859.
37. Câmara Municipal de Lisboa. Geodados - Recenseamento Comercial https://geodados.cm-lisboa.pt/datasets/6c40669f14c941779b48e2861767ead4_13 (accessed Sep 11, 2020).
38. Berry, B. J. *Geography of Market Centres and Retail Distribution*; Prentice-Hall: London, **1967**.
39. Berry, B. J. L.; Garrison, W. L. A Note on Central Place Theory and the Range of a Good. *Econ. Geogr.* **1958**, *34* (4), 304. https://doi.org/10.2307/142348.

40. Berry, B. J. L.; Garrison, W. L. The Functional Bases of the Central Place Hierarchy. *Econ. Geogr.* **1958**, *34* (2), 145. https://doi.org/10.2307/142299.
41. Thurstain-Goodwin, M.; Unwin, D. Defining and Delineating the Central Areas of Towns for Statistical Monitoring Using Continuous Surface Representations. *Trans. GIS* **2000**, *4* (4), 305–317. https://doi.org/10.1111/1467-9671.00058.
42. Batty, M.; Besussi, E.; Maat, K.; Harts, J. J. Representing Multifunctional Cities: Density and Diversity in Space and Time. *Built Environ.* **2004**, *30* (4), 324–337. https://doi.org/10.2148/benv.30.4.324.57156.
43. Porta, S.; Strano, E.; Iacoviello, V.; Messora, R.; Latora, V.; Cardillo, A.; Wang, F.; Scellato, S. Street Centrality and Densities of Retail and Services in Bologna, Italy. *Environ. Plan. B Plan. Des.* **2009**, *36* (3), 450–465. https://doi.org/10.1068/b34098.
44. Piovani, D.; Molinero, C.; Wilson, A. Urban Retail Location: Insights from Percolation Theory and Spatial Interaction Modeling. *PLoS One* **2017**, *12* (10), 1–13. https://doi.org/10.1371/journal.pone.0185787.
45. Zachariadis, V.; Vargas-ruiz, C.; Serras, J.; Ferguson, P.; Ferguson, P. Decoding Retail Location : A Primer for the Age of Big Data and Social Media. *Cupum 2015* **2015**, No. July. https://doi.org/10.13140/RG.2.1.1826.3520.
46. Saraiva, M.; Marques, T. S.; Ribeiro, D.; Ribeiro, P. Multi-Diversity Clusters of Commercial Activities in the City of Porto : How Neighbourhood Concentrations Shape the Municipal Hierarchy. In *La Geografía de las Redes Económicas Y la Geografía Económica en Red*; Logroño, M. P. A., Marques, T. S., Santos, H., Eds.; Faculdade de Letras da Universidade do Porto, Asociación de Geógrafos Españoles: Porto, **2020**; pp 209–217. https://doi.org/10.21747/9789898969460/geoa20.
47. Nelson, P. Information and Consumer Behavior. *J. Polit. Econ.* **1970**, *78* (2), 311–329. https://doi.org/10.1086/259630.
48. Zhen, F.; Du, X.; Cao, J.; Mokhtarian, P. L. The Association between Spatial Attributes and E-Shopping in the Shopping Process for Search Goods and Experience Goods: Evidence from Nanjing. *J. Transp. Geogr.* **2018**, *66* (November 2017), 291–299. https://doi.org/10.1016/j.jtrangeo.2017.11.007.
49. Schmid, B.; Axhausen, K. W. In-Store or Online Shopping of Search and Experience Goods: A Hybrid Choice Approach. *J. Choice Model.* **2018**, *31* (March 2018), 156–180. https://doi.org/10.1016/j.jocm.2018.03.001.
50. Jensen, P. A Network-Based Prediction of Retail Stores Commercial Categories and Optimal Locations. *Phys. Rev. E Stat. Nonlinear, Soft Matter Phys.* **2006**, *74*.
51. Sevtsuk, A. Location and Agglomeration: The Distribution of Retail and Food Businesses in Dense Urban Environments. *J. Plan. Educ. Res.* **2014**, *34* (4), 374–393. https://doi.org/10.1177/0739456X14550401.
52. Câmara Municipal de Lisboa. *Recenseamento Dos Estabelecimentos de Comércio a Retalho e Restauração e Bebidas Da Cidade de Lisboa*; CML: Lisboa, **2009**.
53. Direção Geral do Território. Carta Administrativa Oficial de Portugal https://www.dgterritorio.gov.pt/cartografia/cartografia-tematica/caop (accessed Feb 14, 2021).
54. Instituto Nacional de Estatística. Página de download de Informação Geográfica http://mapas.ine.pt/download/index2011.phtml (accessed Aug 1, 2020).
55. Oliveira, V.; Pinho, P. Lisbon. *Cities* **2010**, *27* (5), 405–419. https://doi.org/10.1016/j.cities.2009.12.008.
56. Gaspar, J. *A dinâmica funcional do centro de Lisboa*; Livros Horizonte: Lisboa, **1985**.
57. Rubalcaba, L.; Gallego, J.; Gallo, M. T.; Garrido, R. Business Services Location and Market Factors in Major European Cities. *Cities* **2013**, *31*, 258–266. https://doi.org/10.1016/j.cities.2012.06.022.
58. Cachinho, H. O Comércio a Retalho Na Cidade de Lisboa. Reestruturacão Económica e Dinâmicas Territoriais. *Finisterra* **1994**, *29* (57), 119–144.
59. Salgueiro, T. B. *Lisboa, periferia e centralidades*; Celta: Oeiras, **2001**.
60. Cachinho, H. Urban Retail Dynamics From Shopping Spaces to Consumer Places. *Retail Plan. Resilient City* **2011**, No. March, 147–168.
61. Openshaw, S. *The Modifiable Areal Unit Problem.*; Geo Books: Norwich, England, **1984**. https://doi.org/10.1002/9781118526729.ch3.

62. ArcGIS. Point Density (Spatial Analyst) https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/point-density.htm (accessed Feb 14, 2021).
63. Jost, L. Entropy and Diversity. *Oikos* **2006**, *113* (2), 363–375. https://doi.org/10.1111/j.2006.0030-1299.14714.x.
64. Simpson, E. H. Measurement of Diversity [16]. *Nature*. 1949, p 688. https://doi.org/10.1038/163688a0.
65. Anselin, L. *Spatial Econometrics : Methods and Models*; Springer Netherlands: Dordrecht, **1988**.
66. Anselin, L. Under the Hood Issues in the Specification and Interpretation of Spatial Regression Models. *Agric. Econ.* **2002**, *27*, 247–267.
67. Anselin, L. Cluster Analysis https://geodacenter.github.io/workbook/7bk_clusters_1a/lab7b.html#fn1 (accessed Dec 20, 2020).
68. Hair, J. F.; Black, W. C.; Babin, B. J.; Anderson, R. E. *Multivariate Data Analysis*; Pearson Education Limited, **2014**.
69. Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2* (2), 165–193. https://doi.org/10.1007/s40745-015-0040-1.
70. Anselin, L. *Exploring Spatial Data with GeoDa: A Workbook*; Urbana-Champaign, Ill., **2005**. https://doi.org/http://www.csiss.org/.
71. Arthur, D.; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms* **2007**, *07-09-Janu*, 1027–1035.
72. Cachinho, H. Consumerscapes and the Resilience Assessment of Urban Retail Systems. *Cities* **2014**, *36*, 131–144. https://doi.org/10.1016/j.cities.2012.10.005.
73. Cócola-Gant, A. Tourism and Commercial Gentrification. *ideal city. Between Myth reality.* **2015**, No. August, 27–29.
74. Alho, A. R.; de Abreu Silva, J. Utilizing Urban Form Characteristics in Urban Logistics Analysis: A Case Study in Lisbon, Portugal. *J. Transp. Geogr.* **2015**, *42* (December), 57–71. https://doi.org/10.1016/j.jtrangeo.2014.11.002.
75. Macário, R. Modeling for Public Policies Inducement of Urban Freight Business Development. In *Freight Transport Modelling*; Ben-Akiva, M., Meersman, H., Voorde, E. Van de, Eds.; Emerald Group Publishing Limited, **2013**; pp 405–432. https://doi.org/10.1108/9781781902868-020.
76. Lösch, A. *The Economics of Location. (Translated by William H. Woglom with the Assistance of Wolfgang F. Stopler in 1954)*; New Haven, Yale University Press, **1940**.
77. Miller, E. J. Integrated Urban Modeling : Past , Present , and Future Integrated Urban Modeling : Past , Present , and Future. *J. Transp. Land Use* **2018**, *11* (1), 387–399.
78. Reigadinha, T.; Godinho, P.; Dias, J. Portuguese Food Retailers – Exploring Three Classic Theories of Retail Location. *J. Retail. Consum. Serv.* **2017**, *34* (September 2016), 102–116. https://doi.org/10.1016/j.jretconser.2016.09.015.
79. Larraz, B.; García-Gómez, E. Depopulation of Toledo's Historical Centre in Spain? Challenge for Local Politics in World Heritage Cities. *Cities* **2020**, *105* (March), 102841. https://doi.org/10.1016/j.cities.2020.102841.