Complex & Intelligent Systems (2022) 8:83–106
https://doi.org/10.1007/s40747-021-00335-8

ORIGINAL ARTICLE

Identification of robust retailing location patterns with complex network approaches

Rosa María Sánchez-Saiz[1] · Virginia Ahedo[1] · José Ignacio Santos[1] · Sergio Gómez[2] · José Manuel Galán[1]

Received: 30 September 2020 / Accepted: 9 March 2021 / Published online: 8 April 2021

Abstract
The problem of location is the cornerstone of strategic decisions in retail management. This decision is usually complex and multidimensional. One of the most relevant success factors is an adequate balanced tenancy, i.e., a complementary ecosystem of retail stores in the surroundings, both in planned and unplanned areas. In this paper, we use network theory to analyze the commercial spatial interactions in all the cities of Castile and Leon (an autonomous community in north-western Spain), Madrid, and Barcelona. Our approach encompasses different proposals both for the definition of the interaction networks and for their subsequent analyses. These methodologies can be used as pre-processing tools to capture features that formalize the relational dimension for location recommendation systems. Our results unveil the retail structure of different urban areas and enable a meaningful comparison between cities and methodologies. In addition, by means of consensus techniques, we identify a robust core of commercial relationships, independent of the particularities of each city, and thus help to distinguish transferable knowledge between cities. The results also suggest greater specialization of commercial space with city size.

Keywords Retailing · Location · Balanced tenancy · Complex networks · Community analysis · Commercial structure

Introduction

The geographical distribution of economic activity is a long-standing problem of great interest in Economics and Industrial Organization [1–3]. In words of Krugman [4], "economic activities are definitely not homogeneously distributed in space". The study of spatial-economic distribution patterns has been addressed by different branches of Economics such as Economic Geography, Urban Economics and Industrial Organization—among others. Part of this research has focused on various theories about how companies take individual spatial-economic decisions: based on potential demand, location of competitors, presence of specific externalities, transport costs, or proximity to customers and suppliers [5–9]. Another relevant approach has emphasized the need for a more global perspective to explain the formation and development of economic aggregation [10–13].

Understanding the spatial distribution of economic activity and its relationship to the problem of location at any scale, is undoubtedly a critical problem for the company itself, for the economy as a whole, and for those responsible for industrial and commercial policies in any region. Both for large companies—for which retailing constitutes the final step of their supply chain—and for SMEs, in-store sales still represent the most important distribution channel in terms of revenue [14]. However, the impact of location on a business's success is even more relevant for retail shops. In fact, the location decision is considered one of the most critical strategic decisions in retailing, especially given how difficult it is for competitors to imitate [14–16].

There are many relevant factors to be taken into account when analyzing both the initial location of the retail area and the subsequent specific location of a retail store within such area. Important variables are the size of the population and its socioeconomic characteristics, the availability of workers, the proximity to suppliers, the possibility of promotion, the base economy of the area, the availability and cost of space, taxes, and other regulations [14, 16]. Once the area has been

✉ José Manuel Galán
jmgalan@ubu.es

1 Departamento de Ingeniería de Organización, Universidad de Burgos, Escuela Politécnica Superior. Edificio A. Avenida Cantabria s/n, 09006 Burgos, Spain

2 Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Avda. Països Catalans 26, 43007 Tarragona, Spain

analyzed, it is essential to evaluate the type of location structure (isolated store, unplanned business district—central, secondary or neighborhood—or planned shopping center), its pedestrian accessibility, vehicle and/or public transportation facilities, the size of the establishment, etc.

In the last survey on location and decision-making carried out in the United Kingdom, ten methods of analysis—arranged in three categories—were included [17]: (i) Comparative: thumb rule, checklist, ratios and analogies; (ii) Predictive: Multiple regression, discriminant analysis, clustering analysis and gravitational models; and (iii) Expert-based: Expert systems and neural networks. The interviews carried out revealed that a large part of location decisions are based on intuition and on the simplest and most subjectivity-dependent methods. This behavior is not surprising since the number of aspects to be considered is large, and, in some cases, difficult to compare or quantify. This has led to an increasing use of multi-criteria decision methods as selection mechanisms, for example, the hierarchical analytical process [18, 19].

At the time of selecting a location, one of the most relevant factors is the nature of the competitive behavior in the sector under consideration. The empirical study of different strategies that consider the location of competitors has been widely investigated. Strategies among retailers in the same sector are usually classified into three types: *avoidance*, in which the retailer seeks a site as far away as possible from competitors, while trying to capture some competitive advantage (nearby population density, accessibility in the transportation network, etc.); *confrontation*, in which the objective is to get as close as possible to competitors; and *predation*, which consists in installing oneself in the gaps left by competitors and trying to capture customers with a price strategy [16]. These different strategies often appear as a result of the trade-off between the increase in the size of the pie to be shared (consequence of the increase in demand due to preference and taste uncertainty and to expectations of lower prices), and the reduction of the share of the pie, resulting from more direct competition in the sector [20].

Notwithstanding, the functioning of businesses within a sector is also significantly influenced by the structure and balance of nearby businesses operating in other sectors, i.e., the *balanced tenancy*. Given that the balanced tenancy can complement one's business ecosystem, it is increasingly seen as an important factor in the localization process [14]; in fact, it is one of the key elements in planned centers and may have boosted synergies in unplanned areas using and being used by recommendation systems of points of interest (POI) [21–23], geolocation-based marketing and location-based social networks (LBSN) [24–28].

Research objective

The main objective of the present paper is to provide a comprehensive analysis of the spatial relations between retail businesses in urban environments, to identify robust patterns at different confidence and resolution levels. To this end, we analyze previous methodologies used to formalize the relational dimension, identify their problems and propose and compare alternatives. We consider the adaptive and strategic aspects, while placing particular emphasis on how the interactions—both positive and negative—between the different businesses determine the configuration of the economic environments in cities. We follow a complex systems methodological approach [29, 30]. Indeed, cities, as highly organized spatial structures, each with individual characteristics consequence of its own historical, economic, geographical and political environments, are considered one of the most representative examples of complex adaptive phenomena [31].

The remainder of this paper is organized as follows. In the next section, we analyze the "Related literature" and establish how our work differs from previous research. Next, we provide some "Theoretical background", initially focusing on a pioneering network approach by Jensen [32, 33] for the analysis of the retailing location structure. Then, we identify some potential technical problems of such approach, and propose two alternative network methods with different implicit assumptions. Subsequently, we present the dataset used in our case study. "Results and discussion" are then provided, first for the analyses conducted at the city level, and then for the consensus approaches used to identify robust commercial interactions at two different threshold levels. Finally, the main "Conclusions" are presented in the last section.

Related literature

Given the importance of the problem of location in retailing, many approaches have been developed to try to address it. From the famous Reilly's law of retail gravitation [34], which states that retail stores "attract" customers by the size of the trading area and inversely by the square of the distance, to its many refinements and extensions—perhaps the most famous of which is the Huff model [35]—and various other model-based methodologies [36, 37].

One of the fundamental problems with location in retailing, and one that classic models do not capture to its full extent, is that determining the success of a particular location is a multidimensional problem including many factors. There is currently a consensus that the business site selection problem is a Multi-Criteria Decision-Making problem (MCDM). Many examples can be found that use this approach to tackle the problem: the selection of a shopping center in Turkey [38], the selection of warehouses for agricultural products [39], the location of gas stations [19, 40], the opening of a new supermarket in Spain [41], the selection of Surface Water Treatment [42] or solar energy plants [43], among many others.

The relational dimension between a particular business and the economic activity in the environment is one of the most influential factors in the decision. As established by Hidalgo et al. [44] in what they call the *principle of relatedness*, economic activities are strongly influenced by others already present in the area. In the context of Geography, this principle is also known as Tobler's First Law of Geography, and establishes that "near things are more related than distant things" [45].

One of the most relevant and pioneering works formalizing and capturing the relational dimension in the location retailing problem is that of Jensen [32, 33]. In his research, he proposes two coefficients to empirically characterize the interactions between retail activities in a given urban region (see the next section for a more detailed explanation). His results are significant in two aspects: in explanatory terms, helping to understand the business dynamics and structure that occur in cities, and from an applied point of view, defining a quality index for potential new locations depending on the category or sector of the business.

Partially influenced by Jensen's work, Karamshuk et al. [46] select a set of characteristics specific to each potential location (e.g., area population, mobility, competitive businesses) to identify the best places to locate new stores. Their work focuses on three well-known food chains (Starbucks, McDonalds and Dunkin' Donuts) in New York. Their methodology is innovative because it joins information from several layers. In particular, they process the dataset to obtain the retail quality of a geographic area from Jensen's metrics. They use them together with other features obtained from a location-based social network such as Foursquare, an approach which has subsequently been applied with other social networks such as Facebook [47] or Baidu [48]. Their results show that geographic quality features obtained from Jensen are one of the best individual metrics for prediction. Using all the features in conjunction with supervised learning algorithms, they are able to identify the relevant potential location areas with good results. This line of work has been continued by Chen et al. [22]. They similarly include location-based social network data to develop a retail store recommendation system for the coffee retail industry using the New York and Tokyo datasets from Foursquare, and the Tainan city dataset from Facebook.

The coefficients proposed by Jensen have also been applied as relevant geographic features to capture the relational quality of potential store sites in the case of bike-sharing station placement [49, 50], and for the planning of new hotels [51]. In all cases they were used to feed location recommendation systems algorithms.

In Rohani and Chua [52], the authors use location data from Google in the Klang Valley area in Malaysia together with other location features—e.g., parking lots, nearby roads, proximity to housing, public transport—, and train a decision tree model to predict suitable sites. In this work, the concept of proximity and the features of nearby businesses are also based on the radius concept of the Jensen's model.

Guo et al. [53] create a system to use the knowledge obtained from the location of stores belonging to chain businesses in certain cities, and use it in other cities where there is no previous experience. Specifically, they use a recommendation system based on collaborative filtering, and successfully apply it using data from hotel chains. One of the problems the authors explicitly mention is that different cities can have other characteristics and rating distributions, so the transfer of knowledge from one to another is not trivial.

Recently, Hidalgo et al. [30] have studied the neighborhood-scale agglomerations of the amenity space from a network perspective. In their work, they identify amenities that are most likely to be found in the same neighborhood from a dataset of 47 US cities. From the data, they establish a model for Boston that allows identifying communities where specific amenities are over- or under-supplied.

Our approach

Our work complements the research mentioned above in two different aspects. Firstly, our results unveil the retail structure of different urban areas, and, to our knowledge, this paper is innovative in the use of a systematic approach that enables a meaningful comparison between diverse cities, complete retail sectors, and complementary methodologies. In addition, employing consensus network techniques, we identify a robust core of commercial relationships, at different resolution levels, partially independent of the particularities of each city. These results advance the types of categories and sectors in which retail knowledge can be transferred among cities. Secondly, part of the frontier research applying information fusion and multidimensional data to the location retailing problem makes use of Jensen's approach as a feature extraction mechanism to feed the learning algorithms and recommender systems. In this paper, we identify some of the technical problems of the methodology, and propose and analyze alternatives that can either be used alternatively or in conjunction with Jensen's metric as tools to characterize the commercial structure of any potential candidate site. They can also be used to formalize the relational dimension of the complementary ecosystem of retail stores in the surroundings in multi-criteria decision methods.

Theoretical background

Several authors have explored the problem of localization of retail stores using a Complex Systems approach. In two influential works, Jensen [32, 33] proposes two different coefficients to quantify the interactions between retail activities using only location data. Such coefficients of interaction are then used to define a network of the relational structure of retail stores, which can be explored using community detection algorithms.

Jensen's coefficients intend to quantify the two most common interactions between retail stores: intragroup (relation between stores in the same commercial category) and intergroup (relation between stores in different categories). The underlying idea is to compute the deviation of the spatial empirical distribution of retail shops from purely random and non-interacting distributions.

To be clear, let us consider the set T of all the stores in a certain area, and focus on two particular types of retail business. Let A be the set of stores of one type, and B the set of stores of the other type. Let $N_S(p,\ r)$ denote the number of stores in set S within a radius r from store p (excluding p itself), i.e., $N_S(p,\ r) = |\{x \in S\backslash\{p\};\ d(p,\ x) \leq r\}|$.

Jensen [33] defines the average local concentration of type A businesses at a certain radius r as $\frac{1}{|A|}\sum_{a\in A}\frac{N_A(a,\ r)}{N_T(a,\ r)}$ (where the fraction 0/0 is assumed to be equal to 1), and the global concentration as $\frac{|A|-1}{|T|-1}$ (where the effect of the focal business is eliminated by subtracting 1).

On the basis of the foregoing, the intra-coefficient M_{AA}, which is intended to measure the independence between points of the same type, is defined by Jensen [33] as the average local concentration divided by the global concentration, which yields:

$$M_{AA} = \frac{|T| - 1}{|A|(|A| - 1)}\sum_{a\in A}\frac{N_A(a,\ r)}{N_T(a,\ r)}. \tag{1}$$

Similarly, the inter-coefficient M_{AB} quantifies the relationship between two different types of retail stores (A and B). Formally, M_{AB} is the ratio between a local concentration of B-type stores around A-type stores defined as $\frac{1}{|A|}\sum_{a\in A}\frac{N_B(a,\ r)}{N_T(a,\ r)-N_A(a,\ r)}$, and the same concentration over the whole area, i.e., $\frac{|B|}{|T|-|A|}$. Thus:

$$M_{AB} = \frac{|T| - |A|}{|A||B|}\sum_{a\in A}\frac{N_B(a,\ r)}{N_T(a,\ r) - N_A(a,\ r)}. \tag{2}$$

For both M_{AA} and M_{AB}, values greater than 1 are interpreted as attraction, whereas lower values imply a repulsion tendency.

The significance of the empirical results is determined by checking against the respective null models proposed by Jensen, which are based on Monte Carlo sampling. More specifically, in the null model for the intra-coefficient, for each commercial category A, M_{AA} is obtained after uniformly randomizing the locations of all A shops over all possible locations, while preserving their total number $|A|$. Besides, the total number of retail stores belonging to the other categories and the location of the commercial premises in the city are also retained (each shop keeps the same number of commercial establishments in its neighborhood; however, their categories may be different).

As for the null model for the inter-coefficient M_{AB}, it is calculated by keeping fixed the location of the A-type retail stores (hence controlling the pattern of the A category) and by randomly and independently redistributing all the other retail shops in the remaining locations (random sampling without replacement of all stores except for those belonging to category A). Henceforth, in the M_{AB} null model, in addition to maintaining the location of the A establishments, we are again preserving the total number of stores belonging to each retail category and the original geospatial distribution of commercial premises.

Thereafter, the empirical value of each coefficient is compared with the percentiles of its respective null distribution—obtained by Monte Carlo sampling—to assess its significance. Finally, regarding the inter-coefficient, in accordance with Marcon and Puech [54], an interaction AB is considered to be significant if and only if both values M_{AB} and M_{BA} are significantly different from their respective null hypotheses.

Eventually, to construct the network of the interactions between retail stores, Jensen's proposal consists of establishing the $\log M_{AB}$ as the weight of the edge between nodes A and B. Hereinafter, Jensen illustrates his methodology on a dataset of the city of Lyon containing information from 8500 stores, and assesses its community structure with an adaptation of Potts algorithm [55], which is designed to handle weighted graphs with both positive and negative weights.

This pioneering methodology ingeniously captures and formalizes the intuition of empirical balanced tenancy in cities. However, it presents some technical problems in its calculation and application, especially in sectors with few commercial establishments, something that occurs with some frequency in small cities. On the one hand, the use of the logarithmic function to convert coefficients into signed weights has two downsides: (i) when the coefficients are 0, the interaction weights become negative infinity, which makes the method impractical unless a large and negative value is arbitrarily taken instead (in this paper, we have replaced the negative infinity value with the floor function of the highest finite repulsion force found in the city); and (ii) the behavior of the logarithmic function is different for positive and negative values; while the behavior for attraction (positive values) grows very slowly, the growth towards repulsion (negative values) is very fast; this asymmetry makes the interpretation of the coefficients difficult, and the detection of communities in the resulting network potentially problematic. On the other hand, in the calculation of the values of the coefficients, indeterminations often appear in the fractions, e.g., if a business is isolated. In these cases, fractions are assumed equal to 1 [33], although one could alternatively envisage taking the fraction 0/0 as equal to 0, thus considering no interactions for isolated stores. Since the sums do not depend on the volume of retail stores, these decisions can have a substantial weight on the results when averaging, sometimes generating artifacts in the results.

Nevertheless, it is important to stress that in the original proposal, part of these inconsistent correlations were removed when assessing the significance of the relationships found between categories. More specifically, in accordance with Jensen [33], although the method is not necessarily symmetric, it is prescriptive to check each relationship in both senses, that is, M_{AB} and M_{BA}, since—as previously stated—an interaction will only be considered significant if both coefficients differ significantly from their null values; otherwise, important artifacts could appear.

Once the interaction matrix is found, Jensen's relational quality index, frequently used as a feature extraction method for recommendation systems and learning algorithms, consists of evaluating the quality of a potential site as the amount of businesses with attraction or repulsion weighted by its signed weight.

Alternative methods proposed

In the present contribution, we propose two alternative measures that solve some of the problems of the intra- and inter-coefficients. Both alternatives are network approaches consisting of inducing a commercial spatial network of the city. In this network, the nodes are the stores, and there is an undirected link between them if they are at a distance lower than the radius proposed in Jensen's model. This network may comprise different unconnected components. Besides, the nodes are endowed with the attribute of the commercial category they belong to. So as to establish the interactions between the different retail types, we count the number of edges in the network where the ends join each pair of categories. The results can be stored in a matrix, initially symmetric, in which each row and column represent a commercial category, and the value represents the number of edges in the city that connect them. Understandably, this number depends on the frequency of the retail store type. To know whether the empirical relationships found in the commercial network of the city are above or below the expected value, or if they can be considered significant, it is necessary to compare them against a null model. At this point, our analysis can be divided into two depending on the null model selected.

In the first case, the null model is a permutation model which assumes that the commercial structure of the city is fixed at a global level (the empirical distribution of commercial premises remains untouched; only their commercial categories are randomly permuted). In Network Theory jargon, it means that the spatial network found empirically is maintained and so are the number of nodes belonging to each category; only the commercial categories of all nodes are randomized. By repeating the permutation process thousands of times, we obtain for each pair of categories a probability distribution function of the number of links between them. After that, for each such null distribution, we obtain the 2.5 and 97.5 percentiles and compare them with the empirical results, keeping just those relationships whose number of edges is outside the interval defined by the 2.5 and 97.5 percentiles. To obtain a signed network of retail types while avoiding the problems associated with the use of logarithms, we propose the use of the Z-score function as a measure of the force of attraction or repulsion. Note that the Z-score function takes the empirical results obtained for each pair of retail categories, subtracts from them the corresponding mean obtained in their respective null distribution, and divides by the standard deviation. In the case of a standard deviation equal to zero, we assume a Z-score of 0.

The other alternative takes the configuration model [56] as the null model. The implicit assumption behind this second approach is that each retail store creates a local commercial structure, which is precisely what is preserved. The configuration model starts with N nodes and a sequence k containing the degrees of all nodes ($k = \left\{k_1,\ k_2, \dots,\ k_n\right\}$); therefore, we have k_i half-edges emanating from each node i. By definition, this model maintains the position of all nodes, as well as their degree. Under these two premises, each sample of the null model is obtained by random rewiring, i.e., random matching the half-edges of each node in the network. This method is widespread in Network Science as it is the null model used to compute the standard modularity in the community detection problem [57, 58], and it is also applied in other contexts such as economic networks [59].

In our case study, we keep the degree of every retail store from the empirical distribution. In addition, we preserve the retailing category attribute of each node, as in this second alternative it is the local interaction structure inherent to each category that we are interested in. Then, we apply random rewiring to get a null model sample as in the standard procedure, with just a slight modification aimed at not creating multiple edges or loops. Again, this process is repeated multiple times (Monte Carlo sampling) to obtain the expected random distributions of the interactions between each pair of retail categories. Afterwards, we calculate the 2.5 and 97.5 percentiles to determine if the empirical relationships found can be considered statistically significant or not. Finally, by calculating the Z-score we can identify the sign of the relationship between categories and quantify its intensity.

At this point, before moving on to our case study, we would like to emphasize that there is no such thing as the best model for all cases. Therefore, below we present a detailed comparative evaluation of the three methods (Jensen, permutation and rewiring) with the aim of helping the reader to select the most suitable model for the problem at hand.

Each model has its own particularities, placing the emphasis on different aspects. Notwithstanding, about half of the information they provide is common (please refer to the Appendix 1 for the details). Hence, the three approaches could be considered complementary to some extent. The main differences between them stem from two different variables: (i) the level at which they operate and (ii) the extent to which they preserve local structure. The experimental results found with each method are a consequence of the weight that each approach puts on the different dimensions for calculating the indexes, and of the heterogeneous spatial-commercial organization that each city may have.

Jensen's model conducts the analysis at category (node) level, aggregating the local concentration ratios obtained for the shops belonging to a given category (in both intra- and inter-coefficients). Consequently, Jensen gives the same importance to pairwise interactions occurring in dense commercial districts than to those taking place between almost isolated stores in outskirts neighborhoods. On their part, both permutation and rewiring models perform the analysis at link level, being the weight of the interaction the total sum of edges between each pair of categories found across the whole city. Thereupon, both permutation and rewiring implicitly give more importance to relationships occurring in big shopping districts (where the number of interactions is likely to be much higher), than to the sparse interactions found in small commercial settings, whose contribution is somewhat diluted when aggregated into the total sum.

Regarding the preservation of the local structure, the null model imposing less implicit assumptions is that of permutation, as it only preserves the number of retail stores falling under each category, i.e., the global commercial structure. Jensen's model, on its part, is halfway between permutation and rewiring as far as the preservation of the local structure is concerned. For the intra coefficient, the null model is permutation-like. However, for the inter-coefficient, the null values are calculated by keeping fixed the location of category A establishments (which implies that their degree remains the same), and by randomly and independently permuting the categories of all the other retail stores (so their degree may change). Therefore, in this case, we only preserve the local structure of category A stores, and the global proportions of the rest of categories. Lastly, in the rewiring model, we maintain the location, category and degree of each store, only randomizing the connections between half-edges. As a consequence, the rewiring model preserves both the global structure of the network (number of stores from each category and their location) as well as the local structure of each node (degree and category).

Table 1 Cities included in the study

City	Population	Number of business	Number of commercial categories
Madrid	3,266,126	29,836	66
Barcelona	1,636,762	33,757	63
Valladolid	298,412	3314	60
Burgos	175,821	2382	64
Salamanca	144,228	2772	65
León	124,303	2156	60
Palencia	78,412	1414	61
Zamora	61,406	1124	56
Ávila	57,744	1012	58
Segovia	51,674	1221	63
Soria	39,398	705	57

Case study

The case study proposed in this paper analyses the nine provincial capitals of Castile and Leon—a northwest region in Spain—together with the two most populated cities of the country: Madrid and Barcelona. Such set of 11 cities constitutes a representative sample of Spanish cities of different sizes (Table 1). The retrieval of information was conducted during 2017, being the Yellow Pages the source of information from which the category and the address of each retailer were obtained. Subsequently, addresses were georeferenced using the MapQuest Application, Open Street Map data and Google Maps API. The radius considered in all the analyses is 100 m similarly to previous studies [32, 33].

The retailers are mapped into the 68 codes of The North American Industry Classification System for small businesses (NAICS) (Table 2) to make the analyses comparable with previous research.

Table 2 Categories of the retail business included in the study

Code	Description	Code	Description
7211	Hotels and other travel accommodations	44613	Optical goods stores
7212	RV parks	4,619	Other health and personal care stores
7213	Rooming and boarding houses	44711	Gasoline stations with convenience stores
7221	Full service restaurants	44719	Gasoline stations w/o conv. Stores
7222	Limited service restaurants	44811	Men's clothing stores
7223	Special food services and catering	44812	Women's clothing stores
7224	Drinking places	44813	Children's and infant's clothing stores
44111	New car dealers	44814	Family clothing stores
44112	Used car dealers	44815	Clothing accessory stores
44121	Recreational vehicle dealers	44819	Other apparel stores
44122	Motorcycle and boat dealers	44821	Shoe stores
44131	Auto parts and accessories	44831	Jewelry stores
44132	Tire dealers	44832	Luggage stores
44211	Furniture stores	45111	Sporting goods stores
44221	Floor covering stores	45112	Hobby, toy, and game stores
44229	Other home furnishing stores	45113	Sewing and needlecraft stores
44311	Appliances and electronics stores	45114	Musical instrument stores
44312	Computer stores	45121	Book stores
44313	Camera and photography stores	45122	Record, tape, and CD stores
44411	Home centers	45211	Department stores
44412	Paint and wallpaper stores	45291	Warehouse superstores
44413	Hardware stores	45299	Other general merchandise stores
44419	Other building materials stores	45311	Florists
44421	Outdoor power equipment stores	45321	Office and stationary stores
44422	Nursery and garden stores	45322	Gift and souvenir stores
44511	Grocery stores	45331	Used merchandise stores
44512	Convenience stores	45391	Pet and pet supply stores
44521	Meat markets	45392	Art dealers
44522	Fish and seafood markets	45393	Mobile home dealers
44523	Fruit and vegetable markets	45399	Other miscellaneous retail stores
44529	Other specialty food markets	45411	Mail order and catalog stores
44531	Liquor stores	45421	Vending machines
44611	Pharmacy and drug stores	45431	Fuel dealers
44612	Cosmetics and beauty stores	45439	Other direct selling establishments

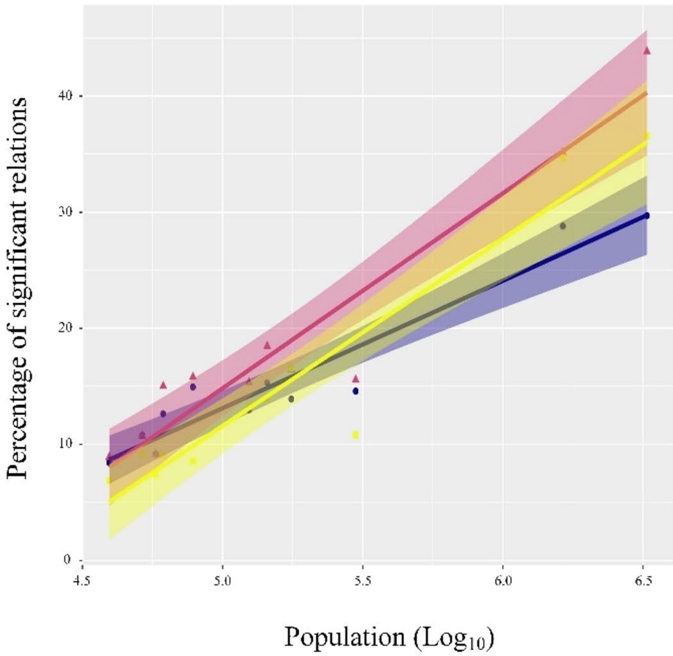
Results and discussion

Assessment of the individual commercial structure of each city and pairwise comparisons

For each of the 11 cities, we have obtained the empirical network of interactions between retail business categories, and we have assessed the significance of the relationships found using the three methods proposed. More precisely, we have analyzed 1000 permutation samples in Jensen's method, 10,000 samples in the *Z*-score permutation method, and 1000 samples in the *Z*-score rewiring method.[1] As a result, for each city, we have obtained three different adjacency matrices (one matrix per method) that summarize the networks of interactions between retail categories at a 0.05 significance level.

Given the variability in the size of the cities in our dataset, we have checked the existence of a possible relationship between population size and the percentage of significative relations found in the city. Our results suggest a higher spatial-commercial organization with city size under all three methods. In all cases, Spearman's rank correlation coefficient suggests a significative monotonic relation between the two variables (see Fig. 1). A linear-log regression model approximately fits the relation. Nonetheless, the number of cities analyzed is relatively small to draw definite conclusions in this regard.

[1] Note that a greater number of samples was obtained for the permutation method since it is less computationally intensive.



	Adjusted R^2	Spearman's rank correlation
Jensen	0.914 ($2.7\ 10^{-6}$)	0.891 ($4.3\ 10^{-4}$)
Z-score permutation	0.908 ($3.6\ 10^{-6}$)	0.909 ($5.6\ 10^{-5}$)
Z-score rewiring	0.904 ($4.3\ 10^{-6}$)	0.882 ($6.7\ 10^{-4}$)

Fig. 1 Percentage of significant empirical relations between retail categories for each city compared with their logarithm of the population size using the three methods proposed. The different shaded areas correspond to the 95% CI of the linear regression fitting included in the figure. On the right, Adjusted R^2 and Spearman's rank correlation of the data and their correspondent p values (in brackets) are provided

Previous results in the literature pointing at some kind of relationships between population density and commercial organization can be found for instance in [33], where the author asserts that in two out of the five communities identified in the network of retail interactions in Lyon, the majority of stores locate according to population density.

Having obtained the networks of significant relationships between retail categories for each city and method, we have analyzed whether such networks present community structure. Usually, communities in networks are defined as nodes that are densely connected to each other and poorly connected to the rest of the graph [60–62]. In our case, however, given that our retail networks are weighted and signed, in the community analysis, we look for nodes that are strongly attracted to each other and far from nodes that repel each other.

Despite the importance of community detection, thus far there are not many algorithms in the literature that deal with the problem of community partitioning in signed weighted networks [60]. In our study, we have used the algorithm proposed by Gómez et al. [63], which was specifically conceived to handle weighted networks with sign, and which has already been successfully applied to detect communities in the retail network from the city of Lyon [63].

Their proposal consists in unfolding the traditional definition of modularity for weighted networks given by Newman and Girvan [57] into two different terms—one accounting for positive weights and the other for the negative ones—which, in turn, are duly weighted by their respective total positive and negative strengths. Their final expression for modularity is the following:

$$Q = \frac{1}{2w^+ + 2w^-} \sum_i \sum_j \left[w_{ij} - \left(\frac{w_i^+ w_j^+}{2w^+} - \frac{w_i^- w_j^-}{2w^-} \right) \right] \times \delta(C_i,\ C_j), \tag{3}$$

where C_i is the community to which node i is assigned and δ is the Kronecker delta function. The relationships between the different terms involved in (3) can be seen in Eqs. (4) and (5) below, where w_i^+ and w_i^- are the positive and negative strengths, and $2w^+$ and $2w^-$ the total positive and negative strengths, respectively,

$$2w^+ = \sum_i w_i^+ = \sum_i \sum_j w_{ij}^+, \tag{4}$$

$$2w^- = \sum_i w_i^- = \sum_i \sum_j w_{ij}^-, \tag{5}$$

where

$$w_{ij}^+ = \max\{0,\ w_{ij}\}, \tag{6}$$

$$w_{ij}^- = \max\{0,\ -w_{ij}\}. \tag{7}$$

After applying the community detection algorithm by Gómez et al. [63] to the 33 networks that we have (11 cities and 3 methods for each city), to compare the communities obtained we have used one of the best known partition comparison metrics hitherto: the Variation of Information (VI).

The problem of comparing two different partitions (clustering partitions, communities in networks, etc.) is well known within the Physics, Statistics and Machine Learning communities. Several metrics have been proposed for such endeavor, many of which find their roots in Information Theory. However, in the light of the comprehensive reviews published on the topic [64], none of the available metrics comes without shortcomings. The most accurate and widely used to date are the Variation of Information (VI) [65] and its normalized version: the Normalized Variation of Information (NVI).

The VI measures the amount of information that we lose and gain respectively when going from one partition to the other. The most common version of the VI formula is the one described in Eq. (8), where R and S stand for the two different partitions under consideration:

$$\mathrm{VI}(R,\ S) = H(R) + H(S) - 2I(R;\ S). \tag{8}$$

More precisely, each partition can be seen as a random variable, thus, we have two random variables which can take R and S values, respectively. In Eq. (8), $H(R)$ and $H(S)$ are the entropies associated with partitions R and S, which can be interpreted as the uncertainty of random variables R and S. Formally, the entropy of R is calculated as follows:

$$H(R) = -\sum_{r=1}^{R} P(r) \log P(r), \tag{9}$$

where $P(r)$ stands for the probability of r class within R partition of the dataset under consideration.

As for $I(R;\ S)$, it is the mutual information between partitions R and S, i.e., the information that one partition has about the other. Intuitively, $I(R;\ S)$ represents the reduction in uncertainty of partition S provided that we already know partition R.

It is worth highlighting that the VI is a distance measure, hence exhibiting larger values the more dissimilar the labelings. Quite outstanding among its properties are its symmetric and nonnegative nature, as well as the fact that it satisfies the triangle inequality. Its main disadvantage, however, is that it lacks a straightforward interpretation in terms of information content.

For the sake of interpretability, the VI can be normalized to obtain a distance that varies between 0 and 1, where 0 indicates perfect coincidence and 1 total mismatch. VI is normalized by $\log n$, its upper bound [65] (where n is the number of nodes).

Given that, as previously stated, the VI and the NVI are distance measures, and that in our case study we are interested in how similar the different commercial structures found in the 11 cities are, to render interpretation easier we have calculated the complementary of the NVI (NVI similarity—SimNVI):

$$\mathrm{SimNVI}\ (R,\ S) = 1 - \frac{\mathrm{VI}(R,\ S)}{\log n}. \tag{10}$$

Another relevant remark is that the VI in particular, and partition comparison metrics in general, are conceived to compare two different partitions of the same dataset. In our case study, which encompasses 11 cities of different sizes, the number of retail categories present in each city does not necessarily coincide (in general, bigger cities comprise a greater number of categories). Therefore, to make the community partitions of the different cities comparable, we performed the partition comparison analysis on the intersection set of the categories of each pair of cities.[2]

In terms of methodologies, our results show some consensus across all methods, being the agreement especially strong between the two Z-score-based methodologies (see Appendix 1 analysis). In regard to the cities, the biggest cities considered in the analysis, i.e., Madrid and Barcelona, present the most similar relation of partitions in many cases, regardless of the methodology used (see Fig. 2).

The differences obtained between Jensen's and Z-score methods imply certain particularities and nuances specific to each technique. These results imply that their use as pre-processing mechanisms to integrate features in learning algorithms can be combined, potentially together with feature selection tools, or, given the technical difficulties of Jensen's method in small cities, as an alternative. On the other hand, differences between cities suggest that knowledge obtained from relationships may not be directly transferable, at least in all cases and in a complete way. Precisely to identify the robust relationships, we analyze the results in the following section.

Assessment of the existence of robust commercial relations across cities

In this section, we study whether robust spatial-commercial relationships exist across all cities. Such relationships would not depend on the individual characteristics of each city, thus being somewhat invariant. To that end, we have used consensus network approaches [66]. Specifically, in our work,

[2] Alternatively, we could have considered the joint set of categories to compare the partitions of the different cities, assuming those sectors that are not present in a given city to be isolated nodes. This procedure may also provide useful information; however, for the present research, in which we try to identify robust patterns present in all cities, the inclusion of minority sectors with very few establishments—which are typically those not shared– could distort the interpretation of the results.

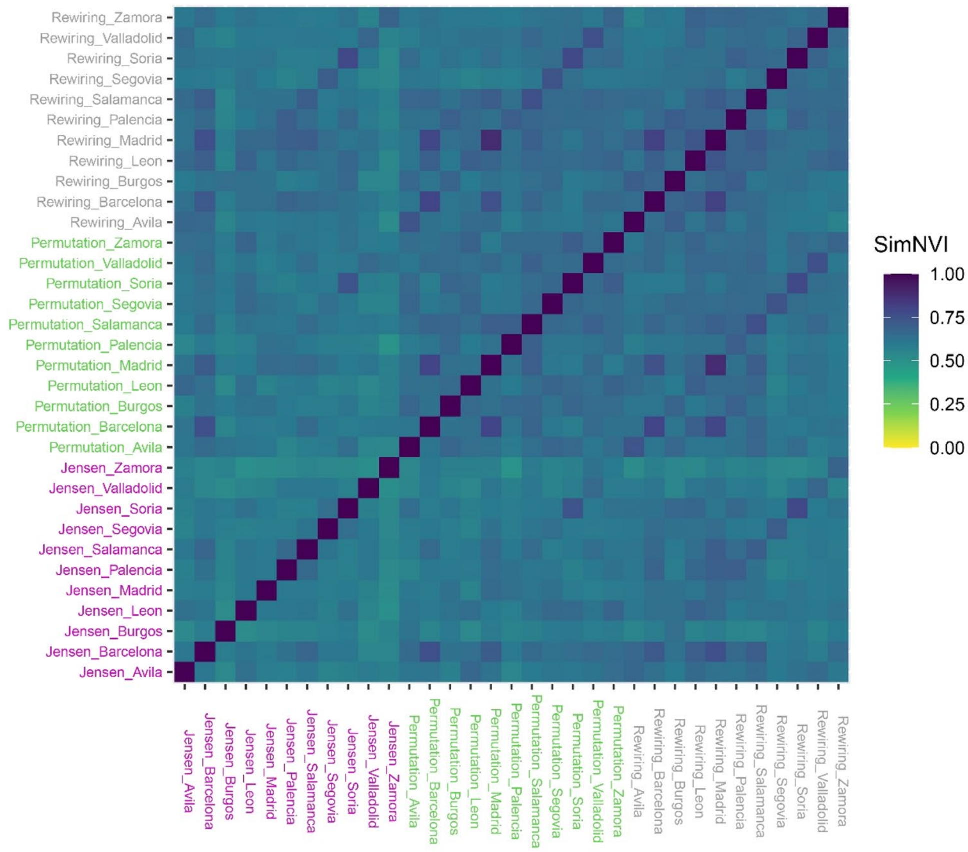


Fig. 2 Matrix of pairwise comparisons of the community partitions obtained with the algorithm from Gómez et al. [63] for the different cities and methods (Jensen, permutation and rewiring). The partition comparison metric selected is the complementary of the normalized variation of information, i.e., 1 – VI/log n (SimNVI). Hence, it has to be interpreted as a similarity measure instead of a distance. Note that all the pairwise comparisons present a SimNVI value above 0.5

we have used two complementary analyses: consensus networks of relationships and consensus networks of partition.

Consensus networks of relationships

For each of the three methods proposed (Jensen, permutation and rewiring), a consensus network of the relationships found across the 11 cities has been constructed (see Fig. 3). In each consensus network of relationships, which is weighted and signed, the nodes are the different retail categories, and the links have a weight attribute that quantifies the strength of the interaction in the following manner: for each method, we pick each pair of retail categories and check city by city if they are linked in the network (i.e., if there is a significant relationship). If the link exists, we add 1 (if it has positive weight) or -1 (if it has negative weight) to the weight of the link between them in the consensus network. Thus, given that we study 11 cities, the absolute value of the maximum weight would be 11. However, it is important to note that so as to avoid biases, instead of building the consensus networks on the intersection set of retail categories present across the 11 cities, we have considered all categories. Hence, the maximum weight possible between each pair of categories is dependent on the number of cities in which the two categories cooccur (so it could be lower than 11).

On the three consensus networks of relationships thus obtained, different thresholds can be established (for instance, the minimum value of a node's degree required to be considered in subsequent analyses). After that, to assess their community structure, partitions can be again calculated according to Gómez et al.'s [63] algorithm.

In Tables 3 and 4, we show the most relevant relationships (both attractive and repulsive) that have been found by means of the consensus networks of relationships.

The complete attraction and repulsion relationships found when we do not impose any threshold are shown in the following adjacency matrices, which also allow us to straightforwardly compare the results obtained with the three methods (see Appendix 2 for a detailed list of the communities detected and the categories belonging to each of them).

Jensen's consensus network of relationships presents an adjacency matrix with few significant repulsive relationships (Fig. 3, top). When analyzed with the community detection algorithm by Gómez et al. [63], four communities are identified. A first community (in black) includes a nucleus of relationships based on food shopping (grocery stores, meat, fish and seafood markets, fruit and vegetable shops), which acts as the most central element in the community; additional

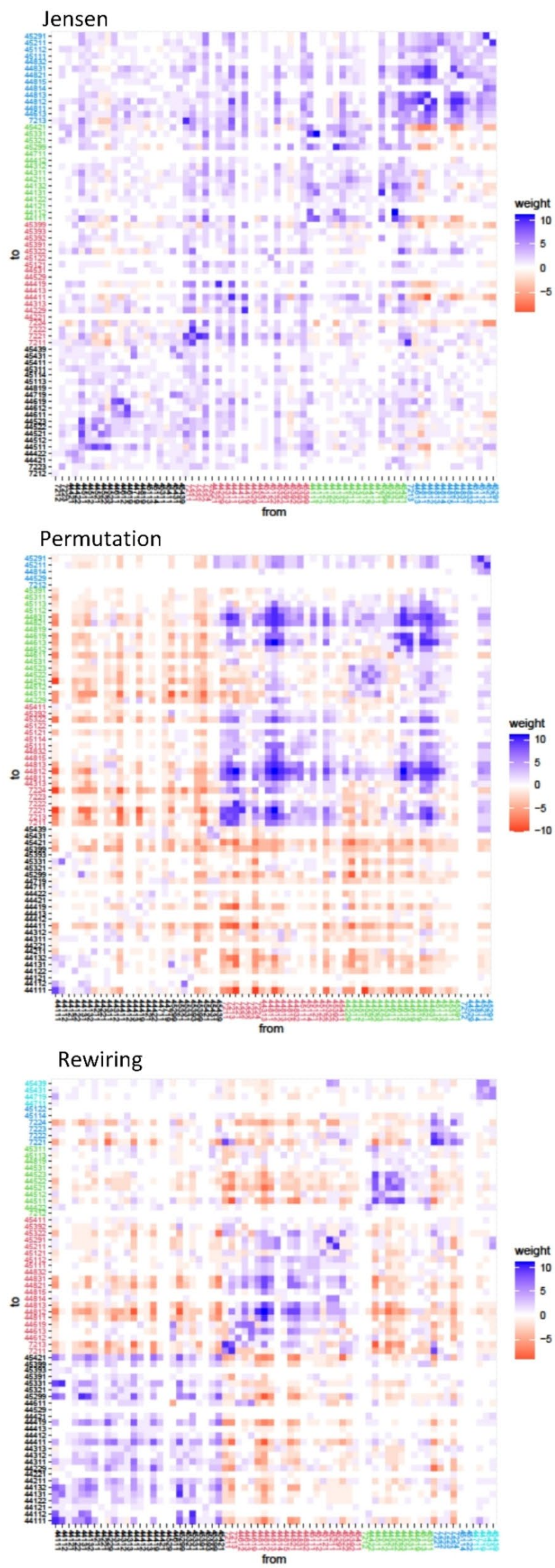


Fig. 3 Adjacency matrices of the consensus networks of significant relationships found across the 11 cities for Jensen, permutation Z-score and rewiring Z-score methods. The gradient scale between blue and red indicates the level of attraction or repulsion found between each pair of categories. The color of the categories (in the axes) indicates the community to which they belong according to the Gómez et al. [63] algorithm

categories with weaker relationships between them are also part of this community. The second community (in red) has two relevant commercial nuclei: one based on accommodation, rooming, drinking places and restaurants; and another core more focused on home centers, paint, wallpaper stores, hardware stores, and other building supplies. Although the relationships between both cores are weak in terms of attraction, both present a relatively intense repulsive behavior with the first cluster. In the third community (in green), there is a set of activities related to the automobile distribution (results show intense attraction among new car dealers, used car dealers, recreational vehicle dealers, auto parts and accessories). This core is linked to different activities such as used merchandise stores and additional categories, all of them sharing a strong repulsion with the last community (in blue). Lastly, the fourth community (in blue) contains a strongly related core of activities associated with clothing (men, women, children); this is the community with the highest repulsion towards all other retailing activities.

The patterns identified and discussed above for Jensen's consensus network of relationships are to some extent common across all configurations obtained with this approach. However, with the other two methods (permutation and rewiring) those patterns are intensified and clearer. Both with permutation and rewiring the structures of attraction and repulsion are more stable, and the division into communities is perceived more clearly.

Remarkably, the partitions of the significance consensus networks obtained with high thresholds (i.e., when we impose a minimum weight between links to be considered in the analyses) tend to be more similar among them regardless of the method used (see a comparison in Appendix 3). This result shows that, although there are differences between each approach stemming from the different hypotheses and null models considered, the most robust patterns of all the methods seem to share a similar core. In the case of permutation and rewiring, one of the differences is precisely identified in the cluster (turquoise blue) associated with gasoline stations (gasoline stations with convenience stores, gasoline stations w/o convenience stores, fuel dealers and other direct selling establishments) which, although in Jensen's network belong to the same community as well, present interrelations that are not as strong and clear as in the other two methods.

The green community in the permutation consensus network includes two nuclei of high positive internal interconnection: one associated with accessories (jewelry, shoes, optics, and cosmetics) and another—partially differentiated from the previous one—which is associated with the food sector. On its part, in the rewiring method, the partitioning

Table 3 Most frequent significant attractive relations between retail categories found in the analyzed cities

From	To	Jensen	Permutation	Rewiring
Women's clothing stores	Men's clothing stores	10	10	10
Men's clothing stores	Women's clothing stores	10	10	10
Warehouse superstores	Department stores	10	10	10
Department stores	Warehouse superstores	10	10	10
Rooming and boarding houses	Hotels and other travel accommodations	10	9	10
Hotels and other travel accommodations	Rooming and boarding houses	10	9	10
Limited service restaurants	Full service restaurants	10	10	9
Full service restaurants	Limited service restaurants	10	10	9
Children's and infant's clothing stores	Women's clothing stores	10	10	9
Shoe stores	Women's clothing stores	10	10	9

Table 4 Most frequent significant repulsive relations between retail categories found in the analyzed cities

From	To	Jensen	Permutation	Rewiring
Women's clothing stores	Home centers	−3	−8	−9
Women's clothing stores	Vending machines	−5	−3	−7
Women's clothing stores	New car dealers	−9	−9	−7
Shoe stores	New car dealers	−7	−7	−6
Rooming and boarding houses	Home centers	−8	−9	−8
Women's clothing stores	Other general merchandise stores	−3	−5	−9
Men's clothing stores	Vending machines	−4	−8	−6
Shoe stores	Vending machines	−5	−10	−6
New car dealers	Full service restaurants	−10	−10	−5
Jewelry stores	Vending machines	−4	−6	−7

algorithm divides them, making the food-related cluster (green) independent, and joining the accessory-related core to clothing, restaurants, and drinking stores (red). The latter relationship between accessories, clothing, restaurants and bars is also perceived in the permutation partition, where a high positive interaction between the accessories and food community (green) and the red one is clearly visible; note that the red community agglutinates activities more related to clothing, restaurants and drinking stores—as in rewiring—but also includes accommodation and rooming—which in the case of rewiring are in a separate community (blue). The black communities in both *Z*-score methods exhibit a strong repulsion towards the other clusters. They concentrate commercial activities associated with the automobile, hardware, and home stores.

Consensus networks of partition

In the construction of consensus networks of partition, for each of the three methods, we set again the retail categories as nodes, and in this case, we calculate the weight of the links as the number of cities in which the two retail categories belong to the same community in the partition obtained with Gómez's method (see Fig. 4). In contrast with the previous analyses, in this approach the network is undirected, and all the weights of the links are positive. Again, different thresholds can be established on these networks in order to choose the scale at which to conduct further analyses.

First, we analyze the structure of the three entire partition consensus networks establishing an initial threshold of 1, i.e., all pairwise relationships which appeared at least in one city out of 11 have been considered in the analysis. The community structure has been assessed with the OSLOM algorithm [67]. This algorithm allows estimating the statistical significance of clusters (potentially hierarchical, overlapping and detecting homeless nodes) with respect to random fluctuations [67]. We have calculated the results using 200,000 runs for the first hierarchical level and 500 for the rest, with a p-value of 0.01, and with initial configurations obtained with Louvain [68], Infomap [69] and copra [70] algorithms.

The community analysis of the partition consensus networks shows a split into 4–5 communities similar to the one obtained for the relationship consensus networks (Fig. 4). In Jensen's partition consensus network, three very significant communities are identified, with some commercial categories overlapping between communities (black nodes in the network), and some homeless nodes (white nodes) which

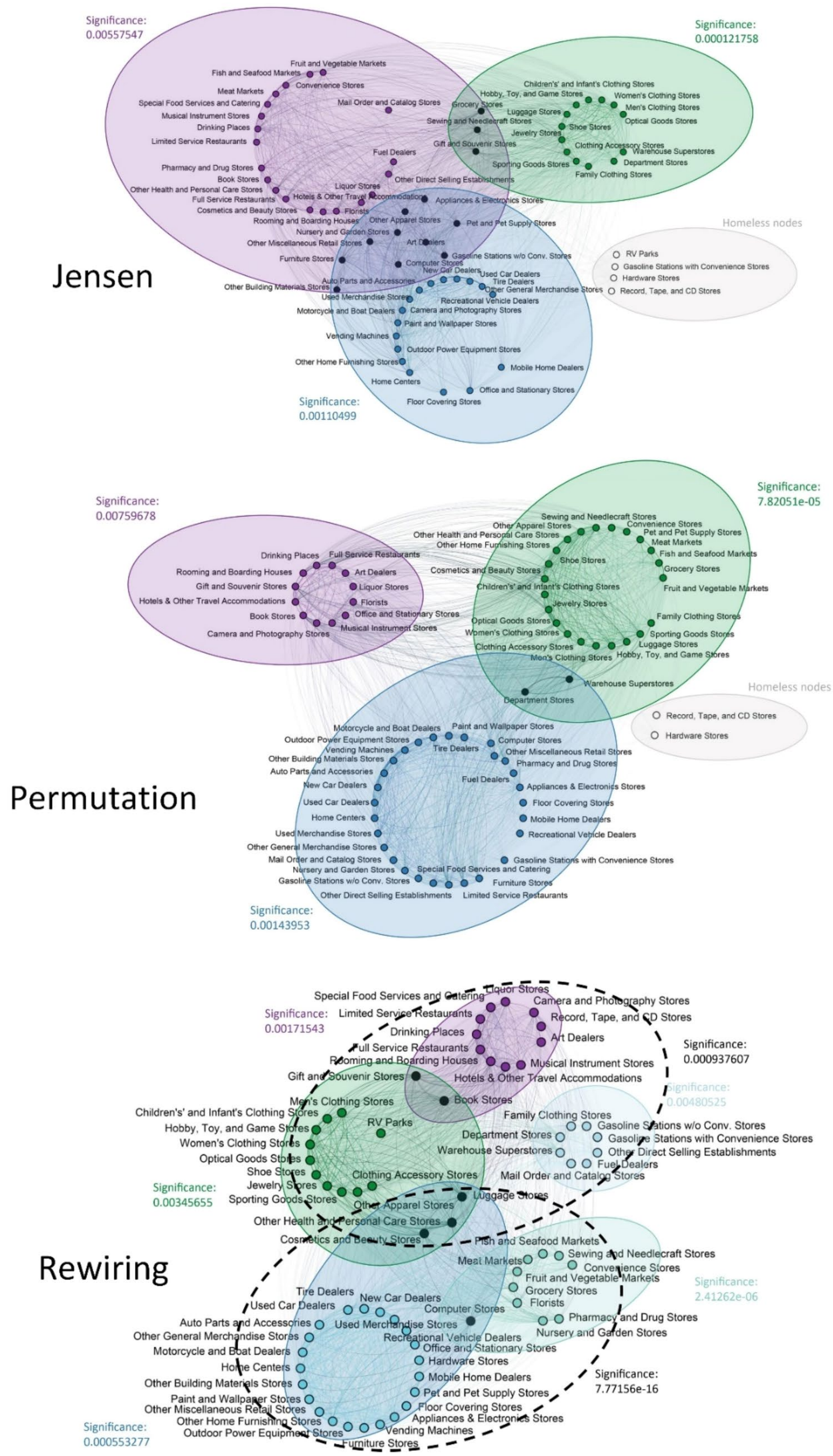


Fig. 4 Visualizations of the partition consensus networks found across the 11 cities for Jensen's, permutation Z-score and rewiring Z-score methods. The color of the nodes indicates the community to which they belong according to the OSLOM algorithm [67]. Black nodes represent overlapping categories that are shared by two or more communities. White nodes indicate homeless, isolated nodes. Dashed lines represent hierarchical structure: communities that comprise others of smaller size. The significance level of each cluster is also included in the figure

have no clear relationship with any community. The permutation partition consensus network presents a very similar structure, also with very significant modules: a community associated with accommodation, rooming, restaurants and drinking stores, another one strongly associated with clothing and complements, and a third one related to the automobile sector, hardware, home, etc. In the case of the rewiring partition consensus network, the structure of the network is more complex and insightful, shedding light on the hierarchical structure between relationships: there are two main communities that, in turn, can be decomposed into three and two smaller clusters, respectively. In the first large community, we find, as in the previous analyses, a set associated with the categories of accommodation, rooming, restaurants and drinking stores, another associated with clothing and accessories, and a third cluster associated with family clothing, department stores and urban gas stations. In the second large community, we find a cluster associated with products related to the automobile, home and hardware, and another linked to the food sector, drugs, and pharmacy.

These results of the entire networks (threshold 1) are interesting, but they give us a very general and aggregated picture of the commercial structure of the cities. Given that part of our objective is to identify the backbone of commercial interactions, at this point we have established a higher threshold: 6, thus considering only relations that occur in most of the cities. Both in consensus networks of relationships and in consensus networks of partition, the similarity of the three methods (Jensen, permutation and rewiring) converges when the threshold is increased (see Appendix 3). Application of the above-mentioned threshold to both consensus networks of relationships and partition produces results that are more specific and detailed than those obtained under the general perspective considering full interaction.

The results of the community analyses conducted on the consensus networks of relationships with threshold 6 can be seen in Fig. 5. Jensen's method identifies eight communities, the permutation eleven and rewiring seven. It is interesting to note that some of the communities in the *Z*-score-based methods are divided as a consequence of different levels of repulsion towards other activities, despite being in general intensively connected between them. It is also worthy of mention that Jensen's method is asymmetrical and that significant positive relationships in one direction can be negative in the opposite direction (although these cases are rare).

As for the consensus networks of partition with threshold 6, the communities obtained with the OSLOM algorithm are much smaller and fragmented, with very high levels of significance and revealing the empirically more robust business to business interactions (see Fig. 6). Although there are still some particularities associated with each method: more fragmentation in Jensen's method, more overlap between communities in the permutation method, and hierarchical structure in the rewiring method, the relationships found are quite similar and insightful.

The common backbone identified by both consensus methods allows the analysis of transferable and stable relationships between cities, and to distinguish them from particular relationships found spuriously, or as a consequence cultural, geographic, size or demographic aspects that have to be analyzed in detail in each particular case.

Conclusions

In retailing, the selection of the most suitable location is considered the most relevant strategic decision. Although this is a multidimensional problem, one key factor for success is the pursuit of balanced tenancy. In this paper, by means of a complex networks approach, we have analyzed from different perspectives the relationships between retail categories across 11 Spanish cities of different sizes. To disentangle the intricacies of the relationships between retail categories, the empirical relationships found have been checked against three different null models, each of them capturing different aspects of interest. Subsequent community analyses conducted on each of the cities for each of the models shed light on their commercial structure, enabling the pairwise comparison of the results obtained for the different cities. Ultimately, consensus approaches have been implemented to check for the existence of robust commercial relationships in retailing.

The main conclusions to be drawn from our analyses may be summarized as follows:

- The empirical interactions between retail categories can be usefully modeled as a network. Different null models can be used to check the significance of the relationships found, and there is no model better than the rest in all cases. Even though more than half of the information they provide is common, each model has its own particularities, placing the emphasis on different aspects. Remarkably, the main differences stem from: (i) the level at which they operate and (ii) the extent to which they preserve the local structure.
- For the particular problem we have considered, the rewiring method might be the most suitable approach, as the preservation of the local structure of each store seems a sensible and plausible assumption, which in turn seems to yield more interpretable results.
- From a practical perspective, these results allow the use of *Z*-score methods as pre-processing techniques, both complementary and/or substitutive to Jensen's, to capture the relational dimension of the potential sites in location recommendation systems.

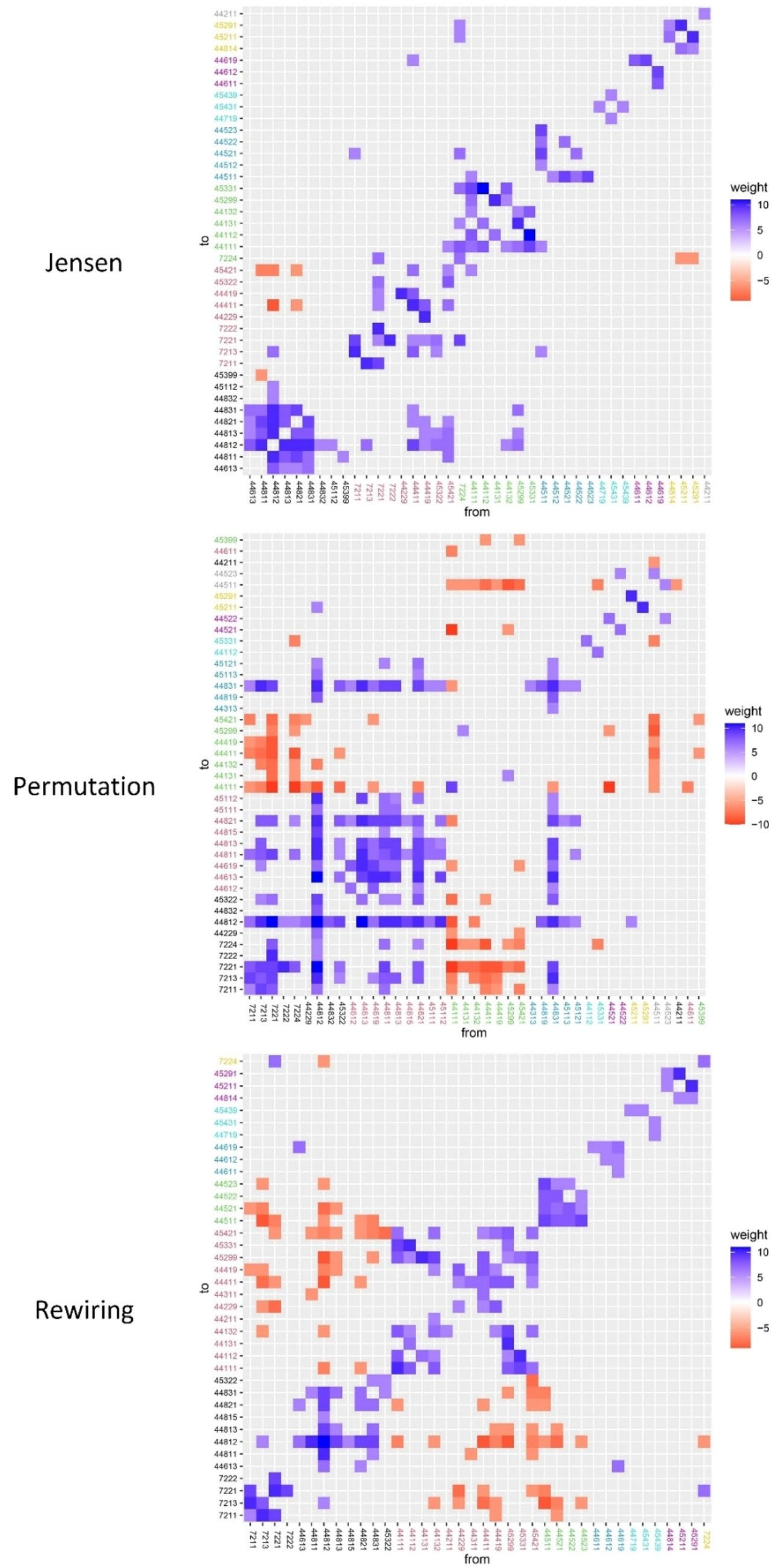


Fig. 5 Adjacency matrices of the consensus networks of relationships found across the 11 cities for Jensen's, permutation Z-score and rewiring Z-score methods with threshold 6 (edges with lower weights in absolute value have been pruned before the analysis). The gradient scale between blue and red indicates the level of attraction or repulsion found between each pair of categories. The color of the categories indicates the community to which they belong according to Gómez et al.'s [63] algorithm

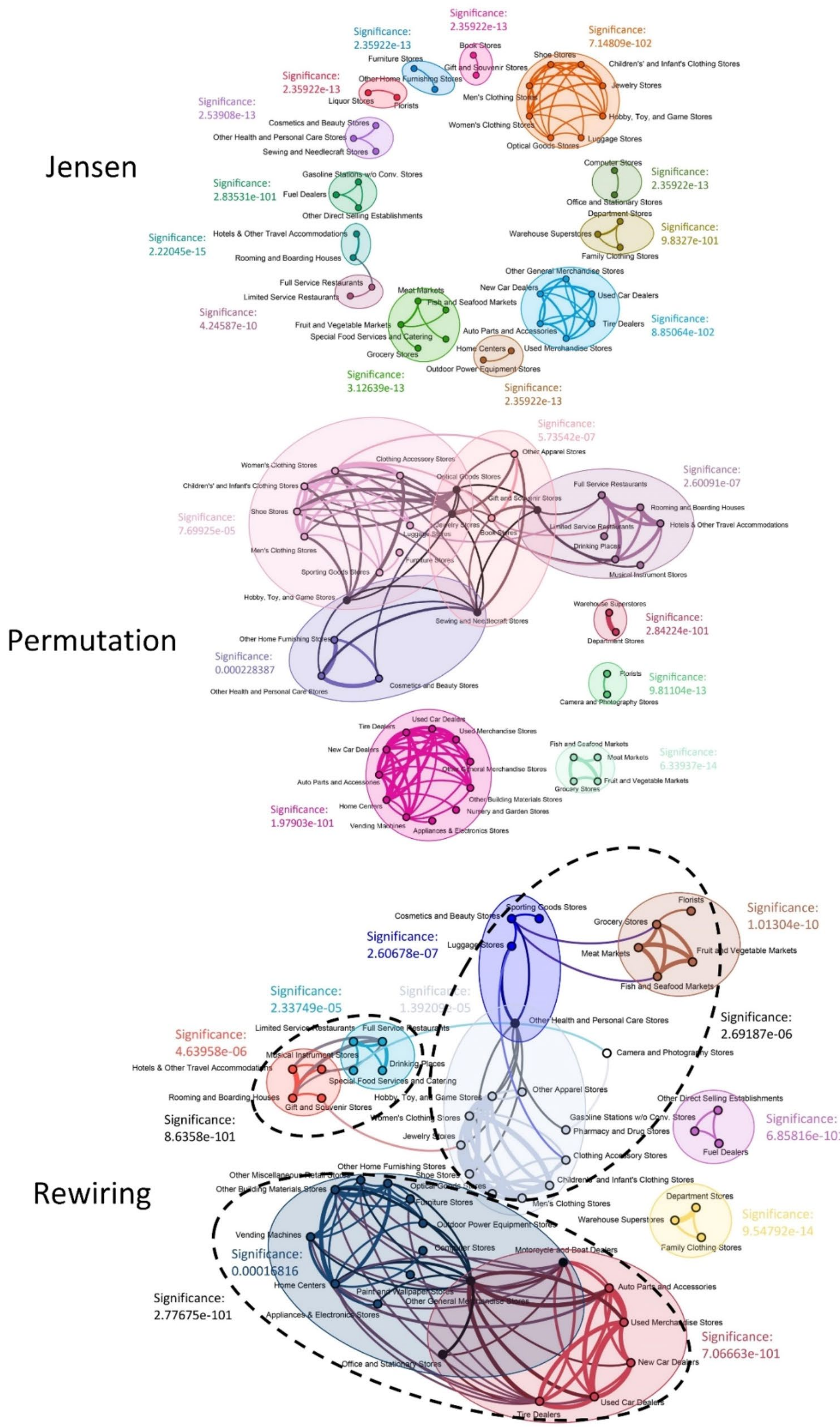


Fig. 6 Visualizations of the partition consensus networks found across the 11 cities for Jensen's, permutation Z-score and rewiring Z-score methods with threshold 6 (edges with lower weights in absolute value have been pruned before the analysis). The color of the nodes indicates the community to which they belong according to the OSLOM algorithm [67]. Black nodes represent overlapping categories that are shared by two or more communities. White nodes indicate homeless, isolated nodes. Dashed lines represent hierarchical structure: communities that comprise other communities of smaller size. The significance level of each cluster is also included in the figure

- Concerning the existence of a robust core of interactions between retail categories common across cities, insightful results have been obtained with the two consensus methods implemented. On the one hand, we have found structure in all the six consensus networks obtained—three for consensus by relationships (Jensen, permutation and rewiring), and three for consensus by partition—regardless of the threshold selected; understandably, the layout and intensity of such relationships present some variations depending on the method used. In addition, for a given threshold, the number of communities found in each consensus network varies as well. However, despite their particularities, both types of consensus networks (with their respective subdivisions: Jensen, permutation and rewiring) tend to converge towards a core as we increase the threshold. On the other hand, it is interesting to note that no consensus methodology is better than the other, since they yield complementary insights. Relationship consensus networks are weighted and signed (they aggregate over all cities while maintaining the sign of the relationship), thus providing information about the attractive or repulsive nature of the consensual interactions. On their part, consensus by partition networks (undirected and unsigned) enable to assess the significance of the different communities found by means of the OSLOM algorithm. Remarkably, in our consensus networks of partition, the higher the threshold imposed, the greater the significance of the communities found.
  These results allow the identification of transferable relationships between cities, and of those that may be specific to particular towns, a result with special interest for its application in recommendation systems with cold-start problems.
- As for the analyses conducted between the size of the different cities considered and their percentage of significant relationships, we conclude using Spearman's rank correlation that under the three methods (Jensen, permutation and rewiring) a positive trend was found. A tentative interpretation of such trend could be that bigger cities present greater commercial specialization, which translates into positively related retail stores being closer in space across all commercial areas and vice versa. Nevertheless, given the limited size of our sample of cities (11), and considering that most of them are of similar size, to obtain conclusive results it would be necessary to include more cities of varied sizes in the analyses.

A final important remark is that retailing patterns are complex phenomena, and hence, the relationships captured by our methods might not be entirely explained by the mechanism of balanced tenancy. Other factors such as population density, space availability, venue rental prices, cultural preferences, etc., (apart from the fact that the dataset may not include the entire population, as not all the local stores are included in the Yellow Pages), could also play an important role and constitute a limitation of this work. In addition, our study focuses on the analysis of spatial patterns found in cities, but it does not distinguish between whether the proximity is a consequence of several categories seeking the same types of places (joint-location), or whether the regularities are due to commercial synergies or other links or relationships that may exist between them (co-location). Analysis crossing georeferenced retailing information with additional layers and longitudinal data might give further insights into this problem. Notwithstanding, we consider that our proposal is a natural choice to identify and quantify robust and stable associations between retail categories.

Appendix 1

See Fig. 7.

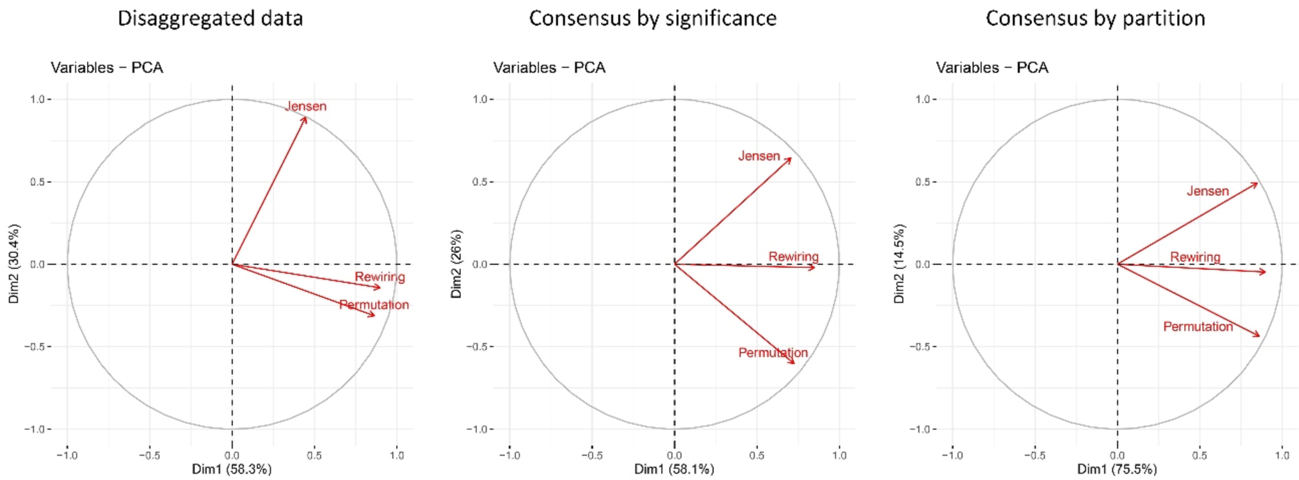


Fig. 7 Results of the principal component analysis to compare the different methods (Jensen, permutation and rewiring) as well as the consensus approaches. In the first PCA the data used are all the pairwise interaction weights of each pair of categories and cities. Results show more similarity between the rewiring and permutation methods. The second and third figures show the differences between methods when consensus techniques are applied. The biggest differences are found between permutation and Jensen, but the importance of the principal dimension reveals a relevant degree of agreement (although some differences as well) among all the three methods when consensus is used

Appendix 2

See Tables 5, 6 and 7.

Table 5 Community partition of the consensus networks of significant relationships found across the 11 cities for Jensen method according to the Gómez et al. [63] algorithm

Code	Description	Community
7212	RV parks	Black
7223	Special food services and catering	Black
44421	Outdoor power equipment stores	Black
44422	Nursery and garden stores	Black
44511	Grocery stores	Black
44512	Convenience stores	Black
44521	Meat markets	Black
44522	Fish and seafood markets	Black
44523	Fruit and vegetable markets	Black
44529	Other specialty food markets	Black
44611	Pharmacy and drug stores	Black
44612	Cosmetics and beauty stores	Black
44619	Other health and personal care stores	Black
44719	Gasoline stations w/o conv. stores	Black
44819	Other apparel stores	Black
45113	Sewing and needlecraft stores	Black
45114	Musical instrument stores	Black
45311	Florists	Black
45411	Mail order and catalog stores	Black
45431	Fuel dealers	Black
45439	Other direct selling establishments	Black
7211	Hotels and other travel accommodations	Red
7221	Full service restaurants	Red
7222	Limited service restaurants	Red
7224	Drinking places	Red
44221	Floor covering stores	Red
44229	Other home furnishing stores	Red
44313	Camera and photography stores	Red
44411	Home centers	Red
44413	Hardware stores	Red
44419	Other building materials stores	Red
44531	Liquor stores	Red
45121	Book stores	Red
45122	Record, tape, and CD stores	Red
45322	Gift and souvenir stores	Red
45391	Pet and pet supply stores	Red
45392	Art dealers	Red
45393	Mobile home dealers	Red
45399	Other miscellaneous retail stores	Red
44111	New car dealers	Green
44112	Used car dealers	Green
44121	Recreational vehicle dealers	Green
44122	Motorcycle and boat dealers	Green
44131	Auto parts and accessories	Green
44132	Tire dealers	Green
44211	Furniture stores	Green
44311	Appliances and electronics stores	Green

Table 5 (continued)

Code	Description	Community
44312	Computer stores	Green
44412	Paint and wallpaper stores	Green
44711	Gasoline stations with convenience stores	Green
45299	Other general merchandise stores	Green
45321	Office and stationary stores	Green
45331	Used merchandise stores	Green
45421	Vending machines	Green
7213	Rooming and boarding houses	Blue
44613	Optical goods stores	Blue
44811	Men's clothing stores	Blue
44812	Women's clothing stores	Blue
44813	Children's and infant's clothing stores	Blue
44814	Family clothing stores	Blue
44815	Clothing accessory stores	Blue
44821	Shoe stores	Blue
44831	Jewelry stores	Blue
44832	Luggage stores	Blue
45111	Sporting goods stores	Blue
45112	Hobby, toy, and game stores	Blue
45211	Department stores	Blue
45291	Warehouse superstores	Blue

The name of the community is the color in the axes in Fig. 3 (top)

Table 6 Community partition of the consensus networks of significant relationships found across the 11 cities for the permutation *Z*-score method according to the Gómez et al. [63] algorithm

Code	Description	Community
44111	New car dealers	Black
44112	Used car dealers	Black
44121	Recreational vehicle dealers	Black
44122	Motorcycle and boat dealers	Black
44131	Auto parts and accessories	Black
44132	Tire dealers	Black
44211	Furniture stores	Black
44221	Floor covering stores	Black
44311	Appliances and electronics stores	Black
44312	Computer stores	Black
44411	Home centers	Black
44412	Paint and wallpaper stores	Black
44413	Hardware stores	Black
44419	Other building materials stores	Black
44421	Outdoor power equipment stores	Black
44422	Nursery and garden stores	Black
44711	Gasoline stations with convenience stores	Black
44719	Gasoline stations w/o conv. stores	Black
45299	Other general merchandise stores	Black
45321	Office and stationary stores	Black

Table 5 (continued)

Code	Description	Community
45331	Used merchandise stores	Black
45393	Mobile home dealers	Black
45399	Other miscellaneous retail stores	Black
45421	Vending machines	Black
45431	Fuel dealers	Black
45439	Other direct selling establishments	Black
7211	Hotels and other travel accommodations	Red
7213	Rooming and boarding houses	Red
7221	Full service restaurants	Red
7222	Limited service restaurants	Red
7223	Special food services and catering	Red
7224	Drinking places	Red
44313	Camera and photography stores	Red
44811	Men's clothing stores	Red
44812	Women's clothing stores	Red
44813	Children's and infant's clothing stores	Red
44815	Clothing accessory stores	Red
44832	Luggage stores	Red
45111	Sporting goods stores	Red
45114	Musical instrument stores	Red
45121	Book stores	Red
45122	Record, tape, and CD stores	Red
45322	Gift and souvenir stores	Red
45392	Art dealers	Red
45411	Mail order and catalog stores	Red
44229	Other home furnishing stores	Green
44511	Grocery stores	Green
44512	Convenience stores	Green
44521	Meat markets	Green
44522	Fish and seafood markets	Green
44523	Fruit and vegetable markets	Green
44531	Liquor stores	Green
44611	Pharmacy and drug stores	Green
44612	Cosmetics and beauty stores	Green
44613	Optical goods stores	Green
44619	Other health and personal care stores	Green
44819	Other apparel stores	Green
44821	Shoe stores	Green
44831	Jewelry stores	Green
45112	Hobby, toy, and game stores	Green
45113	Sewing and needlecraft stores	Green
45311	Florists	Green
45391	Pet and pet supply stores	Green
7212	RV parks	Blue
44529	Other specialty food markets	Blue
44814	Family clothing stores	Blue
45211	Department stores	Blue
45291	Warehouse superstores	Blue

The name of the community is the color in the axes in Fig. 3 (middle)

Table 7 Community partition of the consensus networks of significant relationships found across the 11 cities for the rewiring *Z*-score method according to the Gómez et al. [63] algorithm

Code	Description	Community
44111	New car dealers	Black
44112	Used car dealers	Black
44121	Recreational vehicle dealers	Black
44122	Motorcycle and boat dealers	Black
44131	Auto parts and accessories	Black
44132	Tire dealers	Black
44211	Furniture stores	Black
44221	Floor covering stores	Black
44229	Other home furnishing stores	Black
44311	Appliances and electronics stores	Black
44312	Computer stores	Black
44313	Camera and photography stores	Black
44411	Home centers	Black
44412	Paint and wallpaper stores	Black
44413	Hardware stores	Black
44419	Other building materials stores	Black
44421	Outdoor power equipment stores	Black
44529	Other specialty food markets	Black
44611	Pharmacy and drug stores	Black
45299	Other general merchandise stores	Black
45321	Office and stationary stores	Black
45331	Used merchandise stores	Black
45391	Pet and pet supply stores	Black
45393	Mobile home dealers	Black
45399	Other miscellaneous retail stores	Black
45421	Vending machines	Black
7211	Hotels and other travel accommodations	Red
7213	Rooming and boarding houses	Red
44612	Cosmetics and beauty stores	Red
44613	Optical goods stores	Red
44619	Other health and personal care stores	Red
44811	Men's clothing stores	Red
44812	Women's clothing stores	Red
44813	Children's and infant's clothing stores	Red
44814	Family clothing stores	Red
44815	Clothing accessory stores	Red
44821	Shoe stores	Red
44831	Jewelry stores	Red
44832	Luggage stores	Red
45111	Sporting goods stores	Red
45112	Hobby, toy, and game stores	Red
45121	Book stores	Red
45211	Department stores	Red
45291	Warehouse superstores	Red
45322	Gift and souvenir stores	Red
45392	Art dealers	Red
45411	Mail order and catalog stores	Red
7212	RV parks	Green
44422	Nursery and garden stores	Green
44511	Grocery stores	Green
44512	Convenience stores	Green
44521	Meat markets	Green
44522	Fish and seafood markets	Green
44523	Fruit and vegetable markets	Green
44531	Liquor stores	Green
44819	Other apparel stores	Green
45113	Sewing and needlecraft stores	Green
45311	Florists	Green
7221	Full service restaurants	Blue
7222	Limited service restaurants	Blue
7223	Special food services and catering	Blue
7224	Drinking places	Blue
45114	Musical instrument stores	Blue
45122	Record, tape, and CD stores	Blue
44711	Gasoline stations with convenience stores	Turquoise blue
44719	Basoline stations w/o conv. stores	Turquoise blue
45431	Fuel dealers	Turquoise blue
45439	Other direct selling establishments	Turquoise blue

The name of the community is the color in the axes in Fig. 3 (bottom)

Appendix 3

See Figs. 8 and 9

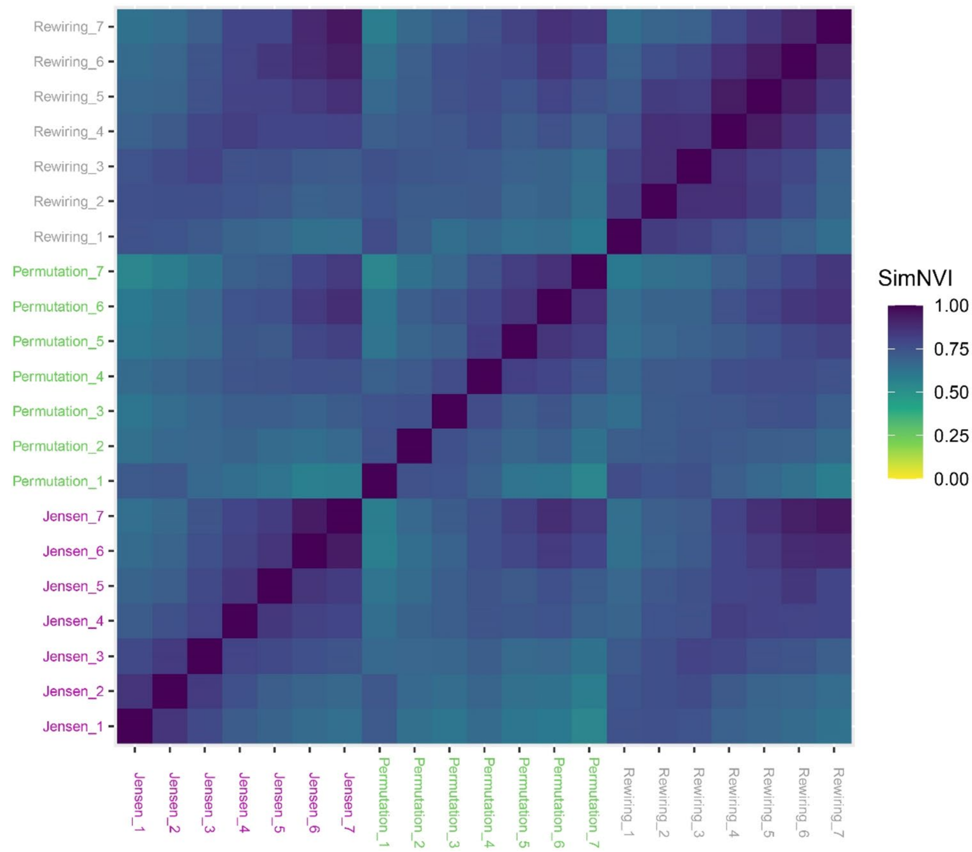


Fig. 8 Matrix of pairwise comparisons of the community partitions obtained with the algorithm from Gómez et al. [63] for the different methods (Jensen, Permutation and Rewiring) and thresholds for the consensus networks of significance. The partition comparison metric selected is the complementary of the normalized variation of information, i.e., 1 – VI/logn (SimNVI). Hence, it has to be interpreted as a similarity measure instead of a distance. Results show that as the threshold is increased there is some convergence between all methods. All the pairwise comparisons are above a threshold of 0.55 of SimNVI

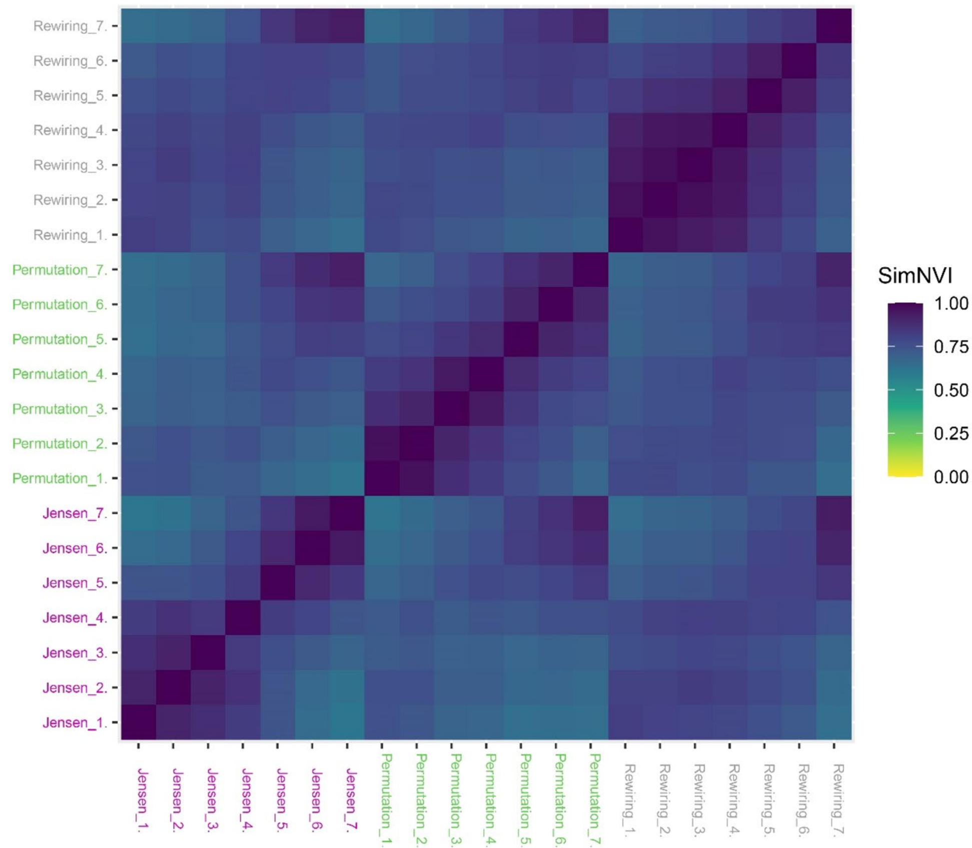


Fig. 9 Matrix of pairwise comparisons of the community partitions at different threshold levels for each method and threshold in the consensus networks of partitions. The output of OSLOM algorithm does not produce partitions (there are different scales, overlapping and nodes not assigned to any community); consequently, it is not easy to compare the influence of different thresholds. To do that, we have used alternative community detection algorithms and then calculated the similarity between partitions using the complementary of the normalized variation of information, i.e., $1 - \mathrm{VI}/\log n$ (SimNVI). The communities have been obtained using Radatools 5.0 maximizing the weighted modularity [71, 72] combining different optimization algorithms and initialization modes. Concretely, we have used spectral optimization [73], extremal optimization [74], Louvain algorithm [68], and a fine-tuning reposition from the best of them using tabu search [75] and fast algorithm [76].The process has been repeated ten times to output the best partition found. Results show that as the threshold is increased there is some convergence between all methods. All the pairwise comparisons are above a threshold of 0.6 of SimNVI

Appendix 4

The networks of this research can be found publicly in the following repository: https://riubu.ubu.es/handle/10259/5585, https://doi.org/10.36443/10259/5585.

Acknowledgements The authors acknowledge financial support from the Spanish Ministry of Science, Innovation and Universities (excellence networks HAR2017-90883-REDC and RED2018-102518-T, and project PGC2018-094754-B-C21), the Junta de Castilla y León—Consejería de Educación through BDNS 425389 and the predoctoral grant awarded to Virginia Ahedo (supported by the European Social Fund), the Generalitat de Catalunya through grant 2017SGR-896 and the Universitat Rovira i Virgili through grant 2019PFR-URV-B2-41. In addition, we acknowledge support from the Santander Supercomputación group (University of Cantabria), that provided access to the Altamira Supercomputer—located at the Institute of Physics of Cantabria (IFCA-CSIC) and member of the Spanish Supercomputing Network—to perform the different simulations/analyses. Lastly, the authors would like to especially thank Dr. Luis Izquierdo for his insightful suggestions to improve the manuscript.

Declarations

Conflict of interest The authors declare that no competing interests exist.

References

1. Weber A (1909) Uber den Standort der Industrien. Russell & Russell, Tübingen
2. Marshall A (1890) Principle of economics. Macmillan, London
3. Marcon E, Puech F (2003) Evaluating the geographic concentration of industries using distance-based methods. J Econ Geogr 3:409–428. https://doi.org/10.1093/jeg/lbg016
4. Krugman P (1991) Geography and trade. MIT Press, London
5. Beguin H (1992) Christaller's central place postulates. Ann Reg Sci 26:209–229. https://doi.org/10.1007/BF01581383
6. Moses LN (1958) Location and the theory of production. Q J Econ 72:259. https://doi.org/10.2307/1880599
7. Wallsten SJ (2001) An empirical test of geographic knowledge spillovers using geographic information systems and firm-level data. Reg Sci Urban Econ 31:571–599. https://doi.org/10.1016/S0166-0462(00)00074-0
8. Bozkaya B, Yanik S, Balcisoy S (2010) A GIS-based optimization framework for competitive multi-facility location-routing problem. Netw Spat Econ 10:297–320. https://doi.org/10.1007/s11067-009-9127-6
9. Brache J, Felzensztein C (2019) Geographical co-location on Chilean SME's export performance. J Bus Res 105:310–321. https://doi.org/10.1016/j.jbusres.2017.11.044
10. Fujita M, Krugman P (2003) The new economic geography: Past, present and the future. Pap Reg Sci 83:139–164
11. Duranton G, Puga D (2004) Micro-foundations of urban agglomeration economies. In: Handbook of regional and urban economics. Elsevier, pp 2063–2117
12. Jara-Figueroa C, Jun B, Glaeser EL, Hidalgo CA (2018) The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. Proc Natl Acad Sci 115:12646–12653. https://doi.org/10.1073/pnas.1800475115
13. Pablo-Martí F, Arauzo-Carod JM (2018) Spatial distribution of economic activities: a network approach. J Econ Interact Coord. https://doi.org/10.1007/s11403-018-0225-8
14. Berman BR, Evans JR, Chatterjee PM (2018) Retail management. A strategic approach. Pearson
15. Ciari F, Löchl M, Axhausen KW (2008) Location decisions of retailers: an agent-based approach. In: 15th International conference on recent advances in retailing and services science, pp 1–34
16. Zentes J, Morschett D, Schramm-Klein H (2012) Strategic retail management. Gabler, Wiesbaden
17. Reynolds J, Wood S (2010) Location decision making in retail firms: evolution and challenge. Int J Retail Distrib Manag 38:828–845. https://doi.org/10.1108/09590551011085939
18. Erbıyık H, Özcan S, Karaboğa K (2012) Retail store location selection problem with multiple analytical hierarchy process of decision making an application in Turkey. Procedia Soc Behav Sci 58:1405–1414. https://doi.org/10.1016/j.sbspro.2012.09.1125
19. Shaikh SA, Memon MA, Prokop M, Kim KS (2020) An AHP/TOPSIS-based approach for an optimal site selection of a commercial opening utilizing geospatial data. In: Proc 2020 IEEE Int Conf Big Data Smart Comput BigComp 2020, pp 295–302. https://doi.org/10.1109/BigComp48618.2020.00-58
20. Konishi H (2005) Concentration of competing retail stores. J Urban Econ 58:488–512. https://doi.org/10.1016/j.jue.2005.08.005
21. Xiong X, Xiong F, Zhao J et al (2020) Dynamic discovery of favorite locations in spatio-temporal social networks. Inf Process Manag 57:102337. https://doi.org/10.1016/j.ipm.2020.102337
22. Chen YM, Chen TY, Chen LC (2020) On a method for location and mobility analytics using location-based services: a case study of retail store recommendation. Online Inf Rev. https://doi.org/10.1108/OIR-10-2017-0292
23. Zheng Y, Capra L, Wolfson O, Yang H (2014) Urban computing. ACM Trans Intell Syst Technol 5:1–55. https://doi.org/10.1145/2629592
24. Ma Y, Mao J, Ba Z, Li G (2020) Location recommendation by combining geographical, categorical, and social preferences with location popularity. Inf Process Manag 57:102251. https://doi.org/10.1016/j.ipm.2020.102251
25. Bao J, Zheng Y, Wilkie D, Mokbel M (2015) Recommendations in location-based social networks: a survey. GeoInformatica 19:525–565. https://doi.org/10.1007/s10707-014-0220-8
26. Valverde-Rebaza JC, Roche M, Poncelet P, de Lopes A (2018) The role of location and social strength for friendship prediction in location-based social networks. Inf Process Manag 54:475–489. https://doi.org/10.1016/j.ipm.2018.02.004
27. Zhang K, Pelechrinis K, Lappas T (2018) Effects of promotions on location-based social media: evidence from foursquare. Int J Electron Commer 22:36–65. https://doi.org/10.1080/10864415.2018.1396118
28. Guo B, Liu Y, Ouyang Y et al (2019) Harnessing the power of the general public for crowdsourced business intelligence: a survey. IEEE Access 7:26606–26630. https://doi.org/10.1109/ACCESS.2019.2901027
29. Monaghan S, Lavelle J, Gunnigle P (2017) Mapping networks: exploring the utility of social network analysis in management research and practice. J Bus Res 76:136–144. https://doi.org/10.1016/j.jbusres.2017.03.020
30. Hidalgo CA, Castañer E, Sevtsuk A (2020) The amenity mix of urban neighborhoods. Habitat Int. https://doi.org/10.1016/j.habitatint.2020.102205
31. Goh S, Choi MY, Lee K, Kim K (2016) How complexity emerges in urban systems: theory of urban morphology. Phys Rev E 93:052309. https://doi.org/10.1103/PhysRevE.93.052309
32. Jensen P (2006) Network-based predictions of retail store commercial categories and optimal locations. Phys Rev E 74:035101. https://doi.org/10.1103/PhysRevE.74.035101
33. Jensen P (2009) Analyzing the localization of retail stores with complex systems tools. In: Adams NM, Robardet C, Siebes A, Boulicaut J-F (eds) Advances in intelligent data analysis VIII. Springer, Berlin, pp 10–20
34. Reilly WJ (1931) The law of retail gravitation. Knickerbocker Press, New York
35. Huff DL (1964) Defining and estimating a trading area. J Mark 28:34–38. https://doi.org/10.2307/1249154
36. Cliquet G, Baray J (2020) Location-based marketing. Wiley
37. Ladle JK, David SD (2009) Retail site selection: A new, innovative model for retail development. Cornell Real Estate Rev 7:1–27
38. Önüt S, Efendigil T, Soner Kara S (2010) A combined fuzzy MCDM approach for selecting shopping center site: an example from Istanbul, Turkey. Expert Syst Appl 37:1973–1980. https://doi.org/10.1016/j.eswa.2009.06.080
39. García JL, Alvarado A, Blanco J et al (2014) Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an analytic hierarchy process. Comput Electron Agric 100:60–69. https://doi.org/10.1016/j.compag.2013.10.009
40. Wang S-P, Lee H-C, Hsieh Y-K (2016) A Multicriteria approach for the optimal location of gasoline stations being transformed as self-service in Taiwan. Math Probl Eng 2016:1–10. https://doi.org/10.1155/2016/8341617

41. Roig-Tierno N, Baviera-Puig A, Buitrago-Vera J, Mas-Verdu F (2013) The retail site location decision process using GIS and the analytical hierarchy process. Appl Geogr 40:191–198. https://doi.org/10.1016/j.apgeog.2013.03.005
42. Choudhury S, Howladar P, Majumder M, Saha AK (2019) Application of novel MCDM for location selection of surface water treatment plant. IEEE Trans Eng Manag. https://doi.org/10.1109/TEM.2019.2938907
43. Çoban V (2020) Solar energy plant project selection with AHP decision-making method based on hesitant fuzzy linguistic evaluation. Complex Intell Syst 6:507–529. https://doi.org/10.1007/s40747-020-00152-5
44. Hidalgo CA et al (2018) The principle of relatedness. In: Morales A, Gershenson C, Braha D, Minai A, Bar-Yam Y (eds) Unifying themes in complex systems IX. ICCS 2018. Springer proceedings in complexity. Springer, Cham. https://doi.org/10.1007/978-3-319-96661-8_46
45. Tobler WR (1979) Cellular geography. Philosophy in geography. Springer Netherlands, Dordrecht, pp 379–386
46. Karamshuk D, Noulas A, Scellato S et al (2013) Geo-spotting: mining online location-based services for optimal retail store placement. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 793–801
47. Lin J, Oentaryo R, Lim E-P et al (2016) Where is the goldmine? In: Proceedings of the 27th ACM conference on hypertext and social media—HT ’16. ACM Press, New York, pp 93–102
48. Xu M, Wang T, Wu Z et al (2016) Store location selection via mining search query logs of baidu maps. arXiv:1606.03662
49. Chen L, Zhang D, Pan G et al (2015) Bike sharing station placement leveraging heterogeneous urban open data. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing—UbiComp ’15. ACM Press, New York, pp 571–575
50. Chen L, Zhang D, Wang L et al (2016) Dynamic cluster-based over-demand prediction in bike sharing systems. In: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing. ACM, New York, pp 841–852
51. Chen J, Yu C, Jin H (2019) Evaluation model for business sites planning based on online and offline datasets. Futur Gener Comput Syst 91:465–474. https://doi.org/10.1016/j.future.2018.08.024
52. Rohani AMBM, Chua F-F (2018) Location analytics for optimal business retail site selection. In: Computational science and its applications—ICCSA 2014 and its applications—ICCSA 2018. Springer International Publishing, pp 392–405
53. Guo B, Li J, Zheng VW et al (2018) CityTransfer. Proc ACM Interact Mobile Wearable Ubiquit Technol 1:1–23. https://doi.org/10.1145/3161411
54. Marcon E, Puech F (2010) Measures of the geographic concentration of industries: improving distance-based methods. J Econ Geogr 10:745–762. https://doi.org/10.1093/jeg/lbp056
55. Reichardt J, Bornholdt S (2004) Detecting fuzzy community structures in complex networks with a Potts model. Phys Rev Lett 93:218701. https://doi.org/10.1103/PhysRevLett.93.218701
56. Bollobás B (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. Eur J Combin 1:311–316. https://doi.org/10.1016/S0195-6698(80)80030-8
57. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113. https://doi.org/10.1103/PhysRevE.69.026113
58. Sarzynska M, Leicht EA, Chowell G, Porter MA (2016) Null models for community detection in spatially embedded, temporal networks. J Complex Netw 4:363–406. https://doi.org/10.1093/comnet/cnv027
59. Fagiolo G, Squartini T, Garlaschelli D (2013) Null models of economic networks: the case of the world trade web. J Econ Interact Coord 8:75–107. https://doi.org/10.1007/s11403-012-0104-7
60. Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44. https://doi.org/10.1016/j.physrep.2016.09.002
61. Porter M, Onnela J-P, Mucha PJ (2009) Communities in networks. Am Math Soc. https://doi.org/10.1016/j.physrep.2009.11.002
62. Danon L, Díaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. J Stat Mech Theory Exp 2005:P09008–P09008. https://doi.org/10.1088/1742-5468/2005/09/P09008
63. Gómez S, Jensen P, Arenas A (2009) Analysis of community structure in networks of correlated data. Phys Rev E Stat Nonlinear Soft Matter Phys 80:16114. https://doi.org/10.1103/PhysRevE.80.016114
64. Vinh NX, Epps J, Bailey J (2010) Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. J Mach Learn Res 11:2837–2854
65. Meilă M (2007) Comparing clusterings—an information based distance. J Multivar Anal 98:873–895. https://doi.org/10.1016/j.jmva.2006.11.013
66. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep 2:336. https://doi.org/10.1038/srep00336
67. Lancichinetti A, Radicchi F, Ramasco JJ, Fortunato S (2011) Finding statistically significant communities in networks. PLoS ONE 6:e18961. https://doi.org/10.1371/journal.pone.0018961
68. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008:P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
69. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA 105:1118–1123. https://doi.org/10.1073/pnas.0706851105
70. Gregory S (2010) Finding overlapping communities in networks by label propagation. New J Phys 12:103018. https://doi.org/10.1088/1367-2630/12/10/103018
71. Newman MEJ (2004) Analysis of weighted networks. Phys Rev E 70:056131. https://doi.org/10.1103/PhysRevE.70.056131
72. Arenas A, Duch J, Fernández A, Gómez S (2007) Size reduction of complex networks preserving modularity. New J Phys 9:176–176. https://doi.org/10.1088/1367-2630/9/6/176
73. Newman MEJ (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103:8577–8582. https://doi.org/10.1073/pnas.0601602103
74. Duch J, Arenas A (2005) Community detection in complex networks using extremal optimization. Phys Rev E 72:027104. https://doi.org/10.1103/PhysRevE.72.027104
75. Arenas A, Fernández A, Gómez S (2008) Analysis of the structure of complex networks at different resolution levels. New J Phys 10:053039. https://doi.org/10.1088/1367-2630/10/5/053039
76. Newman MEJ (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69:066133. https://doi.org/10.1103/PhysRevE.69.066133