Knowledge Transfer in Commercial Feature Extraction for the Retail Store Location Problem

VIRGINIA AHEDO, JOSÉ IGNACIO SANTOS, AND JOSÉ MANUEL GALÁN
Departamento de Ingeniería de Organización, Universidad de Burgos, 09006 Burgos, Spain

Corresponding author: José Manuel Galán (jmgalan@ubu.es)

ABSTRACT Location is the most important strategic decision in retailing. The location problem is markedly complex and multicriteria. One of the key factors to consider is the so-called balanced tenancy —i.e., the degree to which neighboring businesses complement each other. There are several network-based methodologies that formalize the notion of balanced tenancy by capturing the spatial interactions between different commercial sectors in cities. Some of these methodologies provide indices that have been successfully used as input features in location recommendation systems. However, from a predictive perspective, it is still unknown which of the indices provides best results. In this work, we analyze the performance of six of these indices on a set of nine Spanish cities. Our results show that the combined use of all of them in an ensemble model such as random forest significantly improves predictive accuracy. In addition, we explore the effect of knowledge transfer between cities from two different perspectives: 1) quantify how much the quality of solutions degrades when the balanced tenancy of a city is explored through the indices obtained from another city; 2) investigate the interest of network consensus approaches for knowledge transfer in retailing.

INDEX TERMS Complex networks, feature extraction, recommendation systems, retail location problem, prediction, knowledge transfer.

I. INTRODUCTION

Location is considered one of the most important strategic decisions in retailing, as it provides a significant competitive advantage that cannot be imitated by competitors. This strategic decision is particularly critical for many Small and Medium-sized Enterprises (SMEs) for which in-store sales still constitute the most important distribution channel [1], [2].

The problem of location is markedly complex and multicriteria, as it is influenced by a plethora of different factors: demographic variables (e.g. population size, population density, age profile), socio-economic characteristics of the area (e.g. base economy, availability of workers, availability and cost of space, commercial ecosystem, competitive behavior of the sector, customer behavior and preferences), legal and fiscal factors (e.g. tax policies, regulations), accessibility (e.g. transport and communications infrastructure),

availability of supplies, climatic and environmental aspects, etc. [2], [3].

Given its relevance, location selection has been extensively studied in the literature. Traditional approaches have typically resorted to offline customer surveys, census data and revenue statistics [4], [5]. However, such methods are extremely time-consuming, expensive and unable to capture the dynamic nature of the market. Fortunately, with the advent of the Information and Communication Technologies (ICTs) and the development of social networks, geolocated mobile services and Big Data, we now have access to plenty of information on the geospatial distribution of stores, their popularity, customer mobility patterns, consumption behaviors, customer opinions, etc. This information, if adequately exploited, can provide valuable insights into the location problem.

Given the complex nature of the problem and the availability of unprecedented volumes of data, it has become common practice to use Multi-Criteria Decision-Making (MCDM) methods to address location decisions [6]–[11]. The embracement of multicriteria approaches, together with the continued development of sophisticated data analysis techniques,

The associate editor coordinating the review of this manuscript and approving it for publication was Yichuan Jiang.

has led to the development of location recommendation systems.

Location recommendation systems can differ in two important dimensions, namely: according to the paradigm they adopt (content-based recommendation, link-analysis-based recommendation or collaborative filtering recommendation), and according to the features they use as inputs [12]. On this topic there exist multiple lines of research, including (i) the development of alternative recommendation approaches, (ii) the implementation of alternative/more efficient versions of pre-existing recommendation systems, and (iii) the improvement of feature extraction, assessment and selection. In this paper, we focus on this last line of research.

Specifically, among the different types of variables that can be used to feed a location recommendation system, here we focus on a set that represents the commercial spatial interactions in a city. This choice is supported by the fact that an adequate balanced tenancy —i.e., an appropriate balance of neighboring businesses operating in different commercial categories and thus complementing one's business— is one of the most relevant success factors in retailing [1]. As Hidalgo *et al.* [13] put it, the interdependence among retail players can be summarized with *the principle of relatedness*, which states that the probability that a new retail store locates in a given commercial area is a function of the number of related activities present in that area. Indeed, the existence of complementary/correlated activities could explain the agglomeration patterns of commercial activities that are often found in cities [14]. Nonetheless, the formalization and evaluation of the commercial spatial interactions between retail shops is not a trivial endeavor. This is mainly due to (i) the diversity of commercial activities —which makes competitive and/or complementary relationships difficult to differentiate; and to (ii) the fact that both types of forces coexist simultaneously in retail clusters, hence making it difficult to determine their individual effect.

Within the framework of retailing, several attempts have been made to formalize the notion of balanced tenancy in terms of business-to-business interactions in a given area. However, to this day no method has proved clearly superior to the rest. Different approaches provide complementary and/or overlapping insights [15]. To our knowledge, the pioneering proposal is Jensen's [16], [17] network approach. In Jensen's seminal works, two metrics were defined to quantify the pairwise inter- and intra-category interactions between retail stores. These two metrics are used to compute a global index (Jensen's Quality Index) that can be employed to assess the quality of a potential location in terms of the attractive or repulsive forces that neighboring stores exert on it. Jensen's Quality Index is very popular in the retail research field, where it is often used as an input for supervised learning models [18]–[20] and for location recommendation systems [20].

In Jensen's approach, the significance of the attractive/repulsive forces between retail categories (e.g., between pharmacy and drug stores, meat markets and computer stores, etc.) is determined by comparison against a null model that assumes a random distribution of retail shops. Nevertheless, it is possible to use alternative null models that present different degrees of preservation of the local and/or global commercial structure of the city, as well as alternative metrics to quantify the statistical deviations from the average commercial behavior of the city [15]. Different alternatives capture different aspects and present their own advantages and shortcomings. Therefore, it is of interest to explore whether one of these indices is better than the rest, or the best approach is to consider all indices simultaneously. This is a question we address at length in this paper.

Besides determining the performance of the different indices as inputs for location recommendation systems, we also explore the transferability of knowledge between cities. This problem is receiving a lot of attention in urban computing [21], not only for retailing but also from a more general perspective. Previous works in this line have explored knowledge transfer between cities for air quality prediction [22] and for human mobility patterns [23], [24], among other applications.

Nonetheless, the issues behind the idea of knowledge transfer are very general and fundamental. In machine learning, many predictive models implicitly assume that the training set from which a pattern is learned comes from the same system on which predictions will be made. However, this is not always the case, for technical, economic, or other reasons. In these scenarios, understanding the limits of transferability is a central issue in decision-making. In the case of cities, the problem can be stated as "can we transfer knowledge from a city where data are sufficient to a city that faces either data insufficiency or label scarcity?" [22]. As far as we know, this question has not yet been addressed for retail location, and we explore it for the first time in the present contribution.

Note that if the number of source cities —i.e., cities from which to transfer knowledge— is greater than one, then there is the additional challenge of determining how to condense the different sources properly. In this work, for instance, we use network consensus techniques to address such information fusion problem.

In short, in the present work we have focused on the individual and/or joint use of the different retail-commercial-interaction indices as either inputs for location recommendation systems or regressors for machine learning predictive models, and we also explore their transferability between cities. Below we detail the specific research questions we address, ascribing them to two main approaches.

A. TECHNIQUES THAT USE INFORMATION FROM THE CITY ITSELF (CITY'S OWN DATA)

Two different lines of inquiry were explored with this approach: (i) the individual use of the different indices and (ii) their joint use.

Regarding the individual use of the different quality indices, we have tried to answer the following questions:

Q1 From an empirical perspective, are there indices that are better than others in terms of the predictive accuracy attained?

Q2 To what extent is the business category of a retail store predictable using its business ecosystem, measured by each of the indices taken individually?

Q3 Are the different indices similar, or do they provide complementary information on the commercial structure of a city?

With respect to their joint use, we have addressed the following question:

Q4 Does the combined use of the different indices improve predictive accuracy?

B. TECHNIQUES THAT USE INFORMATION FROM OTHER CITIES — KNOWLEDGE TRANSFER

Within this second approach, we have explored two different possibilities: (i) the use of commercial information from cities other than the target city, and (ii) the joint use of commercial information from both other cities and the target city itself. In particular, the set of research questions addressed may be summarized as follows:

Q5 Is the information on the commercial spatial structure of a city transferable to other cities?

Q6 City-to-city knowledge transfer: Is good predictive accuracy attained by transferring data from a single source city to a single target city?

Q7 Within the framework of knowledge transfer, which of the predictive patterns identified are specific of the city itself, and which are related to the commercial structure shared across cities?

Q8 Consensus knowledge transfer: What is the predictive accuracy obtained by transferring data from multiple source cities to a single target city (both including or not the city's own data)? Is it better than the accuracy of transferring data from just a single city?

The remainder of this paper is organized as follows. In section II, we cover the theoretical background necessary to fully understand the present contribution, which comprises: (A) the different methods for modeling balanced tenancy, (B) the different quality indices, (C) the Mean Reciprocal Rank, (D) the different machine learning tools used to evaluate the joint performance of the different quality indices and (E) consensus approaches. In section III, we describe the dataset used in the analyses. In section IV, we explain how we have formalized the different research questions and the results obtained. More specifically, this section is articulated around two main blocks: (A) "Techniques that use information from the city itself — city's own data", and (B) "Techniques that use information from other cities — Knowledge transfer"; the different research questions are covered under their respective block. Finally, section V presents the conclusions and limitations of the present contribution.

II. THEORETICAL BACKGROUND
A. METHODS FOR MODELING BALANCED TENANCY

In this work, we use the three network methods described in [15] for modeling balanced tenancy, namely *Jensen*, *permutation* and *rewiring*. Notably, the three of them work at two different levels: (i) the network of commercial interactions between retail shops, and (ii) the network of interactions between the different commercial categories. In all cases, the network of commercial interactions between retail shops is defined in terms of the radius concept proposed by Jensen [16], [17]. Specifically, the different retail stores constitute the nodes, and an undirected link is created between each shop p and all the stores x that are located within a radius $r = 100$ m from it. The three approaches differ in the methodology for obtaining the networks of interactions between commercial categories.

In the case of *Jensen* [16], [17], the formalization of the network of interactions between commercial categories is based on two interaction coefficients: the intra-category (M_{AA}) and the inter-category (M_{AB}) coefficients. These two coefficients quantify the deviation of the empirical spatial distribution of retail stores from purely random distributions. More precisely, the intra-category coefficient is defined as the average local concentration of type-A stores divided by their global concentration:

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

where T is the set of all the stores in a certain area, A are the stores of category A and $N_s\,(p, r)$ represents the number of stores in set S within a radius r from shop p.

The inter-category coefficient is defined as the quotient of the local concentration of type-B stores around type-A stores to their respective global concentration:

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

Please note that, in the chosen nomenclature, A and B are just a generic representation of two different types of business categories, and act merely as indices. Therefore, they have to be duly substituted to explore all possible pairwise commercial interactions in the set of commercial categories considered.

These interaction coefficients are subsequently used to define the weight of the link between each pair of retail categories. Specifically, the different commercial categories constitute the nodes, and the logarithm of the inter- and/or intra-category coefficients constitutes the weight of the corresponding links. All these weight values are then collected in what is known as Jensen's matrix of interactions.

The significance of the attractive or repulsive commercial relationships thus identified is determined against two different null models via Monte Carlo sampling. In particular, the null model for the intra-category coefficient is the uniform randomization of all category-A shops over all possible

locations. In contrast, the null model for the inter-category coefficient is the random distribution of all retail shops except for those from category A, whose location is kept fixed. In both cases, the number of shops from each commercial category is maintained.

As regards the *permutation* and *rewiring* methods, both use the Z-score function to transform the network of commercial interactions between retail shops into the network of interactions between commercial categories [15]. More precisely, the Z-score function (3) determines the weight of the link between each pair of retail categories as follows:

$$Z_{AB} = \frac{x_{AB} - \overline{x_{AB}^{null\_model}}}{s_{AB}^{null\_model}} \qquad (3)$$

where x_{AB} is the empirical number of links between retail shops from category A and retail shops from category B, and $\overline{x_{AB}^{null_{model}}}$ and $s_{AB}^{null_{model}}$ are, respectively, the mean and the standard deviation of the corresponding null distribution of the number of links between these two categories. The *permutation* and *rewiring* approaches differ in the null models that they use to obtain such null distributions, which present different levels of preservation of the original commercial structure of the city. In the null model of the *permutation* method, the city's global commercial structure is maintained, being just the commercial categories of the different stores permuted. In contrast, in the null model from the *rewiring* method, it is the local commercial structure of each retail store that is preserved. Specifically, it maintains the position, commercial category and degree of each store while conducting random rewiring, i.e., randomly matching the half edges emanating from each node in the network. Note that, as in the case of *Jensen*, the Z-score values obtained with the *permutation* and *rewiring* approaches are also stored in interaction matrices.

The significance of the interactions found between retail categories is determined in the same way for all three methods. In particular, the processes of the different null models are repeated a sufficiently large number of times to obtain for each pair of categories a probability distribution of the values of the interaction coefficients in the case of *Jensen*, and of the Z-score for *permutation* and *rewiring*. For each of those probability distributions, the 2.5 and 97.5 percentiles are obtained and compared with the empirical values; only those interactions whose empirical value is outside the interval defined by the 2.5 and 97.5 percentiles are kept.

B. DIFFERENT QUALITY INDICES

As pointed in the Introduction, in addition to defining the inter- and intra-category coefficients, Jensen proposed a quality index —Jensen's Quality Index— to assess the suitability of a potential location in terms of the attractive and/or repulsive forces that the neighboring stores exert on it [16]. Since Jensen's Quality Index has been very well received within the framework of location recommendation systems, here we define the quality indices corresponding to the

permutation and *rewiring* methods, as well as an alternative to Jensen's original quality —what we have termed Jensen's Quality Raw— and its corresponding versions for *permutation* and *rewiring*. The interest of these new indices is to be found in that, as noted in [15], Jensen's original approach presents some technical problems that could lead to the generation of artifacts in the results. Furthermore, since the different approaches may capture different aspects of the same problem, their joint use could be beneficial in location recommendation algorithms.

The idea behind Jensen's Quality Index is that a location that resembles the average location of the shops from a given commercial category may be a good site for a new retail store from that category. More precisely, Jensen's Quality Index of a certain location (x, y) for activity i is defined as follows [16]:

$$Q_{JENSEN_i}(x, y) = \sum_{j=1}^{N} a_{ij}\left(nei_{ij}(x, y) - \overline{nei_{ij}}\right) \qquad (4)$$

where N is the total number of commercial categories, $nei_{ij}(x, y)$ represents the number of neighboring stores of category j around (x, y), $\overline{nei_{ij}}$ is the average number of neighbors of category j that the shops of category i have, and $a_{ij} = \log(M_{ij})$ is the corresponding value of the Jensen's matrix of interactions between commercial categories (the radius is considered constant during all the analyses).

The conceptualization of Jensen's Quality Index can be extrapolated quite straightforwardly to the methods of *permutation* and *rewiring*. Specifically, given that the three methods consist in obtaining an interaction matrix between commercial categories, and that Jensen's Quality Index uses the elements of such matrix as weighting factors, here we define the Permutation Quality Index and the Rewiring Quality Index analogously. Concretely, our definitions consist in changing the value of the weighting factor (a_{ij}) from the logarithm of Jensen's coefficients to the corresponding Z-score values obtained for each pair of categories with the *permutation* or *rewiring* method. At this point, it is important to note that the three quality indices are calculated on their respective interaction matrices *after assessing the significance of the relationships found* —i.e., they are obtained from interaction matrices that only include the relationships that were found to be significant against their respective null models.

In a subsequent step, we defined what we call *Raw Quality Indices*, which differ from the previous ones in that they do not take as reference the average number of neighbors of category j that shops of category i have ($\overline{nei_{ij}}$). Thus, we define Jensen's Raw Quality Index as follows:

$$Q_{JENSEN-RAW_i}(x, y) \equiv \sum_{j=1}^{N} a_{ij}\left(nei_{ij}(x, y)\right) \qquad (5)$$

Again, the Permutation and Rewiring Raw Quality Indices can be obtained by simply changing the weighting factors a_{ij} in (5) as explained above.

One last remark regarding the calculation of the different indices is that, in this study, we have considered 68 commercial categories, as we have worked with the North American Industry Classification for Small business (NAICS) standard [25] to make our results comparable with previous works [15], [16], [26], [27].

C. INDEX PERFORMANCE ASSESSMENT: MEAN RECIPROCAL RANK (MRR)

In the problem of location recommendation there is no single correct answer. A simple and adequate approach to assess the performance of location recommendation systems that output a ranking of the most suitable commercial categories for a given location is to identify the position of the actual commercial category in that ranking.

Following this approach, we have evaluated the performance of the different quality indices by means of the Mean Reciprocal Rank (MRR), which is the average of the reciprocal ranks of the results obtained for a sample of locations Q:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \qquad (6)$$

D. JOINT USE OF THE DIFFERENT QUALITY INDICES

In this contribution, we address the joint use of the different quality indices using a supervised learning approach. Specifically, we opted for a random forest classifier on grounds of its good performance and because it facilitates conducting variable importance analyses. Such analyses are very useful to determine the possible complementarities between indices. Both the random forest algorithm and the different types of variable importance analyses are detailed below.

1) RANDOM FOREST

In machine learning, ensembles are methods that combine the predictions of multiple learning algorithms to increase predictive accuracy beyond that of the constituent algorithms alone. In fact, they are based on the assumption that, in prediction, combining different base models is a better strategy than using them separately [28], [29]. Random forests are one of the most popular ensemble methods that exist, as they have proved to perform remarkably well in a wide range of different problems [30], [31]. Here we use random forests for classification. In terms of its internal logic, a random forest classifier consists of different classification trees that are built on different bootstrapped samples of the training set; this facet, together with the random subspace method, serves to minimize the correlations between trees [30], [31]. More precisely, for each individual tree, the random subspace method considers at each split a fresh random sample of predictors —instead of the full set of them— to prevent trees from being too similar in the presence of strong predictors. Consequently, these individual trees present high variance and low bias, an aspect that is compensated by averaging the resulting predictions of the different trees. As a result,

the overall variance of the model is reduced, and thence, a good global bias-variance tradeoff is achieved. The resulting random forest is characterized by its robustness against correlated predictors and by the fact that increasing the number of trees in the forest does not result in overfitting [34].

2) VARIABLE IMPORTANCE (GINI)

As stated in the previous section, random forests have a good bias-variance tradeoff and therefore achieve good predictive accuracy. However, they do so at the expense of interpretability. For this reason, random-forest-based variable importance analyses are commonly used to shed light on their inner workings. In very broad terms, these random-forest-based variable importance analyses can be divided into two main groups: (i) individual and (ii) group variable importance.

Within the framework of individual variable importance analyses, the permutation method proposed by Breiman [33] is probably the most pervasive of all of them. Specifically, it determines the relative importance of each predictor by quantifying the impact of randomly permuting its values on the final classification accuracy of the model. Notwithstanding, accuracy is not a suitable metric for location recommendation systems, as they typically provide a ranking of different possibilities instead of a single output. It is precisely for this reason that we proposed the MRR as an adequate performance index for them. Therefore, an accuracy-based importance metric does not seem to be the best choice for our problem. Consequently, we propose to use the Gini variable importance instead.

Gini variable importance is inspired by the node-impurity-minimization criterion that is used when growing a classification tree. Such criterion selects, for each split, the variable and split point that induce the greatest decrease in node impurity —i.e., in the heterogeneity of classes at that node [34]. In fact, the idea underlying Gini variable importance is that important variables are those that, when used for splitting, result in a large decrease in node impurity measured by the Gini impurity. Gini impurity is defined as follows:

$$\sum_{i=1}^{C} f_i (1 - f_i) \qquad (7)$$

where f_i is the relative frequency of class i in the node under consideration, and C is the total number of classes. The decrease of impurity is the difference between a node's impurity and the weighted sum of the impurity values of the two child nodes —the weights being the number of observations in these two child nodes. [34], [35] The Gini variable importance of predictor X_j is calculated as the sum of the decrease in the Gini impurity in all nodes of the forest that split over X_j, normalized by the number of trees [36].

Gini importance is not recommended when dealing with both numerical and categorical predictors, when predictors vary in their scale of measurement, or if all predictors are categorical but their number of classes is markedly different, as it is known to be biased in favor of variables with many

possible split points —i.e., continuous or high cardinality variables [37]. However, these caveats do not constitute a problem in our case study, since all our quality indices (Jensen Quality, Permutation Quality, Rewiring Quality and their raw versions) are continuous, operate with the same scale, and their individual importance has been obtained using the 'impurity_corrected' importance measure from *ranger* package [38] —which is unbiased in terms of the number of categories and category frequencies [37].

Thus far, we have described the different individual variable importance analysis methods. As regards group variable importance analyses, they are the extension of the foregoing approaches to the case of variables being grouped under different categories. More specifically, Gini group variable importance is defined as the direct sum of the importance of the individual variables in each category.

E. CONSENSUS APPROACHES

In Network Science, different consensus approaches exist depending on the objective pursued [39], [40]. Within the framework of network methods that model the notion of balanced tenancy, consensus techniques have also been proposed. In particular, in [15] two consensus methods were defined to assess whether a robust core of commercial relationships was shared across different cities. From these two methods, in this contribution we rely on the consensus networks of relationships. In such networks —which are weighted and signed— the nodes are the different retail categories, and the weights of the links are calculated as follows: for each pair of commercial categories, we check across all cities considered if a significant relationship exists between them, and if so, we add 1 if the relationship is positive or -1 if it is negative. Thereupon, the absolute value of the maximum and the minimum possible weight would be equal to the total number of cities under consideration. Notably, different consensus networks of relationships can be obtained depending on the method used to model the interactions between commercial categories: consensus network of relationships by *Jensen*, *permutation* or *rewiring*.

As for the number of commercial categories considered to obtain the consensus networks in this contribution, we have again worked with the 68 NAICS commercial categories (instead of the intersection of them present across all the cities in the study).

Finally, another relevant aspect in the framework of consensus networks of relationships is the use of thresholds — i.e., the minimum weight of a link to be considered in subsequent analyses. If, for instance our threshold is 5, only those relationships found in at least five cities or more will be taken into account.

III. DATASET

The dataset used for the present study consists of the nine provincial capitals of Castile and Leon —an autonomous community in northwestern Spain [41]. It should be noted that the selected capitals are representative of small and medium-sized Spanish cities. As for the commercial information of each city —i.e., the commercial category and address of each retail store— it was extracted from the Yellow Pages during 2017. Finally, all addresses were georeferenced using the MapQuest Application, Open Street Map data and the Google Maps API.

IV. PROBLEM FORMALIZATION AND RESULTS

A. TECHNIQUES THAT USE INFORMATION FROM THE CITY ITSELF (CITY'S OWN DATA)

1) INDIVIDUAL ASSESSMENT OF THE DIFFERENT QUALITY INDICES ON EACH CITY'S OWN DATA

To assess the individual potential of each of the proposed quality indices as input features for location recommendation systems, the problem was formalized as follows.

For each city, the network of commercial interactions between retail shops was obtained. Subsequently, a k-fold cross-validation approach was selected to determine the performance of each of the quality indices. K-fold cross-validation is one of the most widespread approaches for the honest evaluation of models/indices, as it consists of dividing the original dataset into k disjoint folds, using $k-1$ folds to train the model/calculate the respective index, and the remaining fold to assess its performance. The process is repeated k times until all folds have been used as test set, with the estimate of the performance being the average value over the k folds. Typical values for k are 5 or 10. However, in our study, for the results to be insightful it is necessary to preserve the empirical commercial structure of the city as much as possible when dividing the data into training and test. In this regard, recall that if we took $k = 10$, for each iteration of the cross-validation loop, we would be removing 10% of the nodes in the original network and their corresponding links, a percentage high enough to substantially distort the empirical commercial interactions between shops. Consequently, we opted for $k = 100$, as it implies that each of the folds contains just 1% of the nodes, thus significantly reducing the impact of removing them from the training set (albeit at the expense of increasing the computational cost of the analyses). From each such training set, both the interaction matrices between commercial categories for the three methods — *Jensen*, *permutation* and *rewiring*— and the frequency vectors of each commercial category —necessary for the computation of the quality indices that are not raw— were obtained. All retail categories were considered in the computation of the interaction matrices (even if some of them only contained a single store in certain cities), and the significance of the empirical interactions was determined against 250 iterations of the respective null models.

Once the interaction matrices were obtained, for each node (i.e., retail store), the different quality indices were calculated for the 68 NAICS business categories and ranked in descending order of quality. To assess the goodness of each index, the ranking position of the true class —i.e., the original commercial category of the eliminated node in the network— was

evaluated. Recall that high positions in the ranking indicate that the quality index captures the actual business typology well, whereas low values indicate the opposite. An additional remark in this regard is that since the commercial categories with only one store were considered in the calculation of the interaction matrices (training set), they were not used in the evaluation process of each node's quality indices (test set), as one single store cannot be simultaneously in both sets.

Fig. 1 provides a schematic diagram of the experimental design used to assess the individual and joint performance of the 6 quality indices on the city's own data. The 100-fold cross-validation process is illustrated for 2 iterations.

The first six columns of Fig. 2 show the distribution of ranking positions obtained for each index in each of the cities evaluated. They are compared with a random distribution, i.e., a distribution in which the ranking position is
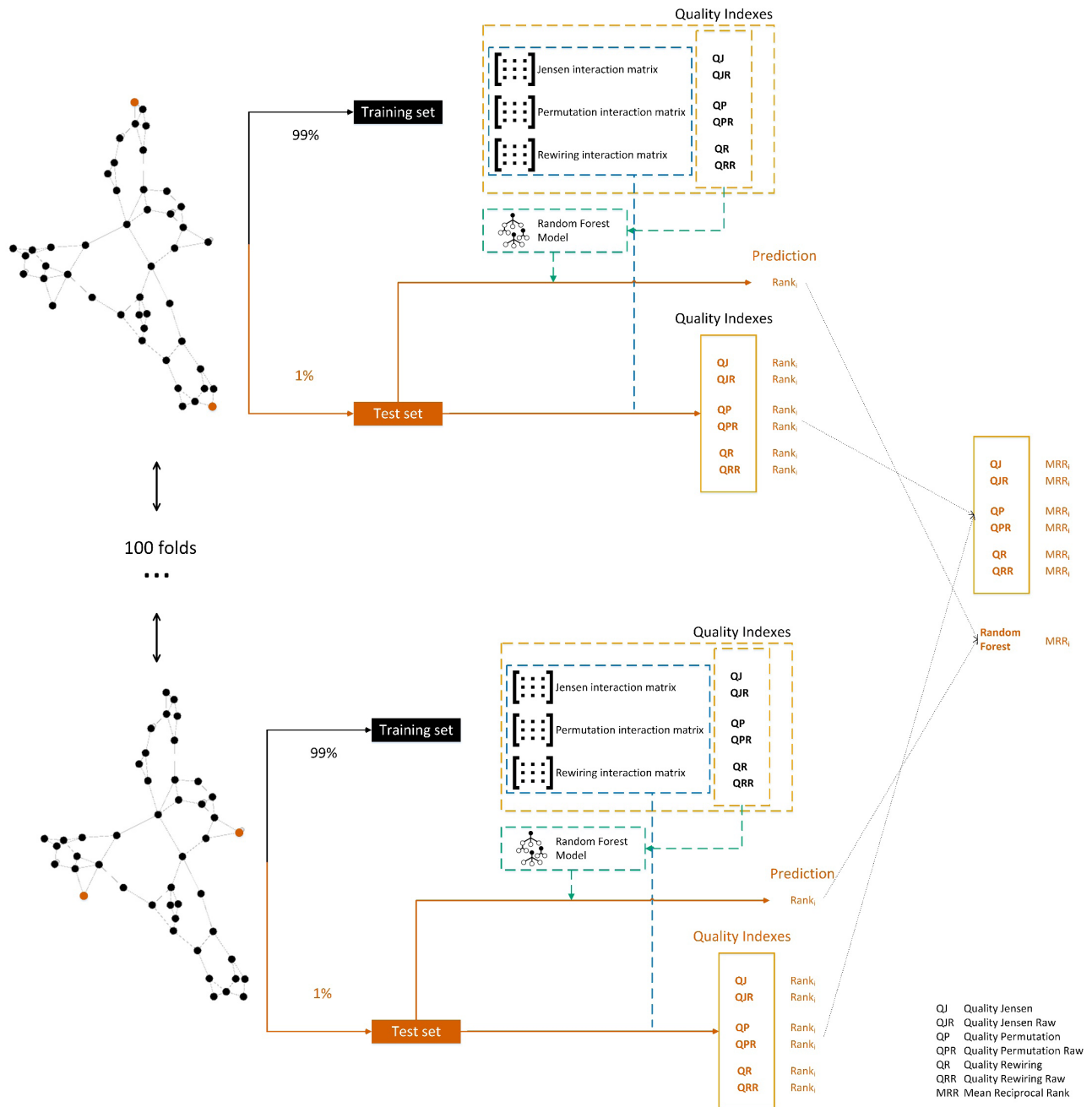
FIGURE 1. Scheme of the analysis framework used to explore the individual and joint performance of the different quality indices on the city's own data. The nodes in the 99 folds that constitute the training set are represented in black. The nodes in the test set (remaining fold) appear in orange. The calculation of the three interaction matrices, the different quality indices and their subsequent assessment on the test set are also duly schematized. Dashed lines indicate calculations made on the training set while solid lines indicate calculations made on the test set. A legend of the different acronyms is provided.
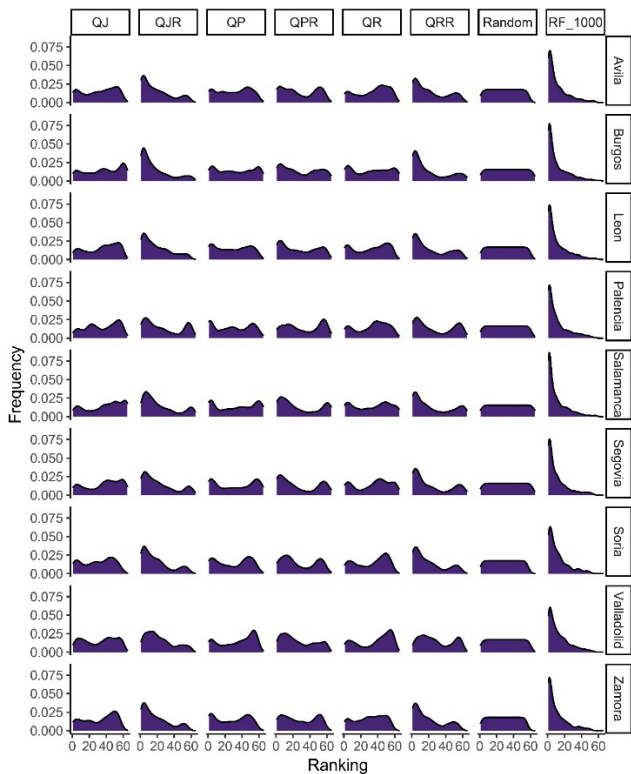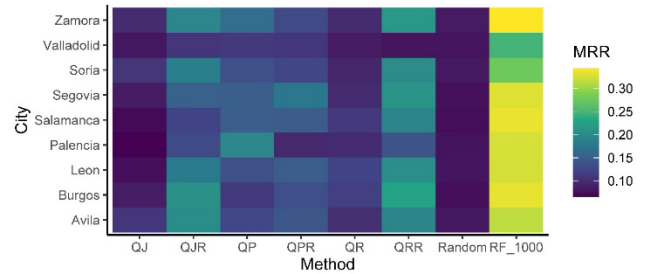
FIGURE 2. Performance (ranking) distribution by method and city. This figure shows the distribution of the position of the correct prediction (actual class) among the 68 NAICS possible classes of the classification problem, for each method and city. A significant performance improvement is attained by combining the different metrics in a random forest, since, as it can be seen, the ranking distribution thus obtained is markedly skewed to the left.

FIGURE 3. Average performance (Mean Reciprocal Rank — MRR) by method and city.

randomized. The ranking distributions of the indices Quality Jensen Raw and Quality Rewiring Raw reveal that these two metrics capture reasonably well the commercial category to be predicted from its commercial environment (Q1).

The information contained in the ranking distributions of Fig. 2 is summarized in Fig. 3, which presents the MRR values obtained by method and city, thus providing a general overview of the average performance of the different metrics. Remarkably, the best indices obtain MRR values consistently higher than 0.2 (i.e., the position of the true class in the ranking is on average the fifth or above), which demonstrates their capacity to capture the spatial context interactions (Q1, Q2).

2) EVALUATION OF THE COMPLEMENTARITY OF THE DIFFERENT QUALITY INDICES AND THEIR JOINT USE THROUGH A CLASSIFICATION APPROACH

Here we (i) determine the possible similarities/ complementarities between the different quality indices and (ii) assess the predictive potential of their combined use. These two research questions can be quite straightforwardly formalized through a classification approach in which the variable to predict is the actual commercial category of each store, and the regressors are the values of the different quality

indices obtained for each retail store (one value per NAICS commercial category and index).

It is important to clarify that to evaluate the individual predictive performance of each quality index, only its 68 values were considered (one value per commercial category), and that their comparison was made by means of a simple rule: the best index is the one with the best MRR. However, when assessing the joint predictive potential of all quality indices, 6 indices · 68 values/index = 408 values were at play, being all of them aggregated in a classification model manifestly more complex and sophisticated than the aforementioned rule.

To ensure both consistency and coherence with the previous results, in the experimental design for the classification approach, we used the same 100 folds as in the research question above (100-fold cross-validation scheme —see Fig. 1). To be clear, we iteratively trained random forests of 1000 trees on datasets made up by 99% of the nodes, and tested them on the remaining 1%. The number of trees was set to 1000, since this value is above the experimental threshold of significant improvement for datasets of this size [42]. The predictions thus obtained were analyzed both in terms of their distribution (Fig. 2) and of the MRR (Fig. 3). These results reveal that the combined use of the different quality indices achieves a significant increase in predictive performance, as the ranking distribution is markedly skewed to the left in all cities, and the MRR values are consistently above 0.3. Accordingly, we can conclude that the different quality indices are complementary, i.e., they capture different aspects of the problem, and hence, their combined use results in better predictive accuracy (Q3, Q4).

Finally, we used the implementation of the unbiased Gini group variable importance by Wright & Ziegler [38] to determine the relative importance of each index within the framework of the previously trained random forests. We used group variable importance instead of individual variable importance since we have 68 values for each quality index (one value per NAICS commercial category), so it is necessary to create six groups of variables (one per quality index) under which we subsume the corresponding 68 values obtained with each of them. The results obtained are presented in Fig. 4 and show that, although the permutation quality indices do not have a very relevant performance in isolation, they are indeed relevant when combined in the classifier. In any case, this analysis

shows that the six measures have a significant contribution, which supports the idea of complementarity (Q3, Q4).
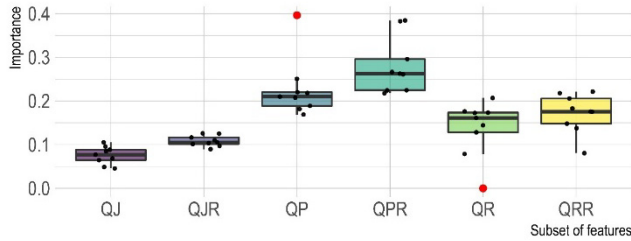
FIGURE 4. Normalized relative group importance (Gini impurity corrected). Outliers are represented as red dots.

B. TECHNIQUES THAT USE INFORMATION FROM OTHER CITIES — KNOWLEDGE TRANSFER

1) TRANSFER OF KNOWLEDGE BETWEEN CITIES

The research question on the transfer of commercial knowledge between cities was explored from two different perspectives: (i) the individual predictive potential of each quality index from the source city with respect to the target city, and (ii) the joint predictive potential of all the indices from the source city to make predictions on the target city.

Given that the values of the six quality indices for each shop and NAICS commercial category in each city were obtained in the previous sections, to explore these research queries it was only necessary to define the experiments, which were set up as follows. The different cities were taken iteratively as either source or target city, so all of them acted as target once. For each target city, the remaining eight cities were considered individually as source; in particular, in a first approach, each of the six quality indices from the source city was taken separately to make predictions on the target city. Finally, to determine their joint predictive potential, all quality indices from the source city were combined by means of a random forest to predict the commercial categories of all nodes in the target city. Recall that, in this case, we have used the random forest classifier again for consistency reasons, i.e., for coherence with the previous analyses.

Fig. 5a presents the MRR values obtained for each city pair with a random forest of 1000 trees —to make the results comparable with those in Fig. 3— trained on the dataset that includes all quality indices from the source city. In comparison with Fig. 3, a clear degradation of performance is observed when using data from another city. However, it is worth highlighting that the overall results are not poor, as the MRRs are consistently above 0.2. Fig. 5b shows the MRR values obtained when using each index individually, and also when considering all quality indices together (RF_1000). The points in blue represent analyses in which the source and the target city do not coincide, whereas the points in red are those obtained with the city's own data —i.e., the results already presented in Figs 2 and 3. In all cases, a loss of precision is observed when data are transferred from one city to another (Q5, Q6).
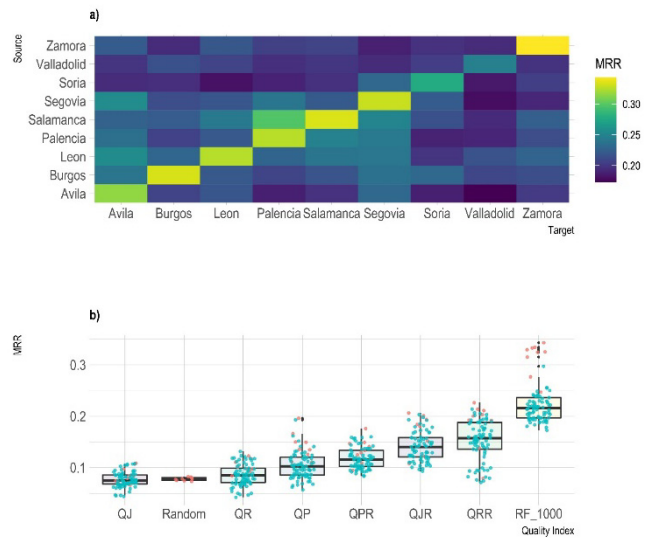
FIGURE 5. a) Mean Reciprocal Rank (MRR) obtained when using all the quality indices of a city (source) as training data for a random forest of 1000 trees to make location predictions in another city. Results are compared with those that used data from the same city (Fig. 3) — ascending diagonal in this figure. b) MRR performance for every method with and without transferred data. Blue dots are results obtained when the source and the target city are different, whereas red dots represent the results of using data from the same city.

Fig. 6 compares the ranking distributions obtained when the random forest is trained on the city's own data with those obtained when it is trained on transferred data. Whereas the performance attained with data from another city is reasonably accurate, it is systematically worse than the performance obtained with the city's own data (Q5, Q6).
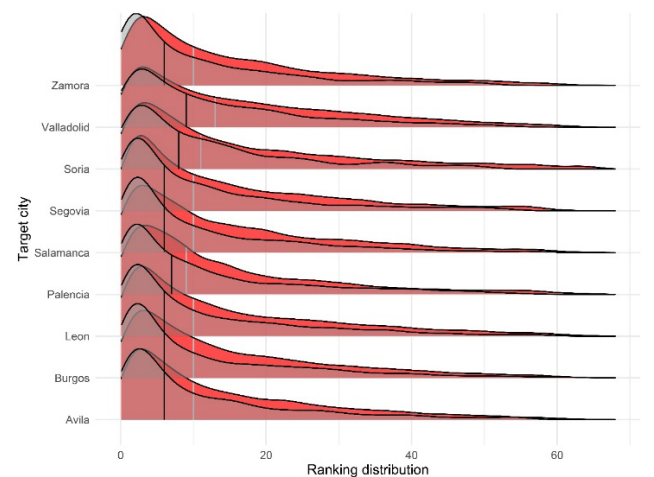
FIGURE 6. Comparison of the performance of own vs. transferred data in the ranking distribution. Gray distributions (median represented in black) represent the results obtained with data from the city itself whereas red distributions (median represented in light grey) are obtained with different source and target cities.

2) IDIOSYNCRASY OF EACH CITY VS. SHARED COMMERCIAL STRUCTURE

In the context of knowledge transfer, an interesting question arises in relation to the predictability of the commercial

category of a given store: which of the identified patterns are specific of the city itself and which are related to the commercial structure shared across cities? Put another way: what is the level of specificity of the commercial pattern of one city compared to that of other cities?

The answer to this question is very relevant from a decision-making perspective. Such information would allow us to determine if it is worth obtaining the commercial data of new urban areas, or, on the contrary, it is more cost-effective to use already available data from other cities to make predictions on these new areas. Specifically, we formalized this question as a multiple linear regression (MLR) problem in which the dependent variable was the MRR —the level of predictability of the commercial category as a function of the spatial context— and three independent explanatory variables were proposed: (i) the level of spatial commercial specialization of the source city, (ii) that of the target city and (iii) a dummy variable indicating whether transferred data were used. Note that the level of commercial spatial specialization is an indicator of the level of organization and structure that each city has. In this regard, previous studies suggest that it is likely to increase with city size [15]. The rationale behind the explanatory variables proposed is that cities that are highly commercially organized may allow us to capture patterns of commercial interactions that can be transferred to other cities, whereas learning patterns from cities with more random structures may be more difficult. Similarly, predicting over target cities with a high level of commercial and spatial organization may be easier than over cities with lower structure.

To estimate each city's level of specialization and commercial organization, we can use the *Jensen*, *permutation* and/or *rewiring* interaction matrices. More precisely, the number of non-zero values in these matrices reflects the number of significant relationships found between commercial categories. Although the underlying assumptions and the weights of the relationships are different in the three approaches, they are highly correlated. Therefore, we decided to condense them by conducting Principal Component Analysis (PCA) and retaining only the first dimension. Note that prior to conducting PCA, we normalized the data without centering it so that the indicator is positive, and hence, more meaningful. Remarkably, in our case study, the first dimension of the PCA explains 98% of the variance, so the loss of information is minimal. In return, by keeping it, we eliminate possible collinearities and increase the interpretability of the results.

Once the commercial spatial organization of all cities was determined, the MLR model was built. Within the framework of MLR, the effect of a dummy variable can be modeled as either additive —leading to a change in the intercept— or multiplicative —leading to a change in the slope of the regression line. In our case, a significant regression equation was found for the additive approach ($F(3,77) = 85.05$, p-value < 2.2e-16), with a Multiple R^2 of 0.7682 and an Adjusted R^2 of 0.7592. These results (see Table 1) demonstrate that the level of commercial organization of the source

city is highly significant. At the same time, the level of commercial organization of the target city would only be considered significant at a 0.1 significance level. As regards the effect of transferring data from one city to another, it is markedly significant in the additive approximation, and can be quantified as a decrease of 0.1 units in the MRR metric (Q7).

TABLE 1. Multiple linear regression coefficient results (additive model).

Coefficients	Estimate	Std. Error	t value Pr(>|t|)
Intercept	0.266310	0.010914	24.402 < 2e-16 ***
Source	0.023346	0.004066	5.741 1.77e-07 ***
Target	0.008087	0.004066	1.989 0.0503 .
Transferred	-0.101069	0.006841	-14.773 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fig. 7 plots the MRR as a function of the commercial spatial specialization of the source city and includes two different linear regressions: one for the case when the data come from the city itself, and the other for the case of transferred data. Although there is a slight change in slope between both regression lines, a multiplicative model for the dummy variable was not found to be significant (Q7).
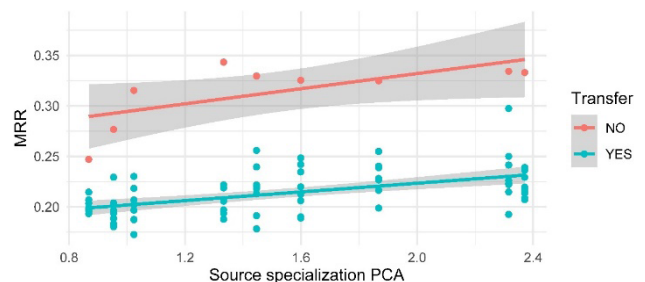

FIGURE 7. Individual linear regressions that have the commercial spatial organization of the source city (obtained through PCA) as input and the MRR obtained with the random forest approach (joint use of all quality indices) as output.

3) CONSENSUS APPROACHES WITH AND WITHOUT THE CITY'S OWN DATA

We adopted the consensus network of relationships approach (consensus network of the interactions between commercial categories) to study whether —and to what extent— the joint use of data from several cities is of interest to predict the commercial categories of retail stores from another city. In particular, to make predictions in a given city, the set of remaining cities was used as the source of information from which the consensus matrices by *Jensen*, *permutation* and *rewiring* were obtained. In this case, no prior division of the data into training and test was made: the interaction matrices were obtained on all available data and then condensed into the consensus matrices.

A hyperparameter in the consensus process is the threshold used on these matrices. In the evaluation procedure,

the threshold was optimized within the range [1], [6] (the upper limit was not set at eight, as this would imply keeping only the interactions of each commercial category with itself). Specifically, the optimization process was performed as follows. First, the consensus matrices were obtained for a given threshold. Then, they were used to calculate the quality indices of all the stores in the eight cities considered for their construction. Afterwards, the whole dataset — i.e., the true commercial categories and the corresponding quality indices of all shops in the eight cities under consideration— were divided into 70% training and 30% validation, doing random resampling five times. Finally, a random forest of 1000 trees was built on the training set, and the MRR metric was calculated on the test set. After this process was conducted over the whole range of thresholds, the threshold with the highest performance —i.e., the optimal threshold value— was selected. For such threshold value, a random forest model was trained on the dataset consisting of the eight cities and evaluated on the ninth city (the test city), whose data had remained hidden from the training process.

The results comparing the performance of the consensus approach with that of the previous alternatives are summarized in Fig. 8. In general, this procedure obtains intermediate results between using the best possible city as data source and using the city's own data (Q8).
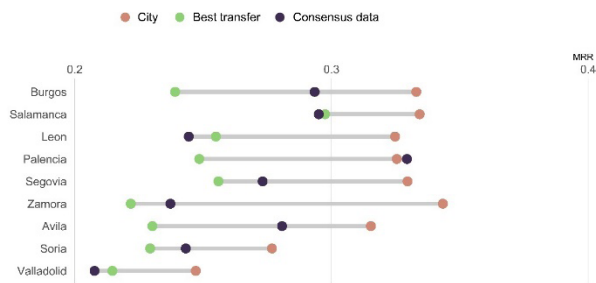
FIGURE 8. Comparison of the MRR values obtained for the different cities depending on the source of information (cities are sorted by empirical spatial specialization). Orange dots represent results obtained with data from the city itself, green dots are the results obtained with the data from the best possible source —other than the city itself— and purple dots represent the results of using consensus data from all the other cities (threshold optimized internally).

To complete this study, we explore the results obtained in the process of resampling and threshold validation (Fig. 9). These results reveal that, in general, the application of thresholds is not effective in predictive terms; a better strategy is to use all available information. A second relevant aspect is that the performance obtained with consensus approaches that include the city's own data in the training process is systematically higher. These findings suggest that consensus-based information fusion processes including all available information from other cities and the city's own data, can be an effective strategy to assess the quality of a site in terms of its balanced tenancy (Q8).
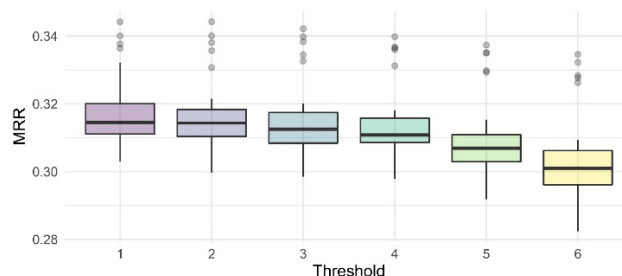
FIGURE 9. MRR as function of the threshold used in the consensus approach. Results were obtained with 1000-tree random forests trained on 70% of the data and tested on the remaining 30%; random sampling was performed five times. In contrast with figure 8, here the city's own data are part of the consensus set.

V. CONCLUSION AND LIMITATIONS

In retailing, location is a crucial aspect for the success of a business. The location problem is multicriteria, with the surrounding commercial ecosystem being a particularly relevant factor in the decision, because of the presence of competitors and the existence of complementary/substitute activities. Of the different metrics that address the problem of optimal retail store placement, Jensen's Quality Index is the one that enjoys the highest popularity in the literature. However, other alternative measures can be used. In this paper, we have empirically analyzed which quality index is best for evaluating the location of a store for a given commercial category.

Our results show that the indices Quality Rewiring Raw and Quality Jensen Raw, when used individually, capture the commercial structure of the city quite substantially (Q1, Q2). Nonetheless, the predictive performance improves significantly with the joint use of the six proposed quality indices combined in a classifier such as random forest (Q4). In addition, our results evidence a high predictability of the commercial typology of a given location based on its commercial ecosystem, i.e., excluding all other factors that also influence the location decision (Q2). From all the above, our findings suggest that the combined use of the six quality indices is markedly beneficial to assess the suitability of potential locations. Given that the highest cost of location selection is predominantly linked to the acquisition of the city's commercial information, once such data are obtained, the most straightforward approach would be to use of the six indices together. This way, one would exploit all the available information in different and complementary ways, thus increasing performance (Q3, Q4).

Given the costly nature of obtaining georeferenced commercial information for entire cities, we also analyzed whether the quality of the prediction is compromised if data from a given city are used to assess the suitability of locations in another city. In this regard, our results reveal a significant loss of predictive accuracy that we quantify statistically at around 0.1 points in MRR. Our analyses indicate that using cities with a high level of commercial organization as a data source has a beneficial effect on prediction (Q5, Q6, Q7).

Finally, this paper explores the use of consensus network techniques to fuse information on the retail structure of several cities to predict the commercial categories of stores in another city. Our results indicate that, in general, consensus approaches improve knowledge transfer with respect to individual transfer from one city to another. Nevertheless, in most cases the improvement attained cannot reach the predictive accuracy of models trained on the city's own data (Q8). Remarkably, consensus approaches that do include information from the city itself without applying any thresholds perform consistently better. Notwithstanding, they entail the cost of mining the information from all possible cities.

To conclude, given that this paper is one of the first to analyze the use of different quality indices for decision support in the retail location problem in a comparative, systematic, and multi-city way, it is also essential to highlight some of its possible limitations.

The first important aspect is related to the data sources used. In this regard, given that not all local stores are included in the Yellow Pages, it may be that not all business categories are equally represented in them.

Another relevant aspect that could improve the results and complete the study would be the inclusion of other spatial aspects (transportation facilities, non-commercial points of interest, etc.) beyond the strictly commercial relationships considered in this work.

In addition, even though our analyses focus on all urban areas in a large Spanish region such as Castile and Leon, the fact is that the nine of them are medium- and small-sized cities. In this vein, the results for larger cities —and therefore with a greater commercial organization— could be different. It is also important to highlight that these cities belong to the same region, so they may share some cultural aspects that could be reflected in their spatial and commercial organization. In this respect, a relevant open question would be to determine the influence of cultural aspects on the transferability of data between cities in different countries and/or continents.

From a machine learning perspective, we have used a classifier —random forest— to combine the quality indices together. The reasons behind the choice of random forest are that it gives good results on a variety of different problems without the need for high-level tuning, parameterization, or preprocessing. Nevertheless, it is possible that a more exhaustive model selection could improve our prediction results.

Finally, we would like to stress again that, even though retail patterns are a complex phenomenon with a strong strategic dimension, the quality indices explored in this work focus exclusively on the identification of commercial spatial interactions and their subsequent use for predictive purposes. In fact, once identified, we use those patterns to make accurate predictions without delving into their generating mechanism. Consequently, in addition to including other relevant factors —such as spatial aspects, cultural phenomena, demographic variables, etc.— it would also be of interest

to adopt alternative approaches —such as game-theoretical approximations— to address the strategic dimension of the problem and thus understand the underlying causes of the observed patterns [43]–[45].

REFERENCES

[1] B. R. Berman, J. R. Evans, and P. M. Chatterjee, *Retail Management. A Strategic Approach*. London, U.K.: Pearson, 2018.

[2] J. Zentes, D. Morschett, and H. Schramm-Klein, *Strategic Retail Management*. Wiesbaden, Germany: Gabler Verlag, 2012.

[3] C. Teller and T. Reutterer, "The evolving concept of retail attractiveness: What makes retail agglomerations attractive when customers shop at them?" *J. Retailing Consum. Services*, vol. 15, no. 3, pp. 127–143, May 2008.

[4] T. Hernández and D. Bennison, "The art and science of retail location decisions," *Int. J. Retail Distrib. Manage.*, vol. 28, no. 8, pp. 357–367, Sep. 2000.

[5] J. Reynolds and S. Wood, "Location decision making in retail firms: Evolution and challenge," *Int. J. Retail Distrib. Manage.*, vol. 38, nos. 11–12, pp. 828–845, Oct. 2010.

[6] S. Önüt, T. Efendigil, and S. S. Kara, "A combined fuzzy MCDM approach for selecting shopping center site: An example from Istanbul, Turkey," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 1973–1980, Mar. 2010.

[7] J. L. García, A. Alvarado, J. Blanco, E. Jiménez, A. A. Maldonado, and G. Cortés, "Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an analytic hierarchy process," *Comput. Electron. Agricult.*, vol. 100, pp. 60–69, Jan. 2014.

[8] N. Roig-Tierno, A. Baviera-Puig, J. Buitrago-Vera, and F. Mas-Verdu, "The retail site location decision process using GIS and the analytical hierarchy process," *Appl. Geogr.*, vol. 40, pp. 191–198, Jun. 2013.

[9] S. A. Shaikh, M. A. Memon, M. Prokop, and K.-S. Kim, "An AHP/TOPSIS-based approach for an optimal site selection of a commercial opening utilizing GeoSpatial data," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2020, pp. 295–302.

[10] S. Choudhury, P. Howladar, M. Majumder, and A. K. Saha, "Application of novel MCDM for location selection of surface water treatment plant," *IEEE Trans. Eng. Manag.*, early access, Sep. 20, 2019, doi: 10.1109/TEM.2019.2938907.

[11] V. Çoban, "Solar energy plant project selection with AHP decision-making method based on hesitant fuzzy linguistic evaluation," *Complex Intell. Syst.*, vol. 6, no. 3, pp. 507–529, Oct. 2020.

[12] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: A survey," *GeoInformatica*, vol. 19, no. 3, pp. 525–565, Jul. 2015.

[13] C. A. Hidalgo, P.-A. Balland, R. Boschma, M. Delgado, M. Feldman, K. Frenken, E. Glaeser, C. He, D. F. Kogler, A. Morrison, F. Neffke, D. Rigby, S. Stern, S. Zheng, and S. Zhu, "The principle of relatedness," in *Unifying Themes in Complex Systems IX*, vol. 2, A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, and Y. Bar-Yam, Eds. Cham, Switzerland: Springer, Jul. 2018, pp. 451–457.

[14] C. A. Hidalgo, E. Castañer, and A. Sevtsuk, "The amenity mix of urban neighborhoods," *Habitat Int.*, vol. 106, Dec. 2020, Art. no. 102205.

[15] R. M. Sánchez-Saiz, V. Ahedo, J. I. Santos, S. Gómez, and J. M. Galán, "Identification of robust retailing location patterns with complex network approaches," *Complex Intell. Syst.*, Oct. 2021, doi: 10.1007/s40747-021-00335-8.

[16] P. Jensen, "Network-based predictions of retail store commercial categories and optimal locations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, Sep. 2006, Art. no. 035101.

[17] P. Jensen, "Analyzing the localization of retail stores with complex systems tools," in *Advances in Intelligent Data Analysis VIII*, N. M. Adams, C. Robardet, A. Siebes, and J.-F. Boulicaut, Eds. Berlin, Germany: Springer, 2009, pp. 10–20.

[18] F. Wang, L. Chen, and W. Pan, "Where to place your next restaurant? Optimal restaurant placement via leveraging user-generated reviews," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 2371–2376.

[19] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo, "Geospotting: Mining online location-based services for optimal retail store placement," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 793–801.

[20] Y. Liu, L. Yao, B. Guo, N. Li, J. Zhang, J. Chen, D. Zhang, Y. Liu, Z. Yu, and S. Zhang, "DeepStore: An interaction-aware Wide&Deep model for store site recommendation with attentional spatial embeddings," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7319–7333, Aug. 2019.

[21] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Oct. 2014.

[22] Y. Wei, Y. Zheng, and Q. Yang, "Transfer knowledge between cities," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1905–1914.

[23] T. He, J. Bao, R. Li, S. Ruan, Y. Li, L. Song, H. He, and Y. Zheng, "What is the human mobility in a new city: Transfer mobility knowledge across cities," in *Proc. Web Conf.*, Apr. 2020, pp. 1355–1365.

[24] R. Jiang, X. Song, Z. Fan, T. Xia, Z. Wang, Q. Chen, Z. Cai, and R. Shibasaki, "Transfer urban human mobility via POI embedding over multiple cities," *ACM/IMS Trans. Data Sci.*, vol. 2, no. 1, pp. 1–26, Jan. 2021.

[25] SC Government of Canada, Tech. Rep. NAICS 2012-44-45, Mar. 2011. [Online]. Available: https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=118464&CVD=118465&CPV=44-45&CST=01012012&CLV=1&MLV=5

[26] P. Jensen, "Analyzing the localization of retail stores with complex systems tools," in *Proc. Int. Symp. Intell. Data Anal.*, in Lecture Notes in Computer Science, vol. 5772, 2009, pp. 10–20.

[27] S. Gómez, P. Jensen, and A. Arenas, "Analysis of community structure in networks of correlated data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 1, p. 16114, Jul. 2009.

[28] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, 2020.

[29] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, "A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities," *Inf. Fusion*, vol. 64, pp. 205–237, Dec. 2020.

[30] P. Probst, M. N. Wright, and A. Boulesteix, "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining Knowl. Discovery*, vol. 9, no. 3, pp. 1–15, May 2019.

[31] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Found. Trends Comput. Graph. Vis.*, vol. 7, nos. 2–3, pp. 81–227, Feb. 2012.

[32] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, 1st ed. New York, NY, USA: Springer, 2013.

[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.

[35] P. Wei, Z. Lu, and J. Song, "Variable importance analysis: A comprehensive review," *Rel. Eng. Syst. Saf.*, vol. 142, pp. 399–432, Oct. 2015.

[36] H. Ishwaran, "The effect of splitting on random forests," *Mach. Learn.*, vol. 99, no. 1, pp. 75–118, Apr. 2015.

[37] S. Nembrini, I. R. König, and M. N. Wright, "The revival of the Gini importance?" *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, Nov. 2018.

[38] M. N. Wright and A. Ziegler, "Ranger: A fast implementation of random forests for high dimensional data in C++ and R," *J. Stat. Softw.*, vol. 77, no. 1, pp. 1–17, 2017.

[39] Y. C. Dong, Q. Zha, H. Zhang, G. Kou, H. Fujita, F. Chiclana, and E. Herrera-Viedma, "Consensus reaching in social network group decision making: Research paradigms and challenges," *Knowl. Based Syst.*, vol. 162, pp. 3–13, Dec. 2018.

[40] A. Lancichinetti and S. Fortunato, "Consensus clustering in complex networks," *Sci. Rep.*, vol. 2, no. 1, p. 336, Dec. 2012.

[41] R. M. Sánchez-Saiz, V. Ahedo, J. I. Santos, S. Gómez, and J. M. Galán, "Dataset of the retailing location networks in the cities of Castile-Leon, Madrid and Barcelona," Dataset, Universidad de Burgos, Burgos, Spain, 2021. [Online]. Available: https://riubu.ubu.es/handle/10259/5585, doi: 10.36443/10259/5585.

[42] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?" in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7376, 2012, pp. 154–168.

[43] H. Konishi, "Concentration of competing retail stores," *J. Urban Econ.*, vol. 58, no. 3, pp. 488–512, Nov. 2005.

[44] V. Marianov and H. A. Eiselt, "On agglomeration in competitive location models," *Ann. Oper. Res.*, vol. 246, nos. 1–2, pp. 31–55, Nov. 2016.

[45] D. Kress and E. Pesch, "Sequential competitive location on networks," *Eur. J. Oper. Res.*, vol. 217, no. 3, pp. 483–499, Mar. 2012.

VIRGINIA AHEDO received the M.S. degree in industrial engineering and the Ph.D. degree in industrial technologies and civil engineering from the Universidad de Burgos, in 2015 and 2021, respectively. She is currently a Research Fellow and an Assistant Lecturer in complex systems and data analysis at the Universidad de Burgos. She has participated as a modeler and data analyst in several Spanish and international research projects that led to publications in renowned multidisciplinary journals. Her main research interests include complex systems modeling, network science, machine learning, operations research, and management engineering.

JOSÉ IGNACIO SANTOS received the B.S. degree in industrial engineering from the Universidad de Valladolid, Spain, the M.S. degree in information systems from the Escuela de Organización Industrial, Madrid, the M.S. degree in applied economics from the Universidad Nacional de Educación a Distancia, Spain, and the Ph.D. degree in industrial and civil engineering from the Universidad de Burgos, Spain. He is currently an Associate Professor in management engineering at the Universidad de Burgos. His main area of expertise is the modeling of complex systems. His research interests include application of methods and techniques for the study of complex systems, including agent-based modeling, complex network theory, and machine learning.

JOSÉ MANUEL GALÁN was born in Valladolid, Spain, in 1977. He received the M.S. degree in industrial engineering from the Universidad de Valladolid, Spain, in 2002, and the Ph.D. degree in industrial and civil engineering from the Universidad de Burgos, Spain, in 2007.

He has been a Full Professor in management engineering at the Universidad de Burgos, since 2019. He is the co-editor of three books and the author of more than 50 articles. His research interests include computational modeling and simulation, network theory and quality, and predictive modeling in industrial and management engineering. He is currently an Academic Editor of the journals *PLOS ONE*, *PeerJ Computer Science*, and *Complexity*, and a member of the Editorial Board of *Modelling*.

• • •