



# Spatial distribution of economic activities: a network approach

Federico Pablo-Martí<sup>1</sup> · Josep-Maria Arauzo-Carod<sup>2</sup> 

Received: 11 March 2015 / Accepted: 9 August 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

The aim of this paper is to analyse co-location patterns of manufactures and service industries at a microgeographic level using Spanish data from the Mercantile Register. Our approach allows us to analyse joint-location and co-location patterns of firms in different industries, and to overcome previous technical constraints in this type of analyses, partially thanks to using homogeneous cells instead of administrative units. This paper contributes to the empirical literature on industry location by developing a multisectorial co-location index computed by comparing differences between observed data about firms' location and randomly generated data. Multisectorial relationships are analyzed by transposing bilateral relations onto an n-dimensional space. Our results show that dispersed industries tend to locate jointly and that industries with lower joint-location patterns have spatial structures similar to those obtained through input–output relationships, suggesting weak role of co-location patterns as interindustry linkages are not the main location determinants.

**Keywords** Clusters · Microgeographic data · Network analysis · Firm location

**JEL Classification** R10 · R12 · R34

---

✉ Josep-Maria Arauzo-Carod  
josepmaria.arauzo@urv.cat  
Federico Pablo-Martí  
federico.pablo@uah.es

<sup>1</sup> Facultad de CC. Económicas y Empresariales, Universidad de Alcalá, Pl. de la Victoria, 2, 28802 Alcalá de Henares, Spain

<sup>2</sup> Department of Economics and CREIP - QURE, Universitat Rovira i Virgili, Av. Universitat, 1, 43204 Reus, Spain

## 1 Introduction

It is important to accurately analyse firm location and co-location patterns in order to better adapt public policies to decisions taken by firms in terms of where to locate and what to do at those locations. Concretely, these patterns may illustrate firms' preferences regarding the type of economic environment that they need, which may be taken into account by policy makers aiming to provide these firms with what they need at a local level. Given that location and co-colocation data may be used for policy purposes, public organisations need access to sufficient and reliable data about location strategies used by firms and industries, as clusters may play a key role in local and regional development (Porter 1998). These data requirements highlight the importance of accurate measurements, both in terms of the methodologies used to identify clusters and the spatial aggregation levels used in this identification.

In terms of methodologies used for clusters measurements, there currently are two main approaches:<sup>1</sup> Industrial Districts and Clusters. The first ones arising mainly from the Italian tradition of districts identified by Becattini in which supramunicipal areas specialize in some activities and generate a strong interaction among all agents involved in these industries (i.e., firms, workers, local banks, public administrations, etc.), and the second ones coming from Porters' contributions in which main interest was on competitive advantages of given areas. Empirical applications arising from them are not exactly of the same characteristics, being that whilst the former approach is more popular mainly due to the standardization of the Sforzi-ISTAT methodology, the latter is potentially easier to use because of its lower data requirements. On account of the advantages and disadvantages of both approaches, in this paper we will use the cluster one, due to both data availability and the shortcomings of the Sforzi-ISTAT methodology (Boix and Galletto 2008; Sforzi and Lorenzini 2002; ISTAT 1996).

In terms of spatial aggregation levels, most analyses of spatial distribution of economic activity have been carried out using extant administrative units (e.g. counties, regions, etc.), but unfortunately, these analyses suffer from the major shortcoming that administrative units vary greatly in size and shape, do not always coincide with real economic areas and are sometimes arbitrary. To deal with these constraints, recent research has started to use ad hoc units (usually smaller), as we do in this paper. These units are created by equally dividing a space into homogeneous squared cells which therefore do not exactly match any extant administrative unit.<sup>2</sup>

Accordingly, the methodology proposed in this paper aims to overcome previous methodological constraints, to obtain more precise results regarding firms' location and co-location patterns and to improve public policy tools aiming to attract new firms. Concretely, in this paper we identify manufacturing and service clusters (from

---

<sup>1</sup> Nevertheless, apart from these methodologies there is as well an extensive list of alternative approaches (see for instance Van Egeraat et al. 2016, for a review). The reason for the distinction between Clusters and Industrial Districts was to use widely known concepts (Industrial Districts and Clusters) in order to group contributions into those that consider some subjective determinants as trust among economic agents (Industrial Districts), and those that only consider some objective characteristics (Clusters).

<sup>2</sup> There are also other approaches such as those that use the stochastic Point Pattern methodology and those that use Neuronal Networks for pattern recognition. However, these approaches are not able to perform the multisectorial analyses that are the goal of this paper.

all sectors) in Spain and we classify these clusters according to the reasons behind clusterization processes; that is, (a) whether firms tend to locate together because they look for the same types of site (regardless of the industry to which they belong to) (i.e., joint-location), or (b) whether firms look to be located close to their supplier/customers in order to optimise commercial exchanges or, simply, for unknown reasons related with interindustry linkages (i.e., co-location). In specific terms, by dividing Spain into homogeneous cells we can check whether each industry follows a concentrated or dispersed pattern and subsequently, whether similar location exist for pairs of industries, so that clusters formed by different industries can also be identified. Finally, once we have identified these industry location patterns, we apply network theory to show the local microstructure of clusters. This paper contributes to the literature by identifying patterns of concentration of manufacturing and service industries in Spain using highly disaggregated geographical data (i.e. the space is divided into 10 km \* 10 km cells) that enables it to fully account for location and co-location patterns of firms in different industries. The methodology used in this paper allows us to overcome previous technical constraints when analysing this geographical concentration of economic activity, as well as the shortcomings caused by the definitions of administrative borders. Additionally, although this is an empirical oriented paper focused in Spain, its results are easily applicable into other institutional settings.<sup>3</sup>

To sum up, in order to properly analyse previously detailed joint-location and co-location issues, the main hypothesis to be tested in this paper is the following: “Whether industry location patterns can be explained in terms of relationships along the supply chain (i.e., co-location) or by common location determinants shared by those industries (i.e., joint-location)”.

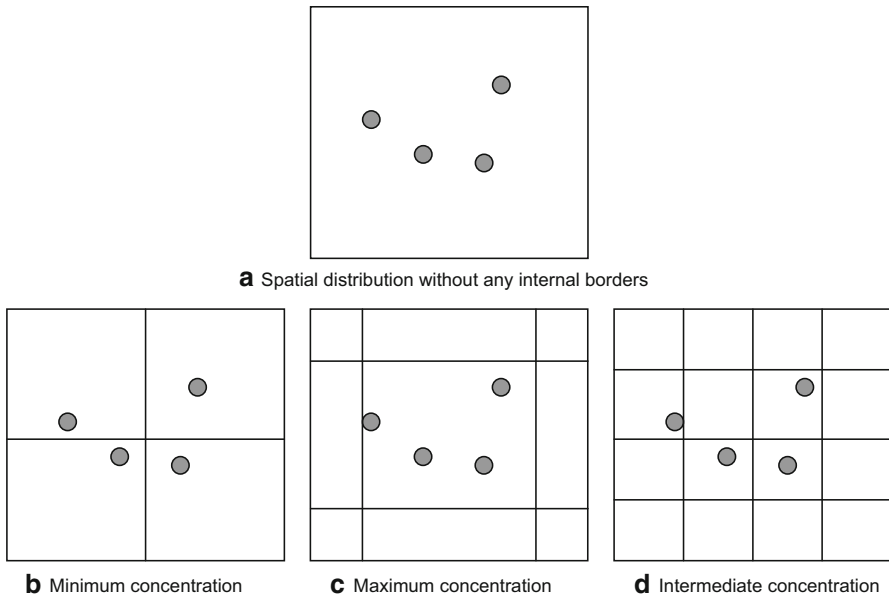
This paper is organised as follows. In the next section we review the main literature on the spatial distribution of economic activity and the spatial units used in empirical analysis. In the third section we explain the data set, we describe and analyse the spatial distribution of firms in Spain and we define the methodology. In the fourth section we present and discuss our main empirical results. In the final section we present our conclusions.

## 2 The spatial distribution of economic activity

If we review the empirical literature on the spatial distribution of economic activity, most researchers agree that there is a high level of concentration, as is shown by Duranton and Overman (2005), Devereaux et al. (2004), Maurel and Sédillot (1999) and Ellison and Glaeser (1997), among others. The Spanish case is roughly the same (i.e., firms tend to cluster in a few sites), as scholars such as Paluzie et al. (2004) and Viladecans (2004) have demonstrated. This concentration pattern is usually explained in terms of agglomeration economies (Marshall 1890), but additional knowledge is needed to determine the motives behind them.

---

<sup>3</sup> In this paper we carry out an introductory analysis from a national (aggregated) point of view, and although an alternative urban approach is also reasonable, we have left urban issues for further contributions.



**Fig. 1** Modifiable area unit problem (MAUP). *Source:* Arbia (2001)

Firms therefore aim to be located both close to similar firms (e.g. firms from the same industry) and to different firms (e.g. firms from another industry). However, although firms look for neighbours, not all neighbours are equally useful, and some could even be useless and harmful. This is why firms sometimes look to be located close to other firms with which they are vertically integrated, because they need to have close linkages with their providers/suppliers. Spatial proximity<sup>4</sup> therefore appears to be a good enough argument for sharing the same location. There are additional reasons that explain why certain types of firms are more likely to be located in the same area than other types of firms. Even if they belong to different industries and have different characteristics, they share the need for specific territorial inputs that push them to the sites where those inputs are available (e.g. skilled human capital, access to main markets, etc.).

Once having discussed firms' location preferences it is important to properly measure them in order to avoid spatial bias, and it is at this point that the *Modifiable Area Unit Problem* (MAUP) appears.<sup>5</sup> Arbia (2001) provides an excellent example of this problem: he portrays a hypothetical distribution of firm location (Fig. 1) in which there are four firms inside the spatial area under analysis (Fig. 1a). Arbia (2001) shows that depending on how spatial borders are designed, this location could result in a minimum concentration pattern (Fig. 1b), in a maximum concentration pattern (Fig. 1c) or an intermediate concentration pattern (Fig. 1d).

Figure 1 shows that spatial aggregation really matters, but in the past this has not been a major concern in empirical analysis mainly due to the lack of sufficiently

<sup>4</sup> In this paper, "spatial proximity" means to be located in the same cell.

<sup>5</sup> See Openshaw and Taylor (1979) for a detailed analysis and Wrigley (1995) for a further review.

disaggregated data;<sup>6</sup> recently, however, researchers have started to gain access to dramatically improved datasets with extended spatial disaggregation. This is the case with our data set, which contains accurate individual information about the location of firms, thus allowing us to technically address previous shortcomings and to freely decide on the way in which a space is disaggregated, regardless of where the administrative (and usually arbitrary) boundaries are. This is particularly important because “(...) *any statistical measure based on spatial aggregates is sensitive to the scale and aggregation problems*” (Arbia 2001, p. 414). As Duranton and Overman (2005, p. 1079) point out, “(...) *any good measure of localization must avoid these aggregation problems*”. Unfortunately, MAUP is not the only problem about measuring spatial distribution of economic activity, as Bickenbach et al. (2013) point out when discussing methodological challenges in order to obtain feasible indicators. In this sense, other issues as industry aggregation and time-varying measures are also relevant.

Given these considerations, our goal is to empirically assess the location patterns of both manufacturing and services firms in Spain, and try to determine if these firms tend to locate close to other firms from the same industry, to firms with close industry linkages (e.g. providers and suppliers, which implies co-location) or to firms that share the same location requirements (e.g. accessibility to inputs, labour and infrastructures, which implies joint-location). Previous contributions have taken a similar approach, as Duranton and Overman (2008, 2005) when using microgeographic (postcode level) data from the Annual Census of Production in the UK to analyse manufacturers. They computed Euclidean distances between every pair of entering establishments, and compared those results with the extant distances between incumbent establishments in order to check for any possible similarities between the location patterns of entrants and incumbent establishments.

The contribution by Duranton and Overman (2008) is the one that guided our approach as they tried to identify two specific situations: the first occurs when firms from different industries locate in the same areas (*joint-localization*);<sup>7</sup> the second occurs when firms from different industries also locate in the same areas because there are certain inter-industry linkages between them (*co-location*). This distinction is extremely important, because it allows us to better understand the location process and therefore to advise firms as to which is the best type of environment (e.g. in terms of spatial characteristics, firms, specialised services, inter-industry linkages, and so on). Following on from this distinction, the term *joint-localization* means that there are some firms (from different industries) that share the same spatial requirements (i.e. they need access to the same type of inputs, services, infrastructures, etc.), which means they tend to locate in the same areas. However, *co-location* is very different, as it implies that firms need to be close to their suppliers/clients, with the result that firms from different industries will cluster together. Even if final decisions in terms of location of firms may be (apparently) the same, economic and policy implications

<sup>6</sup> The influence of spatial units on the analysis of firm location has been studied in Arauzo-Carod and Manjón-Antolín (2004) and Arauzo-Carod (2008). See Olsen (2002) for a discussion of the units to be used in geographical economics.

<sup>7</sup> Although Duranton and Overman (2008) use the terms joint-localization and co-location, for the sake of simplicity, we will refer to them simply as joint-location and co-location.

of joint-location and co-location are quite different and, consequently, deserve to be analysed.

### 3 Data and methodology

#### 3.1 Data

Our data set refers to 2006 and comprises Spanish manufacturing, services and agriculture firms.<sup>8</sup> The source of this database is the SABI (*Sistema de Análisis de Balances Ibéricos*), which uses data from the Mercantile Register, including balance sheets and income and expenditure accounts, and has been extensively used for locational purposes.<sup>9</sup> For each firm, we also know the number of employees, the industry to which it belongs (the three-digit NACE code), and its sales and assets, among other variables. We also have detailed information about the firm's geographical location, information which is particularly relevant for the purposes of this paper. Nevertheless, the SABI dataset also has two important shortcomings. The first concerns the sample. Although the number of firms is very high (e.g. 581,712 firms for the 2007 edition), microfirms and self-employed individuals are not taken into account, despite that fact that it is reasonable to assume that the spatial distribution of those activities is similar to that of the firms that are included. The second concerns the nature of the units; SABI only covers firms, not establishments,<sup>10</sup> and the latter are more appropriate for analyzing the spatial distribution of economic activity. In any case, since SABI covers most of the economic activity carried out in Spain, these disadvantages are easily overcome.<sup>11</sup> Additionally, in a similar way to Blöch et al. (2011) and Titze et al. (2011) we use the Spanish input–output tables to determine whether the geographical proximity of firms belonging to different industries can be explained by inter-industry linkages (i.e. supply chains across industries) or by the need to access similar areas.<sup>12</sup>

#### 3.2 Identification of joint-location and co-location patterns

The methodology proposed in this paper (X co-location index) partially follows the contributions of Duranton and Overman (2005), Brenner (2004, 2006) and Ellison and Glaeser (1997), but also improves on those approaches in several ways.

<sup>8</sup> It is worth noting that the data set refers to firms (not establishments) and that each firm could have more than one establishment. Nevertheless, not taking into account multi-plant firms is not a major problem, as only around 1% of firms in Spain are multi-plant (see Jofre-Monseny et al. 2015, for details).

<sup>9</sup> Albert et al. (2012) perform a quite similar analysis using the same dataset, which has also been used by other scholars including Acosta et al. (2011) and Boix and Galletto (2008).

<sup>10</sup> Other alternative statistical sources such as the *Censo de Locales* (INE) are not currently updated, although having firms as observation units instead of establishments also provides useful information since it highlights the role of municipalities when firms are choosing where to locate their headquarters.

<sup>11</sup> There are alternative datasets such as DIRCE (INE) but the geographical location of the firms is highly spatially aggregated.

<sup>12</sup> The Spanish Input Output matrix is from 2000 and covers all economic industries at 2 digits of NACE classification (Spanish National Statistics Institute).

First, we divide the space into homogeneous 10 km \* 10 km cells. This size was decided in order to avoid computational constraints (smaller sizes entailed a huge increase in computational capacity in order to deal with a larger amount of spatial units) and to obtain a cell big enough to have several firms from different industries. Although alternative sizes were also feasible<sup>13</sup> we considered that the selected size was appropriate both from the computational (the calculations allowed) and the economic (it enabled us to consider small and homogeneous areas regardless of the administrative aggregations) point of view.<sup>14</sup> This is quite different from the strategies followed by other researchers, who have used administrative units (Brenner 2004, 2006; Ellison and Glaeser 1997) or the distance between firms (Duranton and Overman 2005). As López-Bazo (2006) points out, these strategies have several shortcomings, which are: the inability to take into account the precise location of firms, the limitations resulting from the specific administrative aggregation levels in each country, the difficulties with comparing the results obtained for different levels of administrative aggregation, the non-economic nature of those administrative units, the size differences across administrative units,<sup>15</sup> the modifiable areal unit problem (MAUP), which can create spurious correlations between variables, and the fact that the administrative divisions do not take into account neighbour effects across units. Using homogeneous cells therefore permits researchers to analyse both large areas such as countries and smaller areas such as cities. Nevertheless, this approach has some shortcomings, of which the most important is the fact that we are only considering the areas where firms are located without taking into account either the size or the number of firms located there.<sup>16</sup> We could partially resolve this disadvantage by reducing the cell size to a certain extent,<sup>17</sup> although if the cell were so small that it contained only one firm, it would not be possible to identify any agglomeration pattern. Consequently, our approach allows us to compare the spatial distribution of firms with random simulations of those distributions to check whether there are any concentrations in the former. There are other alternatives, such as kernel-smoothing (see Barlet et al. 2009; Duranton and Overman 2005 and Silverman 1986), but using kernels is not feasible when using units such as 10 km \* 10 km cells like those used in this paper; i.e. kernels could be a good strategy for urban areas because they allow smooth contiguous areas (i.e., between cells), but smoothing has already been performed with the cells for larger areas, and additional smoothing could homogenise non-adjacent and heterogeneous areas. We therefore decided not to use kernel-smoothing.

<sup>13</sup> We tested several alternative cell sizes (5 km \* 5 km, 20 km \* 20 km, 25 km \* 25 km, and 50 km \* 50 km) but the results were essentially unchanged.

<sup>14</sup> The cells' size is obviously related to the total area analysed. A similar approach is used by Larsson and Öner (2014) in their analysis of Swedish retail markets in metropolitan areas. Nevertheless, as they focus on smaller areas the cells also become smaller, at 250 m \* 250 m.

<sup>15</sup> By way of an example, municipality sizes in Spain range from 5 to 1000 km<sup>2</sup>.

<sup>16</sup> This is a (simple) starting point that could be easily improved by taking into account the intensity of land use, considering certain indicators such as the number of jobs, the production value and levels of sales, among others. This should allow the expected results to be compared with real results to determine the number of jobs, for example. However, there are also some (potential) limitations regarding data accuracy.

<sup>17</sup> Nevertheless, the main problems concern the heterogeneity of firm size, so it seems that a better solution would be to use the size of firms (e.g. the number of employees) rather than just the number (or the existence) of firms.

Second, we create industry specific maps that depart from the firms' georeferenced data. This approach is similar to those used by Duranton and Overman (2005); however, their maps consider distances between firms, whereas we focus on the areas occupied by firms. Although our dataset (SABI) provides data at a 3-digit level, we have decided to use data at a 2-digit level. Using alternative measures to the NACE code (i.e. in terms of the number of digits to be considered) may obviously bias the results, as more aggregated classifications may imply connections with almost all industries and more disaggregated ones may imply a limited number of connections (see Hoffmann et al. 2016, for a discussion). In order to overcome this potential shortcoming, we decided to use data from the 2-digit NACE code as this level enables better identification of industry specificities and use of reliable sectorial data, to focus on smaller areas, and to maintain enough interindustry linkages.

Third, we create multiple random industry specific maps with two conditions: *i*) the total number of firms in each industry remains constant and *ii*) the total number of firms in each cell remains constant.<sup>18</sup> We thereby compare the observed spatial distribution of firms at industry level with that one expected if firms would locate using same location criteria than all industries. As a result, if the real data shows a cell with only one firm, our simulations will also show this cell with one firm, although the industry will appear as a random variable depending on industry distribution.

Fourth, we compare at industry level the actual number of cells with firms (according to real data) with the expected number of cells with firms, and we obtain a concentration index<sup>19</sup> similar to that of Ellison and Glaeser (1997), but with some important differences. In particular, our methodology does not focus on agglomeration issues, which allows us to analyse industry distribution. Furthermore, whereas our index is centred at 1 (values below 1 indicate concentration and values over 1 indicate dispersion), Ellison and Glaeser's (1997) index ranges between zero and infinite, which means that they arbitrarily define the concentration threshold.

Fifth, we generalize our approach to several industries. Methodologically, this is quite similar to the approach with only one sector, but now this is no longer a concentration index but instead a co-location index, as we analyse whether or not a group of industries tends to locate together.

Sixth, network analysis is used to show the local microstructure of industries in a similar way as what is called "equivalence" in social network analysis (e.g. whether industries A and B have the same linkages, with industries C, D, etc.).<sup>20</sup>

<sup>18</sup> This latter requirement implies that firms localise randomly inside "occupied" cells (i.e. areas where real firms are located) as suggested by Duranton and Overman (2008). This approach means that firms are expected to be located only in those places that are available for economic activity (as the real data shows). Unfortunately, a major shortcoming of this approach is that it assumes that firms could be located elsewhere with other firms, regardless of the industry they are involved in, which is not as realistic (especially at a 2/3-digit level). An extension of this study (and a possible solution to this shortcoming) would be to regard manufacturing, services and agricultural firms as being located with other firms from the fields of manufacturing, services and agriculture, respectively.

<sup>19</sup> A literature review of classical indexes used in regional analyses can be found in Isard (1960), whilst notable recent approaches to the spatial distribution of industries are those by Hoffmann et al. (2016), Bickenbach et al. (2013) or Mulligan and Schmidt (2005), among others.

<sup>20</sup> The idea about "equivalence" was suggested by a reviewer of this paper.



To sum up, we identify (i) how many cells ( $X$ ) contain firms from industry  $y$  (i.e. this is the “real” spatial distribution of firms); (ii) the expected number of cells (the mean of simulations) where firms from industry  $y$  should appear if they were randomly spatially distributed (according to the total number of firms in each industry); (iii) and a co-location index ( $X$  co-location index) that relates these measurements to each other. Concretely, we are especially interested in that  $X$  co-location index:

$$X \text{ co-location index} = \frac{\text{Observed number of cells occupied by firms of } X \text{ industries}}{\text{Mean of simulated number of cells occupied by firms of } X \text{ industries}}$$

This  $X$  co-location index has a structure similar to z-score but we prefer  $X$  co-location index because it highlights the divergence with respect to non-normalized data, and that this is of more interest from an economic point of view. Significance of divergence should be approached with caution because baseline distribution is not theoretical and is only a particular case referred to a specific place and time (according to that absolute divergence seems to be more appropriate in order to make comparisons).

This methodology can complement previous approaches based on distribution comparisons (Brenner 2004, 2006; Ellison and Glaeser 1997) and on distance distributions (Duranton and Overman 2005) and thus enable industry to better understand the determinants behind the spatial distribution of firms.

## 4 Main results

Our main results show that the location of firms is driven by several industry-specific determinants (i.e. whether the firm belongs to a manufacturing or services activity or to a specific industry within these sectors) and also by their technological level. In some vertically integrated industries (e.g. the automobile industry), reducing distance to providers/suppliers is a key issue, whereas other types of industries do not need such spatial proximity. Additionally, industries also differ in terms of whether they are concentrated in a small number of areas or on the contrary, dispersed over a large number of areas.

Table 1 illustrates the expected spatial distributions of firms across regular cells<sup>21</sup> (according to the number of firms in each industry) and the real (observed) spatial distribution of such firms. In this table  $X = 1$  because it refers to single industries (it would be  $X = 2$  in case of pairs of industries,  $X = 3$  in case of trios of industries, and so on). This index can be understood in the following way: if the  $X$  co-location index is significantly  $< 1$ , this means that the industry  $y$  appears in fewer cells than expected (i.e. this industry is spatially concentrated in a smaller number of cells); and if the  $X$  co-location index significantly  $> 1$ , this means that the industry  $y$  appears in more cells than expected (according to the mean of random distributions), which means that this industry is spatially dispersed. This indicates that there is a certain location pattern that should be analyzed in order to determine whether or not it is a cluster (i.e. whether or not firms from industry  $y$  tend to locate together). It is important to note that this

<sup>21</sup> These regular cells have an area of 100 km<sup>2</sup> (10 km \* 10 km).

**Table 1** Concentration patterns of firms at a single industry level (X1 co-location index). *Source:* own calculations

Code	Industry	X	Mean	SD	X1 co-location index	X - 2S	X + 2S	Pattern
22	Financial intermediation	882	1480.11	17.6811804	0.59590166	1444.74764	1515.47236	Concentrated
6	Paper and publishing	947	1494.58	17.9619013	0.63362282	1458.6562	1530.5038	Concentrated
13	Office machinery, computers and medical equipment, precision and optical instruments	324	502.86	13.3553001	0.64431452	476.1494	529.5706	Concentrated
26	Education	790	1209.17	17.5580164	0.65334072	1174.05397	1244.28603	Concentrated
14	Electrical machinery and apparatus	520	782.36	14.6890463	0.66465566	752.981907	811.738093	Concentrated
24	Business services	1360	1979.03	21.1557261	0.68720535	1936.71855	2021.34145	Concentrated
23	Real estate activities	1957	2803.29	18.8970069	0.69810829	2765.49599	2841.08401	Concentrated
28	Other services	1375	1819.52	21.5638493	0.75569381	1776.3923	1862.6477	Concentrated
12	Machinery and equipment	820	1076	17.2533118	0.76208178	1041.49338	1110.50662	Concentrated
4	Textiles, leather clothes and shoes	1169	1523.26	17.384319	0.76743301	1488.49136	1558.02864	Concentrated
27	Health and veterinary activities, social services	1122	1458.21	20.5029168	0.7694365	1417.20417	1499.21583	Concentrated
8	Rubber and plastic products	698	903.5	19.1498609	0.77255119	865.200278	941.799722	Concentrated
25	Public administration	141	179.3	7.24812759	0.78639152	164.803745	193.796255	Concentrated
7	Chemical products	734	837.17	14.8691634	0.87676338	807.431673	866.908327	Concentrated
10	Basic metals	567	629.55	16.7460986	0.90064332	596.057803	663.042197	Concentrated
15	Transport materials	668	726.47	16.8111645	0.91951491	692.847671	760.092329	Concentrated
19	Trade and repair	2888	3035.78	16.6336521	0.95132058	3002.5127	3069.0473	Concentrated

Table 1 continued

Code	Industry	X	Mean	SD	X1 co-location index	X - 2S	X + 2S	Pattern
16	Recycling	349	359.69	9.90020406	0.97027996	339.889592	379.490408	Undetermined
11	Fabricated metal products	1682	1701.7	19.8267751	0.98842334	1662.04645	1741.35355	Undetermined
21	Transport and communications	2090	2034.14	19.9479221	1.02746124	1994.24416	2074.03584	Dispersed
17	Construction	2706	2585.57	21.9674944	1.04657774	2541.63501	2629.50499	Dispersed
20	Hotels and restaurants	2238	2136.5	20.4181045	1.04750761	2095.66379	2177.33621	Dispersed
18	Electricity and water distribution	795	739.43	15.2674838	1.07515248	708.895032	769.964968	Dispersed
5	Wood, furniture and other manufactures	1734	1610.89	20.5956232	1.07642359	1569.69875	1652.08125	Dispersed
9	Non-metallic mineral products	1297	1125.88	18.1566027	1.15198778	1089.56679	1162.19321	Dispersed
2	Extractive activities	1152	823.16	15.7015858	1.39948491	791.756828	854.563172	Dispersed
1	Agriculture and fishing	2409	1691.54	20.5354682	1.42414604	1650.46906	1732.61094	Dispersed
3	Food, beverages and tobacco	2236	1540.31	20.5001577	1.45165584	1499.30968	1581.31032	Dispersed

X - 2S equals X minus 2 standard deviations and X + 2S equals X plus 2 standard deviations

index has asynthetic properties, given that with a small number of cells the variability could be quite important and the results could consequently be easily biased.

On a technological level, it seems that the lower the technological level of the industry, the higher the spatial dispersion (Table 1). High-tech firms therefore tend to be more spatially concentrated than low-tech firms.<sup>22</sup> This is logical, since the markets and resources of such firms tend to be concentrated in a few areas, which means there is no logical reason for a dispersion pattern.

Our results for the differences between manufacturing and services (Table 1) are even clearer than those of previous studies, and show that whereas most services activities show high concentration levels (e.g. Financial intermediation, Education, Business services, etc.), manufacturing activities are more dispersed (Agriculture and fishing, Food, beverages and tobacco, etc.). These results reflect the spatial distribution of population and economic activity and the production and distribution requirements of manufacturing and services. In specific terms, most services need face-to-face interactions and their location decisions are therefore strongly motivated by the previous locations of their customers (both firms and individuals) who typically tend to concentrate in a small number of locations. By contrast, manufacturers can transport their goods easily, which means that those interactions are not essential and that these firms can locate elsewhere. To sum up, Table 1 provides empirical evidence at a single-industry level for how industry specificities (i.e. manufacturing vs. services and high-tech vs. low-tech) helps us to understand firms' location patterns.

However, this situation becomes more complicated if we take into account the location patterns of more than one industry. The next step is therefore to check for the existence and extent of industry clusters by checking whether pairs of industries (or groups of three or four industries) tend to be located close to each other (X2 co-location index). Table 2 summarises the main findings, and shows a selection of all the possible combinations of pairs of industries, but given that there are 378 possible combinations of industry pairs, here we only show the results for the 10 pairs with the lowest X2 co-location index values, for the 10 pairs with the highest index values, and for 10 intermediate cases. The data from this table is similar to that in Table 1 and for each of these pairs we include: the codes of industries  $y$  and  $i$  respectively; the number of times ( $X$ ) that firms from industry  $y$  and industry  $i$  appear together inside the same cell; the expected number of times (Mean) that firms from industry  $y$  and industry  $i$  should appear together inside the same cell if they were randomly spatially distributed (according to the total number of firms for both industries); and the X2 co-location index that relates these measurements to one another (i.e.,  $\text{Index} = X/\text{Mean}$  of simulations). This Index can be understood in the following way: if the X2 co-location index is significantly  $< 1$ , this means that this industry combination ( $y$  and  $i$ ) appears fewer times than expected and (for reasons that we will analyze later) this pair of firms tends not to be located in the same areas; and if the X2 co-location index is significantly  $> 1$ , this means that this industry combination appears more times than expected, so this pair of industries locates in the same areas (they cluster together).

<sup>22</sup> As an example, indices of high-tech industries such as Office machinery, computers and medical equipment, Precision and optical instruments (0.644) and Electrical machinery and apparatus (0.664) are clearly lower than those of some low-tech industries such as Food, beverages and tobacco (1.452) and Agriculture and fishing (1.424).

**Table 2** Concentration patterns of firms for pairs of industries. *Source:* own calculations

Industry code y	Industry code i	X	Mean	SD	X2 co-location index	X - 2S	X + 2S	Attraction versus Repulsion
<i>The 10 pairs of industries with the lowest X2 co-location index values</i>								
4	22	639	1092.83	12.749	0.585	1067.333	1118.327	Repulsion
22	23	835	1424.44	16.580	0.586	1391.280	1457.600	Repulsion
14	22	391	662.3	12.630	0.590	637.039	687.561	Repulsion
22	24	748	1259.35	15.338	0.594	1228.675	1290.025	Repulsion
14	26	361	606.94	10.773	0.595	585.394	628.486	Repulsion
6	22	651	1080.17	14.169	0.603	1051.833	1108.507	Repulsion
12	26	464	769.81	13.134	0.603	743.542	796.078	Repulsion
4	26	574	948.17	14.318	0.605	919.534	976.806	Repulsion
19	22	878	1449.2	17.590	0.606	1414.021	1484.379	Repulsion
22	26	569	934.68	13.485	0.609	907.709	961.651	Repulsion
<i>The 10 pairs of industries with intermediate X2 co-location index values</i>								
3	6	874	1108.37	14.899	0.789	1078.572	1138.168	-
19	25	141	178.69	7.340	0.789	164.010	193.370	-
4	11	943	1190.99	15.325	0.792	1160.340	1221.640	-
9	22	703	887.45	13.918	0.792	859.614	915.286	-
5	25	134	168.98	7.589	0.793	153.802	184.158	-
20	25	139	174.6	7.549	0.796	159.502	189.698	-
17	25	141	177.03	7.259	0.796	162.512	191.548	-

Table 2 continued

Industry code y	Industry code i	X	Mean	SD	X2 co-location index	X - 2S	X + 2S	Attraction versus Repulsion
11	27	918	1152.4	16.483	0.797	1119.435	1185.365	-
21	28	1190	1493.26	16.838	0.797	1459.583	1526.937	-
12	18	440	549.95	11.876	0.800	526.198	573.702	-
<i>The 10 pairs of industries with the highest X2 co-location index values</i>								
1	17	2013	1572.54	18.303	1.280	1535.934	1609.146	Attraction
1	19	2120	1648.27	20.232	1.286	1607.805	1688.735	Attraction
2	9	785	609.29	11.853	1.288	585.584	632.996	Attraction
2	17	1059	799.41	15.180	1.325	769.049	829.771	Attraction
2	19	1100	815.4	15.300	1.349	784.801	845.999	Attraction
3	17	1957	1446.37	19.204	1.353	1407.963	1484.777	Attraction
3	19	2042	1506.4	19.019	1.356	1468.361	1544.439	Attraction
1	2	990	723.86	13.682	1.368	696.496	751.224	Attraction
2	3	985	700.7	13.457	1.406	673.787	727.613	Attraction
1	3	1773	1193.15	15.684	1.486	1161.782	1224.518	Attraction

High values of the X2 co-location index therefore suggest that there is a cluster (the two industries locate together because they have strong inter-industry linkages) or it could be an example of co-location (the two industries locate together because they need the same type of economic environment but do not have any kind of inter-industry relationship). The procedure to be followed is first to identify such location patterns, and second to distinguish between the aforementioned proximity explanations.

The 28 industries can be distributed into 378 potential pairs. Most of these (324) show a X2 co-location index  $< 1$ , which means that these pairs of industries appear fewer times than expected. In contrast, only 54 pairs show a X2 co-location index  $> 1$ , which indicates a cluster or a significant co-location. In any case, it is important to note that dispersed industries (with a high X1 co-location index) tend to locate together, while the concentrated industries (those with lower X1 co-location index values) tend not to be located together. Consequently, concentration/dispersion patterns should also be taken into account when analysing (potential) cluster sites. Interestingly, highly co-located pairs of industries very often include mature manufacturing activities as Agriculture and fishing; Food, beverages and tobacco; Extraction activities; Wood, furniture and other manufacturing activities; Non-metallic mineral products; Construction; and Trade and repair. Meanwhile, low co-location industries include regularly service skilled industries as Financial intermediation and Education.

As we have assumed that interindustry linkages may be potential determinants of strong co-location levels, it is important to analyse data from input–output tables in order to identify the (potential) interindustry linkages behind co-location patterns. Table 3 shows the top 10 and bottom 10 pairs of industries in terms of the X2 co-location index, but including inter-industry linkages (i.e. through input–output tables) for intermediate consumption between pairs of industries in order to distinguish between a co-location and a co-location pattern. We assume that if two pairs of industries are linked by such inter-industry intermediate consumption and they are closely located, then there is a co-location pattern, whilst if there is no such relationship but they are also closely located, then their location patterns can be explained in terms of joint-location.

Our results show no clear pattern in terms of inter-industry linkages as inter-industry linkages are randomly found between pairs of industries. This type of data therefore cannot be used to explain firm co-location behaviour (or the absence thereof) in terms of such linkages; that is, it seems that intermediate consumption is not driving co-location patterns. In particular, the bottom of the table shows that some pairs of industries have important linkages (e.g. 34.51% of the intermediate consumption of Agriculture and fishing comes from Food, beverages and tobacco, and 15.84% of the intermediate goods sold by this industry go to Agriculture and fishing) whereas others do not have such linkages or that those linkages are much weaker (e.g. Extractive activities and Food, beverages and tobacco, Extractive activities and Non-metallic mineral products, etc.). Finally, the top of the table presents a similar picture: while some of the pairs of industries achieve important inter-industry linkages (e.g. Financial intermediation and Business services, Trade and repair and Financial intermediation, etc.) others are less well linked (e.g. Machinery and equipment and Education, Electrical machinery and apparatus and Education, etc.). We should consequently look for other joint-location explanations apart from classical inter-industry linkages.

**Table 3** Inter-industry linkages according to the X2 co-location index. *Source:* Spanish input–output table (National Statistics Institute) and own calculations

Industry x buys (a)	Industry y sells (b)	Purchases x to y (c)	Total purchases x (d)	Total sales y (e)	(c/d) (%)	(c/e) (%)	X2 co-location index
<i>The 10 industries with the lowest X2 co-location index values</i>							
4	22	258.10	12,305.50	16,868.80	2.10	1.53	0.585
22	23	635.90	8520.00	18,743.00	7.46	3.39	0.586
14	22	135.50	8143.10	16,868.80	1.66	0.80	0.590
22	24	3869.60	8520.00	59,803.40	45.42	6.47	0.594
14	26	20.60	8143.10	1597.20	0.25	1.29	0.595
6	22	220.40	11,111.00	16,868.80	1.98	1.31	0.603
12	26	17.20	8694.10	1597.20	0.20	1.08	0.603
4	26	58.80	12,305.50	1597.20	0.48	3.68	0.605
19	22	1838.10	40,752.90	16,868.80	4.51	10.90	0.606
22	26	35.90	8520.00	1597.20	0.42	2.25	0.609
<i>The 10 industries with the highest X2 co-location index values</i>							
1	17	212.40	13,773.00	43,515.60	1.54	0.49	1.280
1	19	1675.00	13,773.00	34,413.70	12.16	4.87	1.286
2	9	29.30	6282.90	16,546.10	0.47	0.18	1.288
2	17	112.90	6282.90	43,515.60	1.80	0.26	1.325
2	19	208.40	6282.90	34,413.70	3.32	0.61	1.349
3	17	225.90	45,829.90	43,515.60	0.49	0.52	1.353
3	19	2504.90	45,829.90	34,413.70	5.47	7.28	1.356
1	2	505.60	13,773.00	13,841.30	3.67	3.65	1.368
2	3	0.40	6282.90	30,001.60	0.01	0.00	1.406
1	3	4752.60	13,773.00	30,001.60	34.51	15.84	1.486

(a) and (b) are industry codes and (c), (d) and (e) are millions of euros



Tables 4 and 5 summarise the main results for co-location patterns from the data portrayed in Tables 2 and 3 about pairs of industries. Specifically, Table 4 shows linkages of each industry in terms of the X2 co-location index. The first three columns (1, 2 and 3) have an X2 co-location index lower than 1 and show (for each industry) the number of industries that do not tend to co-locate. This behaviour ranges from strong aversion (row 1) to neutrality (row 3). Finally, row 4 indicates the number of industries for which there is joint-location evidence.

Table 4 shows that the main industries with a strong tendency not to co-locate are Office machinery, Computers and medical equipment, precision and optical instruments, Electrical machinery and apparatus, Financial intermediation and Education. As these are all high-tech industries, it seems that inter-industry spillovers are of little importance for these knowledge-intensive industries and therefore do not drive firms to locate in nearby areas. This is a surprising result, as these innovative industries need to interact with other firms, although our data suggests that these interactions are mainly intra-industrial rather than inter-industrial. The industries that show low joint-location levels<sup>23</sup> are quite heterogeneous and include both medium and high-tech activities and finally, there is a small number of low-tech industries with strong levels of co-location (i.e., Agriculture and fishing; Extractive activities; Food, beverages and tobacco and Non-metallic mineral products).

Table 5 summarises information regarding potential industry co-colocation into four types of industrial co-location relationships (i.e., according to X2 co-location index). In specific terms, types 1, 2 and 3 include combinations of two industries which do not tend to be located together, whilst type 4 is for co-located pairs of industries. These results show that co-location is particularly important for some mature manufacturing industries (Non-metallic mineral products and Fabricated metal products), Construction and most low and medium-tech services, whilst it is rare for most high-tech services.

In previous sections, we have presented empirical evidence for concentration patterns at a single industry level (Table 1), and for joint-location and co-location for pairs of industries (Tables 2, 3, 4, 5), but inter-industry relationships are sometimes more complex and involve more than a couple of industries. In these cases, the inter-industry linkages are quite difficult to understand using the previous indexes and other quantitative measures. “Direct relationships” (e.g. *vis-à-vis*) can be easily checked, but situations such as “the friends of my friends are also my friends” (e.g., “indirect relationships”) are much more complex. That is why we have decided to approach joint-location and co-location from a multidimensional perspective, rather than from a *vis-à-vis* one. This approach provides a better understanding of previous location indexes and a more realistic approach to joint-location and co-location patterns.

Concretely, we defined an n-dimension Euclidean space in which industries are located using their X2 co-location index as coordinates.<sup>24</sup> Using these coordinates, we carried out a cluster analysis using the Ward procedure (Ward 1963) to identify

<sup>23</sup> Specifically, Textiles, leather clothes and shoes, Paper and publishing, Rubber and plastic products, Machinery and equipment, Business services and Health and veterinary activities.

<sup>24</sup> If an industry is joint-located with the same industry the value of the index is 1. In terms of the novelty of this proposal, it is similar to Correspondence Analysis, but the approach proposed by this paper is more appropriate because there are no information losses linked to dimensionality reduction. Among other

**Table 4** Co-location relationships by industries (X2 co-location index). *Source:* own research

Code	Industry	(1) Below median (0.787)	(2) Between median (0.787) and mean (0.840)	(3) Between mean (0.840) and 1	(4) Over 1
1	Agriculture and fishing	5	2	9	11
2	Extractive activities	2	2	11	12
3	Food, beverages and tobacco	4	2	9	12
4	Textiles, leather clothes and shoes	18	3	5	0
5	Wood, furniture and other manufactures	5	6	6	10
6	Paper and publishing	21	4	2	0
7	Chemical products	9	4	14	0
8	Rubber and plastic products	17	5	5	0
9	Non-metallic mineral products	2	4	9	12
10	Basic metals	6	4	17	0
11	Fabricated metal products	7	6	7	7
12	Machinery and equipment	16	4	7	0
13	Office machinery, computers and medical equipment, precision and optical instruments	26	1	0	0
14	Electrical machinery and apparatus	26	0	1	0
15	Transport and communications	5	3	18	1
16	Recycling	1	0	24	2
17	Construction	12	1	5	9
18	Electricity and water distribution	5	4	9	9
19	Trade and repair	12	1	8	6
20	Hotels and restaurants	10	3	8	6
21	Transport and communications	10	3	7	7
22	Financial intermediation	23	2	1	0
23	Real estate activities	14	4	6	3
24	Business services	18	2	7	0
25	Public administration	12	12	3	0
26	Education	24	2	1	0
27	Health and veterinary activities, social services	17	4	6	0
28	Other services	15	4	7	1

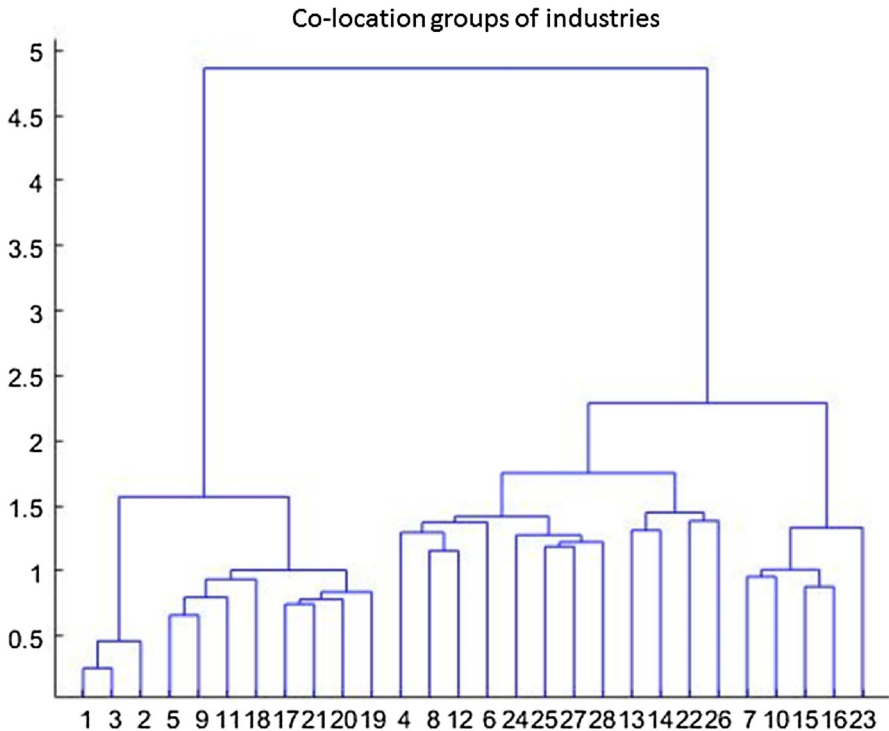
**Table 5** Types of co-location among industries (X2 co-location index). *Source:* own research

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	4	4	3	4	1	3	2	4	3	4	3	1	1	3	4	4	4	4	4	4	4	1	3	3	2
2		4	3	4	3	3	3	4	3	4	3	1	1	3	3	3	4	4	4	4	4	2	4	3	3
3			3	4	2	3	3	4	3	4	3	1	1	3	4	4	4	4	4	4	4	1	4	3	2
4				2	1	1	1	3	1	2	1	1	1	1	3	3	1	2	1	1	1	0	1	1	1
5					1	3	2	4	3	4	2	1	1	3	3	4	4	4	4	4	4	1	3	2	2
6						1	1	1	1	1	1	1	1	2	3	1	2	1	1	1	1	1	1	1	1
7							2	3	3	3	2	1	1	3	3	3	3	3	3	3	3	1	2	1	2
8								3	3	2	1	1	1	2	3	1	1	1	1	1	1	1	1	1	1
9									3	4	3	1	1	4	3	4	4	4	4	4	4	2	4	3	2
10											3	3	1	1	3	3	3	3	3	3	3	1	3	2	2
11												2	1	1	3	3	4	4	3	3	3	1	2	1	1
12													1	1	3	3	1	2	1	1	1	1	1	1	1
13														1	1	1	1	1	1	1	1	1	1	1	2
14															1	3	1	1	1	1	1	1	1	1	1
15																3	3	3	3	3	3	1	3	3	3
16																	3	3	3	3	3	3	3	3	3
17																		4	3	4	4	1	1	1	2
18																			4	3	4	1	3	3	2
19																					3	3	1	1	2
20																						3	1	2	1

Table 5 continued

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
21																					1	2	1	1	2	
22																							1	1	1	1
23																								1	1	1
24																									1	1
25																										1
26																										
27																										
28																										

Type 1 means that the X2 co-location index is below the median (0.787), type 2 means the X2 co-location index is between the median (0.787) and the mean (0.840), type 3 means the X2 co-location index is between the mean (0.840) and 1, and type 4 means the X2 co-location index is over 1



**Fig. 2** Co-location groups of industries. *Source:* own calculations

different typologies of industries based on their proximity, and our results show that industries with similar characteristics tend to be grouped.<sup>25</sup>

The results obtained are shown in Fig. 2 using a dendrogram that identifies four groups (I, II, III, and IV respectively).

In specific terms, group I includes Office machinery, computers and medical, precision and optical equipment, Electrical machinery and apparatus, Financial intermediation and Education. These are concentrated industries with X1 co-location indexes lower than 0.66 that are not especially attracted by any other industry.

Group II includes Textiles, leather clothes and shoes, Paper and publishing, Rubber and plastic products, Machinery and equipment, Business services, Public administration, Health and veterinary activities, social services, and Other services. These industries have concentration levels slightly lower than the previous group (i.e., between 0.67 and 0.79), except for Paper and publishing, which is more concentrated

Footnote 24 continued

advantages, we may highlight that there are no random biases in measures derived from criteria used for histogram construction given that spatial position is unique.

<sup>25</sup> One reviewer suggested that our method could be similar to that of correspondence analysis, but we consider that though there are some similarities our proposal is more appropriate because there are no information losses linked to dimensionality reduction. Among other advantages of our proposal, we may highlight that there are no random biases in measures derived from criteria used for histogram construction given that spatial position is unique. Consequently, although if this is not a correspondence analysis in that case we would prefer Hellinger distance rather than Chi2 (Rao 1995).

(i.e., 0.63). They tend not to locate with other industries. Generally speaking, when analysing joint-location patterns for pairs of industries there is no evidence of these collocations, but for groups of three and four industries (X3 and X4 co-location indexes) Business services and Other services tend to locate jointly with group IV industries.

Group III includes Chemical products, Basic metals, Transport materials, Recycling and Real estate activities. Most of these industries are not concentrated (the indexes range between 0.87 and 0.97) and do not show significant joint-location patterns. The exception is Real estate activities, which has stronger concentration values (0.70) and tends to locate jointly with industries in group IV.

Finally, group IV includes Agriculture and fishing, Extractive activities, Food, beverages and tobacco, Wood, furniture and other manufactures, Non-metallic mineral products, Fabricated metal products, Construction, Electricity and water distribution, Trade and repair, Hotels and restaurants and Transport and communications. This is a large group of mainly dispersed industries (only Fabricated metal products and Trade and repair have indexes lower than 1, i.e. 0.99 and 0.95, respectively), but in spite of their dispersed patterns, industries from group IV tend to be strongly located with other industries, and particularly those in the same group. In specific terms, 95.4% of X2 joint-locations are within industries from this group, 80.5% are for X3 joint-locations and 43.8% are for X4 joint-locations. The main reported co-locations with industries not belonging to group IV are with Business services, Other services and Real estate activities.

There are important similarities between the results obtained from high X co-location indexes (3 and 4) and those obtained using clusters of positions in a multidimensional space. Concretely, the previous grouping roughly indicates an inverse relationship between the technological level of industries and their joint-location patterns, in a similar way to the one we identified for pairs of industries: high-tech industries do not show joint-location patterns whilst medium and low-tech industries do.

By using the positions of each industry in an n-dimensional space it is possible to establish a network, taking into account the distance between them. Considered in detail, this is a complete network in which all nodes (i.e., industries) are interconnected and where short (long) distances imply similar (different) joint-location structures. We use distances between industries instead of the X2 joint-location index because it avoids problems related with the different behavior of the index for values higher and lower than 1, in a similar way to networks with negative weights.

Figure 3 shows the relationship between bilateral joint-location indexes (horizontal axis) and multidimensional distances (vertical axis). The star colors indicate the groups to which each pairs of industries belong (a circle is used instead of a star when both industries belong to the same group). This figure is closely linked with Fig. 2, as the red dots are group I industries in Fig. 2, the yellow dots are group II industries, the green dots are group III industries and the purple dots are group IV industries. The group IV industries are those with a joint-location pattern (i.e., values of X2 joint-location index larger than 1) and minimal distances between them (small distances). The negative slope of these dots indicates that the higher the X2 joint-location index, the greater the similarity between location patterns.

Given that we are dealing with a complete network, graphically illustrating interindustry linkages is a difficult task and we need to seek alternative strategies.

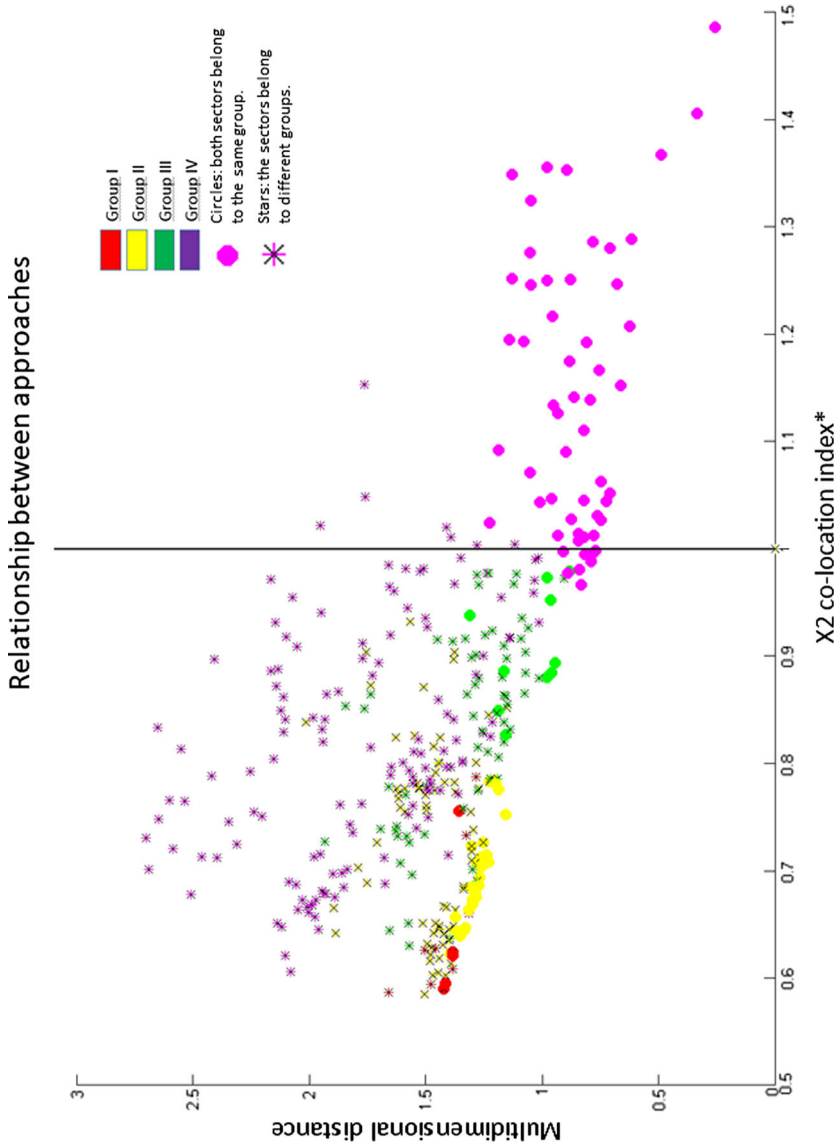


Fig. 3 Relationship between approaches

In this sense, Minimum Spanning Trees (MST) (Graham and Hell 1985) enables us to connect all the nodes in the network without any cycles, and to minimise the total edge weight.

Figure 4 shows that industries are closely positioned around four industries in group I (Office machinery, computers and medical equipment, precision and optical, Electrical machinery and apparatus, Financial intermediation and Education). Concretely, Office machinery is linked with dispersed primary and manufacturing activities with high joint-location levels, Education is linked with concentrated manufacturing industries that do not locate jointly, Financial intermediation is mainly linked with service and final demand industries (which includes industries from groups II and III that are jointly located with group IV). Finally, Electrical machinery and apparatus is mainly isolated, but links the groups centered around Education and Financial intermediation.

As highlighted previously, multilateral comparison of joint location and co-location indexes is a complicated task as their potential determinants are not the same. Nevertheless, as it is reasonable to assume that interindustry linkages may have a key role, we have compared industries' ranks in co-location terms with their position in terms of interindustry linkages using data from input–output tables. In specific terms, we defined two  $n$ -dimension spaces where the industries are positioned on each axis according to their percentage of interindustry sales/purchases relationships with the other industries. For comparison purposes, we recalculated the X2 co-location indexes, fitting values between 0 and 1. After performing these calculations, we computed the distances between the position of each industry in joint-location terms and the position in terms of sales/purchases, and we then selected the shorter distance as co-location may be caused by advantages of proximity to a customer/supplier.

Figure 5 shows that industries which no significant joint-location pattern have a similar pattern in terms of interindustry linkages (i.e. they have an even distribution in both cases<sup>26</sup>). Meanwhile, among industries that tend to locate jointly with other industries, there are fewer similarities between their locational structure and their industry mix. Whilst the shortest distances are found on group I, the longest distances are found on group IV. This result confirms our previous findings, and suggests that there is no clear empirical evidence of col-location as interindustry linkages are not the main determinants of joint-location, although further research should carry out robustness checks for alternative spatial and industry aggregation levels.

Our results suggest that inter-industry knowledge spillovers among high-tech firms may not be transmitted through physical proximity but by alternative means, as firms in high-tech industries are those with the lowest joint-location levels. Nevertheless, as these industries are highly concentrated, it seems that face-to-face knowledge flows come from firms belonging to the same industry. At the same time, there are some low and medium-tech-industries that locate in a dispersed way and jointly with firms from other industries. All in all, joint-location and co-location patterns indicate that most high-tech firms benefit mainly from specialised environments, whilst medium and low-tech firms prefer diversified ones. These results may seem counterintuitive, but they are reasonable if we take into account *i*) the medium knowledge level of most

<sup>26</sup> Using a geometric perspective, they are close to the bisectriz.



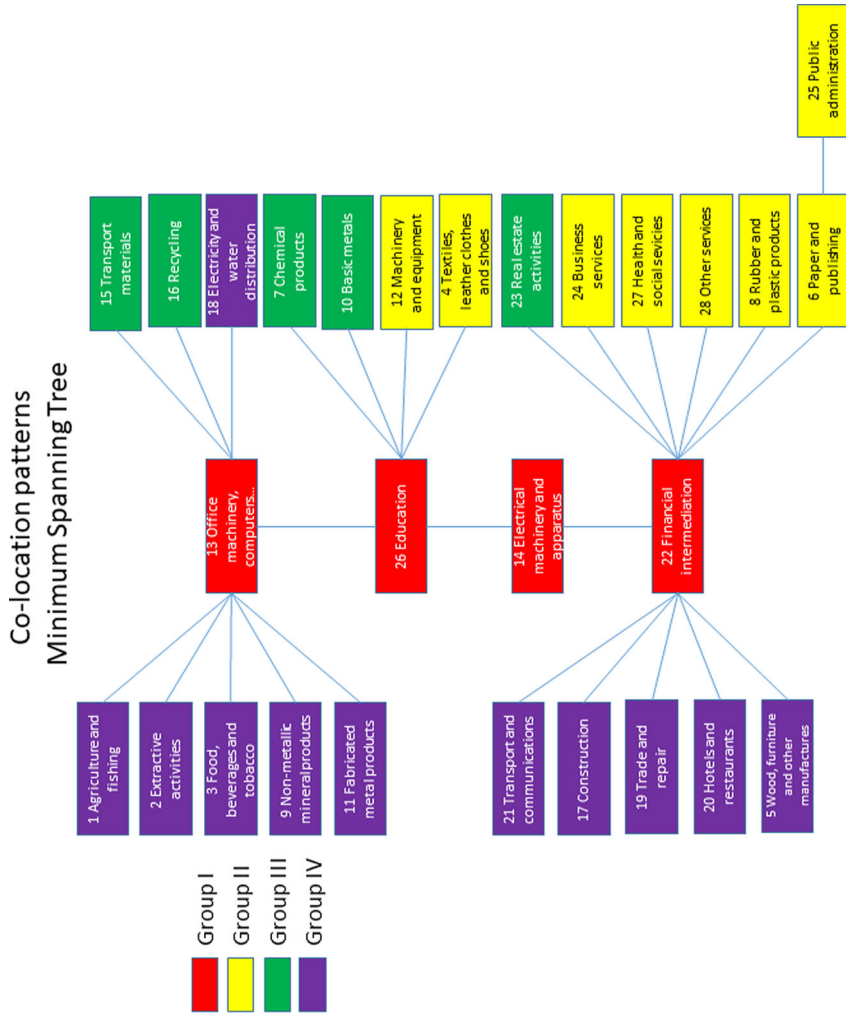


Fig. 4 Co-location patterns minimum spanning tree

Co-location and interindustry linkages

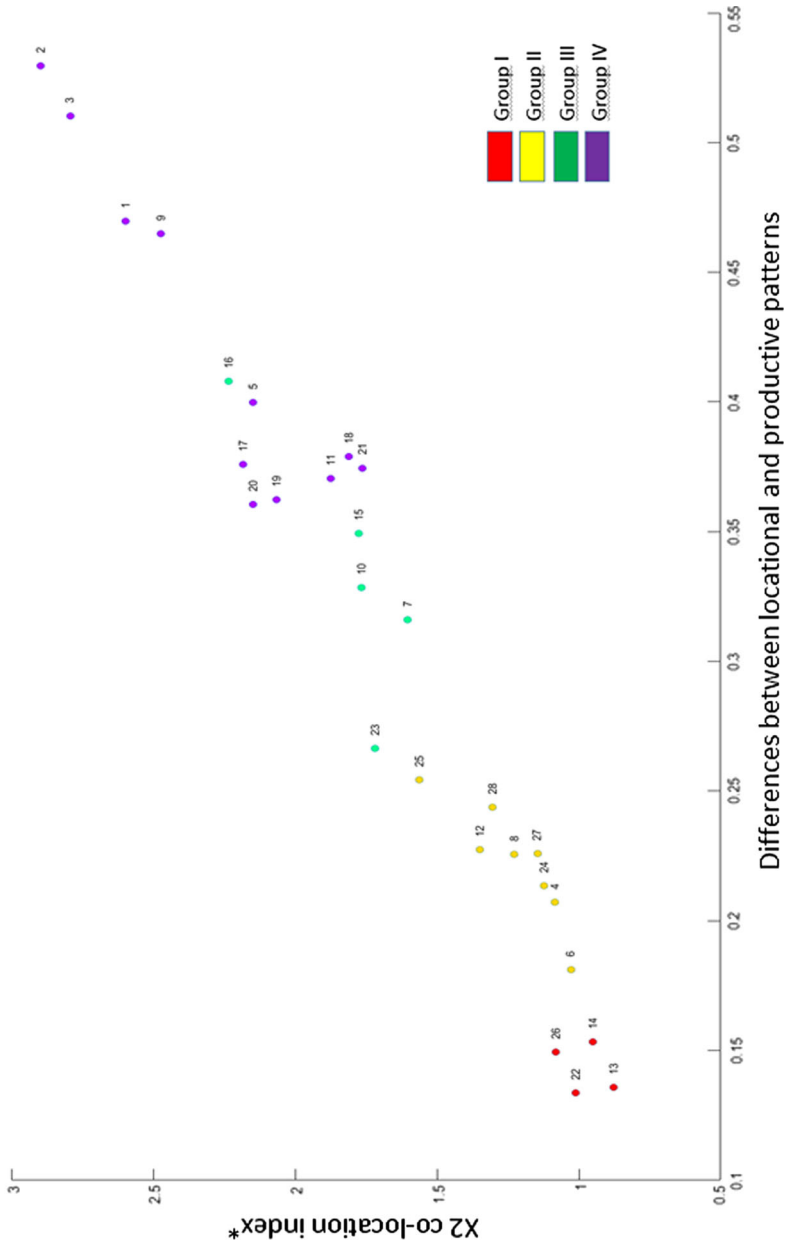


Fig. 5 Co-location and interindustry linkage

Spanish firms, and *ii*) the empirical evidence for the role of skilled human capital and knowledge spillovers in firms' location decisions (see Arauzo-Carod 2005, 2013).

## 5 Conclusions

This paper has contributed to the extant literature on joint-location and co-location by designing a procedure for identifying groups of industries that tend to locate together and analysing whether this behaviour can be explained in terms of relationships along the supply chain (i.e., co-location) or by common location determinants shared by those industries (i.e., joint-location). This distinction enables a detailed analysis of firm location determinants, and our results show that there are strong industry-level determinants that guide location decisions, favouring sites around firms in the same industry or on the contrary, favouring joint-location with firms from other industries. Additionally, there are also industry determinants that result in concentration in few locations or on the contrary, to dispersion over a large number of areas. Likewise, these results indicate that proximity to other (the same) economic activities matters and that firms need "specific" neighbours in order to maximise their performance, as it seems reasonable to argue that if location patterns are rationally driven they may exert a positive effect over firms' performance. Nevertheless, we leave the specific analysis about performance for future research.

The methodology proposed in this paper provides a better explanation of the procedures driving firms' location patterns, but much more work needs to be done in this area, particularly on identifying cluster size and improving the capture of cluster borders. As this methodology involves dividing spaces into homogeneous cells of equal size, it is important to handle it with care, because cell size influences the number and characteristics of the identified clusters. Specifically, bigger cells are more likely to contain a cluster, whereas smaller cells are more likely to have fewer inter-industrial clusters because the number of firms in each cell will be smaller. Given that in this paper we have assumed equal land areas for all the areas, it would appear that using flexible sizes is a better strategy and is therefore a promising line for future research, but that is obviously beyond the scope of this paper. In any event, putting aside administrative borders allows us to concentrate on proximity and its effects on firms' location decisions.

This is merely an initial attempt to better identify the forces behind firms' location patterns at an industry level that may be them towards cluster formation. However, this is just a starting point and further work needs to be done, in particular to cover industry-specific characteristics that influence the location decisions of firms. We therefore plan to extend our analysis of specific types of clusters (both specialised and diversified) and to cover several types of urban/rural environments that are hypothesised to influence such agglomerative behaviour. Finally, as we mentioned above, industry aggregation is also important and despite the computational constraints that make it unfeasible to work with such disaggregate industry-levels, we need to carry out further research to accurately determine whether our results are robust at different industry aggregation levels (e.g. the 3-digit level).

**Acknowledgements** We are grateful to the seminar participants at the 12th EUNIP Conference (Universitat Rovira i Virgili) and at the International Meeting on Regional Science (Universidad de Badajoz), and especially to Anxo Sánchez for his valuable comments. Any errors are, of course, our own.

**Funding** This paper was partially funded by ECO2014-55553-P, ECO2013-42310-R, the “*Xarxa de Referència d’R+D+I en Economia i polítiques Públiques*”, the SGR programme (2014 SGR 299) of the Government of Catalonia and the R&D Spanish project ‘CITiTALENT’, CSO2016-74888-C4-4-R (AEI/FEDER, UE).

## Appendix

See Table 6.

**Table 6** List of industries.  
Source: SABI

Code	Industry
1	Agriculture and fishing
2	Extractive activities
3	Food, beverages and tobacco
4	Textiles, leather clothes and shoes
5	Wood, furniture and other manufactures
6	Paper and publishing
7	Chemical products
8	Rubber and plastic products
9	Non-metallic mineral products
10	Basic metals
11	Fabricated metal products
12	Machinery and equipment
13	Office machinery, computers and medical equipment, precision and optical instruments
14	Electrical machinery and apparatus
15	Transport materials
16	Recycling
17	Construction
18	Electricity and water distribution
19	Trade and repair
20	Hotels and restaurants
21	Transport and communications
22	Financial intermediation
23	Real estate activities
24	Business services
25	Public administration
26	Education
27	Health and veterinary activities, social services
28	Other services

## References

- Acosta M, Coronado D, Flores E (2011) University spillovers and new business location in high-technology sectors: Spanish evidence. *Small Bus Econ* 36:365–376
- Albert JM, Casanova MR, Orts V (2012) Spatial location patterns of Spanish manufacturing firms. *Pap Reg Sci* 91(1):107–136
- Arauzo-Carod JM (2005) Determinants of industrial location: an application for Catalan municipalities. *Pap Reg Sci* 84:105–120
- Arauzo-Carod JM (2008) Industrial location at a local level: comments on the territorial level of the analysis. *Tijdschr Econ Soc* 99:193–208
- Arauzo-Carod JM (2013) Location determinants of new firms: does skill level of human capital really matter? *Growth Change* 44(1):118–148
- Arauzo-Carod JM, Manjón-Antolín M (2004) Firm size and geographical aggregation: an empirical appraisal in industrial location. *Small Bus Econ* 22:299–312
- Arbia G (2001) Modeling the geography of economic activities on a continuous space. *Pap Reg Sci* 80:411–424
- Barlet M, Briant A, Crusson L (2009) Location patterns of services in France: a distance-based approach. Working paper, Paris School of Economics
- Bickenbach F, Bode E, Krieger-Boden C (2013) Closing the gap between absolute and relative measures of localization, concentration or specialization. *Pap Reg Sci* 92(3):465–479
- Blöch F, Theis FJ, Vega-Redondo F, Fisher EO (2011) Vertex centralities in input-output networks reveal the structure of modern economies. *Phys Rev E* 83:1–9
- Boix R, Galletto V (2008) Marshallian industrial districts in Spain. *Sci Reg Ital J Reg Sci* 7(3):29–52
- Brenner T (2004) *Local industrial clusters: existence, emergence and evolution*. Routledge, London
- Brenner T (2006) Identification of local industrial clusters in Germany. *Reg Stud* 40(9):991–1004
- Devereaux M, Griffith R, Simpson H (2004) The geographic distribution of production activity in the UK. *Reg Sci Urban Econ* 34(5):533–564
- Durantón G, Overman HG (2005) Testing for localization using microgeographic data. *Rev Econ Stud* 72:1077–1106
- Durantón G, Overman HG (2008) Exploring the detailed location patterns of U.K. manufacturing Industries using microgeographic data. *J Reg Sci* 48(1):213–243
- Ellison G, Glaeser EL (1997) Geographic concentration in US manufacturing industries: a dartboard approach. *J Polit Econ* 195:889–927
- Graham RL, Hell P (1985) On the history of the minimum spanning tree problem. *Ann Hist Comput* 7(1):43–57
- Hoffmann J, Hirsch S, Simons J (2017) Identification of spatial agglomerations in the German food processing industry. *Pap Reg Sci* 96(1):139–162
- Isard W (1960) *Methods of regional analysis*. MIT Press, Cambridge
- ISTAT (1996) *Rapporto annuale. La situazione del paese nel 1995*. ISTAT, Rome
- Jofre-Monseny J, Sanchez-Vidal M, Viladecans-Marsal E (2015) Big plant closures and agglomeration economies. SERC discussion paper no. SERCDP0179
- Larsson JP, Öner Ö (2014) Location and co-location in retail: a probabilistic approach using geo-coded data for metropolitan retail markets. *Ann Reg Sci* 52:385–408
- López-Bazo E (2006) Definición de la metodología de detección e identificación de clusters industriales en España. Dirección General de la Pequeña y Mediana Empresa (DGPYME), Madrid
- Marshall A (1890) *Principles of economics*. MacMillan, New York
- Maurel F, Sédillot B (1999) A measure of the geographic concentration in French manufacturing industries. *Reg Sci Urban Econ* 29:575–604
- Mulligan GF, Schmidt C (2005) A note on localization and specialization. *Growth Change* 36(4):565–576
- Olsen J (2002) On the units of geographical economics. *Geoforum* 33:153–164
- Openshaw S, Taylor PJ (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N (ed) *Statistical applications in the spatial sciences*. Pion, London, pp 127–144
- Paluzie E, Pons J, Tirado D (2004) The geographical concentration of industry across Spanish regions, 1856–1995. *Rev Reg Res* 24(2):143–160
- Porter M (1998) Clusters and the new economics of competition. *Harvard Bus Rev* 76(6):77–90

- Rao CR (1995) A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Questiio* 19(1–3):23–63
- Sforzi F, Lorenzini F (2002) I distretti industriali. In: IPI-Ministero delle Attività Produttive, L'esperienza italiana dei distretti industriali, Rome, pp 20–33
- Silverman B (1986) Density estimation for statistics and data analysis. Chapman & Hall, London
- Titze M, Brachert M, Kubis A (2011) The identification of regional industrial clusters using qualitative input–output analysis (QIOA). *Reg Stud* 45:89–102
- Van Egeraat C, Morgenroth E, Kroes R, Curran D, Gleeson J (2018) A measure for identifying substantial geographic concentrations. *Pap Reg Sci* 97(2):281–300
- Viladecans E (2004) Agglomeration economies and industrial location: city-level evidence. *J Econ Geogr* 4(5):565–582
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Wrigley N (1995) Revisiting the modifiable areal unit problem and the ecological fallacy. In: Cliff AD, Gould PR, Hoare AG, Thrift NJ (eds) *Diffusing geography*. Wiley, Oxford, pp 49–71