

Mapping industrial patterns in spatial agglomeration: A SOM approach to Italian industrial districts [☆]



Vittorio Carlei ^{a,*}, Massimiliano Nuccio ^b

^a Università Gabriele d'Annunzio, Dipartimento di Scienze Economico-Quantitative e Filosofico-Educative, Viale Pindaro, 42, Pescara 65127, Italy

^b Leuphana Universität Lüneburg, Scharnhorststrasse 1, 21335 Lüneburg, Germany

ARTICLE INFO

Article history:

Received 24 February 2013

Available online 14 December 2013

Keywords:

Self-organizing maps
Pattern recognition
Spatial agglomeration
Industrial districts

ABSTRACT

The paper presents a new approach based on Self-Organizing Maps (SOM) and a new index called Relative Industrial Relevance (RIR) to discover, track and analyze spatial agglomeration of economic activities. By comparing patterns of local employment, this methodology shows how the local supply of human capital can explain the advantages generating spatial agglomerations. The reference case for this research is Italy, which has developed one of the most remarkable and studied example of spatial agglomerations, the Industrial Districts (IDs). IDs are traditionally identified by indexes which measure the physical concentration of firms belonging to a given industry, but are unable to seize the overall productive structure of the local economy. Employing the Italian Clothing Industry as test bed, the approach proposed in this paper identifies spatial agglomerations in terms of industry patterns and not of industry concentration. This methodology can offer a new basis to analyze the multiple pattern of local development.

© 2013 Published by Elsevier B.V.

1. Introduction

This paper develops a methodology based on Self-Organizing Maps (SOM) and introduces a new index to identify spatial agglomeration of economic activities, namely *Relative Industrial Relevance* (RIR). The methodology can be applied to different sizes of economic regions, to different industries and at different levels of industry classification. The topological clustering provided by the SOM is able to define industrial patterns and also can measure the relative relevance of a given industry with Component Planes (Kohonen, 2001).

We define the RIR index as the value assumed by the corresponding feature of the SOM Processing Element (PE), which quantifies the relevance of that industry in the local economies belonging to the same PE.¹ From an economic perspective, the RIR can select those PEs which mirror patterns of labour factor supply as a proxy of localised human capital. We are therefore able to measure the ability of this pattern to be an original determinant of spatial agglomeration.

Applying this methodology on highly disaggregated cross-sectional spatial data² the paper demonstrates that a rich taxonomy

of spatial agglomerations emerges even when the focus is on a single industry. We claim that RIR, differently from usual concentration-based indicators, is able to identify different forms of spatial agglomeration and local patterns of industrial co-agglomeration.

The determinants of agglomeration may depend on specialization of input, labour market pooling and knowledge spillovers (Rosenthal and Strange, 2001) and drive to a concentration of specialized industries over a defined geographical space. One way to characterize the socio-economic complexity of spatial agglomerations is what are commonly known as Marshallian externalities. Their meaning cannot be properly understood at either firm or market level, but looking at the intermediate layer where non-market interaction among players typically occurs. According to the Marshall–Arrow–Romer (MAR) approach, spillovers within the same industry favour innovation, whereas Jacobs (1969) contends that diversification of processes and production stimulates innovation and, therefore, production and employment (Van der Panne, 2004). This opposition mirrors the market's local structure. In MAR models, market power underpins innovation because it allows firms in the industry to incorporate innovation revenues through their localised monopoly. On the contrary, according to Jacobs (1969), higher market competition stimulates creativity, thereby fostering innovation.

Although a lot of economic research has been conducted on this topic to support alternatively one of the former positions, most of the literature is built on measures of concentration. Birckenbach and Bode (2008) have shown that measures of concentration and specialization explore two sides of the same

[☆] This paper has been recommended for acceptance by S. Sarkar.

* Corresponding author. Tel.: +39 0645595383.

E-mail addresses: carlei@unich.it (V. Carlei), massimiliano.nuccio@gmail.com (M. Nuccio).

¹ The PE computed by the SOM algorithm is a pattern which clusters similar samples of the dataset.

² Number of employees per industry in a given local economy.

coin, and can be nested in measures of the localization of an economy (Cutrini, 2010), which capture both concentration and specialization simultaneously.

Whatever position is endorsed, there is a consensus that local human capital and knowledge spillovers are crucial for spatial agglomerations to achieve competitive advantage. The cornerstone of our approach is that the spatial agglomeration of firms can be explained also in terms of local supply of human capital, measured by similar patterns of labour factor employed.

The paper is organized as follows. Section 2 reviews and summarizes measurement of Italian IDs, also presenting a critique that contributes to the development of the analytical method applied in this paper. Section 3 explains the methodology of spatial pattern recognition by employing the SOM algorithm. Section 4 develops some empirical results of the proposed approach building on the Italian case.

2. Limitation in identifying and measuring IDs

Italian Industrial Districts (IDs) are a peculiar example of spatial agglomerations and have some distinctive characteristics that explain their enduring success (Becattini, 1990; Becattini et al., 2003). Examining IDs in relation to simple causal mechanisms involves a tendency to isolate, recognize and acknowledge only those examples that are especially self-evident, for example, in presenting exceptionally high concentrations of firms belonging to the same industry. We intend to compare the results of our methodology to the evidence of Italian IDs since Italy has traditionally considered industrial agglomerations of SMEs to be as an alternative model of industrial development and has set up an international benchmark.

The methodology usually followed to map IDs (De Blasio et al., 2009) departs from the settlement of Local Labour Market Areas (LLMAs), that is, aggregations of municipalities based on the analysis of workers' transportation flows. Daily commuting between home and the workplace is a good proxy for the boundaries of the local "community of people and firms" (Sforzi and Lorenzini, 2002). LLMAs borders are drawn using algorithms that maximize "territorial self-containment" (i.e., relatively closed systems with respect to commuting flows), which originate in urban development studies.³

IDs are sorted from LLMAs using a four-step technique, based on the coefficients of spatial concentration, the so-called Sforzi algorithm (Sforzi, 1990, 2009):

1. localization of "manufacturing" LLMAs (i.e. LLMAs that present a relatively high incidence of workers in the manufacturing sector compared to the overall workforce);
2. selection of SME⁴ LLMAs among "manufacturing" LLMAs (i.e., "manufacturing" LLMAs that present a relatively high incidence of workers employed in SMEs compared to the total manufacturing workforce);
3. choice of the main industry of a given LLMA (i.e., selecting the industry that presents a relatively higher incidence of workers employed compared to the overall manufacturing workforce);
4. level of ID specialization (i.e. whether at least 50% of overall LLMA workers are employed in SMEs operating in the main industry).

³ IDs are identified firstly by ISTAT, the Italian National Agency of Statistics. ISTAT counted 156 IDs out of 686 LLMAs in 2001 and 199 out of 784 in 1999.

⁴ EU adopted a recommendation that defined a small firm as having less than 50 employees and a medium-sized as having less than 250 (European Commission 1992, 1996).

This methodology is not universally accepted in the literature. For instance, Tattara (2001, 2002) argues that the use of LLMAs in identifying IDs is misleading. Since the Sforzi algorithm (Sforzi, 1990, 2009) covers a wide range of different phenomena and, although exogenous, is not consistent with the IDs definition. Moreover, it does not clearly distinguish between firms belonging or not belonging to IDs, which makes it impossible to define a control group for an empirical analysis. Non-IDs firms are residual to IDs and show a variety of openness/closedness, and non-IDs firms are situated in undeveloped regions, industrial *milieux* and metropolitan areas. Viesti (2005) stresses some difficulties in surveying LLMAs that can invalidate IDs because of the increasing variance in demographic dimension and artificial Census borders, the huge variability in urban attractive potential and the lack of recognition by locals. Further critical issues raised on IDs refer to:

1. the miscalculation of relative manufacturing employment on total employment: when the incidence of services (in cities) or agriculture is high, LLMAs cannot capture a measure of industrialization;
2. the rigid definition of SMEs (productive plants smaller than 250 workers), which does not consider the concentration of the same ownership over different smaller plants;
3. the potential distortion of a relative measure of industrial specialization;
4. the lack of surveys on non-IDs productions, e.g. agro-food, business service, tourism, ICT, etc.

Calafati and Compagnucci (2005) claim that the LLMA method is misleading for Italian local development. LLMAs rely only on "commuting for labour" which is not necessarily the most important reason for territorial interaction and is built on an implicit and not-replicable algorithm. This methodology does not explain the poli-centric nature of Italian urban development, the local organization of integrated local economies, and the hierarchy and strength of the inter-dependency between Councils.

Iuzzolino (2003) remarks the absolute arbitrariness of the geographic and industrial extents of IDs. First, since LLMAs are exogenous to the Census and their borders depend on the commuting flows in any industry, their borders are extremely variable because of modern transport networks which enable ever longer commuting distances. Second, specialized industries are arbitrarily and exogenously defined. Third, even the limit of 250 employees is exogenous because it comes from the EU legislation on fiscal incentives for firms. It does not consider that, on the one hand, corporate assets can corrupt the SME survey and, on the other, since firms' dimension changes significantly according to different technologies and markets, big enterprises are *de facto* excluded in studying IDs.

In the rigorous exercise of IDs mapping undertaken by Iuzzolino (2003), the author derives an index for the measurement of economies of agglomeration (γ) building on the work of Ellison and Glaser (1997). This index estimates the probability that two plants belonging to the same industry co-localize in the same area. Therefore, the index of specialization of a given single area i for a given industry p is:

$$\gamma_i = \frac{(s_i - x_i)^2 - s_i^2 h_i (1 - \sum_{i=1}^n x_i^2)}{(1 - \sum_{i=1}^n x_i^2)(1 - s_i^2 h_i)}, \quad (1)$$

where γ_i is the combination of two elements:

1. *Raw spatial concentration* (G) of the industry p , which uses Gini's concentration index. Dividing the territory in n areas, and denoting by Z_i^p the number of employees in the industry p in the area i and by Z^p the size of the national employee pool in the same industry for all the areas, we have:

$$G = \sum_{i=1}^n (G_i) = \sum_{i=1}^n (s_i^p - x_i)^2, \quad (2)$$

with

$$s_i^p = \frac{z_i^p}{Z^p} \quad \text{and} \quad x_i = \frac{\sum_p z_i^p}{\sum_p Z^p}. \quad (3)$$

2. *Herfindahl Index (H)* of the industry p measuring plant employment shares. Given m plants for the industry p and z_j^p employees in the plant j , and considering all the i areas, we can calculate H :

$$H = \sum_{j=1}^m \frac{(z_j^p)^2}{(Z^p)^2}, \quad (4)$$

Therefore, H can also be written as the combination of the local components h_i and s_i :⁵

$$H = \sum_i h_i s_i^2. \quad (6)$$

For a given industry p , as h_i grows with the inequality in firms' dimension, it decreases the quantity of gross specialization G of that industry when this depends on a strong concentration of employees in the biggest plants. By setting:

$$\sum_i \gamma_i = \gamma, \quad (7)$$

and

$$T = \left(1 - \sum_i x_i^2\right), \quad (8)$$

we obtain the following synthetic index:

$$\gamma = \frac{G - TH}{T(1 - H)}. \quad (9)$$

Iuzzolino (2003) adds two further conditions to enhance the characterization of districts. First of all, since the contribution of two given areas to the index γ_i does not depend on the distance between them, the author puts together a conjoint of elementary geographic units into an agglomerated space (S) where the area with the highest γ_i value is the "centre of agglomeration". Under this condition, Iuzzolino invites to consider a "strongly agglomerating" space or IDs as "a continuum of territories, specialized in a given industry, and such that at least one of them presents a level of specialization significantly stronger than the expected value under the assumption of absence of agglomerative advantages".⁶ Secondly, the mapping exercise should capture not only those externalities that are generated by geographic proximity, but also those deriving from intra- and inter-

⁵ If k_i is the number of plants in the area i , multiplying and dividing Eq. (4) by the square of the total number of employees for a given industry in the area z_i^p , we obtain:

$$H = \sum_{j=1}^{k_1} \frac{(z_j^1)^2}{z_1^2} \left(\frac{z_1^2}{Z^2}\right) + \dots + \sum_{j=k_{i-1}+1}^{k_i} \frac{(z_j^i)^2}{z_i^2} \left(\frac{z_i^2}{Z^2}\right) + \dots + \sum_{j=k_{n-1}+1}^m \frac{(z_j^n)^2}{z_n^2} \left(\frac{z_n^2}{Z^2}\right). \quad (5)$$

⁶ For each area it is possible to test if the level of raw specialization statistically consistent with the absence of agglomeration advantages (Iuzzolino, 2003). Formally, we select the local areas where:

$$G_i > \mu_i + 2\sigma_i. \quad (10)$$

Writing k_i for the number of plants located in the i^{th} area, the following threshold of agglomerative intensity should result statistically significant:

$$\gamma_i > 2 \frac{\left(s_i^2 h_i H k_i - s_i^4 \sum_{j=1}^{k_i} \frac{z_j^4}{z_i^4}\right)^{\frac{1}{2}}}{T(1 - H)}. \quad (11)$$

industrial links existing among different production chains. Using the ISTAT input–output matrices, Iuzzolino (2003) proposes to define macro-industries made up of branches of activities with a high level of internal exchanges and with a limited degree of self-containment. One then proceeds to the estimation of the employees for each given component industry in each geographical area (municipality) and, finally, carries out a cluster analysis which employs individual component industries as variables to be aggregated.

Contrary to the limitations of the above mentioned approach, our methodology based on SOM algorithm presents several advantages. First of all, the determination of spatial agglomerations relies on specialization parameters for the geographical units, thus allowing to overcome LLMA and all of their limitations. Secondly, the importance of a given industry in the local economy is measured by taking into account the pattern of the whole distribution of employment, rather than by focusing on a single concentration/threshold value. Thirdly, the relevant industrial areas are defined by discovering endogenously alternative patterns of spatial agglomerations of industries. As noted by Patuelli et al. (2006) the analysis of social phenomena, and spatial in particular, can be very explicative by employing non-linear systems, some of which are widely used especially in evolutionary approaches. The SOM algorithm accomplishes these purposes, since it does not define any particular a priori on the training process. The result is not supervised, although driven by relationships implicit in the observed samples, which deform the space to preserve the topological relationships among the data.

Finally, the ability to map any relationship between features and samples, allows extracting information even when linear relationships are not present, but it is reasonable to make conjectures about more complex relationships.

3. Industrial patterns recognition: spatial agglomeration with SOM neural networks

SOMs are among the most important and widely used neural network architectures. They were developed by Kohonen (2001); for a collection of state-of-the-art applications to geographical analysis see Agarwal and Skupin (2008). The key element of a SOM network is the *Kohonen Layer* (KL), which is made up of spatially ordered *Processing Elements* (PEs) or neurons. The global state of the layer evolves during the learning process, identifying each PE as a representative pattern of the input data with an unsupervised learning technique.⁷ (see Fig. 1).

A vector is associated with the generic PE in the KL, whose elements are the weights relative to the patterns identified. The weight vector associated with the generic PE_r in the KL is indicated by

$$W_r = (W_{r,1}, W_{r,2}, \dots, W_{r,p}, W_{r,p+1}, \dots, W_{r,N-1}, W_{r,N}). \quad (12)$$

"Closeness" of vectors can be expressed in terms of several different possible metrics. Although several definitions of a neighborhood are available, a convenient metric is based on the Euclidean distance d among the PEs on the KL.⁸ For each given PE in the KL, there is a set of samples (input data), which, if submitted iteratively to the SOM during the learning process, makes the given PE the most representative pattern of them. These samples define a region over the KL: near-by input data tend to map onto the same PE, or more generally within the same neighborhood of PE. Neighborhood is defined in the two dimensional space of the KL, whatever the dimensionality of the

⁷ Training a neural network model essentially means selecting one model from the set of allowed models (or, in a Bayesian framework, determining a distribution over the set of allowed models) that minimizes the cost criterion.

⁸ The function $h(d)$ describes how the generic PE updates starting from the associated weights array W_r and from the size of the error $X - W_r$, with X as a generic input vector. In our case h is a Gaussian distribution.

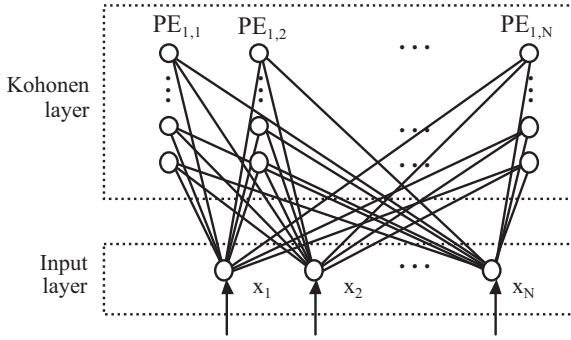


Fig. 1. SOM with N variables and the PEs in the 2-D Kohonen layer (KL).

input space. The learning criterion, therefore, cannot be geometrically invariant to the dimension of the input space.

The clustering with SOM keeps the relations of similarities across the observed objects, defining a continuous space differently populated in terms of frequency of the original objects.

In this paper we present a new methodology based on the SOM algorithm to draw spatial agglomerations that do not reflect the prevailing approach based on concentration analysis. We trained a SOM to define spatial agglomerations by taking municipalities as the geographic samples and the number of employees for each industry as its features.

The dataset to train the SOM algorithm is obtained from a matrix X , whose entries x_{ij} are i -samples of the local economies and j -features of the employees number for each industry. We rescaled X into a new matrix Y whose entries y_{ij} are defined as:

$$y_{ij} = \frac{e^{\delta_{ij}}}{1 + e^{\delta_{ij}}}, \quad (13)$$

with⁹

$$\delta_{ij} = \alpha_i + \beta_i \cdot u_{ij}, \quad (14)$$

where

$$\alpha_i = \frac{1}{\max_j(x_{ij})}, \quad (15)$$

and

$$\beta_i = \frac{\min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})} \quad (16)$$

α_i and β_i are the two vectors of parameters for the logistic function to rescale the matrix X by rows to preserve the key properties of the local distribution of employees across the different industries for each sample. Both samples and features can be employed at different levels of aggregation: the former can be Municipality, Province, Region and Country; the latter can refer to different industries and various classifications, consistently with the number of digit per each industry. During the training of the SOM each sample is represented as a pattern whose properties capture how the labour factor is locally allocated across the industries.

At the end of the training process of the matrix Y , we obtain the $RIR_{r,p}$ index for a given industry p defined as the value of the weight $w_{r,p}$ in the vector W_r associated with the respective PE_r . Formally,

$$RIR_{r,p} = w_{r,p}, \quad (17)$$

⁹ The logistic function enables a non-linear scaling which tends to emphasize differences among the central values of the distribution, while being relatively unaffected by differentials that occur at its extremes.

$$RIR_p = W_p \quad (18)$$

with

$$W_p = \{w_{r,p}\}_{r=1}^k \quad (19)$$

and

$$k = 5 * \sqrt{(n)}^{10} \quad (20)$$

where n is the number of training samples and k the number of the PEs.¹⁰

Therefore, the RIR represents the convergence value of W , by which the SOM algorithm has reconstructed the relationships between the given industry and all the others through examination of observed data and the consequent training process.¹¹ Contrary to the results of spatial analysis based on the concentration of a single industry, high values of the RIR define spatial agglomerations of geographic samples with a relevant similarity in terms of industrial patterns. With RIR we achieved three main results:

1. a new measure of relevance for a given industry p on a given PE_r : high values of $RIR_{r,p}$ suggest that a given industry has a relatively greater weight than the others;
2. a new relation between high levels of RIR_p and similarities of the geographic samples in terms of industrial patterns (the neighborhood of PEs on the KL);
3. a new definition of spatial agglomeration, based on the membership to the same neighborhood of the KL.

4. Spatial agglomeration and industrial pattern

4.1. The case of the Italian Clothing Industry

In this section we present an application of the above mentioned methodology to the Italian case. Employees are organised in 57 industrial divisions according with the ATECO 2-digit classification.¹² The dataset gathers the overall number of the Italian municipalities, covering therefore the entire country.¹³ As a result, the matrix X_{ij} has $i = 1 \dots 8101$ municipalities and $j = 1 \dots 57$ industries.

The training process of the dataset of the matrix Y obtained from Eqs. (13)–(16) is performed with the following outputs and parameters:

1. The training process requires a map size (number of PEs) specification. Larger map sizes result in more detailed patterns; smaller map sizes result in more general patterns. Since larger array size results in more patterns, there is a trade off between compressing information into few patterns and accuracy. The number of PEs corresponds to Eq. (20): the general formula suggested by Kohonen in the Matlab SOM Toolbox (Vesanto et al., 2000) is a good compromise in terms of errors (Table 1).
2. The shape of the lattice has been made by hexagons, which have performed better than rectangles in terms of TE (Table 1).

¹⁰ This is the general approach suggested by Kohonen to train the SOM available at <http://www.cis.hut.fi/somtoolbox/documentation>. In general, there is an inverse relationship between the number of PEs and the quantization error (QE), as well as a direct relationship between a high number of PEs and the topographic error (TE) (Kiviluoto, 1996).

¹¹ The values of the RIR index are in the range [0, 1].

¹² ATECO 2002 is a hierarchical coding system adapted from the NACE- Statistical Classification of the Economic Activity in the European Union (Eurostat) covering different levels of classification.

¹³ Data are available in the census “VIII Censimento Generale dell’industria e dei servizi” provided by the Italian statistical agency ISTAT (2001). ISTAT conducted a census of 4.3 mln firms and institutions, for a total 4.7 mln plants and 19.4 mln employees.

Table 1

Quantitative measures of mapping quality: average quantization error (QE) and topographic error (TE).

PEs	Shape	Random	Initialization	Linear	Initialization
		Q. Err.	T. Err.	Q. Err.	T. Err.
20	Rect	0,683	0,110	0.657	0.022
100	Rect	0,652	0,070	0.625	0.079
300	Rect	0,592	0,137	0.58	0.154
442	Rect	0,559	0,127	0.550	0.179
20	Hexa	0,675	0,109	0.66	0.011
100	Hexa	0,654	0,083	0.627	0.037
300	Hexa	0,593	0,086	0.585	0.081
442	Hexa	0,56	0,085	0.556	0.114

3. The map grid side-lengths are determined by the ratio between eigenvalues of the training data: $k = 26 * 17 = 442$.
4. The initialization is performed by a linear approach (Kohonen, 2001, 142), where the weight vectors are initialized in an orderly fashion along the linear subspace spanned by the two principal eigenvectors of the input data set. The eigenvectors can be calculated using Gram-Schmidt procedure (Vesanto, 1999).¹⁴
5. The training mode is Batch, which is computationally more efficient than the sequential version (Vesanto et al., 2000).
6. The matrix W obtained converges to the weights $w_{r,p}$, with $r = 1 \dots 442$ and $p = 1 \dots 57$.¹⁵

Fig. 2 compares the distribution of RIR_p showing quite similar results for two initialization methods (random and linear), two shapes of the lattice (rectangular and hexagonal) and different map grid dimensions (20 100 300 and 442). In Table 1 two measures to evaluate the quality of the SOM are benchmarked by varying the learning parameters, the map topology and the map size. QE is the average distance between each data vector and its PE, and reveals how good the map can fit the input data (continuity). Unlike the average QE, TE considers the structure of the map and measures the topological preservation (resolution), calculating the proportion of all data vectors for which first and second PEs are not adjacent vectors (Kiviluoto, 1996). In our experiment we decided on a medium size map (442 PEs) based on Kohonen approach, since small maps present higher errors and for large maps unnecessary folds occur. Hexagonal lattice is better than rectangular with larger SOM, since obviously offers more continuity. Linear initialization is more functional to the aim of the study because guarantee the stability of the results.

According to Vesanto (1999), the most widely used technique to show the cluster structure of the SOM is the Unified Distance Matrix (U-Matrix). A matrix of distances between the weight vectors of adjacent PEs on the map is calculated and visualized (Fig. 3). Black hexagons indicate the location of the map units ($k = 442$) and hexagons between them show the actual values of the U-Matrix in the high-dimensional spaces. Larger black hexagons highlight PEs with higher number of samples. The RGB code measures the distance between PEs: cold colors (blue) suggest similarity and topological proximity, while warm colors (red) point out higher diversity and distance among samples. Clusters of samples can be identified on the grid as blue areas with red borders. The U-Matrix is relatively homogeneous in its bottom part, but in the middle and upper part presents several aggregates of PEs, like islands emerging from the sea. Component Planes (CPs) is a

representation that visualizes relative component values in the weight vectors of the SOM (Kohonen, 2001). We employ CPs to each of the 57 RIR_p observing the distribution of their values over the KL. The CP in Fig. 3 describes the RIR_p of the manufacturing of clothing, preparation and dyeing of fur (Clothing Industry),¹⁶ showing two disjoint patterns of PEs where the $RIR_{p=18}$ values are especially high. PEs with high $RIR_{p=18}$ values are grouped in two distinct neighborhoods and point out two distinct patterns of industry mix (CLO1 and CLO2).

4.2. Concentration vs. RIR

This paragraph offers a benchmark between high values of $RIR_{r,p}$ and G_i detected on each area i in the same industry p . In particular, the following comparative analysis is aimed at:

1. characterizing the industrial patterns of agglomeration detected on the basis of the $RIR_{p=18}$;
2. studying the differences between agglomerations based on $RIR_{p=18}$ and those detected using a concentration index such as Gini G_i ;
3. evaluating the differences across various agglomerations based on $RIR_{p=18}$.

The experiment was conducted as follow:

1. We select one out of the 57 industrial divisions. In particular, the choice of the Clothing Industry ($p = 18$) points at one of the most famous manufacturing industries of the Made in Italy, which typically tends to agglomerate in IDs.
2. According to Eq. (17) we obtain the $RIR_{r,p}$ for each i^{th} municipality belonging to the same PE_r .
3. According to Eq. (2) we calculate

$$G_i = \left(\frac{z_i^p}{Z^p} - \frac{\sum_p z_i^p}{\sum_p Z^p} \right)^2. \quad (21)$$

with $i = 1 \dots 8101$ Italian municipalities.

4. We compare the first 200 municipalities with the highest value of $RIR_{r,p}$ as well as G_i .
5. We highlight the 200 municipalities sample onto the Italian geographical map to single out spatial agglomerations. Since two distinct neighborhoods emerged in Fig. 3, they are visualized in different colors (CLO1 in blue and CLO2 in green).

Plotting on the map of Italy the values of $RIR_{p=18}$ for each 8101 municipalities (Fig. 4), we can observe spatial agglomerations driven by municipalities with the highest $RIR_{p=18}$ (black spots on a gray scale). Although the RIR_p has not been built on any geographical information, Fig. 4 shows a clear correspondence between industrial patterns and forms of spatial agglomeration. Municipalities which allocate alike their labour factor across industries belong to the same PE_r .

The benchmark analysis compares the two approaches to explore the corresponding forms of agglomeration. Concentration-based agglomeration is determined by one specific industry (e.g. the Clothing Industry) and is independent from the characteristics of the remaining 56 industries. This type of agglomeration is highly specialized and has a very high concentration of employees working in that industry. RIR-based agglomeration is defined by patterns of the corresponding allocations of employees across industries. This second type of agglomeration reflects analogies in the overall structure of industry across the municipalities, which

¹⁴ In general, this approach guarantees more stable results in terms of PEs neighborhood.

¹⁵ j and p have the same length $N = 57$ (the number of industries of the input matrix Y).

¹⁶ In the 2-digit ATECO 2002, this industry corresponds to the industrial division $p = 18$.

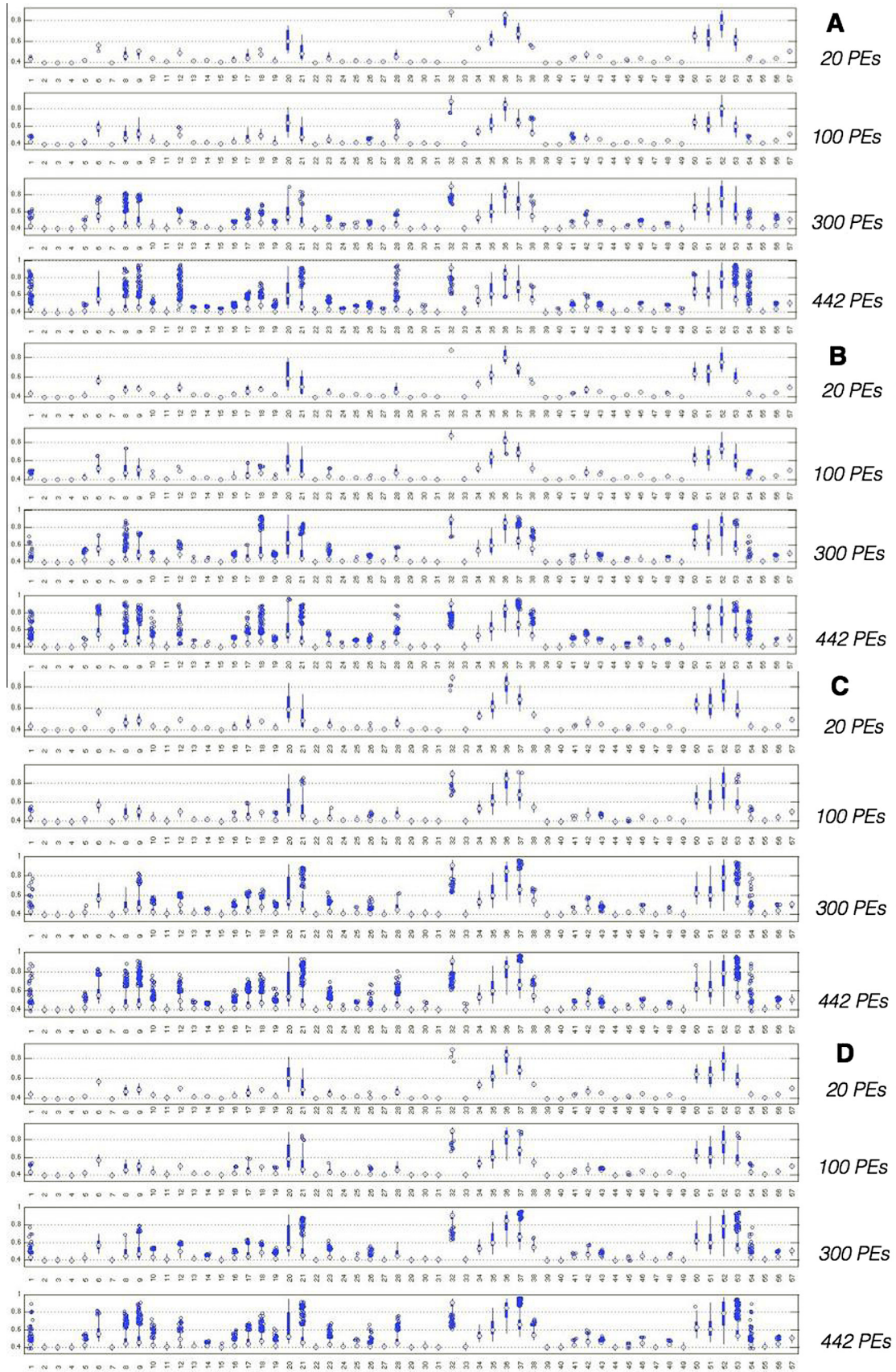


Fig. 2. Boxplot analysis of the 57 RIRp. (A) random initialization with rectangular shape, (B) random initialization with hexagonal shape, (C) linear initialization with rectangular shape, (D) linear initialization with hexagonal shape.

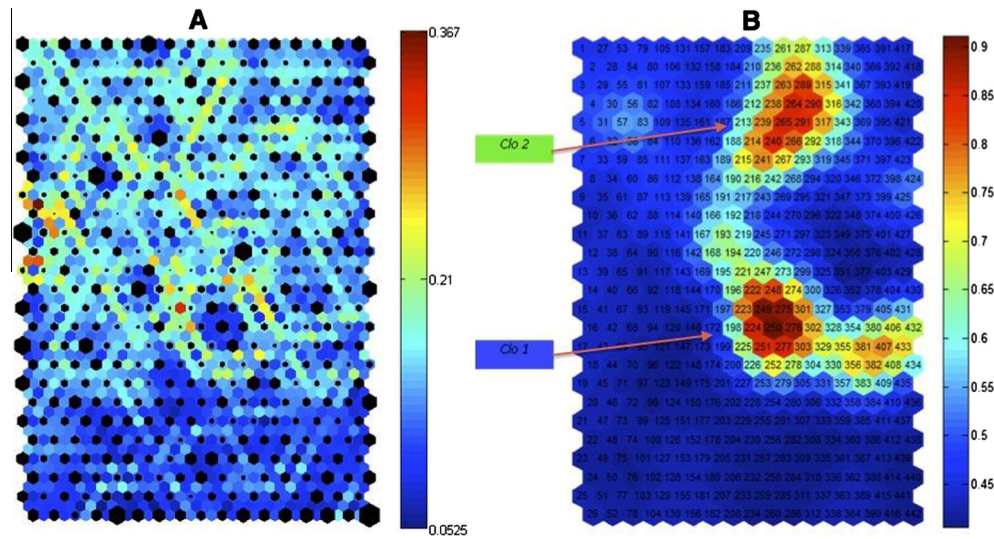


Fig. 3. (A) Unified Distance Matrix (U-Matrix) with density of municipalities in the PEs (RGB code) and (B) the Component Plane for the Clothing Industry (ATECO 2002, $p = 18$, ISTAT Census 2001) in RGB code. Two main clusters ($CLO1$ and $CLO2$) clearly emerge with a high RIR value (red hexagons). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

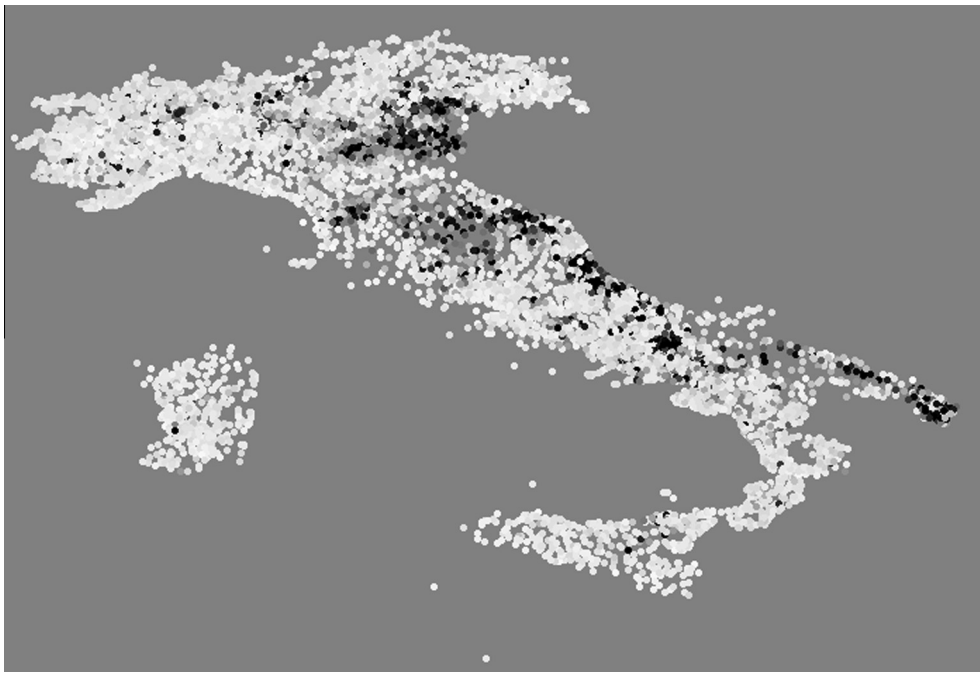


Fig. 4. Distribution of the $RIR_{p=18}$ (Clothing Industry) across Italian municipalities in gray scale (ATECO 2002, n.18, ISTAT Census 2001).

does not necessarily imply the existence of a predominant industry in absolute terms (i.e. in terms of high levels of concentration). In some areas the two forms of agglomeration are in geographical proximity and sometimes they overlap.

At least three distinct typologies emerge from this analysis. The first type (A) relates to situations where there is an exclusive presence of concentration-based agglomerations. Starting from Northern Italy, the area between the provinces of Vicenza and Treviso in Veneto, a region with an established tradition of IDs, presents a remarkable concentration of employees in the Clothing Industry, and a scarcity of municipalities that display similarities in terms of RIR values (Fig. 5). This type of agglomeration, based on concentration only, can be defined as a *specialized vertical district*: that is, an ID characterized by a clear prevalence of vertically

integrated SMEs. In these single-focus, vertically integrated areas, only a few enterprises in the industry operate in the final market, most being sub-contractors of and providers for other firms in the spatial agglomeration.

The second type (B) includes only similarity-based spatial agglomerations and no or very few municipalities with high concentrations of employment in the Clothing Industry. For instance, the area between the provinces of Avellino and Benevento in Campania, a southern region with a very low level of classical ID entrepreneurial culture, would not register as hosting an ID were it to be analyzed exclusively in terms of industrial concentration. Using the industrial pattern analysis we can identify a dense spatial agglomeration of municipalities with high homogeneity of labour factor allocation (Fig. 6). These industrial patterns, not

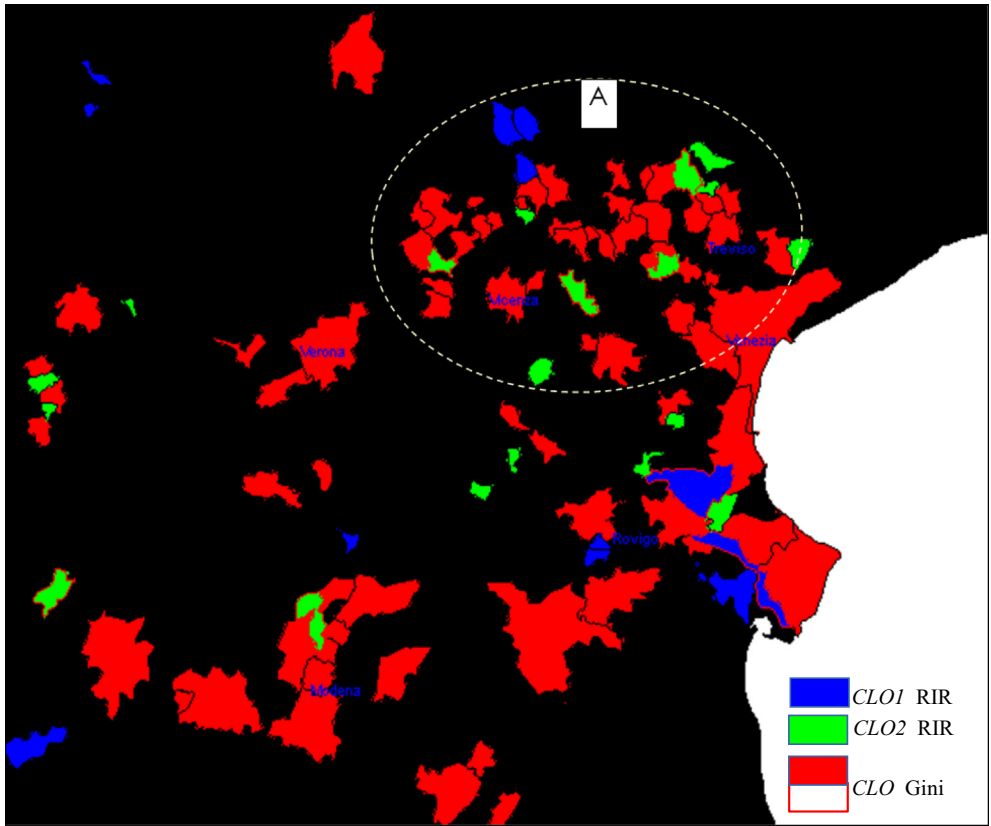


Fig. 5. Benchmarking Gini and RIR agglomerations in Clothing Industry: the North-Eastern Italy. A-type agglomerations are mainly based on the Gini Index.

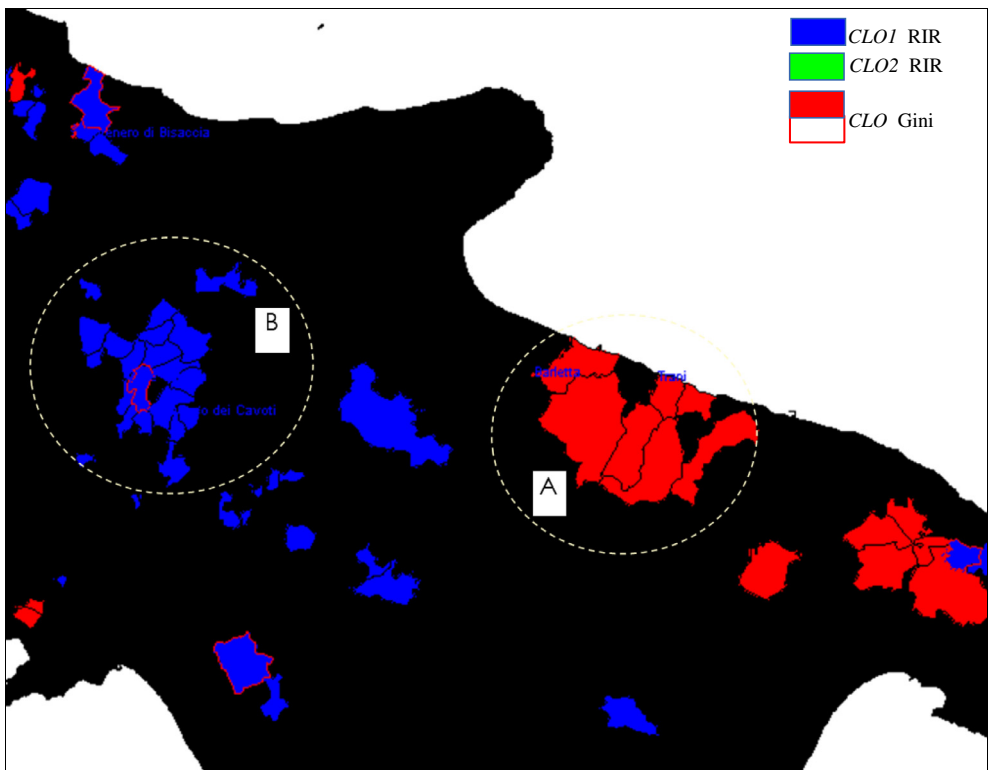


Fig. 6. Benchmarking Gini and RIR agglomerations in the Clothing Industry: Campania and Puglia regions. B-type agglomerations are mainly based on the RIR Index.

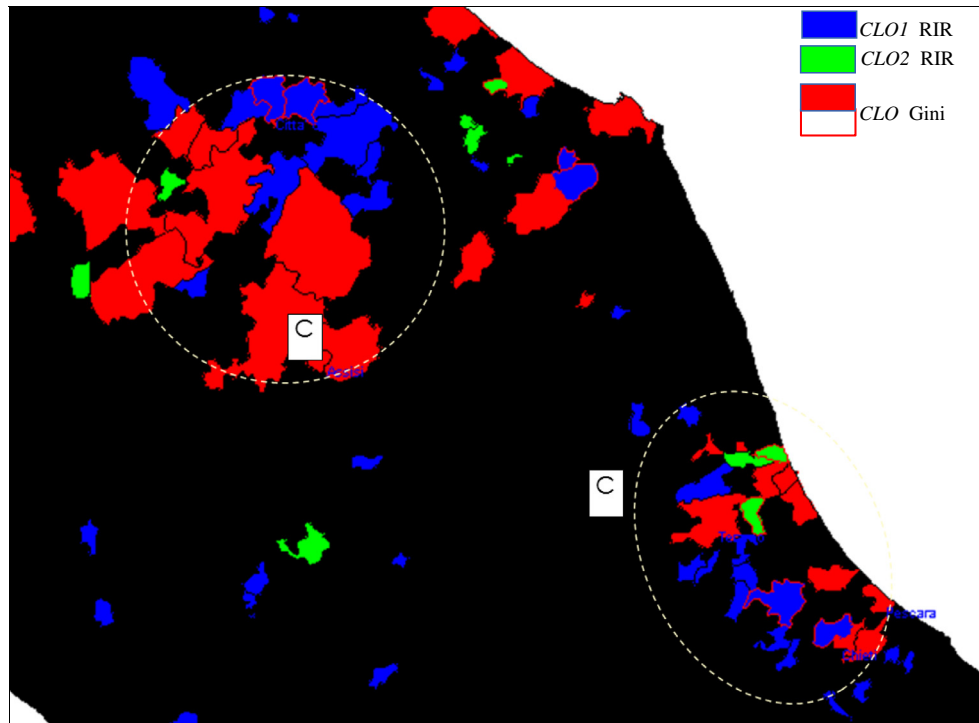


Fig. 7. Benchmarking Gini and RIR agglomerations in the Clothing Industry: the Central Italy. C-type agglomerations are hybrid territories where Gini and RIR municipalities seem to be complementary.

identified by concentration analysis, emerge as distinctive spatial agglomerations can be grouped under the umbrella of *complex horizontal district*. In such regions the Clothing Industry is not sufficiently concentrated to drive the local economy on its own or to qualify as a national champion, but nevertheless plays a quite leading role in the local economy due to its structural complementarity with other activities which highlights some form of horizontal integration.

The third type (C) relates to situations of partial overlap between concentration-based and RIR-based agglomerations, that is, *complementary district*. For example, the regions of Umbria and Abruzzo, in the center of the peninsula, have a relatively well established presence of traditional concentration-based IDs, but only in certain parts of their territories. We can identify high concentration-based agglomerations in the coastal area of Abruzzo as well as around Perugia in Umbria, and significant levels of RIR-based agglomerations in the Teramo-Chieti hinterland in Abruzzo, as well as around Città di Castello in Umbria. Complementary district corresponds to agglomerations with several functional specializations in the northern parts of both regions, and more structurally uniform agglomerations in the southern parts (Fig. 7). The preliminary analysis at the national level suggests a fairly wide diffusion of these hybrid patterns in several industries.

This kind of analysis can be applied to different territorial scales, macro-regional systems, regional systems, cluster systems, inter-county clusters, remote clusters, urban and suburban belts. Assuming the availability of appropriate data, they can be exploited to study local development paths at the desired level of resolution. The difference from traditional approaches is that the analysis includes multiple features and does not focus on a single geographical or sector-based point of view, but is able to explore and characterize the whole spectrum of industrial interdependencies.

5. Conclusion

This paper contributes to investigate the process of regionalization of economic activities based on the geographic distribution of labour factor as a proxy of the local supply of human capital. Patterns of agglomeration are neither theoretically fixed nor methodologically predetermined, since data are processed by searching for bottom-up aggregates emerging from the adaptation to changing market and non-market conditions in the local economy. The taxonomy of agglomerations proposed by using this SOM-based approach is extracted endogenously from the available data. We have shown that identifying agglomerations based on the concentration of a particular industry in a specific location does not take account of the fact that spatial patterns of agglomeration can shape coherent industrial structures in several different ways.

The contribution provided by this paper should be considered methodological: the primary goal was to illustrate how SOM techniques enable analysis of similarity in terms of supply of local human capital, which allows the definition and visualization of the industrial structures of spatial agglomeration not emerging from concentration-based analysis. The empirical results should be taken as examples; they refer only to the case of the Italian Clothing Industry. However, even this rather specific case shows that industrial patterns can be as relevant as concentration-based agglomerations for understanding some properties of the spatial organization of Italian IDs. Especially relevant is that the nature of these spatial agglomerations is determined by the whole allocation structure of labour factor in the local economy and not just by the characteristics of the prevailing industry, possibly benchmarked against country averages.

Also, the research highlights the difficulty involved in analyzing IDs on the basis of a strictly reductionist approach that attempts to

explain their formation, dynamics and structural evolution in terms of returns to scale, agglomeration and localization economies. This problem cannot be resolved by extending the variables to include knowledge-related assets in order to interpret the transfer of knowledge or failures to reach a critical mass of human capital. The limitations to the circulation of knowledge are not technological or infrastructural in nature, but are related to the fact that knowledge assets are socially situated, managed, and transmitted (Hakanson, 2005), and that their effectiveness depends on the successful and context-specific integration of social and economic factors. The attraction and retention of human capital is related to problems of social sustainability and quality of life (Rappaport, 2009), which do not find easily causal explanations (Storper and Scott, 2009). These sources of non-economic structural complexity encompass alternative organizational principles leading to further types of agglomerations.

Prospective analytical works could provide a broader characterization of Italian agglomeration typologies starting from an in-depth analysis of the complete spectrum of Italian industries and their structural interdependencies. We would expect that the repertoire of types would increase. It is possible that different municipalities belonging to the same similarity-based agglomeration might play complementary and subtly differentiated roles. The fundamental opposition between concentration-based and RIR-based agglomerations may not be exhaustive and alternative organizational principles might be identified with an open-minded and technically appropriate scanning of the available evidence.

References

- Agarwal, P., Skupin, A. (Eds.), 2008. *Self-Organising Maps. Applications in Geographical Information Science*. Wiley, Chichester.
- Becattini, G., 1990. The Marshallian industrial district as a socio-economic notion. In: Pyke, F., Becattini, G., Sengenberger, W. (Eds.), *Industrial Districts and Inter-Firm Cooperation in Italy*, International Institute for Labor Studies, Geneva, pp. 75–107.
- Becattini, G., Bellandi, M., Dei Ottati, G., Sforzi, M. (Eds.), 2003. *From Industrial Districts to Local Development: An Itinerary of Research*. Edward Elgar.
- Birckenbach, F., Bode, E., 2008. Disproportionality measures of concentration specialization and localization. *Int. Regional Sci. Rev.* 31 (4), 359–388.
- Calafati, A., Compagnucci, F., 2005. Oltre i sistemi locali del lavoro. *Economia Marche* 24, 51–76.
- Cutrini, E., 2010. specialization and concentration from a twofold geographical perspective: evidence from Europe. *Regional Stud.* 44 (3), 315–336.
- De Blasio, G., Omiccioli, F., Signorini, 2009. Measuring the district effect. In: Becattini, G., Bellandi, M., De Propris, L. (Eds.), *A Handbook of Industrial Districts*. Edward Elgar.
- Ellison, G., Glaser, E., 1997. Geographic concentration in us manufacturing industries: a dartboard approach. *J. Political Economy* 105 (5), 889–927.
- Hakanson, L., 2005. Epistemic communities and cluster dynamics: on the role of knowledge in industrial districts. *Ind. Innov.* 12 (4), 433–463.
- Iuzzolino, G., 2003. Identifying the geographical agglomerations of manufacturing industries. In: *Local Economies and Internationalization in Italy Conference*. Available at: <http://ssrn.com/abstract=2160413> or <http://dx.doi.org/10.2139/ssrn.2160413>.
- Jacobs, J., 1969. *The Economy of Cities*. Vintage Books, New York.
- Kiviluoto, K., 1996. Topology preservation in self-organizing maps. In: *IEEE International Conference on Neural Networks*. IEEE, pp. 294–299.
- Kohonen, T., 2001. *Self-Organizing Maps*, third ed. Springer Verlag, Berlin, Heidelberg.
- Patuelli, R., Reggiani, A., Nijkamp, P., Blien, U., 2006. New neural network methods for forecasting regional employment: an analysis of German labour markets. *Spatial Econ. Anal.* 1 (1), 7–30.
- Rappaport, J., 2009. The increasing importance of quality of life. *J. Econ. Geography* 9 (6), 779–804.
- Rosenthal, S., Strange, W., 2001. The determinants of agglomeration. *J. Urban Econ.* 50 (2), 191–229.
- Sforzi, F., 1990. The quantitative importance of Marshallian Industrial Districts in the Italian Economy. In: Pyke, F., Becattini, G., Sengenberger, W. (Eds.), *Industrial Districts and Inter-Firm Cooperation in Italy*, International Institute for Labor Studies, Geneva, pp. 75–107.
- Sforzi, F., 2009. The empirical evidence of industrial districts in Italy. In: Becattini, G., Bellandi, M., De Propris, L. (Eds.), *A Handbook of Industrial Districts*. Edward Elgar, pp. 327–342.
- Sforzi, F., Lorenzini, F., 2002. I distretti industriali. In: *L'esperienza italiana dei distretti industriali*, IPI and Ministry of Productive Activities, Rome, pp. 20–33.
- Storper, M., Scott, A.J., 2009. Rethinking human capital, creativity and urban growth. *J. Econ. Geography* 9 (2), 147–167.
- Tattara, G., 2001. L'efficienza dei distretti industriali: una ricerca condotta dal servizio studi della Banca d'Italia. *Economia e Società Regionale* 16 (4), 114–144.
- Tattara, G., 2002. L'efficienza dei distretti industriali: una risposta a Omiccioli e Signorini. *Economia e Società Regionale* 17 (1), 260–264.
- Van der Panne, G., 2004. Agglomeration externalities: Marshall versus Jacobs. *J. Evol. Econ.* 14 (5), 593–604.
- Vesanto, J., 1999. SOM-based data visualization methods. *Intell. Data Anal.* 3 (2), 111–126.
- Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. *SOM Toolbox for Matlab 5*, report, Helsinki Univ. of Technol., Helsinki, Finland. Available at: <http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf>.
- Viesti, G., 2005. Distretti industriali e agglomerazioni territoriali in Italia. Lo stato delle conoscenze e i problemi di ricerca. *Argomenti* 14, 85–116.