Social Indicators Research (2024) 173:9–51
https://doi.org/10.1007/s11205-021-02836-y

ORIGINAL RESEARCH

Measuring Competitiveness at NUTS3 Level and Territorial Partitioning of the Italian Provinces

Pierpaolo D'Urso[1] · **Livia De Giovanni**[2] · **Francesca G. M. Sica**[3] · **Vincenzina Vitale**[1]

Accepted: 17 October 2021 / Published online: 6 October 2022

Abstract

In this paper we propose a dashboard of indicators of territorial attractiveness at NUTS3 level in the framework of the EU Regional Competitiveness Index (RCI). Then, the Fuzzy C-Medoids Clustering model with multivariate data and *contiguity* constraints is applied for partitioning the Italian provinces (NUTS3). The novelty is the territorial level analized, and the identification of the elementary indicators at the basis of the construction of the eleven composite competitiveness pillars. The positioning of the Italian provinces is deeply analyzed. The clusters obtained with and without contraints are compared. The obtained partition may play an important role in the design of policies at the NUTS3 level, a route already considered by the Italian government. The analysis developed and the related set of indicators at NUTS3 level constitute an information base that could be effectively used for the implementation of the National Recovery and Resilience Plan (NRRP).

Keywords Competitiveness · Territorial attractiveness · NUTS3 · Spatial constraints · Fuzzy partitioning around medoids

✉ Livia De Giovanni
ldegiovanni@luiss.it

Pierpaolo D'Urso
pierpaolo.durso@uniroma1.it

Francesca G. M. Sica
f.sica@confindustria.it

Vincenzina Vitale
vincenzina.vitale@uniroma1.it

[1] Department of Social Sciences and Economics, Sapienza University of Rome, Rome, Italy

[2] Department of Political Sciences, Luiss university and Data Lab Luiss, Rome, Italy

[3] Economic Research Department, Confindustria and Data Lab Luiss, Rome, Italy

1 Introduction

The term "territorial attractiveness" is a binomial shared by economists and economic geographers to identify a series of assets with which the territories are equipped. The intensity of individual assets and a favorable combination of different assets can represent an attractive factor to direct preferences towards a given territory rather than another for residential and productive settlements, respectively of private citizens (residential attractiveness) of foreign and national investors (productive attractiveness). Less universally accepted is the use, or rather the abuse of the concept of territorial competitiveness. Unlike the concepts of "utility" and "efficiency", competitiveness is not a basic construct in economics and analyses of competitiveness have in general no fundamentals that are strictly anchored to economic theory. From a macroeconomic point of view, various official definitions of territorial (country) competitiveness can be found featuring at least one of the following elements: economic performance, in terms of productivity growth rate and real income; international trade in goods and services; sustainability, understood as long-term sustainable achievements. In the European Competitiveness Report (2000) we find the following: "An economy is competitive if its population can enjoy high and rising standards of living and high employment on a sustainable basis. More precisely, the level of economic activity should not cause an unsustainable external balance of the economy nor should it compromise the welfare of future generations". If at the sectoral level the adaptation of the concept does not present any problems whatsoever, at the macroeconomic level some conceptual dyscrasias arise. The basic idea of the supporters of extending the micro concept of corporate competitiveness to the whole country is that this can be considered as the sum of the companies that operate there, or as a single large company that is operating on international markets with an ever increasing number of competitors (Porter 2004; Rucinska and Rucinsky 2007). It is precisely because of the similarity between company and country that economists consider the translation of the concept from the micro to the macro level as unacceptable (Krugman 1994). On a closer inspection, the implicit analogy between business and territory is for many economists meaningless, as competition between countries cannot, for obvious reasons, lead to the expulsion or suppression of the less competitive ones (Krugman 1994). On the contrary, the success of a territory (like a country or a region within a country) may in general benefit its neighboring territories thanks to the effects of positive spillovers. In essence, the competitive game between countries is not zero-sum, but rather a plus-sum game. The success of a country or region creates more than destroys the opportunities for others and as known, trade among nations is not a game "without results" (Psofogiorgos and Metaxas 2016).[1] The concept of regional competitiveness, adopted by the European Commission (EC) when drawing up the Regional Competitiveness Index (RCI, from now on), lies somewhere between the microeconomic concept (firm) and the

[1] Among the economists, the fiercest opponent to the concept of competitiveness of a country is Paul Krugman (Krugman 1994) who defines country competitiveness as a dangerous obsession with politicians when they claim to put it at the top of their priority agenda. The main argument of the MIT professor is that competitiveness is in itself an empty word and acquires its meaning only by referring to productivity ("... a poetic way of saying productivity ... "). In fact, the most commonly used single indicator of competitiveness at the country level is the labour cost per unit of product (ULC) calculated as the ratio between unit labour cost (per worker or per hour worked) and labor productivity (added value for worker or per hour worked). If productivity is certainly to be considered as a key factor of a country's competitiveness, the link between competitiveness and well-being is a mutual one. Empirical evidence highlights the virtuous circle between productivity-competitiveness-income per capita, considering that the most competitive countries in international rankings are also those characterized by a higher standard of living measured by per capita income.

macroeconomic one (country). “Regional competitiveness can be defined as the ability to offer an attractive and sustainable environment for firms and residents to live and work.” If, therefore, competitiveness is the ability to offer an attractive environment, then the two concepts of competitiveness and attractiveness end up merging into one another (Davies et al. 2000).

The measures of attractiveness proposed here, the “pillars”, represent dimensions or aspects of attratctiveness. Each pillar is obtained through techniques of multivariate statistical analysis as the synthesis of a plurality of indicators, so that both the causes, input, and the effects, outcome, of attractiveness in the territory are captured transversally. The comparative evaluation makes it possible to carry out a precise anamnesis of the territory through the “components” of the pillars and then to define the “cure” with the formulation of policy proposals tailored to each territory. The methodological approach for the construction of the pillars is not new, but has been borrowed from the Regional Competitiveness Index (RCI) of the European Commission. The originality of the work consists in the lower territorial level, that has influenced the choice of indicators within each pillar.

Unfortunately, the information available at the territorial level provided by official statistics is published in different databases depending on the topic and is therefore dispersed in many information sources. And yet they are fundamental for an exhaustive and in-depth reading of local specificities. Local specificities are preparatory to the formulation of local policies aimed at raising the potential attractiveness.

The clustering procedure adopted enjoys the benefits connected to Fuzzy clustering and to Partitioning Around Medoids (PAM). Due to the difficulty in identifying a clear boundary between clusters in real applications involving territorial units, i.e. provinces even belonging to the same region, fuzzy clustering is more attractive than the hard clustering methods (D’Urso 2014. The PAM approach allows for more appealing and easy to interpret results of the final partition (Kaufman and Rousseeuw 2005), determining real and not virtual representatives of the clusters.

In this paper we propose a dashboard of indicators of territorial attractiveness at NUTS3 level in the framework of the EU Regional Competitiveness Index (RCI). Then, the Fuzzy C-Medoids Clustering model with multivariate data and *contiguity* constraints is applied for partitioning the Italian provinces (NUTS3). The novelty is the territorial level analized, and the identification of the elementary indicators at the basis of the construction of the eleven composite competitiveness pillars.

The paper is structured as follows. In Sect. 2 the competitiveness indicators at NUTS3 level and related pre-processing are presented. In Sect. 3 the clustering model is introduced. In Sect. 4 the model is used for clustering the Italian provinces. Section 5 presents the Conclusions.

2 Indicator of Competiveness at NUTS3 Level (Provinces)

The Regional Competitiveness Index (RCI) (Annoni and Dijkstra 2019) is composed of eleven pillars that describe the different aspects of competitiveness. They are classified into three groups (subindexes): *Basic*, *Efficiency* and *Innovation*.

The Basic group includes five pillars: (1) Institutions; (2) Macroeconomic Stability; (3) Infrastructure; (4) Health; (5) Basic Education. These represent the key basic drivers of all types of economies.

The Efficiency group includes three pillars: (6) Higher Education; (7) Labor Market Efficiency; (8) Market Size.

The Innovation group includes three pillars: (9) Technological Readiness; (10) Business Sophistication; (11) Innovation.

The pillars are composite variables. The complete list of all candidate indicators at the NUTS2 level can be found in The EU Regional Competitiveness index 2019 (Annoni and Dijkstra 2019). The partition of the European regions (NUTS2) with respect to the Basic, Efficiency and Innovation subindexes has been analized in D'Urso et al. (2019b).

In the data warahouse of the National Institute of Statistics (Istat) there is no theme specifically dedicated to the territory but it is possible to download from each macro theme the territorial detail through the customization options of the default layout and analyze the phenomena of interest from a triple perspective:

- Spatial: to analyze the relative positioning of the territories (regions and provinces);
- Temporal: to grasp the evolution of a given phenomenon over time at a national and territorial level (region or province);
- Sectoral: to analyze productive specialization and its territorial articulation.

For this reason, the collection of quantitative territorial data at the provincial level ("NUTS3" European glossary, "Small regions" OECD glossary) was the most challenging phase of this analysis due to the difficulty of finding updated and transversal data on the various themes of interest in a single source of information. Thanks to the fusion of a number of official national (Istat, Unioncamere, Bank of Italy, Cnel) and international (Eurostat, OECD) information sources, the number of variables collected was quite large, but the creation of a complete territorial database required careful prior selection based on the criterion of relevance to the eleven dimensions chosen to describe the phenomenon of attractiveness. In the end, over 150 indicators were selected for each territorial unit and catalogued in each pillar. This second phase of systematization of the data collected was easier because it was possible to move along a path already traced and regularly updated in scientific work in Europe. The selection of the elementary indicators and their subsequent cataloguing within the pillars was inspired, in fact, by the methodology published in the reports of the European Commission to calculate the RCI (Annoni and Dijkstra 2019) and of the Word Economic Forum to calculate the Global Competitiveness Index. The originality of this study is twofold and consists, on the one hand, in having replicated at the NUTS3 provincial level the measurement approach now consolidated at the regional level (NUTS2) and, on the other, in having included exclusively indicators referring to the provinces. It must be said that this has been made possible by the Istat initiative to elaborate Equitable and Sustainable Well-being not only at the national level but also at the level of the territories (BES of the territories) thanks to which a rich set of indicators for each of the twelve domains in which the BES has been articulated has been made available to the government and citizens with coverage of all 110 provincial administrative units.

In the paper, to obtain the pillars, the RCI methodology is used.

Firstly the indicators describing each of the eleven attractiveness aspects for the Italian provinces are identified. To correct for different range and measurement units, weighted z-scores are adopted using the provinces' population sizes as weights. The Principal Component Analysis (PCA) is used to select the indicators within each pillar. Then the eleven pillars are computed as a simple average of the selected indicators in each pillar, and next the subindexes Basic, Efficiency, Innovation are computed as a simple average of the pillars in each subindex. The use of simple averages in the two steps is based on the Principal Component Analysis, used to check for the internal consistency of the indicators within each pillar and to determine the sign (positive or negative) of the indicators. The conditions to be verified to use only one pillar - obtained as a simple average of the indicators measuring that pillar - are that each pillar shows a unique, most relevant principal component accounting for a large amount of variance and that all the indicators contribute to approximatively the same extent to the first principal component.

The sources utilized are institutes of official statistics with the exception of "Fondazione Etica su dati Amministrazione Trasparente".[2]

The selected indicators in the pillars of the Basic group are presented in Table 1.

Pillar I - Institutions Recognition of the role of institutions in shaping a country's "fate" has gained relevance as a result of a new strand of research that identifies institutions as another cause of differentials in the development rates of economies in addition to traditional factors (Acemoglu et al. 2001). The empirical literature has emphasized the links between institutional soundness and the following aspects of an economic system: resolution of market failures and improved efficiency (Streeck and Schmitter 1991); reduction in transaction costs (North 1990); stimulation of innovation and productivity (Putnam 2000).

What are Institutions? According to Douglass North (North 1990): "are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction". Two important characteristics emerge from the definition: 1. the human component ("humanly devised") that overlaps with other factors such as natural geographic factors that are beyond human control; 2. constraints on human behavior ("the rules of the game" setting "constraints" on human behavior). Candidate indicators to measure the "institutions" dimension must be able to capture the quality and efficiency of institutions and the regulatory environment that impacts on the ease of "doing business". Other indicators capture the phenomenon of corruption through an *ad hoc* module included by Istat in the 2015-2016 Citizens' Security Survey (NUTS3 level).

Pillar II - Macroeconomic stability A situation of sound finances at the local level is essential for the public operator to receive confidence in its solvency from private operators, whether they are consumers or producers of goods and services. The risk of financial imbalances impacts on confidence which is, in turn, crucial to raising the rate of investment in the long term, a fundamental ingredient for preserving the competitiveness of an area.

Pillar III - Infrastructure The fourth industrial revolution is making possible, thanks to digital technology, a closer connection between production systems located in different places. This paradigm shift influences the competitiveness factors of the territories by making logistics enter the top ten of the winning elements, not only as storage and sorting, but increasingly as an ancillary and accessory service to production and as an advanced service with high technological content. Modern and functioning infrastructures contribute in fact to increase the economic efficiency and the social equity through the maximization of local economic potential (Rodriguez-Pose and Crescenzi 2008). In addition, they promote accessibility to other regions and countries, contributing to the integration of peripheral areas. Others authors (Lopez-Claros et al. 2007) emphasize the key role of infrastructure in determining the location of economic activities and in influencing the development of certain types of productive activities. The impact on the competitiveness of territories is conveyed by the increase in economic efficiency.

Pillar IV - Health Health is a crucial dimension for the well-being of the citizens who reside in a territory and for this reason an *ad hoc* pillar is dedicated to it that describes the

[2] I.stat (Istat), Istat (Bes), Istat (Indicatori territoriali per le Politiche di sviluppo), Istat (A misura di comune), Istat (Companies Permanent Census), Istat (ASIA), Istat (COEWEB), Istat (BES), Minister of Justice (DG-STAT), Fondazione Etica su dati Amministrazione Trasparente, OECD (PISA), Minister of Economic Development, UIBM database.

Table 1 Indicators of the subindex Basic

Subindex	Pillar	Indicator (source, year)
Basic	Institutions	Pending trials (2016) (reversed) (Ministry of Justice, DG-STAT)
Basic	Institutions	Trial duration (reversed) (2016) (Ministry of Justice, DG-STAT)
Basic	Institutions	Vote participation (BES-Istat, average 2004, 2009, 2014, 2019)
Basic	Institutions	Female municipal administrators out of total local administrators (BES-Istat, 2018)
Basic	Institutions	Social relations intensity: Non profit organizations, per 10000 population (BES-Istat, 2017)
Basic	Institutions	Administrative capacity (NUTS3 level) (Fondazione Etica su dati Amministrazione Trasparente, 2020)
Basic	Institutions	Corruption Last 3 years (I.stat-Justice and Security, 2015–2016))
Basic	Institutions	Bribe Health (I.stat-Justice and Security, 2015–2016)
Basic	Institutions	Bribe Assistance (I.stat-Justice and Security, 2015–2016)
Basic	Institutions	Bribe Education (I.stat-Justice and Security, 2015–2016)
Basic	Institutions	Bribe Job (I.stat-Justice and Security, 2015–2016)
Basic	Institutions	Bribe administration (I.stat-Justice and Security, 2015–2016)
Basic	Macroeconomic stability	Surplus (deficit) of administration in relation to current revenues (I.stat Public Adm. and Private Inst., 2017)
Basic	Macroeconomic stability	Collection capacity (BES-Istat, 2017)
Basic	Macroeconomic stability	Interest expenses in relation to current revenues (reversed) (I.stat Public Adm. and Private Inst., 2017)
Basic	Infrastructure	Accessibility (travel times) index towards urban and logistic nodes (reversed) (Istat Indicators for Development, 2013)
Basic	Infrastructure	Seats km offered by all modes of transport per inhabitant (BES-Istat, 2018)
Basic	Infrastructure	Annual passenger density in local public transport and airports per inhabitant (BES-Istat,, 2017)
Basic	Infrastructure	Car-sharing: availability of vehicles per 100 thousand inhabitants (Istat-Urban environment, 2017)
Basic	Health	Life expectancy at birth, average number of years (BES-Istat, 2018)
Basic	Health	Infant mortality per 1.000 live births (BES-Istat, 2017)
Basic	Health	Cancer mortality (20–64 years) - standardized rates per 10.000 residents (reversed) (BES-Istat, 2017)
Basic	Health	Hospital outmigration to other region for ordinary acute hospitalizations (BES-Istat, 2018) (reversed)
Basic	Basic education	Vocational (vocational) graduates: technical and vocational graduates (Eurostat), 2018
Basic	Basic education	Students' reading proficiency level - mean score (OECD - PISA, 2018)
Basic	Basic education	Students' numeracy proficiency level - mean score (OECD - PISA, 2018)
Basic	Basic education	Underachievement rate in reading (reversed) (Invalsi, 2019)
Basic	Basic education	Underachievement rate in numeracy (reversed) Invalsi, 2019)

health conditions of the population. A healthy workforce is a key factor for the increase of the rate of activity in the labor market and for the increase in labor productivity at the regional and national levels (Official Journal of the European Union). Of course, the link with competitiveness is indirect in that mediated by the impact of healthy living conditions.

Pillar V - Basic Education Unlike the availability of natural resources, the endowment of human capital of an area, is not fixed but can be increased by investing in education which, in turn, produces a return that from the private point of view proves to be higher than other forms of investment available to households, who must decide how to allocate their financial capital between alternative investments (Coleman 1988). There are a number of empirical studies demonstrating the existence of a positive association between educational quality and economic growth (Hanushek and Woessmann 2007). International tests of learning outcomes from primary school to adults at work aim to capture the quality of the human capital compared to quantitative measures. There are also empirical evidences that adult competences applied at work enhance labor productivity at company level and activate the virtuous circle from human capital to a strong, sustainable and balanced growth by disseminating new technologies and work-organization practices. The transition from a traditional knowledge-based to a competence-based educational-training system is by now unavoidable. The quality of education is measured by the results obtained in cognitive tests, whose purpose is to assess not only "knowledge" but also theoretical knowledge. The most widely used test for measuring skills is PISA, which stands for Programme for International Student Assessment, an OECD initiative that, scheduled every three years, measures the reading, mathematics and science skills of 15-year-old students.

The selected indicators in the pillars of the Efficiency group are presented in Table 2.

Pillar VI - Higher Education The contribution of education to productivity and growth has been extensively studied. Knowledge and innovation-based economies need well-educated, adaptable human capital and an education system capable of transmitting not only theoretical knowledge but also practical skills and, hence, competencies. In a context increasingly permeated by knowledge, universities and businesses play a decisive role: the former because they are typically the places where knowledge is cultivated, accumulated and transmitted; the latter because they have the task of applying the results of research to production techniques, products and business organization.

Pillar VII - Labor Market Efficiency An efficient and flexible labor market favors an optimal allocation of resources (Lopez-Claros et al. 2007) which is reflected in the attractiveness of an area that is a precondition for its competitiveness understood as competition that is triggered between territories in order to catalyze the preferences of potential "users" of the area, as investors (new or existing) who must evaluate the best location for their production facilities, but also as citizens who must decide where to live. Employment and unemployment rates provide information on the level of activity in the local labor market, while a long-term unemployment rate is a symptom of the existence of structural problems. The differential in employment rates between women and men is an important aspect and signals a lack of reconciliation between work and family life, the burden of which falls on women who are often forced to leave the labor market and swell the ranks of the inactive.

Pillar VIII - Market size The pillar describes the potential outlet market available to firms: the larger the market, the greater the possibility of exploiting economies of scale and benefiting from the gains from them in terms of reduced fixed costs. Market size encourages entrepreneurship and fosters innovation. The problem is not so much the availability of a large market but rather the accessibility to it. The potential of the market is captured in terms of absolute values of population, Gross Domestic Product and spending capacity.

The selected indicators in the pillars of the Innovation group are presented in Table 3.

Pillar IX - Technological Readiness This dimension captures the degree to which households and businesses are using ICT technologies. The Fourth Industrial Revolution is

Table 2 Indicators of the subindex Efficiency

Subindex	Pillar	Indicator (source, year)
Efficiency	Higher education	Percentage incidence of tertiary graduates 25–39 (BES-Istat, 2019)
Efficiency	Higher education	Transition to tertiary education (BES-Istat, 2017)
Efficiency	Higher education	Life Long Learning (I.stat, 2018)
Efficiency	Higher education	Early school leavers (BES-Istat, 2017)
Efficiency	Higher education	Stem graduates (author elaboration on I.stat, 2018)
Efficiency	Labor market efficiency	Employment rate 15–64 years (I.stat, 2019)
Efficiency	Labor market efficiency	Gender Gap - employment rate (author elaboration on I.stat, 2019)
Efficiency	Labor market efficiency	Missing work participation rate (BES-Istat, 2019) (reversed)
Efficiency	Labor market efficiency	Gender Gap - missing work participation (author elaboration on BES-Istat, 2019) (reversed)
Efficiency	Labor market efficiency	Share 15–24 not in education, employment, training (NEET) (BES-Istat, 2019) (reversed)
Efficiency	Labor market efficiency	Labor productivity (author elaboration on I.stat, 2017)
Efficiency	Labor market efficiency	Formal Job (reversed) (BES-Istat, 2018)
Efficiency	Labor market efficiency	Fatal accidents at work (reversed) (BES-Istat, 2017)
Efficiency	Labor market efficiency	Wages of tertiary graduates (I.stat, 2017)
Efficiency	Market size	Provincial GDP year 2017-Constant prices base year 2015 (I.stat, 2017)
Efficiency	Market size	Population (I.stat, 2020)
Efficiency	Market size	Distance of 2017 GDP from pre-crisis GDP levels - index numbers 2007=100 (author elaboration on I.stat, 2017)
Efficiency	Market size	Potential market in terms of GDP provincial incidence on Italy GDP (author elaboration on I.stat, 2017)
Efficiency	Market size	Propensity to export (author elaboration on I.stat, 2017)
Efficiency	Market size	Propensity to import (author elaboration on I.stat, 2017)
Efficiency	Market size	Non-performing loans loans (BES-Istat, 2019) (reversed)

changing the way we produce under the banner of the three "v's": volume, velocity, variety. Increasingly high production volumes, greater speed in the production of goods and services and, finally, wider variety of products. Compared to previous revolutions, with digital technology both the time lapse between discovery, application and diffusion of innovations and the distance between things, people and countries have become much shorter thanks to connectivity. The way in which new information and communication technologies are used by a firm's workers depends closely on the degree of penetration and diffusion of these technologies in everyday life. Empirical evidence shows how the adoption and diffusion of ERP (Enterprise Resource Planning) and CRM (Customer Relationship Management) applications is strongly dependent on the size of the firm, but a crucial role is played by the level of education of employees rather than of the entrepreneur.

Pillar X - Business Sophistication The degree of maturity of the productive system provides an indication of the level of productivity achieved by the area in response to competitive pressure from other areas, including those beyond its borders. Specialization in sectors with high added value, such as industry, contributes to raising territorial competitiveness.

Table 3 Indicators of the subindex Innovation

Subindex	Pillar	Indicator (source, year)
Innovation	Technological readiness	Ultrabroadband penetration (Indicators for Development-Istat, 2017)
Innovation	Technological readiness	Number of firms registered in the innovative SME section (Ministry of Economic Development, 2019)
Innovation	Technological readiness	Manufacturing specialization in high-tech sectors (A misura di Comune-Istat, 2015)
Innovation	Technological readiness	Active enterprises with 3 and more employees engaged in Innovation projects (Companies Census-Istat, 2018)
Innovation	Technological readiness	Active enterprises with 3 and more employees using digital platforms (Companies Census-Istat, 2018)
Innovation	Technological readiness	Number of online services made available to citizens by the local PA (Istat, 2018)
Innovation	Business sophistication	Business fragmentation: percentage share of micro, small and medium-sized enterprises (ASIA-Istat, 2018)
Innovation	Business sophistication	Agriculture, forestry and fishing specialization index - value added (I.stat National Accounts, 2017) (reversed)
Innovation	Business sophistication	Industry specialization index - value-added (I.stat National Accounts, 2017)
Innovation	Business sophistication	Construction specialization index - value added (I.stat National Accounts, 2017) (reversed)
Innovation	Business sophistication	Services specialization index - value added (I.stat National Accounts, 2017)(reversed)
Innovation	Business sophistication	Entrepreneurship intensity per thousand inhabitants (I.stat, 2018)
Innovation	Business sophistication	Number of total businesses registered in the cultural production system (ASIA-Istat, 2018)
Innovation	Business sophistication	Degree of openness to foreign trade (author elaboration on I.stat, 2017)
Innovation	Innovation	Propensity to patent - applications filled at the European Patent Office (EPO) (UIBM database, 2016)
Innovation	Innovation	Propensity to patent - number of patents applications to the Italian Patent Office (UIBM database, 2018)
Innovation	Innovation	Registered patents to the Italian Patent Office (UIBM database, 2018)
Innovation	Innovation	Registered trademarks by province of registration in Italy (UIBM database, 2018)
Innovation	Innovation	Brain gain/drain or mobility of Italian graduates (25–39 years) (BES-Istat, 2017)
Innovation	Innovation	Cultural business employees as a percentage of total active business employees (ASIA-Istat, 2017)

Pillar XI - Innovation Innovation is the true engine of growth. More than costs, more than the availability of raw materials, more than geographical location, innovation is the key factor in the competitiveness of a country and a territory, especially the developed ones, as underlined by Lopez-Claros et al. (2007). In its annual report, the World Bank highlights the positive correlation between knowledge and growth and underlines how the fastest growing economies are also those with a higher Knowledge Economy Index (KEI). Unlike developing areas, where it is the increase in domestic consumption induced by the rise in the standard of living that drives GDP growth, in mature economies growth is fueled by technological innovation that stimulates the replacement of existing goods through the creation of new or higher performance goods: the faster the replacement of goods, the higher the growth rate. For innovation to spread throughout the territorial economy, the institutional environment must be sufficiently pervasive to create collaborative relationships between knowledge infrastructures (universities and research centers) and the firms that must apply the results of innovation to processes and products (Cantwell 2006). Empirical research shows that knowledge production is quite concentrated (Audretsch and Feldman 1996), so innovative firms tend to locate in settings with specialized human capital, which in turn tends to accumulate further in areas that are vibrant in terms of innovation.

For detailed description of the indicators for each pillar see the Sect. 5 (Appendix).

The values of the subindexes Basic, Efficiency and Innovation for the 106 regions are presented in Table 4.

With respect to the Basic subindex, the first ten provinces are Milano, Trento, Venezia, Treviso, Bologna, Lecco, Firenze, Monza Brianza, Padova, Udine; the last ten are Siracusa, Caltanissetta, Barletta Andria Trani, Foggia, Cosenza, Catanzaro, Salerno, Caserta, Crotone, Benevento.

With respect to the Efficiency subindex, the first ten provinces are Milano, Bologna, Trieste, Roma, Parma, Firenze, Torino, Modena, Bolzano, Padova; the last ten are Catania, Vibo Valentia, Agrigento, Reggio Calabria, Trapani, Ragusa, Enna, Siracusa, Crotone, Caltanissetta.

With respect to the Innovation subindex, the first ten provinces are Milano, Bologna, Torino, Modena, Vicenza, Firenze, Roma, Trieste, Parma, Pordenone; the last ten are Foggia, Crotone, Isernia, Nuoro, Barletta Andria Trani, Rieti, Oristano, Enna, Caltanissetta, Agrigento.

3 Fuzzy Clustering with Multivariate Data and *contiguity* Constraints

The data set can be represented as a spatial data matrix (D'Urso 2000, 2004, 2005) as:

$$\mathbf{X} \equiv \{x_{ij} : i = 1, \ldots, I;\ j = 1, \ldots, J\} \tag{1}$$

where i indicates the generic unit (geographical area or region, i.e. the province), j the variable (i.e. the pillar); x_{ij} is the value of the j-th variable observed for the i-th unit, or alternatively as follows:

$$\mathbf{x}_i \equiv \{x_{ij} :\ j = 1, \ldots, J\}. \tag{2}$$

Furthermore, we also assume to have K additional information on spatial location of the units, i.e. K different levels of contiguity. In particular, we can consider K $(I \times I)$ symmetric data matrices $\mathbf{P}_k$ $(k = 1, \ldots, K)$, whose generic entry $p_{kii'}$ is a measure of a particular kind of spatial proximity between the i-th and i'-th units $(i, i' = 1, \ldots, I)$ (Pham 2001; Coppi et al. 2010). In the literature, there are different ways of defining neighbourhood and consequently there are different ways of constructing proximity matrices among spatial units (Gordon 1999; Páez and Scott 2005). Two of the most common definitions are based on connectivity, i.e. travel time or distance between pairs of units, and physical contiguity. Contiguity can be specified in several ways. For instance, two spatial units can be contiguous either if they are adjacent (neighbours) or if they belong to the same macro-area, even if they are not adjacent. In this case, $\mathbf{P}$ is constructed as a symmetric matrix with 0

Table 4 Basic, Efficiency, Innovation by province

Region	Province	B	E	I	Region	Province	B	E	I
Piemonte	Torino	0.46	0.62	0.75	Toscana	Lucca	−0.09	-0.08	-0.11
Piemonte	Vercelli	− 0.01	− 0.09	− 0.27	Toscana	Pistoia	0.20	− 0.21	− 0.34
Piemonte	Novara	0.05	0.20	0.20	Toscana	Firenze	0.52	0.65	0.55
Piemonte	Cuneo	0.30	0.12	− 0.13	Toscana	Livorno	− 0.01	− 0.19	− 0.13
Piemonte	Asti	0.26	− 0.06	− 0.35	Toscana	Pisa	0.27	0.42	0.17
Piemonte	Alessandria	− 0.27	0.06	− 0.25	Toscana	Arezzo	0.18	-0.23	− 0.11
Piemonte	Biella	− 0.08	0.26	− 0.27	Toscana	Siena	0.31	0.18	0.07
Piemonte	Verbano C.O.	−0.07	− 0.13	− 0.62	Toscana	Grosseto	−0.05	− 0.32	− 0.73
Valle d’Aosta	Aosta	0.27	− 0.08	− 0.24	Toscana	Prato	0.44	0.13	− 0.14
Liguria	Imperia	−0.02	− 0.49	− 0.83	Umbria	Perugia	0.02	0.03	− 0.23
Liguria	Savona	0.09	0.13	− 0.29	Umbria	Terni	− 0.34	− 0.20	− 0.44
Liguria	Genova	0.12	0.37	−0.09	Lazio	Viterbo	− 0.60	− 0.29	−0 .75
Liguria	La Spezia	− 0.12	0.04	− 0.27	Lazio	Rieti	− 0.55	− 0.40	− 1.04
Lombardia	Varese	0.37	0.16	0.29	Lazio	Roma	− 0.11	0.72	0.53
Lombardia	Como	0.43	0.01	0.12	Lazio	Latina	− 0.52	− 0.32	− 0.42
Lombardia	Sondrio	0.03	0.09	− 0.47	Lazio	Frosinone	− 0.59	− 0.49	− 0.62
Lombardia	Milano	0.94	1.57	1.97	Campania	Caserta	− 0.94	− 0.87	− 0.91
Lombardia	Bergamo	0.28	0.24	0.09	Campania	Benevento	− 1.09	− 0.56	− 0.64
Lombardia	Brescia	0.39	0.15	0.08	Campania	Napoli	− 0.46	− 0.72	− 0.36
Lombardia	Pavia	0.18	0.37	0.18	Campania	Avellino	−0.58	− 0.37	− 0.72
Lombardia	Cremona	0.19	0.07	0.03	Campania	Salerno	0.93	0.67	0.52
Lombardia	Mantova	0.03	− 0.14	− 0.17	Abruzzo	L’Aquila	− 0.39	0.02	− 0.76
Lombardia	Lecco	0.57	0.19	0.14	Abruzzo	Teramo	− 0.70	− 0.19	− 0.54
Lombardia	Lodi	0.19	− 0.08	− 0.24	Abruzzo	Pescara	− 0.57	− 0.20	− 0.19
Lombardia	Monza Brianza	0.49	0.39	0.38	Abruzzo	Chieti	− 0.47	− 0.21	− 0.47
Trentino Alto Adige	Bolzano	0.27	0.47	0.18	Molise	Campobasso	− 0.41	− 0.54	− 0.83
Trentino Alto Adige	Trento	0.72	0.44	0.37	Molise	Isernia	− 0.48	− 0.17	− 1.00
Veneto	Verona	0.38	0.29	0.22	Puglia	Foggia	− 0.84	− 0.93	− 0.95

Table 4 continued

Region	Province	B	E	I	Region	Province	B	E	I
Veneto	Vicenza	0.47	0.26	0.57	Puglia	Bari	− 0.14	− 0.37	− 0.23
Veneto	Belluno	0.36	0.16	− 0.11	Puglia	Taranto	− 0.42	− 0.86	− 0.54
Veneto	Treviso	0.61	0.25	0.26	Puglia	Brindisi	− 0.40	− 0.82	− 0.79
Veneto	Venezia	0.71	0.35	− 0.04	Puglia	Lecce	− 0.65	− 0.70	− 0.64
Veneto	Padova	0.48	0.47	0.35	Puglia	Barletta A.T.	− 0.78	− 0.94	− 1.01
Veneto	Rovigo	0.13	0.09	− 0.55	Basilicata	Potenza	− 0.57	− 0.62	− 0.37
Friuli Venezia Giulia	Udine	0.47	0.42	0.10	Basilicata	Matera	− 0.58	− 0.49	− 0.71
Friuli Venezia Giulia	Gorizia	0.16	− 0.26	0.05	Calabria	Cosenza	− 0.87	− 0.84	− 0.76
Friuli Venezia Giulia	Trieste	0.34	0.78	0.49	Calabria	Catanzaro	− 0.88	− 0.43	− 0.64
Friuli Venezia Giulia	Pordenone	0.32	0.10	0.41	Calabria	Reggio Calabria	− 0.68	− 1.04	− 0.89
Emilia Romagna	Piacenza	0.01	0.38	0.03	Calabria	Crotone	− 1.05	− 1.24	− 0.96
Emilia Romagna	Parma	0.06	0.70	0.44	Calabria	Vibo Valentia	− 0.69	− 1.00	− 0.71
Emilia Romagna	Reggio Emilia	0.43	0.13	0.26	Sicilia	Trapani	− 0.54	− 1.09	− 0.93
Emilia Romagna	Modena	0.20	0.48	0.69	Sicilia	Palermo	− 0.40	− 0.83	− 0.55
Emilia Romagna	Bologna	0.59	0.83	0.92	Sicilia	Messina	− 0.67	− 0.91	− 0.72
Emilia Romagna	Ferrara	− 0.11	0.43	− 0.05	Sicilia	Agrigento	− 0.71	− 1.02	− 1.25
Emilia Romagna	Ravenna	0.44	0.07	0.11	Sicilia	Caltanissetta	− 0.73	− 1.51	− 1.25
Emilia Romagna	Forli Cesena	0.38	0.10	− 0.01	Sicilia	Enna	− 0.51	− 1.12	− 1.20
Emilia Romagna	Rimini	0.22	− 0.27	0.31	Sicilia	Catania	− 0.48	− 0.99	− 0.59
Marche	Pesaro Urbino	− 0.12	− 0.02	− 0.01	Sicilia	Ragusa	− 0.45	− 1.11	− 0.66
Marche	Ancona	0.21	0.10	0.39	Sicilia	Siracusa	− 0.72	− 1.17	− 0.53
Marche	Macerata	0.13	− 0.09	− 0.25	Sardegna	Sassari	− 0.32	− 0.58	− 0.76
Marche	Ascoli Piceno	0.06	0.00	− 0.08	Sardegna	Nuoro	− 0.49	− 0.74	− 1.00
Marche	Fermo	− 0.02	− 0.14	− 0.06	Sardegna	Cagliari	0.24	0.06	− 0.30
Toscana	Massa Carrara	− 0.25	− 0.06	− 0.37	Sardegna	Oristano	− 0.21	− 0.41	− 1.04

diagonal elements and with off-diagonal elements given by:

$$p_{ii'} = \begin{cases} 1 & \text{if } i \text{ is contiguous to } i' \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, I, \; i \neq i'. \tag{3}$$

The clustering procedure adopted enjoys the benefits connected to Fuzzy clustering and to Partitioning Around Medoids (PAM). Due to the difficulty in identifying a clear boundary between provinces even belonging to the same region, fuzzy clustering is more attractive than the hard clustering methods. In addition, the memberships indicate whether there is a second-best cluster almost as good as the best one, a scenario which hard clustering methods cannot uncover (Everitt et al. 2011). For more details, see D'Urso (2014).

Following a Partitioning-Around-Medoids (Pham 2001, Kaufman and Rousseeuw (2005)) approach in a fuzzy framework, the Fuzzy *C*-Medoids (FCMd) (FCMd, Krishnapuram et al. 2001) clustering algorithm is adopted, thanks of its great advantage of obtaining non-fictitious representative spatial units (i.e. the medoids) as final result. This allows for more appealing and easy to interpret results of the final partition (Kaufman and Rousseeuw 2005). From a computational perspective, fuzzy clustering algorithms are generally more efficient (dramatic changes in the value of cluster membership are less likely to occur in estimation procedures) and they are less affected by both local optima and convergence problems (Everitt et al. 2001; Hwang et al. 2007).

Dealing with spatial data, effects between adjacent units have to be taken into account. Since there could be different, say K $(K \geq 1)$, definitions of proximity, K spatial penalty terms are added to the objective function.

3.1 The Clustering Model

Following Pham (2001); Coppi et al. (2010); D'Urso et al. (2019a), the Fuzzy C-Medoids clustering algorithm with multivariate data and *contiguity* constraints is then formalised as follows:

$$\begin{aligned} &\min : \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} d(\mathbf{x}_i, \widetilde{\mathbf{x}}_c) + \sum_{k=1}^{K}\frac{\beta_k}{2}\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m}\sum_{i'=1}^{I}\sum_{c'\in C_c} p_{kii'}u_{i'c'}^{m} \\ &s.t. \sum_{c=1}^{C} u_{ic} = 1, \; u_{ic} \geq 0 \end{aligned} \tag{4}$$

where $\mathbf{x}_i$ and $\widetilde{\mathbf{x}}_c$ represents the multivariate i-th spatial unit and c-th spatial medoid $(c = 1, \ldots, C)$, respectively; $d(\cdot,\cdot)$ is the squared euclidean distance; $m > 1$ is the fuzziness parameter; $\beta_k \geq 0$ is the tuning parameter of the k-th spatial information; $p_{kii'}$ is the generic element of the $(I \times I)$ "proximity" matrix $\mathbf{P}_k$; C_c is the set of the C clusters, with the exclusion of cluster c; u_{ic} is the membership degree of the unit i to the cluster c.

The optimal iterative solution of the objective function in 4 is:

$$u_{ic} = \frac{\left[d(\mathbf{x}_i, \widetilde{\mathbf{x}}_c) + \sum\limits_{k=1}^{K}\beta_k\sum\limits_{i'=1}^{I}\sum\limits_{c'\in C_c} p_{kii'}u_{i'c'}^{m}\right]^{-\frac{1}{m-1}}}{\sum\limits_{c'=1}^{C}\left[d(\mathbf{x}_i, \widetilde{\mathbf{x}}_{c'}) + \sum\limits_{k=1}^{K}\beta_k\sum\limits_{i'=1}^{I}\sum\limits_{c''\in C_{c'}} p_{kii'}u_{i'c''}^{m}\right]^{-\frac{1}{m-1}}}. \tag{5}$$

The first term in (4) is the within cluster dispersion due to the multivariate features. The second (spatial dependent) term in (4) suitably allows the objective function to incorporate spatial information. The optimization of the objective function in (4) ensures that the cohesion within clusters is maximized and that the spatial autocorrelation existing in the data at hand is properly coped with.

The second (spatial dependent) term in (4) is the sum of K $(K \geq 1)$ spatial penalty terms (Pham 2001; Coppi et al. 2010), one for each definition of proximity among areas considered. In this way, the clustering model captures the information connected to the different levels of the proximity or "contiguity" (multilevel proximity or multilevel "contiguity"). For instance, we can consider the simple case in which the units, i.e. provinces, and macroareas, i.e. regions, are considered. In this specific case, two kinds of proximity ("contiguity") can be defined, proximity ("contiguity") among provinces (level 1 proximity or level 1 "contiguity") and proximity among regions (level 2 proximity or level 2 "contiguity") which the provinces belong to. Therefore, different scenarios can be identified: (1) two provinces (a_1 and a_2) are close to each other (level 1 proximity or level 1 "contiguity") and they belong to the same region (level 2 proximity or level 2 "contiguity"); (2) two provinces (a_1 and b_1) are close to each other (level 1 proximity or level 1 "contiguity") but they don't belong to the same region; (3) two provinces (a_1 and a_3) are not close to each other but they belong to the same region (level 2 proximity or level 2 "contiguity"); (4) two provinces (a_1 and b_2) are not close to each other and they don't belong to the same region.

In each spatial penalty term, two parameters are relevant, the proximity matrix $\mathbf{P}_k$, and the tuning parameter β_k. The role of the k-th proximity matrix is to increase the membership degree of unit i in cluster c and, at the same time, to increase the membership degrees of the units that are connected, in some way, to i in cluster c, while reducing these membership degrees in the other clusters. We define this spatial smoothing as neighbouring effect, where, as previously observed, the concept of neighbour is vast enough to encompass different types of connectivity between areas. The tuning parameter β_k can enhance the neighbouring effect due to $\mathbf{P}_k$ if the spatial autocorrelation between units is high, i.e., if the features of a spatial unit display a certain degree of concordance with those of the "neighbour". Otherwise, β_k could counterbalance, if not neutralise at all, the neighbouring effect, if there is relatively low spatial autocorrelation between areas. The choice of the value of β_k is data dependent. As observed by Coppi et al. (2010), the choice should be made according to a measure of a within cluster spatial autocorrelation (see Sect. 3.3), to avoid that the spatial smoothing induced by the proximity matrix overcome the cluster separation. Indeed, an excessively high value of one or more β_k's could constraint all "neighbour" units to be classified in one cluster, regardless the features observed.

An heuristic procedure for a suitable choice of β_k is described in Sect. 3.3.

3.2 Validity Measure

In general, internal validity measures provide useful guidelines in the identification of the best partition (as suggested by Handl et al. 2005; D'Urso 2015). A suitable measure for fuzzy clustering algorithm has been proposed by Xie and Beni (1991).

The Xie and Beni cluster validity index (Xie and Beni 1991) is the ratio between compactness and separation among clusters and it can be expressed as:

$$XB = \frac{\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^{m} d(\mathbf{x}_i, \widetilde{\mathbf{x}}_c)}{I \min_{p \neq q} x(\widetilde{\mathbf{x}}_p, \widetilde{\mathbf{x}}_q)} \tag{6}$$

where $(p, q) \in \{1..., C\}$. The smaller XB, the more compact and separate are the clusters.

3.3 Spatial Autocorrelation

As deeply analized in Coppi et al. (2010), the optimal choice of the value of the parameter β is a very complex issue. It has to be set exogenously by means of an heuristic procedure based on the spatial autocorrelation measure introduced in Coppi et al. (2010), that could be seen as a generalization of the Moran's index. For a chosen value of C and m and $k = 1$, the algorithm is run for increasing values of β (chosen in a suitable interval): the optimal β value is that maximizes the within cluster spatial autocorrelation. Properly, it maximizes the Global Moran overall spatial autocorrelation measure $\rho_{overall}$ that, for a given partition, is computed as follows:

$$\rho_{overall} = \frac{\sum_{c=1}^{C} \rho_c s_c}{I} \tag{7}$$

where $s_c = \sum_{i=1}^{I} u_{ic}$.

The ρ_c, the spatial autocorrelation measure for the c-th cluster, is computed as:

$$\rho_c = \frac{tr\left[\mathbf{X}'\mathbf{U}_c^{\frac{1}{2}}\mathbf{P}\mathbf{U}_c^{\frac{1}{2}}\mathbf{X}\right]}{tr\left[\mathbf{X}'\mathbf{U}_c^{\frac{1}{2}} diag(\mathbf{P}'\mathbf{P})\mathbf{U}_c^{\frac{1}{2}}\mathbf{X}\right]} \tag{8}$$

where $\mathbf{U}_c$ is the square diagonal matrix (of order I) of the membership degrees of cluster c, and $\mathbf{P}$ is the spatial contiguity matrix. The operator $diag(\cdot)$ creates a diagonal matrix whose elements in the main diagonal are the same as those of the square matrix in the argument. If P is a contiguity matrix with 0/1 values, every diagonal element contains the number of neighboring units for the associated spatial unit.

As for Moran's index, also for $\rho_{overall}$, a value of 1 (-1) identifies a perfect positive (negative) autocorrelation, while 0 indicates the absence of autocorrelation. Therefore, to higher values of the $\rho_{overall}$ corresponds a better spatial assignment of the units to the clusters. An heuristic procedure for a suitable choice of β consists in running the clustering model for increasing values of β, and choosing that value β_{opt} such that $\rho_{overall}$ is maximal.

Moreover, the Fuzzy Moran's index, as the Moran's index, can be interpreted as a measure of spatial spill-over effect (Ma et al. 2015; Yang 2012). In the literature, the spatial spill-over effect is considered as the indirect or unintentional effect that a geographical area exerts on other neighbour areas (Yang and Fik 2014). A positive spill-over effect is obtained when an area benefits of their neighbours influence due to the existence of spatial externalities across area.