The Knowledge Complexity of the European Metropolitan Areas: Selecting and Clustering Their Hidden Features

Carlo Bottai*

Martina Iori†

Cities are key places of economic activity, as they produce an enormous amount of wealth compared to the land they cover. Their study is, therefore, of primary importance in understanding the success of nations. Given the many interactions among people that happen within them, cities are well described as complex evolving systems, and a thorough analysis of their economy should be able to deal with this complexity. A likely candidate to grasp the reality of complex evolving systems, such the economy of cities, is the Economic Complexity framework (Hidalgo and Hausmann, 2009), given its capacity to synthesize a large amount of information into a single index.

We use patent data to compute the knowledge complexity index (KCI) of European metropolitan areas and describe their economy in terms of their innovative potential. Interpreted as a dimensionality-reduction algorithm, as proposed by Mealy et al. (2019), KCI helps to filter out the background noise from the abundant information produced by the interactions that happen within cities. By extending the work by van Dam et al. (2021), we highlight the relevance of going beyond the first leading eigenvector, to the analysis of which the rest of the literature is limited. We define clusters of similar cities, based on the additional dimensions obtained through this dimensionality-reduction procedure. The introduction of clusters dramatically increases the predicting power of KCI. Under this lens, the Economic Complexity framework is more than a single index: it is a powerful methodology to reveal the organized complexity hidden behind the large amount of chaotic information produced by out-of-equilibrium economic systems such as cities.

JEL codes: O34; O47; O18

Keywords: Economic Complexity; Labor Productivity; Metropolitan Areas; Dimensionality Reduction; Clustering

*Department of Economics, Management and Statistics, University of Milano–Bicocca, Milano (IT); carlo.bottai@unimib.it

†Institute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Pisa (IT); martina.iori@santannapisa.it

This text has been prepared to appear in the *Handbook of Complexity Economics*, Ping Chen, Wolfram Elsner, Andreas Pyka (Eds.), Routledge, 2023.

We thank the anonymous reviewers for their comments and suggestions which resulted in a substantial improvement of the text.

1 Introduction

The role of urbanization in economic development attracted increasing popularity in recent years (OECD, 2015; UN-Habitat, 2016). To understand why, it is perhaps sufficient to say that, worldwide, urbanization increased from 30% to 50% in the last 50 years and is expected to keep growing (UN-DESA, 2019). Moreover, while metropolitan areas cover only a tiny fraction of the planet, they are highly productive places, fundamental for national competitiveness in global markets. For instance, between 2000 and 2012, metropolitan areas accounted for about 45% of the EU15 Gross Domestic Product (GDP), although covering only 10% of its land (own elaboration from OECD, 2013b).

For this reason, an understanding of cities' economic performance is key to incentivizing countries' economic growth. However, it is not straightforward to capture cities' dynamics, and standard (*neoclassical*) economic-policy tools seem not well suited to deal with their complexity. This chapter contributes to show that a *complexity economics* perspective is more suitable for the analysis of cities' economy.

Neoclassical and complexity economics correspond to distinct ontological claims about the world (Arthur, 1999; 2021) and, like oil and water, cannot mix with each other (Fontana, 2010). Indeed, neoclassical economics describes an economic system as composed of some (perfectly and boundlessly rational) representative agents who, in facing several well-defined problems, behave consistently with the aggregate outcome of their actions (Arthur, 2021). Without the intervention of some extra-economic factor, the outcome of such a "well-functioning machine" will be a timeless equilibrium, where there cannot be growth, if not in quantitative terms (Schumpeter, 1911). On the contrary, complexity economics looks at economies as an evolving system in which novelty emerges from within because of the *creative reactions* of its agents to macro-level out-of-equilibrium conditions (Antonelli, 2015; Schumpeter, 1947). Agents need to collectively contribute to develop a knowledge base constituted of a coherent scaffolding of technologies, institutions, firms, routines, etc. In this way, this *emergent* environment that they co-create through their (decentralized) efforts will guide them toward mutually satisfactory ends, in a continuous feedback loop process.

Metropolitan areas hardly fit with the neoclassical paradigm, being perfect examples of out-of-equilibrium systems (Prigogine, 1977). Conversely, cities are *complex evolving systems* with many interacting physical and social components (Batty, 2013; Jacobs, 1961).

Here, we analyze cities by focusing on their technological endowment. Since new knowledge is generated through the recombination of existing knowledge pieces (Arthur, 2009), the study of the technical knowledge available in a city is crucial to foresee its future economic development potential. While cities are the *locus* of countries' technological progress , not all cities are able to translate their technological development efforts into a more competitive economic system. This chapter shows how the so-called *economic complexity framework* (E.C. for brevity), applied to cities and their technological capabilities, can offer a useful tool to synthesize complex information – resulting from multiple interactions among cities' actors – and describe how cities evolve.

The E.C. framework has been introduced by Hidalgo and Hausmann in 2009 to describe countries' competitiveness in terms of exported products. More recent applications of this framework to the knowledge base of metropolitan areas led to the definition of the Knowledge Complexity Index (Balland et al., 2019; Balland and Rigby, 2017), an indicator that summarizes the knowledge and innovative potential of a city.[1]

Interpreted as a *dimensionality-reduction* algorithm (Gomez-Lievano, 2018; Mealy et al., 2019), this index filters out random noise from the abundant information produced by cities.

Focusing on this interpretation helps to fit the E.C. framework within a complex systems' perspective. In line with the definition of *organized complexity*, (Jacobs, 1961; Weaver, 1948), the E.C. methodology is an attempt to overcome the limitations of the *science of simplicity* approach of the usual production functions that oversimplifies the problems by taking for granted given inputs, such as capital and labor. As highlighted by Hidalgo (2021), the E.C. framework proposes, instead, to let places reveal the abstract factors of production they are endowed with. And, as shown by a growing literature, by organizing the abundant information provided by such fine-grained databases, the E.C. framework proved to accurately predict the performance and growth perspectives of local economies.[2]

Here, we provide an example of the potential of this approach in describing city economic development. We apply the KCI to study the evolution of the European metropolitan areas between 2004 and 2008. We compute the index by collecting data on European Patent Office (EPO) patent applications from 214 Metropolitan Regions of EU28 and EFTA countries. By interpreting the E.C. methodology as a dimensionality-reduction algorithm and preserving additional dimensions resulting from this procedure, we show that it is possible to complement the KCI with additional information and improve its ability in describing the role of technical knowledge for the competitiveness of these metropolitan areas.

The chapter is organized as follows. Section 2 presents the motivation and a review of the previous literature. Section 3 introduces the Knowledge Complexity Index. Section 4 details the data and discusses the results. Lastly, Section 5 concludes.

2 Motivation and literature background

In *complex evolving systems*, the "complexity" may arise, among other factors, from the large number of variables involved. Even though low-dimensional complex systems exist, when you face complex phenomena, reducing the dimensionality of the system helps in dealing with the problem of interest.

Cities and technical change Since technical progress is the only factor able to grow in per-capita terms, investing in research and innovation is considered the fundamental endogenous engine of economic development (Foray, 2004; Jones and Romer, 2010; Lucas, 1988; Romer, 1990; Schumpeter, 1911). In a globalized economy, national prosperity

[1] In this context, we analyze cities' knowledge base and technologies instead of countries' economy as a whole and exported products. Therefore, we call Knowledge Complexity Index (KCI) what in the original framework by Hidalgo and Hausmann (2009) was called Economic Complexity Index (ECI), and Technological Complexity Index (TCI) what was called Product Complexity Index (PCI). However, even if the terminology differs, the methodology is exactly the same as in the original contribution of 2009.

[2] For a similar perspective on complexity economics, see also Lü et al. (2023) in this book.

substantially depends on the capacity of a country to continually commercialize new and high-value (even unique) products, services, and processes (Porter, 1990; 1998).

In other words, the most competitive nations are those able to mobilize a significant amount of diversified knowledge to achieve technical progress. However, technical progress depends on three distinct knowledge sources (Hausmann, 2016): *embodied*, *codified*, and *tacit knowledge* (i.e., *know-how*). This latter is the stickiest since it moves with the people and requires long training sessions to be transferred from person to person (Breschi and Lissoni, 2009; Jaffe et al., 1993; Polanyi, 1966). Therefore, the mobilization of know-how can be seen as the key competitive advantage of a nation (Hausmann et al., 2014). Moreover, from a complexity economics perspective, it is not merely a fact of how diverse the knowledge base of an economy is, but also of how much this pool is organized and how the elements interact with each other.[3]

Links between *urbanization* and *technical progress* (Florida, 2002; Giersch, 1995; Glaeser, 2011; Henderson and Thisse, 2004) and between *urban density* and *inventiveness* (Berkes and Gaetani, 2020; Bettencourt et al., 2007; Carlino et al., 2007; Moretti, 2019) have been repetitively shown. Given the limited human capacity of acquiring and storing know-how, the production of highly complex products requires institutions that gather different knowledge elements, scattered among several brains, and steer them towards a collective effort (Henrich, 2004; Mokyr, 2002; Richerson and Boyd, 2005). And cities are, in knowledge-based capitalism, the most important of such institutions (Bettencourt, 2013; Glaeser, 1999). Even more, they are «not just the containers where innovation and entrepreneurship happen, they are the key mechanisms which enable them» (Florida et al., 2017, p. 93), first and foremost because of their capacity to contain and organize such a diverse pool of technical knowledge towards creative outputs.

Therefore, we can consider cities as the core of nations' competitiveness since they provide an institutional scaffolding enabling the production of a diversified pool of activities, among which some know-how-intensive ones. The latter, being rare and hard to imitate, provides nations with lasting competitive advantages.

Economic Complexity The E.C. framework, proposed by Hidalgo and Hausmann (2009), described national economies through country exports. The two authors reinterpreted the countries-products bipartite network as the sign left by a tripartite network connecting countries to the *capabilities* they have and products to the capabilities they require.[4] In this way, the authors showed that it is possible to indirectly measure these capabilities by looking at who produces what.[5] Building on this intuition, the authors introduced the so-called Economic Complexity Index (ECI): an index aiming to measure how much know-how an economy is able to mobilize in its productive effort (Hausmann et al., 2014). Since then, this index has been widely used in many contexts (see Balland et al., 2022; Hidalgo, 2021, for a comprehensive review).

[3]About knowledge from a complexity economics perspective, see also Hidalgo (2023) in this book.

[4]Capabilities are chunks of knowledge needed to achieve a goal. For the purpose of this chapter, one can read it indistinguishably from know-how, even though they are two separate concepts. See Aistleitner et al. (2021) for an extensive review of the concept.

[5]To this seminal paper many other works based on the same intuition followed and two main streams of literature can be identified. On the one hand, Hausmann, Hidalgo, and their co-authors have defined the so-called Economic Complexity Index (Hausmann et al., 2014). On the other hand, Pietronero and his co-authors proposed a measure called Fitness (Cristelli et al., 2013; Tacchella et al., 2012).

Specifically, Balland and Rigby (2017) and Pintar and Scherngell (2021) applied the E.C. framework to technological knowledge at the metropolitan level by defining a Knowledge Complexity Index (KCI), i.e. the ECI computed on the city-knowledge bipartite network. Results offer evidence, both in the US and Europe, of the influence of knowledge complexity on cities' capacity to generate new knowledge and, ultimately, grow in the long run. Antonelli et al. (2017) and Antonelli et al. (2020) looked at the knowledge complexity of European regions and showed that more complex regions generate new technical knowledge with more ease, while the evidence about a nexus with productivity enhancements is less straightforward than expected. Lastly, Petralia et al. (2017) observe that, along their development path, countries tend to move towards more complex and valuable technological domains. Overall, these studies showed that the type of regional knowledge-based activities, as well as the structural characteristics of this complex bundle, matter to produce new knowledge and, ultimately, influence the pace and directionality of an economy's growth path. At the same time, they highlighted the primary role of the relative scarcity of each element. A bundle of knowledge items able to yield strong Jacobian externalities will include many rare activities. Since only regions with many high-skilled individuals and specific technical competencies will develop sophisticated – and thus rarer – technologies, these regions are expected to be the most competitive ones.

About its interpretation, it is common to see the E.C. framework read as a generalized notion of diversity. However, more recently, Gomez-Lievano (2018) and Mealy et al. (2019) showed that it is, ultimately, a dimensionality-reduction algorithm. Consequently, the index is not a measure of how much two economic systems are diversified within themselves, but it captures how much the two are similar to each other in terms of specialization pattern.

As recently highlighted by Hidalgo (2021), the ECI as a dimensionality-reduction technique is also an alternative to traditional economics approaches that isolate the components of an aggregated output, like the GDP, assuming the nature of its inputs, such as capital and labor. Unlike these approaches, the ECI learns, from fine-grained databases about the "behavior" of several economic systems, which are the "abstract factors of production" each place is endowed with.

Moreover, this interpretation aligns the E.C. framework with the broader literature on *network science* and *complex systems* methods for economics, which frequently suggests tools that separate random noise from the underlying signal of interest (Hidalgo, 2021; Pugliese and Tübke, 2019).

3 The Knowledge Complexity Index

According to the theory previously exposed, we can extract information about cities' knowledge complexity from a city-by-technology matrix $\mathbf{Q}$. Following the literature in economics of innovation (Griliches, 1990; Hall et al., 2001), we use patent information to proxy regional knowledge production so that each cell of the matrix, Q_{ck}, counts the (fractional) number of citation-weighted[6] patent applications, in the technological

[6] In this way, we assign a higher weight to more relevant patents (see section 4.1).

domain k, of inventors located in city c.[7]

Firstly, we use the Revealed Technological Advantage (RTA) to find the portfolio of technological (relative) specializations of each city (Soete and Wyatt, 1983). We transpose this information into a binary bi-adjacency matrix, $\mathbf{M} = [m_{ck}]$, stating whether a city c reveals a comparative advantage in technology k or not. That is, $m_{ck} = 1$ if $RTA(\mathbf{Q})_{ck} \geq 1$, and 0 otherwise, where $RTA(\mathbf{Q})_{ck} = \frac{q_{ck}}{q_{c\cdot}} / \frac{q_{\cdot k}}{q_{\cdot\cdot}}$; $q_{c\cdot} = \sum_k q_{ck}$; $q_{\cdot k} = \sum_c q_{ck}$; and $q_{\cdot\cdot} = \sum_k \sum_c q_{ck}$.

Secondly, we define two indices, the Knowledge Complexity Index ($\overrightarrow{KCI}$) and the Technological Complexity Index ($\overrightarrow{TCI}$), that are deeply related to each other. As proposed by Hausmann et al. (2014), $\overrightarrow{KCI}$ is the *eigenvector*, $\vec{v}_2$, associated with the second-largest *eigenvalue*, λ_2, that solves the problem:[8]

$$\widetilde{\mathbf{M}}\mathbf{v} = \lambda \mathbf{v}, \tag{1}$$

where $\widetilde{\mathbf{M}} = \mathbf{D}_c^{-1}\,\mathbf{M}\,\mathbf{D}_k^{-1}\,\mathbf{M}^T$ is a stochastic matrix of pairwise similarities between cities; $\mathbf{D}_c = \mathrm{diag}(m_{c\cdot})$ and $\mathbf{D}_k = \mathrm{diag}(m_{\cdot k})$ are diagonal matrices; $m_{c\cdot} = \sum_k m_{ck}$ is the number of technological domains in which a city has a comparative advantage (city *diversity*); and $m_{\cdot k} = \sum_c m_{ck}$ is the number of cities having a comparative advantage in a technological domain (technology *ubiquity*). The values are standardized, so that:

$$\overrightarrow{KCI} = \frac{\vec{v}_2 - \langle \vec{v}_2 \rangle}{sd(\vec{v}_2)}.$$

As well, it is possible to solve a symmetric problem with respect to the technological domains, $\widehat{\mathbf{M}}\mathbf{u} = \lambda \mathbf{u}$, where $\widehat{\mathbf{M}} = \mathbf{D}_k^{-1}\,\mathbf{M}^T\,\mathbf{D}_c^{-1}\,\mathbf{M}$ is a stochastic matrix of pairwise similarities between technological domains. Its solution leads to the Technological Complexity Index, $\overrightarrow{TCI} = \frac{\vec{u}_2 - \langle \vec{u}_2 \rangle}{sd(\vec{u}_2)}$, where TCI_k is the *complexity* of the technological domain k.[9]

Economic Complexity and Correspondence Analysis As explained by Mealy et al. (2019), Eq. 1 is equivalent to the problem solved by Correspondence Analysis (CA), a multivariate statistical method for analyzing relationships between two categorical variables (Greenacre, 1984; Hill, 1974). Like Principal Component Analysis, CA decomposes the χ^2 statistic associated with the contingency table $\mathbf{M}$ into orthogonal axes. CA can be used to summarize the association between rows' and columns' categories of a contingency table in a lower-dimensional space.

Let us define the *specialization pattern* of city c as $m_c = [m_{c1}/m_{c.} \ldots m_{cK}/m_{c.}]$, where each element is its propensity to patent in the k-th technological domain. Geometrically, CA defines a vector space where the distance between the *specialization pattern* of two

[7] Although we are aware of the many concerns about the use of patents as generic indicators of inventive activity –principally, that not all inventions are patented and that some relevant types of inventions cannot be patented at all (Griliches, 1990)– we nevertheless see, in line with the literature, patents as a useful proxy for inventions in empirical studies (Jaffe and Trajtenberg, 2002)

[8] For a detailed description of the E.C. methodology with an in-depth discussion about how we can interpret the meaning of *eigenvector* within it, see also Hidalgo (2023) in this book.

[9] The vector $\vec{v}_2$ is supposed to positively correlate with the city *diversity*. Otherwise, $\vec{v}_2 \equiv -\vec{v}_2$. In this case, also $\vec{u}_2 \equiv -\vec{u}_2$. Moreover, when KCI is compared with other subsequent eigenvectors ($\vec{v}_3$, $\vec{v}_4$, etc.), we standardize also these last as we do for the $\vec{v}_2$.

cities is defined as $d^2(m_c, m_{c'}) = \sum_{k=1}^{K} \left(m_{ck}/m_{c.} - m_{c'k}/m_{c'.}\right)^2 / \left(m_{.k}/m_{..}\right)$, known as "$\chi^2$ distance" (Greenacre and Hastie, 1987). The same can be said about the localization patterns of two technological domains. This space can be approximated by reducing its dimensions to a subset of the *eigenvectors* in which it has been decomposed. From this perspective, the KCI and TCI are proportional to the best one-dimensional approximations of the cities and technologies specialization patterns space, respectively: i.e., $\text{KCI}_c - \text{KCI}_{c'} \propto d(m_c, m_{c'})$ and $\text{TCI}_k - \text{TCI}_{k'} \propto d(m_k, m_{k'})$.

Economic Complexity as similarity in low-dimensions To interpret the E.C. methodology as CA (van Dam et al., 2021; Mealy et al., 2019) reveals that KCI captures how similar two cities are, in terms of technological specialization. As well, TCI provides information on how similar two technological domains are, in terms of which cities show a propensity to patent in such domains. This is true, the more the first axis in which CA decomposed **M** represent most of the variance (*total inertia*) present in the data; i.e., the higher the eigenvalues associated, respectively, to KCI and TCI.

Correspondence Analysis and clustering However, in accordance with this CA interpretation, to explain a higher share of the variance we can consider additional eigenvector solutions of Eq. 1 ($\vec{v}_3$, $\vec{v}_4$, etc.) to further identify cities' common specialization patterns (or hidden features), as proposed by van Dam et al. (2021).

To retrieve the maximum amount of information and preserve, at the same time, the convenience of having a single index for each city, we define clusters of cities based on these additional orthogonal dimensions. We use a *k-means* algorithm to define clusters, and we determine the optimal number of clusters in the data based on Hartigan's rule (Hartigan, 1975).[10] Then, we combine the KCI with the belonging of a city to these clusters to better describe the determinants of cities' competitiveness. To avoid data over-fitting and select only meaningful clusters, we reduce the sample of additional eigenvalues to those that, cumulatively (including the KCI), explain the 20% of the *total inertia*.

4 KCI and cities' competitiveness

We now apply these measures to show that the KCI is a proper tool to disentangle the complexity of cities and illustrate what is the role of the additional dimensions in better describing these complex systems.[11] In particular, we study the ability of the KCI to describe the competitiveness of cities in the near future, as measured by labor productivity (LP).[12] We also provide evidence on the advantages of the combined use of the KCI and clustering to improve the explanatory power of the KCI. Specifically,

[10] This index compares data variability in different levels of hierarchical clusters and selects the number of clusters that maximize the distance between them.

[11] The analysis has been performed using the R package SCCA (van Dam et al., n.d.).

[12] Since we defined a competitive city as one able to continuously innovate, we believe that labor productivity is able to capture cities' competitiveness, at least partly. This is because, only by increasing the efficiency in the use of the inputs required by the pre-existing activities, a city can free the resources needed to introduce new products in the economy. Anyhow, we acknowledge the narrowness of this shortcut, partly justifiable by the fact that this empirical exercise is only illustrative of a more general idea, which is the true core of this chapter.

we show that the KCI is remarkably associated with labor productivity only if we look at a subset of metropolitan areas, which we can easily identify using the additional information provided by the eigenvectors after the leading one, i.e. the KCI itself.

4.1 Data

We fetch 246,644 patent applications submitted to the European Patent Office (EPO) by EU28 and EFTA inventors between 2004 and 2008. This time-window has been chosen with two aims in mind. Firstly, our measures rely on citations that patents received in the five-year period that follows their publication date, as discussed later in more detail. Therefore, we have to limit our analysis to patents published before 2013 (due to data availability and completeness). Secondly, the paper aims to provide a methodological contribution: the key point of the chapter is that the KCI, if properly interpreted, is a useful tool to summarize the large amount of information produced by a complex evolving system (e.g., cities). Therefore, to avoid further complications to the analysis, we selected a time-window that, to the best of our knowledge, is considered not particularly turbulent by the macroeconomics literature.

Patents are attributed to 443 NUTS-3 regions, belonging to 30 different countries, by inventor residence, using REGPAT (OECD, 2020). We aggregate NUTS-3 regions into 214 Metropolitan Regions as defined by the Eurostat, excluding (fractional) patents located outside Metropolitan Regions.[13] Patents of Metropolitan Regions account for 85% of the entire European patent production.

Each patent is classified, by patent offices, into several (hierarchical) technological classes, following the Cooperative Patent Classification (CPC). In the analysis, we use 621 sub-classes (4 digits) to proxy technological capabilities.

Patents vary enormously in their importance or value, and hence, simple patent counts are problematic as proxies of innovative output. To partly correct this issue, we follow Pintar and Scherngell (2021), and we weigh each patent by the number of citations, q, it received from other EPO patents in the five years after the publication date.[14] To account for the differences in citation patterns over the years and among technological domains, we discount the citations (n_{ptf}) received by patent p by the average number of citations received by patents applied in the same year, t, and belonging to the same technical field, f: $w_{ptf} = n_{ptf} + 1 / \mathbb{E}(n_{tf}) + 1$.[15]

[13] According to the definition provided by the OECD and the EU (2012), an urban area is a "functional economic unit" formed by a densely inhabited city and its "related" commuting zone. Consequently, the Eurostat defines Metropolitan Regions combining several NUTS-3 regions designed to represent, overall, at least 250,000 inhabitants, commuter belts around an urban core included. In the analysis, we use Metropolitan Regions based on the NUTS 2013 classification.

[14] Pintar and Scherngell (2021) use only citations coming from patents assigned to locations other than the Metropolitan Region of the focal patent since they would like to approximate for knowledge something similar to export data, in analogy with the original work by Hidalgo and Hausmann. Instead, we use any citation, as counted by the OECD's PATENT QUALITY database (OECD, 2013a). Indeed, we aim to follow a well-established tradition that uses patent citations to correct for differences in the innovativeness value of each patent (Trajtenberg, 1990). Facing this issue is particularly significant in this context since rare knowledge domains can be as such either because more complex or because less useful than others.

[15] While in the rest of the chapter we use CPC sub-classes to proxy the technological domain of a patent, the citations are normalized with respect to the average patent belonging to the same technical field, f, as defined in Schmoch (2008) and subsequent updates. This is convenient, among other reasons,

Following the previous literature, we aggregate citation-weighted patents over five years, because the patenting patterns of each metropolitan area vary significantly from year to year. As well, we dropped regions belonging to the twentieth percentile, in terms of fractionally counted patent applications in five years (88 patents) and CPC sub-classes belonging to the fifth percentile (6 cities), to remove the less informative part of the co-occurrence matrix.

Finally, we retrieve information on the gross value added at current prices and the number of employees at the NUTS-3 level in 2009 from ARDECO (Eurostat, 2020). These data are not available for Switzerland and Iceland, so we cannot include these countries in the final estimations, while we still consider their patents to compute the KCI. We aggregate NUTS-3 regions into Metropolitan Regions, as previously done for patents. Then, we define Labor Productivity as the Gross Value Added per employee. The average Labor Productivity in EU28 countries and Norway in 2009 is € 48.57 per employee, ranging from a minimum of 7.03 to a maximum of 98.64 euros.

4.2 Analysis and discussion

Knowledge complexity as sorting Fig. 1a represents the specialization pattern (matrix **M**) of a Metropolitan Region (row) in each technological domain (column). The rows (columns) have been sorted according to the *diversity* (*ubiquity*) of each city (technology). The matrix shows a peculiar quasi-triangular shape, well-known in ecology as *nestedness* and already observed in several economic contexts: from trade data to industrial sectors, occupations, and patents; from countries to regions, to cities (Antonelli et al., 2017; Balland and Rigby, 2017; Bustos et al., 2012; Mealy et al., 2019; Saracco et al., 2015).

Nested bipartite networks tend to be *disassortative* (Jonhson et al., 2013), so that the average ubiquity of the activities present in a place tends to correlate negatively with the diversity of such a place. Fig. 1c confirms a clear disassortativity in our data. In light of the literature discussed in Section 2, this property reflects the idea that know-how is sticky and tends to diffuse slowly, from the location where it has been produced to other places. Moreover, it diffuses neither to all areas nor in all technological domains at the same pace (Petralia et al., 2017). Therefore, cities that patent more, tend to be more diversified (Spearman's rank correlation: 0.75) and to patent also in domains that are rarer than the average.

Based on the previous discussion, this observation is somewhat surprising. Indeed, we would expect that cities will try to increase their competitiveness by patenting only in the rarest technological domains that are feasible given their capabilities. Fig. 1b partly reconciles with this hypothesis. By sorting the rows and columns of the matrix by their KCI and TCI, instead of by their *diversity* and *ubiquity*, we observe that high-KCI cities patent, preferably, in high-TCI domains, as already observed for trade data (Saracco et al., 2015; Schetter, 2019; Straka, 2018). In other words, Fig. 1 shows why it makes sense to look at the patenting-basket *diversity* of a city to describe its level of

because in the OECD PATENT QUALITY database each patent is characterized by only one technical field but by multiple technological domains. Moreover, to preserve in the analysis patents that receive no citations, we add 1 to both the terms of the fraction. Lastly, for both the numerator and the denominator, we limit the count to the first five years from the publication date of each patent, to consider the different probability of receiving a citation among the latest and more ancient patents.

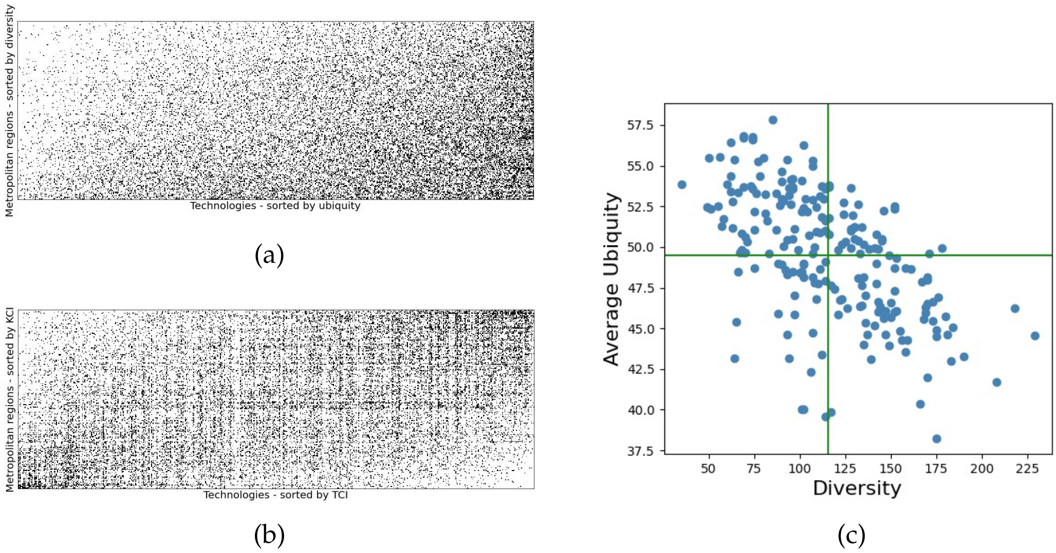


Figure 1: **(a) M** matrix with axes sorted by *diversity* and *ubiquity*. Black squares indicate the presence of cities' technological comparative advantages in certain technological domains, while white squares signal their absence. **(b) M** matrix with axes sorted by KCI and TCI. **(c)** City diversity and average ubiquity of the technological domains in which it patents.

competitiveness. But, at the same time, it also shows why the KCI is better than *diversity* at ranking cities from the least to the most competitive ones.

The equivalence between KCI (TCI) and what is known as *dual scaling* explain why they order the data along the diagonal of the matrix (Greenacre, 1984; Nishisato, 1980; 2006). Dual scaling, essentially equivalent to CA, provides a way of obtaining quantitative scale values for categorical data, like a contingency table. These scalar values are determined in such a way that the data, weighted by KCI and TCI, attain the maximum Pearson correlation. As said, to learn something about which capabilities are present in a city, it is more informative to know which kind of economic activities are performed in it than to know the size of the economic activities performed there. The E.C. framework helps to transform this qualitative information into a quantitative one, so that it can be used in subsequent empirical models, like regression analysis.

Knowledge complexity as clustering As described in Section 3, it is worth including CA additional axes in the analysis to retrieve as much information as possible. Fig. 2 presents the inertia associated with the first twenty non-trivial axes (hidden features). This figure shows that the first axis (i.e., the KCI) accounts for 3% of the total variance in the data, while the first thirteen axes explain 20% of the total inertia. It is worth noticing that the ECI computed on the countries-exports data, as in the original application by Hidalgo and Hausmann (2009), explains a similar share (3.5%) of the total variance in the data (van Dam et al., 2021).

As explained in Section 3, we can use the additional CA axes to retrieve more information about **M**. Fig. 3 shows two scatter plots that combine information about the KCI and the first two additional axes ($\vec{v}_3$ and $\vec{v}_4$) of each European Metropolitan Regions.

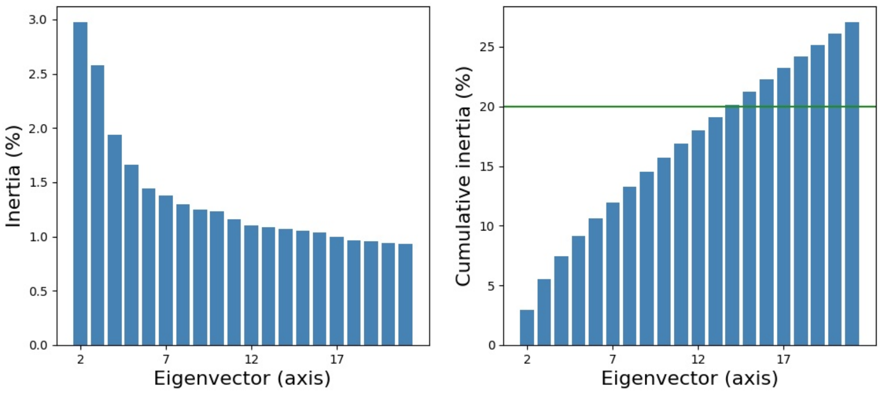


Figure 2: Inertia (left panel) and cumulative inertia (right panel) corresponding to the first twenty orthogonal axes from CA (the first eigenvalue is the trivial solution, so it is not included in the figure). To compute clusters, we select the axes that explain, cumulatively, the 20% of the inertia in the data (red line in the right panel).

Let us look at the three examples: Brussels (BE), Eindhoven (NL), and Uppsala (SE). The figure suggests that different axes capture different characteristics in terms of technological specialization patterns. All three areas show a very high KCI, and they are even more similar looking at the hidden feature captured by $\vec{v}_4$. On the contrary, these three metropolitan areas differ substantially (and so, are very far apart) if we focus on the information carried by $\vec{v}_3$. These observations suggest that they must have something in common (as shown by $\vec{v}_2$ and $\vec{v}_4$), but they also differ in some other characteristics (as captured by $\vec{v}_3$).[16]

To better capture these differences and similarities, we select the additional CA axes that explain 20% of the total inertia and compute metropolitan cities' clusters based on these additional dimensions. The optimal number of clusters, computed with Hartigan's rule, is equal to 5 and the resulting clusters are mapped in Fig 4. If we consider the three metropolitan areas in the example above, in terms of clusters the differences between them result in their belonging to different clusters: Brussels is in *cluster 2*; Eindhoven in *cluster 4*; while Uppsala in *cluster 1*.

Combining sorting and clustering By looking at the relationship between the KCI and Labor Productivity, Fig. 5a shows that the KCI of a city is positively associated with its (future) productivity. However, a linear regression between the two variables explains only 3% of the variance, as signaled by the R^2. At first sight, this might suggest that the KCI captures only partially the competitiveness of Metropolitan Regions. However, the KCI is only one – even though the most important – dimension that captures similarities in our data.[17] The idea of including clusters in the analysis is to complement KCI and increase the amount of information provided by this index.

[16]For a complete picture, we should look at all possible eigenvectors in which the matrix has been decomposed. But the first three are already enough to highlight the point that is relevant here.

[17]It is worth noticing that we are including the KCI as the only dependent variable, without further controls and fixed effects.

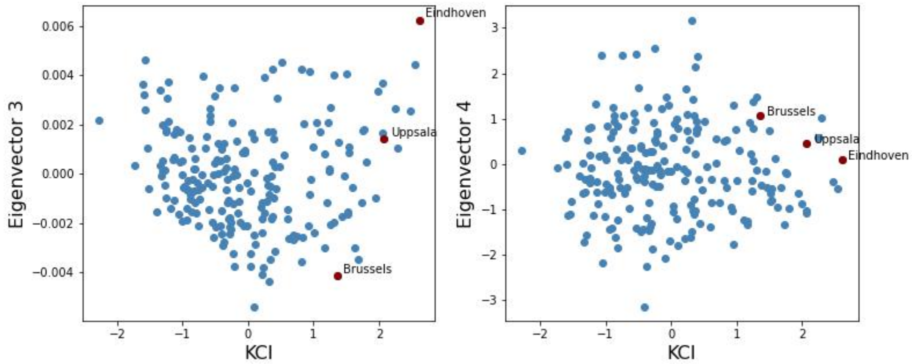


Figure 3: KCI vs v_3 (Eigenvector 3) and KCI vs v_4 (Eigenvector 4). Each dot corresponds to a metropolitan area in the period 2004-2008. A few examples of metropolitan areas are highlighted in red.

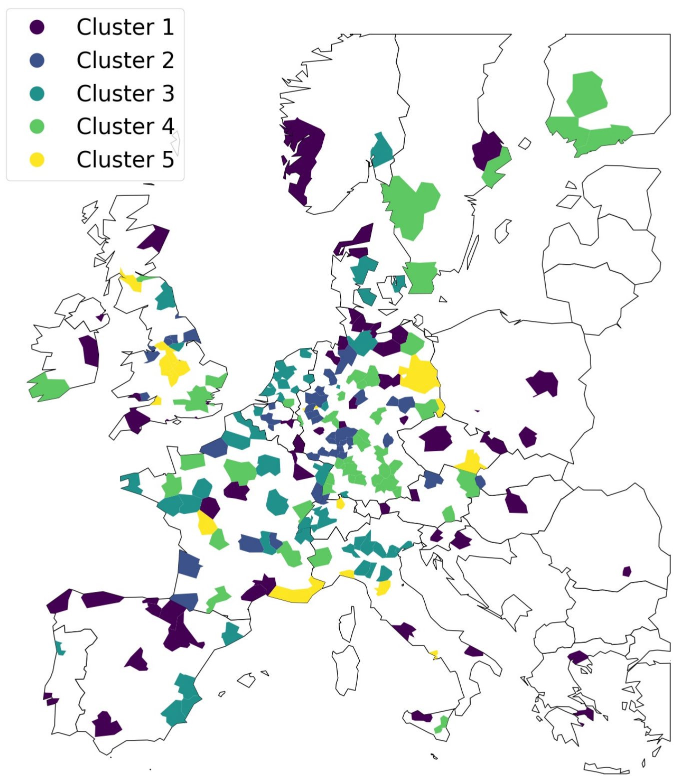
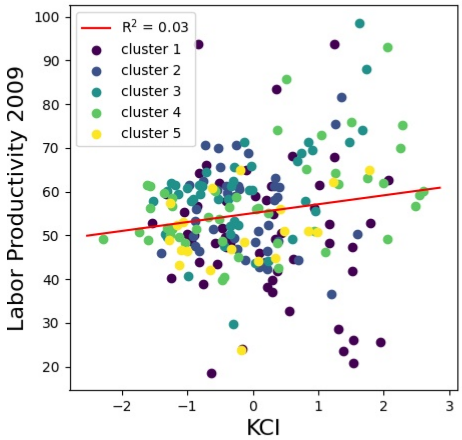
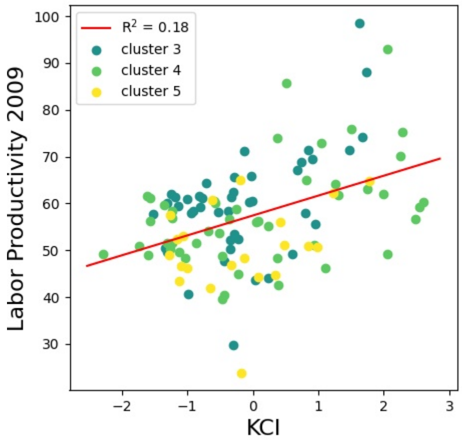


Figure 4: Classification of the European Metropolitan Regions in five clusters based on the additional CA axes.



(a) All metropolitan regions



(b) Selected clusters

Figure 5: Relationship between the KCI and Labor Productivity for all metropolitan areas (left panel) and selected clusters (right panel). The colors signal the clusters, and the red lines reflect linear regression results.

Since CA results in a list of several orthogonal axes, we can increase the amount of information provided by KCI by considering these additional dimensions and creating clusters of similar cities based on these common features. If we include in the analysis the clusters resulting from the additional CA axes (different colors in Fig. 5a), we observe that the prediction power of the KCI on the Labor Productivity is heterogeneous across clusters. This evidence suggests that the explanatory power of the KCI in capturing cities' similarity in terms of technical progress might vary according to other variables and that the definition of clusters captures, at least partially, these relevant dimensions. In particular, the explanatory power of the KCI and its association with Labor Productivity is higher in clusters 3, 4, and 5. By selecting only the Metropolitan Regions belonging to these clusters, we remove outliers (present especially in cluster 1) and we substantially improve the effectiveness of the KCI in capturing cities' competitiveness, as shown in Fig. 5b. In this case, the explanatory power of the linear regression between the KCI and Labor Productivity rises to 18%.

Even though an overall and satisfactory comprehension of the differences in KCI explanatory powers from cluster to cluster goes beyond the scope of this chapter, Fig. 6 might offer a tentative explanation. Given the higher relevance of patenting for the manufacturing sectors compared with sectors like services, we summarize the distribution of the proportion of Gross Value Added (GVA) in manufacturing (NACE sectors B-E) over the total GVA of a Metropolitan Region. Clusters 1 and 5 are the ones with a lower proportion of GVA in the manufacturing sector. Cluster 4, instead, shows both the highest average value and the highest dispersion, compared with the other cities' clusters. Therefore, the plot seems to suggest that the lower explanatory power of KCI compared to clusters 1 and 5 is due to the lower proportion of manufacturing activities in the cities belonging to such sub-samples. Instead, concerning cluster 2, from the map in Fig. 3, we can observe that it covers most of the area around the Ruhr (Germany). This area saw a substantial process of deindustrialization, moving from

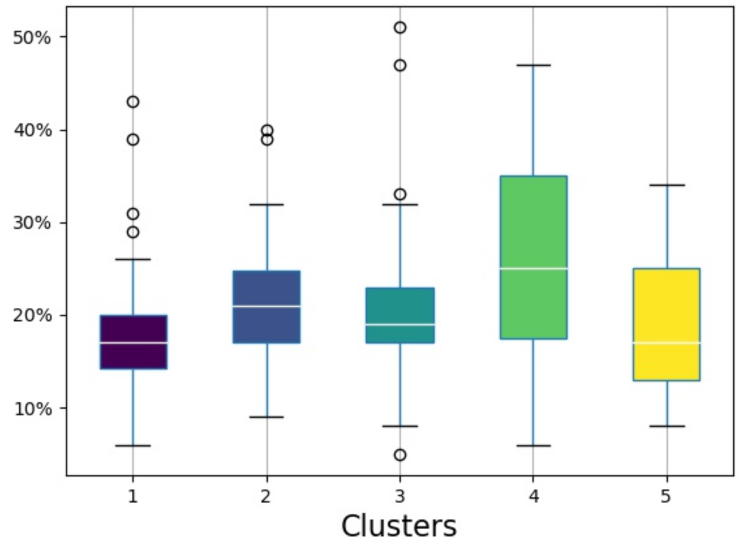


Figure 6: Gross Value Added (GVA) in manufacturing sectors over the total (%).

manufacturing to services in the last decades. Therefore, we might question, in a speculative way, the ability of the specialization patterns based on patents in mapping the current capabilities owned by these metropolitan areas.

Overall, we can say that the KCI is a promising tool for describing cities' economy, but not for all Metropolitan Regions. The use of information provided by the additional CA axes allows selecting regions for which this tool is meaningful and captures relevant technological specialization patterns. Therefore, the introduction of clusters that further summarize common patterns and technological characteristics improves the KCI itself and can help us in understanding future cities' competitiveness. Despite the KCI captures only 3% of the total variance in the data, it is, indeed, surprisingly able to predict the potential growth of a Metropolitan Region once we control, through the definition of clusters, for common characteristics of cities, as shown in Fig. 5b.

5 Conclusions

This chapter shows the relevance of the E.C. methodology to capture the hidden structures of complex evolving systems, such as the economy of a metropolitan area.

Despite its young age, this methodology developed quickly in the last decade and found important applications in the geography of innovation literature (Antonelli et al., 2017; 2020; Balland and Rigby, 2017; Petralia et al., 2017; Pintar and Scherngell, 2021). Moreover, given its capacity to synthesize a large amount of information into a single index, it showed high potential as an economic-policy tool (Balland et al., 2019; Mealy and Coyle, 2021; Pugliese and Tübke, 2019).

With respect to the most of cited literature, we embrace and stay closer to the interpretation of the E.C. methodology as a dimensionality-reduction algorithm, as proposed by Mealy et al. (2019). By extending the work by van Dam et al. (2021), we show the relevance of additional dimensions beyond the first leading eigenvector, to the analysis of which the rest of the literature is limited. In this way, we showed that the E.C. framework offers more than a single indicator. Instead, as recently underlined also by Hidalgo (2021), it is a powerful methodology to reveal the different facets of economic systems, extracting them from granular data. More specifically, we have shown that, by introducing clusters based on additional CA axes, we can identify for which cities the KCI is able to capture relevant technological specialization patterns, and for which it does not. While the KCI explains only a fraction of the total variance in the data, the introduction of clusters, and therefore of the additional information provided by the dimensionality reduction algorithm, dramatically increases the predicting power of this tool.

Furthermore, the reading of the E.C. methodology as a dimensionality-reduction algorithm lets us not only to reconcile it with the broader complexity economics literature and help proper interpretations of the empirical findings. It also helps to keep developing this promising tool. On this trail, we believe that, as shown in this chapter, the use of an even further amount of information, contained in the eigenvectors after the leading one (what we called KCI), can offer further development to this literature.

References

Aistleitner, M., C. Gräbner, and A. Hornykewycz (2021). "Theory and empirics of capability accumulation: Implications for macroeconomic modeling". *Research Policy* 50.6, p. 104258. DOI: 10.1016/j.respol.2021.104258.

Antonelli, C. (2015). "Innovation as a creative response: A reappraisal of the Schumpeterian legacy". *History of Economic Ideas* 23.2, pp. 99–118. DOI: 10.1400/236719.

Antonelli, C., F. Crespi, C. A. Mongeau Ospina, and G. Scellato (2017). "Knowledge composition, Jacobs externalities and innovation performance in European regions". *Regional Studies* 51.11, pp. 1708–1720. DOI: 10.1080/00343404.2016.1217405.

Antonelli, C., F. Crespi, and F. Quatraro (2020). "Knowledge complexity and the mechanisms of knowledge generation and exploitation: The European evidence". *Research Policy*, p. 104081. DOI: 10.1016/j.respol.2020.104081.

Arthur, W. B. (1999). "Complexity and the economy". *Science* 284.5411, pp. 107–109. DOI: 10.1126/science.284.5411.107.

– (2009). *The Nature of Technology: What It Is and How It Evolves*. Free Press. ISBN: 9781416544050.

– (2021). "Foundations of complexity economics". *Nature Reviews Physics* 3.2, pp. 136–145. DOI: 10.1038/s42254-020-00273-3.

Balland, P.-A., R. A. Boschma, J. Crespo, and D. L. Rigby (2019). "Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification". *Regional Studies* 53.9, pp. 1252–1268. DOI: 10.1080/00343404.2018.1437900.

Balland, P.-A., T. Broekel, D. Diodato, E. Giuliani, R. Hausmann, N. O'Clery, and D. Rigby (2022). "The new paradigm of economic complexity". *Research Policy* 51.3, p. 104450. DOI: 10.1016/j.respol.2021.104450.

Balland, P.-A. and D. L. Rigby (2017). "The geography of complex knowledge". *Economic Geography* 93.1, pp. 1–23. DOI: 10.1080/00130095.2016.1205947.

Batty, M. (2013). *The New Science of Cities*. The MIT Press. ISBN: 9780262019521.

Berkes, E. and R. Gaetani (2020). "The geography of unconventional innovation". *The Economic Journal*, ueaa111. DOI: 10.1093/ej/ueaa111.

Bettencourt, L. M. A. (2013). "The origins of scaling in cities". *Science* 340.6139, pp. 1438–1441. DOI: 10.1126/science.1235823.

Bettencourt, L. M. A., J. Lobo, and D. Strumsky (2007). "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size". *Research Policy* 36.1, pp. 107–120. DOI: 10.1016/j.respol.2006.09.026.

Breschi, S. and F. Lissoni (2009). "Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows". *Journal of Economic Geography* 9.4, pp. 439–468. DOI: 10.1093/jeg/lbp008.

Bustos, S., C. Gomez, R. Hausmann, and C. A. Hidalgo (2012). "The dynamics of nestedness predicts the evolution of industrial ecosystems". *PLOS ONE* 7.11, e49393. DOI: 10.1371/journal.pone.0049393.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). "Urban density and the rate of invention". *Journal of Urban Economics* 61.3, pp. 389–419. DOI: 10.1016/j.jue.2006.08.003.

Cristelli, M., A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero (2013). "Measuring the intangibles: A metrics for the economic complexity of countries and products". *PLOS ONE* 8.8, e70726. DOI: 10.1371/journal.pone.0070726.

van Dam, A., M. Dekker, K. van Eijden, J. de Bruin, and R. Schram (n.d.). *Spectral Clustering Correspondence Analysis in R*. Version 0.1.0. URL: https://github.com/UtrechtUniversity/SCCA.

van Dam, A., M. Dekker, I. Morales-Castilla, M. Á. Rodrı'guez, D. Wichmann, and M. Baudena (2021). "Correspondence analysis, spectral clustering and graph embedding: Applications to ecology and economic complexity". *Scientific Reports* 11.1, p. 8926. DOI: 10.1038/s41598-021-87971-9.

Eurostat (2020). *ARDECO database*. Version 2020/04. URL: https://knowledge4policy.ec.europa.eu/territorial/ardeco-online_en (visited on 01/31/2021).

Florida, R. L. (2002). *The Rise of the Creative Class. And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books. ISBN: 9780465024766.

Florida, R. L., P. Adler, and C. Mellander (2017). "The city as innovation machine". *Regional Studies* 51.1, pp. 86–96. DOI: 10.1080/00343404.2016.1255324.

Fontana, M. (2010). "Can neoclassical economics handle complexity? The fallacy of the oil spot dynamic". *Journal of Economic Behavior & Organization* 76.3, pp. 584–596. DOI: 10.1016/j.jebo.2010.08.010.

Foray, D. (2004). *Economics of Knowledge*. The MIT Press. ISBN: 9780262062398.

Giersch, H., ed. (1995). *Urban Agglomeration and Economic Growth*. Springer. ISBN: 9783642793998.

Glaeser, E. L. (1999). "Learning in cities". *Journal of Urban Economics* 46.2, pp. 254–277. DOI: 10.1006/juec.1998.2121.

– (2011). *Triumph of the City. How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Macmillan. ISBN: 9780230709386.

Gomez-Lievano, A. (2018). "Methods and Concepts in Economic Complexity". mimeo.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press. ISBN: 0122990501.

Greenacre, M. J. and T. Hastie (1987). "The geometric interpretation of correspondence analysis". *Journal of the American Statistical Association* 82.398, pp. 437–447. DOI: 10.1080/01621459.1987.10478446.

Griliches, Z. (1990). "Patent statistics as economic indicators: A survey". *Journal of Economic Literature* 28.4, pp. 1661–1707.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). *The NBER patent citation data file: Lessons, insights and methodological tools*. Working Paper 8498. National Bureau of Economic Research. DOI: 10.3386/w8498.

Hartigan, J. A. (1975). *Clustering Algorithms*. J. Wiley & Sons. ISBN: 9780471356455.

Hausmann, R. (2016). "Economic development and the accumulation of know-how". *Welsh Economic Review* 24, pp. 13–16. DOI: 10.18573/j.2016.10049.

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, and M. A. Yildirim (2014). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. 2nd ed. (1st ed. 2011). The MIT Press. ISBN: 9780262525428.

Henderson, J. V. and J.-F. Thisse, eds. (2004). *Handbook of Regional and Urban Economics*. Vol. 4. North-Holland. ISBN: 9780444509673.

Henrich, J. (2004). "Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses—The Tasmanian case". *American Antiquity* 69.2, pp. 197–214. DOI: 10.2307/4128416.

Hidalgo, C. A. (2021). "Economic complexity theory and applications". *Nature Reviews Physics*. DOI: 10.1038/s42254-020-00275-1.

– (2023). "Knowledge is Non-fungible". In: *Handbook of Complexity Economics*. Ed. by P. Chen, W. Elsner, and A. Pyka. London and New York: Routledge. Chap. II.1.3.

Hidalgo, C. A. and R. Hausmann (2009). "The building blocks of economic complexity". *Proceedings of the National Academy of Sciences of the United States of America* 106.26, pp. 10570–10575. DOI: 10.1073/pnas.0900943106.

Hill, M. O. (1974). "Correspondence Analysis: A neglected multivariate method". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 23.3, pp. 340–354. DOI: 10.2307/2347127.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. Random House.

Jaffe, A. B. and M. Trajtenberg (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. The MIT Press. ISBN: 9780262100953.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations". *The Quarterly Journal of Economics* 108.3, pp. 577–598. DOI: 10.2307/2118401.

Jones, C. I. and P. M. Romer (2010). "The new Kaldor facts: Ideas, institutions, population, and human capital". *American Economic Journal. Macroeconomics* 2.1, pp. 224–245. DOI: 10.1257/mac.2.1.224.

Jonhson, S., V. Domínguez-García, and M. A. Muñoz (2013). "Factors determining nestedness in complex networks". *PLOS ONE* 8.9, pp. 1–7. DOI: 10.1371/journal.pone.0074025.

Lü, L., S. Xu, and N. Xu (2023). "Complexity Science in the Application of Big Data Economics". In: *Handbook of Complexity Economics*. Ed. by P. Chen, W. Elsner, and A. Pyka. London and New York: Routledge. Chap. II.2.3.

Lucas, R. E. (1988). "On the mechanics of economic development". *Journal of Monetary Economics* 22.1, pp. 3–42. DOI: 10.1016/0304-3932(88)90168-7.

Mealy, P. and D. Coyle (2021). "To them that hath: Economic complexity and local industrial strategy in the UK". *International Tax and Public Finance*. DOI: 10.1007/s10797-021-09667-0.

Mealy, P., J. D. Farmer, and A. Teytelboym (2019). "Interpreting economic complexity". *Science Advances* 5.1, eaau1705. DOI: 10.1126/sciadv.aau1705.

Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press. ISBN: 9780691094830.

Moretti, E. (2019). *The effect of high-tech clusters on the productivity of top inventors*. 26270. Working Papers of the National Bureau of Economic Research. DOI: 10.3386/w26270.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto (CA): University of Toronto Press. DOI: 10.3138/9781487577995.

– (2006). *Multidimensional Nonlinear Descriptive Analysis*. Chapman and Hall/CRC. DOI: 10.1201/9781420011203.

Organization for Economic Co-operation and Development (2012). *Redefining "Urban": A New Way to Measure Metropolitan Areas*. OECD Publishing. ISBN: 9789264174054.

Organization for Economic Co-operation and Development (2013a). *Measuring Patent Quality: Indicators of Technological and Economic Value*. DOI: 10.1787/5k4522wkw1r8-en.
– (2013b). *Metropolitan Areas*. OECD Regional Statistics (database). DOI: 10.1787/data-00531-en. URL: https://www.oecd-ilibrary.org/content/data/data-00531-en (visited on 05/31/2018).
– (2015). *The Metropolitan Century: Understanding Urbanisation and its Consequences*. Paris: OECD Publishing. ISBN: 978-92-64-22873-3. DOI: 10.1787/9789264228733-en.
– (2020). *REGPAT database*. Version 2020/01.
Petralia, S., P.-A. Balland, and A. Morrison (2017). "Climbing the ladder of technological development". *Research Policy* 46.5, pp. 956–969. DOI: 10.1016/j.respol.2017.03.012.
Pintar, N. and T. Scherngell (2021). "The complex nature of regional knowledge production: Evidence on European regions". *Research Policy*, p. 104170. DOI: 10.1016/j.respol.2020.104170.
Polanyi, M. (1966). *The Tacit Dimension*. Routledge.
Porter, M. E. (1990). *The Competitive Advantage of Nations*. Free Press. ISBN: 9780029253618.
– (1998). *On Competition*. Harvard Business School Publishing. ISBN: 0875847951.
Prigogine, I. (1977). *Nobel Lecture*. Nobel Media AB. URL: https://www.nobelprize.org/prizes/chemistry/1977/prigogine/lecture/ (visited on 04/19/2021).
Pugliese, E. and A. Tübke (2019). *Economic Complexity to Address Current Challenges in Innovation Systems: A Novel Empirical Strategy for Regional Development*. Industrial R&I – JRC Policy Insights.
Richerson, P. J. and R. Boyd (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. The University of Chicago Press.
Romer, P. M. (1990). "Endogenous technological change". *Journal of Political Economy* 98.5, Part 2, S71–S102. DOI: 10.1086/261725.
Saracco, F., R. Di Clemente, A. Gabrielli, and T. Squartini (2015). "Randomizing bipartite networks: The case of the world trade web". *Scientific Reports* 5.10595. DOI: 10.1038/srep10595.
Schetter, U. (2019). *A Structural Ranking of Economic Complexity*. 119. Working Papers of the Center for International Development at Harvard University.
Schmoch, U. (2008). *Concept of a Technology Classification for Country Comparisons*. World Intellectual Property Organisation (WIPO).
Schumpeter, J. A. (1911). *Theorie der Wirtschaftlichen Entwicklung*. Translated by Redvers Opie as *The Theory of Economic Development* (1934) from the 2nd German edition of 1926. Duncker & Humblot.
– (1947). "The Creative Response in Economic History". *The Journal of Economic History* 7.2, pp. 149–159. DOI: 10.1017/S0022050700054279.
Soete, L. G. and S. M. E. Wyatt (1983). "The use of foreign patenting as an internationally comparable science and technology output indicator". *Scientometrics* 5.1, pp. 31–54. DOI: 10.1007/bf02097176.
Straka, M. J. (2018). "Entropy-Based Methods for the Statistical Validation of Bipartite Networks". PhD thesis. IMT School for Advanced Studies.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2012). "A new metrics for countries' fitness and products' complexity". *Scientific Reports* 2.723, pp. 1–7. DOI: 10.1038/srep00723.

Trajtenberg, M. (1990). "A penny for your quotes: Patent citations and the value of innovations". *The RAND Journal of Economics* 21.1, pp. 172–187. DOI: 10.2307/2555502.

United Nations Human Settlements Programme (2016). *World Cities Report 2016. Urbanization and Development: Emerging Futures*. United Nations Publications. ISBN: 9211327083.

United Nations, Department of Economic and Social Affairs (2019). *World Urbanization Prospects 2018. Highlights*. United Nations Publications. ISBN: 9211483182.

Weaver, W. (1948). "Science and complexity". *American Scientist* 36.4, pp. 536–544.