UNIVERSITÀ DEGLI STUDI DI TORINO

CARLO BOTTAI

ESSAYS ON THE ECONOMICS AND GEOGRAPHY OF TECHNOLOGICAL DIVERSIFICATION

Vilfredo Pareto Doctorate in Economics
Cycle XXXI

Contents

Measures of Regional Technological Diversity
A Critical Review

Carlo Bottai

Department of Economics and Statistics *Cognetti de Martiis*, University of Turin
BRICK, Collegio Carlo Alberto

Abstract

Economists and Geographers have been debating long since about the role of the regional economies diversity on their economic performance. If the existence of a diversity-development nexus is well established and acknowledged, the measurement of the concept in empirical studies is problematic. Moreover, most recently *diversity* has been framed as a facetted concept, with different and interrelated dimension: *variety*, *balance*, *disparity*, and *rarity*. This has problematised even more the measurement and operationalisation issues. In this paper, firstly some of the measures more broadly used in the empirical literature –Related-Unrelated Variety, Coherence, Economic Complexity, and Fitness– are presented. Then, each is critically reviewed, and its main limitations highlighted. Lastly, some solutions to overcome these drawbacks are proposed, introducing alternative some indices: namely, Evenness and its within-between-groups decomposition; a Coherence measure based on a *null model* that constraints both the margins of the regions-technologies occurrence matrix; Complexity and Fitness indices computed on weighted matrices; and a Rarity-weighted diversity measure. And, using data about the patenting activity of the European regions, each of the measures is compared to its possible alternative, discussing the results. Lastly, an empirical application that fits within the so-called regional *resilience* literature is proposed as a tool to test an clarify the ideas introduced. The findings suggest that, by solving the issues and drawbacks raised, it is possible to distinguish more clearly, in the empirical applications, between different components of the regional *diversity.* And that this helps to go deeply in the analysis of the diversity-development nexus, accounting for the contribution of each of the various aspects that compose such a faceted concept.

Keywords— Regional economic development, Technology, Diversity, Measures and indices

1 Introduction

The role of regional diversity for the development potentials of an economic system has been analysed and discussed long since in the Economic Geography literature. If the existence of a diversity-development nexus is broadly accepted, the measurement of the concept in empirical studies is problematic. Moreover, most recently diversity has been framed as a facetted concept, with different and interrelated dimension: *variety*, *balance*, *disparity*. Moreover, the *rarity* of the elements of the regional *knowledge capital* has been identified as an additional orthogonal dimension that must be accounted together with the regional (technological) *diversity* to understand the performance and development potentials of regional economies. This has made even more problematic the measurement of this concept in empirical applications.

In this paper we will review and critically analyse the following measures that have been introduced in the empirical literature to operationalise some of the ideas just briefly remembered: (i) Related-Unrelated Variety (Frenken et al. 2007);[1] (ii) Coherence (Nesta and Saviotti 2005); (iii) Economic Complexity (Hidalgo and Hausmann 2009) and Fitness (Cristelli et al. 2013; Tacchella et al. 2012). The first is a decomposition of an entropy index that splits it in a *within-groups diversity*, that measures how much the productions of a region can be grouped in highly related blocks, and a *between-groups* one, that accounts for a stronger type of *diversity* among these blocks. The second is a measure of the complementarity level among the different elements of a set, or of the average epistemic relatedness of any technological domain that a region has developed to any other developed element. Therefore, it focuses on the *complementarities* among the elements of the system, which is the true power source in a *recombinant innovation* framework. The latter two, even though in a different way, take into account not only the diversity of the production of an economic system, but also the ubiquity of these production among all the other economies considered, with the idea that a combination of the two dimensions will be a good indicator of the *capabilities* available in a given economic system.

The paper is organised in four main sections. Sec. 2 frames the literature about regional diversity and economic development within a classification mainly based on what has been highlighted within Science of Science about diversity and interdisciplinarity. Sec. 3 reviews the measures just remembered. While in Sec. 5 the main critical issues of each of these indices will be highlighted and some solution to each issue will be proposed. Moreover, using data about the patenting activity of the European regions, I will show which advantages each alternative measure proposed offers. Lastly, Sec. 6 proposes an empirical application of the ideas and measured introduced. This exercise frames within the so-called *resilience* literature that aims at identifying the preconditions that helped some European regions to react better than others in front of the recent Great Recession. The findings of this last section corroborate the analysis carried on in the previous part. Therefore, the results of this paper suggest that, by solving the issues and drawbacks raised, it is possible to distinguish more clearly, in the empirical applications, between different components of the regional *diversity*. And, given the classification proposed in Sec. 2, this opportunity seems something desirable to go more deeply in the analysis of the diversity-development nexus, accounting for the contribution of each of the various aspects that compose such a faceted concept.[2]

---

[1] To avoid as much as possible terminological confusions I will use the word *variety* to speak about Stirling's concept, and the word Variety, for the idea proposed in Frenken et al. 2007.

[2] All the measures introduced in the paper have been included in an R package, `ReKS`, freely available on the Internet. See `https://github.com/n3ssuno/ReKS` and Appendix A.

2 Diversity and regional economic development

Economists and Geographers have been debating long since about the economic role of the regional economies diversification structure and its effects on the performance of such economic systems. With a Schumpeterian framework in mind, we can say that *economic development* is something more than *economic growth*, since it is not just a question of increasing the output through an increase of production inputs or of their productivity, but it consists also in the introduction of new kind of activities and products in the economic system. Indeed, as Lucas (1993, p. 263) said «[a] growth miracle sustained for a period of decades clearly must thus involve the continual introduction of new goods, not merely continued learning on a fixed set of goods» and as Schumpeter (1983, ch. 2 n. 6) reminds us, you can «[a]dd successively as many mail coaches as you please, you will never get a railway thereby». Also Pasinetti (1993) went in the same direction when he stated that an economy has to increase its variety over time in order to trigger productive gains and absorb structural unemployment due to the combination of product innovation and technical progress in production. And, as Hidalgo and Hausmann (2009) remind us, since Adam Smith, the wealth of nations has been related to the division of labour, and this also because it is related to the *complexity* that emerges from the interactions among the individual economic agents. In other words, a larger and more diversified economy is supposed to be wealthier and to have higher labour productivity, also thanks to the division of labour and knowledge that it allows (Metcalfe 2010; Smith 1776). Therefore, as underlined by Saviotti (1996), the increase in variety of activities and goods available in an economy is one of the fundamental trends in economic development. Therefore, we can say that the existence of a diversity-development nexus is broadly acknowledged. And even though the causal direction of this relationship is still unclear, this is not only far beyond the scope of this paper, but it is also likely that we are in front of a case of circular causation that a Complex Adaptive Systems approach is more suitable to describe.

Regional economic diversity has been defined as «the presence in an area of a great number of different types of industries» (Rodgers 1957, p. 16) or as «the extent to which the economic activity of a region is distributed among a number of categories» (Parr 1965, p. 22). Likewise, Saviotti (1996, p. 92) defined *variety* as «the number of *distinguishable* types of actors, activities and outputs required to characterise an economic system», and citing Pielou (1977) he highlights the similarity of this idea with biologists mean speaking of *diversity.* And these ideas have been recently extended also to the regional knowledge bases, not least because of the increasing economic relevance of knowledge, technology and innovation for economic performance in advanced countries (Freeman and Soete 1997).

As underlined by Saviotti (1996, p. 142), if we depict an economic system as a collection of elements, a simple measure that accounts for this type of regional diversity, that stays close to Information Theory and to a *recombinant innovation* framework as implicitly done in what follows, is

$$\text{variety} = \log_2 n,$$

where n is the number of distinguishable available elements in the economic system. However, more recently this type of definitions have been questioned as too simplistic and the term has been opened up, looking more carefully at the different aspects of regional diversity. An economic system is not just composed of "distinguishable elements". A knowledge-based economy, *knowledge capital* is not a homogeneous substance.[3] And, even though both Marshall and Jacobs externalities account for the positive effects of the regional clustering of activities, the latter is considered particularly relevant for growth and development in the knowledge-based economic

[3]The *knowledge capital* is defined as the stock of knowledge that firms or regions accumulate over time, and is mainly measured by the literature either as R&D expenditures, or as number of patents applications.

systems, and it focuses exactly on the composition of activities of the region. Moreover, the different elements of this bundle, as productive inputs, have to be used in different combinations, so that also the relations and proportions between them matter. Therefore, a growing number of papers are looking at the composition of this bundle of physical and human resources available in each geographical area, as well as to it structural characteristics, to qualify better regional diversity, no more as a monad, but as a facetted concept. In particular, following Antonelli et al. (2017), besides *variety*, two structural characteristics of the regional knowledge base –*relatedness* and *rarity*– let Jacobs knowledge externalities exert their positive –pecuniary– effects, reducing the costs of knowledge both as input and output.

In line with the seminal paper by Stirling (2007) and the contributions of Rafols and coauthors (Rafols 2014; Rafols and Meyer 2010), we can identify three major characteristics of the regional knowledge base: *variety*, *balance*, and *disparity*. In this framework, the *diversity* of the knowledge capital of a region grows if any of these aspects increases. In Stirling's terminology, *variety* is nothing more than the number of technological domains in which a region has some competences. Instead, *balance* measures the evenness of the distribution of elements across categories. Under this point of view, the more the elements of the set belongs to just one or few groups, the less the set is diversified. In Economic Geography empirical literature, the Shannon Entropy index has been broadly used to capture this idea, but, as clarified below, this type of measure confounds *variety* and *balance* in a unique measure.[4] Regarding what has been called in Ecology, Science of Science and interdisciplinarity studies *disparity*, the concept is closely related to what Economic Geographers have called *relatedness*. It measures the degree to which the categories of the elements of the regional knowledge capital are different from each other. The more related they are, the less we will be prone to say that (everything else being equal) the regional capabilities structure is diversified.

Moreover, as underlined by an even more recent literature (Antonelli et al. 2017; Balland and Rigby 2017; Hausmann and Hidalgo 2011), there is a second dimension of the knowledge base of regional economies that must be considered together with *diversity*. Indeed, the *rarity* of each of the technological items of these bundles of capabilities offers some useful information to understand the economic value of each domain.

2.1 Variety

With a specific focus on the technological component of the economic systems, the analysis of their variety was pioneered by Archibugi and Pianta (1992a) and Pianta and Meliciani (1996) who investigated the role of technological specialisation across patent classes at the country level. Their findings are partly in contrast with what just said about the sectoral organisation of the economies. Indeed, the more performing countries seem to be these whose technological structure is more focused on the few highly productive technological domains in which each country has developed some competitive advantages. Moreover, only the bigger countries turn out to be able to diversify in many different fields –and this in line with the just highlighted strong linkage that the division of knowledge puts between the technological diversity and the population size.

2.2 Unrelated balance

As clarified by Attaran (1986), regional diversification can be thought in analogy with a portfolio diversification strategy (see also Barth et al. 1975; Conroy 1972, 1975b, among others). And the same idea, that the more diversified a region's industrial structure is, the less subject to

[4] Another widely used measure is the Simpson index (also known as HerfindahlHirschman index), but we will not consider it in this paper.

fluctuations caused by changes in extra-regional factors its economy will be, has been debated since long within Regional Business Cycle literature (see e.g. Hoover and Fisher 1949; Nourse 1968; Richardson 1969). This because a more specialised economic system will be less able to cushion adverse cyclical effects and idiosyncratic external shocks, like oil prices changes or the introduction of new technologies; and because the «market for its speciality might be undercut by discovery of new and cheaper supply sources, by improvements in production elsewhere, by improvement in transportation, or by shifts in demand» (Attaran 1984, p. 2). Therefore, make the other factors equal, the higher the *balance* of the different capabilities groups developed by a region, the higher the protection against idiosyncratic shocks. Conversely, a highly diversified structure, but in which most of the activities are concentrated within only one (or few) of the technological classes developed by the region (or groups of these) offers a low protection against this type of external shocks, because the small other groups will not be able to compensate for the negative performance of the bigger technological domain(s) if shocked.

Moreover, this last type of diversification let the local economic system able to give rise to both strong Marshallian and Jacobs externalities, that are very effective and powerful in the short-term (Castaldi et al. 2014). Therefore, we expect that a lower *unrelated balance*, helps to reduce the risk to be hit by sectoral shocks and, by helping stronger related recombinations, increases regional performance in after-shock periods.

However, as remembered by Boschma (2015), specialised regions can be considered less at risk to be shocked, since they are exposed to a smaller number of possible idiosyncratic shocks.[5] And the effectiveness of the diversification strategy will also depend on the degree of inter-relatedness between the components of the bundle of activities and technological domains in which a region is involved (Conroy 1975a; Diodato and Weterings 2015). Tab. 1 summarises these points.

Table 1: Variety and unrelated balance

	Low unrelated balance	High unrelated balance
Low variety	*Mono-specialisation* Low probability to be hit by idiosyncratic shocks, but lack of protection against them once shocked; and strong Marshall externalities	*Multi-specialisation* Almost null protection against idiosyncratic shocks, but strong Marshall externalities
High variety	*Unbalanced diversity* Low protection against idiosyncratic shocks, but strong Marshall and Jacobs externalities	*Balanced diversity* High protection against idiosyncratic shocks, strong Jacobs, but low Marshall externalities

2.3 Related balance and organised complexity

At list starting from the seminal contribution of Frenken et al. (2007), many Evolutionary Economic Geographers and Economists of Innovation argue that it is not variety *per se*, but the *coherence* among the pieces of knowledge in the available stock, that helps the process of innovation through *knowledge recombination* (e.g., Antonelli et al. 2010; Frenken and Boschma 2007; Quatraro 2010). And this is true looking at both the sectoral and technological composition of a given local economic system. Indeed, in line with the Multi-product (Teece 1980, 1982; Willig 1979) and the Resource-based (Penrose 1959) theories of the firm and also with the Hirschman's theory of Economic development (Hirschman 1958), it is expected that an increase in the *diversity* of the structural composition of a country or region will help its economic growth. But,

[5] A topic well known in Network Science, that has widely underlined the advantages and disadvantages of different topological structures of networks against external attacks (Albert et al. 2000).

since the advantages of a higher diversity happens mainly thanks to the *spatially constrained external economies of scope* or *of complexity* (Parr 2002) that occurs within it,[6] some degree of *relatedness* among these different productions is needed for this to happen (Montgomery 1979; Montgomery and Hariharan 1991; Nesta and Saviotti 2005; Ramanujam and Varadarajan 1989; Teece et al. 1994). In other words, within a given geographical area there are some localised but shareable resources that can be exploited by firms and other economic agents, in combination to their internal resources,[7] so to achieve their productive purposes (Panzar and Willig 1981; Teece 1980). In order to exploit the spatially concentrated external *economies of scope* and *complexity* that arise from these shared physical and human resources, and more in general to exploit the productive services performed both by each piece of knowledge *per se* and by their combined use, it is needed at least some degree of *relatedness* among these *internal* and *external* resources. Indeed, if the external resources are not enough similar to the internal ones, they will not be useful or understandable and the two sources cannot be combined together in order to exploit their mutual Edgeworth complementarities (Weber 2005).

Under these conditions, the heterogeneous and complementary knowledge components are available at low absorption costs, and this makes easier and cheaper to get technological innovations via a recombination of these items. At the firm level, Teece et al. (1994) have underlined that the reason for enterprises to keep as much *relatedness* as possible when they enter into new business lines is that *diversification* comes at costs. In order to contain this increase in costs, firms must devote part of their focus towards integrating these new sets of activities, competencies and technological knowledge with pre-existing ones. Therefore, *diversification* inherently calls for some sort of integration to increase the *relatedness* of the firms activities and the underlying knowledge base (Breschi et al. 2003), since it is well recognised that *economies of scope* or *complexity* arise when similar productive sequences are shared among several business lines or where, the productive activities across businesses, vertically integrate complementary activities and competencies.

Although it may seem that this reasoning unduly confuse a firm and a local economic system as if they were the same, whenever there were economies that can be exploited only by co-localised firms, although achievable using contracts, we can see so strong analogy among the two that can justify it: «when [...] contracts can be devised for sharing of inputs by independent firms and when the sharing also requires spatial proximity, we have the case of a spatially concentrated external economy of scope, which represents, in essence, an agglomeration economy of the urbanization type» (Parr 2002, p. 159). If these two requirements are satisfied, also local economic systems can contain the costs of development-through-diversification thanks to the access to *external economies* that can arise from a *lateral* or *vertical integration* of their activities and the justification for nonrandom *relatedness* in the process of *diversification* apply also to these spatially concentrated economies.[8]

[6] As explained by Parr (2002, p. 155), with respect to spatially constrained economies external to a single firm «[w]hereas scope is concerned with the multiproduct nature of the output, the dimension of complexity refers to the multiprocess or the multi-input nature of production and, more generally, to the fact that a firms production involves several technologically separable stages». Therefore, while the former type gives rise to a production structure with multiple end products (*lateral integration*), the latter gives rise to a structure characterised by several stages or processes needed to get the end product (*vertical integration*).

[7] I mean a positive level complementarity both within and between these two fundamental components (Antonelli and Colombelli 2015; Johansson and Lööf 2014; Patrucco 2008, 2009). Indeed, since no economic actor is able to command the whole existing knowledge, the Recombination Generation hypothesis (Weitzman 1996, 1998) implies a multiplicative relationship both between knowledge pieces and, at the firm level, between internal and external knowledge. This means, at the level of a local economy, that the more the firms co-localised within a common neighbourhood share similar characteristics with each other, the easier will be for one of those to get access to the internal resources of the other firms, and then the more powerful the possibility to recombine these last with their own resources to get something new at a lower cost.

[8] Moreover, the expansion process of a regional economy can happen also through imitation. In this case, we

Therefore, looking at branching-out expansion of a regional economy you must consider that *diversification* necessarily happens at cost, since different branches cannot fully use others' sharable inputs –knowledge, in particular–, products and byproducts. Indeed, it has been shown –using main different data sources and in main different contexts– that there exists a *relatedness principle* that essentially says that «the probability that a region enters (or exits) an economic activity [is] a function of the number of related activities present in that location» (Hidalgo et al. 2018, p. 452).

Moreover, stronger Jacobs knowledge externalities are found when the composition of the knowledge base of an economic system exhibits high levels of *organised complexity* (Antonelli 2011; Jacobs 1961; Schumpeter 1947) and, as such, «is able to provide cheaper access to and use conditions of the stock of quasi-public knowledge that is necessary for the recombinant generation of technological knowledge» (Antonelli et al. 2017, p. 1710).

Lastly, to help further developments of the system these costs, measurable for example through *information entropy*, must be contained with the organisation of the internal structure of the system (Hidalgo 2015; Mokyr 2002).[9] The main mechanism that let this costs containment is a direct consequence of this type of expansion. Indeed, an expansion in related sectors leads to a *path dependent* growth of the local economic systems, that gives rise to a lower growth of the entropy of the system, if compared to an ergodic process of fully-at-random expansion. In other words, *relatedness* and *knowledge coherence* help to contain the costs of an expansion-through-diversification.

Therefore, a higher *organised complexity* helps regions both in static and dynamic sense, by reducing the access cost to knowledge external to the firm, but already internal to the region, and by reducing the costs of search and access to the knowledge external to both the single firm and the region in which it is located. Tab. 2 outlines this idea.

Table 2: Variety, related balance and organized complexity

	Low relatedness Disorganised complexity	High relatedness Organised complexity
Low variety	*Incoherent specialisation* Low Marshallian and Jacobian externalities	*Coherent specialisation* Low recombination costs with high short term growth potentials.
High variety	*Incoherent diversity* Few opportunities to develop in the long term.	*Coherent diversity* Strong and powerful Marshallian and Jacobian externalities

2.4 Rarity

The argumentation proposed so far link the size of the bundle of knowledge capabilities of a region and its structural characteristics, with the resilience capacity and the growth rates of this economic system. However, what we are able to measure is the technological composition

must look at knowledge that is external not only to the single company, but also to the whole region. Also in this case, the access to external knowledge (with this different meaning) is not for free, so that again the closer to the internal one, the cheaper to copy, reuse and integrate with the existing one it will be. Indeed, it will be easy to understand it, to appraise its quality, to absorb it through cheaper and faster learning processes.

[9] This idea that an organised economic system is a key factor needed for firms to react *creatively* in front of out-of-equilibrium conditions (Schumpeter 1947) and therefore for economies to grow (or to *develop*, in Schumpeterian terms) is also present in Antonelli (2013, 2015), Antonelli and Ferraris (2011) and Antonelli and Scellato (2013), in a try to mix up the Schumpeterian and Marshallian legacies together (Metcalfe 2007, 2010).

of these economies, in terms of their patents stock. Slightly less than ten years ago a seminal contribution of Hidalgo and Hausmann (2009) has proposed to re-interpret the countries-exports bipartite network as the sign left by a tripartite network connecting countries to the capabilities they have and products to the capabilities they require. In this way the authors showed that it is possible to indirectly measure these *capabilities* by looking at their productions and activities.[10]

As Antonelli et al. (2017) have highlighted, a major contribution of this approach is to be able to qualify the composition of an economic system in terms of the *rarity* of its elements. Not only does the number of different activities of the regional knowledge bases and the structural characteristics of this complex bundle matter, but also the relative scarcity of each element has a primary role. Hence, the composition of the bundle of activities that are likely to engender high-level Jacobs knowledge externalities has to be qualified in terms of the rarity of its components. A bundle of knowledge items able to yield strong Jacobs externalities will include many rare activities. As said by Balland and Rigby (2017, p. 2), «For many firms and regions of the industrialized world, competitive advantage hinges on the production of high-value, nonubiquitous, complex and tacit knowledge».

Therefore, this idea seems an appropriate approach to grasp the pecuniary effects of the *organised complexity* of a system in terms of Jacobs knowledge externalities. And following the schema summarised by Tab. 3 we can say that

> [w]hen the variety of the bundle of activities is high, but it is able to include only ubiquitous products and competencies, the levels of Jacobs externalities are low. When the variety of the bundle of activities is high and the bundle includes rare items, there is strong likelihood that the levels of Jacobs externalities are high. When the variety of the bundle is low and includes only ubiquitous items, the levels of Jacobs externalities are deemed to stay low. When, finally, the variety of the bundle is small, but includes rare items, the levels of Jacobs externalities are likely to exhibit high levels of variance because, on the one hand, the limited variety reduces the working of the recombinant generation of technological knowledge, but, on the other hand, it can yield rare combinations that characterize the generation of radical new knowledge that yield high profits and total factor productivity increases with positive effects on output growth –Antonelli et al. 2017, p. 1711.

In other words, Antonelli et al. (2017) have shown that, next to the role played by the Marshall knowledge externalities (Antonelli and Colombelli 2015), also the Jacobs knowledge externalities have a key role in shaping knowledge generation at the regional level. In their opinion, this effect can be grasped by analysing the role played by qualified variety in knowledge composition as captured by indexes like the ones just said. Said differently, the *complexity* measures provide

[10]This first seminal paper has been followed by many other works basically based on the same key concepts, 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 and Hidalgo 2011; Hausmann et al. 2014). On the other hand, Pietronero and his coauthors proposed a similar measure called Fitness (Cristelli et al. 2013; Tacchella et al. 2012). All these founding contributions looked at the products exported by countries, while here the focus is on the patents produced by each European region. An important difference to be taken into account is that, while for trade data have been considered the value of the exports in monetary terms (even though the matrix is then put in a binary form using the Balassa index criterion), for patent data the quantities (number of patents in each technological class) have been used. A drawback of this choice is that it is less clear, in principle, the reason why some patents are less ubiquitous than others. For trade data, we are sure that the more rare items have a positive demand as well, and so we can say that this higher rarity is a signal of a higher complexity of the product considered. Conversely, for patent data, it can be that the higher rarity is due to a lower relevance, and demand, for this technological domain. So that more rare classes can be simply less useful, and not more complex ones. However, since patenting is not costless, and patents are supposed to be novelty, usefulness, and non-obviousness, we can be more confident that the quantity of the patents in a technological domain is an estimate of its value. I thank Ricardo Hausmann for making me realise this difference.

a synthetic indicator of the *diversity* of a region and of the *ubiquity* of its knowledge items (at the same time both inputs and outputs). So doing they are supposed to map different regions according to their ability to develop sophisticated, and thus more rare, technologies emerging where a large number of high-skilled individuals and specific technological competences are available.

Table 3: Variety and rarity

	Low rarity	High rarity
Low variety	*Poor specialisation* Low and poor Jacobs externalities	*Hyperspecialisation* Low but rich Jacobs externalities, thanks to the command of rare knowledge inputs
High variety	*Unqualified diversity* Strong but poor Jacobs externalities	*Qualified diversity* Strong and rich Jacobs externalities

3 Measures of regional (technological) diversity

In this section, we will define some of the measures more broadly used in the empirical literature to operationalise each of the dimensions and characteristics of the regional knowledge capital remembered in the previous section.

3.1 Related-Unrelated Variety

Since long *entropy indices* have been used as regional indicators to test whether industrial diversity reduces unemployment and promote growth (Attaran 1984, 1986; Hackbart and Anderson 1975).

Shannon entropy is a non-parametric statistical tool that, broadly speaking, describes the dishomogeneity of a distribution. We have maximum entropy when these particles move completely at random, while we have minimum entropy when all the particles are bounded on a given area. The former case can be thought as a flat probability distribution, while in the latter the probability to find a particle will be positive only in one area of the space: more in general, the higher the skewness of the probability distribution, the lower the entropy of the system. As explained by Frenken

> Entropy is thus a macroscopic measure at the level of a distribution that indicates the degree of randomness in the macro-dynamics underlying the frequency distribution. As such, entropy can be used as a variety measure of frequency distributions of technological design. [...] Maximum entropy corresponds to the case in which all designs occur with the same frequency. [...] A skewed distribution occurs when some designs dominate the product population. In that case, the frequency of some designs is high, while the frequency of most designs is low or zero —Frenken 2006, p. 69.

Indeed, even though it could seem counter-intuitive at a first glance, we will have the maximum disorder in case of a homogeneous distribution of the occurrences across all the possible events or classes of events. However, the reason why this is the case is straightforward once we look at the probabilistic interpretation of the entropy provided by Information Theory, starting from Claude Shannon (1948). In the 1960s, Henri Theil developed several applications of Information

Theory in Economics and other Social Sciences (Theil 1967, 1972), and as Frenken (2006, p. 70) remembers us «[t]he entropy formula expresses the expected information content or uncertainty of a probability distribution». Indeed, in Information Theory, the term *entropy* refers to information we do not have about a system, and so it is a measure of the uncertainty or unpredictability of that system: in other words, the higher the entropy, the more the system will be able to surprise us and, conversely, once we have received a new piece of information about the structure of the system, its entropy will diminish.

Since the occurrence of events with smaller probability is least expected, their realisation yields more information. Therefore, a measure of information h should be a decreasing function of p_i. Shannon (1948) proposed a logarithmic function

$$h(p_i) = \log_2\left(\frac{1}{p_i}\right)$$

which decreases from ∞ to 0 for p_i ranging from 0 to 1. The function reflects the idea that the lower the probability of an event occurring, the higher the amount of information of a message stating that the event occurred.

The expected information content of a probability distribution, called *entropy*, is derived by weighing the information values $h(p_i)$ by their respective probabilities

$$H = \sum_{i=i}^{n} p_i \log_2\left(\frac{1}{p_i}\right), \qquad \text{with } p_i \log_2\left(\frac{1}{p_i}\right) = 0, \text{ if } p_i = 0.$$

Therefore, $H \in [0; \log_2 n]$ and it will be minimised when only one event has a positive probability of happening, while it will reach its maximum when all states are equally probable. Moreover, we can notice that its maximum is an increasing function of the possible elements, but it increases in a decreasing way. Theil (1972) remarks that the entropy concept is similar to the variance of a random variable whose values are real numbers. The main difference is that entropy applies to quantitative rather than qualitative values, and, as such, depends exclusively on the probabilities of possible events.

3.1.1 Entropy Decomposition Theorem

Among others, one of the reasons of success of entropy as a measure of *diversity* is that, differently from most of the others proposed and thanks to its *additivity* property, it is decomposable in two sub-components: the within-groups entropy and the between-groups one (Attaran and Zwick 1987; Theil 1972; Zajdenweber 1972).

Let $E_1, \ldots, E_n$ be events that happen with probability $p_1, \ldots, p_n$, respectively. Assume that they can be aggregated in G groups, $S_1, \ldots, S_G$, so that each event exclusively falls under one of these sets. The probability that one event of the set S_g occurs is

$$P_g = \sum_{i \in S_g} p_i.$$

Therefore, the *between-group entropy* is given by

$$H_0 = \sum_{g=1}^{G} P_g \log_2\left(\frac{1}{P_g}\right).$$

Furthermore, it is possible to prove that the entropy H can be decomposed in two parts

$$H = \sum_{g=1}^{G} P_g \log_2\left(\frac{1}{P_g}\right) + \sum_{g=1}^{G} P_g \left(\sum_{i\in S_g} \frac{p_i}{P_g}\log_2\left(\frac{P_g}{p_i}\right)\right)$$
$$= H_0 + \sum_{g=1}^{G} P_g H_g,$$
$$\text{with } H_g = \sum_{i\in S_g} \frac{p_i}{P_g}\log_2\left(\frac{P_g}{p_i}\right), \qquad g = 1,\ldots,G,$$

where H_g is the entropy within the set S_g and the second right-hand term of the equation is the average *within-group entropy.*

Within the Economic Geography literature, this decomposition has been made famous by Frenken et al. (2007). The authors called *Related Variety* the former and *Unrelated Variety* the latter. In short, they claimed that Related Variety measures the stronger knowledge spillovers possible among related sub-sectors. And that Unrelated Variety estimates the benefits of having a wide portfolio of uncorrelated sectors that protect an economic system against idiosyncratic shocks. As explained by Content and Frenken (2016, p. 2097), the concept was introduced by Frenken et al. (2007) precisely «in an attempt to resolve an earlier empirical question put forward by Glaeser, Kallal, Scheinkman, and Shleifer (1992) whether regions benefit most from being specialized or being diversified». By disentangling diversity in two types, the authors claimed that it is not diversity as such, but diversity in related industries that enhances knowledge spillovers and has positive effects on employment growth thus highlighting spillovers among sectors that are cognitively proximate. In other words, Frenken et al. (2007) agreed with Jacobs that innovation is essentially a *recombinant* process, so that a more *diversified* structure helps a region to grow quickly and strongly. However, the notion of *relatedness* let them to take into account that some pieces of knowledge and artefacts are much easier to recombine together than others. That is, some sort of *specialisation* is helpful alike, even though this happens not in terms of just one sector or technological domain, but around a group of industries and technologies similar to each other. [11]

The original hypothesis advanced by Frenken et al. (2007), and tested by most of the following literature (see Content and Frenken 2016 for a comprehensive review), was that Related Variety would spur employment growth, as new combinations lead to new products or services (product innovation), and so to new jobs: roughly, the mechanism proposed goes from Related Variety to incremental innovations and then to employment growth. Conversely, the MAR localisation economies stemming from the spatial concentration of firms in the same industry would help process innovation, as specialised knowledge is used to optimise production processes in existing value chains: such innovations spur labour productivity and do not necessarily lead to employment growth. The same paper argued also that, instead, Unrelated Variety is expected to decrease unemployment growth. In this respect, Unrelated Variety can be described as a measure of risk-spreading that appeases the effects of an external sector-specific shock in demand: specialisation in one or in few (related) sectors will result in the opposite scenario, as the region

[11] As explained by Teece (1980), *economies of scope* make product diversification efficient only if they are based on the common and recurrent use of proprietary know-how or on an indivisible physical asset. When "translated" into a regional framework, each of the business branches of the multi-product firm is the firms located in the region. Therefore, the *geographies of scope* can work only if the different pieces of technological knowledge are not too different, so that their recombination can happen at low costs.

is exposed to the probability of a severe slowdown if a key sector will be hit by the shock.[12]

Despite the broad application this index has found in the empirical literature, an important drawback of the Related-Unrelated Variety index, already well documented by the literature (see e.g., Boschma et al. 2012; Content and Frenken 2016; Rocchetta and Mina 2019), is that the distinction between the two components is based on the assumption that any pair of entities included within the same group are generally more closely related, or more similar, to each other than any pair of entities included in two different groups, with the following assumption that it will be possible to observe stronger knowledge spillovers within these groups, than between groups. But, since the grouping is based on the tree structure of a classification system –like the Statistical Classification of Economic Activities in the European Community (NACE) or, as done here, the International Patent Classification (IPC)–, the results are highly dependent on that hierarchical structure, too. For this reason, this measure will be able to capture essentially only the components of the technological relatedness incorporated in this tree structure –and for this reason Boschma et al. (2012) call it a measure of "*ex ante* relatedness"–, while it underestimates other broader notions of relatedness –like the epistemic similarity between two knowledge items, or the complementarities between pairs of technological components once combined together.

3.2 Regional Coherence

For this reason, other measures of *ex post* relatedness are frequently coupled to the Related Variety index, since they are able to capture different aspects of the *relatedness* of the regional knowledge base. A possible measure of *ex post relatedness* strength is called *Coherence*. It is possible to define the Coherence of the regional knowledge base (or *knowledge integration*) as the extent to which the technologies held by firms, workers and other economic actors within a geographical area are related to each other (Nesta and Saviotti 2005, 2006). The idea of *knowledge integration* highlights the fundamental role of the knowledge capital dishomogeneity: indeed, this last characteristic is a production service in itself, through the combinatorial opportunities it offers and the non-random character of the knowledge accumulation and articulation. Therefore, *knowledge integration* is the expression that something fundamentally not at random guides the accumulation and formation of the regional knowledge capital (Henderson 1994). Therefore, this measure explores a quite different aspect of the composition of the regional knowledge bases compared to what is done by the Related Variety index. Looking at the patents developed within a geographical area, and according to the empirical studies that explored this dimension, an economic or technological system has better performance for high levels of technological coherence (among others, Antonelli et al. 2010; Quatraro 2010; Rocchetta and Mina 2019). Indeed, as said, it is expected that some degree of *relatedness* helps the recombination of the regional knowledge base components, and so the growth-through-innovation of the area. Moreover, the existence of complementarities between these different components is expected to enhance the regional productivity, because a region with a more integrated knowledge base will be able to exploit easily and strongly the synergies between its (complementary) competences.

Operatively, the construction of the Coherence index requires two steps. Firstly, it is needed to collect information about the Relatedness between the different technological components. And in a second step, the mean degree of knowledge relatedness within each region is supposed to provide a measure of the Coherence of its knowledge base.

[12] Consistently, Saviotti and Frenken (2008) shown that while *related* (export) *variety* helps in the short run, *unrelated* (export) *variety* only promotes growth in the long run. The authors hypothesise that this is due to the fact that Related Variety produces only *incremental innovations*; on the contrary, Unrelated Variety is harder to recombine, but if successful, it can lead to completely new industries (*radical innovation*) sustaining long-term growth.

3.2.1 The survivor measure of relatedness

The first step exploits the measure of Relatedness developed by Teece et al. (1994). This measure is based on the so-called *survivor principle*; i.e., the idea that economic competition leads to the disappearance of relatively inefficient combinations of businesses, and so that the observed combinations signal the existence of same complementarities between them.

Let the technological universe consist of $k = 1, \ldots, K$ patent applications. Let $P_{ik} = 1$ ($P_{lk} = 1$) if patent k is assigned to technology i (l), and 0 otherwise, with $i, l = 1, \ldots, n$. The number J_{il} of observed joint occurrences of technologies i and l is $\sum_k P_{ik} P_{lk}$. We can build the square symmetrical co-occurrences matrix of technological classes as

$$\underset{(n \times n)}{\hat{\mathbf{\Omega}}} = \begin{bmatrix} \hat{J}_{11} & \cdots & \hat{J}_{i1} & \cdots & \hat{J}_{n1} \\ \vdots & \ddots & & & \vdots \\ \hat{J}_{1j} & & \hat{J}_{ij} & & \hat{J}_{nj} \\ \vdots & & & \ddots & \vdots \\ \hat{J}_{1n} & \cdots & \hat{J}_{in} & \cdots & \hat{J}_{nn} \end{bmatrix}.$$

As explained by van Eck and Waltman (2009), the number of co-occurrences of two elements (patent classes, industrial sectors, etc.) can be seen as the result of two independent effects: a *similarity* and a *size effect*. Since we are interested in measuring the former, and not the latter, we need of a way to exclude this last from our index. That is to say, we need to benchmark value accounting for regional idiosyncratic effects. As exemplified by Bottazzi and Pirino (2010, p. 5), these effects are, in economic terms, both the fact that the observed joint presence of two patent classes within a region can be due to chance –since the bigger the field, the higher the probability of an observed co-occurrence–, or the effect of long-term *path dependencies* of the regional diversification evolution, so that the joint presence of the two technological domains cannot be interpreted as a true signal of their complementarity, that would make them more valuable if used together as input of a knowledge production function.

As usual also in other streams of literature, like Scientometrics and Network Analysis, we can do so by computing the expected value of each of these joint occurrences under a random distribution assumption, and compare the observed level with this last. In particular, the procedure chosen by Teece et al. (1994) is to assume that the number j of patents assigned to both technologies i and l is the realisation of a random variable J_{il} that follows a hypergeometric distribution,[13] which mean and variance are, respectively,

$$\mu_{il} = E[J_{il}] = \frac{O_i O_l}{K},$$

$$\sigma_{il}^2 = \mu_{il} \left(\frac{K - O_i}{K} \right) \left(\frac{K - O_l}{K - 1} \right),$$

where $O_i = \sum_l J_{il}$ and $O_l = \sum_i J_{il}$. If the actual number $\hat{J}_{il}$ of co-occurrences observed between two technologies i and l greatly exceeds the expected value μ_{il} of random technological co-occurrence, then the two technologies are highly *related* (and the opposite): there must be a strong, non-casual relationship between the two technology classes. Hence, the measure of relatedness for a pair of technological classes is defined as

$$t_{il} = \frac{\hat{J}_{il} - \mu_{il}}{\sigma_{il}},$$

[13]It describes the probability of j successes in O_l draws, without replacement, from a finite population of size K that contains exactly O_i objects with that feature, wherein each draw is either a success or a failure.

with $t_{il} \in (-\infty; +\infty)$.

Since large values of the t-statistic, t, are very unlikely under the *null hypothesis*, we can assume this as a signal of "deterministic" mechanisms that make the two domains to appear together more often than expected, and we can call this signal *similarity.*

As underlined by Nesta (2008) the interpretation of the *relatedness* measure is different if we apply it to the activities of a firm (or region), as done by Teece et al. (1994), or to its technology classes, as in Antonelli et al. (2010), Bottazzi and Pirino (2010), Nesta and Saviotti (2005), Quatraro (2010) and Rocchetta and Mina (2019). In the first case, the prominent reason for *related diversification* lies in the possibility for the firm (region) to exploit common competencies shared in a variety of business lines. Instead, *technological relatedness* says that the utilisation of a technology implies that of another one in order to perform a specific set of activities, not reducible to their independent use. For this reason, *technological relatedness* is considered a signal of the complementarity of the services rendered by the join combination of two different technologies: and this, as said above, is exactly what we would like to capture through this measure.

3.2.2 The measure of regional coherence

After having measured the *relatedness* between pairs of technologies (or sectors), Teece et al. (1994) suggest the weighted average relatedness WAR_i of technology (sector) i with respect to all other technologies (sectors) within the firm,

$$WAR_i = \frac{\sum_{l \neq i} t_{il} p_l}{\sum_{l \neq i} p_l},$$

as a measure of the expected relatedness of technology i with respect to any given technologies randomly chosen within the firm. WAR_i may be either positive or negative, the former (latter) indicating that technology i is closely (weakly) related to all other technologies within the firm. Lastly, following Nesta and Saviotti (2005), it is possible to define the Coherence of the firm's knowledge base as the weighted average of its WAR_i measures

$$\mathcal{C} = \sum_{i=1}^{I} WAR_i \, \frac{p_i}{\sum_i p_i}.$$

This is an estimate of the average relatedness of any technology (sector) randomly chosen within the firm with respect to any other technology. As for the WAR, a positive level of Coherence means that the firm's technologies (sectors) in which the firm has developed competencies are globally well related (and the opposite). The same measures can be applied, *mutatis mutandis*, to regions (see e.g., Quatraro 2010).

3.3 Complexity indices, regional diversity and technological rarity

The last group of measures is composed by the Economic Complexity Index, recently proposed by Hidalgo and Hausmann (2009), and the index of Fitness, suggested by Pietronero and his coauthors (Cristelli et al. 2013; Tacchella et al. 2012). As underlined by Hausmann and Hidalgo (2011), these measures, even though in a different way, take into account not only the *diversity* of the production of an economic system, as done by the previously explored measures, but also the *ubiquity* of these production among all the other economies considered, with the idea that a combination of the two dimensions will be a good indicator of the capabilities available in a given economic system. Essentially the same idea has been expressed by Antonelli et al. (2017) and Balland and Rigby (2017) about technology at the regional level. Not only does the number of different activities of the regional knowledge bases and the structural characteristics of this complex bundle matter, but also the relative scarcity of each element has a primary role. A bundle of knowledge items able to yield strong Jacobs externalities will include many rare activities, and since only regions with a large number of high-skilled individuals and specific technological competences will be able to develop sophisticated, and thus more rare, technologies, these regions will be the most competitive ones.

3.3.1 Technological Complexity Index

In order to compute the so-called Technological Complexity Index, we need to transpose the data in a binary bi-adjacency matrix whose layers are the regions, on one side, and the technological classes, on the other. The procedure followed by most of the literature (Antonelli et al. 2017; Balland and Rigby 2017; Hidalgo and Hausmann 2009) is to apply the so-called Revealed Technological Advantages approach (Archibugi and Pianta 1992b; Balassa 1961; Soete and Wyatt 1983) so that

$$M(r,i) \equiv \begin{cases} 1 & \text{if } RTA_{ri} \geq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $RTA_{ri} = \frac{P_{ri}}{P_{r\cdot}} / \frac{P_{\cdot i}}{P_{\cdot\cdot}}$, $P_{r\cdot} = \sum_{i=1}^{I} P_{ri}$, $P_{\cdot i} = \sum_{r=1}^{R} P_{ri}$, $P_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{r=1}^{R} P_{ri}$, and P_{ri} is the number of patent applications of region r in technology i.

Moreover, we need to define

$$\overrightarrow{K}_r \equiv \sum_i M_{ri},$$

$$\overrightarrow{K}_i \equiv \sum_r M_{ri},$$

where $\overrightarrow{K}_r$ is the vector of regional *diversity*, and $\overrightarrow{K}_i$ is the vector of technological (sectoral) *ubiquities*.[14]

The Technological Complexity Index, $\overrightarrow{TCI}$, as proposed by Hausmann et al. (2014) is then the *eigenvector* associated with the second largest *eigenvalue* of the following matrix:[15]

$$\tilde{M} = \text{diag}\left(\frac{1}{\overrightarrow{K}_i}\right) M \,\text{diag}\left(\frac{1}{\overrightarrow{K}_r}\right) M'$$

3.3.2 Regional Fitness

The method proposed in Cristelli et al. (2013) and Tacchella et al. (2012) use the same basic element of the previous one: $M(r,i)$. The idea of the procedure is to squeeze more information from the bi-adjacency matrix, exploiting the nested structure observed in the trade data, in a fashion similar to the procedure proposed by Zhou et al. (2007) and to the Google's PageRank algorithm. Indeed, a structure of this kind, suggests that those countries (regions) with a higher *diversity*, and those products (tech. classes) with a lower *ubiquity* provide less information than their opposite cases. Indeed, a product exported by most of the countries, and among those also by the ones with few exports, very likely will require a low level of sophistication. Therefore, the

[14] In other words, they are the degree distributions of the two layers of the network described by the bi-adjacency matrix $M(R \times I)$.

[15] This vector so obtained, $\overrightarrow{K}$, is assumed to be positively correlated with the levels of regional diversity, $\overrightarrow{K}_r$. Otherwise, it is needed to apply the following transformation: $\overrightarrow{K} = -\overrightarrow{K}$. Moreover, the values are standardised, so that $\overrightarrow{TCI} = \frac{\overrightarrow{K} - \langle\overrightarrow{K}\rangle}{sd(\overrightarrow{K})}$.

non-linearity in the algorithm proposed is such that the information that a product is produced by some scarcely diversified (and so scarcely competitive) countries is sufficient to assign a lower complexity level to that product. In other words «the only possibility for a product to have a high qualitative level (or complexity) is to be produced only by highly competitive countries» (Tacchella et al. 2012, p. 1).

The iterative method starts by settings the initial conditions as $\tilde{F}_r^0 = 1, \forall r$ and $\tilde{Q}_i^0 = 1, \forall i$. Then it is composed of two steps in each iteration ($n > 0$):

$$\begin{cases} \tilde{F}_r^{(n)} = \sum_i M_{ri} Q_i^{(n-1)}, \\ \tilde{Q}_i^{(n)} = \frac{1}{\sum_r M_{ri}\left(1/F_r^{(n-1)}\right)}; \end{cases} \quad \rightarrow \quad \begin{cases} F_r^{(n)} = \tilde{F}_r^{(n)} / \langle \tilde{F}_r^{(n)} \rangle_r, \\ Q_i^{(n)} = \tilde{Q}_i^{(n)} / \langle \tilde{Q}_i^{(n)} \rangle_i; \end{cases}$$

4 Data

The previous section has introduced some measures widely used in the empirical literature to capture the idea of economic and technological diversity, in its different aspects. Instead, in Sec. 5, I will look at these indices, highlighting for each of them some major drawbacks that affect them. The main differences will be discussed using data about the patenting activity of the European regions. These data come from the OECD REGPAT databases (version 2018/03). I assigned a patent application to a region on the bases of the inventors' addresses, I took all the NUTS2 regions of the EU28 with the exception of the "overseas territories" of Spain, France and Portugal, and I used any patent with priority year between 2000 and 2013.[16] Following the literature each year is the aggregation of all the applications happened in the previous 5 years.[17] I also decided to cut away the cases in which I counted less than 10 patents in 5 years in a given region. In the end, I have an unbalanced and hierarchically structured panel database of 256 NUTS2 regions, and 27 countries (EU28 with the exception of Croatia).[18] As a preliminary analysis, if the data are represented as bipartite networks, these yearly graphs remain substantially stable over time under the structural point of view (Tab. 5).

Instead, in Sec. 6 I will use the measure proposed in the two following sections in an empirical analysis that fits within the so-called *resilience* literature. This exercise will be used as a tool to test and explain in practice the issues raised in Sec. 5. Apart from the patent data just said, all the other data used in the last section are from the Eurostat Regio database. Moreover, while in Sec. 5 the measures will be computed at 3 (classes), 4 (subclasses), or 7 (main groups) digits of the International Patent Classification (IPC),[19] in the last section each measure considered is computed using 4 IPC class digits, while the decomposition of the Entropy and Evenness indices use 1 digit as macro-class. The data refer to 247 NUTS 2 regions and 26 countries of the European Union.[20] Summary statistics of the main variable used are reported in Tab. 6 (see Tab. 4 for the list of variables considered and the symbols used to represent each of them henceforth).

[16]In line with the literature, I chose to look at the priority year to have data that are as much informative as possible about the moment of the invention. Moreover, since it is common knowledge that it takes 3–5 years for a patent to be approved, I truncated the time series in 2013 for precautionary reasons.

[17]As explained by Nesta and Saviotti (2006, p. 630, n. 3) «[t]his compensates for the fact that learning processes are time consuming, due to certain rigidities in firms' technological competencies».

[18]All the data uses the NUTS 2013 classification. I reclassified the NUTS 2 codes of the area of London that in the database provided by the OECD were still in the 2010 version, differently from the other codes that had been reclassified according to the 2013 definition.

[19]Each time the choice will be properly signalled.

[20]The whole Croatia and Slovenia have been excluded because of data unavailability. Moreover, apart from the "overseas territories" of Spain, France and Portugal, also the following NUTS 2 regions has been excluded because the data about some of the variables were unavailable: DED4 (Chemnitz), DED5 (Leipzig), UKI3 (Inner London - West), UKI4 (Inner London - East), UKI5 (Outer London - East and North East), UKI6 (Outer London

Table 4: Explanation of the symbols used in the regression tables.

Symbol	Variable
E	Employment level
density	Population density
HC	Human Capital index, defined as the share of people who have successfully completed a tertiary level education and are employed in a S&T (science and technology) occupation (HRSTC)
RTA	N. of technological domains in which a region has Revealed Technological Advantages grater than one.
ETP	Total (undecomposed) Shannon Entropy
RV	Related Variety
UV	Unrelated Variety
CT	Coherence based on t-statistic
EVS	Total (undecomposed) Shannon Evenness
RE	Related Evenness
UE	Unrelated Evenness
CP	Coherence based on p-value
CX	Weighted Technological Complexity Index
CX>0	Dummy variable that is true for regions with a Weighted Complexity Index above its mean, and false otherwise (see Fig. 7a)
FX	Weighted Fitness
RWD	Rarity-weighted diversity

Table 5: Network statistics. The values have been computed through the `bipartite` and `ineq` R Packages (Dormann 2011; Dormann et al. 2009, 2008; R Core Team 2018). OECD RegPat, NUTS2, 3 digits IPCs, 2000–2013. R = regions; TC = technological classes.

	2000	2001	2002	2003	2004	2005	2006
n. regions	239	243	247	251	257	258	260
n. tech. classes	121	121	121	121	121	121	122
n. links	17830	18136	18452	18729	18993	19109	19218
sum. weights	488204	526235	556005	571773	578535	572626	571208
linkage density	58.69	58.69	58.82	59.16	59.55	59.82	60.66
weighted connectance	0.16	0.16	0.16	0.16	0.16	0.16	0.16
cluster coef.	0.64	0.64	0.65	0.63	0.63	0.63	0.63
cluster coef. (R)	0.89	0.90	0.90	0.90	0.90	0.90	0.89
cluster coef. (TC)	0.82	0.82	0.83	0.83	0.82	0.83	0.83
weighted nestedness	0.80	0.80	0.81	0.81	0.81	0.81	0.82
weighted NODF	68.49	69.11	69.52	69.65	69.93	69.90	69.91
Est. power law exp. (R)	0.44	0.39	0.39	0.37	0.39	0.41	0.42
Est. power law exp. (TC)	0.82	0.75	0.72	0.69	0.76	0.72	0.23
Gini deg. dist. (R)	0.26	0.26	0.26	0.26	0.27	0.27	0.27
Gini deg. dist. (TC)	0.18	0.17	0.17	0.17	0.18	0.18	0.19
Gini strength dist. (R)	0.68	0.68	0.69	0.69	0.69	0.69	0.69
Gini strength dist. (TC)	0.66	0.66	0.67	0.67	0.67	0.67	0.67
	2007	2008	2009	2010	2011	2012	2013
n. regions	261	262	262	260	260	261	259
n. tech. classes	122	122	122	122	122	122	122
n. links	19460	19522	19612	19769	19990	20135	20388
sum. weights	571714	572202	571770	577031	580629	578368	577985
linkage density	61.35	61.81	62.37	63.03	63.47	63.86	64.34
weighted connectance	0.16	0.16	0.16	0.17	0.17	0.17	0.17
cluster coef.	0.64	0.64	0.63	0.64	0.65	0.66	0.68
cluster coef. (R)	0.89	0.89	0.88	0.89	0.89	0.88	0.89
cluster coef. (TC)	0.83	0.84	0.84	0.85	0.85	0.86	0.86
weighted nestedness	0.82	0.82	0.82	0.81	0.81	0.81	0.81
weighted NODF	69.96	69.98	69.83	70.02	70.10	70.12	69.69
Est. power law exp. (R)	0.43	0.46	0.46	0.49	0.48	0.42	0.48
Est. power law exp. (TC)	0.23	0.27	0.27	0.30	0.33	0.36	0.52
Gini deg. dist. (R)	0.26	0.26	0.25	0.24	0.23	0.23	0.22
Gini deg. dist. (TC)	0.19	0.19	0.19	0.19	0.19	0.19	0.19
Gini strength dist. (R)	0.68	0.68	0.68	0.67	0.67	0.67	0.66
Gini strength dist. (TC)	0.67	0.67	0.66	0.66	0.66	0.66	0.65

Table 6: Descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
density	247	316.006	643.768	3	74.2	293.6	6,366
E	247	828.164	660.621	14.100	424.800	1,060.200	5,177.100
HC	247	15.145	4.421	6.600	11.750	17.800	31.200
RTA	247	124.810	64.171	11	73.5	173.5	270
ETP	247	6.293	1.127	3.323	5.780	7.149	7.890
RV	247	3.743	0.977	1.084	3.153	4.485	5.095
UV	247	2.550	0.237	1.335	2.458	2.701	2.880
CT	247	7.044	0.401	6.147	6.713	7.327	8.214
EVS	247	0.872	0.064	0.707	0.836	0.917	1.000
RE	247	0.507	0.054	0.285	0.486	0.546	0.639
UE	247	0.873	0.055	0.654	0.845	0.911	0.994
CP	247	0.367	0.104	0.060	0.305	0.433	0.707
CX	247	0.005	1.011	−3.822	−0.498	0.714	1.614
FX	247	1.028	0.837	0.016	0.314	1.536	4.931
RWD	247	1.814	1.355	0.066	0.561	2.898	5.341

5 Major issues affecting diversity measures

As said, while in Sec. 3 I introduced some widely used measures of economic and technological diversity, in this section I will highlight for each of these indices some major drawbacks that affect them. In each case, a possible solution will be introduced and, using data about the patenting activity of the European regions, the advantages of each solution suggested will be tested.

5.1 Revealed Technological Advantages

As remembered in the first section of the paper, we can define the technological *variety* of a region as $\log_2 n$. However, this index does not account for some randomness in the patents development at the regional level. A way to purge the data from this effect is to compare the number of patents observed in each region-technological domain with a null model that maximise the randomness of the distribution under given constraints. This is what the Revealed Technology Advantage (RTA) index, already introduced in the previous section, does. In other words, it provides an indication of the relative specialisation of a given geographical area in selected technological domains, taking into account of how diversified a region is and how ubiquitous a technological domain is. The index is equal to zero when a given region holds no patent in a specific domain; is equal to 1 when the region's share in the sector equals its share in all fields (i.e., the region is not specialised in the domain); and above 1 when a positive specialisation is observed. Therefore, in the empirical application proposed in the following section we will use this index to capture the idea of *variety*.

- South), UKI7 (Outer London - West and North West), BG34 (Yugoiztochen Planning Region), EL51 (Eastern Macedonia and Thrace), EL53 (Western Macedonia), EL62 (Ionian Islands), EL41 (North Aegean), EL42 (South Aegean), RO22 (Sud-Est), RO31 (Sud - Muntenia).

5.2 Entropy and Evenness

Using the terminology of Stirling (2007), Entropy fails to sharply distinguish between the *variety* and *balance* of a regional knowledge structure. Indeed, the index grows, not only if the items are distributed more uniformly among the possible technological domains developed by the region (*balance*), but also with the number of technological domains developed by a region (*variety*). In other words, this measure is affected by a *size effect* that should be accounted if we want to distinguish the two phenomena in an empirical investigation.

Not being able to distinguish between these two components, the index induces to fallacious interpretations of the results of the empirical analyses. Besides, a major drawback that follows is the dependence of the results from the level of aggregation of the technological domains chosen. Indeed, almost by definition, the *variety* of technological classes developed within a region grows, once we increase the number of digits at which we compute the index. This means that the Entropy will be higher, the lower is the aggregation level and that we cannot expect a linear correlation between the index computed at different levels of aggregation (Fig. 1a).

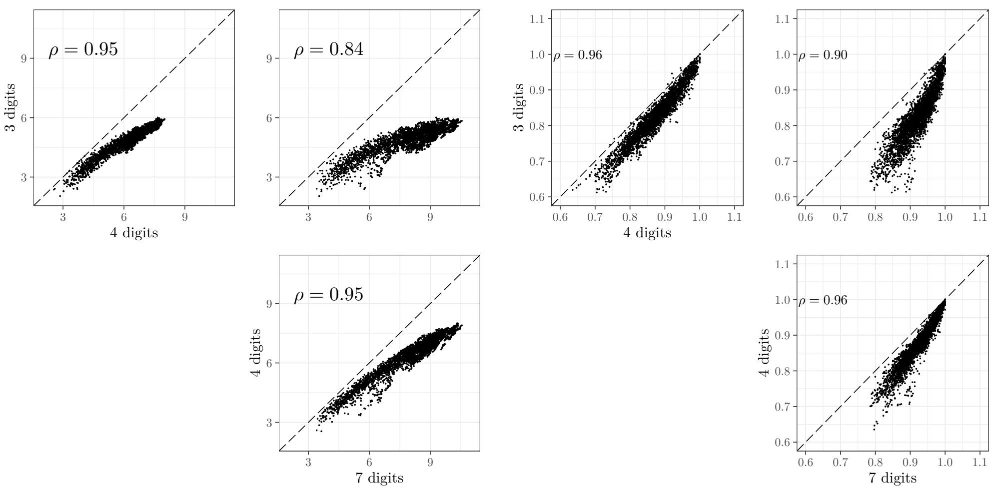
A possible solution, proposed by Stirling following Pielou (1969), is the use of the Shannon Evenness index, defined as

$$E = H/\log_2 n.$$

Since the theoretical maximum of the Shannon Entropy is $\log_2 n$, this index normalises it between 0 and 1, letting comparisons between sets of different size (n) possible and meaningful. As shown in Fig. 1b, this measure has also the secondary advantage, compared to the Entropy, of being less dependent on the technological domain aggregation level chosen in the analysis, helping the comparison of different empirical analyses computed at different levels of the patent-classes tree.

Figure 1: Correlation between the Entropy and the Evenness computed at different levels of aggregation of the patent classes (IPC 3, 4, and 7 digits, respectively). The plot on the bottom-right shows the Spearman Rank correlations. Data: OECD RegPat, NUTS2, 2000–2013.

(a) Entropy (b) Evenness



Evenness decomposition As well as Entropy, also its Related-Unrelated Variety decomposition confounds the two effects named by Stirling *variety* and *balance*, so that these two indices grow with the number of technological domains "owned" by the region, and not only with the evenness of the distribution of the patents developed in the area over the macro-classes, or the average evenness of their distribution over the micro-classes within each macro-class. Some papers tried to overcome this issue normalising the Related and Unrelated Variety by $\log_2 n$, as done just above for the overall Entropy, or just by n (Lee 2017; Lengyel and Szakálná Kanó 2013). These solutions have the advantage to preserve the perfect decomposability of the index. However, to normalise the two indices we need to account for their theoretical maximum. For Unrelated Variety, since this is just the Shannon Entropy of the relative size of each macro-class considered, its maximum is $\log_2 G$. Instead, Related Variety grows with the number of micro-classes, but decreases with the number of macro-class in which these last are grouped. In particular, it will reach its maximum when we have only one macro-group and each micro-group is equally likely. In this last case, the Related Variety and the total Shannon Entropy will be the same and equal to $\log_2 n$. In what follows, I will call these two components of the Shannon Evenness, Related and Unrelated Evenness, in analogy with the Frenken et al. (2007) Related-Unrelated Variety.

5.3 Null models underlying Coherence

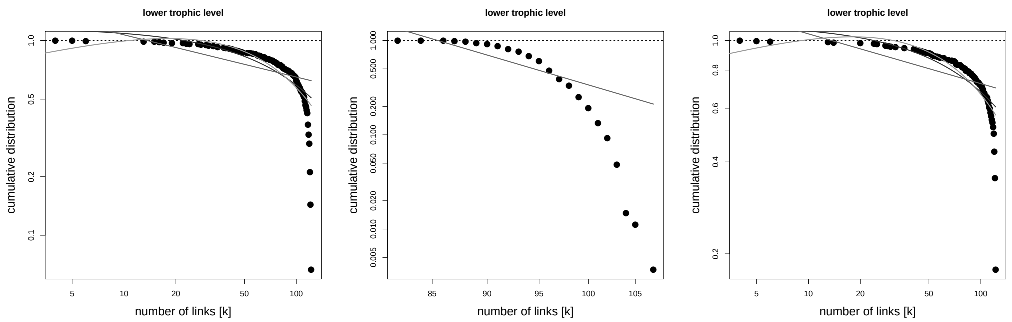
First of all, following Bottazzi and Pirino (2010, pp. 5–6), we can identify two drawbacks about the *relatedness* measure proposed by Teece et al. (1994). Both, in the end, having to do with the distribution of the patent applications among the regions, and the effect that an uneven distribution has on the projection of the bipartite network on the *technology* layers.[21] Indeed, the more skewed the distribution of the number of technological classes in each region is, the more likely is to observe very high numbers of the t-statistic –since it is based on a normal approximation. Therefore, this statistic is not, in general, a valid tool to detect possible deterministic effects, since its reliability depends on the characteristics of the underlying distribution. Moreover, what is chosen as a *null model* in the seminal contribution of 1994 is what Bottazzi and Pirino have called *H2 model*, i.e., it constrains the column sums (O_l) and the total number of links (K), but not the row sums (O_i), that are instead random variables. But this choice is quite arbitrary, and not theory based. Because of that, following a long tradition in Ecology, it should be better to put a constraint on the distribution of both layers (Gotelli 2001; Gotelli and Graves 1996). As shown in Fig. 2, by preserving only the strength distribution (columns sum) of the technological domains layer of the regions-technologies co-occurrence matrix, both the degree and the strength distributions of the regions layer are completely different from the ones of the observed data. This suggests that the H2 *null hypothesis*, that underlays the t-statistic relatedness method introduced in the previous section, is not a proper way to produce simulated data against which to compare the observed ones. Indeed, a key constraint (i.e., the distribution of the number of patents developed by each region) seems not respected and reproduced in the simulated data. Consequently, it is likely that the Coherence measures built on the type of relatedness measure already introduced will be biased, underestimating the average relatedness of some regions and overestimation the one of some others.

Lastly, another limitation of the z-scores implicitly used by the model presented so far, is that they are still affected by what I have called before *size effect*, since the numerator of the statistic grows faster than its denominator.[22]

[21] This is a topic well known also in the Network Analysis literature, since the value of many statistics computed on the projections of a bipartite network depends on the loss of information and distortions that happens through the projection procedure itself (Padrón et al. 2011).

[22] This, for example, is clearly highlighted also by Alstott et al. (2016, p. 454). In the paper, the authors deflate the z-score to correct for that. The use of p-values solves the problem directly, without the need of the deflation.

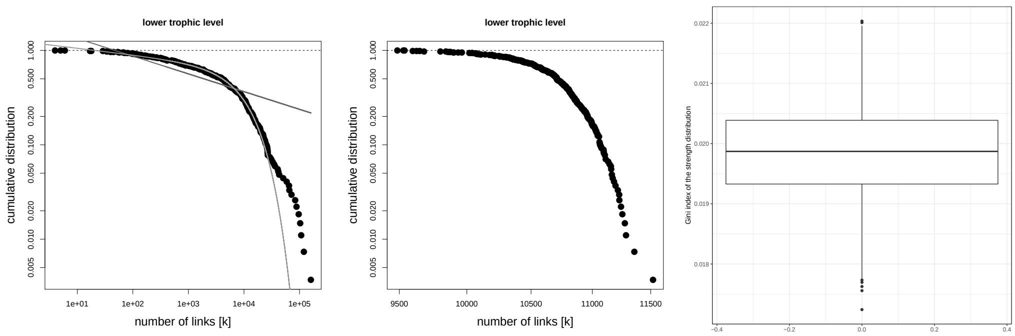
Figure 2: Degree and strength distributions. Data: OECD RegPat, NUTS2, IPC3, 2000–2013. The simulations have been done using the `vegan` R Package (Oksanen et al. 2018; R Core Team 2018). The plots and estimates has been obtained using the `bipartite` R package (Dormann 2011; Dormann et al. 2009, 2008; R Core Team 2018).



(a) Degree distribution of the regions in the observed data.

(b) Degree distribution of the regions of one of the simulations that preserves only the strength of the technologies layers of the bi-adjacency matrix.

(c) Degree distribution of the regions of one of the simulations that preserves the strength of both layers of the bi-adjacency matrix.



(d) Strength distribution of the regions in the observed data.

(e) Strength distribution of the regions of one of the simulations that preserves only the strength of the technologies layers of the bi-adjacency matrix.

(f) Distribution of the Gini index of the strength distribution of the regions preserving only the strength of the technologies layers of the bi-adjacency matrix considering 1000 simulations. Gini index of the observed data = 0.7.

The p-value Coherence A possible solution –similar to another proposed by Nesta (2008, Appendix A)– to overcome the three issues just raised about the *relatedness* matrix, is to substitute the inference based on the value of the statistic J_{il} –or any other possible similar statistic we can think about– with the inference based on its p-score (Bottazzi and Pirino 2010).

The procedure consists of three steps:

i Randomise the empirical bi-adjacency matrix hundreds of times, constraining the degree (strength) distribution of both its layers and the total number of links (weights);[23]

ii Compute the J_{il} statistic on each of these simulations;

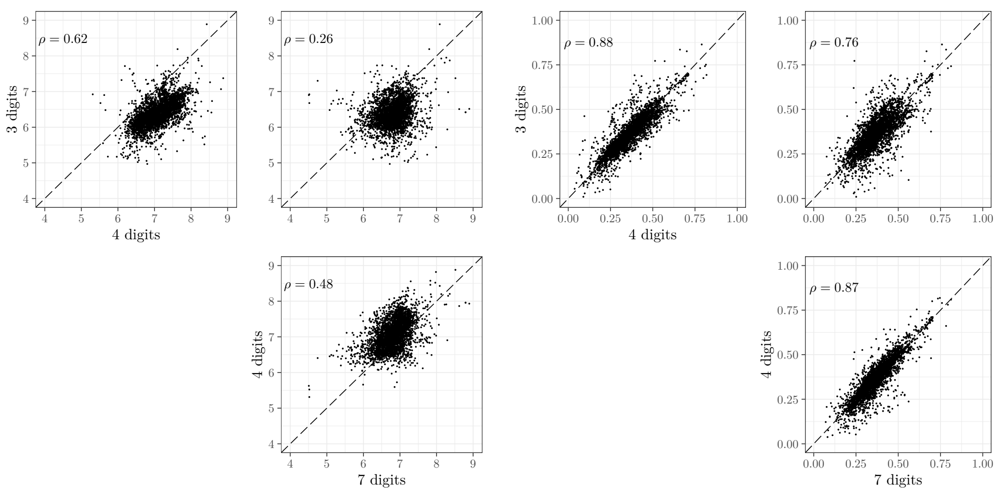
iii Compute $p_{il}(J, H4) = Pr[\hat{J}_{il} \geq J_{il}|H4]$.

The p-values so obtained can then be used, in the same way of the t-statistics as shown above, to obtain a Coherence measure, that is not dependent on the size and the form of the degree (strength) distributions of the empirical data, and that is also less affected by the statistic chosen as a bipartite network projection device.

Interestingly enough, as shown in Fig. 3, while the Coherence index based on the t-statistics depends on the aggregation level of IPC at which we compute it –because the deeper we go in the IPC classes tree, the more the number of classes, by definition–, this is not for the same measure computed on the basis of the p-values. This is what has been previously accounted as *size effect*. Therefore, another good reason to favour this last measure instead of the other variant is that the results will not be affected by the IPC-level aggregation choice, helping the comparability of the results between different empirical exercises.

Figure 3: Correlation of the Coherence index at different levels of aggregation of the IPC classes: 3 dig. VS 4 dig. (top left); 3 dig. VS 4 dig. (top right); 4 dig. VS 7 dig. (bottom left); Spearman Rank correlations (bottom right). Data: OECD RegPat, NUTS2, 2000–2013.

(a) t-statistics based Coherence. (b) p-values based Coherence.



[23]I used, for this procedure, the `quasiswap` algorithm provided by the `vegan` R Package (Oksanen et al. 2018; R Core Team 2018), but in case of a weighted bipartite network, a repeated reshuffling of one of the two columns of the edge list of the graph converges to the same result.

5.4 Complexity and Fitness indices for weighted matrices

Since the Economic Complexity Index (ECI) has been introduced in literature with respect to data about the products-countries trade data, its use (and usefulness) as a measure once applied to technological domains-regions data should be carefully evaluated. An important issue immediately emerges from the observation of the structural characteristics of the bipartite network that represent the patent data. In particular, looking at Fig. 4, we can see a clear triangular-like shape in some of the regions-tech. classes occurrence matrices. But this type of nested structure almost disappears in Fig. 4b, i.e. exactly for the $M(r,i)$ matrix used by Hidalgo and Hausmann (2009) in their original paper. In other words, for patent data, it seems that the use of the binarization algorithm *à la* Balassa risk to break down an important structural characteristic of the network under investigation. Therefore, I chose to use an alternative statistic:

$$MW(r,i) = (\arctan RTA_{ri}) \,/\, \frac{\pi}{2}. \tag{2}$$

The use of the RTA, being equivalent to the Weighted Configuration Model introduced by Serrano and Boguñá (2005), helps to account for spurious observations due to the *size-effect* of both the layers of a weighted bipartite network –in a similar fashion of what discussed above about the *relatedness* measure. Thus, it must be preferred to the direct use of observed values (Fig. 4a), since the empirical matrix is compared to a random *null* under the assumption of independence between the strength of any pair of nodes, constraining the strength distribution of both the layers of the graph.[24]

Moreover, since in many known empirical networks the nodes weights are correlated with the respective node degree–and this is the case also for the patent data here considered (Fig. 5)–, as highlighted by Serrano and Boguñá (2005, p. 102), the removal of the binarization of the weights can introduce a significant loss of information about the true structure of the graph. So, both in general and in this specific case, the use of an RTA-based statistic –like the one of Fig. 4d– should be preferred to a simpler binarization method –like the one depicted in Fig. 4c.

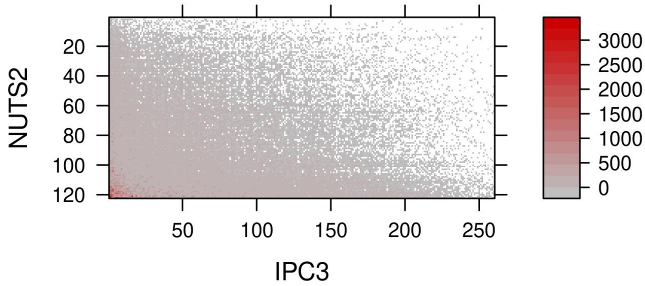
But, since the RTA values are spread over a very large (non-negative) range, I choose to take the arctangent of the values so obtained. In this way, the values are squeezed on a smaller range of possible values. Moreover, the transformations reduce the distance between very high value while preserving a higher distance (in relative terms) for small numbers. Since the RTA values are nothing else if not an estimate of how bigger is the empirical strength of a link, compared with the *null model*, it is more useful to preserve small numbers than very big ones that add not so much to the basic observation that "the observed value is truly unpredictable under the chosen null hypothesis".[25] Lastly, the division by $\pi/2$ is useful only to have values in $0, 1$. In this way I am able to compute the algorithm explained above on a weighted bipartite graph.

As well as for the Complexity index, I computed also the Regional Fitness measure on a weighted co-occurrence matrix, as introduced just above. In this case, the observations provided above about the nested structure of the occurrence matrix are even more important than in the case of the Complexity index, since the rationale itself of the Fitness as a useful measure is based on the triangular-like structure of the trade data matrix.

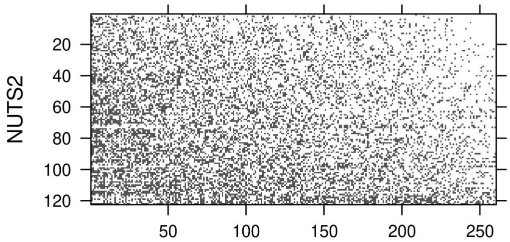
[24]See Bottazzi and Pirino 2010 for a classification of the four possible *null models* (or *null hypotheses*) for bi-adjacency matrices (i.e., bipartite networks). Essentially, the so-called H1 algorithm reshuffles the values of the binary matrix, preserving only the total number of links of the network. At the opposite end, under the H4 procedure, not only the sum of the links of the original and the simulated matrices is the same, but also the degree distribution of the two layers of the reshuffled network (i.e., matrix row and column sums, respectively) are kept equal to the original one. Halfway, the H2 and H3 *null models* preserve the column and row sums of the empirical data, respectively, besides network *density*. Therefore, the Coherence proposed by Teece et al. (1994) can be classified as based on an H2 *null model*. Also Hausmann and Hidalgo (2011) provide essentially the same classification.

[25]I thank Martina Iori for making me realise this useful property of the arctangent.

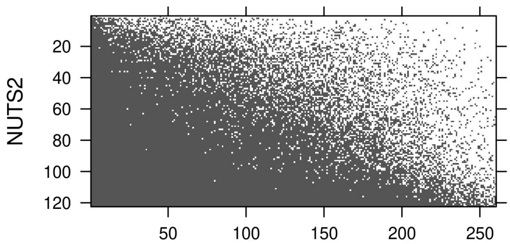
Figure 4: Regions in rows and tech classes in columns. The columns (rows) are ordered from left to right (from bottom to top) according to the node strength, The colours (when present) represent the weight of the link. Data: OECD RegPat, NUTS2, 3 digits IPCs, 2010.



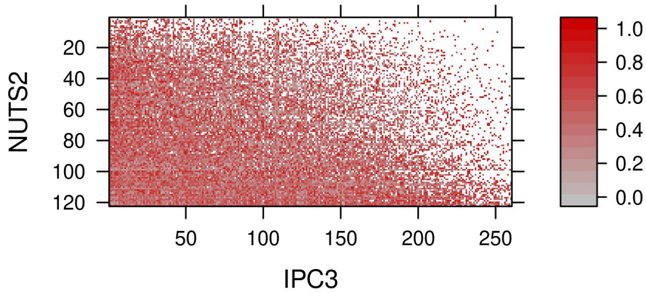
(a) Weighted matrix that represents the number of patents that each region have developed in each tech class.



(b) Binary matrix above named as $M(r, i)$.



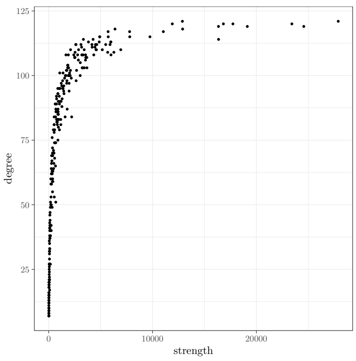
(c) Binary matrix in which a region-tech. class is 1 each time at least 10 patents are observed in a 5-years time span considering a specific pair.



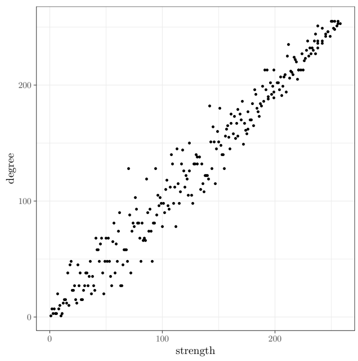
(d) Weighted matrix above named as $MW(r, i)$.

Figure 5: Weights-degrees correlations. Data: OECD RegPat, NUTS2, 3 digits IPCs, 2005.

(a) Correlation



(b) Rank Correlation



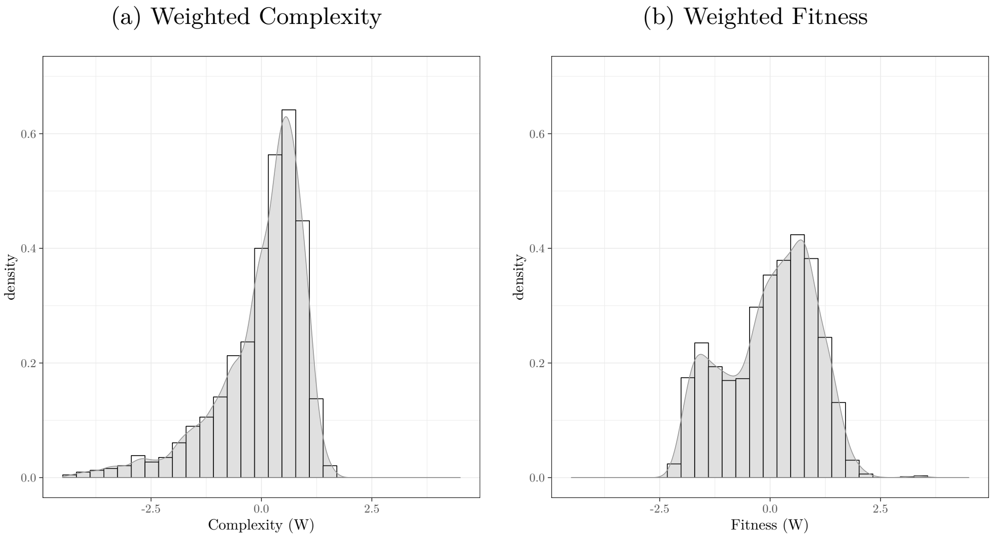
Regional Fitness convergence issue The Fitness index share, in general, most of the problems highlighted above introducing the Economic Complexity index *à la* Hidalgo-Hausmann. But the main problem of this measure is the lack of convergence in many cases. In particular, on the database here used it does not converge for the whole time span 2006–2011 (Tab. 7).

Distribution of Complexity and Fitness On the other hand, as shown in Fig. 6, the distribution of the Complexity index, compared to the Fitness one, is extremely skewed, and with fatter tails.

Table 7: Number of interactions that the (Weighted) Fitness algorithm took to converge. If the algorithm has never converged to a stable value "NO" is reported.

Year	Iterations or Convergence	Year	Iterations or Convergence
2000	26	2007	NO
2001	26	2008	NO
2002	26	2009	NO
2003	27	2010	NO
2004	27	2011	NO
2005	30	2012	43
2006	NO	2013	23

Figure 6: Distribution of the Weighted Complexity and Weighted Fitness algorithms (rescaled data). Data: OECD RegPat, NUTS2, 3 digits IPCs, 2000–2005 and 2012–2013. The period 2006–2011 has been excluded since the Fitness algorithm has not converged.



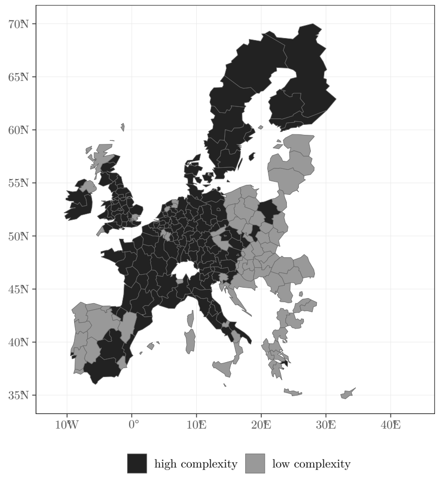
	Complexity (W)	Fitness (W)
Skewness	-1.52	-0.29
Kurtosis	5.57	2.26

5.4.1 The Regional Technological Complexity Index as a classifier

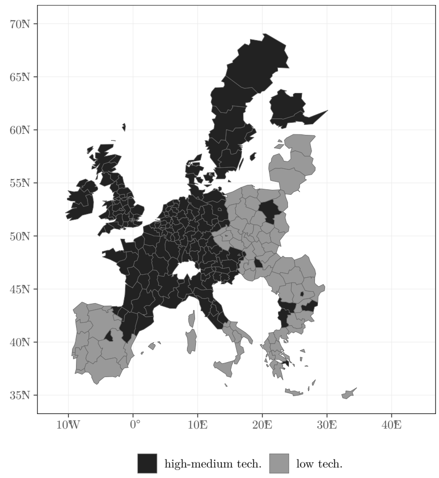
Following Mealy et al. (2017), we can also think at the index just presented as analogous to the clustering algorithm proposed by Shi and Malik (2000), which partitions a similarity graph into two balanced components that are internally similar and externally dissimilar. The regions with a positive normalised Complexity are more similar to each other than with these regions with a negative value of the index, and the opposite. Indeed, if we look at the map in Fig. 7 we can see that the method partitions quite clearly the EU regions in two groups. If we compare the classification so obtained with the one provided by Wintjes and Hollanders (2010) –commuting the original classification in a binary one as reported in Fig. 7c– we have that the in about 83% of the cases the classifications coincide over the period 2000-2013 considered. This seems in line with the interpretation proposed by Mealy et al. (2017), and calls for further examinations of the index, that seems able to capture the type of technological knowledge developed by the region, more than its technological *diversity*, as claimed by the previous empirical literature.

Figure 7: Classifications of the EU regions based on their technological capabilities. Data: EuroGeographics for the administrative boundaries.

(a) ECI greater (lower) than zero. Data: OECD RegPat, NUTS2, 3 digits IPCs, 2005.



(b) Binary classification based on the one provided by Wintjes and Hollanders (2010).



(c) Different classifications.

Wintjes and Hollanders (2010)	Cortinovis and van Oort (2015)	Used here
Metropolitan knowledge-intensive services regions Public knowledge centers High-tech regions	High tech. regime	High-medium tech. regime
Knowledge-absorbing regions Skilled technology regions	Medium tech. regime	
Traditional Southern regions Skilled industrial Eastern EU regions	Low tech. regime	Low tech. regime

5.5 Rarity-weighted regional diversity

After having analysed and described the Technological Complexity Index *à la* Hidalgo-Hausmann, Antonelli et al. (2017) introduced the idea that it is possible to use an interaction term of the first two iterations of the Method of Moments (Hidalgo and Hausmann 2009) as an index of rarity-weighted variety of the technological knowledge base of a given region. In particular, the authors proposed to use the ratio of the regional *diversity* over the weighted average *ubiquity* of the technological classes "owned" by a region:

$$\left(\frac{\mathrm{RD}}{\mathrm{WATU}}\right)_r = \left(\sum_{i=1}^{I} M_{ri}\right) \Big/ \left(\frac{1}{\sum_{i=1}^{I} M_{ri}} M_{ri} \sum_{r=1}^{R} M_{ri}\right).$$

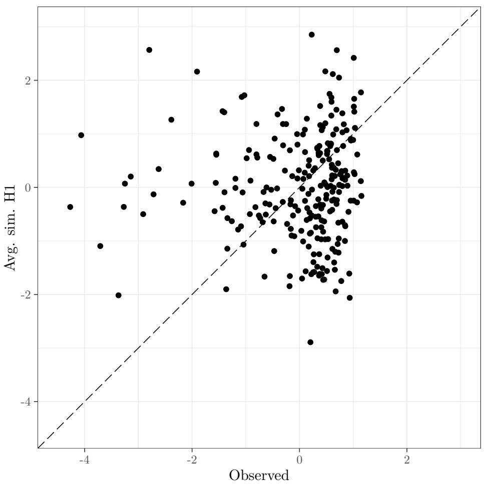
Since the denominator measures the average ubiquity of the technological domains "owned" by a region, a higher value of this fraction means that the region has a more diversified patent portfolio, and that the technological domains possess by the region are also rare ones.

This last measure seems quite interesting. Firstly, its a more direct way to capture the question of the *rarity* before mentioned, compared to the Economic Complexity and Fitness indices above introduced. Moreover, as shown in Fig. 8, the two measures seem to be essentially driven by the degree distributions of the layers of the bi-adjacency matrix. Indeed, if we take the average values of the two measures for 100 simulations in which the values are reshuffled constraining the two degree distributions (H4 null model), there is a strong correlation between the measure computed on the empirical data and the average of the measure computed on the simulated matrices. Conversely, by randomising the matrix imposing no other constraint than on the total weights sum (H1), the correlation completely disappears for both the measures introduced before.[26]

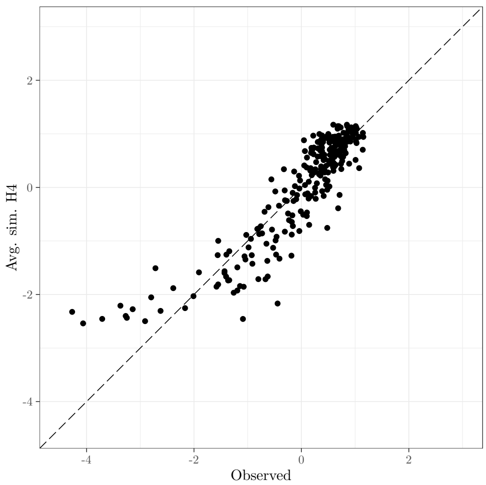
[26] I used the `quasiswap` (H4 *null model*) and `r00` (H1 *null model*) algorithms provided by the `vegan` R Package to get these results (Oksanen et al. 2018; R Core Team 2018).

Figure 8: Regional Weighted Complexity and Fitness indices. Observed data vs. null models. Data: OECD RegPat, NUTS2, 3 digits IPCs, 2005.

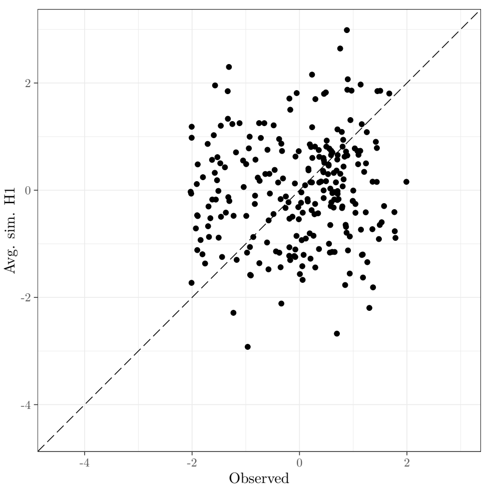
(a) Complexity.
Observed data vs. Average H1 simulations.



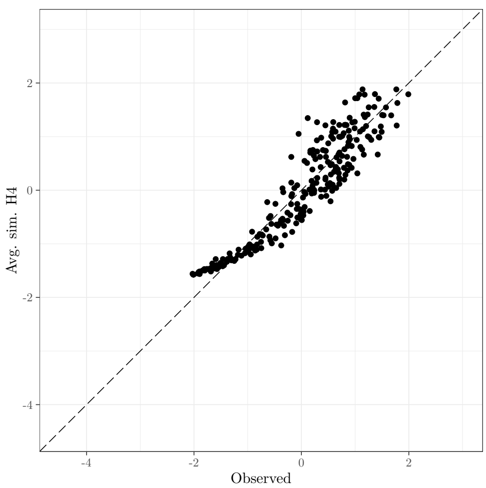
(b) Complexity.
Observed data vs. Average H4 simulations.



(c) Fitness.
Observed data vs. Average H1 simulations.



(d) Fitness.
Observed data vs. Average H4 simulations.



6 An empirical application of the diversity measures

This last section connects the ideas and indices introduced in the previous sections. Sec. 2 framed the regional economic development and regional (technological) diversity nexus within a classification that identifies three fundamental components of *diversity*, that are interrelated but distinct dimensions of this faceted concept. Moreover, a fourth aspect –the *rarity* of the elements of the regional *knowledge capital*– is identified as a fundamental orthogonal dimension that must be accounted together with the *diversity* of the bundle of productive resources of a region to understand and explain the evolutionary possibilities of an economic system. Sec. 3 introduced some measures broadly used in the empirical literature to capture and operationalise the concepts exposed in the previous section. Lastly, Sec. 5 discussed some prominent limitations and issues connected with each of the measures introduced, providing some solution to each of the drawbacks highlighted. In this last section, the measures proposed in the previous two parts will be used in an empirical exercise that looks at that so-called *resilience* literature. Even though some tentative interpretations of the results will be raised, the main aim of this analysis is to test and explore the issues raised in Sec. 5. Therefore, it has to be viewed as a tool that integrates and corroborates the analysis carried on in the previous section.

6.1 EU regional resilience capacity differentials during the Great Recession

As said, this last section fits within a body of literature recently boosted by the observed differences in the capacity of regions to recover from the recessionary period the affected the European economies from 2008 onward. Indeed, the recent financial and economic crisis of 2008–2010 hit Europe with particular strength, so much to deserve the name of Great Recession. Not only it has been the stronger –for magnitude and duration– since the 1930s, but the European Union was affected even more than other advanced economies, like the US (Aujean et al. 2015). Moreover, it has broken off a very long period of sustained economic and employment growth of the European area.

The consequences of the shock have followed different paths in each country: some showed a stronger effect in terms of GDP compared to the one in employment rate –for example in Germany and Italy–; while in others –like Spain– the crisis has had very strong short-run effects also on employment (Aujean et al. 2015, p. 45). Furthermore, there is a growing gap within the EU between these countries that experienced a *double dip* recession in 2012 –in particular Italy, Spain, Portugal, Greece, Slovenia and Finland– and the others.[27]

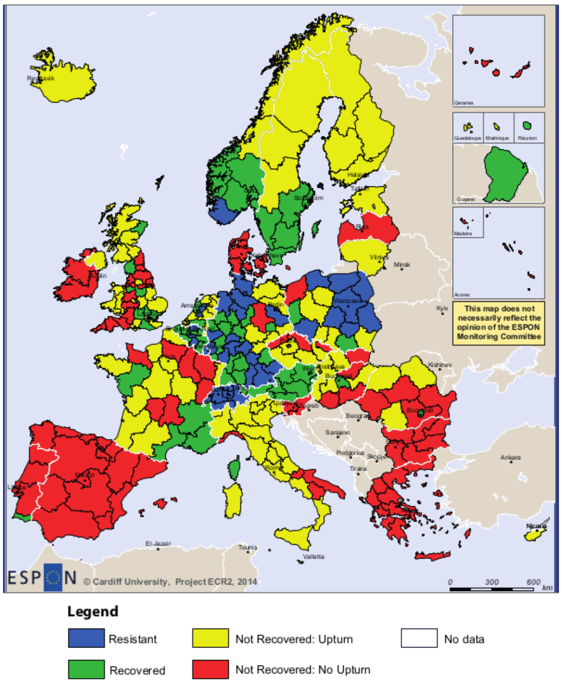
Also looking at the sub-national level, it is possible to see differences within countries and among regions of different countries. As shown in some 2014 publication of the European Program ESPON, the shock has had an asymmetric impact in territorial terms within the EU, with areas that have not been hit in any way by the crisis and others which have shown at least a relevant decrease of their GDP level or also a decline in employment terms.

Moreover, as highlighted by the 2014 ESPON report, there have been important differences also in the way in which the *recovery* has happened in this last group of regions (Fig. 9). Some experienced a swift return to pre-crisis levels of employment and output, while others enter in a more than five years long period of sustained stagnation. With this respect, the Great Recession has stopped and abruptly reversed a long term trend of convergence that the European regions showed both in GDP and employment rates terms. In 2014, about one-third of the European regions still have to experience the end of employment loss and economic decline, and another

[27]However, this last phenomenon is beyond the scope of this section, that looks only at the first phase that followed the shock, that the literature call *resistance*.

third has had not recovered the pre-crisis employment levels, even though it was no more subject to the downturn. Conversely, about 25% of the NUTS 2 areas, even though hit by the crisis, had recovered to their pre-crisis peak before 2014. Furthermore, there is even a tenth of regions which has not experienced any fall in employment or output whatsoever and even continued to grow even during the downturn period.

Figure 9: Distribution of Regional employment resilience (peak year to 2011). Source: ESPON 2014b.



Spurred by the effects of this severe economic downturn, and in particular by these geographical differences in the capacity to react and overcome the crisis, economists and geographers have tried to propose new ideas and measures with the purpose of deeply understanding the key structural and institutional factors that have helped some regions and countries more than others in the recovery phase from the Great Recession. Focusing in particular on the Evolutionary Economic Geography literature, the concept of *resilience*, and more specifically *adaptive resilience*, has emerged as the main theoretical and analytical framework to look at these differences (Boschma 2015; Martin 2012; Martin and Sunley 2015; Reggiani et al. 2002; Simmie and Martin 2010).

The idea of *adaptive resilience*, borrowed from the Complex Adaptive Systems literature, differs from other types of *resilience capacity* that a region can show (Martin 2012), since it looks at their *reinvention* and not, or at least not only, at their *resurrection*. Indeed, it can be thought as «the ability of a system to undergo anticipatory or reactive reorganisation of form and/or function so as to minimise the impact of a destabilizing shock» (Martin 2012, p. 5). Therefore, *adaptive resilience* refers not only to the capacity of a localised economic system to absorb the effects of external and unpredictable shocks in the first recessionary phase, but also to the ability of its industrial and technological structure, and the underlying knowledge base, to react against it through both *adaptability* and *creativity*. In conclusion, the focus is on the *adaptability* (more than on the *adaptation*) of a complex system to the new external environment, the characteristics of which were not fully predictable in advance (Boschma 2015; Grabher 1993; Pike et al. 2010).

As highlighted by Martin (2012), we can identify at least four, distinct though interrelated, dimensions of regional *resilience*: *resistance*, *recovery*, *re-orientation*, and *renewal.* The analysis proposed in this section focuses on the first of these components, the regional *resistance*, that Martin (2012, p. 11) has defined as «the vulnerability or sensitivity of a regional economy to disturbances and disruptions, such as recessions».

In particular, with respect to the last economic downturn, it comes out that the more resilient regions have been the more diversified ones. Also the embeddedness in the international markets, the endowment of an innovative and high-skilled workforce and the presence of urban centres, seems to have played a major role in pushing the adaptive and reactive capacity of regions against the crisis (ESPON 2014a). More specifically, the European Program ESPON (2014b, p. 12 and 25) reported that «Regions which specialise in a narrow range of sectors are more likely to be vulnerable than more diversified regional economies. They risk suffering permanent reductions in the numbers of firms and jobs. However, no territorial endowments or public policies can fully insulate regions from the impacts of global economic crises or guarantee their recovery. [...] This points towards a greater emphasis on place-based policy approaches to build adaptability to withstand and recover from exogenous economic shocks». Moreover, «Employment rates are significantly higher in urban areas in many European countries. [...] Often, boosting education levels in an area is seen as a means to combat unemployment and even as a route to recovery. However, simply increasing the extent of educational qualifications does not appear to confer greater levels of resilience. Indeed, resilience is rather a long-term phenomenon; it cannot be easily conjured through short-term actions. Places with more stable long-term growth patterns tend to be more resilient. This points to a key role for long-term policy actions in building resilience».

As said, it is expected that a higher *adaptability* of a regional economy helps its *resilience* and *resistance* capacity. And, as highlighted by Grabher

> Adaptability crucially depends on the availability of unspecific and uncommitted capacities that can be put to a variety of unforeseeable uses: redundancy. Redundancy enables social systems not just to adapt to specific environmental changes but to question the appropriateness of adaptation. It is this kind of self-questioning ability that underpins the activities of systems capable of learning to learn and self-organize
> —Grabher 1993, p. 265.

Therefore, following a dense and still growing stream of literature, in this section, we will look specifically at *diversity* as a driver of regional economic performance and as a supportive element for its *resilience*, focusing on the *resistance* phases. Indeed, the capabilities heterogeneity of an economic system has been identified, by many scholars, as a key factor in explaining the differences in economic patterns followed by regions and in their output levels, particularly when they are affected by recessionary shocks. Moreover, more recently the literature has also clarified the key role played by the *relatedness* between these different elements and the *rarity* of the knowledge items, as already highlighted in Sec. 2.

6.2 The effect of technological diversity of the resistance of the European regions

Using a simple proxy proposed e.g. by Cappelli et al. (2018), we can plot the *resistance* of the EU regions as the difference (in logarithms) between the employment rate in 2007 and the minimum level experienced by the region in the following period. Therefore, I propose to use the *instantaneous growth rate* of the employment levels of each European region as a proxy the *resistance* capacity of the European regions

$$\text{resistance}_r = \frac{1}{s_r} \log \frac{\min(\mathrm{E}_{r,2008\text{–}2012})}{\mathrm{E}_{r,2007}}, \tag{3}$$

where $E_{r,t}$ is the level of employment of region r in year t, while s is the time span between 2007 and the year in which the employment reaches its minimum in the window 2008–2012 (immediately-post-crisis *peak*). I prefer this slightly different version of the measure used by Cappelli et al. (2018), since I see are more *resistant* a region that takes more time than another to reach the same (minimum) employment level. Fig. 10 shows a map of the spatial patterns of the index of regional *resistance* that will be used in the analysis that follows.

6.2.1 Empirical strategy

Therefore, in this last part of the paper I will estimate the effect of the *diversity* of the *knowledge capital* (in its different components) on the employment *resistance* of the European regions against the Great Recession shock. This will be a way to test the different measures proposed in the previous sections, exploring them within an empirical exercise. I will use Weighted Least Squares (WLS) to estimate the following equation

$$\text{resistance}_r = \alpha + \beta_0 \log(E_{r,2007}) + \beta_1 \log(\text{HC}) + \beta_2 X_r + \beta_3 D_r + \varepsilon,$$

in which I used as weights the regional population density. The measure on the left-hand side is the index with which we operationalise the regional *resistance* (Eq. 3) and, as explained above, considers the years from 2007 to 2012. Instead, on the right-hand side of the equation, all the variables used refer to the year 2006. The only exception is the logarithm of the employment level, that refers to 2007, since it controls for Solow-style convergence of regional employment.[28] The other control variable included accounts for the human capital (HC) of the region. The human capital level is proxied by the share of people who have successfully completed a tertiary level education and are employed in a S&T (science and technology) occupation (HRSTC). The D_r group of variables is composed of two dummies. The `Capital` dummy variable controls for the expected out-performances of the capital city's region of each country thanks to several factors like the higher concentration of public sector activities, research institutes and high value-added activities (Dijkstra et al. 2015; Hoekman et al. 2009). While the `EU15` dummy accounts for the belonging of a region to the first 15 nations that joined the European Union, since these are expected to be nations in a more mature stage of capitalism. Lastly, the X_r group collects all the variables main variables of interest that capture the different aspects of regional technological *diversity*. Tab. 4 reports the explanation of the symbols used in the regression tables.

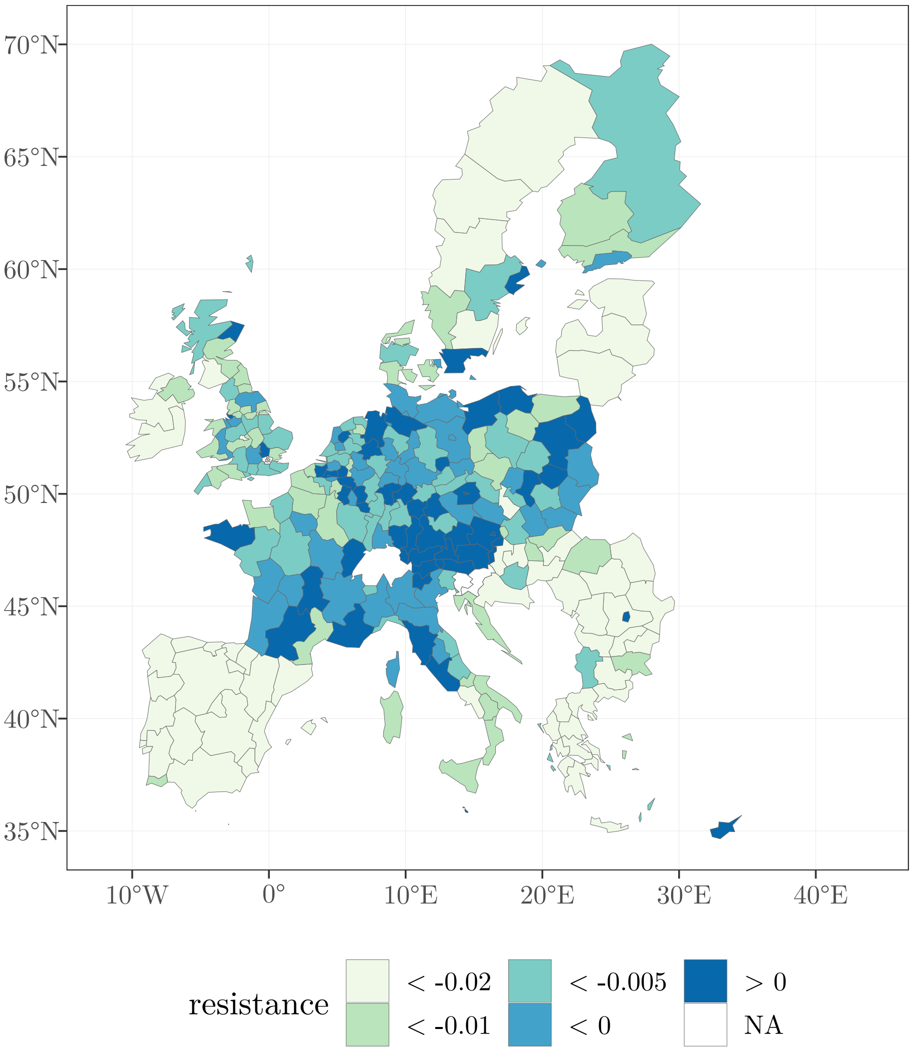
6.2.2 Results

The results of the analysis are reported in Tab. 8 and 9.[29] We can see that the use of the Evenness (and its two decomposed parts), as well as of the Coherence are significantly different

[28] Indeed, it is possible to rewrite the equation as $\log(\min E_{r,2008\text{–}2012}) = \alpha + (\beta_0+1)\log(E_{r,2007}) + \beta_1 \log(\text{HC}) + \beta_2 X_r + \beta_3 D_r + \varepsilon$. Therefore, we can also say that we are estimating the determinants of the minimum employment level between 2008 and 2012, controlling for employment in 2007.

[29] Instead, Tab. 18–22 report the regression tables for the same models commented here, with standard errors clustered at the country level. Most of the results are confirmed, and the fact that instead none of the results of the table that uses the measures introduced in Sec. 3 are no longer significant seems another signal of the importance of the normalisations proposed in Sec. 5.

Figure 10: Regional (employment) resistance. Data: Eurostat.



from zero once we include the RTA-based *variety* measure in the regressions. Conversely, the significance of the estimates of the Shannon Entropy, Related-Unrelated Variety, and t-statistic Coherence is more strongly affected by the introduction of the *variety* measure in the regressions. This is confirmed also looking at the Variance Inflation Factors (VIF) of the terms for the first two groups of regressions (Tab. 10 and 11). The values clearly suggest the existence of severe multicollinearity between the Shannon Entropy and the RTA-based *variety*, as well as between the Related Variety and this last index. These results are in line with the analysis carried on in Sec. 5. The use of the measures introduced in this last section helps to discriminate the (positive) effect of the *variety* per se on the *resistance* capacity of the European regions, from the other effects due to the evenness of the elements that compose the *knowledge capital* of each region. About this last point, even though this goes beyond the aims of this paper, we can risk an interpretation of the results just shown. From Tab. 9, we can say that, once controlled for the level of regional *variety*, the more *resistant* regions have been those endowed, before the arrival of the shock, with a *knowledge capital* more focused on few of the macro-groups of technological domains in which they have shown the ability to develop patents. At the same time, the results show also a negative relationship between the average within-groups Evenness and its *resistance* capacity. Lastly, the analysis suggests that those areas endowed with a higher average epistemic similarity between the items of its *knowledge capital* show a stronger *resistance* in terms of employment levels against an exogenous shock.

About the *rarity*, the three measures proposed in Sec. 5 are introduced in the regression one at a time, together with the other *diversity* measures (see Tab. 12–14). In the first of these tables, the Weighted Technological Complexity Index is used as a binary classifier that has been shown able to identify these regions with capabilities in high-medium technology sectors (see Fig. 7a). The results do not show any significant effect. The analysis seems to say that regions that belonged to this group in the year before the shock do not show a higher *resistance* capacity, compared to the other regions. This seems in contrast with the existing theoretical and empirical literature. However, as already said, the results here shown wants to be a way to corroborate the analysis developed in the previous sections of the paper, and they are too preliminary to derive some ultimate precept about the phenomenon analysed. Indeed, the findings from the other two *rarity* indices seem in line with the expectations of the literature. In these last two tables the *variety* index is not included, since all the measures that account for *rarity* are supposed to be a synthetic indicator that combines together this last dimension with the "pure" regional diversity. The results show that regions characterised by higher levels of Fitness or Rarity-weighted diversity before being shocked show a better reaction in terms of employment *resistance*. About the other *diversity* measures, the introduction of the `CX>0` variable does not affect significantly the results already discussed, both in terms of significance and of punctual estimate. Conversely, when the `FX` or the `RWD` measures are introduced, the other variables become non-significant, or their credibility is seriously questioned. Therefore, this seems to say that the measure that tries to combine *rarity* and *diversity* in a synthetic index are able to identify better the pre-conditions for a stronger regional *resistance* capacity. However, Tab. 15–17 show that there seems to be no or mild multicollinearity issues in this group of regressions.[30] This go, once again, in the direction of a confirm that the measures proposed in Sec. 5 capture aspects of the regional technological *diversity* that are not the *variety*.

[30]The only exception is log(RTA) in model (3) of Tab. 12, but a value of 4.64 is in any case below any rule of thumb proposed in the literature to identify strong multicollinearity cases (Kutner et al. 2005; Sheather 2008).

Table 8: Regressions using the indices introduced in Sec. 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(E)	−0.010***	−0.012***	−0.012***	−0.010***	−0.009***	−0.013***	−0.013***	−0.013***	−0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(HC)	0.008**	0.006	0.004	0.008**	0.011**	0.003	0.005	−0.00004	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(ETP)		0.036***				−0.048*			
		(0.010)				(0.028)			
log(RV)			0.026***				−0.015		
			(0.006)				(0.019)		
log(UV)				0.003				−0.045***	
				(0.014)				(0.016)	
log(CT)					0.055***				0.038*
					(0.020)				(0.020)
log(RTA)						0.025***	0.018**	0.017***	0.011***
						(0.008)	(0.008)	(0.003)	(0.003)
E15	0.005*	−0.001	−0.003	0.005*	0.004	−0.006	−0.005	−0.007**	−0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Capital	0.015***	0.016***	0.017***	0.015***	0.016***	0.015***	0.015***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Const.	0.033**	−0.008	0.030**	0.031*	−0.085*	0.045*	0.004	0.043**	−0.068
	(0.014)	(0.017)	(0.013)	(0.018)	(0.045)	(0.024)	(0.017)	(0.018)	(0.044)
Obs.	247	247	247	247	247	247	247	247	247
R^2	0.366	0.400	0.406	0.366	0.385	0.424	0.419	0.435	0.426
Adj. R^2	0.356	0.387	0.393	0.353	0.372	0.410	0.404	0.421	0.411

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 9: Regressions using the indices introduced in Sec. 5

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(E)	-0.010^{***}	-0.011^{***}	-0.010^{***}	-0.009^{***}	-0.011^{***}	-0.014^{***}	-0.014^{***}	-0.012^{***}	-0.014^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(HC)	0.008^{**}	0.002	0.008^{**}	0.006	0.009^{**}	−0.002	0.002	−0.001	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(EVS)		-0.068^{***}				-0.069^{***}			
		(0.017)				(0.017)			
log(RE)			0.003				-0.051^{***}		
			(0.013)				(0.015)		
log(UE)				−0.027				-0.049^{***}	
				(0.017)				(0.017)	
log(CP)					0.016^{***}				0.013^{***}
					(0.004)				(0.004)
log(RTA)						0.012^{***}	0.019^{***}	0.014^{***}	0.011^{***}
						(0.003)	(0.003)	(0.003)	(0.003)
E15	0.005^{*}	−0.00000	0.005	0.005	0.005^{*}	-0.010^{***}	-0.007^{**}	-0.007^{*}	−0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
Capital	0.015^{***}	0.014^{***}	0.015^{***}	0.015^{***}	0.015^{***}	0.015^{***}	0.014^{***}	0.016^{***}	0.016^{***}
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Const.	0.033^{**}	0.050^{***}	0.036^{**}	0.031^{**}	0.054^{***}	0.029^{**}	-0.039^{*}	0.006	0.032^{**}
	(0.014)	(0.014)	(0.017)	(0.014)	(0.014)	(0.014)	(0.020)	(0.014)	(0.015)
Obs.	247	247	247	247	247	247	247	247	247
R^2	0.366	0.404	0.366	0.373	0.403	0.455	0.444	0.436	0.442
Adj. R^2	0.356	0.391	0.353	0.360	0.390	0.442	0.430	0.422	0.428

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 10: VIF of the variables included in the regressions of Tab. 8. The numbers in the columns correspond to the name of the models of the regression table.

	(6)	(7)	(8)	(9)
log(E)	1.24	1.27	1.26	1.35
log(HC)	1.68	1.69	1.88	1.75
log(ETP)	16.12			
log(RV)		17.64		
log(UV)			1.67	
log(CT)				1.32
log(RTA)	18.72	18.69	3.09	2.27
EU15	1.92	1.85	2.00	1.85
Capital	1.67	1.74	1.63	1.66

Table 11: VIF of the variables included in the regressions of Tab. 9. The numbers in the columns correspond to the name of the models of the regression table.

	(6)	(7)	(8)	(9)
log(E)	1.27	1.33	1.29	1.27
log(HC)	1.91	1.69	1.95	1.66
log(EVS)	1.72			
log(RE)		1.92		
log(UE)			1.37	
log(CP)				1.09
log(RTA)	2.18	3.56	2.31	2.24
EU15	2.11	1.96	1.94	1.88
Capital	1.62	1.96	1.62	1.61

Table 12: Regressions with Complexity used as a clustering algorithm

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.012^{***}	-0.013^{***}	-0.014^{***}	-0.012^{***}	-0.013^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(HC)	0.004	-0.003	0.001	-0.0003	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(EVS)		-0.070^{***}			
		(0.017)			
log(RE)			-0.056^{***}		
			(0.015)		
log(UE)				-0.043^{**}	
				(0.018)	
log(CP)					0.013^{***}
					(0.004)
CX>0	-0.005^{*}	-0.005^{**}	-0.006^{**}	-0.003	-0.005^{*}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
log(RTA)	0.014^{***}	0.015^{***}	0.023^{***}	0.015^{***}	0.013^{***}
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
E15	-0.003	-0.008^{**}	-0.006^{*}	-0.006	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Capital	0.015^{***}	0.013^{***}	0.012^{***}	0.016^{***}	0.015^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Const.	0.001	0.017	-0.058^{***}	0.00003	0.022
	(0.015)	(0.015)	(0.022)	(0.015)	(0.016)
Obs.	247	247	247	247	247
R^2	0.425	0.465	0.456	0.438	0.449
Adj. R^2	0.411	0.449	0.440	0.422	0.433

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Regressions with Fitness

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.015^{***}	-0.015^{***}	-0.015^{***}	-0.014^{***}	-0.015^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log(HC)	0.003	-0.001	0.002	-0.001	0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(EVS)		-0.053^{***}			
		(0.017)			
log(RE)			-0.025^{*}		
			(0.013)		
log(UE)				-0.042^{**}	
				(0.017)	
log(CP)					0.010^{**}
					(0.004)
log(FX)	0.008^{***}	0.007^{***}	0.009^{***}	0.009^{***}	0.007^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
E15	-0.006	-0.008^{**}	-0.006^{*}	-0.007^{**}	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Capital	0.016^{***}	0.015^{***}	0.015^{***}	0.016^{***}	0.016^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Const.	0.089^{***}	0.095^{***}	0.079^{***}	0.090^{***}	0.093^{***}
	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)
Obs.	247	247	247	247	247
R^2	0.425	0.446	0.433	0.440	0.439
Adj. R^2	0.413	0.433	0.419	0.426	0.425

Note: *p<0.1; **p<0.05; ***p<0.01

Table 14: Regressions with Rarity-weighted diversity

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.016^{***}	-0.016^{***}	-0.016^{***}	-0.015^{***}	-0.016^{***}
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
log(HC)	-0.002	-0.004	-0.003	-0.006	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(EVS)		-0.033^{*}			
		(0.017)			
log(RE)			-0.029^{**}		
			(0.012)		
log(UE)				-0.038^{**}	
				(0.016)	
log(CP)					0.007^{*}
					(0.004)
log(RWD)	0.013^{***}	0.011^{***}	0.014^{***}	0.013^{***}	0.012^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
E15	-0.012^{***}	-0.013^{***}	-0.012^{***}	-0.013^{***}	-0.010^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Capital	0.016^{***}	0.016^{***}	0.015^{***}	0.017^{***}	0.016^{***}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Const.	0.107^{***}	0.109^{***}	0.094^{***}	0.107^{***}	0.111^{***}
	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)
Obs.	247	247	247	247	247
R^2	0.477	0.485	0.489	0.489	0.484
Adj. R^2	0.466	0.472	0.476	0.477	0.471

Note: *p<0.1; **p<0.05; ***p<0.01

Table 15: VIF of the variables included in the regressions of Tab. 13. The numbers in the columns correspond to the name of the models of the regression table.

	(2)	(3)	(4)	(5)
log(E)	1.31	1.35	1.32	1.30
log(HC)	1.92	1.69	1.97	1.66
log(EVS)	1.75			
log(RE)		1.97		
log(UE)			1.45	
log(CP)				1.09
log(CX)	1.94	1.97	2.03	1.92
log(RTA)	2.87	4.64	2.89	2.90
EU15	2.13	1.98	1.94	1.88
Capital	1.76	1.85	1.77	1.73

Table 16: VIF of the variables included in the regressions of Tab. 13. The numbers in the columns correspond to the name of the models of the regression table.

	(2)	(3)	(4)	(5)
log(E)	1.53	1.59	1.58	1.53
log(HC)	1.89	1.70	1.96	1.71
log(EVS)	1.80			
log(RE)		1.40		
log(UE)			1.33	
log(CP)				1.19
log(FX)	2.77	3.17	2.73	2.98
EU15	2.02	1.92	1.97	2.03
Capital	1.62	1.68	1.61	1.60

Table 17: VIF of the variables included in the regressions of Tab. 14. The numbers in the columns correspond to the name of the models of the regression table.

	(2)	(3)	(4)	(5)
log(E)	1.45	1.48	1.52	1.45
log(HC)	1.94	1.82	2.05	1.85
log(EVS)	1.93			
log(RE)		1.33		
log(UE)			1.30	
log(CP)				1.20
log(RWD)	3.30	3.34	2.97	3.33
EU15	2.16	2.13	2.16	2.28
Capital	1.64	1.67	1.62	1.61

Table 18: Regressions using the indices introduced in Sec. 3 with SE clustered at the country level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(E)	−0.010***	−0.012***	−0.012***	−0.010***	−0.009***	−0.013***	−0.013***	−0.013***	−0.012***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
log(HC)	0.008	0.006	0.004	0.008	0.011	0.003	0.005	−0.00004	0.006
	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)	(0.007)
log(ETP)		0.036				−0.048			
		(0.026)				(0.061)			
log(RV)			0.026				−0.015		
			(0.016)				(0.022)		
log(UV)				0.003				−0.045	
				(0.029)				(0.036)	
log(CT)					0.055				0.038
					(0.042)				(0.041)
log(RTA)						0.025	0.018**	0.017**	0.011*
						(0.017)	(0.009)	(0.008)	(0.006)
E15	0.005	−0.001	−0.003	0.005	0.004	−0.006	−0.005	−0.007	−0.004
	(0.006)	(0.008)	(0.008)	(0.006)	(0.006)	(0.010)	(0.009)	(0.010)	(0.008)
Capital	0.015**	0.016***	0.017***	0.015**	0.016***	0.015***	0.015***	0.017***	0.017***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Const.	0.033**	−0.008	0.030*	0.031	−0.085	0.045	0.004	0.043*	−0.068
	(0.016)	(0.033)	(0.016)	(0.029)	(0.092)	(0.044)	(0.021)	(0.026)	(0.087)
Obs.	247	247	247	247	247	247	247	247	247
R^2	0.366	0.400	0.406	0.366	0.385	0.424	0.419	0.435	0.426
Adj. R^2	0.356	0.387	0.393	0.353	0.372	0.410	0.404	0.421	0.411

Note: *p<0.1; **p<0.05; ***p<0.01

Table 19: Regressions using the indices introduced in Sec. 5 with SE clustered at the country level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(E)	-0.010^{***}	-0.011^{***}	-0.010^{***}	-0.009^{***}	-0.011^{***}	-0.014^{***}	-0.014^{***}	-0.012^{***}	-0.014^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
log(HC)	0.008	0.002	0.008	0.006	0.009	-0.002	0.002	-0.001	0.005
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)
log(EVS)		-0.068^{*}				-0.069^{**}			
		(0.036)				(0.035)			
log(RE)			0.003				-0.051^{*}		
			(0.025)				(0.026)		
log(UE)				-0.027				-0.049	
				(0.033)				(0.036)	
log(CP)					0.016^{**}				0.013^{**}
					(0.007)				(0.005)
log(RTA)						0.012^{**}	0.019^{***}	0.014^{**}	0.011^{*}
						(0.006)	(0.008)	(0.007)	(0.006)
E15	0.005	-0.00000	0.005	0.005	0.005	-0.010	-0.007	-0.007	-0.003
	(0.006)	(0.008)	(0.006)	(0.006)	(0.006)	(0.011)	(0.010)	(0.010)	(0.009)
Capital	0.015^{**}	0.014^{**}	0.015^{**}	0.015^{**}	0.015^{**}	0.015^{**}	0.014^{**}	0.016^{***}	0.016^{***}
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Const.	0.033^{**}	0.050^{***}	0.036	0.031^{*}	0.054^{***}	0.029^{*}	-0.039	0.006	0.032
	(0.016)	(0.017)	(0.026)	(0.017)	(0.021)	(0.017)	(0.036)	(0.022)	(0.020)
Obs.	247	247	247	247	247	247	247	247	247
R^2	0.366	0.404	0.366	0.373	0.403	0.455	0.444	0.436	0.442
Adj> R^2	0.356	0.391	0.353	0.360	0.390	0.442	0.430	0.422	0.428

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 20: Regressions with Complexity used as a clustering algorithm with SE clustered at the country level.

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.012^{***}	-0.013^{***}	-0.014^{***}	-0.012^{***}	-0.013^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(HC)	0.004	-0.003	0.001	-0.0003	0.005
	(0.006)	(0.006)	(0.005)	(0.007)	(0.007)
log(EVS)		-0.070^{**}			
		(0.033)			
log(RE)			-0.056^{**}		
			(0.022)		
log(UE)				-0.043	
				(0.041)	
log(CP)					0.013^{**}
					(0.005)
CX>0	-0.005	-0.005	-0.006	-0.003	-0.005
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
log(RTA)	0.014^{**}	0.015^{**}	0.023^{***}	0.015^{**}	0.013^{**}
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
E15	-0.003	-0.008	-0.006	-0.006	-0.002
	(0.010)	(0.012)	(0.011)	(0.011)	(0.010)
Capital	0.015^{***}	0.013^{***}	0.012^{***}	0.016^{***}	0.015^{***}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Const.	0.001	0.017	-0.058^{*}	0.00003	0.022
	(0.025)	(0.025)	(0.032)	(0.026)	(0.027)
Obs.	247	247	247	247	247
R^2	0.425	0.465	0.456	0.438	0.449
Adj. R^2	0.411	0.449	0.440	0.422	0.433

Note: *p<0.1; **p<0.05; ***p<0.01

Table 21: Regressions with Fitness with SE clustered at the country level.

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.015^{***}	-0.015^{***}	-0.015^{***}	-0.014^{***}	-0.015^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(HC)	0.003	-0.001	0.002	-0.001	0.004
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
log(EVS)		-0.053^{*}			
		(0.031)			
log(RE)			-0.025		
			(0.028)		
log(UE)				-0.042	
				(0.034)	
log(CP)					0.010^{**}
					(0.005)
log(FX)	0.008^{*}	0.007^{*}	0.009^{*}	0.009^{*}	0.007
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
E15	-0.006	-0.008	-0.006	-0.007	-0.004
	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)
Capital	0.016^{**}	0.015^{**}	0.015^{**}	0.016^{***}	0.016^{**}
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Const.	0.089^{***}	0.095^{***}	0.079^{***}	0.090^{***}	0.093^{***}
	(0.033)	(0.032)	(0.031)	(0.033)	(0.033)
Obs.	247	247	247	247	247
R^2	0.425	0.446	0.433	0.440	0.439
Adj. R^2	0.413	0.433	0.419	0.426	0.425

Note: *p<0.1; **p<0.05; ***p<0.01

Table 22: Regressions with Rarity-weighted diversity with SE clustered at the country level.

	(1)	(2)	(3)	(4)	(5)
log(E)	-0.016^{***}	-0.016^{***}	-0.016^{***}	-0.015^{***}	-0.016^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log(HC)	-0.002	-0.004	-0.003	-0.006	-0.001
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
log(EVS)		-0.033			
		(0.035)			
log(RE)			-0.029		
			(0.021)		
log(UE)				-0.038	
				(0.032)	
log(CP)					0.007^{*}
					(0.004)
log(RWD)	0.013^{***}	0.011^{***}	0.014^{***}	0.013^{***}	0.012^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
E15	-0.012	-0.013	-0.012	-0.013	-0.010
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Capital	0.016^{***}	0.016^{***}	0.015^{***}	0.017^{***}	0.016^{***}
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
Const.	0.107^{***}	0.109^{***}	0.094^{***}	0.107^{***}	0.111^{***}
	(0.029)	(0.028)	(0.028)	(0.029)	(0.029)
Obs.	247	247	247	247	247
R^2	0.477	0.485	0.489	0.489	0.484
Adj. R^2	0.466	0.472	0.476	0.477	0.471

Note: *p<0.1; **p<0.05; ***p<0.01

7 Conclusions

Even though the debate on the role of the regional diversity and its effects on the economic performance of these economic systems is going on long since, the measurement of this characteristic of the regional economic structure is still an open question. Moreover, more recently, the idea of *diversity* has been opened up, trying to identify different aspects of this bundle of dimensions: namely, *variety*, *balance*, *relatedness*, and *rarity.* This has introduced new challenges for its measurement, because some measures, like the ones shown in the paper that are among the most used in the literature, confound more than one these aspects together, so that their interpretability in the results of an empirical exercise is not that clear-cut as largely assumed.

In the paper, I have critically reviewed the main measures proposed in the literature, highlighting their main limitations and drawbacks, particularly with respect to the technological aspect of the regional diversity. One main topic investigated is the dependence of most of the indices considered on the *size-effect* –i.e., on the fact that they grow with the *variety*, and not just with what they are supposed to measure, being that *balance* or *relatedness.* A second main issue explored is the dependence of some of the measures introduced on the structural characteristics of the occurrence matrix on which they are computed. Moreover, it has been shown that the problems exposed are also connected to the question of the aggregation level at which to compute the measures, since for example at a deeper level of the tree-structured classifications we will observe, structurally, a higher *variety.*

The exploration of each measure and the comparison between the different methods and definitions carried on in Sec. 5, as well as the results of the empirical application proposed in Sec. 6, show that the solutions proposed for each of the issues raised about the indices introduced in Sec. 3 are helpful tools for the empirical analysis of the effect of the *variety*, *balance*, and *relatedness* of the capability structure of a region, as well as of the *rarity* of the items of this bundle, on its economic performance. Therefore, the results of this paper suggest that, by solving the problems and limitations exposed, we will be able to reach more clear and reliable interpretations of the results, as well as an easier comparability across different empirical investigations. And, even though not conclusive, the analysis developed this paper goes in this direction.

Acknowledgements

First of all, I have to thank my supervisor, Aldo Geuna, for the comments and the guide he provided me with this paper. I am also indebted with Ricardo Hausmann, Frank Neffke, and all the resident and visiting Fellows of the Growth Lab – CID, Harvard University, whose critiques and advises have radically improved the quality of this work. Moreover, my gratitude goes to the members of the RENIR project, in particular Riccardo Cappelli and Ben Armstrong who comment previous versions of this paper during the group meetings. As well, I thank all the participants of the INET-YSI Innovation, Economic Complexity, and Economic Geography Workshop for the comments and critiques. In particular, my thoughts go to Cesar Hidalgo, Mercedes Delgado, and Andrea Morrison. And I am grateful to Francesco Capone, Niccolò Innocenti, and all the other participants of the 1st International Workshop on Rethinking Clusters to have fostered and guided me in restructuring the paper. I thank the participants of the conferences GeoInno 2018 and WICK#5 as well, for all the precious comments that provided me. Lastly, I am indebted to Martina Iori, Fabrizio Fusillo, and Silvia Rocchetta to the support they provide me during the construction of the paper, with constant comments, opinions and critiques that have definitely improved the quality of the work. Any errors are my own.

References

Albert, Réka, Hawoong Jeong and Albert-László Barabási (July 2000). 'Error and attack tolerance of complex networks'. *Nature* 406.6794, 378–382.

Alstott, Jeff, Giorgio Triulzi, Bowen Yan and Jianxi Luo (2016). 'Mapping Technology Space by Normalizing Patent Networks'. *Scientometrics* 110.1, 443–479. DOI: 10.1007/s11192-016-2107-y.

Antonelli, Cristiano (2011). 'The Economic Complexity of Technological Change: Knowledge Interaction and Path Dependence'. In: *Handbook on the Economic Complexity of Technological Change*. Ed. by Cristiano Antonelli. Cheltenham (UK) and Northampton (MA): Edward Elgar.

— (2013). *The Economic Complexity of Innovation as a Creative Response*. Working Paper 26. Dipartimento di Economia "S. Cognetti de Martiis", Università degli Studi di Torino and LEI & BRICK, Collegio Carlo Alberto.

— (2015). 'Innovation as a Creative Response: A Reappraisal of the Schumpeterian Legacy'. *History of Economic Ideas* 23.2, 99–118.

Antonelli, Cristiano and Alessandra Colombelli (2015). 'External and Internal Knowledge in the Knowledge Generation Function'. *Industry and Innovation* 22.4, 273–298. DOI: 10.1080/13662716.2015.1049864.

Antonelli, Cristiano, Francesco Crespi, Christian A. Mongeau Ospina and Giuseppe Scellato (2017). 'Knowledge Composition, Jacobs Externalities and Innovation Performance in European Regions'. *Regional Studies* 51 (11), 1708–1720. DOI: 10.1080/00343404.2016.1217405.

Antonelli, Cristiano and Gianluigi Ferraris (2011). 'Innovation as an Emerging System Property'. *Journal of Artificial Societies and Social Simulation* 14.2, 1. URL: http://jasss.soc.surrey.ac.uk/14/2/1.html.

Antonelli, Cristiano, Jackie Krafft and Francesco Quatraro (Mar. 2010). 'Recombinant Knowledge and Growth: The Case of ICTs'. *Structural Change and Economic Dynamics* 21 (1), 50–69. DOI: 10.1016/j.strueco.2009.12.001.

Antonelli, Cristiano and Giuseppe Scellato (Jan. 2013). 'Complexity and Technological Change: Knowledge Interactions and Firm Level Total Factor Productivity'. *Journal of Evolutionary Economics* 23.1, 77–96.

Archibugi, Daniele and Mario Pianta (Feb. 1992a). 'Specialization and Size of Technological Activities in Industrial Countries: The Analysis of Patent Data'. *Research Policy* 21.1, 79–93. DOI: 10.1016/0048-7333(92)90028-3.

— (1992b). *The Technological Specialization of Advanced Countries. A Report to the EEC on International Science and Technology Activities*. Dordrecht (NL): Kluwer Academic Publishers. ISBN: 9780792317500. DOI: 10.1007/978-94-015-7999-5.

Attaran, Mohsen (1984). 'The Relation of Economic Diversity to Levels, Growth Rates, and Stability of Unemployment and Income'. PhD thesis. Portland State University.

— (July 1986). 'Industrial Diversity and Economic Performance in U.S. Areas'. *The Annals of Regional Science* 20.2, 44–54. DOI: 10.1007/bf01287240.

Attaran, Mohsen and Martin Zwick (Autumn 1987). 'Entropy and Other Measures of Industrial Diversification'. *Quarterly Journal of Business and Economics* 26.4, 17–34.

Aujean, Laurent, Virginia Maestri, Filip Tanay and Céline Thévenot (2015). 'The Legacy of the Crisis: Resilience and Challenges'. In: *Employment and Social Developments in Europe 2014*. European Union. Chap. 1.

Balassa, Bela (1961). *The Theory of Economic Integration*.

Balland, Pierre-Alexandre and David L. Rigby (2017). 'The Geography of Complex Knowledge'. *Economic Geography* 93.1, 1–23. DOI: 10.1080/00130095.2016.1205947.

Barth, James, John Kraft and Philip Wiest (Apr. 1975). 'A Portfolio Theoretic Approach to Industrial Diversification and Regional Employment'. *Journal of Regional Science* 15.1, 9–15. DOI: 10.1111/j.1467-9787.1975.tb01127.x.

Bates, Douglas and Martin Maechler (2018). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-14. URL: https://CRAN.R-project.org/package=Matrix.

Boschma, Ron A. (2015). 'Towards an Evolutionary Perspective on Regional Resilience'. *Regional Studies* 49 (5), 733–751. DOI: 10.1080/00343404.2014.959481.

Boschma, Ron A. and Simona Iammarino (July 2009). 'Related Variety, Trade Linkages, and Regional Growth in Italy'. *Economic Geography* 85.3, 289–311. DOI: 10.1111/j.1944-8287.2009.01034.x.

Boschma, Ron A., Asier Minondo and Mikel Navarro (June 2012). 'Related Variety and Regional Growth in Spain'. *Papers in Regional Science* 91 (2), 241–256. DOI: 10.1111/j.1435-5957.2011.00387.x.

Bottazzi, Giulio and Davide Pirino (Nov. 2010). 'Measuring Industry Relatedness and Corporate Coherence'. *SSRN Electronic Journal*, 1–24. DOI: 10.2139/ssrn.1831479.

Breschi, Stefano, Francesco Lissoni and Franco Malerba (Jan. 2003). 'Knowledge-relatedness in Firm Technological Diversification'. *Research Policy* 32 (1), 69–87. DOI: 10.1016/s0048-7333(02)00004-5.

Cappelli, Riccardo, Fabio Montobbio and Andrea Morrison (July 2018). *Unemployment Resistance Across EU Regions: The Role of Technological and Human Capital*. Research rep. 18.31. Papers in Evolutionary Economic Geography, Utrecht University. URL: https://ideas.repec.org/p/egu/wpaper/1831.html.

Castaldi, Carolina, Koen Frenken and Bart Los (2014). 'Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting'. *Regional Studies* 49 (5), 767–781. DOI: 10.1080/00343404.2014.940305.

Conroy, Michael E. (1972). 'Optimal Regional Industrial Diversification: A Portfolio-Analytic Approach'. PhD thesis. University of Illinois at Urbana-Champaign.

— (1975a). *Regional Economic Diversification*. New York: Praeger.

— (Jan. 1975b). 'The Concept and Measurement of Regional Industrial Diversification'. *Southern Economic Journal* 41.3, 492–505. ISSN: 00384038.

Content, Jeroen and Koen Frenken (Oct. 2016). 'Related Variety and Economic Development: A Literature Review'. *European Planning Studies* 24.12, 2097–2112. DOI: 10.1080/09654313.2016.1246517.

Cortinovis, Nicola and Frank van Oort (Oct. 2015). 'Variety, Economic Growth and Knowledge Intensity of European Regions: A Spatial Panel Analysis'. *The Annals of Regional Science* 55.1, 7–32. DOI: 10.1007/s00168-015-0680-2.

Cristelli, Matthieu, Andrea Gabrielli, Andrea Tacchella, Guido Caldarelli and Luciano Pietronero (Aug. 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.

Dijkstra, Lewis, Enrique Garcilazo and Philip McCann (Sept. 2015). 'The Effects of the Global Financial Crisis on European Regions and Cities'. *Journal of Economic Geography* 15 (5), 935–949. DOI: 10.1093/jeg/lbv032.

Diodato, Dario and Anet B. R. Weterings (July 2015). 'The Resilience of Regional Labour Markets to Economic Shocks: Exploring the Role of Interactions Among Firms and Workers'. *Journal of Economic Geography* 15 (4), 723–742. DOI: 10.1093/jeg/lbu030.

Dormann, Carsten F. (2011). 'How to Be a Specialist? Quantifying Specialisation in Pollination Networks'. *Network Biology* 1.1, 1–20.

Dormann, Carsten F., Jochen Fründ, Nico Blüthgen and Bernd Gruber (2009). 'Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks'. *The Open Ecology Journal* 2.1, 7–24. DOI: 10.2174/1874213000902010007.

Dormann, Carsten F., Bernd Gruber and Jochen Fründ (2008). 'Introducing the bipartite Package: Analysing Ecological Networks'. *R News* 8.2, 8–11.

van Eck, Nees Jan and Ludo Waltman (2009). 'How to Normalize Cooccurrence Data? An Analysis of Some Well-known Similarity Measures'. *Journal of the American Society for Information Science and Technology* 60.8, 1635–1651. DOI: 10.1002/asi.21075.

ESPON (2014a). *Territorial Dynamics in Europe: Economic Crisis and the Resilience of Regions.* Research rep. 12. Territorial Observations. ESPON 2013 Programme. URL: https://www.espon.eu/topics-policy/publications/territorial-observations/economic-crisis-and-resilience-regions.

— (2014b). *Territories Finding a New Momentum: Evidence for Policy Development, Growth and Investment.* Research rep. 3. ESPON Synthesis Report. ESPON 2013 Programme. URL: https://www.espon.eu/topics-policy/publications/synthesis-reports/third-espon-2013-synthesis-report.

Freeman, Christopher and Luc G. Soete (1997). *The Economics of Industrial Innovation.* 3rd ed. An earlier version of the book was published by Penguin Modern Economics Texts in 1974. The MIT Press.

Frenken, Koen (2006). *Innovation, Evolution and Complexity Theory.* Cheltenham (UK) and Northampton (MA): Edward Elgar. ISBN: 9781843761976.

— (2007). 'Entropy statistics and information theory'. In: *Elgar Companion to Neo-Schumpeterian Economics.* Ed. by Horst Hanusch and Andreas Pyka. Cheltenham (UK) and Northampton (MA): Edward Elgar. Chap. 33, 544–555.

Frenken, Koen and Ron A. Boschma (Sept. 2007). 'A Theoretical Framework for Evolutionary Economic Geography: Industrial Dynamics and Urban Growth as a Branching Process'. *Journal of Economic Geography* 7 (5), 635–649. DOI: 10.1093/jeg/lbm018.

Frenken, Koen, Frank Van Oort and Thijs Verburg (2007). 'Related Variety, Unrelated Variety and Regional Economic Growth'. *Regional Studies* 41 (5), 685–697. DOI: 10.1080/00343400601120296.

Gotelli, Nicholas J. (2001). 'Research Frontiers in Null Model Analysis'. *Global Ecology and Biogeography* 10.4, 337–343. DOI: 10.1046/j.1466-822x.2001.00249.x.

Gotelli, Nicholas J. and Gary R. Graves (1996). *Null Models in Ecology.* Washington (DC): Smithsonian Institution Press. URL: https://repository.si.edu/handle/10088/7782.

Grabher, Gernot (1993). 'The Weakness of Strong Ties: The Lock-in of Regional Development in the Ruhr Area'. In: *The Embedded Firm: On the Socioeconomics of Industrial Networks.* Ed. by Gernot Grabher. London & New York: Routledge. Chap. 12, 255–277.

Hackbart, Merlin M. and Donald A. Anderson (Nov. 1975). 'On Measuring Economic Diversification'. *Land Economics* 51.4, 374–378.

Hausmann, Ricardo and César A. Hidalgo (Dec. 2011). 'The Network Structure of Economic Output'. *Journal of Economic Growth* 16 (4), 309–342. DOI: 10.1007/s10887-011-9071-4.

Hausmann, Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jimenez, Alexander Simoes and Muhammed A. Yildirim (2014). *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* First edition 2011. The MIT Press.

Henderson, Rebecca (1994). 'The Evolution of Integrative Capability: Innovation in Cardiovascular Drug Discovery'. *Industrial and Corporate Change* 3 (3), 607–630. DOI: 10.1093/icc/3.3.607.

Hidalgo, César A. (2015). *Why Information Grows: The Evolution of Order, from Atoms to Economies.* New York: Basic Books.

Hidalgo, César A., Pierre-Alexandre Balland, Ron A. Boschma, Mercedes Delgado, Maryann P. Feldman, Koen Frenken, Edward L. Glaeser, Canfei He, Dieter F. Kogler, Andrea Morrison, Frank Neffke, David L. Rigby, Scott Stern, Siqi Zheng and Shengjun Zhu (2018). 'The Principle of Relatedness'. In: *International Conference on Complex Systems (ICCS 2018). Unifying Themes in Complex Systems IX*. Ed. by Alfredo J. Morales, Carlos Gershenson, Dan Braha, Ali A. Minai and Yaneer Bar-Yam. Springer Proceedings in Complexity. Cham (CH): Springer, 451–457. ISBN: 978-3-319-96661-8. DOI: 10.1007/978-3-319-96661-8_46.

Hidalgo, César A. and Ricardo Hausmann (June 2009). 'The Building Blocks of Economic Complexity'. *Proceedings of the National Academy of Sciences of the United States of America* 106.26, 10570–10575. DOI: 10.1073/pnas.0900943106.

Hirschman, Albert O. (1958). *The Strategy of Economic Development*. New Haven (CT): Yale University Press.

Hoekman, Jarno, Koen Frenken and Frank van Oort (Sept. 2009). 'The Geography of Collaborative Knowledge Production in Europe'. *The Annals of Regional Science* 43 (3), 721–738. DOI: 10.1007/s00168-008-0252-9.

Hoover, Edgar M. and Joseph L. Fisher (1949). 'Research in Regional Economic Growth'. In: *Problems in the Study of Economic Growth*. National Bureau of Economic Research. Chap. 5, 173–250.

Jacobs, Jane (1961). *The Death and Life of Great American Cities*. New York: Random House.

Johansson, Börje and Hans Lööf (2014). 'Innovation Strategies Combining Internal and External Knowledge'. In: *Routledge Handbook of the Economics of Knowledge*. Ed. by Cristiano Antonelli and Albert N. Link. London: Routledge. Chap. 4, 29–52.

Kutner, Michael H., Christopher J. Nachtsheim, John Neter and William Li (2005). *Applied Linear Regression Models*. 5th ed. McGraw-Hill.

Lee, Dongheon (June 2017). 'Industrial Variety and Structural Change in Korean Regional Manufacturing, 1992-2004'. *Growth and Change* 48 (2), 246–264. DOI: 10.1111/grow.12184.

Lengyel, Balázs and Izabella Szakálná Kanó (2013). 'Related Variety and Regional Growth in Hungary: Towards a Transition Economy Approach'. *Regional Statistics: journal of the Hungarian Central Statistical Office* 3 (1), 98–116.

Lucas, Robert E. (Mar. 1993). 'Making a Miracle'. *Econometrica* 61.2, 251. DOI: 10.2307/2951551.

Martin, Ron (Jan. 2012). 'Regional Economic Resilience, Hysteresis and Recessionary Shocks'. *Journal of Economic Geography* 12 (1), 1–32.

Martin, Ron and Peter Sunley (Jan. 2015). 'On the Notion of Regional Economic Resilience: Conceptualization and Explanation'. *Journal of Economic Geography* 15 (1), 1–42. DOI: 10.1093/jeg/lbu015.

Mealy, Penny, J. Doyne Farmer and Alexander Teytelboym (2017). 'A New Interpretation of the Economic Complexity Index'. URL: https://www.researchgate.net/publication/321228210_A_New_Interpretation_of_the_Economic_Complexity_Index.

Metcalfe, J. Stanley (2007). 'Alfred Marshall and the General Theory of Evolutionary Economics'. *History of Economic Ideas* 15.1, 81–110.

— (2010). 'Complexity and Emergence in Economics: The Road from Smith to Hayek (Via Marshall and Schumpeter)'. *History of Economic Ideas* 18.2, 45–75.

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

Montgomery, Cynthia A. (1979). 'Diversification, Market Structure, and Firm Performance: An Extension of Rumelt's Model'. PhD thesis. Purdue University.

Montgomery, Cynthia A. and S. Hariharan (Jan. 1991). 'Diversified Expansion by Large Established Firms'. *Journal of Economic Behavior & Organization* 15 (1), 71–89. DOI: 10.1016/0167-2681(91)90005-i.

Nesta, Lionel (Sept. 2008). 'Knowledge and Productivity in the World's Largest Manufacturing Corporations'. *Journal of Economic Behavior & Organization* 67.3-4, 886–902. DOI: 10.1016/j.jebo.2007.08.006.

Nesta, Lionel and Pier Paolo Saviotti (Mar. 2005). 'Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry'. *Journal of Industrial Economics* 53.1, 123–142. DOI: 10.1111/j.0022-1821.2005.00248.x.

— (Aug. 2006). 'Firm Knowledge and Market Value in Biotechnology'. *Industrial and Corporate Change* 15.4, 625–652. DOI: 10.1093/icc/dtl007.

Nourse, Hugh (1968). *Regional Economics*. New York: McGraw-Hill.

OECD (Mar. 2018). *REGPAT database*.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2018). *vegan: Community Ecology Package*. R package version 2.5-2. URL: https://CRAN.R-project.org/package=vegan.

Padrón, Benigno, Manuel Nogales and Anna Traveset (2011). 'Alternative Approaches of Transforming Bimodal into Unimodal Mutualistic Networks: The Usefulness of Preserving Weighted Information'. *Basic and Applied Ecology* 12, 713–721. DOI: 10.1016/j.baae.2011.09.004.

Panzar, John C. and Robert D. Willig (May 1981). 'Economies of Scope'. *The American Economic Review* 71.2, 268–272.

Parr, John B. (Nov. 1965). 'Specialization, Diversification and Regional Development'. *The Professional Geographer* 17 (6), 21–25. DOI: 10.1111/j.0033-0124.1965.021_h.x.

— (Apr. 2002). 'Missing Elements in the Analysis of Agglomeration Economies'. *International Regional Science Review* 25.2, 151–168. DOI: 10.1177/016001702762481221.

Pasinetti, Luigi Lodovico (1993). *Structural Economic Dynamics: A Theory of the Economic Consequences of Human Learning*. Cambridge: Cambridge University Press.

Patrucco, Pier Paolo (Feb. 2008). 'The economics of collective knowledge and technological communication'. *The Journal of Technology Transfer* 33.6, 579–599.

— (Apr. 2009). 'Collective knowledge production costs and the dynamics of technological systems'. *Economics of Innovation and New Technology* 18.3, 295–310.

Penrose, Edith Tilton (1959). *The Theory of the Growth of the Firm*. New York: John Wiley and Sons.

Pianta, Mario and Valentina Meliciani (1996). 'Technological Specialization and Economic Performance in OECD Countries'. *Technology Analysis & Strategic Management* 8.2, 157–174. DOI: 10.1080/09537329608524241.

Pielou, Evelyn Chrystalla (1969). *An Introduction to Mathematical Ecology*. Wiley.

— (1977). *Mathematical Ecology*. Wiley.

Pike, Andy, Stuart Dawley and John Tomaney (Mar. 2010). 'Resilience, Adaptation and Adaptability'. *Cambridge Journal of Regions, Economy and Society* 3 (1), 59–70. DOI: 10.1093/cjres/rsq001.

Quatraro, Francesco (Dec. 2010). 'Knowledge Coherence, Variety and Economic Growth: Manufacturing Evidence from Italian Regions'. *Research Policy* 39.10, 1289–1302. DOI: 10.1016/j.respol.2010.09.005.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rafols, Ismael (2014). ‘Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence’. In: *Measuring Scholarly Impact*. Ed. by Ying Ding, Ronald Rousseau and Wolfram Dietmar. Springer. Chap. 8, 169–190. DOI: 10.1007/978-3-319-10377-8_8.

Rafols, Ismael and Martin Meyer (2010). ‘Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience’. *Scientometrics* 82, 263–287. DOI: 10.1007/s11192-009-0041-y.

Ramanujam, Vasudevan and P. Varadarajan (Nov. 1989). ‘Research on Corporate Diversification: A Synthesis’. *Strategic Management Journal* 10.6, 523–551. DOI: 10.1002/smj.4250100603.

Reggiani, Aura, Thomas De Graaff and Peter Nijkamp (June 2002). ‘Resilience: An Evolutionary Approach to Spatial Economic Systems’. *Networks and Spatial Economics* 2 (2), 211–229. DOI: 10.1023/a:1015377515690.

Richardson, Harry W. (1969). *Regional Economics: Location Theory, Urban Structure, and Regional Change*. New York: Praeger.

Rocchetta, Silvia and Andrea Mina (Mar. 2019). ‘Technological Coherence and the Adaptive Resilience of Regional Economies’. *Regional Studies*, 1–14. DOI: 10.1080/00343404.2019.1577552.

Rodgers, Allan (1957). ‘Some Aspects of Industrial Diversification in the United States’. *Economic Geography* 33 (1), 16–30. DOI: 10.2307/142564.

Saviotti, Pier Paolo (1996). *Technological Evolution, Variety and the Economy*. Cheltenham (UK) and Brookfield (VT): Edward Elgar. ISBN: 1852787740.

Saviotti, Pier Paolo and Koen Frenken (Apr. 2008). ‘Export Variety and the Economic Performance of Countries’. *Journal of Evolutionary Economics* 18 (2), 201–218. DOI: 10.1007/s00191-007-0081-5.

Schumpeter, Joseph A. (Nov. 1947). ‘The Creative Response in Economic History’. *The Journal of Economic History* 7.2, 149–159.

— (1983). *The Theory of Economic Development. An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Trans. from the German by Redvers Opie. Transaction Books.

Serrano, M. Ángeles and Marián Boguñá (2005). ‘Weighted Configuration Model’. In: *Science of Complex Networks. From Biology to the Internet and WWW*. CNET 2004. Ed. by J. F. F. Mendes, S. N. Dorogovtsev, A. Povolotsky, F. V. Abreu and J. G. Oliveira. Vol. 776. AIP Conference Proceedings. American Institute of Physics, 101–107. ISBN: 0-7354-0262-0. DOI: 10.1063/1.1985381. URL: https://arxiv.org/abs/cond-mat/0501750 (visited on 23/07/2018).

Shannon, Claude E. (July 1948). ‘A Mathematical Theory of Communication’. *Bell System Technical Journal* 27.3, 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Sheather, Simon J. (2008). *A Modern Approach to Regression with R*. New York (NY): Springer. ISBN: 0387096078.

Shi, Jianbo and Jitendra Malik (Aug. 2000). ‘Normalized Cuts and Image Segmentation’. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8, 888–905. DOI: 10.1109/34.868688.

Simmie, James and Ron Martin (Mar. 2010). ‘The Economic Resilience of Regions: Towards an Evolutionary Approach’. *Cambridge Journal of Regions, Economy and Society* 3 (1), 27–43. DOI: 10.1093/cjres/rsp029.

Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Soete, Luc G. and Sally M. E. Wyatt (Jan. 1983). ‘The Use of Foreign Patenting as an Internationally Comparable Science and Technology Output Indicator’. *Scientometrics* 5 (1), 31–54. DOI: 10.1007/bf02097176.

Stirling, Andy (Aug. 2007). 'A General Framework for Analysing Diversity in Science, Technology and Society'. *Journal of The Royal Society Interface* 4 (15), 707–719. DOI: 10.1098/rsif.2007.0213.

Tacchella, Andrea, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli and Luciano Pietronero (2012). 'A New Metrics for Countries' Fitness and Products' Complexity'. *Scientific Reports* 2.723, 1–7. DOI: 10.1038/srep00723.

— (Aug. 2013). 'Economic Complexity: Conceptual Grounding of a New Metrics for Global Competitiveness'. *Journal of Economic Dynamics and Control* 37 (8), 1683–1691. DOI: 10.1016/j.jedc.2013.04.006.

Teece, David J. (Sept. 1980). 'Economies of Scope and the Scope of the Enterprise'. *Journal of Economic Behavior & Organization* 1 (3), 223–247. DOI: 10.1016/0167-2681(80)90002-5.

— (Mar. 1982). 'Towards an Economic Theory of the Multiproduct Firm'. *Journal of Economic Behavior & Organization* 3 (1), 39–63. DOI: 10.1016/0167-2681(82)90003-8.

Teece, David J., Richard Rumelt, Giovanni Dosi and Sidney G. Winter (Jan. 1994). 'Understanding Corporate Coherence: Theory and Evidence'. *Journal of Economic Behavior & Organization* 23 (1), 1–30. DOI: 10.1016/0167-2681(94)90094-9.

Theil, Henri (1967). *Economics and Information Theory*. Amsterdam: North-Holland.

— (1972). *Statistical Decomposition Analysis*. Amsterdam: North-Holland.

Weber, Christian E. (June 2005). 'Edgeworth on Complementarity, or Edgeworth, Auspitz-Lieben, and Pareto De-Homogenized'. *History of Political Economy* 37 (2), 293–307. DOI: 10.1215/00182702-37-2-293.

Weitzman, Martin L. (May 1996). 'Hybridizing Growth Theory'. *The American Economic Review* 86.2, 207–212.

— (May 1998). 'Recombinant Growth'. *The Quarterly Journal of Economics* 113 (2), 331–360. DOI: 10.1162/003355398555595.

Willig, Robert D. (May 1979). 'Multiproduct Technology and Market Structure'. *The American Economic Review* 69.2, 346–351.

Wintjes, René and Hugo Hollanders (2010). *The Regional Impact of Technological Change in 2020*. Research rep. European Commission.

Zadjenweber (1972). 'Une Application de la Théorie de l'Information à l'Économie: La Mesure de la Concentration'. *Revue d'Economie Politique* 82, 486–510.

Zajdenweber, Daniel (1972). 'Une Application de la Théorie de l'Information à l'Économie: La Mesure de la Concentration'. *Revue d'Économie Politique* 82.3, 486–510.

Zhou, Tao, Jie Ren, Matú Medo and Yi-Cheng Zhang (Oct. 2007). 'Bipartite Network Projection and Personal Recommendation'. *Physical Review E* 76 (4), 046115. DOI: 10.1103/PhysRevE.76.046115.

A The ReKS R Package functions

In the following I provides the code in R of the functions useful to reproduce the results of the paper (R Core Team 2018). They require the R Packages *Matrix* and *vegan* to work (Bates and Maechler 2018; Oksanen et al. 2018).

Information Entropy The function will return the information entropy of a given vector that reports the absolute frequency of each of the possible types/groups of observations of the database. See Frenken 2007; Shannon 1948; Theil 1967, 1972.

```
entropy <- function(data) {
    freqs <- .get_freqs(data)
    etp <- -sum(freqs * log2(freqs))
    etp[!is.finite(etp)] <- 0
    return(etp)
}
```

Related and Unrelated Variety The two following functions return the information entropy of a given vector of frequencies decomposed in two parts: a between-groups and a within-groups one. Moreover, it provides you also the probability of each group and the entropy of each of the groups. See Attaran 1986; Boschma and Iammarino 2009; Content and Frenken 2016; Frenken 2007; Frenken et al. 2007; Quatraro 2010; Rocchetta and Mina 2019; Theil 1972; Zadjenweber 1972.

```
entropy_decomposition <- function(data, groups) {
    Pg <- by(.get_freqs(data), groups, sum)
    BG <- entropy(Pg)
    WG <- entropy(data) - BG
    by_group <- log2(Pg) + 1/Pg * by(data, groups, entropy)
    etp_dcp <- list(BG = BG,
                    WG = WG,
                    by_group = by_group,
                    Pg = Pg)
    return(etp_dcp)
}
```

```
entropy_decomposition_panel <- function(data, kng_nbr, kng_dim_upper,
                                        geo_dim, time_dim) {

    # Preliminary transformations and checks ---------------
    data <- as.data.frame(data)
    if (!all(complete.cases(data))) {
        warning(paste('There is some non complete row in the database.\n',
                      'I cannot guarrenty you about the results.'))
    }
    kng_nbr <- deparse(substitute(kng_nbr))
    geo_dim <- deparse(substitute(geo_dim))
    kng_dim_upper <- deparse(substitute(kng_dim_upper))
```

```
    time_dim <- deparse(substitute(time_dim))

    # Decomposed entropy ---------------
    dd <- split(data[, kng_nbr],
                list(data[, time_dim], data[, geo_dim]))
    ddnt <- sapply(names(dd), function(s) strsplit(s, "[.]")[[1]][1])
    ddng <- sapply(names(dd), function(s) strsplit(s, "[.]")[[1]][2])
    obs_list <- 1:length(dd)
    entropy_total <- sapply(obs_list,
                            function(x) entropy(dd[[x]]))
    entropy_total <- cbind.data.frame(ddnt, ddng, entropy_total)
    colnames(entropy_total) <- c(time_dim, geo_dim, "entropy.total")
    grps <- split(data[, kng_dim_upper],
                  list(data[, time_dim], data[, geo_dim]))
    entropy_decomposed <- sapply(obs_list,
                                 function(x)
                                     entropy_decomposition(dd[[x]],
                                                           grps[[x]]))
    entropy_decomposed <- matrix(unlist(entropy_decomposed[1:2,]),
                                 ncol = 2, byrow = T)
    entropy_decomposed <- cbind.data.frame(ddnt, ddng, entropy_decomposed)
    colnames(entropy_decomposed) <- c(time_dim, geo_dim,
                                      "entropy.between", "entropy.within")
    entropy <- merge(entropy_total, entropy_decomposed)
    tl <- levels(entropy[, time_dim])
    tn <- entropy[, time_dim]
    entropy[, time_dim] <- as.numeric(tl)[tn]

    measure <- c("entropy.total", "entropy.between", "entropy.within")

    class(entropy) <- c("reks_entropy", "data.frame")
    attr(entropy, 'geo_dim') <- geo_dim
    attr(entropy, 'kng_dim_upper') <- kng_dim_upper
    attr(entropy, 'time_dim') <- time_dim
    attr(entropy, 'measure') <- measure

    return(entropy)
}
```

Related and Unrelated Evenness

```
evenness_decomposition <- function (data, groups) {
    Pg <- by(ReKS:::.get_freqs(data), groups, sum)
    BG <- entropy(Pg)
    WG <- entropy(data) - BG
    BG <- BG / log2(length(Pg))
    WG <- WG / log2(length(data))
    etp_dcp <- list(BG = BG, WG = WG, Pg = Pg)
    return(etp_dcp)
```

```
}
```

```
evenness_decomposition_panel <- function (data, kng_nbr, kng_dim_upper,
                                          geo_dim, time_dim) {
    data <- as.data.frame(data)
    if (!all(complete.cases(data))) {
        warning(paste("There is some non complete row in the database.\n",
                      "I cannot guarrenty you about the results."))
    }
    kng_nbr <- deparse(substitute(kng_nbr))
    geo_dim <- deparse(substitute(geo_dim))
    kng_dim_upper <- deparse(substitute(kng_dim_upper))
    time_dim <- deparse(substitute(time_dim))
    dd <- split(data[, kng_nbr], list(data[, time_dim], data[, geo_dim]))
    ddnt <- sapply(names(dd), function(s) strsplit(s, "[.]")[[1]][1])
    ddng <- sapply(names(dd), function(s) strsplit(s, "[.]")[[1]][2])
    obs_list <- 1:length(dd)
    evenness_total <- sapply(obs_list, function(x) {
        entropy(dd[[x]]) / log2(length(dd[[x]]))
    })
    evenness_total <- cbind.data.frame(ddnt, ddng, evenness_total)
    colnames(evenness_total) <- c(time_dim, geo_dim, "evenness.total")
    grps <- split(data[, kng_dim_upper], list(data[, time_dim],
                                              data[, geo_dim]))
    evenness_decomposed <- sapply(obs_list,
                              function(x) evenness_decomposition(dd[[x]],
                                                                 grps[[x]]))
    evenness_decomposed <- matrix(unlist(evenness_decomposed[1:2, ]),
                              ncol = 2, byrow = T)
    evenness_decomposed <- cbind.data.frame(ddnt, ddng, evenness_decomposed)
    colnames(evenness_decomposed) <- c(time_dim, geo_dim, "evenness.between",
                                       "evenness.within")
    evenness <- merge(evenness_total, evenness_decomposed)
    tl <- levels(evenness[, time_dim])
    tn <- evenness[, time_dim]
    evenness[, time_dim] <- as.numeric(tl)[tn]
    measure <- c("evenness.total", "evenness.between", "evenness.within")
    return(evenness)
}
```

Regional Coherence Index The function computes the so called Coherence index. See Bottazzi and Pirino 2010; Nesta and Saviotti 2005, 2006; Quatraro 2010; Rocchetta and Mina 2019; Teece et al. 1994.

```
    coherence <- function(occurrence_mtx, relatedness_mtx) {
        if (!requireNamespace("Matrix", quietly = TRUE))
            stop(paste0('Package \"Matrix\" needed for this function to work. ',
```

```
                    'Please install it.'), call. = FALSE)

    geo_dim <- attr(occurrence_mtx, 'geo_dim')
    kng_dim <- attr(occurrence_mtx, 'kng_dim')
    if (is.list(occurrence_mtx))
        time_dim <- attr(occurrence_mtx, 'time_dim')
    measure <- "Coherence"

    coherence_crossSection <- function(occurrence_mtx, relatedness_mtx) {

        # Preliminary operations, checks and transformations ----------------
        oc_mtx_names <- colnames(occurrence_mtx)
        rl_mtx_names <- colnames(relatedness_mtx)
        if (dim(occurrence_mtx)[[2]] != sum(dim(relatedness_mtx)) / 2) {
            names_tbr <- setdiff(rl_mtx_names, oc_mtx_names)
            if (length(names_tbr) != 0)
                relatedness_mtx <- relatedness_mtx[
                    -which(rownames(relatedness_mtx) %in% names_tbr),
                    -which(colnames(relatedness_mtx) %in% names_tbr)]
        }
        if (dim(occurrence_mtx)[[2]] != sum(dim(relatedness_mtx)) / 2)
            stop(paste('There is some problem, because the two matrices',
                       'considered have a different number of',
                       'columns.'), call. = FALSE)
        rl_mtx_names <- colnames(relatedness_mtx)
        if (any(oc_mtx_names != rl_mtx_names)) {
            oc_mtx_names <- oc_mtx_names[, order(colnames(oc_mtx_names))]
            relatedness_mtx <- relatedness_mtx[order(rownames(relatedness_mtx)),
                                               order(colnames(relatedness_mtx))]
        }
        if (any(oc_mtx_names != rl_mtx_names))
            stop(paste('There is some problem, because the there is no perfect',
                       'correspondence between the column names of the two',
                       'matrices considered.'), call. = FALSE)
        ones <- !Matrix::diag(TRUE,
                              nrow = nrow(relatedness_mtx),
                              ncol = nrow(relatedness_mtx))

        # Waighted Average Relatedness
        WAR_num <- Matrix::tcrossprod(occurrence_mtx, relatedness_mtx)
        WAR_den <- Matrix::tcrossprod(occurrence_mtx, ones)
        WAR <- WAR_num / WAR_den

        # Coherence
        C_num <- WAR * occurrence_mtx
        C_den <- Matrix::rowSums(occurrence_mtx)
        C <- Matrix::rowSums(C_num / C_den)
        C[which(is.nan(C))] <- 0
```

```
        C <- cbind.data.frame(names(C), unlist(C))
        colnames(C) <- c(geo_dim, measure)

        return(C)
    }

    coherence_panel <- function(occurrence_mtx, relatedness_mtx) {
        time_span <- names(occurrence_mtx)
        C <- lapply(as.character(time_span),
                    function(y) coherence_crossSection(occurrence_mtx[[y]],
                                                       relatedness_mtx))
        yrs <- unlist(mapply(rep, time_span, lapply(C, nrow)))
        C <- do.call("rbind", C)
        C <- cbind.data.frame(C, yrs)
        C <- C[, c(1, 3, 2)]
        colnames(C) <- c(geo_dim, time_dim, measure)

        return(C)
    }

    fntn <- ifelse(is.list(occurrence_mtx),
                   "coherence_panel",
                   "coherence_crossSection")
    C <- do.call(fntn, list(occurrence_mtx, relatedness_mtx))

    # final steps
    # class(R) <- c("reks_coherence", "data.frame")
    attr(C, 'geo_dim') <- geo_dim
    attr(C, 'kng_dim') <- kng_dim
    if (is.list(occurrence_mtx))
        attr(C, 'time_dim') <- time_dim
    attr(C, 'measure') <- measure

    return(C)
}
```

Regional Complexity Index The function computes Regional Knowledge Complexity Index *a là* Hidalgo-Hausmann in each given year. See Antonelli et al. 2017; Balland and Rigby 2017; Hidalgo and Hausmann 2009.

```
complexity_hh <- function(occurrence_mtx,
                          rta = TRUE, binary = TRUE, scale = TRUE) {
    if (!requireNamespace("Matrix", quietly = TRUE))
        stop(paste0('Package \"Matrix\" needed for this function to work. ',
                    'Please install it.'), call. = FALSE)

    geo_dim <- attr(occurrence_mtx, 'geo_dim')
    kng_dim <- attr(occurrence_mtx, 'kng_dim')
    if (is.list(occurrence_mtx))
```

```
    time_dim <- attr(occurrence_mtx, 'time_dim')
measure <- "Complexity"

complexity_hh_crossSection <- function(occurrence_mtx) {
    if (any(Matrix::rowSums(occurrence_mtx) == 0)) {
        occurrence_mtx <- occurrence_mtx[-Matrix::which(
            Matrix::rowSums(occurrence_mtx) == 0), ]
    }
    if (any(Matrix::colSums(occurrence_mtx) == 0)) {
        occurrence_mtx <- occurrence_mtx[, -Matrix::which(
            Matrix::colSums(occurrence_mtx) == 0)]
    }
    rnms <- rownames(occurrence_mtx)
    if (isTRUE(rta))
        occurrence_mtx <- rta(occurrence_mtx, binary = binary)
    if (isTRUE(binary))
        occurrence_mtx <- Matrix::Matrix(ifelse(occurrence_mtx > 0, 1, 0),
                                         nrow = nrow(occurrence_mtx))
    du <- ReKS:::.get_du(Matrix::as.matrix(occurrence_mtx))

    mm_tilde <- Matrix::t(Matrix::t(occurrence_mtx) / du$ubiquity)
    mm_tilde <- Matrix::tcrossprod(mm_tilde, occurrence_mtx)
    mm_tilde <- mm_tilde / du$diversification

    if (!all(round(Matrix::rowSums(mm_tilde)) == 1)) {
        stop(paste("The matrix is not row-stochastic.\n",
                   "It is not possible to compute the measure."))
    }
    if (round(Re(as.complex(eigen(mm_tilde)$value[1]))) != 1) {
        stop(paste("The first eigen-value is different from 1.",
                   "It is not possible to compute the measure."))
    }

    RKCI <- eigen(mm_tilde)$vectors
    if (dim(RKCI)[2] >= 2) {
        RKCI <- RKCI[, 2]
        RKCI <- Re(as.complex(RKCI))
        if (cor(RKCI, du$diversification,
                use = "pairwise.complete.obs", method = "spearman") < 0) {
            RKCI <- -RKCI
        }
    } else {
        RKCI <- NA
    }
    RKCI <- cbind.data.frame(rnms,
                             RKCI)
    colnames(RKCI) <- c(geo_dim, measure)
    if (scale == TRUE) {
        RKCI[, measure] <- scale(as.numeric(RKCI[, measure]))
```

```
            # warning('The values of the index have been standardised.')
        }

        return(RKCI)
    }

    complexity_hh_panel <- function(occurrence_mtx) {
        time_span <- names(occurrence_mtx)
        RKCI <- lapply(as.character(time_span),
                       function(y) complexity_hh_crossSection(occurrence_mtx[[y]]))
        yrs <- unlist(mapply(rep, time_span, lapply(RKCI, nrow)))
        RKCI <- do.call("rbind", RKCI)
        RKCI <- cbind.data.frame(RKCI, yrs)
        RKCI <- RKCI[, c(1, 3, 2)]
        colnames(RKCI) <- c(geo_dim, time_dim, measure)

        return(RKCI)
    }

    fntn <- ifelse(is.list(occurrence_mtx),
                   "complexity_hh_panel",
                   "complexity_hh_crossSection")
    Cx <- do.call(fntn, list(occurrence_mtx))

    # final steps
    # class(R) <- c("reks_complexity_hh", "data.frame")
    attr(Cx, 'geo_dim') <- geo_dim
    attr(Cx, 'kng_dim') <- kng_dim
    if (is.list(occurrence_mtx))
        attr(Cx, 'time_dim') <- time_dim
    attr(Cx, 'measure') <- measure
    # attr(Cx, 'diversity') <- du$diversification
    # attr(Cx, 'ubiquity') <- du$ubiquity
    attr(Cx, 'standardised') <- scale
    attr(Cx, "RTA") <- rta
    attr(Cx, "binary") <- binary

    return(Cx)
}
```

Regional Complexity Index The function computes the Regional Knowledge Fitness Index *a là* Tacchella, Cristelli, Caldarelli, Gabrielli and Pietronero (i.e. *competitiveness*) of each given geographical area considered, in each year provided. See Cristelli et al. 2013; Tacchella et al. 2012, 2013.

```
fitness_tccgp <- function(occurrence_mtx,
                          rta = TRUE, binary = TRUE, scale = FALSE) {
    if (!requireNamespace("Matrix", quietly = TRUE))
        stop(paste0('Package \"Matrix\" needed for this function to work. ',
```

```r
                'Please install it.'), call. = FALSE)

    geo_dim <- attr(occurrence_mtx, 'geo_dim')
    kng_dim <- attr(occurrence_mtx, 'kng_dim')
    if (is.list(occurrence_mtx))
        time_dim <- attr(occurrence_mtx, 'time_dim')
    measure <- "Fitness"

    fitness_tccgp_crossSection <- function(occurrence_mtx) {
        if (any(Matrix::rowSums(occurrence_mtx) == 0)) {
            occurrence_mtx <- occurrence_mtx[-Matrix::which(
                Matrix::rowSums(occurrence_mtx) == 0), ]
        }
        if (any(Matrix::colSums(occurrence_mtx) == 0)) {
            occurrence_mtx <- occurrence_mtx[, -Matrix::which(
                Matrix::colSums(occurrence_mtx) == 0)]
        }
        rnms <- rownames(occurrence_mtx)
        if (isTRUE(rta))
            occurrence_mtx <- rta(occurrence_mtx, binary = binary)
        if (isTRUE(binary))
            occurrence_mtx <- Matrix::Matrix(ifelse(occurrence_mtx > 0, 1, 0),
                                             nrow = nrow(occurrence_mtx))

        # This is not needed for the algorithm, but still it can be useful to hav
        #  this information stored for future purpouses.
        du <- ReKS:::.get_du(Matrix::as.matrix(occurrence_mtx))

        RKFI <- as(rep(1, nrow(occurrence_mtx)), "sparseVector")
        RKCI <- as(rep(1, ncol(occurrence_mtx)), "sparseVector")
        i <- 0
        while (TRUE) {
            RKFI1 <- Matrix::t(occurrence_mtx) * RKCI
            RKFI1 <- Matrix::rowSums(Matrix::t(RKFI1))
            RKCI1 <- 1 / Matrix::rowSums(Matrix::t(occurrence_mtx / RKFI))
            # Normalisation needed to avoid possible divergences
            #  due to the hyperbolic nature of the second equation
            RKFI1 <- RKFI1 / mean(RKFI1)
            RKCI1 <- RKCI1 / mean(RKCI1)
            if (all((RKFI - RKFI1) < 0.0000000001) &
                all((RKCI - RKCI1) < 0.0000000001)) {
                RKFI <- RKFI1
                RKCI <- RKCI1
                convergence <- TRUE
                break()
            }
            if (i >= 200) {
                RKFI <- rep(as.numeric(NA), nrow(occurrence_mtx))
                RKCI <- rep(as.numeric(NA), ncol(occurrence_mtx))
```

```
            names(RKFI) <- names(RKFI1)
            names(RKCI) <- names(RKCI1)
            convergence <- FALSE
            warning(paste0('The algorithm failed to converge.\n',
                           'Maybe your matrix is not triangular ',
                           'as expected.\nYou can check it using ',
                           'image(occurrence_mtx, useAbs = FALSE)'))
            break()
        }
        RKFI <- RKFI1
        RKCI <- RKCI1
        i <- i + 1
    }
    RKFI <- cbind.data.frame(rnms,
                             RKFI)
    colnames(RKFI) <- c(geo_dim, measure)
    if (scale == TRUE) {
        RKFI[, measure] <- scale(as.numeric(RKFI[, measure]))
        # warning('The values of the index have been standardised.')
    }
    gc()

    attr(RKFI, "iterations") <- i
    attr(RKFI, "convergence") <- convergence
    attr(RKFI, 'diversification') <- du$diversification
    attr(RKFI, 'ubiquity') <- du$ubiquity

    return(RKFI)
}

fitness_tccgp_panel <- function(occurrence_mtx) {
    time_span <- names(occurrence_mtx)
    RKFI <- lapply(as.character(time_span),
                   function(y) {
                       Fx <- fitness_tccgp_crossSection(occurrence_mtx[[y]])
                       iterations <- attr(Fx, "iterations")
                       convergence <- attr(Fx, "convergence")
                       return(list(Fx,
                                   iterations, convergence))
                                   # diversification, ubiquity))
                   })
    iterations <- sapply(RKFI, "[", 2)
    names(iterations) <- time_span
    iterations <- do.call("rbind", iterations)
    convergence <- sapply(RKFI, "[", 3)
    names(convergence) <- time_span
    convergence <- do.call("rbind", convergence)

    RKFI <- sapply(RKFI, "[", 1)
```

```
        yrs <- unlist(mapply(rep, time_span, lapply(RKFI, nrow)))
        RKFI <- do.call("rbind.data.frame", RKFI)
        RKFI <- cbind.data.frame(RKFI, yrs)
        RKFI <- RKFI[, c(1, 3, 2)]
        colnames(RKFI) <- c(geo_dim, time_dim, measure)

        attr(RKFI, "iterations") <- iterations
        attr(RKFI, "convergence") <- convergence

        return(RKFI)
    }
    fntn <- ifelse(is.list(occurrence_mtx),
                   "fitness_tccgp_panel",
                   "fitness_tccgp_crossSection")
    Fx <- do.call(fntn, list(occurrence_mtx))

    # final steps
    attr(Fx, 'geo_dim') <- geo_dim
    attr(Fx, 'kng_dim') <- kng_dim
    if (is.list(occurrence_mtx))
        attr(Fx, 'time_dim') <- time_dim
    attr(Fx, 'measure') <- measure
    attr(Fx, 'standardised') <- scale
    attr(Fx, "RTA") <- rta
    attr(Fx, "binary") <- binary
    # attr(Fx, "iterations") <- i
    # attr(Fx, "convergence") <- convergence

    return(Fx)
}
```

Other functions The followings are other functions used internally by the previous ones, or useful to compute the objects used by them as inputs.

```
.get_du <- function(biadj_matrix) {
    du <- list()
    du$diversification <- rowSums(biadj_matrix)
    du$ubiquity   <- colSums(biadj_matrix)
    return(du)
}
.get_freqs <- function(data) {
    if (sum(as.numeric(data))==1) {
        warning('I assume you provided me a list of relative frequences')
        return(data)
    } else {
        warning(paste('I assume you provided me a list of absolute values.\n',
                      'I internally transformed them in relative frequencies.\n',
                      'Otherwise check in the original data',
                      ' why their sum is not 1.'))
```

```
        freqs <- data/sum(as.numeric(data))
    }
    return(freqs)
}
occurrence_matrix <- function(data, geo_dim, kng_dim,
                              kng_nbr = NULL,
                              time_dim = NULL,
                              binary_mode = "none") {
    if (!requireNamespace("Matrix", quietly = TRUE)) {
        stop(paste0("Package \"Matrix\" needed for this function to work. ",
                    "Please install it."), call. = FALSE)
    }
    # Preliminary controls ----------------
    geo_dim <- deparse(substitute(geo_dim))
    kng_dim <- deparse(substitute(kng_dim))
    kng_nbr <- deparse(substitute(kng_nbr))
    time_dim <- deparse(substitute(time_dim))
    if (kng_nbr == "NULL")
        kng_nbr <- NULL
    if (time_dim == "NULL")
        time_dim <- NULL
    if (binary_mode != 'simple' & is.null(kng_nbr)) {
        stop(paste0('Either you specify that you want a \"simple\" matrix, ',
                    'or you have to specify a column for the number of ',
                    'pieces of knowledge'))
    }
    data <- as.data.frame(data)
    if (is.null(kng_nbr)) {
        data <- unique(data[, c(geo_dim, kng_dim, time_dim)])
    } else {
        if (is.null(time_dim)) {
            if (anyDuplicated(data[, c(geo_dim, kng_dim)])) {
                frml <- formula(paste(kng_nbr, "~",
                                      geo_dim, "+", kng_dim))
                data <- aggregate(formula = frml, data = data, FUN = sum)
                warning(paste('Since there are duplicated cases, ',
                              'the function has collapsed them, ',
                              'by summing the number of pieces of knowledge'),
                        call. = F)
            }
        } else {
            if (anyDuplicated(data[, c(geo_dim, kng_dim, time_dim)])) {
                frml <- formula(paste(kng_nbr, "~",
                                      geo_dim, "+", kng_dim, "+", time_dim))
                data <- aggregate(formula = frml, data = data, FUN = sum)
                warning(paste('Since there are duplicated cases, ',
                              'the function has collapsed them, ',
                              'by summing the number of pieces of knowledge'),
                        call. = F)
```

```r
            }
        }
    }
    get_mtx <- function(data, frml, binary_mode) {
        bm <- xtabs(formula = formula(frml),
                    data = data,
                    sparse = TRUE)
        bm <- switch(binary_mode,
                     RTA = rta(bm, binary = TRUE),
                     RCA = rta(bm, binary = TRUE),
                     simple = as(bm, "ngCMatrix"),
                     none = bm)
        return(bm)
    }

    # Main function -----------------
    frml <- ifelse(is.null(kng_nbr),
                   paste("~", geo_dim, "+", kng_dim),
                   paste(kng_nbr, "~", geo_dim, "+", kng_dim))
    if (is.null(time_dim))
        BM <- get_mtx(data, frml, binary_mode)
    else {
        time_span <- unique(data[, time_dim])
        BM <- lapply(time_span, function(y)
            get_mtx(data[which(data[, time_dim] == y), ], frml, binary_mode))
        names(BM) <- time_span
    }

    # Closing operations
    attr(BM, "geo_dim") <- geo_dim
    attr(BM, "kng_dim") <- kng_dim
    if (!is.null(time_dim))
        attr(BM, "time_dim") <- time_dim
    attr(BM, "binary")  <- ifelse(binary_mode == "none", FALSE, TRUE)
    if (binary_mode == 'RTA')
        attr(BM, "binary_mode") <- 'RTA'
    if (binary_mode == 'RCA')
        attr(BM, "binary_mode") <- 'RCA'
    if (binary_mode == 'simple')
        attr(BM, "binary_mode") <- 'simple'

    return(BM)
}
relatedness <- function(adj_mtx, output_statistic = "t",
                        is_binary = NULL,
                        fixedmar = "both", seed = Sys.time(), nSim = 1000) {
    if (!requireNamespace("Matrix", quietly = TRUE)) {
        stop(paste0("Package \"Matrix\" needed for this function to work. ",
                    "Please install it."), call. = FALSE)
```

```r
}

# Preliminary operations
adj_mtx <- as(adj_mtx, "Matrix")
geo_dim <- attr(adj_mtx, 'geo_dim')
kng_dim <- attr(adj_mtx, 'kng_dim')

# t-stat
relatedness_t <- function(...) {
    rnms <- rownames(adj_mtx)
    cnms <- colnames(adj_mtx)
    adj_mtx[Matrix::which(adj_mtx > 0)] <- 1
    if (any(adj_mtx@x != 1))
        stop(paste("It is not possible to transform the matrix into ",
                   "a binary (0/1) one"))
    Nr <- nrow(adj_mtx)
    J <- Matrix::crossprod(adj_mtx)
    Matrix::diag(J) <- 0
    n <- Matrix::colSums(adj_mtx)
    mu <- Matrix::tcrossprod(n)
    mu <- mu / Nr
    s2 <- Matrix::tcrossprod((1 - (n / Nr)), ((Nr - n) / (Nr - 1)))
    s2 <- mu * s2
    t <- (J - mu) / sqrt(s2)
    Matrix::diag(t) <- 0
    rownames(t) <- colnames(t) <- cnms

    return(t)
}

# p-value
relatedness_p <- function(...) {
    if (!requireNamespace("vegan", quietly = TRUE)) {
        stop(paste0("Package \"vegan\" needed for this function to work. ",
                   "Please install it."), call. = FALSE)
    }
    rnms <- rownames(adj_mtx)
    cnms <- colnames(adj_mtx)
    isBinary <- ifelse((is.null(is_binary) &&
                            all(adj_mtx@x %in% c(0, 1, FALSE, TRUE))) ||
                          isTRUE(is_binary), TRUE, FALSE)
    if (isBinary)
        adj_mtx <- as(adj_mtx, "ngCMatrix")
    set.seed(seed)
    adj_mtx_null_models <- vegan::permatswap(adj_mtx,
                                             fixedmar = fixedmar,
                                             mtype = ifelse(isBinary,
                                                            "prab",
                                                            "count"),
```

```
                                                    times = nSim)
        J_hat <- Matrix::crossprod(adj_mtx)
        J <- lapply(adj_mtx_null_models$perm,
                    function(m) Matrix::crossprod(as(m, class(adj_mtx)[[1]])))
        p <- lapply(J, function(m) J_hat >= m)
        p <- Reduce("+", p)
        p <- p / nSim
        Matrix::diag(p) <- 0
        pPlus <- pmax(as.vector(2 * p - 1), rep(0, nrow(p) * ncol(p)))
        pPlus <- Matrix(pPlus, nrow = nrow(p), ncol = ncol(p))
        pMinus <- pmin(as.vector(2 * p - 1), rep(0, nrow(p) * ncol(p))) * (-1)
        pMinus <- Matrix(pMinus, nrow = nrow(p), ncol = ncol(p))
        rownames(p) <- colnames(p) <- cnms
        rownames(pPlus) <- colnames(pPlus) <- cnms
        rownames(pMinus) <- colnames(pMinus) <- cnms

        return(list(p = p, pPlus = pPlus, pMinus = pMinus))
    }

    R <- switch(output_statistic,
                t = relatedness_t(),
                p = relatedness_p(),
                stop('\"output_statistic\" can be one of \"t\", or \"p\"'))

    attr(R, output_statistic) <- output_statistic
    attr(R, 'geo_dim') <- geo_dim
    attr(R, 'kng_dim') <- kng_dim

    return(R)
}
rta <- function(data, binary = FALSE) {
    if (!requireNamespace("Matrix", quietly = TRUE)) {
        stop(paste0("Package \"Matrix\" needed for this function to work. ",
                    "Please install it."), call. = FALSE)
    }
    RA <- Matrix::t(
        Matrix::t(
            data / Matrix::rowSums(data)) /
            (Matrix::colSums(data) / sum(data)))
    if (isTRUE(binary)) {
        RA <- as(RA, "ngCMatrix")
    }

    return(RA)
}
```

Cluster Heterogeneity and Cluster Evolution
The *Business Community* of Turin, 1883–1907 as a Case of Cluster Transformation

Carlo Bottai

Department of Economics and Statistics *Cognetti de Martiis*, University of Turin
BRICK, Collegio Carlo Alberto

Abstract

The aim of the paper is to verify whether or not, in consequence of the 1889 economic shock, Turin was characterised by a radical restructuration of the existing economic backbone. In line with some recent theoretical contribution on *cluster life-cycle*, the evolution of the city's economy – framed as a Complex Adaptive System– is supposed to be driven by its capabilities *diversity* dynamic, and the *fictional expectations* of the entrepreneurs are seen as the engine that guides this last process of change in both directions. The analysis suggests that the 1890s downturn increased the manoeuvring space available to the entrepreneurs for a reorganisation of the sectoral structure of the city, from a textile-based to a mechanics-centric economy.

Keywords— Urban Economic History, Clusters evolution, Development and changes, Entrepreneurship, Uncertainty, Imagination

1 Introduction

In what follows, I will try to explore the bidirectional nexus between the structure of the relationship among the (relevant) actors of a given (localised) economic system and its techno-structural texture. Specifically, I will analyse the radical sectoral restructuring of the Turin cluster at the end of the Nineteenth century as a case of cluster *transformation*, or *reinvention*, that led the city –in consequence of a financial shock– to move from a mainly textile-based economy to an automotive-centred one.[1]

In particular, in line with a recent contribution by Denney et al. (2018), I will frame a theory that primarily rests upon Feldman et al. (2005), Martin and Sunley (2011), and Menzel and Fornahl (2010). In this framework, the evolution of a *cluster*, framed as a Complex Adaptive System is driven by its capabilities *diversity* dynamic, and the entrepreneurial activity is seen as the engine of this last dynamic.

Specifically, following Menzel and Fornahl (2010), I will sketch a model in which the cluster evolution is driven by its level of heterogeneity. Moreover, I will also add that, over periods of high uncertainty, one of the main entrepreneurial functions is to discover the specific cost structure of the local economy (Hausmann and Rodrik 2003). In so doing, *fictional expectations* about the future (Beckert 2016) are both a fundamental tool that guide businessmen in exploring and expanding the possibility space of economic systems, and a decentralised mechanism that facilitates the reconciliation of a business cluster towards a new thematic core, helping to restrain the level of heterogeneity after the *exploration* phase (March 1991). In other words, it is possible to consider *fictional expectations* of Schumpeterian-Knightian entrepreneurs (Knight 1921; Schumpeter 1939) as an important source of the cycles in the level of *diversity*, in its waxing, as well as waning phase.

Lastly, I will test the hypothesis about the role of the heterogeneity of the cluster on its evolution on data about the economy of Turin at the end of the Nineteenth century. Indeed, the economic structure of Turin could be seen as a case in which industrial agglomerations played an opposite role in two different historical contexts. During the so-called Italian take-off of the first decade of the Twentieth century, the city was able to catch the new technological opportunities, moving from a textile-based economy to a more modern engineering-based one. Conversely, a hundred years later, the city was not able to get out unhurt from the "automotive-centric" *technological paradigm* (Dosi 1982) that had ensured it a substantial and lasting economic prosperity. This, by the way, shows that the independence of the technological and the cluster life cycles, and so proving the importance of a cluster evolution theory based on a supernaturalistic metaphor able to account for the renewal and revival capacity of such systems (Martin and Sunley 2011).

The analysis shows that the sectoral structuration of the local *business community* was weaker in the period that followed the financial crisis of 1889, and

[1] Menzel and Fornahl (2010, p. 218–219 and fig. 4) identify three types of *cluster revival*. They call *adaptation* a movement that involves an upgrade of the knowledge base through knowledge inflows coming from the outside of the cluster's borders. Instead, we will have a *renewal* when the revival happens through the integration of different pieces of the local knowledge base. Lastly, a diversification into new activities that builds on the local knowledge base is named *transformation*.

that the opposite happened about a decade later, in overlap with the birth and steep growth of the automotive industry in the city. In other words, the average sectoral *diversity* within the clusters of firms significantly increased in the recessionary period, compared to the pre- and post-crisis levels, because the exploration opportunity costs, as well as the possibility to successfully introduce new narratives about the future of the business cluster, are lower in period of high ontological uncertainty.

The paper is organised as follows. Sec. 2 will frame the theory, trying to reconcile the different streams of literature here briefly presented. Then, in Sec. 3 the economic situation of Turin at the end of the Nineteenth century is sketched, following the key points of the interpretation provided by Balbo (2007). In so doing, the links between the historical case and the theory proposed in the previous section are highlighted. Lastly, in Sec. 4 I propose the main research hypothesis of the paper, I present the empirical approach and the data, and I discuss the main findings of this contribution.

2 A model of cluster evolution and cluster renewal

In line with the arguments by Penrose (1952) about the several limitations a life-cycle model has in describing the development of firms, the classical *cluster life-cycle* model has been recently questioned. In particular, Martin and Sunley (2011) have stated this biological metaphor as too simplistic. The idea of an evolutionary path, from a *birth* to a *decline*, or even a *death*, is in contrast with many empirical cases. Indeed, a cluster –that is not an object *per se*, but only a useful concept (Menzel and Fornahl 2010)– can *resurrect* or *reinvent* itself, renewing some of its constituent elements (see e.g., Denney et al. 2018; Grabher 1993; Klepper 2007; Tappi 2005).

This is not to say that we cannot use an evolutionary perspective on the development of (urban) economic systems (Boschma and Frenken 2006, 2018; Boschma and Lambooy 1999; Boschma and Martin 2007, 2010).[2] What is questioned is the use of strong biological analogies and metaphors in economics. Conversely, a notion of *generalised Darwinism* is at the very core of the model proposed in the following (Aldrich et al. 2008; Essletzbichler and Rigby 2010). In the model, a population of businessmen try and innovate (in terms of opening of new markets) under conditions of uncertainty, and in doing so they propose new narratives and mental representations of future states of the world (*variation*); while others imitate the incumbents, agreeing or welcoming the more successful among the alternative and competing imagined future proposed (*inheritance*); and the environment in which all of them live and are embedded influence the success of each of these competing envisaged courses of events (*selection*), being in turn modified by the actions of these same agents.

We think a *business cluster* nothing more than as an ecology of firms and businessmen, or better as «an agglomeration of mutually reinforcing firms and aligned interests» (Feldman et al. 2005, p. 132). And, following Menzel and Fornahl (2010, p. 210), the theoretical framework here proposed argues that its

[2]About Evolutionary Economics more in general, see also Dosi et al. 1988; Metcalfe and Foster 2004; Nelson and Winter 1982; Witt 2008, among others.

evolution depends on the growth and decline of heterogeneity within the cluster: «clusters decline when they lose their ability to adjust to a changing environment and [...] this ability depends on the diversity of knowledge in the cluster». At the centre of the model proposed by the two authors, there are the different elements that compose the cluster: first and foremost the entrepreneurs and their companies, but also the other organisations, institutions, and the formal and informal networks among them. In particular, businessmen are the agents of change: «[b]y starting companies, entrepreneurs [...] draw on existing resources in the local environment and, in turn, add new resources to the environment that others can draw upon» (Feldman et al. 2005, p. 130). The transformations and modifications that happens at the level of these individual elements (micro-level) are the engines of the (macro-level) cluster dynamic. They are able to increase (decrease) the heterogeneity of the cluster itself by modifying its spatial and thematic boundaries (Porter 1998, p. 78), primarily through the introduction of new firms, skills, and capabilities from the outside (in both dimensions).

Lastly, as Feldman et al. (2005), we join the Schumpeterian tradition in saying that shocks or discontinuities –as the financial crisis that followed the "Banca Romana scandal", and that shocked Turin at the end of the Nineteenth century– are triggering events that motivate businessmen to take the risk to bring innovation to the market; i.e., to move from being latent to active entrepreneurs (Schumpeter 1942, pp. 82–83). In recessionary periods, the level of uncertainty within the system grows, the structural constraints weaken, and in the end the degrees of freedom for individual agency rise. The discontinuities generate temporary out-of-equilibrium conditions in the system, that can be exploited by businessmen to *creatively* react to the crisis by introducing novelties in the system in front of expected excess profits (Antonelli 2015; Schumpeter 1947). In particular, Lane and Maxfield (2005, p. 10) defined *ontological uncertainty* as a situation in which each actor has her own beliefs about «what kinds of entities inhabit their world; what kinds of interactions these entities can have among themselves; how the entities and their interaction modes change as a result of these interactions».[3]

> Ontological uncertainty, [...] resists the formation of propositions about relevant future consequences. The entities and relations of which such propositions would have to be composed are simply not known at the time the propositions would have to be formulated – that is, during the extended present in which action happens —Lane and Maxfield 2005, pp. 10–11.

In such circumstances different (maybe many) alternative possibilities are in the disposal of each entrepreneur, and it is hard to choose among them. Therefore, because the possibility space is unclearly defined, a primary entrepreneurial function is precisely to explore this space for new opportunities, feasible in the context they are embedded in, but yet unexploited. But since the true level of these profits can be discovered, by definition, only ex-post, and since the entrepreneurial creative reactions generate a path-dependent process that nullifies the possibility of a proper comparison between alternative courses of

[3] Similarly, Dequech (2006, p. 112) defined *ontological uncertainty* as a condition «characterized by the possibility of creativity and non-predetermined structural change. The list of possible events is not predetermined or knowable ex ante, as the future is yet to be created».

action, it is hard to orient oneself by use of a constraint optimisation rule, as standard in Neoclassical Economics. An alternative decision-making strategy available to the entrepreneurs is their very human ability to create and visualise plausible images of the future for each of the paths that they have in front. In other words, following Schumpeter we can say that entrepreneurs are guided by opportunities and expected profits in their *creative* reactions, but that their decisions are not grounded on a fully rational maximisation of the opportunities they face, under given constraints, but by a comparison of reasonable imagined courses of events that they create in their mind before actually plan their actions.

Therefore, firstly and as underlined by Hausmann and Rodrik (2003), to look at the economic *fundamentals* –the endowments of an economy in terms of natural resources, labour, physical and human capital, along with its institutions' quality– is not enough if we want to explain their development paths, because is likely that similar systems under this point of view, will follow different specialisation patterns. The authors propose an alternative mechanism in which the range of goods that an economy ends up producing is determined not only by its *fundamentals*, but also by the effectiveness of the *cost discovery* activity of its entrepreneurs. In other words, we can say that, under conditions of ontological uncertainty, a key and preliminary problem that entrepreneurs have to solve, in order to introduce some novelty in the system, is the exploration of the cost structure that characterises each economic system.

Compared to Hausmann and Rodrik, here we argue that entrepreneurs face a *creative process* more than a *discovery process* (Buchanan and Vanberg 1991). Indeed, as highlighted by Foster and Metcalfe (2012, p. 421), conventional economic decision-making based on a constraint optimisation rule «cannot approximate economic decision-making when there is uncertainty [...]. This is the typical state in which technological, organizational and institutional changes occur [...]. The presence of uncertainty does not prevent economic behaviour from occurring. On the contrary, we observe much creative, cooperative and competitive behaviour in states of uncertainty and the result is 'economic evolution' which is characterised by increases in organised complexity in economic systems and accompanying rises in wealth and per capita income». A way to break the possible deadlock that ontological uncertainty can generate is through what Jens Beckert has recently called *fictional expectations* (Beckert 1996, 2013a,b, 2016). As the author explains,

> By "fictions" I refer to images of some future state of the world or course of events that are cognitively accessible in the present through mental representation. [...] Actors are motivated in their actions by the imagined future state and organize their activities based on these mental representations. [...] Fictional expectations in the economy take narrative form as stories, theories, and discourses. Since these representations are not confined to empirical reality, fictionality is also a source of creativity in the economy. Including fictionality in a theory of decision-making opens up a way to an understanding of the microfoundations of the economy's dynamics and growth —Beckert 2013b, p. 220.

And Beckert continues explaining that

> [...] approaches in economic sociology see action as being anchored in networks, institutions, and cultural scripts that direct

> choices [...] These approaches usually make the uncertainty and indeterminacy of decision situations the starting point of their reasoning and bring to the fore the need for actors to *interpret* the social situation.
> But what informs these interpretations of the situation? I suggest [...] "fictions". [...] While fictions help in "overlooking" uncertainty in decision-making by providing *seemingly* good reasons for specific decisions, they are at the same time also a source of the uncertainty they are responding to, because the plethora of possible imaginaries of the future provides an overabundance of options and can bring about novelty by shaping action in unpredictable ways —Beckert 2013b, p. 222.

This type of solution for generation action in the face of *ontological uncertainty* is essentially the same that David Lane and Robert Maxfield themselves identify when they write that

> actors facing ontological uncertainty [...] *cannot* use the value of future consequences to generate appropriate action in the extended present. Narrative logic provides an alternative mode to generate such action, through its immanent link between character and denouement —Lane and Maxfield 2005, p. 15.

Also John Foster and Stanley Metcalfe have highlighted

> Humans have vivid imaginations. [...] Entrepreneurial behaviour involves an imagined novel product or service that can be delivered by an imagined productive structure that yields imagined wealth and/or power. Once such an aspiration is emotionally locked in, then knowledge, information and logic are applied to try to creatively achieve the aspiration —Foster and Metcalfe 2012, p. 428.

And even Robert Shiller has recently advocated a more extensive use of what he has called *narratives* in Economics, because

> When in doubt as to how to behave in an ambiguous situation, people may think back to narratives and adopt a role as if acting in a play they have seen before. The narratives have the ability to produce *social norms* that partially govern our activities, including our economic actions —Shiller 2017, p. 972.

Therefore, in a first recovery phase that follows a shock, these mechanisms of *exploration* (March 1991) will increase the level of *diversity* within the system, either in terms of firms that operates in different markets, or at least in terms of competing alternative visions of the future evolution of the localised economic system. At a later time, the systemic conditions, and in particular the contextual –thematically and spatially defined– institutions and organisations, as well as the networks among them, become the critical factor of success or failure of the entrepreneurial activity (Feldman et al. 2005; Schumpeter 1947). In particular, environments both full of *external economies* and characterised by a structure of relations among the agents that encourages some degree of cooperation –i.e. *flexible* and *coherent*– act as essential factors in the further development of a cluster. *Flexible*, because it must be able to quickly change, counteracting exogenous shocks and adapting itself to mutations of the environment in which the system is immersed.[4] *Coherent* –i.e., organised–, because, we can depict innovations as the outcome of a collective effort (or the emergent property of a complex system), instead of as the heroic action of a single individual (Richerson and Boyd 2005, p. 50; see also Allen 1983; MacLeod 1988; MacLeod and Nuvolari 2010; Nuvolari 2004). Thus, if the people who find themselves, for some historical reasons, in a certain context –and therefore have to act within it, subject to its constraints and opportunities– possess a irreconcilable and diverging vision about future states of the world, the efforts they make to try to tackle the substantive uncertainty that characterises any innovation process (Dosi and Egidi 1991) will be solved in a dispersion of the forces involved, rather than to be more than proportional compared to the individual efforts, regardless of the quantity of available external knowledge present in that particular context, or even worse to will lead the system to a *lock-in* phenomenon.

On the contrary, if properly tuned, these systemic-level devices act as *focal points* (Berger and Luckmann 1966; Schelling 1990), that help to reduce the uncertainty of the system and give rise to an *auto-catalytic* process, with powerful *increasing returns* and synergistic efforts among the not coordinated efforts of each entrepreneur. As highlighted by Antonelli (2011), an *organised complexity* is a systemic condition needful to have *creative reactions* of the agents in out-of-equilibrium situations.[5] Indeed, if, on the one hand, a higher *diversity* and heterogeneity of economic systems is a driver of change, a too diversified structure is, on the other hand, an obstacle to change, since reduces the feedback loops potential and the strength of the increasing returns within the *cluster*.[6] A feeling of belonging to a shared (imagined) community is truly useful whenever a decentralised and chaotic system must be driven out of trouble by means of an evolutionary process within a substantially uncertain environment. But this

[4]This characteristic makes the system efficient in the long term, even though in the short run it could be the opposite, since some degree of resources *slack* both helps an evolutionary process of the system and are its unavoidable byproduct. For example, Nohria and Gulati (1996) have shown the existence of an inverted U-shape relationship between *slack* and innovation at the firm level, and we can assume that a similar mechanism will be true also at a more aggregate level (like regions or business groups). Moreover, the positive role of *slack* and other byproducts for further technological evolution has been clarified by Joel Mokyr (see e.g., Mokyr 2000). In some recent paper, like Andriani and Carignani (2014), Bonifati (2013), and Dew and Sarasvathy (2016), Stephen Gould's work on *spandrel* and *exaptation* (Gould and Lewontin 1979; Gould et al. 1982) has been applied to technological change. In particular, Dew et al. (2004, p. 71) have highlighted how *exaptation* is «a missing but central concept that links the development of technology, the entrepreneurial creation of new markets and the concept of Knightian uncertainty». That is to say, markets can also (and not rarely) be set up by suppliers, not only by consumers: in that case, the function of a product pre-exists to its "being needed". Or even, we can say that *needs* and *functions* are independent of one another as *problems* and *solutions* are in the Garbage Can Model (Cohen et al. 1972).

[5]As explained by Miller and Page (2007, p. 48 and 50), with *disorganised complexity* «the interactions of the local entities tend to smooth each other out». Conversely, with *organised complexity* «interactions are not independent, feedback can enter the system. Feedback fundamentally alters the dynamics of a system. [...] With positive feedback, changes get amplified leading to instability». See also Jacobs 1961; Schumpeter 1947; Weaver 1961.

[6]Looking at the other side of the coin, Boschma (2005, p. 71) has argued that «too much and too little *proximity* are both detrimental to learning and innovation. That is, to function properly, proximity requires some, but a not too great, distance between actors or organizations».

also means that only one (or few) of the possible imagined alternative futures will be actually realised as a *collective action* (or, in other words, as an *emergent property* of the system). The outcome of this dialectic dynamic –from the *micro* to the *macro*, and backward– is an Andersonian-like *Imagined Community*, or what can be called an *Imagined Cluster* (Beckert 2016; Anderson 1991).[7]

Which one of the possible open alternatives will be actually realised depends on a series of factors. Partly it will the result of the contextual specific characteristics: e.g., the availability of some necessary natural resources, or of a specialised workforce pool. Partly it will be influenced by some systemic hysteresis and legacy of the past. And partly by *path dependence* like mechanisms that amplify small initial advantages of some entrepreneur in terms of power and rental position due to economic and extra-economic factors. With respective to this last point, the success of an entrepreneurial idea is therefore dependent, among other things, on factors such as the capacity of the people that have visioned it to convince others to share and support those products of their imagination until they became reality (Beckert 2016; Gramsci 1948–1951), or their ability to modify the context and its institutional setting (DiMaggio 1988).[8]

Once some of the ideas have established themselves as dominant in the local context, as Maskell and Malmberg (2007) have shown, entrepreneurial myopic behaviours are able to explain the self-reinforcing mechanisms that lead to a reduction of the level of heterogeneity of the cluster components.

Therefore, in the first phase that follows a shock, clusters have to deal with an *information* issue, if we want to follow Hausmann and Rodrik (2003), or with a *creation* issue, if we want to stay closer to the model here proposed. Under conditions of *ontological uncertainty*, the *fictional expectations* of the businessmen in the cluster are a way to explore and expand the *possibility space* of the whole environment.[9] Instead, in the second phase remembered above, the

[7]This *imagined community* is a way to reduce the uncertainty, by diminishing the information entropy of the system. A topic that is well known by the Institutional Economic Historians. For example, Douglass North (2005) explained that flexible cultural scaffolding helps a sustained and lasting growth of the economic system, since institutions that are faster in adapting to changed systemic conditions help local (economic) organisations to face unexpected situations. Also Joel Mokyr (1990, 1994, 2002, 2008) –by referring also to the work of Avner Greif (2006) about culture influence on economic development– underlines the specific role that formal and informal institutions have on the dynamic of the technological side of national or regional economic systems. Specifically, in his recent 2016 book, Mokyr has highlighted the role that a small intellectual elite had as the bases of the Industrial Revolution. In his opinion, the *Republic of Letters* were a cultural and institutional setting able to overcome the two *market failures* of the *market for ideas*: the weakness of positive incentives due to the well known *public good* characteristics of knowledge and the excessive strength of negative incentives arising from the fact that new ideas often degrade the value of the human capital of the existing orthodoxy and thus intellectual innovation will be resisted and some times persecuted as apostasy or heresy.

[8]Here the institutions are seen not as mere *constraints*, as in North (1990), but also as *opportunities*, so that the *problem space* and the *solution space* are no more linked through clear relations and proportions between their elements, but are part of a unique *possibility space* in which problems and solutions are independent of each other (Cohen et al. 1972), or at least there exist different levels of strength among the institutional constraints (Landesmann and Scazzieri 1996; Pasinetti 2007).

[9]The core argument raised by Hausmann and Rodrik (2003) is precisely that there is an issue of under-investments in the entrepreneurial self-discovery process, since *information externalities* are produced and followers can freely exploit incumbents' cost discovery investments.

main issue that has to be solved is of *coordination* (Rodrik 1996, 2004). Also in this case, the use of *imagination* is of some help. The feeling of belonging to an *imagined community* that *narratives* help to create is a means to coordinate the action of independent agents towards a common goal.

3 The *business community* of Turin, 1883–1907

The economic structure of Turin could be seen as a case in which industrial agglomerations played an opposite role in two different historical contexts. During the so-called Italian take-off of the first decade of the Twentieth century, the city was able to catch the new technological opportunities, moving from a textile-based economy to a more modern engineering-based one. Conversely, a hundred years later, the city was not able to get out unhurt from the "automotive-centric" *technological paradigm* (Dosi 1982) that had ensured it a substantial and lasting economic prosperity. This, by the way, shows that the independence of the technological and the cluster life cycles, and so proving the importance of a cluster evolution theory based on a super-naturalistic metaphor able to account for the renewal and revival capacity of such systems (Martin and Sunley 2011).

Looking at the first of these historical events, the data available at the macro level clearly show us the transition happened. For example, we can refer to some recent contributions that have analysed the GDP specialisation of the Italian regions (Province)[10] in the years 1871–1911 (Basile and Ciccarelli 2018; Ciccarelli and Proietti 2013), exploiting the newly available data produced by Ciccarelli and Fenoaltea (2013).[11]

Firstly, in Fig. 1 we can see that Turin was one of the few Italian Province that exhibited a positive growth in value added (VA) in the period 1871–1911. Moreover, Fig. 2 shows the level of *relative industrialisation* in each of the Italian regions (*Province*) in each of the census years in the period considered, where the index is defined as the rate of the fraction of the VA of the Provincia with respect to the Italian VA over the share of the male population over age 15 within this geographic unit. The figure highlights how Turin was constantly one of the more relatively industrialised Italian Province over the whole period. Considering also that, in the same time window, the population of the urbanised area of Turin more than doubled, from $199,476$ to $427,106$ inhabitants,[12] it is immediately clear that it represented a period a relevant socio-economic transformation.

Following Ciccarelli and Proietti (2013), we can compute the specialisation of each Provincia in each manufacturing sector (s_{ij}), to look at the changes in

[10]Similar to what is nowadays classified as a NUTS 3 level geographical aggregation.

[11]These data come from one of the strands of the Italian sub-national cliometric literature originated by Stefano Fenoaltea more than a decade ago and consolidated by a set of subsequent co-authored works (see in particular Ciccarelli and Fenoaltea 2009, 2014). With respect to other streams, it considers more in-depth the internal composition of industrial sectors. Indeed, within this framework, whenever the historical sources allow one to do so, product-specific value-added estimates are first provided and the value-added estimates for the whole industrial sector are then obtained by adding its elementary components previously determined.

[12]A growth rate of about 114.1%, against an average population growth of the 15 larger Italian urbanised areas of about 79.9%. Source: Istat, *Serie storiche della popolazione*, Tavola 2.20.

Figure 1: Province that shown an increase (left) or decrease (right) of their VA between 1871 and 1911. Source: Ciccarelli and Fenoaltea (2013, p. 71).

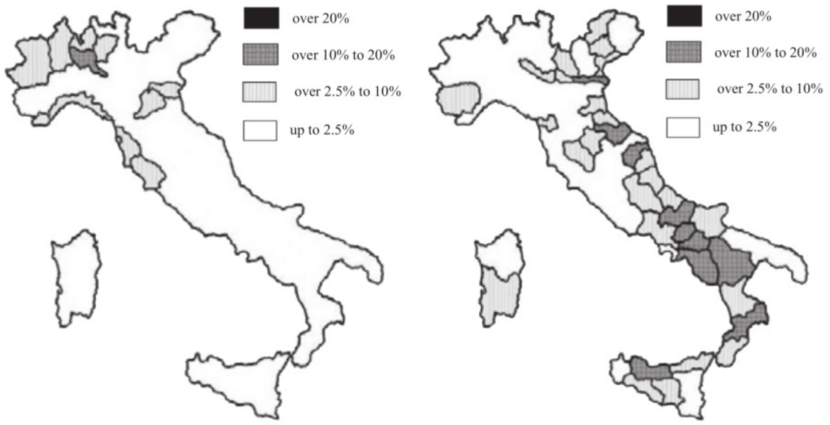
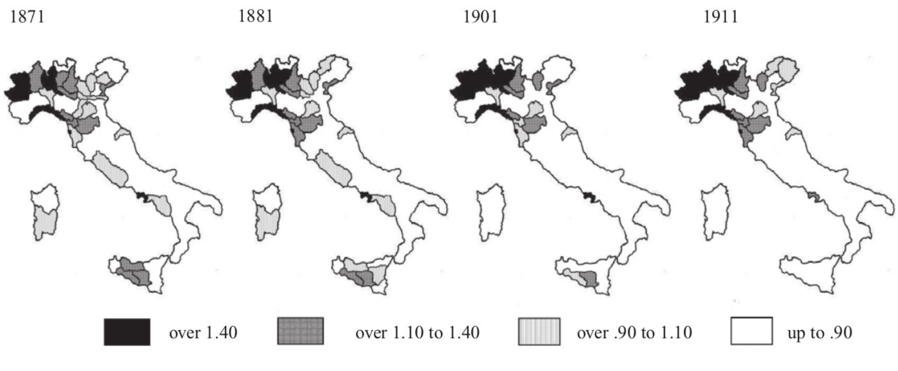


Figure 2: Industrialisation level of the Italian Provinces by year. Source: Ciccarelli and Fenoaltea (2013, p. 65).



the relative relevance of each of them in terms of VA (see Tab. 1 and 2). Where

$$s_{ij} = \frac{VA_{ij}}{\sum_j VA_{ij}} - \frac{\sum_i VA_{ij}}{\sum_{ij} VA_{ij}},$$

for each Provincia i and sector j. So that $s_{ij} \in (-1, 1)$.

Fig. 3 shows how the specialisation of the Provincia of Turin moved from more traditional sectors and mature technologies to some of the leading ones at that time, even though more strongly in the 1910s. Instead, Fig. 4 and 5 compare, respectively, the relative industrial specialisation of Turin with the one of other Italian Province in the same years you can find the figures for the other Province of Piedmont and for the other two Province of the so-called *Triangolo Industriale*.

Table 1: Decennial rates of change of the VA of the Provincia of Turin by manufacturing sector. Data: Ciccarelli and Fenoaltea (2013). I excluded Tobacco, Mining and Sundry manufacturing. Legend: Food (foodstuffs); Text (textile); Cloth (clothing); Leath (leather); Wood (wood); Mtmk (metal making); Eng (engineering); Nmmp (non-metallic mineral products); Chem (chemicals and rubber); Paper (paper and printing).

Years	Food	Text	Cloth	Leath	Wood
1871–1881	0.44	32.67	48.72	42.00	22.95
1881–1901	17.48	73.13	34.48	27.46	24.67
1901–1911	17.05	31.52	41.84	-4.55	61.61
1871–1911	14.67	82.43	64.10	27.50	49.18

Years	Mtmk	Eng	Nmmp	Chem	Paper
1871–1881	140.00	63.28	100.00	89.47	54.84
1881–1901	116.67	17.46	52.08	59.72	62.50
1901–1911	120.00	142.20	148.98	137.97	79.63
1871–1911	415.00	108.40	229.17	222.37	131.45

Table 2: Average decennial rates of change of the VA of all the Italian Province by manufacturing sector. Legend and data: see Tab. 1.

Years	Food	Text	Cloth	Leath	Wood
1871–1881	7.98	18.53	27.18	30.23	11.18
1881–1901	15.53	47.77	22.20	27.89	31.77
1901–1911	28.44	32.25	40.75	3.41	56.25
1871–1911	20.44	51.63	39.62	27.45	46.03

Years	Mtmk	Eng	Nmmp	Chem	Paper
1871–1881	120.83	30.53	39.15	38.16	50.00
1881–1901	87.42	25.23	26.90	47.83	60.30
1901–1911	169.57	77.51	141.71	114.90	96.75
1871–1911	384.03	62.15	104.31	120.23	137.77

Figure 3: Relative industrial specialisation index of the Provincia of Turin: 1871 (up-left), 1881 (up-right), 1901 (down-left), 1911 (down-right). I re-map the index from (-1;1) to (0,1). Legend and data: see Tab. 1.

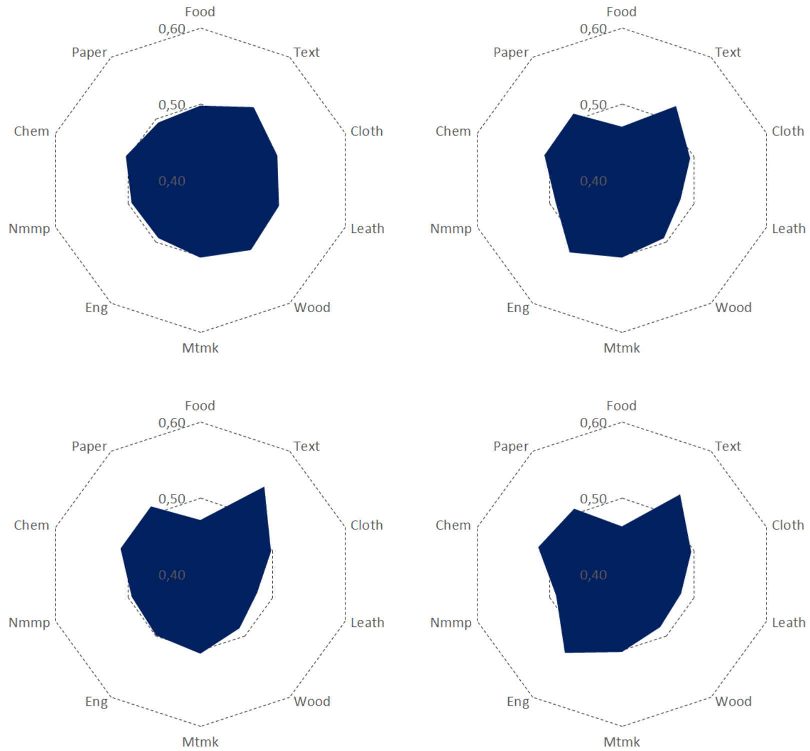


Figure 4: Relative industrial specialisation index of the other Provincie of Piedmont (Alessandria, Cuneo, Novara): 1871, 1881, 1901 and 1911. The blue line represents Turin. Legend and data: see Tab. 1.

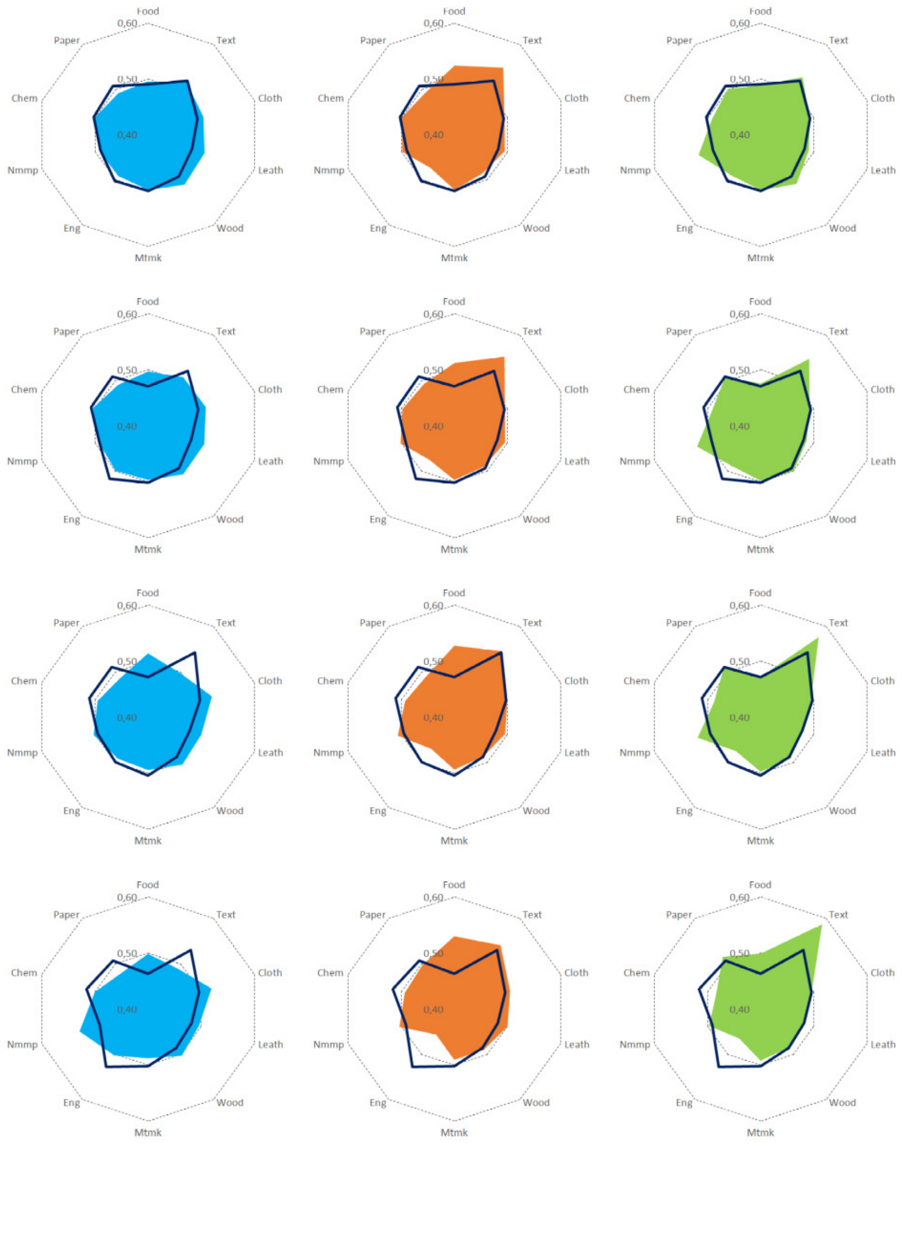
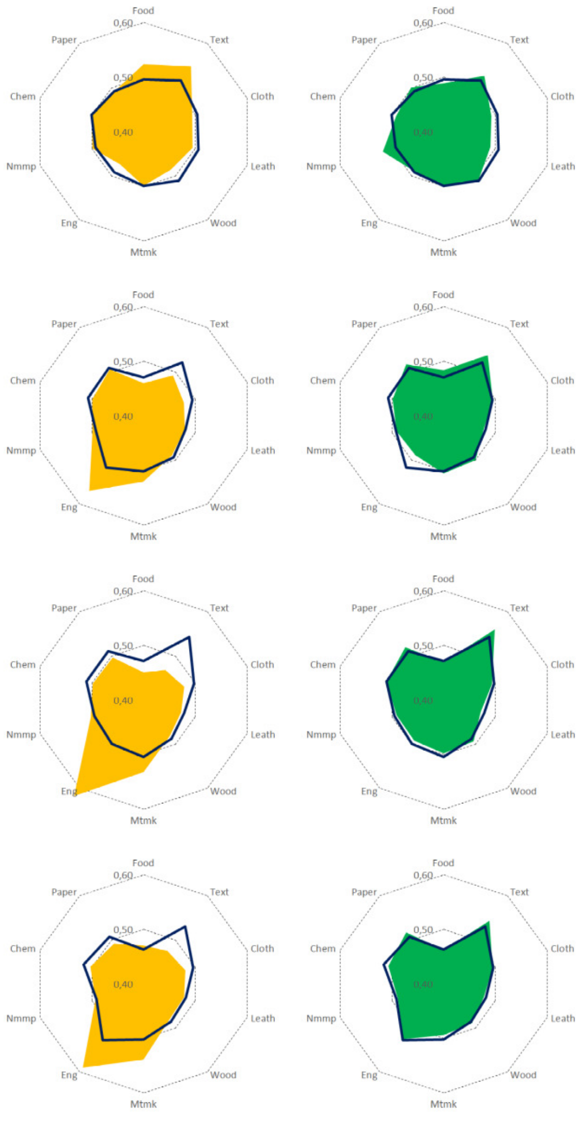


Figure 5: Relative industrial specialisation index of the other Provincie of the so-called *Triangolo Industriale* (Genoa, Milan): 1871, 1881, 1901 and 1911. The blue line represents Turin. Legend and data: see Tab. 1.



However, even though we agree with Ciccarelli and Fenoaltea (2013, p. 68) that macroeconomic and extra-economic factors influenced the success of some Italian Province in catching an economic growth path and that, in particular, «[w]ater, falling water, was the king-maker, and the western Alps provided a disproportionate share of the Italian total», in this paper we want to look at the microeconomic factors that helped these events. Indeed, not all the local economic systems of the North of Italy succeeded and not all of them followed the same path in terms of sectoral specialisation. For example, Fig. 4 shows that, compared to Turin, the specialisation of Cuneo in foodstuffs and textile increases over time, while Novara (the other industrialised area of the region, as shown by Fig. 2) specialised strongly in textile, with a weaker focus on engineering and chemicals and rubber compared to Turin. Instead, Fig. 5 suggests that Turin and Milan followed a similar specialisation pattern in the time period considered, while Genoa strongly focused on the engineering and metal making sectors.

These data, even though extremely interesting and informative, do not provide us of an explanation of what happened at the micro level, and so about the reasons of the successful transition of the economic system of Turin at the down of the Nineteenth century. In order to try to fill this gap, I will exploit the database about the *business community* in Turin between 1883 and 1907 recently built by Ivan Balbo (2007), also with the hope of increasing its underestimated value.[13] Indeed, these data refer to a quite interesting case of techno-structural change in the Italian economic history: the decline of the textile-centric economy at the end of the Nineteenth century and the rise of the automotive industry as the new dominant industrial sector in Turin since the first decade of the Twentieth century.[14]

The central point of the analysis carried out by Ivan Balbo (2007) is to reconsider –in accordance with other recent works; see e.g., Rugafiori 2003– the apparent discontinuity that was described by the traditional historical literature about the economy of Turin at the dawn of the Twentieth century. Indeed, the economy of this city in the years 1889–1894 was characterised by a strong economic crisis, which was involving the whole national economic system (see e.g., Di Martino 2012). Looking at the whole capitalisation of the firms active in Turin at the time as a proxy of the economic cycle of the city, Fig. 6 shows a clear downfall in 1889–1890, followed by a stagnation of about 7 years (1891–1898). As well, Fig. 7 provides a similar feeling, showing a negative trend in the birth of new firms in the recessionary period (1889–1894). Conversely, in the first years of the Twentieth century, both the figures show a strong rebound (1898-1902 in particular), characterised also by the creation of many new firms. Moreover, Turin seems having taken a well defined industrial profile, in which the automotive industry, and the Fiat in particular, has a key and advanced position. More precisely, is exactly the sudden birth of the automotive industry to be seen, by the traditional historiography, as the solution of the crisis began at the end of the 1880s.

In order to show that, during the 1890s economic crisis, Turin was not characterised by a radical break of the existing economic structure, unlike what has been claimed by the traditional historical literature, Balbo tried to highlight

[13] A *business community* is a set of entrepreneurs and firms linked together by social and economic relationships. In this regard, you can see the works by Stefania Licini (1998, 1999).

[14] I will describe the data more in details at p. 89.

Figure 6: Total shares of the companies active in Turin from 1883 to 1906 at constant prices (1902 = 1.0 Lire, as in Balbo 2007, using data from Istat). The grey area, here as the following, highlights the window 1889–1894 considered by the historical literature as the crisis years. The solid line uses the data as reported in the source. In the other two lines, if missing, the middle years is imputed as equal to the last observed, wherever we know the capital shares of a company in two non-contiguous years. If more values are recorded in a year, the dotted line takes the maximum one, while the dashed uses the minimum.

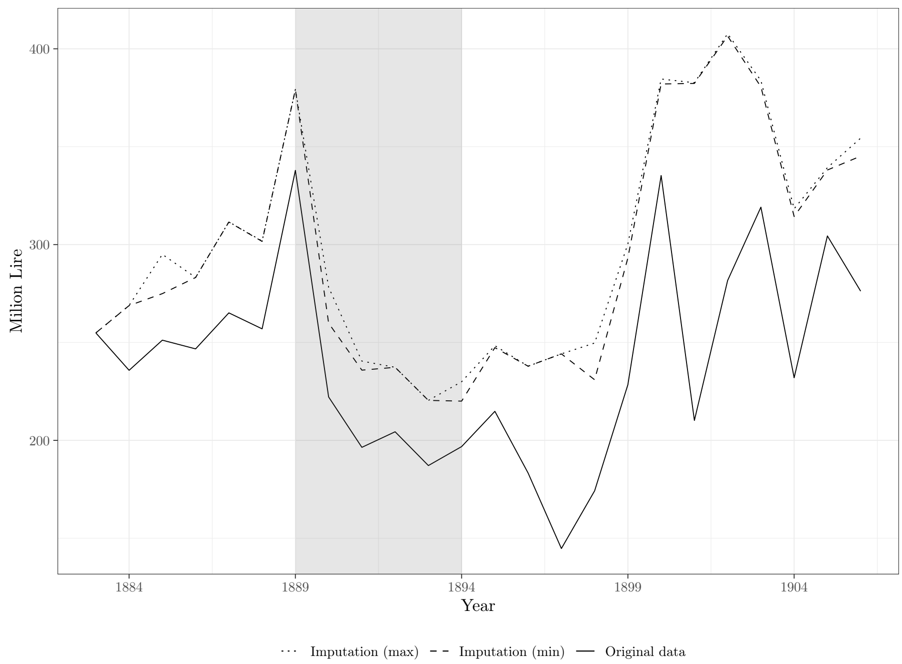
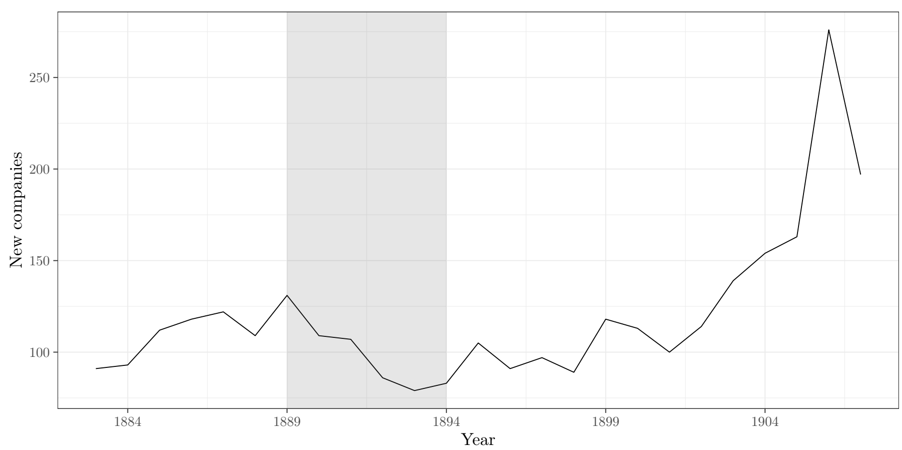


Figure 7: Number of legal documents having as object the establishment of a new company (*costituzione*).



which resources let a gradual and continuous evolution of the city-system from the first to the second configuration. In particular, the historian tried to answer two connected questions. On the one hand, the author wanted to identify which resources ensured the persistence of the economic system of Turin in spite of the land and real estate speculation, connected with the "Banca Romana scandal" of 1893 and that produced a bank crisis in Turin. On the other hand, he intended to discover the fractures that led to the automotive boom in the first years of the Nineteenth century, and more in general to the new economic configuration of the so-called Giolittian Era.

The conclusions of the historical enquiry have identified three agents, endogenous to the economic system of Turin, like the ones that ensured the continuity of the intersectoral relationships and of the financial flows during the economic crisis of the last to decades of the Nineteenth century: the private bankers with their relationships network build on cross-shareholdings; the cotton Protestant entrepreneurs whose relationships were based on this common membership to this religious minority and made of family and marriage ties; and the mechanical engineering industry that was going to take a sort of industrial district shape. The analytical tool chosen by Balbo, the so-called *business community*, arises as particularly useful to highlight the main point that these three elements have in common: they were able to generate external economies to each specific firm, but that were, on the contrary, internal to the local *business community* to which such firms belonged. These externalities, based on the concept of *trust*, were the essential component able to overcome information asymmetries and to help the circulation of trust also during the crisis.

Conversely, Balbo's research shows how other three elements –the electric power industry; the universal bank (*banca mista*); and the automotive industry– act as discontinuity factors during the crisis. Therefore, the main critical question is to try to account for how these novelties can be introduced within a landscape otherwise characterised by a substantial continuity. Furthermore, the author was also interested in analysing which changes these breaking elements have made on the *business community* of Turin.

Therefore, as I clarified above, the core of the analysis is the relationships among the people belonging to the *business community*:[15] mainly their family ties; their common belong to a religious or cultural minority; and the fact of sitting in the same Board of Directors.

The relationship network is described as a means that helps to transfer capitals and skills among the local industrial sectors, and –as widely explained by Balbo– the *trust* is a key element that helps the circulation of these resources among the sectors and along the time. To focus on this point is useful to investigate the circular flows of investments within the local economic system:[16]

[15] A systemic description, similar to the one proposed by Balbo, is used also in other books like De Benedetti (1990) and Rugafiori (1994, 1999).

[16] To have put at the heart of the analysis the cobweb of socio-economic connections, seen as a way to relocate capitals and skills among industries, helps the author to investigate the circular web of investments of the city at the turn of the crisis: How it happened that it was stayed alive during the crisis? How this episode changed it? Balbo put special care in understanding if the *business community* has been able to act in an organic and integrated way. Indeed, such a configuration is seen as the only one able to guarantee an auto-catalytic chain drive, through which the development of one component induces a similar evolution of other portions of the system. Conversely, a progressive disarticulation of the relations among the structures is able to negatively affect the regional economic performances.

which elements helped it to survive during the economic crisis and how it has been changed by the breaking novelties introduced?

In Balbo's book clearly emerges the idea that economic change is not only a process guided by the technological change, but also by changes at the level of the social structure –i.e., changes that affect the direct and indirect relationships among the relevant economic actors. Moreover, the historical analysis summarised above suggests a strong inter-relationship between these two components of the economic change process. On the one hand, the introduction of some (technological) novelty in the economic system leads the localised relational network to try to adapt itself to the new conditions. At the same time, an alteration of the topology of that network influences the ability of the system to help the diffusion of some innovation or, even, to induce further innovations.

In this regard, one of the most significant contributions to the economic history literature given by Balbo is to have shown how the three above said discontinuity elements do not produce a destruction of the economic fabric. Indeed, the data analysed by Balbo suggest that the endogenous agents, who have carried on an active investment circulation also during the crisis period, do not disappear abruptly. On the contrary, we observe an evolutionary process: the *business community* opens up to the outside, welcoming new members, and the private bankers and labour intensive firms survive longer than in other economic centres, like Milan or Genoa, discovering new spaces and reinventing new roles for themselves. We can even say that the network of personal and sectoral interdependencies –i.e., the *business community*– thickens and becomes more diversified. And, more in general, that we are in front of a case of *cluster reinvention*, of the type theorised above.

4 The heterogeneity of the Turin cluster and its evolution

In particular, in line with the cluster evolution model previously exposed, we expect to observe an increase in the heterogeneity of the cluster composition during the 1889–1894 crisis, since the higher uncertainty that characterises such a situation relaxes the constraints that condition the *space of possibilities* of the entrepreneurial agency, and increase the potential profitability that can derive for the introduction of a new fictional paradigm, i.e. of an entrepreneurial *creative reaction* to out-of-equilibrium conditions in their environment. We expect as well to observe a return to a pre-crisis higher conformity once that most of the *opportunities* for the introduction of new competing collective investment opportunities are closed, and the constraints of the socio-economic structure are stronger again. These hypotheses will be tested assuming that the more the communities in which the entrepreneurial network can be subdivided are able to explain the sectoral composition of the economy, the more heterogeneous the system is (and the opposite).

Data These propositions will be tested using data based on the legal documents[17] (*Atti di società*) deposited at the local court (*Archivio di Stato*), as

[17]These documents refer to events that affect the history of each company, like their constitution, dissolution, changes in the Board of Directors, and changes in the level of capitalisation.

prescribed by the new Italian commercial law (*Codice di commercio*) of 1882,[18] by the companies active in Turin in the period 1883–1907.[19] The database was built by Ivan Balbo, through the digitalisation of about 9,650 different legal documents. With respect to this latter, before having analysed them I have cleaned the data, trying to fix as many problems and errors as possible. For that reason, the descriptive statistics can differ from the ones shown by Balbo, Rugafiori and Ottaviano.[20] The final data set is made of data about 3,550 corporations and 17,900 businessmen or companies who own shares of these corporations or are in some way involved in them (e.g. as Chief Executive Officer, etc.). After an intensive data cleansing procedure of the name of the people and of the companies available in the data set, and after having re-link each company backwards at each name change, I attributed a unique identifier to each node of the network.[21]

Moreover, the data just described can be represented as a *bipartite graph* (*affiliation network*), where the nodes of one set are the companies and the nodes of the other are the people which have some role in that companies. The undirected link among them will be a legal document in which two of these elements (one for each disjoint set) appear together.[22] Given the temporal dimension of the data set, it is possible to have 25 of these graphs. Nevertheless, I preferred to sum up the observation of two years together so that the first wave will be composed by the data of the years 1883 and 1884, the second wave of the data of the years 1884 and 1885, and so on (so two subsequent waves have a year in common). This increases the stability of the observations, in particular on the layer of people that, by construction of the database, very easily appear and disappear in year by year only because they were not named in any legal document, and not because they were not still there.

Analysis As said, we would like to compare this last classification based on the industry of each firm's activity, to an alternative one, based on the en-

[18] See Ivan Balbo (2007). *Torino Oltre la Crisi: Una 'Business Community' tra Otto e Novecento*. Italian. Bologna: Il Mulino and Paride Rugafiori and Chiara Ottaviano (2008). *La Business Community a Torino 1883–1907*. CD-ROM. Torino. The former describes the data, while in the latter you can find the data I used. Moreover, Appendix A reports and explains the main variable available in the dataset, and shows some descriptive statistics.

[19] The final year was chosen has been set since in 1907 Turin was shocked by a market crash, that represented a relevant external event of Schumpeterian selection for the local economic system.

[20] I discussed the divergences with the original authors and they are with me in saying that, even though less adherent to the original paper documents, the version used here is more useful for statistical purposes and perhaps even closer to the historical truth.

[21] In particular, a person has been identified using (in order) his or her family name, first name, father name and birth-place. Whenever the last criterion was not available, I used only the first three, and so on. Instead, a company has been identified using its name (*ditta*), with the caveat just said.

[22] The network represents the structure of the economic system as it appears on the basis of the information provided by the Atti di Società. Therefore, differently from the 2007 book by Balbo, the information contained in the database refers only to the links generated by the co-presence of two individuals –in any possible role– within the same legal act of a given company. All the extremely interesting links of any other nature analysed in the book –first and foremost, the familiar, co-religionary and cultural minorities ties– have been reconstructed by the author in a less systematic and more conversational way, and consequently have not been considered in the present analysis.

trepreneurs' activity.[23] To do so, firstly the bipartite network will be partitioned in sub-groups (*communities*) using the BCFinder algorithm developed by Lehmann et al. (2008).[24] Differently from most of the *community detection* methods proposed in the literature,[25] it does not operate on the projected network,[26] but on the bipartite structure as it is. This has many advantages, compared with the project-and-divide procedure. Indeed, as explained by the authors

> The conceptual simplicity of the one-mode projection comes at a high cost. First of all, the procedure typically eradicates the sparsity of the E matrix; this is especially problematic, when constructing the adjacency matrix for the smaller set of nodes, in the case where one of the node sets is significantly larger than the other. Secondly, much of the information present in the bipartite state becomes encoded in the weights of the adjacency matrix. However, due to technical difficulties regarding the analysis of weighted matrices and the high link density of the one-mode projections [. . .], these matrices are usually thresholded such that only entries higher than some threshold are retained. Similarly, the diagonal of the one-mode adjacency matrices is usually set to zero, since self-links are not of interest in the subsequent network analysis –Lehmann et al. 2008, p. 2.

And, at list the first of the two points, match perfectly with the data used here.

Specifically, the procedure assumes that a typical community consists of several complete sub-bigraphs that tend to share many of their nodes. In other words, a community is formed by a group of nodes that are completely connected to one another. In particular, it will be chosen the biggest possible group among the observed ones. In this way, a group of companies will be part of the same *community* the more people are named in the legal documents of all these businesses at the same time.[27]

Having at hand two alternative classifications of the companies active in the *business cluster*, we can compare how much they agree to each other in each year considered. To measure the quantity of information shared by the two repartitions, we will use the index of Mutual Information index, defined as (Cover and Thomas 2006)

$$\mathrm{MI}(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}.$$

[23]In particular, the industrial sector of each company is the ones provided in the database, that is based on a re-classification made by Balbo. Indeed, the original legal documents contain information about the industry of each firm that was self-reported by the firm itself. Instead, Balbo used the Istat industries classification of 1911 matching what declared with the more similar class available (plus some extra classes that the author introduced for reasons of convenience). In particular, in the following analysis, only the first level of the classification has been considered.

[24]I used the software provided by the authors themselves, and that can be found at `http://www.imm.dtu.dk/~sljo/bcfinder/`.

[25]See Fortunato 2010; Porter et al. 2009 for a review.

[26]Broadly speaking, a bipartite network projection can be defined as the shadow of the links that connect people and companies on the surface on which the companies (in this case) lie.

[27]A specific aspect that must be taken into account in reading the results, is that this algorithm does not assign a company to a unique community, but some companies will bridge the communities, being part of both.

Intuitively, the Mutual Information index measures the information that two variables X and Y share. That is, the index capture how much the knowledge on one of the two variable reduces the uncertainty about the other. In particular, if the two variable are independent, $\mathrm{MI}(X, Y) = 0$, since $p(x, y) = p(x)p(y)$. If, at the other extreme, it is possible to define one of the variables as a function of the other (say $Y = f(X)$) the Mutual Information is equal to the Entropy index of the variables ($H(X)$), since all the lacking information depend on the uncertainty about one of the variables.

But, since we want to compare different years, we need to normalise this measure. Indeed, in different years, we will have a different number of communities, and a different number of elements in these groups. Likewise, also the number of industrial sectors and the number of companies in each of them vary year by year. Therefore, we will compare the grouping obtained on the bases of the information provided by the sectoral composition of the economy, and the one obtained by partitioning the network as explained in the previous paragraph, through a Normalised Mutual Information index defined as (Strehl and Ghosh 2002)

$$\mathrm{NMI}(X, Y) = \frac{\mathrm{MI}(X, Y)}{\sqrt{H(X)H(Y)}}.$$

Thus, if we think Mutual Information in analogy with the *covariance*, and Entropy as a *variance*-like index, the NMI can be compared to the Pearson correlation coefficient.

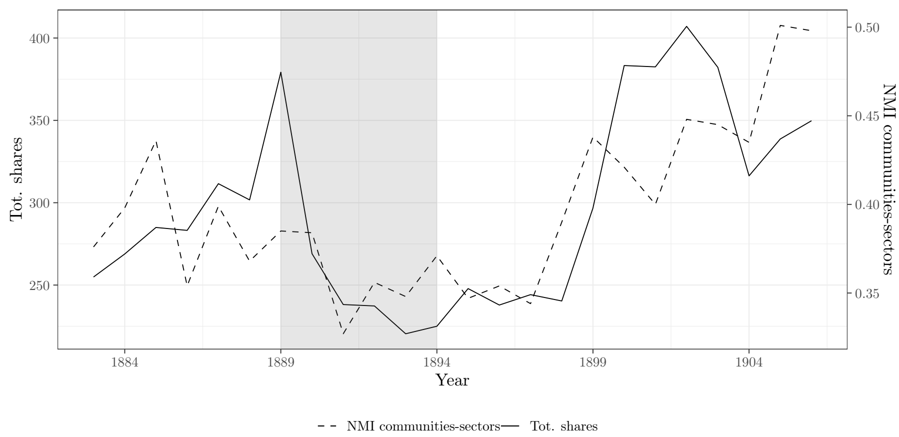
Results As shown by Fig. 8, in the crisis period the information about the communities that can be identified in the network is less informative about their classification based on the industrial sector of belonging. Conversely, in the pre-crisis, and (even more) in the post-crisis periods, only weak ties bridge the sectors, and most of the network structure is explained by the fact that people tend to be involved only in businesses of a specific sector.

Furthermore, if we compare the series of the measure just described with the one of the total capitalisation of the system, the two show a common trend, with a Spearman's rank correlation of 75% (Fig. 9). The crisis, as proxied by the abrupt decrease in the total capital stock, is linked to an increase of the heterogeneity of the sectoral composition of the economic system. Similarly, the rebound of about a decade later shows a common increase of both the variables: at the same time both the homogeneity the capitalisation increased as hypothesised by the theoretical model proposed in the paper.

Therefore, the analysis suggests that the 1890s downturn increased the manoeuvring space available to the entrepreneurs for a reorganisation of the sectoral structure of the city, from a textile-based to a mechanics-centric economy. It seems that, in the period that followed the shock, the propensity to bridge different sectors increases compared to the pre-shock period. Conversely, in line with the theoretical model proposed, once the crisis has been re-absorbed the businessmen come back to a more conformative behaviour, and the number of links across industries shrunk, and only a few *weak ties* bridged them.

Figure 8: Normalised Mutual Information between the classification of companies according to their industrial sector, and the one obtained through the community detection method.

Figure 9: Normalised Mutual Information between the classification of companies according to their industrial sector, and the one obtained through the community detection method (dashed line, left scale), and total capitalisation of the economic system (solid line, right scale). The solid line of this plot is the average of the dashed and the dotted lines of Fig. 6.



5 Conclusions

The paper wants to be both a theoretical and a historical contribution. As we saw, the radical sectoral restructuring of the Turin cluster at the down of the Nineteenth century can be described as a case of cluster *renewal*, in which the city was able, combining an exploitation of some internal forces, with the incorporation of new elements from the outside of its spatial and thematic boundaries, to reinvent itself. In line with recent theoretical contributions, I argued that this evolutionary path asks for a *cluster evolution* model alternative to the classical *life-cycle* one, that, based as it is on a biological metaphor of *birth-growth-maturity-decline-death*, is not able to account for the possibility of (un-natural) cluster *revivals*. Therefore, the paper has framed this historical event within an alternative theoretical model in which the dynamic of the cluster is driven by the dynamical cycle of its heterogeneity. Moreover, in the paper it is proposed that this last dynamic is driven by the entrepreneurial activity. In particular, the *fictional expectations* of these latter agents are seen as able to explain both the entrepreneurial ability to react to out-of-equilibrium systemic conditions in unclear situations –like Turin in the post-1889 shock–, but also to justify the return to a conformative phase. Under the historical point of view, the paper adds to the previous *verbal representations* of Balbo (2007) a more data-grounded proof of the role of the heterogeneity cycle in driving the urban economic system out of the deep recessionary period followed to an exogenous shock, but that was also linked to a weariness of the technological cycle around which the *business community* was organised. The analysis of the historical data about the *business community* of Turin for the period 1883–1907 shows that the 1890s downturn temporarily increased the manoeuvring space available to the entrepreneurs for a reorganisation of the sectoral structure of the city, from a textile-based to a mechanics-centric economy, and also that this observed increased heterogeneity come back, at least to the pre-crisis level. Moreover, it seems to suggest the existence of a simultaneity between the *cluster cycle* and what could be called *heterogeneity cycle*.

In the end, the conclusions of the paper are interesting and promising, but still preliminary. Among possible others, a key aspect that has to be developed to strengthen these findings is a proper null model. To compare the empirical data to this null hypothesis will let us able to understand if the conclusions are mainly driven by trivial characteristics of the network, like the degree distribution of the two layers of the graph, or by something more peculiar and significant. Moreover, much of the theory proposed is still completely unproven, also because the data available have been shown to be dirtier than expected and a relevant portion of the information exploited in Balbo's book is not contained in the data, but based on less structured information sources. To deal with this limitation will require the availability of more detailed data on the entrepreneurial decision-making and on the economic performance of the firms, that are maybe unavailable at all. However, I hope this paper has opened the door to future improvements and further developments that will be able to provide important policy implications about the way in which a local economy can reinvent itself and create new development paths.

Acknowledgements

First and foremost, I would like to thank Dr. Ivan Balbo not only for his priceless, decennial, effort in digitising and cleaning the database which I have benefited from in writing this paper, but also for the time he spent helping me to move in the data and in the socio-economic history of Turin. Then, I have to express my gratitude to my supervisor, Prof. Aldo Geuna, for making me known of the database, and for the support, the kind pushes, and the many comments, without which I would not have reached this result. I am also indebted to Prof. Nicolo Bellanca for making me aware, many years ago, of the topic of the *imagined* and *fictionality* in Economics. Moreover, I thank Maria Del Rio Chanona for the hint to look at the Mutual Information index to test the theoretical expectations of the paper. And I am grateful to Prof. David Wolfe, Dr. Steven Denney and Travis Southin that share with my a preliminary version of their paper on the evolution of the Toronto ICT cluster, during one of the meetings of RENIR, that decisively influenced me in writing this work. Lastly, my gratitude goes to the participants of the LabNet internal seminar and the Fellows of the Complexity Economics group of INET@Oxford, whose comments and critiques to a preliminary version of this paper have helped me to significantly improve it. Any errors are my own.

References

Aldrich, Howard E., Geoffrey M. Hodgson, David L. Hull, Thorbjørn Knudsen, Joel Mokyr, and Viktor J. Vanberg (Oct. 2008). "In Defence of Generalized Darwinism". *Journal of Evolutionary Economics* 18 (5), 577–596. DOI: 10.1007/s00191-008-0110-z.

Allen, Robert C. (Mar. 1983). "Collective Invention". *Journal of Economic Behavior & Organization* 4 (1), 1–24. DOI: 10.1016/0167-2681(83)90023-9.

Andriani, Pierpaolo and Giuseppe Carignani (Nov. 2014). "Modular Exaptation: A Missing Link in the Synthesis of Artificial Form". *Research Policy* 43.9, 1608–1620. DOI: 10.1016/j.respol.2014.04.009.

Antonelli, Cristiano (2011). "The Economic Complexity of Technological Change: Knowledge Interaction and Path Dependence". In: *Handbook on the Economic Complexity of Technological Change*. Ed. by Cristiano Antonelli. Edward Elgar.

— (2015). "Innovation as a Creative Response: A Reappraisal of the Schumpeterian Legacy". *History of Economic Ideas* 23.2, 99–118.

Balbo, Ivan (2007). *Torino Oltre la Crisi: Una 'Business Community' tra Otto e Novecento*. Italian. Bologna: Il Mulino.

Basile, Roberto and Carlo Ciccarelli (May 2018). "The Location of the Italian Manufacturing Industry, 1871–1911: A Sectoral Analysis". *Journal of Economic Geography* 18 (3), 627–661. DOI: 10.1093/jeg/lbx033.

Beckert, Jens (Dec. 1996). "What Is Sociological about Economic Sociology? Uncertainty and the Embeddedness of Economic Action". *Theory and Society* 25.6, 803–840.

— (2013a). "Capitalism as a System of Expectations: Toward a Sociological Microfoundation of Political Economy". *Politics & Society* 41.3, 323–350. DOI: 10.1177/0032329213493750.

— (Mar. 2013b). "Imagined Futures: Fictional Expectations in the Economy". *Theory and Society* 42.3, 219–240. DOI: 10.1007/s11186-013-9191-2.

— (2016). *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Harvard University Press. ISBN: 0674088824.

Berger, Peter L. and Thomas Luckmann (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City (NY): Anchor Books. ISBN: 0385058985.

Bonifati, Giovanni (Jan. 2013). "Exaptation and Emerging Degeneracy in Innovation Processes". *Economics of Innovation and New Technology* 22.1, 1–21. DOI: 10.1080/10438599.2012.689674.

Boschma, Ron A. (Feb. 2005). "Proximity and Innovation: a Critical Assessment". *Regional Studies* 39 (1), 61–74. DOI: 10.1080/0034340052000320887.

Boschma, Ron A. and Koen Frenken (June 2006). "Why Is Economic Geography Not an Evolutionary Science? Towards an Evolutionary Economic Geography". *Journal of Economic Geography* 6 (3), 273–302. DOI: 10.1093/jeg/lbi022.

— (2018). "Evolutionary Economic Geography". In: *The New Oxford Handbook of Economic Geography*. Ed. by Gordon L. Clark, Maryann P. Feldman, Meric S. Gertler, and Dariusz Wójcik. 2nd ed. Oxford (UK): Oxford University Press. Chap. 11, 213–229.

Boschma, Ron A. and Jan G. Lambooy (Dec. 1999). "Evolutionary economics and economic geography". *Journal of Evolutionary Economics* 9.4, 411–429.

Boschma, Ron A. and Ron Martin (June 2007). "Editorial: Constructing an Evolutionary Economic Geography". *Journal of Economic Geography* 7 (5), 537–548. DOI: 10.1093/jeg/lbm021.

— eds. (2010). *The Handbook of Evolutionary Economic Geography*. Edward Elgar. Chap. 1, 3–39.

Buchanan, James M. and Viktor J. Vanberg (Oct. 1991). "The Market as a Creative Process". *Economics and Philosophy* 7 (2), 167–186. DOI: 10.1017/s0266267100001383.

Ciccarelli, Carlo and Stefano Fenoaltea (2009). "Le Industrie non Manifatturiere". Italian. In: *La Produzione Industriale delle Regioni d'Italia, 1861–1913: Una Ricostruzione Quantitativa*. Roma: Banca d'Italia.

— (Apr. 2013). "Through the Magnifying Glass: Provincial Aspects of Industrial Growth in Post-Unification Italy". *The Economic History Review* 66.1, 57–85.

— (2014). "Le Industrie Estrattivo-Manifatturiere". Italian. In: *La Produzione Industriale delle Regioni d'Italia, 1861–1913: Una Ricostruzione Quantitativa*. Roma: Banca d'Italia.

Ciccarelli, Carlo and Tommaso Proietti (Nov. 2013). "Patterns of industrial specialisation in post-Unification Italy". *Scandinavian Economic History Review* 61.3, 259–286.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark E. J. Newman (Nov. 2009). "Power-Law Distributions in Empirical Data". *SIAM Review* 51.4, 661–703.

Cohen, Michael D., James G. March, and J. P. Olsen (1972). "A Garbage Can Model of Organizational Choice". *Administrative Science Quarterly* 17.1, 1–25.

Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. Wiley. ISBN: 0471241954.

Csardi, Gabor and Tamas Nepusz (2006). "The igraph Software Package for Complex Network Research". *InterJournal* Complex Systems. URL: http://igraph.org.

De Benedetti, Augusto (1990). "Il sistema industriale (1880–1940)". Italian. In: *Storia d'Italia. La Campania*. Ed. by Paolo Macry and Pasquale Villani. Regioni. Einaudi.

Denney, Steven, Travis Southin, and David A. Wolfe (2018). "Transforming a Cluster While Building a Firm: Entrepreneurs and the Evolution of Toronto's ICT Cluster". mimeo.

Dequech, David (Jan. 2006). "The New Institutional Economics and the Theory of Behaviour Under Uncertainty". *Journal of Economic Behavior & Organization* 59.1, 109–131.

Dew, Nicholas and Saras D. Sarasvathy (Jan. 2016). "Exaptation and Niche Construction: Behavioral Insights for an Evolutionary Theory". *Industrial and Corporate Change* 25.1, 167–179. DOI: 10.1093/icc/dtv051.

Dew, Nicholas, Saras D. Sarasvathy, and Sankaran Venkataraman (Jan. 2004). "The Economic Implications of Exaptation". *Journal of Evolutionary Economics* 14.1, 69–84. DOI: 10.1007/s00191-003-0180-x.

Di Martino, Paolo (June 2012). "'Rome Wasn't Built in a Day': Lobbies, Institutions and Speculation in the 1880s Building Fever". *Urban History* 39.03, 471–489.

DiMaggio, Paul J. (1988). "Interest and agency in institutional theory". In: *Institutional Patterns and Organizations*. Ed. by Lynne G. Zucker. Cambridge (MA): Ballinger, 3–22.

Dormann, Carsten F. (2011). "How to Be a Specialist? Quantifying Specialisation in Pollination Networks". *Network Biology* 1.1, 1–20.

Dormann, Carsten F., Jochen Fründ, Nico Blüthgen, and Bernd Gruber (2009). "Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks". *The Open Ecology Journal* 2.1, 7–24. DOI: 10.2174/1874213000902010007.

Dormann, Carsten F., Bernd Gruber, and Jochen Fründ (2008). "Introducing the bipartite Package: Analysing Ecological Networks". *R News* 8.2, 8–11.

Dosi, Giovanni (June 1982). "Technological paradigms and technological trajectories". *Research Policy* 11.3, 147–162.

Dosi, Giovanni and Massimo Egidi (June 1991). "Substantive and Procedural Uncertainty". *Journal of Evolutionary Economics* 1.2, 145–168.

Dosi, Giovanni, Christopher Freeman, Richard R. Nelson, Gerald Silverberg, and Luc Soete, eds. (1988). *Technical Change and Economic Theory*. London and New York: Pinter Publisher.

Essletzbichler, Jürgen and David L. Rigby (2010). "Generalized Darwinism and evolutionary economic geography". In: *The Handbook of Evolutionary Economic Geography*. Ed. by Ron A. Boschma and Ron Martin. Cheltenham (UK) and Northampton (MA): Edward Elgar. Chap. 2, 43–61.

Feldman, Maryann P., Johanna Francis, and Janet Bercovitz (2005). "Creating a Cluster While Building a Firm: Entrepreneurs and the Formation of Industrial Clusters". *Regional Studies* 39.1, 129–141. DOI: 10.1080/0034340052000320888.

Fortunato, Santo (Feb. 2010). "Community detection in graphs". *Physics Reports* 486 (3-5), 75–174. DOI: 10.1016/j.physrep.2009.11.002.

Foster, John and J. Stanley Metcalfe (May 2012). "Economic Emergence: An Evolutionary Economic Perspective". *Journal of Economic Behavior & Organization* 82.2-3, 420–432. DOI: 10.1016/j.jebo.2011.09.008.

Gould, Stephen Jay and Richard C. Lewontin (Sept. 1979). "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme". *Proceedings of the Royal Society B: Biological Sciences* 205.1161, 581–598. DOI: 10.1098/rspb.1979.0086.

Gould, Stephen Jay and Elizabeth S. Vrba (1982). "Exaptation – A Missing Term in the Science of Form". *Paleobiology* 8.1, 4–15. DOI: 10.1017/S0094837300004310.

Grabher, Gernot (1993). "The Weakness of Strong Ties: The Lock-in of Regional Development in the Ruhr Area". In: *The Embedded Firm: On the Socioeconomics of Industrial Networks*. Ed. by Gernot Grabher. London & New York: Routledge. Chap. 12, 255–277.

Gramsci, Antonio (1948–1951). *Quaderni del Carcere*. Italian. Ed. by Felice Platone. Torino: Einaudi.

Greif, Avner (2006). *Institutions and the Path to the Modern Economy*. Cambridge University Press.

Hausmann, Ricardo and Dani Rodrik (Dec. 2003). "Economic Development as Self-discovery". *Journal of Development Economics* 72 (2), 603–633. DOI: 10.1016/s0304-3878(03)00124-x.

Istat (n.d.[a]). *Il valore della moneta in Italia dal 1861 al 2015*. URL: https://www.istat.it/it/archivio/184142 (visited on 03/18/2019).

— (n.d.[b]). *Serie storiche*. URL: http://seriestoriche.istat.it/ (visited on 03/17/2019).

Jacobs, Jane (1961). *The Death and Life of Great American Cities*. New York: Random House.

Klepper, Steven (2007). "The evolution of geographic structure in new industries". In: *Applied Evolutionary Economics and Economic Geography*. Cheltenham (UK): Edward Elgar. Chap. 4, 69–92.

Knight, Frank H. (1921). *Risk, Uncertainty and Profit*. Boston and New York: Houghton Mifflin.

Landesmann, Michael A. and Roberto Scazzieri (1996). "Introduction: Production and economic dynamics". In: *Production and Economic Dynamics*. Ed. by Michael A. Landesmann and Roberto Scazzieri. Cambridge: Cambridge University Press, 1–30.

Lane, David A. and Robert R. Maxfield (Mar. 2005). "Ontological Uncertainty and Innovation". *Journal of Evolutionary Economics* 15 (1), 3–50. DOI: 10.1007/s00191-004-0227-7.

Lehmann, Sune, Martin Schwartz, and Lars Kai Hansen (2008). "Biclique Communities". *Physical Review E* 78.1, 016108. DOI: 10.1103/physreve.78.016108.

Licini, Stefania (1998). "Banca e Credito a Milano nella Prima Fase dell'Industrializzazione (1840 - 1880)". Italian. In: *Antonio Allievi. Dalle Scienze Civili alla Pratica del Credito*. Ed. by Enrico Decleva. Laterza.

— (1999). "Ricchi, Ricchezza e Sviluppo Industriale. La Business Community Milanese dell'Ottocento". Italian. *Annali di Storia dell'Impresa* 10, 525–557.

MacLeod, Christine (1988). *Inventing the Industrial Revolution*. Cambridge (UK): Cambridge University Press. DOI: 10.1017/cbo9780511522673.

MacLeod, Christine and Alessandro Nuvolari (Feb. 2010). *Patents and Industrialization: An Historical Overview of the British Case, 1624-1907*. Laboratory of Economics and Management (LEM) Papers Series 04. Sant'Anna School of Advanced Studies.

March, James G. (Feb. 1991). "Exploration and Exploitation in Organizational Learning". *Organization Science* 2.1, 71–87.

Martin, Ron and Peter Sunley (2011). "Conceptualizing Cluster Evolution: Beyond the Life Cycle Model?" *Regional Studies* 45.10, 1299–1318. DOI: 10.1080/00343404.2011.622263.

Maskell, Peter and Anders Malmberg (2007). "Myopia, Knowledge Development and Cluster Evolution". *Journal of Economic Geography* 7.5, 603–618. DOI: 10.1093/jeg/lbm020.

Menzel, Max-Peter and Dirk Fornahl (Feb. 2010). "Cluster Life Cycles: Dimensions and Rationales of Cluster Evolution". *Industrial and Corporate Change* 19 (1), 205–238. DOI: 10.1093/icc/dtp036.

Metcalfe, J. Stanley and John Foster, eds. (2004). *Evolution and Economic Complexity*. Edward Elgar. ISBN: 1843765268.

Miller, John H. and Scott E. Page (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press.

Mokyr, Joel (1990). *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford: Oxford University Press.

— (1994). "Cardwell's Law and the Political Economy of Technological Progress". *Research Policy* 23.5, 561–574.

— (2000). "Evolutionary Phenomena in Technological Change". In: *Technological Innovation as an Evolutionary Process*. Ed. by John Ziman. Cambridge University Press. Chap. 5, 52–65.

— (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton, NJ: Princeton University Press.

— (2008). "The Institutional Origins of the Industrial Revolution". In: *Institutions and Economic Performance*. Ed. by Helpman. Harvard University Press.

— (2016). *A Culture of Growth*. Princeton University Press.

Nelson, Richard R. and Sidney G. Winter (1982). *An Evolutionary Theory of Economic Change*. Cambridge (MA): Harvard University Press.

Nohria, Nitin and Ranjay Gulati (Oct. 1996). "Is Slack Good or Bad for Innovation?" *The Academy of Management Journal* 39.5, 1245–1264.

North, Douglass C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

— (2005). *Understanding the Process of Economic Change*. Princeton University Press.

Nuvolari, Alessandro (May 2004). "Collective Invention During the British Industrial Revolution: The Case of the Cornish Pumping Engine". *Cambridge Journal of Economics* 28 (3), 347–363. DOI: 10.1093/cje/28.3.347.

O'Gorman Anderson, Benedict Richard (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. 1st ed. 1983. Verso.

Pasinetti, Luigi Lodovico (2007). *Keynes and the Cambridge Keynesians: A Revolution in Economics to Be Accomplished*. Cambridge University Press.

Penrose, Edith Tilton (1952). "Biological Analogies in the Theory of the Firm". *The American Economic Review* 42.5, 804–819.

Porter, Mason Alexander, Jukka-Pekka Onnela, and Peter J. Mucha (Feb. 2009). "Communities in Networks". *Notices of the American Mathematical Society* 56 (9).

Porter, Michael E. (1998). "Clusters and the New Economics of Competition". *Harvard Business Review* 76, 77–90. URL: https://hbr.org/1998/11/clusters-and-the-new-economics-of-competition (visited on 09/26/2018).

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Richerson, Peter J. and Robert Boyd (2005). *Not by Genes Alone: How Culture Transformed Human Evolution.* Chicago (IL): The University of Chicago Press.

Rodrik, Dani (Feb. 1996). "Coordination Failures and Government Policy: A Model With Applications to East Asia and Eastern Europe". *Journal of International Economics* 40 (1-2), 1–22. DOI: 10.1016/0022-1996(95)01386-5.

— (2004). *Industrial Policy for the Twenty-First Century.* Research rep. 4767. CEPR Discussion Paper.

Rugafiori, Paride (1994). "Ascesa e declino di un sistema imprenditoriale". Italian. In: *Storia d'Italia. La Liguria.* Ed. by Antonio Gibelli and Paride Rugafiori. Regioni. Einaudi.

— (1999). "Alle Origini della Fiat: Imprese e Imprenditori in Piemonte (1870-1900)". Italian. In: *Grande Impresa e Sviluppo Italiano: Studi per i Cento Anni della Fiat.* Ed. by Cesare Annibaldi and Giuseppe Berta. Vol. 1. 2 vols. Bologna: Il Mulino. Chap. 3, 135–183. ISBN: 8815070885.

— (2003). "Una città-fabbrica?" In: *Una Scuola, una Città: 1852–2002, i 150 Anni di Vita dell'Istituto 'Germano Sommeiller' di Torino.* Ed. by Angelo D'Orsi. ITCS Germano Sommeiller, 57–79.

Rugafiori, Paride and Chiara Ottaviano (2008). *La Business Community a Torino 1883–1907.* CD-ROM. Torino.

Schelling, Thomas C. (1990). *The Strategy of Conflict.* Cambridge (MA): Harvard University Press. ISBN: 0674840313.

Schumpeter, Joseph A. (1939). *Business Cycles. A Theoretical, Historical, and Statistical Analysis of the Capitalist Process.* 2 vols. New York & London: McGraw-Hill.

— (1942). *Capitalism, Socialism, and Democracy.* Harper & Brothers.

— (Nov. 1947). "The Creative Response in Economic History". *The Journal of Economic History* 7.2, 149–159.

Shiller, Robert J. (Apr. 2017). "Narrative Economics". *American Economic Review* 107.4, 967–1004. DOI: 10.1257/aer.107.4.967.

Strehl, Alexander and Joydeep Ghosh (2002). "Cluster Ensembles: A Knowledge Reuse Framework for Combining Multiple Partitions". *Journal of Machine Learning Research* 3, 583–617.

Tappi, Deborah (2005). "Clusters, Adaptation and Extroversion: A Cognitive and Entrepreneurial Analysis of the Marche Music Cluster". *European Urban and Regional Studies* 12.3, 289–307. DOI: 10.1177/0969776405056591.

Weaver, Warren (Jan. 1961). "A Quarter Century in the Natural Sciences". *Public Health Reports (1896-1970)* 76.1. A previous version of this paper can be found in the Rockefeller Foundation's annual report of 1958, 57–65. DOI: 10.2307/4591061.

Witt, Ulrich (2008). "Evolutionary Economics". In: *The New Palgrave Dictionary of Economics.* Ed. by Steven N. Durlauf and Lawrence E. Blume. London (UK): Palgrave Macmillan, 1873–1879.

A Description of the variables available in the database and some descriptive statistics

A.1 List of the available variables

Tables 3–4 describe the available variables. It must be noticed that in each row of the original data some variable refer to the person identified by `PersonID`, while some other to the company identified by `CompanyID`: here I split them in the three tables only for convenience. For the same reason, the table of variables referring to people is split in two.

Table 3: The available variables, common to companies and people, and their description. The data I added by scratch are marked with a *.

Var. name	Description
`ID`	Unique identifier of the row.
`ID_ATTO`	Unique identifier of the legal document from which the data of the row come from.
`ANNO_DEP`	Year in which the legal document was produced
`TIPO_ATTO`	Object of the legal document. See 8 for further information about the possible cases.

Table 4: The available variables that refer to a company and their description. The data I added by scratch are marked with a *.

Var. name	Description
`DEN_RAG`	Name of the company to which the data refer. I worked a little bit on these data so that changes in the name of the company do not impact on its history as a production unit. Shortly, I took the names in `NEWDEN_RAG` and I assign to any company with this new name the old name it has before the rename.
`TIPO_SOC`	It is the type of company (*forma societaria* in Italian). See 7 for further information about the possible cases.
`CAPIT_SOC`	It is the share capital of the company.
`CompanyID`*	It is a unique identifier for each company. It is based uniquely on the information provided by the `DEN_RAG` column. The first number is equal to $max($`PersonId`$) + 1$.
`CAT1level2`* `CAT2level2`* `CAT3level2`*	These three categories were created by scratch, based on the `CAT1`, `CAT2`, `CAT3` original ones and other information that Balbo provided me, not contained in the original data set. In short, they say the industry in which the company undertakes its activities. It is written like that "first level industry specification > second level industry specification" (only in some case there is only the first level). The categories are mainly based on the classification of the Italian official "Censimento Industriale" of 1911 (that is the closest to the period under analysis).

Table 5: The available variables that refer to a person and their description. The data I added by scratch are marked with a *.

Var. name	Description
`COGNOME`	Surname of the person to which the data refer. If it is a legal person this is its name (*Ditta* in Italian). It is the merge of the columns `COGNOME` and `DITTA` in the original data.
`NOME`	First name of the person to which the data refer. If it is a legal person this is NA by definition.
`SESSO`*	It is a categorical variable. It is "M" or "F" if the person is a male or a female, respectively; "Ditta" if the row refers to a legal person. "Famiglia" and "Eredi" refer to particular cases in which the row do not refer to a single person, but to a group of people: a family or the heirs of a person, respectively. Normally this last case happens only in the year of death of the shareholder of the company. The first two categories were already there in the original data. The third category was determined by looking at whether the original data in the column `DITTA` were NA or not. The last two categories were determined by looking at the surname of the person that looked like "John Doe (Ditta)".
`PATERNITA`	It is the name of the father of the person to which the row refers. It helps to identify uniquely a person.
`LUOGO_NA`	It is the place –most of the times it is the municipality– where the person was born. I cleaned it up the strings looking only at its internal consistency, in order to use it for uniquely identifying a person, more than to use it as a datum *per se*: if you want to know more precisely the municipality in which a person was born, you can find other useful information in the `LUOGO_NA-PLUS` column.
`COD_NAZN`	It says the nation in which the person was born. More or less it can be written as its citizenship. Theoretically, when it is NA it should be that the person was born in Italy.
`RESIDENZA`	It is the place where the person lives.
`PROF`	It is the job (broadly defined) of the person. See 9 for further information about the possible cases.
`NOTA.SOCIO`	Additional information about the person, like if he is the representative of someone else, if he is the heir of someone else previously involved in the company, and so on.

(continue ...)

Table 6: The available variables that refer to a person and their description. The data I added by scratch are marked with a *.

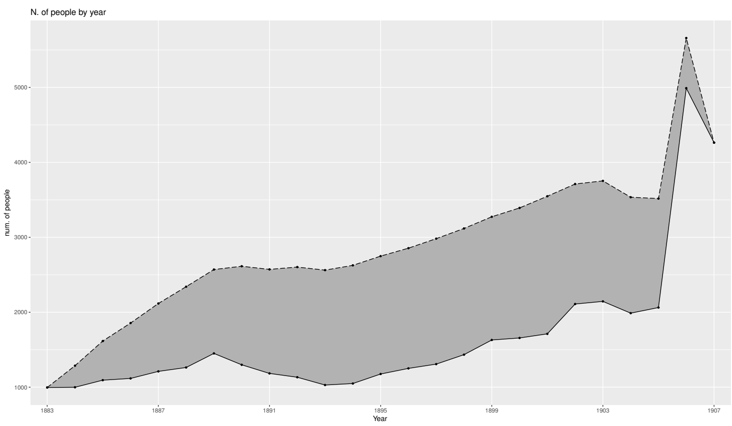
(...continued)

Var. name	Description
`RUOLO`	It is the role carried out by the person within the company. See 10 for further information about the possible cases. By the moment an edge among a person an a company exists regardless of the role this person carried out in the company. Obviously, this is a strong assumption, particularly because of the interpretative implications this choice has when the bipartite network is projected on one of the two layers. Indeed, not all of these roles can be considered equally likely to be the *underlying* of a relationship between two people or two companies, but it is not easy to decide which it is and which it is.
`TITOLI`	It is the noble or honorary title of the person. See 11 for further information about the possible cases. This is the only categorical variable with a precise hierarchy present in the data set. Indeed, assuming noble titles more relevant then honorary titles, you can put them in increasing order of relevance as: Cavaliere < Cavaliere ufficiale < Commendatore < Grand'ufficiale < Cavaliere ufficiale della Corona d'Italia < Commendatore ufficiale della Corona d'Italia < Cavaliere ufficiale della Legione d'onore < Nobile/Nobildonna < Cavaliere (cadetto di famiglia nobile) < Barone/Baronessa < Visconte < Conte/Contessa < Marchese/Marchesa < Duca < Principe.
`PersonID`*	It is a unique identifier of each person. It is based on the information provided by the following columns: `NOME`, `COGNOME`, `PATERNITA`, `LUOGO_NA`. When in a column there is more than one different case and some rows have a missing value, all these last rows are collected together. E.g., suppose there were three people with the same name and surname. Suppose we know the name of the father of two of them, and these two last names are different from one another. Assume that the third name is unknown. Then each row will have a different PersonId number. This case was reduced as much as possible trying to fix all the possible errors in the data.

A.2 Descriptive statistics

Entrepreneurs and enterprises We can see a common trend both in the number of people and companies (Fig. 10-11): a little increase rate in the first years; than a decrease in the years of the crisis; a faster and faster recovery; and in the last years a peak (1906) and a new decline (1907).

Figure 10: Number of people (entrepreneurs) observed in the data set in each year. The solid line shows the observations as they are in the data. Conversely, the dashed line shows the number of people by year assuming that each person were present in any year between the first and the last time it is observed in the data. The "truth" is somewhere in between (grey area).



Legal form of the companies Tab. 7 shows the distribution of the legal forms of the companies contained in the database. Looking at Fig. 12, you can see a clear shift from *Società in Nome Collettivo* (SNC) to *Società Anonima* (SA)[28] and, partially, *Società in Accomandita Semplice* (SAS).

Industrial sectors In terms of the number of enterprises, Fig. 13 clearly shows a decrease in the dimension of the textiles sector and, conversely, a strong increase in the metalmaking industry. The decrease in the banking sector is due to the fact that there was a restructuring of the sector with a concentration of the companies in Milan: that fact does not mean a decrease in activities of banks and insurance companies in Turin, but only that these companies were based in Milan, so that they were no longer obliged to declare their activities to the Court of Turin.

Number of links In Fig. 14 you can see the number of nodes and links by wave. All the plots show a similar trend: an increase; then a decrease in the years after the crisis; a rebound; and a strong recovery in the last years.

Degree distribution To have a visual inspection of the distribution of the degree, I plotted the degree distribution of three sample waves (1885–1886,

[28]They corresponds more or less to what is nowadays called *Società per Azioni* (SpA).

Figure 11: Number of companies (enterprises) observed in the data set in each year. The solid line shows the observations as they are in the data. Conversely, the dashed line shows the number of companies by year assuming that each company were present in any year between the first and the last time it is observed in the data. The "truth" is somewhere in between (grey area).

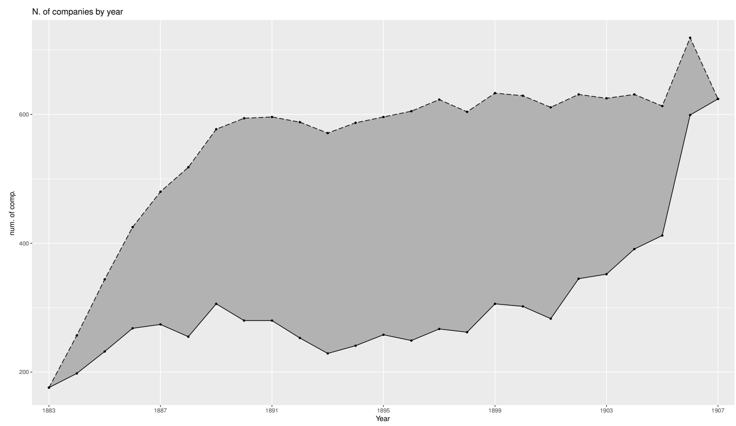


Table 7: The types of legal forms each company can have and the share of each.

	Type	Share
1	Società anonima	44.12
2	Società in nome collettivo	24.70
3	Società in accomandita semplice	9.75
4	Società in accomandita per azioni	4.05
5	Società anonima in liquidazione	3.45
6	Società in accomandita per carature	2.41
7	Società di fatto	1.97
8	Ditta individuale	0.78
9	Società in nome collettivo di fatto	0.28
10	Società in nome collettivo in liquidazione	0.25
11	Società in nome collettivo per carature	0.25
12	Società in accomandita semplice in liquidazione	0.21
13	Società cooperativa	0.07
14	Società in accomandita per azioni in liquidazione	0.06
15	Associazione in partecipazione	0.02
16	Società in accomandita semplice di fatto	0.02

Figure 12: Proportion of companies by legal type in each year.

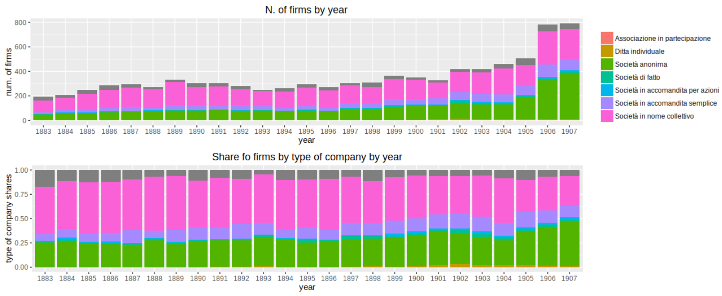


Figure 13: Share of firms in each industrial sector by year.

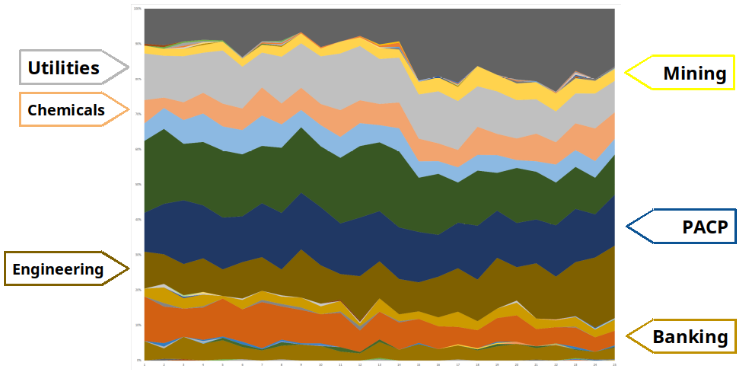
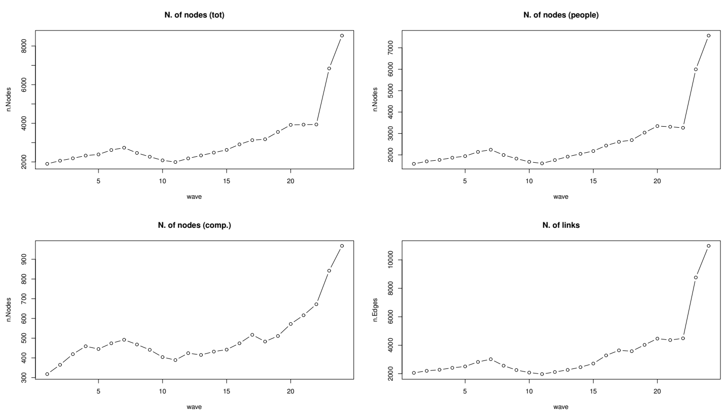


Figure 14: Number of nodes and links by wave.



1895–1896, 1905–1906): see Fig. 15–17.[29]

Figure 18 shows the trend of the Gini index, i.e. the inequality in the degree distribution of the two layers of the network. The inverse of this value is compared with the estimated value of the power law coefficient (see above). The use of the Gini index of the degree distribution is interesting because, instead of the estimated value of the power law coefficient, it does not assumes *a priori* any particular form of the distribution. The strong correlation among the two measures, at least for the *people* side, induces us to be more confident about the fact that the distribution is well approximated by a power law.

Figure 15: Degree distribution for each of the layers for the wave 1885-1886.

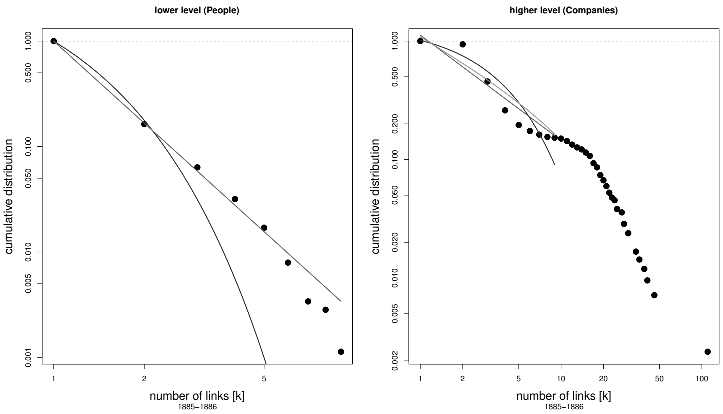
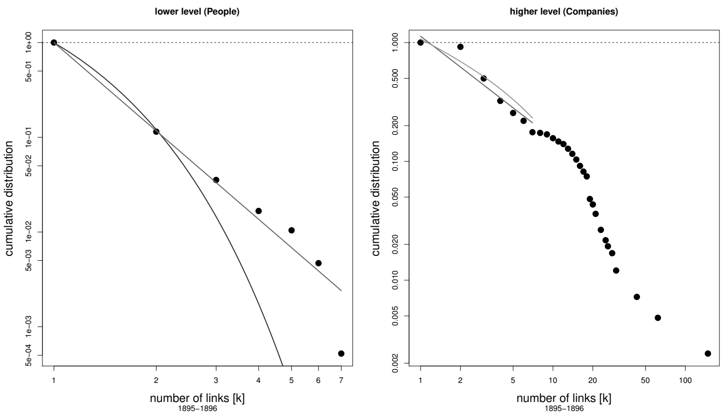


Figure 16: Degree distribution for each of the layers for the wave 1895-1896.



Document type Tab. 8 reports the most importat type of legal acts contained in the database. The most represented ones are the establishments of new

[29]While most of the other times I used the `igraph` R package, here I used the `bipartite` package, which estimates the coefficients of the distributions following the procedure proposed in Aaron Clauset et al. (Nov. 2009). "Power-Law Distributions in Empirical Data". *SIAM Review* 51.4, 661–703 (Csardi and Nepusz 2006; Dormann 2011; Dormann et al. 2009, 2008; R Core Team 2018).

Figure 17: Degree distribution for each of the layers for the wave 1905-1906.

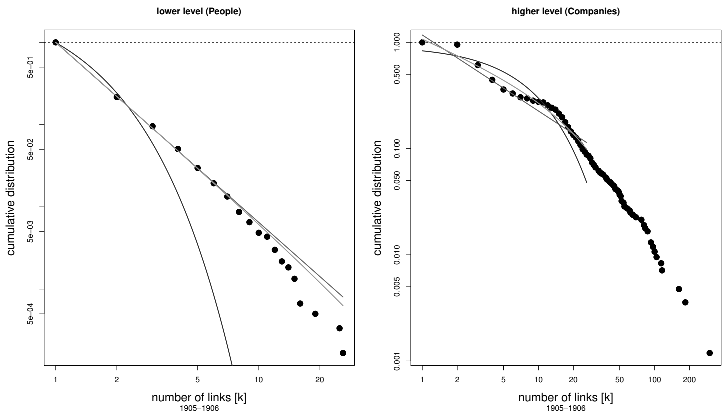
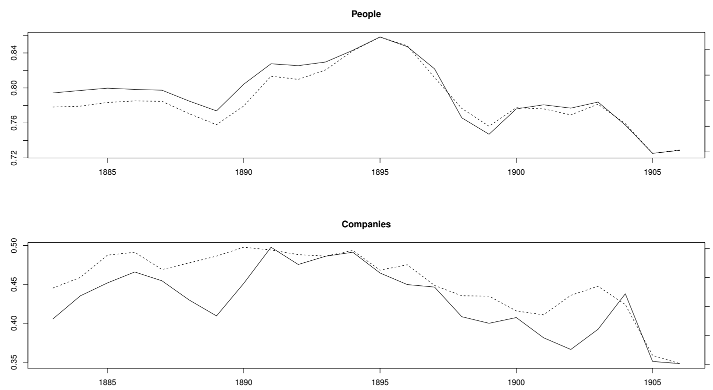


Figure 18: The solid line represents the inverse of the Gini index by wave. The dashed line refers to the estimated power law exponent by wave.



companies (*costituzione*), followed by the acts that companies have to do each year during the annual shareholders meeting (*verbale*, *relazione CdA*, *bilancio*) and by the dissolution of the company (*scioglimento*).

Table 8: The types of object of legal documents and the share of each (only these with share $>$ .5). Overall, there are 286 possible cases (NA included).

	Type	Share
1	Costituzione	31.99
2	Verbale di assemblea ordinaria, relazione cda, presentazione del bilancio	14.36
3	Scioglimento	11.00
4	Vcda - Assegnazione cariche direttive	2.80
5	Accettazione di nomina a consigliere	2.63
6	Conferimento procura	2.55
7	Esclusione o recesso	2.03
8	Vas - Scioglimento e messa in liquidazione	1.70
9	Proroga	1.61
10	Vas - Riduzione capitale	1.32
11	Vao e vas - Modifiche minori	1.17
12	Vcda - Aumento capitale	1.10
13	Vas - Aumento capitale	1.04
14	Vas - Modifiche minori	0.99
15	Messa in liquidazione e nomina liquidatore	0.85
16	Ammissione socio e aumento capitale	0.83
17	Vao e vas - Riduzione capitale	0.78
18	Vao e vas - Trasferimento sede	0.70
19	Vao e vas - Aumento capitale	0.65
20	Vcda - Precisazione poteri di figure direttive	0.63
21	Vas - Trasferimento sede	0.62
22	Vao e vas - Scioglimento e messa in liquidazione	0.62
23	Vao e vas - Delibera di aumento di capitale	0.59
24	Vas - Continuazione società	0.59
25	Vas - Proroga società	0.56

Occupations Tab. 9 shows the most represented occupations of the people nominated by the legal documents. The distribution is very disperse and skewed, also because the names do not follow an official classification, but were self-representations of the people themselves.

Roles Tab. 10 shows the reasons why people were nominated in the legal documents (i.e., the role they had). Also in this case, as in the previus one, the number of possible types is very large.

Noble and honorary titles Tab. 11 reports distribution of the noble and honorary titles of the people nominated in the legal acts.

Table 9: The types of occupation people do and the share of each (only these with share > .5). Overall, there are 452 possible cases (NA included).

	Job	Share
1	Commerciante	6.15
2	Avvocato	5.13
3	Ingegnere	4.58
4	Industriale	3.55
5	Negoziante	2.32
6	Possidente	1.81
7	Dottore	1.72
8	Ragioniere	1.70
9	Benestante	1.13
10	Geometra	0.78
11	Professore	0.62
12	Proprietario	0.59
13	Impiegato	0.56
14	Banchiere	0.50

Table 10: The types of role people do and the share of each (only these with share > .5). Overall, there are 240 possible cases (NA included).

	Role	Share
1	Consigliere	7.96
2	Liquidatore	5.25
3	Sindaco	5.07
4	Accomandante	4.78
5	Scrutatore	3.96
6	Sindaco supplente	3.74
7	Accomandatario/Gerente	3.03
8	Mandatario generale/Institore	1.83
9	Legale rappresentante	1.46
10	Scrutatore e maggiore azionista	1.35
11	Presidente CdA	1.15
12	Segretario assemblea	1.10
13	Direttore amministrativo	1.02
14	Direttore tecnico	0.89
15	Presidente assemblea e CdA	0.77
16	Direttore	0.74
17	Gestore cassa e corrispondenza, amministratore contabilità	0.74
18	Amministratore delegato	0.69
19	Vicepresidente CdA	0.66
20	Presidente assemblea	0.56
21	Segretario CdA	0.54

Table 11: The types of noble or honorary title people have and the share of each.

	Title	Share
1	Cavaliere	8.83
2	Commendatore	2.34
3	Conte/Contessa	1.42
4	Cavaliere ufficiale	0.54
5	Barone/Baronessa	0.50
6	Marchese/Marchesa	0.46
7	Nobile/Nobildonna	0.29
8	Cavaliere (cadetto di famiglia nobile)	0.13
9	Principe	0.03
10	Cavaliere ufficiale della Corona d'Italia	0.02
11	Duca	0.02
12	Cavaliere ufficiale della Legione d'onore	0.02
13	Grand'ufficiale	0.02
14	Commendatore ufficiale della Corona d'Italia	0.01
15	Visconte	0.01

On Scaling of Technological Knowledge Production
How Population Size Affects the Knowledge Recombination Capacity of EU Metropolitan Regions

Carlo Bottai

Department of Economics and Statistics *Cognetti de Martiis*, University of Turin
BRICK, Collegio Carlo Alberto

Abstract

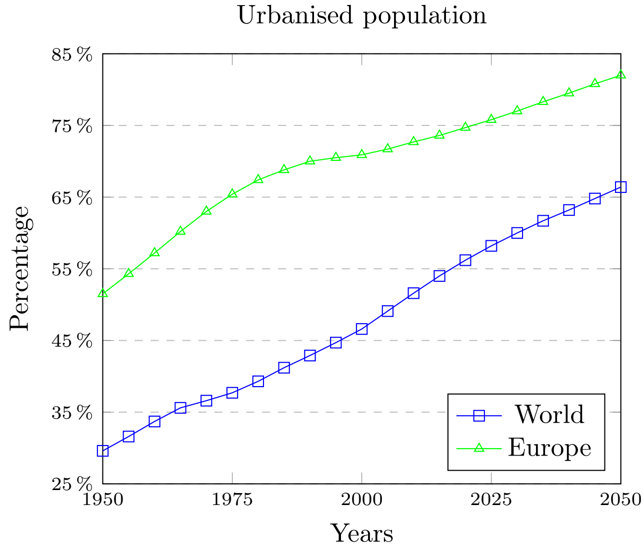
Cities are major *inventions* producers, and the *urban scaling* literature has shown super-linear relationships between their population size and invention performance. However, we still know very little about why the production of some types of technological knowledge agglomerate more than others. This paper estimates the scaling patterns of European Metropolitan Regions in terms of patenting activity. Moreover, it shows that patents with a higher *diversity* of technological domains in their *backward citations* tend to be more concentrated than others in large urban centres. The latter result is compatible with a recently proposed model that explains the emergence of scaling laws combining insights from Complexity Economics and Cultural Evolution. Only large populations are able to accumulate and preserve a *diversity* wide enough to have all the elements needed by the more sophisticated recipes.

Keywords— Agglomeration economies, Urban Economics, Knowledge-based economies, Recombinant innovation, Division of knowledge, Diversity, Europe, Urban scaling, Metropolitan areas, Technological classes, Patents

1 Introduction

The role of cities and metropolitan areas for socio-economic development has attracted increasing popularity in recent years, both as a subject and object of policy action and as a tool for scientific research (UN-Habitat 2016; OECD 2015). To understand why, it is perhaps sufficient to say that the world's urbanised population increased from about 30% in the 1950s to about 50% in the 2010s, and this share is projected to keep growing, even in European countries (Fig. 1). Moreover, cities and metropolitan areas have been recognised as major economic contributors in many national economies, and as key players of the global markets. For example, as shown by Tab. 1, across EU15 countries, metropolitan areas with more than 500,000 inhabitants, although cover only 10% of the land, but account for roughly 40% of the population and about 45% of gross domestic product (GDP).[1]

Figure 1: Urbanised population data and projections. Source: UN data.



In particular, cities are primary *loci* of production of *innovation*, that since long is considered as the key endogenous engine of economic growth and development (Schumpeter 1911). Specifically, in explaining the capacity of an economy to introduce novelties in the system, the Endogenous Growth theory has assigned a pivotal position to *knowledge spillovers* among individuals and firms (Freeman and Soete 1997; Lucas 1988; Romer 1990). And these *spillovers* have been proven to be geographically localised (Breschi and Lissoni 2003; Jaffe et al. 1993; Verspagen and Schoenmakers 2004), so that they are expectedly stronger within cities (Glaeser et al. 1992).

Looking specifically at technological knowledge, it has been shown that most of its production happens in few larger metropolitan areas and the production of those industries where new economic knowledge plays an important role tend to be more geographically concentrated, mainly due to the influence of knowledge spillovers on the production management (Audretsch and Feldman 1996). For example, Tab. 1 shows that the Metropolitan Regions of the EU15 produce on average 57.7% of the European patent applications. Likewise, Wichmann Matthiessen et al.

[1] As underlined by the OECD (2015, p. 37), the "success" of a city can be defined in many different ways, and a metropolitan area that performs well in one of these dimensions can do differently under other points of view. Here we are only speaking about the economic performance of cities, with a specific focus on technological knowledge production.

Table 1: OECD-EU Functional Urban Areas (FUAs) with more than 500,000 inhabitants. EU15 (Luxembourg excluded). 2000-2012 average values. Cumulative sum of the national value shares (above) and min. and max. values (below). Source: OECD (2013). For patents, only the data come from the Eurostat and refer to the Metropolitan Regions.

Nation	FUAs	Pop.	GDP	Surf.	Productivity	Patents
	N	%	%	%	ratio	%
Austria	3	45.8	51.2	19.0	1.08	47.2
		7.1 - 31.4	7.2 - 36.2	3.7 - 11.0	1.03 - 1.17	
Belgium	4	44.4	52.1	22.6	1.03	60.3
		5.3 - 22.6	4.7 - 31.2	3.4 - 10.8	0.80 - 1.40	
Germany	24	38.3	44.4	15.0	1.18	57.9
		0.6 - 5.3	0.6 - 5.4	0.1 - 1.8	0.87 - 1.63	
Denmark	1	35.9	40.8	9.5	1.11	74.9
		—	—	—	—	
Greece	2	41.1	51.4	2.3	1.13	
		8.6 - 32.5	7.0 - 44.5	1.1 - 1.3	0.90 - 1.36	
Spain	8	35.6	40.8	6.8	1.08	
		1.5 - 13.8	1.3 - 18.4	0.2 - 2.4	0.91 - 1.27	
Finland	1	26.9	36.2	2.1	1.17	47.5
		—	—	—	—	
France	15	39.9	51.2	10.1	1.04	
		0.8 - 18.6	0.7 - 29.7	0.2 - 2.2	0.87 - 1.49	
Ireland	1	36.3	46.6	7.0	1.23	56.9
		—	—	—	—	
Italy	11	30.9	35.7	6.7	1.16	45.1
		0.9 - 6.8	0.7 - 10.8	0.2 - 1.9	0.92 - 1.47	
Netherlands	5	36.4	41.7	19.0	1.16	68.0
		4.2 - 13.9	4.1 - 16.2	0.9 - 8.3	1.02 - 1.21	
Portugal	2	38.4	49.1	5.4	1.20	
		12.3 - 26.0	11.7 - 37.4	1.0 - 4.3	0.94 - 1.46	
Sweden	3	37.4	44.9	3.5	1.06	69.2
		7.0 - 21.0	6.2 - 29.4	0.8 - 1.7	0.92 - 1.29	
UK	15	40.1	47.3	9.4	1.00	50.3
		0.8 - 18.5	0.6 - 26.8	0.2 - 2.8	0.80 - 1.41	
	95	37.7	45,2	9.9	1.12	57.7

(2010) have shown that most of the scientific research is produced within the largest cities of the World. Moreover, as *social reactors* –to use the words of Bettencourt (2013a)– cities are dynamic and powerful entrepreneurial environments, where the *creative class* and STEM people concentrate and generate positive spillovers (Brunow et al. 2018; Florida 2002). And, more in general, not only have highly educated people, that are disproportionately more present in cities, a higher productivity by themselves, but also produce positive spillovers that increase others' productivity, being these last high or low skilled workers, no matters.

This confirms the key role of cities also within the *knowledge-based economies*. In this context, the creation and utilisation of knowledge become the key factors affecting the competitiveness of firms, regions, and countries (Freeman and Soete 1997), and more specifically *codified knowledge* raised in relevance compared to the past (Abramovitz and David 1996). As underlined by Porter (1998, p. 78), «[c]ompetition in today's economy is far more dynamic. Companies can mitigate many input-cost disadvantages through global sourcing, rendering the old notion of comparative advantage less relevant. Instead, competitive advantage rests on making more productive use of inputs, which requires continual innovation. Untangling the paradox of location in a global economy reveals a number of key insights about how companies continually create competitive advantage. What happens *inside* companies is important, but clusters reveal that the immediate business environment *outside* companies plays a vital role as well». Concluding, innovation and competitive success in many fields are geographically concentrated. Moreover, as underlined by Arthur (1996), the more advanced and valuable industries of knowledge-based economies are characterised by *increasing returns to scale*, so that production can concentrate in few, small and dense places like cities, but also that production concentrate in few of them much more than in all the others. In other words, if knowledge-based economy promised to make all the places equal, it has actually made some more equal than the others.

More in general, there is a clear, lasting, and strong link between urbanisation processes and economic development (OECD 2015, Ch. 1). But, if this connection is undoubted, determining whether urbanisation causes economic growth or, conversely, whether the opposite direction holds is very difficult to untangle. Duranton and Puga (2004) have summarised in the categories of *sharing*, *matching* and *learning* the advantages that derive to firms from *agglomeration*. In the opposite direction, the disproportionately high concentration of knowledge, innovation and economic activities (both in size and diversity) in cities increases their attractiveness for educated, highly skilled, entrepreneurial and creative individuals who, by locating in urban centres, increase in turn the knowledge spillovers (Feldman and Florida 1994; Glaeser 1999). In the end, as underlined by Bettencourt et al. (2007b, p. 108), «[t]his seemingly spontaneous process, whereby knowledge produces growth and growth attracts knowledge, is the engine whereby urban centres sustain their continuous development through unfolding innovation», so much that it is probably more effective and useful to describe this positive association as a *feedback-loop* phenomenon, of the kind typically produced by Complex Adaptive Systems, as cities have been described (Batty 2008; Bettencourt and West 2010; Jacobs 1961). This means that we cannot analyse the system but as a whole: under the theoretical point of view, it is surely relevant to analyse each cause separately; but empirically this can be roughly impossible, and maybe it can be more effective to look at something more aggregate and subtract it all at once.

In more details, according to the so-called *urban scaling* literature, the relationship between their size and many macroscopic variables turns out to be describable by power-law distributions (Bettencourt 2013b; Bettencourt et al. 2007a). This suggests that general scale-free laws may govern cities agglomeration dynamics, while geographical, historical and cultural features play only a secondary (insofar as decisive) role (Bettencourt and West 2010, p. 912). In particular, it has been shown that many economic phenomena that require the recombination of existing knowledge scale super-linearly with city size. But, the greater knowledge diversity of large cities also provides more opportunities for distinct knowledge combinations and for the *exploration* of new combinations, so that larger cities, usually, host more different economic activities than smaller towns (See, among others, Mori et al. 2008; Schiff 2015; Youn et al. 2016). Therefore, if the literature is concordant in showing increasing returns from population size in cities, we still have much to learn about which type of economic activities concentrate more than others and why they do so.

This paper focuses in particular on the scaling of technological knowledge, as captured by patent production, in the EU larger cities,[2] and the rest of the paper is organised as follows. Sec. 2 presents the biological allometry literature. A specific focus is on a recent contribution by West, Brown and Enquist, to which the urban scaling literature is strongly connected, so that this last is then introduced by analogy. It follows the presentation of alternative models that explain the urban scaling laws observed (Sec. 3). In particular, I will focus on a theory recently proposed by Gomez-Lievano et al. (2016), that better fits with the narrative and empirical findings of this paper, since it accommodates the existence of a multiplicity of scaling laws for the same macro-phenomena, due to the quantity of capabilities required by each sub-domain to be produced (or generated). Sec. 4, following a by then well-established literature, shows that European cities show a super-linear relationships between their population size and invention performance. Sec. 5 focuses on the scaling of technological knowledge and opens the black box even more, showing that the production of some types of technological knowledge agglomerate more than others. The empirical analysis shows that, in line with previous works (Balland et al. 2018; Mewes 2019), technological domains whose production require the recombination of a larger number of complementary useful knowledge components *scales* more than other domains. Lastly, the main conclusions of the paper are summarised and highlighted, and its limitations analysed.

2 Urban Scaling

More, if not all, complex systems tend to show scaling patterns. That is, to show power-law correlations or distributions of some key characteristics of the system. These properties are the signal of dynamic processes in place. More specifically, the idea behind *scaling* is to look at how properties of objects change when their size changes; or more in general to look at some proportionality between given properties of such objects. If these relationships do not change at different scales of the objects we can say that they are *scale-invariant* or *self-similar*.[3] A kind of such self-similar properties, known as *allometry*, is usually expressed as

$$Y = Y_0 M^{\beta} \tag{1}$$

where Y is some observable quantity of the phenomenon of interest (like metabolic rate in Biology, or GDP in Economics), Y_0 is a normalisation constant, M is the scale of the object considered (like the mass of an organism, or the population of a city), and β is the scaling exponent.

[2]Although I am 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 important types of inventions cannot be patented at all (Griliches 1990; Pavitt 1985), I nevertheless see, in line with the literature, patents as a useful proxy for invention in empirical studies (Hall et al. 2005; Jaffe and Trajtenberg 2002; Kortum and Lerner 2000).

[3]Two famous examples that show such property are *fractals* and *power-laws*, so that it is quite common to speak about fractal-like properties or to refer to power-laws in a broader sense, even though in the narrow sense the objects considered are not truly scale-invariant. This is even more true in the case of empirical power-laws, where it is usual to employ this word to refer to anything that shows some linear in a log-log plot. As a matter of fact, with a finite amount of data it is very hard, if not impossible, to distinguish them for other processes, like a log-normal, that mimic this signature behaviour, so that, with real data, such straightness is a necessary, but not sufficient, condition for the data following a power-law relation.

Biological allometry In biology, similar allometric relationships have been known since long.[4] In particular, a well known one, named *Kleiber's law*, says that the metabolic rates (or rate of energy use) of mammals –and a wider range of living things, more in general– scale according to their masses raised to a $\frac{3}{4}$ power (Kleiber 1932).[5] This relationship can be surprising at first glance. Considering that all biological organisms are made up of essentially the same basic materials and structures, with cells growing linearly with body mass, one might expect a $\beta = 1$. On the basis of conventional Euclidean geometric scaling, a second plausible guess could be to observe something proportional to $\frac{1}{3}$.[6] In particular, since the organism has to disperse its metabolic heat through the surface of its body, one might expect a $\beta = \frac{2}{3}$, on the basis of a naïve surface-to-volume relationship. However, as underlined by Brown et al. (2000, p. 4), «organisms do not usually exhibit such simple geometric scaling. This is because there are powerful constraints on structure and function that do not allow organisms to maintain the same geometric relationships among their components as size changes over several orders of magnitude». If this is a decent explanation for *allometry*, instead of *isometry*, the why about the quarter-power scaling widely observed in biology questioned researchers for more than half a century.

A nowadays largely accepted theoretical explanation has been provided in 1999 by West, Brown and Enquist (WBE).[7] Their evolutionary-based arguments rest upon two conjectures: that organisms have been selected to maximise fitness by maximising metabolic capacity, and that this has been achieved by stretching the surfaces through which it happens an exchange of resources between the organisms and its environment; and that this happens under the constraint of maintaining a compact shape, so that the time and resistance for delivery of resources to the whole body is minimised. In particular, their theory grounds on three fundamental hypotheses about the "circulatory system" of animals and plants –i.e., the network through which they convey energy, nutrients and other materials throughout their entire body. They stated that those networks (i) branch to reach all parts of an organism, producing a fractal-like shape;[8] (ii) have terminal units, like capillaries, that do not vary with size; and (iii) have evolved so that total resistance is minimised, so to reduce the energy required to distribute resources. Moving from these premises, they reach to the empirically observed quarter-power sub-linear exponents above said. In conclusion, they argue that the larger the organism, the more efficient the system that can be constructed to provide energy. However, this incremental efficiency is constraint by the interplay between some physical and geometric constraints implicit in the principles above said; and in particular by the fact that, unlike true fractals, self-similarity of living organisms, even if true at many orders of magnitude, is not true endlessly, being limited both from the bottom and from the top.

The significance and usefulness of scaling laws can be appreciated looking at the final words of the authors themselves:

> On the one hand, this is the testimony to the power of natural selection, which

[4] The term *allometry* itself was coined by the biologist Huxley in the 1930s.

[5] Over time also other biological rates and times –such as heart rates, reproductive rates, blood circulation times, and lifetimes– have been observed to scale with quarter powers; mainly $\beta = \frac{3}{4}$, $\pm\frac{1}{4}$ and $\frac{3}{8}$, depending on the variable considered. See Brown et al. (2000) for a review.

[6] In a perfectly *isometric* object –like a sphere, or any objects of self-similar shapes– all volume-based properties would scale proportionally to its body mass, all area-based properties would change with mass to the power of $\frac{2}{3}$, and all length-based properties would vary with mass to the power of $\frac{1}{3}$; assuming body mass to be a 3-dimensional property, and being the others 3-, 2- or 1-dimensional properties, respectively.

[7] See also West et al. (2000) for an exhaustive exposition of the model.

[8] More in general, this hypothesis follows from the fact that such a space-filling hierarchical branching structure would let the network to supply the entire volume of the organism. It is possible to find a full explanation of how such a structure implies a fractal-like one in West et al. (2000).

has exploited variations on this fractal theme to produce the incredible variety of biological form and function. On the other hand, it is testimony to the severe geometric and physical constraints on metabolic processes, which have dictated that all of these organisms obey a common set of quarter-power scaling laws. Fractal geometry has literally given life an added dimension —West et al. 1999, p. 1679.

Said differently, there are two tangled ways to look at *allometry*. Sometimes the differences among organisms that, at first sight, seem very profound, may be traced back to unique and simple causes. And it can be of interest to the researcher to highlight these "universal laws", as in WBE and other competing models –i.e., the *scaling* is «a tool for revealing underlying dynamics and structure» as highlighted by Bettencourt et al. (2007a, p. 7302). At the same time, the exclusion and removal of that common cause allow highlighting some more profound and less striking differences among these objects. That is, scaling laws can be seen as a *null model* against which to compare and test each empirical case, without the use of *a priori* theoretical assumptions, as we do whenever we look at *per capita* indicators (see e.g., Alves et al. 2015; Bettencourt et al. 2010; Katz 2006; Lobo et al. 2013). Then, following the so-called *Method of residues*, a second step will be to look at explanations for the deviations from the expectations based on the general laws just said (Coleman 1964, ch. 15).

Urban allometry More recently, West, Bettencourt, and other colleagues have found that cities, which have long been compared to living things, obey scaling laws similar to the ones above described for biological systems (Bettencourt et al. 2007a, 2010).[9] So they can conclude that «despite appearances, cities are approximately scaled versions of one another» and that «size is the major determinant of most characteristics of a city; history, geography and design have secondary roles» (Bettencourt and West 2010, pp. 913, 912). However, even though cities are Complex Adaptive Systems that integrate countless elements and constituent parts in precarious and fragile balances not less than the biological systems in which the scaling exponents have been originally observed,

> [cities] are much more than giant organisms or anthills: they rely on long-range, complex exchanges of people, goods and knowledge. They are invariably magnets for creative and innovative individuals, and stimulants for economic growth, wealth production and new ideas – none of which have analogues in biology —Bettencourt and West 2010, p. 913.

Nevertheless, as in biology with body mass, it has been empirically shown that many socio-economic urban phenomena grow super-linearly with population size, while the use of total resources tends to scale sub-linearly with it. Or, more in general, we can say that we observe *scaling*, in the sense that the counts of people engaged in (or suffering from) each phenomenon scale as a power of population size. In particular, Bettencourt et al. (2007a) have identified a tripartite taxonomy:

Linear The macro-variables usually associated with individual human needs (job, house, household water consumption) show an exponent $\beta \simeq 1$, i.e. in economic terms, these variables are, roughly, population-size inelastic.

Sub-linear Material quantities display *economies of scale* associated with infrastructure, so that the $\beta < 1$.

[9] To be more precise, *urban allometry* has already been studied in a more distant past –see e.g., Batty and Longley 1994, ch. 9–, but it is only in the last decade that the topic has been analysed intensively and widely.

Super-linear Socio-economic dimensions associated with the intrinsically social nature of cities, such as information, innovation or wealth, show *increasing returns* with population size, which means a $\beta > 1$.

But, while in biology the network principles underlying the *economies of scale* imply that the pace of life slows down with size, and constrain growth –i.e., animals reach a stable size at maturity– (West et al. 2001), the pace of urban life increases with population size, and cities growth boundlessly.[10] Continuous adaptation, not equilibrium, is the rule, and this is basically due to the feedback-loop mechanisms, based on the social interactions, that happens within cities (Bettencourt et al. 2007a, pp. 7304–7305). Moreover,

> [t]he most striking feature of the data is perhaps the many urban indicators that scale superlinearly ($\beta > 1$). These indicators reflect unique social characteristics with no equivalent in biology and are the quantitative expression that knowledge spillovers drive growth, that such spillovers, in turn, drive urban agglomeration, and that larger cities are associated with higher levels of productivity —Bettencourt et al. 2007a, p. 7303.

3 Theoretical foundations of the urban scaling laws

Many different mechanisms have been proposed to explain the origins of scaling. These explanations can be seen as competing, or as complementary ways to look at a faceted issue from different perspectives.[11] In Economic Geography, scaling laws are often attributed to productivity enhances that comes from *economies of scale*, labour mobility (high-skilled labour, in particular), *knowledge spillovers*, and other effects resulting from the *economies of agglomeration* (Krugman 1991; Storper 2011). More recently, other models that look at cities as Complex Adaptive Systems have been suggested. A first group describes scaling as the result of how lines relate to surfaces, and surfaces to volumes. One of the most famous theories that explain biological scaling, that we have just briefly reviewed, is part of this group (West et al. 1999). But, even though also some urban scaling theories belong to it, they are few and mostly related to physical infrastructures (Samaniego and Moses 2008). Conversely, most of the urban scaling literature belong to a second group of theories that describe the underlying phenomena as a complex network and derive the scaling properties from the fact that interactions in a network are expected to grow faster than linearly with the number of agents (Arbesman et al. 2009; Bettencourt 2013b; Pan et al. 2013). However, as highlighted by Gomez-Lievano and Patterson-Lomba (2018), both those groups of theories are not able to account for different scaling patterns across phenomena, if not postulating the existence of a specific and peculiar network structure for each phenomenon, different from the others.

On the other hand, there exist at least two other theoretical explanations that let us account for different scaling patterns for similar phenomena in an easier way. Pumain et al. (2006) have proposed an evolutionary theory of urban systems that leads to the scaling laws empirically observed. In their model, the largest cities became larger because successful in adopting many successive innovations. The feedback-loop remembered above, that links urbanisation processes

[10] To be clear, both the urban life pace acceleration and the open-ended growth, are not positive in themselves. There is also the downside, well documented by the "urban scaling" branch: for example, diseases spread faster; and the rate of major innovations has to accelerate, but this comes up against the limited available resources (Bettencourt et al. 2007a). In the end, cities are the solution to many current humankind problems; but they are open issues, at the same time.

[11] I am more in favour of this last point of view. Moreover, since a comparison of these alternatives is far beyond the scope of this paper, we will assume it as true in what follows.

and economic development, is such that knowledge produces growth and growth attracts knowledge, and is the engine by which cities persist in their development path through unfolding innovation. If we classify technologies in *leading*, *mature*, and *old*, based on the stage of the life cycle that they have reached, the authors posit that the technologies at the top of current innovation cycle will show a super-linear scaling pattern (i.e., *increasing returns* to urban scale), while to more mature ones will be subject to *constant* or *decreasing returns*, so that they will be more equally distributed within the urban system (or even, they will be more concentrated in the smaller urban centres).

Lastly, Gomez-Lievano et al. (2016) have recently proposed a further possible explanation, closely related to the model formalised by Hausmann and Hidalgo (2011). The authors bring together ideas from Economic Complexity (Arthur 1999; Fleming and Sorenson 2001; Hidalgo and Hausmann 2009; Metcalfe 2010) and Cultural Evolution literature (Boyd et al. 2013; Henrich 2015; Henrich et al. 2016; Richerson and Boyd 2005). The first key hypothesis of this model is that socio-economic phenomena occur only if in a given environment a multiplicity of complementary inputs are simultaneously brought together and combined following a specific recipe, and its appearance will be all the more likely the more each of these elements is present. (Hausmann and Hidalgo 2011; Kremer 1993). The environment is characterised by an intense *division of knowledge* and each person is the carrier of some of these specific and complementary pieces. In so doing, the authors assume that any urban phenomenon is an *emergent property* of a complex system of interacting individuals, and cannot be ascribed to any specific person. Indeed, as Richerson and Boyd remind us

> Even the greatest human innovators are, in the great schema of things, midgets standing on the shoulders of a vast pyramid of other midgets. The evolution of language, artifacts, and institutions can be divided up into many small steps, and during each step the changes are relatively modest. No single innovator contributes more than a small portion of the total, as any single gene substitution contributes only marginally to a complex genetic adaptation. The limited imitative capacity of other animals seems to prevent the cumulative evolution of complex cultural feature
> —Richerson and Boyd 2005, p. 50.

This leads to look at the *inventive* and *innovative* processes more as of *collective learning* than of *individual* or *social learning* (Gomez-Lievano et al. 2019a,b; Vanberg 1994). It is not how much each person knows that matters, but how much the system collectively knows (or, how many different *letters* are available in the system). And the mechanism of coordination provided by the environment in which specialised individuals live has a key role, as well. In other words, systemic conditions matters, and only *organised complexities* provide the favourable conditions under which individuals and organisations are able to react positively to out-of-equilibrium conditions, introducing *novelties* in the system (Antonelli 2011, 2015; Metcalfe 2010; Schumpeter 1947), and cities are mechanisms that facilitate this coordination process.

The second and only other key hypothesis of the model proposed by Gomez-Lievano et al. (2016) is that, as within models of Cultural Evolution, the number of factors in the environments –i.e., the *letters* that can be recombined to form different *words*, to be consistent with the *scrabble metaphor*–[12] is a function of population size (Henrich 2004; Kline and Boyd 2010; Kobayashi and Aoki 2012; Powell et al. 2009; Shennan 2001).[13] Within these models, cultural elements (among

[12]See n. 16

[13]If we look at Anthropology literature, the empirical evidence that supports this hypothesis is well established and wide in terms of factors considered: from tools to beliefs, among others (Bromham et al. 2015; Collard et al. 2013; Kline and Boyd 2010). And has been shown that small populations are not able to maintain the more complex elements of their culture (Diamond 1997). But recent evidence goes in its favour also for what concerns

which *technology*) accumulate and evolve following a Darwinian process that involves *inheritance*, *selection* and *variation* (Aldrich et al. 2008), and in particular it has been advanced the hypothesis that the *diversity* of different factors present in a group people grows logarithmically with its size (Henrich 2004). The scaling law of Eq. 1 descends mathematically from the two fundamental hypotheses just remembered, and the model easily accommodates the fact that the different phenomena show specific scaling patterns, depending on how much articulated their receipt is. Said otherwise, following this model we can expect that only large populations will be able to accumulate and preserve a *diversity* of "letters" large enough to possess all the complementary elements required to compose the more sophisticated "words". In particular, the model predicts that, as the complexity of a specific phenomenon increases, its prevalence (Y_0) will decrease, while its scaling exponent (β) and its variance among cities of the same size will increase.

4 Scaling laws of the EU Metropolitan Regions

Following a well-established tradition, we can take Eq. 1 in logarithms, so that it can be rewritten as

$$\log Y = \log Y_0 + \beta \log M, \tag{2}$$

where $\log(\cdot)$ is the natural logarithm. In this way, it is possible to estimate Eq. 2 using a standard Ordinary Least Squares (OLS) model like

$$\log Y_m = \log Y_0 + \beta \log N_m + \varepsilon_m. \tag{3}$$

To take into account the many country differences that still exist within the EU, I will consider a slightly modified version of Eq. 3, where I introduced a random effect that captures the country specificities.[14] Moreover, to help the reader in the interpretation of the results, I will test the significance of the estimated slope against the hypothesis of linearity (H0: $\hat{\beta} = 1$). Therefore, I estimate the following model

$$\log Y_{m,c} = \log Y_0 + \beta \log N_{m,c} + (v_c + \varepsilon_{m,c}).$$

Therefore, all three indices of economic performance correlate with population size in the EU28 Metropolitan Areas, as expected (Fig. 2). Specifically, the three variables scale super-linearly with population size (Tab. 2), with exponents compatible with previous findings about Europe: the punctual estimates in Bettencourt and Lobo (2016) are 1.17 for GDP; 1.02 for employment; and 1.13 for patents. The significance of the exponent for patent applications is very low, due to the higher variance visible also in the plot. This is perfectly in line with previous studies (Bettencourt et al. 2007b) and with the theory proposed by Gomez-Lievano et al. (2016). Indeed, this last predicts a higher variance for more *complex* activities, i.e. phenomena that need the recombination of a lot of different pieces of knowledge,[15] and patent production is surely of this kind. And this is also in line with previous findings about the scientific production in terms of papers (Nomaler et al. 2014).

contemporary cities, even though with more mixing results (Bettencourt et al. 2014; Brummitt et al. 2012; Youn et al. 2016).

[14] A likelihood ratio test of model reduction confirms the relevance of the specification proposed for all the three dependent variables considered. The same holds also for the "disaggregated" models proposed in the following section (Snijders and Bosker 2011, p. 91).

[15] See n. 16.

Figure 2: Population size vs GDP, employment, or patent applications. EU28 Metropolitan regions. Population, GDP and employment in 2008. Patent applications from 2004 to 2008. Source: Eurostat (tables `met_pjanaggr3`, `met_10r_3gdp`, `met_10r_3emp`, `met_pat_eptot`).

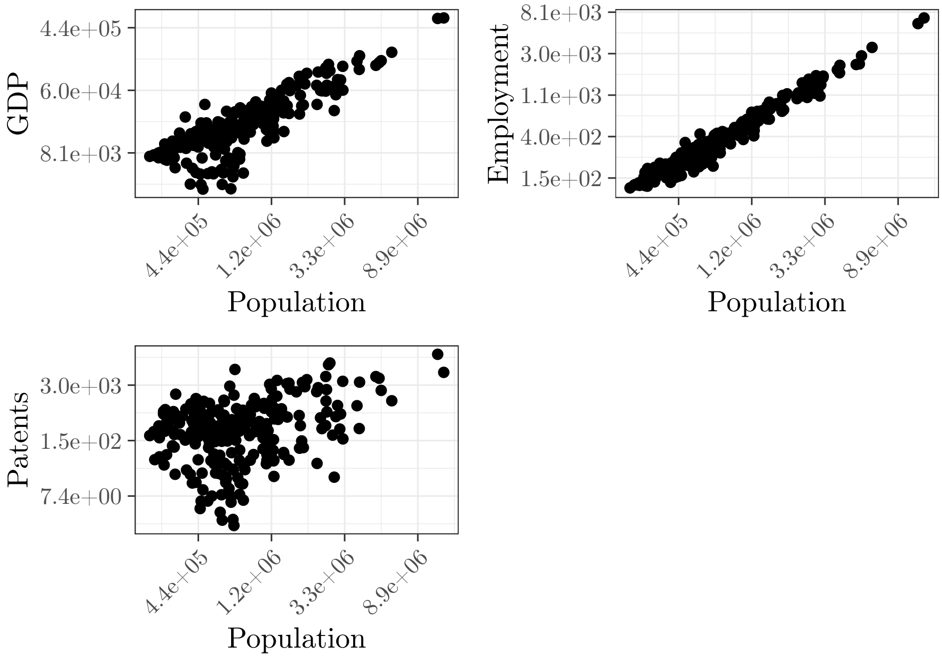


Table 2: Population size vs GDP, employment, or patent applications. EU28 Metropolitan regions. Population, GDP and employment in 2008. Patent applications from 2004 to 2008. Source: see Fig. 2. For the slope H0: $\hat{\beta} = 1$.

	Dependent variable:		
	log(GDP)	log(EMP)	log(PCT)
	(1)	(2)	(3)
log(POP)	1.123***	1.032**	1.123 .
	(0.021)	(0.012)	(0.070)
Constant	−5.485***	−8.067***	−10.372***
	(0.303)	(0.159)	(1.001)
Observations	248	248	248
R^2	0.957	0.990	0.560
Adjusted R^2	0.957	0.990	0.558
F Statistic	5,447.206***	24,382.350***	311.086***
Note:	. p<0.1; * p<0.05; ** p<0.01; *** p<0.001		

5 Technological knowledge production in Europe

Therefore, if it is true that some phenomena scales more than others because of the number of different and complementary capabilities that must be combined together to produce (or obtain) it, we can open the black box even more. Looking specifically at technological knowledge production, we can inspect the scaling patterns of different technological domains. The production of patents in those domains that require the recombination of previous knowledge from a wider set of different and complementary domains is expected to concentrate in larger cities, more than the production in other domains.

This is in line with some recent findings. Specifically, Balland et al. (2018) have shown that more *complex* economic activities concentrate more in large US metropolitan areas.[16] The authors explain that this type of processes requires a deeper *division of knowledge* (Metcalfe 2010; Warsh 2007), so laying on high coordination costs that cities can solve more easily (Scott and Storper 2014). While Mewes (2019) focused on *atypical combinations* of technological knowledge (Uzzi et al. 2013), finding that, over time, they have increasingly concentrated in the larger cities of the US. Thus, the author concludes that large metropolitan areas lead the technological progress not only in *quantitative* but also in *qualitative* terms. Lastly, with respect to scientific production, also Nomaler et al. (2014) have shown the existence of different scaling exponents among disciplines, but they did not investigate more on these differences.

Indeed, as highlighted by Jacobs (1961), *diversity* is the engine that generates new ideas and cultural items, and cities are social complex systems that facilitate the interaction among people, and so the help the functioning of this engine. But economic processes that require the combination of many diverse needed inputs will be subject to a deeper *division of knowledge*, and thus, will operate more efficiently in large cities, that provide organisational, institutional and cultural mechanisms that reduce the coordination costs of such a bundle of complementary inputs. This is in line with what highlighted by models of Cultural Evolution, so that I will push for an explanation that refers to the division of knowledge and evolutionary mechanisms of *variation* and *selection*, more than to explanations that refer to the density of spatially constraint interaction networks, as in most of the *urban scaling* literature. This not because I do not trust in this type of models, but because of the intrinsic limitations they have in framing more than one scaling pattern for the same phenomenon of interest, making them not the more suitable at corroborating the theory I am interested in.

In particular, the invention and innovation processes have been described (explicitly or implicitly) by many authors as a recombinant one, in which the existing knowledge elements (internal and external to each single firm and person considered) are recombined to produce something new (Arthur 2009; Nelson and Winter 1982; Schumpeter 1939; Weitzman 1998; Youn et al. 2015). But, we can expect that this recombination of already existing knowledge happens in different ways from one technological domain to another. Some domains will require the combination of pieces that are quite far apart from one another, while others will recombine elements that are, on average, more similar to one another. And, we can assume that this knowledge resides in peoples' brain, and that each person owns a different set of these portions in which the knowledge

[16] This use of the word *complexity* refers to the so-called *scrabble metaphor* (Hausmann et al. 2014, p. 20; see also Hausmann and Hidalgo 2011) and can be defined, following Fleming and Sorenson (2001, p. 1019), as «the interaction of size and interdependence». Using the words of Gomez-Lievano and Patterson-Lomba (2019), «[i]t may be argued that what it is meant by 'complexity' in this setting is really 'complicatedness'. Despite being aware of the mantra "'complexity' does not mean 'complicated"', in this case the word "complexity" is useful to describe the fact that social phenomena are the emergent result of many parts interacting locally». To avoid this confusion, in this paper I will prefer the idea of *diversity* to the one of *complexity*, but, if we use the definition just proposed, I see a very strong connection between the two concepts. And, in any case, you have to read the word *complexity* with this definition in mind whenever you will find the word in the text that follows.

is fragmented (division of knowledge). Consequently, we can expect a higher scaling exponent for these technological domains in which the production of new patents requires, on average, the recombination of a greater number of different pieces of knowledge, since the presence of more inventors in a city will provide the city with a larger set of different (and potentially complementary) elements to draw from (i.e., different "letters", if want to use the so-called *scrabble metaphor*) in order to recombine them in something new. In other words, we can expect that in larger cities people interact with more people. And since more people means, on average, more diverse people in terms of knowledge, larger cities will be more able to produce new knowledge in those technological domains that require the recombination of more diverse knowledge pieces.

We will test this theory measuring the *diversity* of each patent in the sample in terms of technological classes in their *backward citations*. I used the so-called Rao-Stirling index (or Integration score) to measure that. This index combines the three dimensions of the *diversity*, as detected by Stirling (2007): *variety* (the number of different elements); *balance* (the proportion of these elements); and *disparity* (the similarity among the elements considered). This index has already been extensively used to measure the interdisciplinarity of scientific papers and groups of researchers (Leydesdorff and Rafols 2011; Porter and Rafols 2009; Rafols and Meyer 2010), and since we are looking for a measure that accounts for the needed recombination efforts required by a patent, I see as key to consider the pairwise distance between the recombined classes, and not just the other two dimensions just remembered. Indeed, the key novelty of the Rao-Stirling diversity index is that it captures not only the number of different classes cited by a patent, or their degree of concentration –as other widely used indicators, like the Herfindahl or Shannon indices, do–, but it takes into account also how different these disciplines are one to the other. For this reason, this measure seems more appropriate than others used in the literature for the purpose of this paper.

In particular, I computed the Rao-Stirling index of the technological classes at 4 digits of the IPC classification of the *backward citations* of each patent in the sample, computing the distance between the classes as the cosine similarity of the 'cited classes'-'citing applications' co-occurrence matrix. That is

$$\mathrm{ST}_{patent} = \sum_{i,j=1}^{N} p_i p_j d_{ij},$$

where

$$d_{ij} = \frac{\sum_{k=1}^{M} p_{i,k} p_{j,k}}{\sqrt{\sum_{k=1}^{M} p_{i,k}^2}\sqrt{\sum_{k=1}^{M} p_{j,k}^2}},$$

N is the number of IPC4 classes, and M is the number of patents considered (Fig. 3 reports the distribution of the index by technological domain). Then, I computed the average Rao-Stirling of the patents produced in a given city in a given technological domain (as defined by Schmoch 2008), $\mathrm{ST}_{m,d}$.

Moreover, in line with most of the Cultural Evolution theory and as suggested by Nomaler et al. (2014, p. 5), we can expect that the production of patents scales super-linearly not only with population size, as shown, but also with the *effective population* (borrowing a term from this literature), that in this case is the number of actual inventors in the city. Fig. 4 and Tab. 3 show that this is the case. Even though at the aggregate level the production of patents scales more steeply with the total population, the R^2 is significantly higher for the other variable. And those last findings are in line with what shown by Bettencourt et al. (2007b), even though in this case there is a super-linear exponent for both the variables, while for the US the authors have shown a sub-linearity in the case of inventors. Therefore, I also collected, for each city-technology couple, the number of patent applications obtained $\mathrm{Pat}_{m,d}$, and the number of inventors $\mathrm{Inv}_{m,d}$.

Figure 3: Box-plots of the distribution of the Rao-Stirling index compute on the backward citations of each patent filed at the EPO and with priority year in 2004–2008. The data are grouped by the IPC35 class of the citing patent (Schmoch 2008) and ordered, from the left to the right, for decreasing average value of the group. Source: PATENTS-ICRIOS.

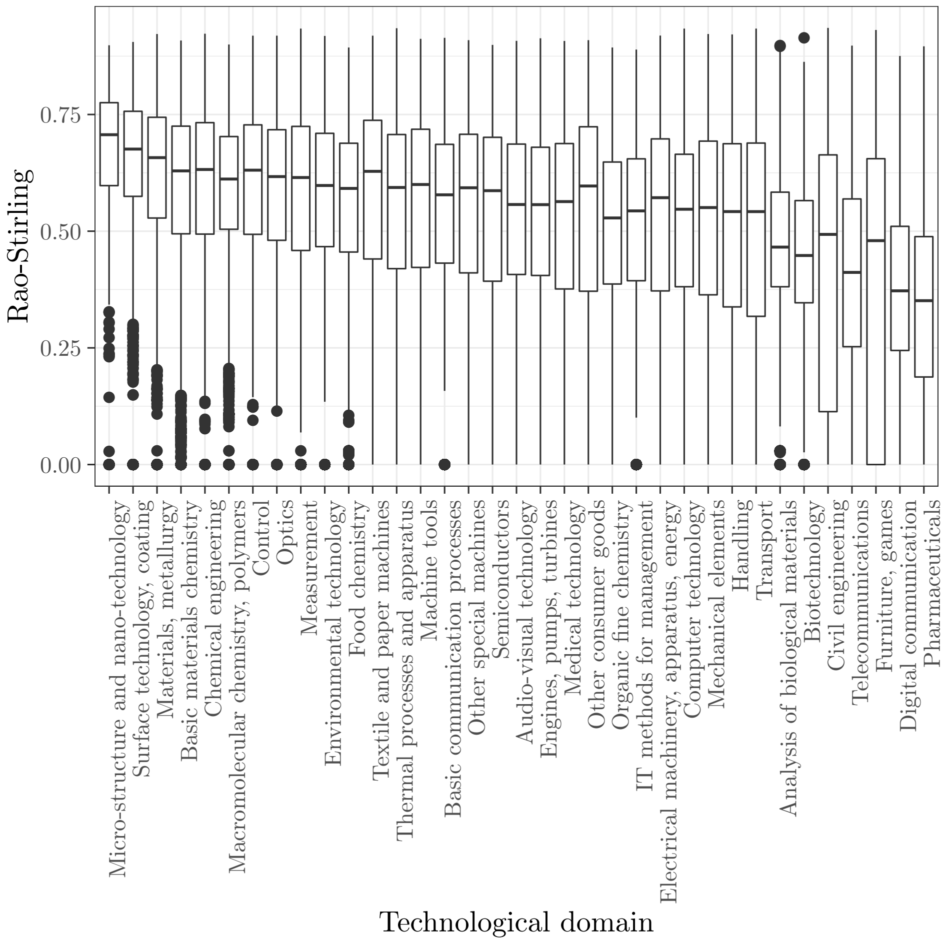
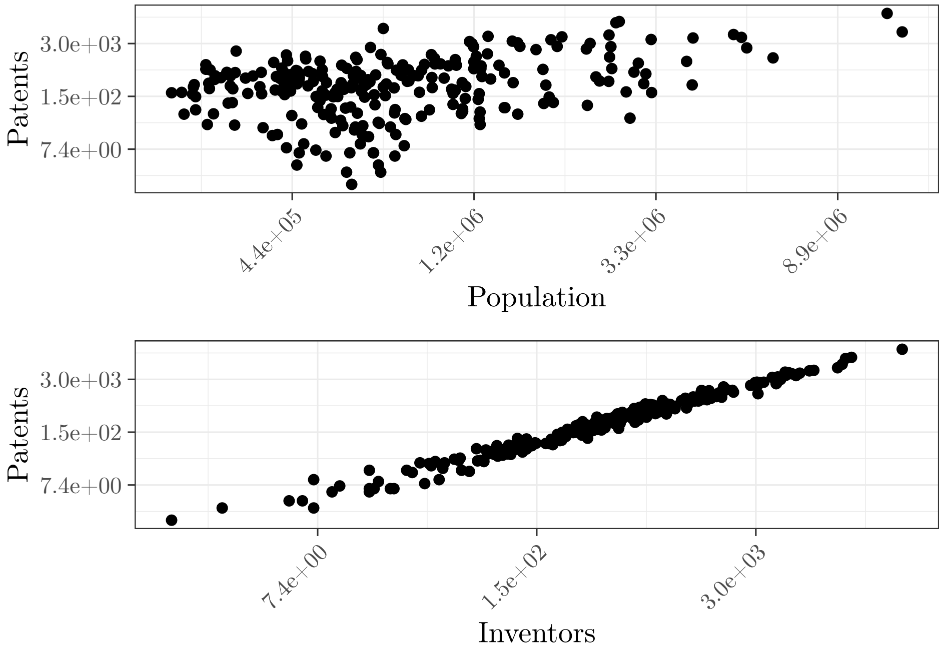


Figure 4: Population size in 2008 or total number of inventors in 2004–2008 vs total patent applications in 2004–2008 of each EU28 Metropolitan Region. Cities whose population size information is missing on the Eurostat database have been excluded. Sources: Eurostat (table `met_pjanaggr3`) and PATENTS-ICRIOS.



Therefore, the empirical strategy adopted in the paper is the following:

$$\log \mathrm{Pat}_{c,m,d} = \log \mathrm{Pat}_0 + \beta_1 \log \mathrm{Inv}_{c,m,d} + \beta_2 \tilde{\mathrm{ST}}_{c,m,d} + \beta_3 \log \mathrm{Inv}_{c,m,d} * \tilde{\mathrm{ST}}_{c,m,d} + (v_c + \varepsilon_{c,m,d}),$$

where $\tilde{\mathrm{ST}}_{c,m,d}$ is the scaled value (z-score) of the Rao-Stirling index.

Data The data used in the analysis are from the PATENTS-ICRIOS database (Coffano and Tarasconi 2014; ICRIOS 2018). I considered all the EPO patent applications in which at least one of the inventors' addresses is in a EU28 country (excluding overseas territories), and that has with priority year in the time span 2004–2008.[17] The geographical location of each patent is based on the location of its inventors (choosing the more represented NUTS3 area, whenever more than one were present). The PATENTS-ICRIOS database has been chosen since I need to count the inventors present in each city. Compared to other available patent data sources, this last contains a more reliable univocal identification of the inventors. Lastly, I considered, as potential *backward citations*, all the patents present in the database and, in the computation of the diversity index I took into account their *subclasses* codes (IPC4). The data obtained

[17] The analysis that follows considers the data as a cross-section. The aggregation of a five-year time window in a unique observation is in line with the literature and compensates for some rigidities present in firms' technological competencies evolution (Nesta and Saviotti 2006, p. 630, n. 3). The choice of 2008 as last year has been for precautionary reasons. Firstly, the 2018 PATENTS-ICRIOS database version is based on original data updated up to 10/2016, so that it could be seen as "unsafe" to look at patent too close to the end of this window. Moreover, I would like to avoid the possible effects of the Great Recession that affected Europe starting from the year 2008, on average.

Table 3: Estimated regressions for population size or total number of inventors vs total number of patents in each EU28 Metropolitan Region. Source: see Fig. 4. For the slope H0: $\hat{\beta} = 1$.

	Dependent variable:	
	$\log \mathrm{Pat}_{c,m}$	
	(1)	(2)
log POP	1.160^{*}	
	(0.074)	
$\log \mathrm{Inv}_{c,m}$		1.035^{**}
		(0.012)
Constant	-10.822^{***}	-0.646^{***}
	(1.042)	(0.075)
Observations	246	246
R^2	0.559	0.976
Adjusted R^2	0.557	0.975
F Statistic	307.568^{***}	$9{,}727.077^{***}$
Note:	. p<0.1; * p<0.05; ** p<0.01; *** p<0.001	

from the PATENTS-ICRIOS database have been, then, aggregated into Metropolitan Regions, which are combinations of NUTS3 regions which represent agglomerations of at least 250,000 inhabitants. These agglomerations were identified using the Urban Audit's Functional Urban Area (FUA) that including the commuter belt around a city, they correct the distortions created by commuting.[18] In the analysis, I considered all the metropolitan areas for which the Eurostat reports its population size for the year 2012, and I also excluded those city-technology observations in which I observed less than 25 inventors. In the end, the sample is composed by 24 countries (all the EU28 with the exception of Bulgaria, Cyprus, Lithuania, and Malta) and 193 Metropolitan Regions (distributed as reported by Tab. 5). Tab. 4 reports the summary statistics of the variables of interest.

[18]The Functional Urban Area (FUA) definition, provided by the OECD and the EU in the 2012 *Redefining "Urban"* report, frame an urban area as a "functional economic unit" formed by a densely inhabited city and its "related" commuting zone. Being a functional definition of the spatial agglomerations, not based on the administrative borders, it provides many advantages, not least the fact that across-countries comparisons are in that way possible. In particular, at least for the EU metropolitan areas, the procedure to define a FUA is the following. At first, the urban cores are identified looking at contiguous 1km 2 grid cells with more than 1,500 inhabitants. Only in a second step, the identified high-density cluster are mapped into small administrative units (Eurostat LAU2 areas), considering as part of the urban core identified all the LAU2 areas whose population is at least for half included in the identified cluster. Then, to account for the polycentric structured urban areas, travel-to-work flows are used to identify the hinterlands whose labour market is highly integrated with the cores. In particular, if more than 15% of the resident population of a core commutes to work in another, the two are joined together. Lastly, the hinterland –i.e., the pool of work-force of the urban labour market, outside the densely inhabited core– is identified and included in the high-density populated area just identified. Specifically, urban hinterlands are defined as all municipalities with at least 15% of their employed residents working in a certain urban core.

Table 4: Descriptive statistics of the variables of the regression analysis reported by Tab. 6. Source: see the main text.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
$\mathrm{Pat}_{c,m,d}$	2,958	131.674	210.913	26	39	133	2,626
$\mathrm{Inv}_{c,m,d}$	2,958	84.968	146.860	4	23	83	2,059
$\mathrm{ST}_{c,m,d}$	2,958	0.509	0.096	0.142	0.444	0.579	0.761

Table 5: Number of Metropolitan Regions considered in the analysis reported by Tab. 6 by country. Source: see the main text.

Country	N_m	Country	N_m	Country	N_m
AT	5	FI	3	NL	9
BE	5	FR	31	PL	2
CZ	3	HR	1	PT	2
DE	56	HU	2	RO	1
DK	4	IE	2	SE	4
EE	1	IT	14	SI	1
EL	1	LU	1	SK	1
ES	16	LV	1	UK	27

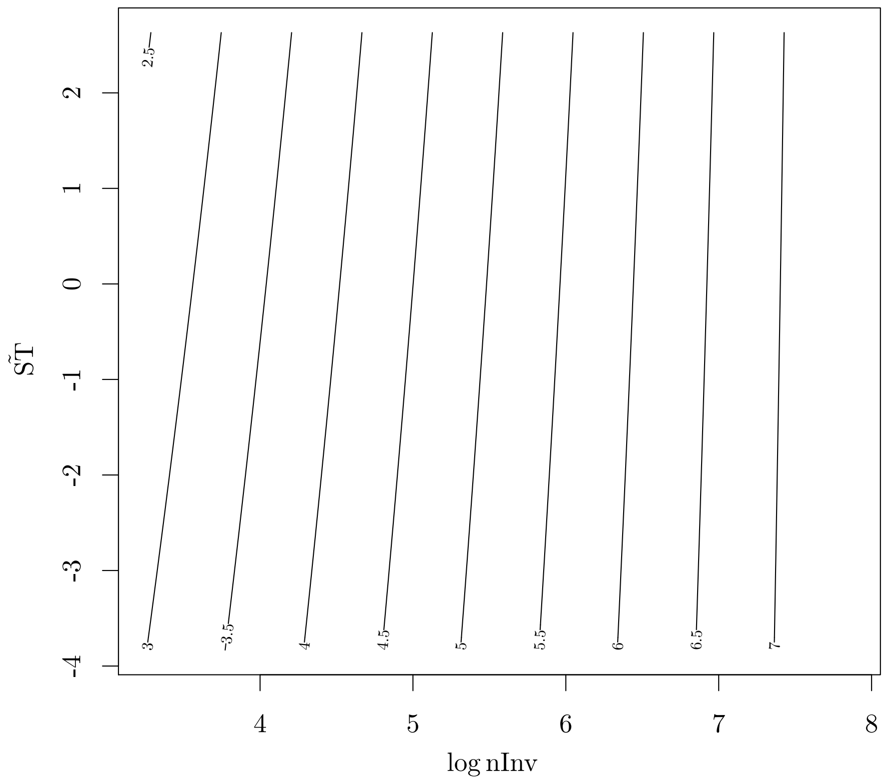
Results As shown by Tab. 6, also at the disaggregated level patent production scales superlinearly with effective population. Moreover, model (2) shows that the production of patents decreases with the increase of the diversity of the patents produced (in terms of needed knowledge to be recombined). Lastly, model (3) shows that more diversified patents tend to scale more with effective population size. A small number of inventors seem not able to produce those patents that require more sophisticated recipes. In terms of the hypothesis raised above, to recombine knowledge pieces that are very different one another requires a larger number of individuals, each contributing with its specific know-how.

Model (3) is then re-estimated using a simple OLS model without country-level random effects, and the interaction between the two predictors is plotted in Fig. 5. We can see how the smaller metropolitan areas are less able to produce those patents that require the recombination of a larger number of different knowledge components. In other words, it seems that those patents that depend more from the recombination of previously existing technological knowledge, and that one average recombine domains more distant one to the other concentrate more in larger cities. This effect is stronger for cities with a small number of inventors, while it seems to disappear for metropolitan areas with a larger effective population. Indeed, the biggest urban agglomerations seem to be not affected by these limitations, and only one scaling law holds for them whatever type of patents we consider. Even though very tentative, this latter observation points in the direction of a nested structure of the urban system. As highlighted by the Economic Complexity literature, few places will be able to produce the items that require a more sophisticated combination of the existing building blocks, and they will be able to produce whatever other items. On the other hand, the vast part of the system will be able to reach only the less valuable part of the production (Cristelli et al. 2013; Hidalgo and Hausmann 2009; Straka et al. 2017; Tacchella et al. 2012).

Table 6: Estimated regressions for the number of inventors, the average Rao-Stirling index, and their interaction effect vs the number of patent applications, by belonging IPC35 class and Metropolitan Region. Source: see the main text.

	Dependent variable:		
	$\log \mathrm{Pat}_{c,m,d}$		
	(1)	(2)	(3)
$\log \mathrm{Inv}_{c,m,d}$	1.044***	1.036***	1.038***
	(0.007)	(0.007)	(0.007)
$\tilde{\mathrm{ST}}_{c,m,d}$		−0.074***	−0.152***
		(0.006)	(0.029)
$\log \mathrm{Inv}_{c,m,d} * \tilde{\mathrm{ST}}_{c,m,d}$			0.018***
			(0.007)
Constant	−0.755***	−0.760***	−0.767***
	(0.050)	(0.056)	(0.056)
Observations	2,958	2,958	2,958
Log Likelihood	−810.485	−737.343	−737.687
Akaike Inf. Crit.	1,628.971	1,484.686	1,487.374
Bayesian Inf. Crit.	1,652.940	1,514.648	1,523.327
Pseudo-R^2 (fixed effects)	0.87	0.86	0.86
Pseudo-R^2 (total)	0.90	0.90	0.91
Note:	. p<0.1; * p<0.05; ** p<0.01; *** p<0.001		

Figure 5: Marginal effects of the regression (3) of Tab. 6. Source: see the main text.



6 Conclusions

Inventive activity is key for economic growth and development, as much as unevenly distributed at the geographical level. Much of the technological knowledge is produced in tiny areas, with few of them responsible for most of the entire production. Following a by then well-established literature, we have seen that European cities show a super-linear relationships between their population size and invention performance.

Then, I opened the black box even more, showing that the production of some types of technological knowledge agglomerate more than others. The empirical findings corroborate the theoretical expectations based on a model proposed by Gomez-Lievano et al. (2016), that, combining Economic Complexity and Cultural Evolution hints, suggests that the small urban agglomerations will be not only less productive in terms of total patent applications produced, but also unable to patent in those technological domains that require the recombination of a wider set different and complementary knowledge inputs.

However, this work has some caveats. Firstly, the model proposed by Pumain et al. (2006) suggests that the different concentration levels of the production of new technological knowledge in different domains depends also on its level of *maturity*. Since the empirical findings shown in this paper do not take into account this fact, we cannot assess which of the effects (if any) is prevalent. And it is anyhow possible that both are in action at the same time, so that if able to disentangle the two, we would see clearer patterns among the different domains. Secondly, in this work I chose to consider as *technological domains* the 35 patent classes defined by Schmoch (2008). To disaggregate the observations more that so has two main drawbacks. On the one hand, it reduces the observations available for each in each group, and so the power of the regression analysis. On the other hand, it also means to abandon a grouping specifically thought to have a technological meaning and to be suitable for cross-country comparisons. However, it would be interesting to have at hands more detailed data under the technological point of view. This would let us explore more accurately the nested structure that seems to emerge from the analysis presented, and that I briefly discussed at the end of the previous section. Thirdly, the use of the residuals of the regression could provide useful hints. First and foremost, due to the feedback-loops that characterise the production of new technological knowledge, we can expect that these that over-perform (under-perform) today will do the same also tomorrow. And it could be the case that this effect is different for different technological domains. Lastly, more diversified knowledge combinations do not (automatically) imply a high technological impact or economic value. Thus, it remains undetermined how more sophisticated combinations relate to the economic performance of cities and how they explain differences between them in terms of well-being and growth potentials. To go in the direction pointed by the four issues raised opens challenging questions whose answers would suggest powerful policy implications. Do *leading technologies* concentrate more in large cities because of path-dependent mechanisms, or are technologies that require more sophisticated recipes to do so more? Do urban systems show a specialisation pattern, or is it true that only the larger cities are able to diversify, covering the whole technological spectrum? Does the evolutionary path of cities as centres of production of new technological knowledge show some sort of Matthew effect? Is the knowledge that requires more sophisticated production processes the more valuable and are large cities over-represented in the more valuable productions? We are confident that future studies will answer these questions.

Acknowledgements

I would like to thank Andres Gomez-Lievano for the priceless discussions we had during my visiting period at the CID at Harvard University, without which this paper would have never been possible. As well, I am indebted to Martina Iori, Fabrizio Fusillo and Francesco Trentini to have supported me during the construction of the paper, with constant comments, opinions and critiques that have definitely improved the quality of the work. I would also like to thank Consuelo R. Nava for the precious contributes to a preliminary project, on which the paper has grown up. Lastly, my gratitude goes also to the members of the RENIR project, to the participants of the WICI conference on "Modelling Complex Urban Environments", and to Frank Neffke to have commented this last preliminary project, indirectly contribution to the present version of the paper. Any errors are my own.

References

Abramovitz, Moses and Paul A. David (1996). 'Technological change and the rise of intangible investments: The US economy's growth-path in the twentieth century'. In: *Employment and Growth in the Knowledge-based Economy*. Paris: OECD, 35–60. ISBN: 9264148132.

Aldrich, Howard E., Geoffrey M. Hodgson, David L. Hull, Thorbjørn Knudsen, Joel Mokyr and Viktor J. Vanberg (Oct. 2008). 'In Defence of Generalized Darwinism'. *Journal of Evolutionary Economics* 18 (5), 577–596. DOI: 10.1007/s00191-008-0110-z.

Alves, Luiz G. A., Renio S. Mendes, Ervin K. Lenzi and Haroldo V. Ribeiro (Sept. 2015). 'Scale-Adjusted Metrics for Predicting the Evolution of Urban Indicators and Quantifying the Performance of Cities'. *PLOS ONE* 10 (9), e0134862. DOI: 10.1371/journal.pone.0134862.

Antonelli, Cristiano (2011). 'The Economic Complexity of Technological Change: Knowledge Interaction and Path Dependence'. In: *Handbook on the Economic Complexity of Technological Change*. Ed. by Cristiano Antonelli. Edward Elgar.

— (2015). 'Innovation as a Creative Response: A Reappraisal of the Schumpeterian Legacy'. *History of Economic Ideas* 23.2, 99–118.

Arbesman, Samuel, Jon M. Kleinberg and Steven H. Strogatz (2009). 'Superlinear Scaling for Innovation in Cities'. *Physical Review E* 79.1, 016115. DOI: 10.1103/physreve.79.016115.

Arthur, W. Brian (July 1996). 'Increasing Returns and the New World of Business'. *Harvard Business Review* 74 (4), 100–109.

— (Apr. 1999). 'Complexity and the Economy'. *Science* 284 (5411), 107–109. DOI: 10.1126/science.284.5411.107.

— (2009). *The Nature of Technology: What It Is and How It Evolves*. New York: Free Press.

Audretsch, David B. and Maryann P. Feldman (June 1996). 'R&D Spillovers and the Geography of Innovation and Production'. *The American Economic Review* 86.3, 630–640.

Balland, Pierre-Alexandre, Cristian Jara-Figuero, Sergio Petralia, Mathieu Steijn, David Rigby and César A. Hidalgo (July 2018). *Complex Economic Activities Concentrate in Large Cities*. arXiv. arXiv. eprint: 1807.07887.

Batty, Michael (Feb. 2008). 'The Size, Scale, and Shape of Cities'. *Science* 319.5864, 769–771. DOI: 10.1126/science.1151419.

Batty, Michael and Paul Longley (1994). *Fractal Cities: A Geometry of Form and Function*. Academic Press. ISBN: 0124555705.

Bettencourt, Luís M. A. (20th June 2013a). *Cities Are a New Kind of Complex System: Part Star, Part Network*. Santa Fe Institute. URL: https://www.santafe.edu/news-center/news/science-bettencourt-cities-framework (visited on 16/03/2018).

Bettencourt, Luís M. A. (June 2013b). 'The Origins of Scaling in Cities'. *Science* 340.6139, 1438–1441. DOI: 10.1126/science.1235823.

Bettencourt, Luís M. A. and José Lobo (Mar. 2016). 'Urban Scaling in Europe'. *Journal of The Royal Society Interface* 13 (116), 1–14. DOI: 10.1098/rsif.2016.0005.

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert and Geoffrey B. West (Apr. 2007a). 'Growth, Innovation, Scaling, and the Pace of Life in Cities'. *Proceedings of the National Academy of Sciences of the United States of America* 104.17, 7301–7306. DOI: 10.1073/pnas.0610172104.

Bettencourt, Luís M. A., José Lobo and Deborah Strumsky (Feb. 2007b). 'Invention in the City: Increasing Returns to Patenting as a Scaling Function of Metropolitan Size'. *Research Policy* 36.1, 107–120. DOI: 10.1016/j.respol.2006.09.026.

Bettencourt, Luís M. A., José Lobo, Deborah Strumsky and Geoffrey B. West (Nov. 2010). 'Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities'. *PLOS ONE* 5 (11), e13541. DOI: 10.1371/journal.pone.0013541.

Bettencourt, Luís M. A., Horacio Samaniego and Hyejin Youn (2014). 'Professional Diversity and the Productivity of Cities'. *Scientific Reports* 4.5393, 1–6. DOI: 10.1038/srep05393.

Bettencourt, Luís M. A. and Geoffrey B. West (Oct. 2010). 'A Unified Theory of Urban Living'. *Nature* 467.7318, 912–913. DOI: 10.1038/467912a.

Boyd, Robert, Peter J. Richerson, Joseph Henrich and J. Lupp (2013). 'The cultural evolution of technology: facts and theories'. In: *Cultural Evolution: Society, Language, and Religion.* Ed. by Peter J. Richerson and Morten H. Christiansen. Vol. 12. Strungmann Forum Reports. Cambridge (MA): The MIT Press.

Breschi, Stefano and Francesco Lissoni (2003). *Mobility and Social Networks: Localised Knowledge Spillovers Revisited*. Research rep. 142. CESPRI Working Paper.

Bromham, Lindell, Xia Hua, Thomas G. Fitzpatrick and Simon J. Greenhill (Feb. 2015). 'Rate of Language Evolution is Affected by Population Size'. *Proceedings of the National Academy of Sciences of the United States of America* 112.7, 2097–2102. DOI: 10.1073/pnas.1419704112.

Brown, James H., Geoffrey B. West and Brian J. Enquist (2000). 'Scaling in Biology: Patterns and Processes, Causes and Consequences'. In: *Scaling in Biology*. Ed. by James H. Brown and Geoffrey B. West. New York: Oxford University Press. Chap. 7, 1–24.

Brummitt, Charles D., Andres Gomez-Lievano, Nicolas Goudemand and Gareth Haslam (2012). *Hunting for Keys to Innovation: The Diversity and Mixing of Occupations Do Not Explain a City's Patent and Economic Productivity.* Complex Systems Summer School Proceedings. Santa Fe Institute.

Brunow, Stephan, Antonia Birkeneder and Andrés Rodríguez-Pose (Aug. 2018). 'Creative and Science Oriented Employees and Firm-level Innovation'. *Cities* 78, 27–38. DOI: 10.1016/j.cities.2018.02.002.

Coffano, Monica and Gianluca Tarasconi (2014). *CRIOS - Patstat Database: Sources, Contents and Access Rules.* Tech. rep. 1. CRIOS Working Paper Series, Bocconi University. DOI: 10.2139/ssrn.2404344.

Coleman, James S. (1964). *Introduction to Mathematical Sociology.* New York: Free Press of Glencoe.

Collard, Mark, April Ruttle, Briggs Buchanan and Michael J. O'Brien (Sept. 2013). 'Population Size and Cultural Evolution in Nonindustrial Food-Producing Societies'. *PLOS ONE* 8 (9), e72628. DOI: 10.1371/journal.pone.0072628.

Cristelli, Matthieu, Andrea Gabrielli, Andrea Tacchella, Guido Caldarelli and Luciano Pietronero (Aug. 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.

Diamond, Jared M. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. New York (NY): W. W. Norton.

Duranton, Gilles and Diego Puga (2004). 'Micro-foundations of Urban Agglomeration Economies'. In: *Handbook of Regional and Urban Economics*. Ed. by J. Vernon Henderson and Jacques-François Thisse. Vol. 4. North-Holland. Chap. 48, 2063–2117. DOI: 10.1016/S1574-0080(04)80005-1.

Feldman, Maryann P. and Richard Florida (June 1994). 'The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States'. *Annals of the Association of American Geographers* 84 (2), 210–229. DOI: 10.1111/j.1467-8306.1994.tb01735.x.

Fleming, Lee and Olav Sorenson (2001). 'Technology as a Complex Adaptive System: Evidence From Patent Data'. *Research Policy* 30.7, 1019–1039.

Florida, Richard L. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community, and Everyday Life*. New York (NY): Basic Books. ISBN: 0465024769.

Freeman, Christopher and Luc G. Soete (1997). *The Economics of Industrial Innovation*. 3rd ed. An earlier version of the book was published by Penguin Modern Economics Texts in 1974. The MIT Press.

Glaeser, Edward L. (Sept. 1999). 'Learning in Cities'. *Journal of Urban Economics* 46 (2), 254–277. DOI: 10.1006/juec.1998.2121.

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman and Andrei Shleifer (Dec. 1992). 'Growth in Cities'. *Journal of Political Economy* 100.6, 1126–1152. DOI: 10.1086/261856.

Gomez-Lievano, Andres, Michele Coscia, Frank Neffke and Ricardo Hausmann (2019a). 'Linking Structure to the Dynamics of Collective Learning'. Unpublished.

Gomez-Lievano, Andres and Oscar Patterson-Lomba (2018). 'The Facets of Complexity of Urban Phenomena, Their Effects on Temporal Dynamics, and Their Association with Economic Performance'. Unpublished.

— (2019). 'The Facets of Complexity of Urban Phenomena, Their Effects on Temporal Dynamics, and Their Association with Economic Performance'. Unpublished.

Gomez-Lievano, Andres, Oscar Patterson-Lomba and Ricardo Hausmann (Dec. 2016). 'Explaining the Prevalence, Scaling and Variance of Urban Phenomena'. *Nature Human Behaviour* 1.1, 0012. DOI: 10.1038/s41562-016-0012.

Gomez-Lievano, Andres, Debborah Strumsky, Joseph Henrich and José Lobo (2019b). 'Invention as Cultural Accumulation'. Unpublished.

Griliches, Zvi (Dec. 1990). 'Patent Statistics as Economic Indicators: A Survey'. *Journal of Economic Literature* 28.4, 1661–1707.

UN-Habitat (2016). *World Cities Report 2016. Urbanization and Development: Emerging Futures*. United Nations. ISBN: 9211327083.

Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg (2005). 'Market Value and Patent Citations'. *RAND Journal of Economics* 36.1, 16–38.

Hausmann, Ricardo and César A. Hidalgo (Dec. 2011). 'The Network Structure of Economic Output'. *Journal of Economic Growth* 16 (4), 309–342. DOI: 10.1007/s10887-011-9071-4.

Hausmann, Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jimenez, Alexander Simoes and Muhammed A. Yildirim (2014). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. First edition 2011. The MIT Press.

Henrich, Joseph (Apr. 2004). 'Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses—The Tasmanian Case'. *American Antiquity* 69.2, 197–214. DOI: 10.2307/4128416.

— (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton (NJ): Princeton University Press.

Henrich, Joseph, Robert Boyd, Maxime Derex, Michelle A. Kline, Alex Mesoudi, Michael Muthukrishna, Adam T. Powell, Stephen J. Shennan and Mark G. Thomas (Nov. 2016). 'Understanding Cumulative Cultural Evolution'. *Proceedings of the National Academy of Sciences of the United States of America* 113.44, E6724–E6725. DOI: 10.1073/pnas.1610005113.

Hidalgo, César A. and Ricardo Hausmann (June 2009). 'The Building Blocks of Economic Complexity'. *Proceedings of the National Academy of Sciences of the United States of America* 106.26, 10570–10575. DOI: 10.1073/pnas.0900943106.

Huxley, Julian S. (1932). *Problems of Relative Growth.* London: Methuen.

ICRIOS (2018). *PATENTS-ICRIOS.* URL: http://download.unibocconi.it/ICRIOS2018/icrios201806.rar.

Jacobs, Jane (1961). *The Death and Life of Great American Cities.* New York: Random House.

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

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson (Aug. 1993). 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations'. *The Quarterly Journal of Economics* 108 (3), 577–598. DOI: 10.2307/2118401.

Katz, J. Sylvan (Sept. 2006). 'Indicators for Complex Innovation Systems'. *Research Policy* 35 (7), 893–909. DOI: 10.1016/j.respol.2006.03.007.

Kleiber, Max (Jan. 1932). 'Body Size and Metabolism'. *Hilgardia* 6.11, 315–353. DOI: 10.3733/hilg.v06n11p315.

Kline, Michelle A. and Robert Boyd (Apr. 2010). 'Population Size Predicts Technological Complexity in Oceania'. *Proceedings of the Royal Society B: Biological Sciences* 277 (1693), 2559–2564. DOI: 10.1098/rspb.2010.0452.

Kobayashi, Yutaka and Kenichi Aoki (Aug. 2012). 'Innovativeness, Population Size and Cumulative Cultural Evolution'. *Theoretical Population Biology* 82 (1), 38–47. DOI: 10.1016/j.tpb.2012.04.001.

Kortum, Samuel and Josh Lerner (2000). 'Assessing the Contribution of Venture Capital to Innovation'. *The RAND Journal of Economics* 31.4, 674–692. DOI: 10.2307/2696354.

Kremer, Michael (Aug. 1993). 'The O-Ring Theory of Economic Development'. *The Quarterly Journal of Economics* 108 (3), 551–575. DOI: 10.2307/2118400.

Krugman, Paul (1991). 'Increasing Returns and Economic Geography'. *Journal of Political Economy* 99.3, 483–499.

Leydesdorff, Loet and Ismael Rafols (Jan. 2011). 'Indicators of the Interdisciplinarity of Journals: Diversity, Centrality, and Citations'. *Journal of Informetrics* 5 (1), 87–100. DOI: 10.1016/j.joi.2010.09.002.

Lobo, José, Luís M. A. Bettencourt, Deborah Strumsky and Geoffrey B. West (Mar. 2013). 'Urban Scaling and the Production Function for Cities'. *PLoS ONE* 8 (3), e58407. DOI: 10.1371/journal.pone.0058407.

Lucas, Robert E. (July 1988). 'On the Mechanics of Economic Development'. *Journal of Monetary Economics* 22 (1), 3–42. DOI: 10.1016/0304-3932(88)90168-7.

Metcalfe, J. Stanley (2010). 'Complexity and Emergence in Economics: The Road from Smith to Hayek (Via Marshall and Schumpeter)'. *History of Economic Ideas* 18.2, 45–75.

Mewes, Lars (2019). 'Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836 to 2010'. *Economic Geography*, 1–21. DOI: 10.1080/00130095.2019.1567261.

Mori, Tomoya, Koji Nishikimi and Tony E. Smith (Feb. 2008). 'The Number-average Size Rule: A New Empirical Relationship Between Industrial Location and City Size'. *Journal of Regional Science* 48 (1), 165–211. DOI: 10.1111/j.1467-9787.2008.00550.x.

Nelson, Richard R. and Sidney G. Winter (1982). *An Evolutionary Theory of Economic Change*. Cambridge (MA): Harvard University Press.

Nesta, Lionel and Pier Paolo Saviotti (Aug. 2006). 'Firm Knowledge and Market Value in Biotechnology'. *Industrial and Corporate Change* 15.4, 625–652. DOI: 10.1093/icc/dtl007.

Nomaler, Önder, Koen Frenken and Gaston Heimeriks (Oct. 2014). 'On Scaling of Scientific Knowledge Production in U.S. Metropolitan Areas'. *PLOS ONE* 9 (10), e110805. DOI: 10.1371/journal.pone.0110805.

OECD (2012). *Redefining "Urban". A New Way to Measure Metropolitan Areas*. OECD. ISBN: 9789264174054.

— (2013). *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 31/05/2018).

— (2015). *The Metropolitan Century: Understanding Urbanisation and its Consequences*. Paris: OECD Publishing. ISBN: 978-92-64-22873-3. DOI: 10.1787/9789264228733-en.

Pan, Wei, Gourab Ghoshal, Coco Krumme, Manuel Cebrian and Alex Pentland (June 2013). 'Urban Characteristics Attributable to Density-driven Tie Formation'. *Nature Communications* 4.1961. DOI: 10.1038/ncomms2961.

Pavitt, Keith (Jan. 1985). 'Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems'. *Scientometrics* 7.1–2, 77–99. DOI: 10.1007/bf02020142.

Porter, Alan L. and Ismael Rafols (Dec. 2009). 'Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields Over Time'. *Scientometrics* 81.3, 719–745. DOI: 10.1007/s11192-008-2197-2.

Porter, Michael E. (1998). 'Clusters and the New Economics of Competition'. *Harvard Business Review* 76, 77–90. URL: https://hbr.org/1998/11/clusters-and-the-new-economics-of-competition (visited on 26/09/2018).

Powell, Adam T., Stephen J. Shennan and Mark G. Thomas (June 2009). 'Late Pleistocene Demography and the Appearance of Modern Human Behavior'. *Science* 324 (5932), 1298–1301. DOI: 10.1126/science.1170165.

Pumain, Denise, Fabien Paulus, Céline Vacchiani-Marcuzzo and José Lobo (July 2006). 'An Evolutionary Theory for Interpreting Urban Scaling Laws'. *Cybergeo: European Journal of Geography*. Systems, Modelling, Geostatistics 343. DOI: 10.4000/cybergeo.2519.

Rafols, Ismael and Martin Meyer (2010). 'Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience'. *Scientometrics* 82, 263–287. DOI: 10.1007/s11192-009-0041-y.

Richerson, Peter J. and Robert Boyd (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago (IL): The University of Chicago Press.

Romer, Paul M. (Oct. 1990). 'Endogenous Technological Change'. *Journal of Political Economy* 98.5, Part 2, S71–S102. DOI: 10.1086/261725.

Samaniego, Horacio and Melanie E. Moses (2008). 'Cities as Organisms: Allometric Scaling of Urban Road Networks'. *Journal of Transport and Land Use* 1.1, 21–39. DOI: 10.5198/jtlu.v1i1.29.

Schiff, Nathan (Nov. 2015). 'Cities and Product Variety: Evidence from Restaurants'. *Journal of Economic Geography* 15 (6), 1085–1123. DOI: 10.1093/jeg/lbu040.

Schmoch, Ulrich (2008). *Concept of a Technology Classification for Country Comparisons*. Research rep. World Intellectual Property Organisation (WIPO).

Schumpeter, Joseph A. (1911). *Theorie der Wirtschaftlichen Entwicklung*. German. translated by Redvers Opie as *The Theory of Economic Development* (1983). Duncker & Humblot.

— (1939). *Business Cycles. A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. 2 vols. New York & London: McGraw-Hill.

Schumpeter, Joseph A. (Nov. 1947). 'The Creative Response in Economic History'. *The Journal of Economic History* 7.2, 149–159.
Scott, Allen J. and Michael Storper (2014). 'The Nature of Cities: The Scope and Limits of Urban Theory'. *International Journal of Urban and Regional Research* 39.1, 1–15. DOI: 10.1111/1468-2427.12134.
Shennan, Stephen J. (Apr. 2001). 'Demography and Cultural Innovation: A Model and its Implications for the Emergence of Modern Human Culture'. *Cambridge Archaeological Journal* 11 (1), 5–16. DOI: 10.1017/s0959774301000014.
Snijders, Tom A. B. and Roel J. Bosker (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* SAGE.
Stirling, Andy (Aug. 2007). 'A General Framework for Analysing Diversity in Science, Technology and Society'. *Journal of The Royal Society Interface* 4 (15), 707–719. DOI: 10.1098/rsif.2007.0213.
Storper, Michael (Mar. 2011). 'Why Do Regions Develop and Change? The Challenge for Geography and Economics'. *Journal of Economic Geography* 11 (2), 333–346. DOI: 10.1093/jeg/lbq033.
Straka, Mika J., Guido Caldarelli and Fabio Saracco (2017). 'Grand Canonical Validation of the Bipartite International Trade Network'. *Physical Review E* 96.2. DOI: 10.1103/physreve.96.022306.
Tacchella, Andrea, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli and Luciano Pietronero (2012). 'A New Metrics for Countries' Fitness and Products' Complexity'. *Scientific Reports* 2.723, 1–7. DOI: 10.1038/srep00723.
Uzzi, Brian, Satyam Mukherjee, Michael Stringer and Ben Jones (Oct. 2013). 'Atypical Combinations and Scientific Impact'. *Science* 342 (6157), 468–472. DOI: 10.1126/science.1240474.
Vanberg, Viktor (1994). 'Cultural Evolution, Collective Learning, and Constitutional Design'. In: *Economic Thought and Political Theory*. Ed. by David Reisman. Dordrecht: Springer, 171–204. DOI: 10.1007/978-94-011-1380-9_7.
Verspagen, Bart and Wilfred Schoenmakers (Jan. 2004). 'The Spatial Dimension of Patenting by Multinational Firms in Europe'. *Journal of Economic Geography* 4 (1), 23–42. DOI: 10.1093/jeg/4.1.23.
Warsh, David (2007). *Knowledge and the Wealth of Nations: A Story of Economic Discovery.* W.W. Norton & Co. ISBN: 0393329887.
Weitzman, Martin L. (May 1998). 'Recombinant Growth'. *The Quarterly Journal of Economics* 113 (2), 331–360. DOI: 10.1162/003355398555595.
West, Geoffrey B., James H. Brown and Brian J. Enquist (June 1999). 'The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms'. *Science* 284.5420, 1677–1679. DOI: 10.1126/science.284.5420.1677.
— (2000). 'The Origin of Universal Scaling Laws in Biology'. In: *Scaling in Biology*. Ed. by James H. Brown and Geoffrey B. West. New York: Oxford University Press. Chap. 2, 87–112.
— (Oct. 2001). 'A General Model for Ontogenetic Growth'. *Nature* 413.6856, 628–631. DOI: 10.1038/35098076.
Wichmann Matthiessen, Christian, Annette Winkel Schwarz and Søren Find (Aug. 2010). 'World Cities of Scientific Knowledge: Systems, Networks and Potential Dynamics. An Analysis Based on Bibliometric Indicators'. *Urban Studies* 47.9, 1879–1897.
Youn, Hyejin, Luís M. A. Bettencourt, José Lobo, Deborah Strumsky, Horacio Samaniego and Geoffrey B. West (Jan. 2016). 'Scaling and Universality in Urban Economic Diversification'. *Journal of The Royal Society Interface* 13 (114), 20150937. DOI: 10.1098/rsif.2015.0937.

Youn, Hyejin, Deborah Strumsky, Luís M. A. Bettencourt and José Lobo (Apr. 2015). 'Invention as a Combinatorial Process: Evidence from US Patents'. *Journal of The Royal Society Interface* 12 (106), 20150272–20150272. DOI: 10.1098/rsif.2015.0272.