RENIR
REsearch Network on Industrial Resilience

RENIR Working Papers

Measures of Regional Technological Diversity: A Critical Review with an Application to Regional Resilience

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

May 2019, n. 11

Measures of Regional Technological Diversity
A Critical Review with an Application to Regional Resilience

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

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.

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 Herfindahl-Hirschman 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 firm's 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 firm's 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

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

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 is exposed to

[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.

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]

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

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

[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 $\overrightarrow{TCI} = \frac{\overrightarrow{K} - \langle \overrightarrow{K} \rangle}{sd(\overrightarrow{K})}$.

by the ones with few exports, very likely will require a low level of sophistication. Therefore, the 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 ($n > 0$):

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

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

[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

(see Tab. 4 for the list of variables considered and the symbols used to represent each of them henceforth).

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

- 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).

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*.

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).

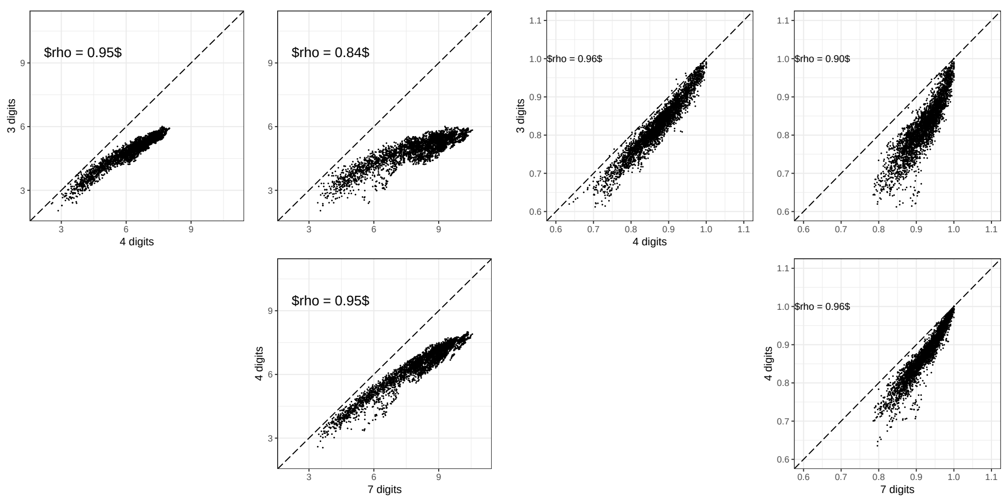
A possible solution, proposed by Stirling following Pielou (1969), is the use of the Shannon Evenness index, defined as

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

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]

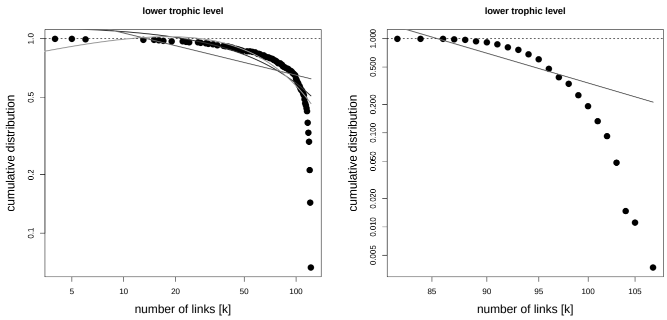
The p-value Coherence A possible solution –similar to another proposed by Nesta (2008, Appendix A)– to overcome the three issues just raised about the *relatedness* matrix, is to substitute the inference based on the value of the statistic J_{il} –or any other possible similar statistic we can think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

The procedure consists of three steps:

[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.

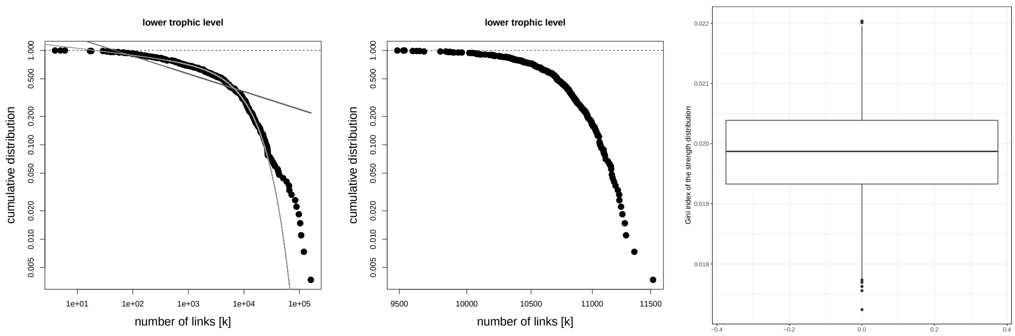
Figure 2: Degree and strength distributions. Data: OECD RegPat, NUTS2, IPC3, 2000–2013. The simulations have been done using the `vegan` R Package (Oksanen et al. 2018; R Core Team 2018). The plots and estimates has been obtained using the `bipartite` R package (Dormann 2011; Dormann et al. 2009, 2008;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

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;

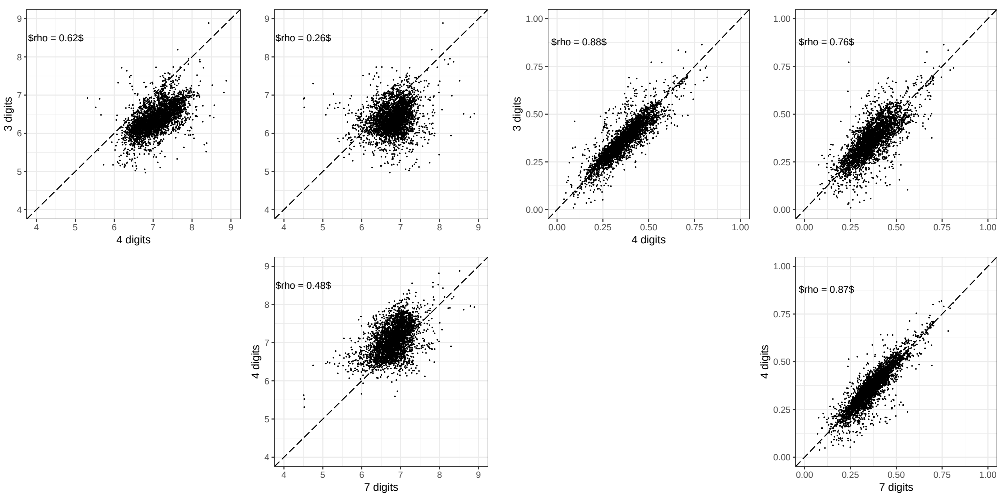
iii Compute $p_{il}(J, H4) = Pr[\hat{J}_{il} \geq J_{il}|H4]$.

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.



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

[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.

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.

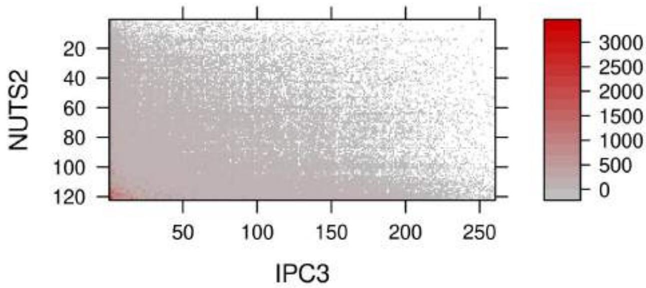
But, since the RTA values are spread over a very large (non-negative) range, I choose to take the arctangent of the values so obtained. In this way, the values are squeezed on a smaller range of possible values. Moreover, the transformations reduce the distance between very high value while preserving 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.

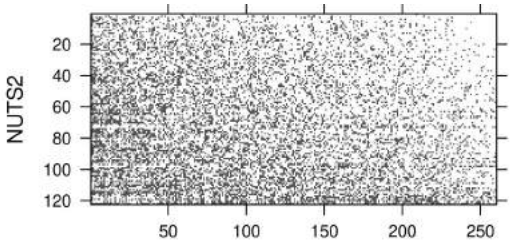
[24]See Bottazzi and Pirino 2010 for a classification of the four possible *null models* (or *null hypotheses*) for bi-adjacency matrices (i.e., bipartite networks). Essentially, the so-called H1 algorithm reshuffles the values of the binary matrix, preserving only the total number of links of the network. 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.

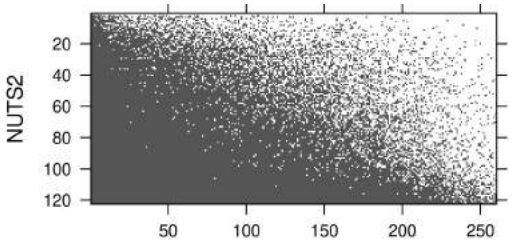
Figure 4: Regions in rows and tech classes in columns. The columns (rows) are ordered from left to right (from bottom to top) according to the node strength, The colours (when present) represent the weight of the link. Data: OECD RegPat, NUTS2, 3 digits IPCs, 2010.



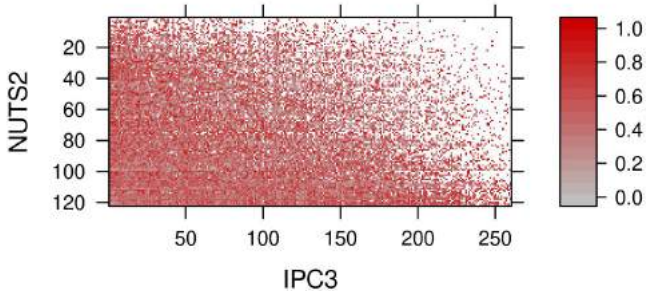
(a) Weighted matrix that represents the number of patents that each region have developed in each tech class.



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



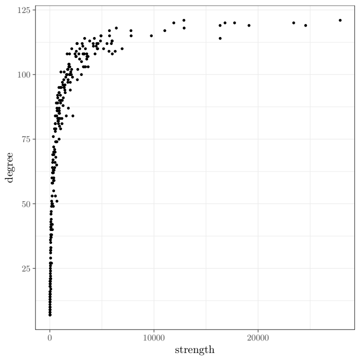
(c) Binary matrix in which a region-tech. class is 1 each time at least 10 patents are observed in a 5-years time span considering a specific pair.



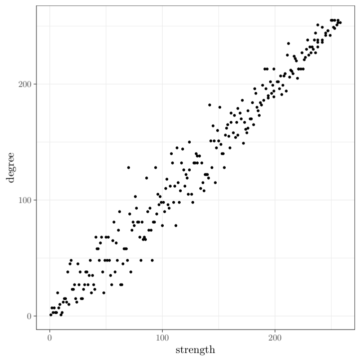
(d) Weighted matrix above named as $MW(r, i)$.

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

(a) Correlation



(b) Rank Correlation



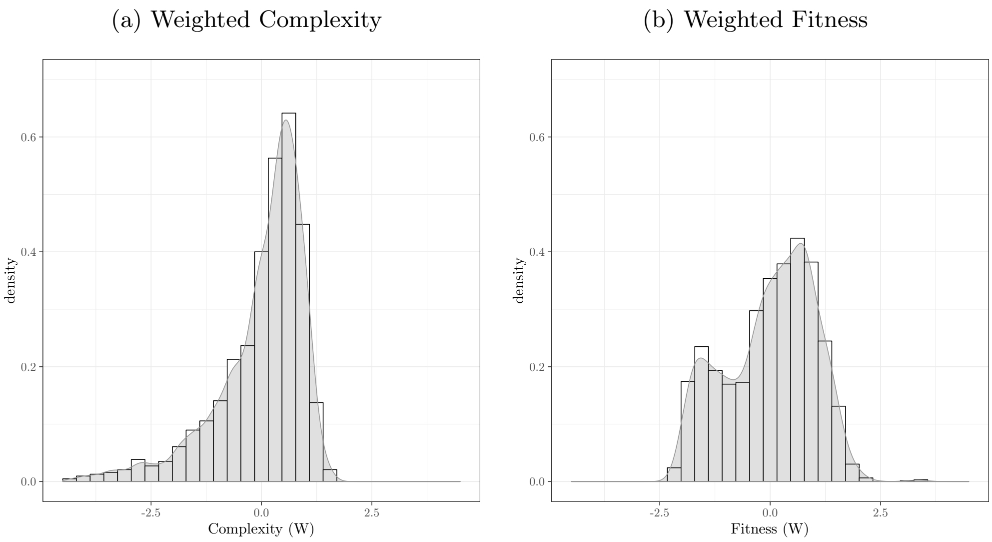
Regional Fitness convergence issue The Fitness index share, in general, most of the problems highlighted above introducing the Economic Complexity index *à la* Hidalgo-Hausmann. But the main problem of this measure is the lack of convergence in many cases. In particular, on the database here used it 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

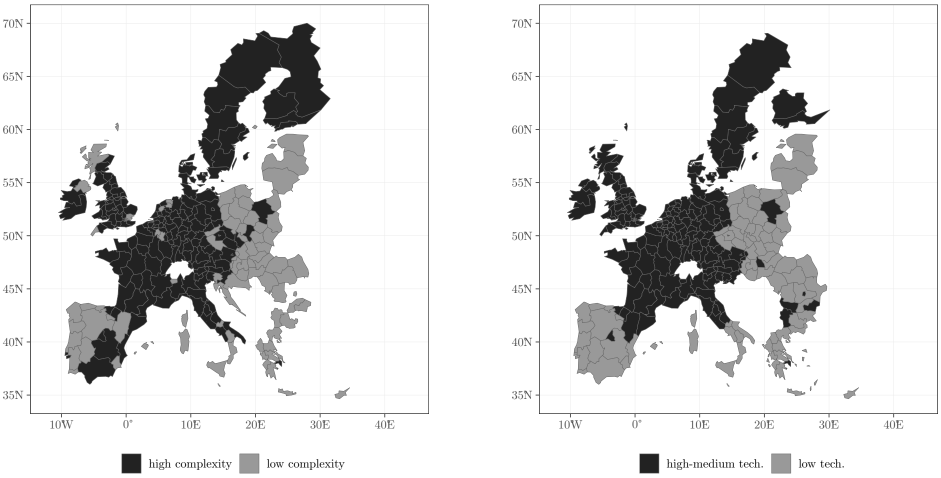
5.4.1 The Regional Technological Complexity Index as a classifier

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:

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

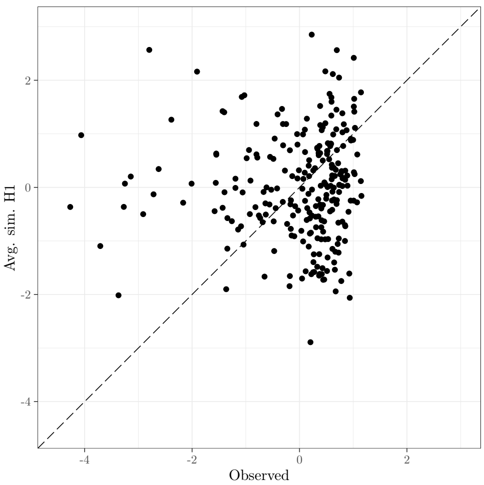
Since the denominator measures the average ubiquity of the technological domains "owned" by a region, a higher value of this fraction means that the region has a more diversified patent portfolio, and that the technological domains possess by the region are also rare ones.

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]

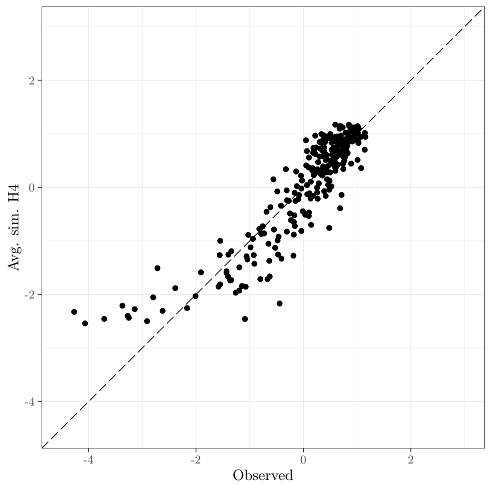
[26] I used the `quasiswap` (H4 *null model*) and `r00` (H1 *null model*) algorithms provided by the `vegan` R Package to get these results (Oksanen et al. 2018; R Core Team 2018).

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

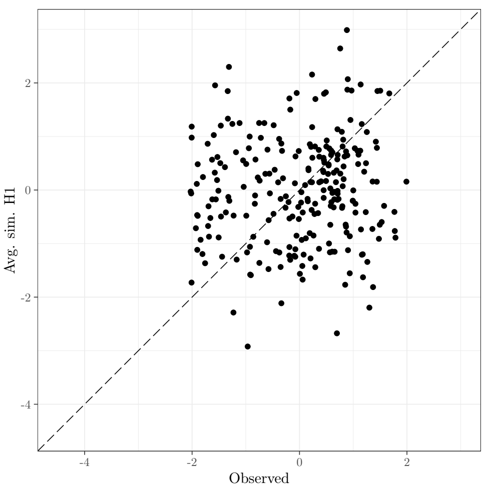
(a) Complexity.
Observed data vs. Average H1 simulations.



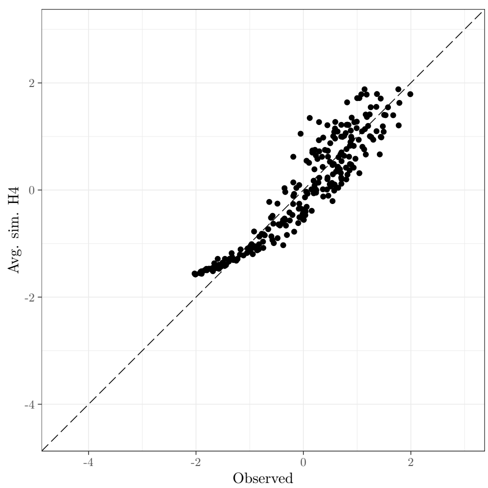
(b) Complexity.
Observed data vs. Average H4 simulations.



(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]

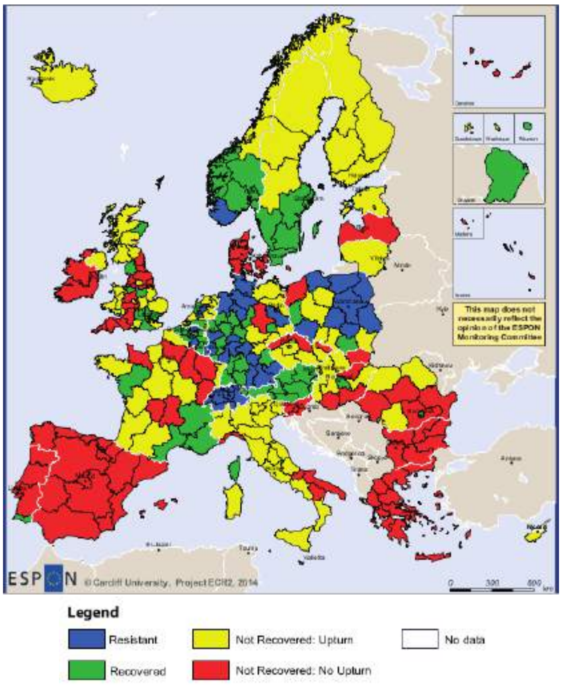
Also looking at the sub-national level, it is possible to see differences within countries and among regions of different countries. As shown in some 2014 publication of the European Program ESPON, the shock has had an asymmetric impact in territorial terms within the EU, with areas that have not been 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.

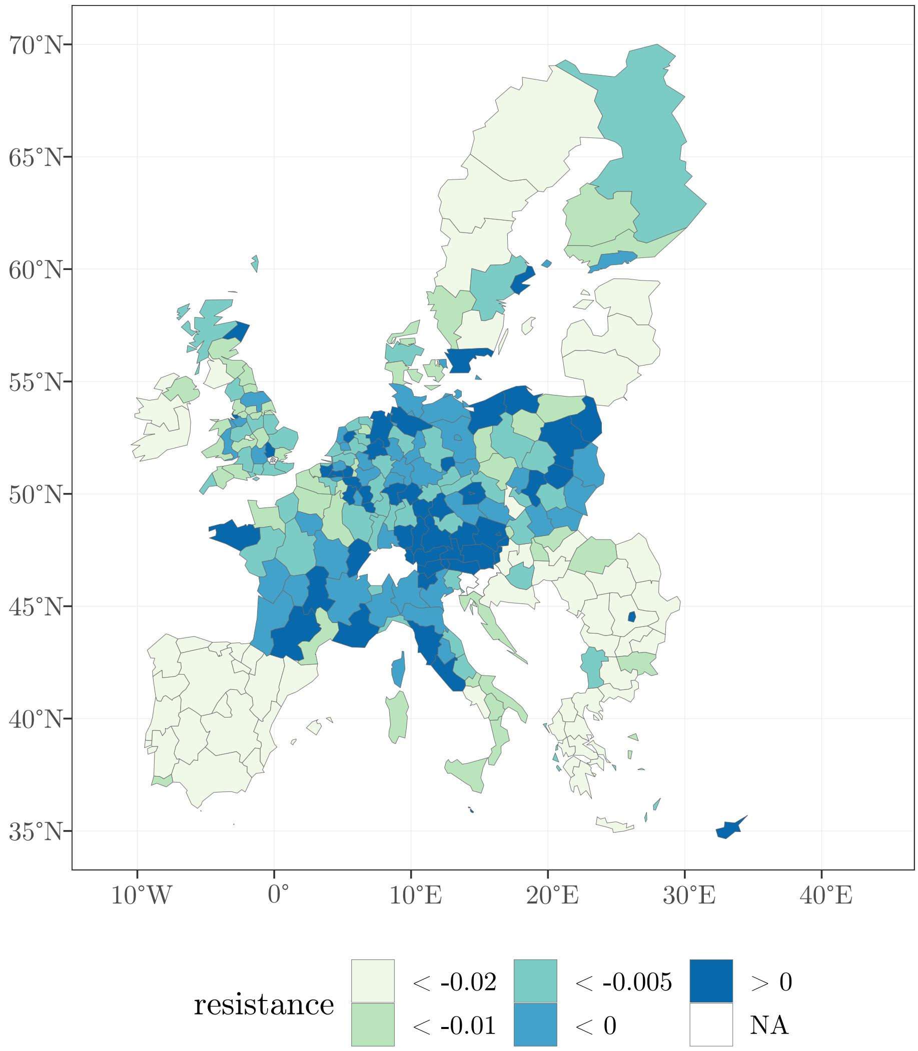
6.2.2 Results

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.

References

Albert, Réka, Hawoong Jeong and Albert-László Barabási (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 (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 (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 (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.

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

Breschi, Stefano, Francesco Lissoni and Franco Malerba (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 (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.

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

Content, Jeroen and Koen Frenken (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 (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 (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 (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 (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 (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 (1975). 'On Measuring Economic Diversification'. *Land Economics* 51.4, 374–378.
Hausmann, Ricardo and César A. Hidalgo (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 (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 (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 (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. (1993). 'Making a Miracle'. *Econometrica* 61.2, 251. DOI: `10.2307/2951551`.

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

Martin, Ron and Peter Sunley (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 (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 (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 (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`.

— (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 (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 (1981). 'Economies of Scope'. *The American Economic Review* 71.2, 268–272.

Parr, John B. (1965). ‘Specialization, Diversification and Regional Development’. *The Professional Geographer* 17 (6), 21–25. DOI: 10.1111/j.0033-0124.1965.021_h.x.
— (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 (2008). ‘The economics of collective knowledge and technological communication’. *The Journal of Technology Transfer* 33.6, 579–599.
— (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 (2010). ‘Resilience, Adaptation and Adaptability’. *Cambridge Journal of Regions, Economy and Society* 3 (1), 59–70. DOI: 10.1093/cjres/rsq001.
Quatraro, Francesco (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 (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 (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 (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 (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. (1947). ‘The Creative Response in Economic History’. *The Journal of Economic History* 7.2, 149–159.

Schumpeter, Joseph A. (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. (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 (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 (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 (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 (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.

— (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. (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.

— (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 (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. (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. (1996). 'Hybridizing Growth Theory'. *The American Economic Review* 86.2, 207–212.

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

Willig, Robert D. (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, Matúš Medo and Yi-Cheng Zhang (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))
```

```r
    # 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

```r
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)
}
```

```r
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.

```r
    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')
```

```r
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"
```

```r
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. ',
                    '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)
            }
        }
    }
```

```
    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)
    }

    # 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)
}
```