Check for updates

Economic complexity theory and applications

César A. Hidalgo [1,2,3]

Abstract | Economic complexity methods have become popular tools in economic geography, international development and innovation studies. Here, I review economic complexity theory and applications, with a particular focus on two streams of literature: the literature on relatedness, which focuses on the evolution of specialization patterns, and the literature on metrics of economic complexity, which uses dimensionality reduction techniques to create metrics of economic sophistication that are predictive of variations in income, economic growth, emissions and income inequality.

Since Adam Smith's pin factory, wealth has been related to the division of knowledge and labour. Yet, even though scholars have long recognized economies as complex systems[1–6], the empirical study of economic complexity only accelerated in the past decade, with the emergence of new data and methods.

Like traditional approaches to economics, economic complexity focuses on the duality between economic inputs and outputs. But, unlike traditional approaches, which either aggregate output — as gross domestic product (GDP) does — or assume the nature of inputs — such as capital, labour and knowledge — economic complexity methods embrace fine-grained data on thousands of economic activities to learn both abstract factors of production and the way they combine into thousands of outputs. This is made possible by applying dimensionality reduction techniques to data on the geography of activities, such as product exports, employment by industry or patents by technology. These techniques — which are related to matrix factorization and are common in machine learning — provide a powerful way to summarize the geography of economic output and can be used to construct predictors of a location's diversification and development potential.

The study of economic complexity accelerated during the last decade thanks to two contributions. The first involved the introduction of metrics of relatedness[7,8], which measure the overall affinity between a specific activity and a location. Relatedness metrics explain path dependencies and predict which activities will grow or decline in a location. They help answer questions such as how 'far' Quito, Kiev or Guadalajara are from having a thriving pharmaceutical industry. The second contribution was the development of metrics of complexity[9]. These use data on the geography of activities (such as exports by country or region, or employment by city and industry) to estimate the availability, diversity and sophistication of the factors or inputs present in an economy. Metrics of complexity extract key information about an economy's capacity to generate[9] and distribute income[10].

Unlike previous approaches to economic growth and development, which attempt to identify individual factors, relatedness and complexity methods are agnostic about the nature of factors. Instead, they try to estimate their combined presence, without making strong assumptions about what these factors may be. For instance, relatedness metrics can be used to estimate the combined presence of inputs that are specific to an activity, no matter if these inputs involve specific forms of labour, capital or institutions. Relatedness can, then, be used to anticipate changes in specialization patterns[7,8,11–18], such as the probability that a location enters or exits an activity.

Complexity metrics apply dimensionality reduction techniques (related to singular value decomposition (SVD); BOX 1) to identify the combinations of factors that best explain the geography of multiple economic activities. Unlike in traditional growth models, which assume the nature of factors, dimensionality reduction techniques can be used to learn factors directly from the data. Economic complexity metrics are useful for predicting economic growth[9,19–25], income inequality[10,26–29] and greenhouse gas emissions[30–33].

Beyond data and methods, the study of economic complexity was motivated by other key trends: the revival of industrial policy, the growth of artificial intelligence (AI) and the development of endogenous growth theory.

Complexity methods grew together with a revival of industrial policy[34,35] and the realization that economic development requires upgrading. Complexity methods help characterize detailed economic structures and provide a quantitative base for industrial policy efforts. Today, these efforts are embodied in Europe's Smart

[1]Center for Collective Learning, Artificial and Natural Intelligence Toulouse Institute (ANITI), University of Toulouse, Toulouse, France.

[2]Alliance Manchester Business School, University of Manchester, Manchester, UK.

[3]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.

e-mail: cesifoti@gmail.com

https://doi.org/10.1038/s42254-020-00275-1

Keywords

- Economic complexity involves the use of machine learning and network techniques to predict and explain the economic trajectories of countries, cities and regions.
- Measures of relatedness — which estimate the affinity between economies and activities — anticipate changes in specialization patterns and explain labour market outcomes, such as income loss and unemployment.
- Economic complexity measures are reduced-dimensionality representations of specialization matrices that explain the geography of hundreds of economic activities.
- Measures of economic complexity explain and predict international and regional variations in income, economic growth, income inequality, gender inequality and greenhouse emissions.
- Economic complexity methods have been validated by studies at multiple geographic scales (from countries to cities) and a variety of economic activities (products, industries, occupations, patents, research papers).
- Relatedness metrics can be unpacked into multiple channels (such as industry, occupation and location-specific knowledge) to understand the drivers of regional diversification.

Specialization Strategy[36–38], China's special economic zones[39–41], Mexico's Smart Diversification strategy or Canada's Superclusters Initiative, to name a few. They include the use of complexity methods to study the structures of economies including the United States[14,42–46], China[26,47–52], Mexico[25,53,54], Canada[55], Russia[56,57], Brazil[58–62], Uruguay[63], Australia[64], Turkey[65–67], Spain[68], Italy[69–73], Paraguay[74] and the United Kingdom[75,76].

The literature in economic complexity also coincides with the rise of machine learning and AI, and can be seen as the use of machine learning in the study of economic geography. Relatedness is akin to recommender systems[77,78], similar to those used to predict clicks or purchases online, but used, instead, to predict the activities that a region is more likely to enter or exit in the future. Likewise, metrics of economic complexity can be seen as the application of dimensionality reduction techniques (such as SVD or principal components) to data on the geography of economic activities. These methods are based on the idea of learning the vectors that are best at explaining the structure of a specialization matrix instead of assuming the nature of the factors of production.

Finally, the study of economic complexity can also be seen as a continuation of endogenous growth theory[79–81]. Endogenous growth theory established that economic growth was the growth of knowledge[79,81]. Knowledge is a non-rival good[79,81] — it can be used simultaneously by multiple people — thus, it is the only productive factor that can grow in per-capita terms. But knowledge is not easy to reproduce or share[82–85]. Its diffusion is limited by geography[82,85], relatedness[7,8,11,19,58,86,87] and social networks[88–90]. Knowledge can also be tacit[91–93] — not codifiable and hard to communicate. Knowledge is also multifarious, being highly specific to an economic task or activity[58,93]. This makes knowledge geographically sticky[94,95]. It follows that the presence of activities in a location carries information about the productive knowledge that a location has accumulated and about the knowledge that the activities present in it require[19,93]. Economic complexity methods attempt to distil that information from fine-grained data.

But the literature on economic complexity is still young. The goal of this Review is to summarize its advances, with a focus on applications of relatedness[7] and complexity[9]. This Review aims to equip those interested in contributing to this literature with a basic understanding of key concepts and contributions. The article is structured as follows. I first put the ideas of economic complexity in a broader scientific context, by connecting them with work in economic geography, innovation and complex systems. Then, I formally introduce the concepts of relatedness and economic complexity. Next, I focus on the applications of relatedness and complexity. This includes the literature unpacking relatedness into multiple dimensions and the literature focused on relatedness and labour. I also review the literature connecting economic complexity to differences in economic growth, income inequality, gender inequality, human development and greenhouse emissions. I finalize by discussing future directions and policy implications.

Complexity and the economy

The science of 'organized complexity'

What is complexity?

In his paper "Science and complexity"[96], the mathematician and AI pioneer Warren Weaver argued that science progressed as people discovered the mathematical languages needed to describe systems of increasing complexity.

For Weaver, science began with the 'science of simplicity.' This involved all systems that could be described using trajectories, such as the motion of a pendulum or the orbit of a planet. Calculus and differential equations were the languages of the science of simplicity, but were ineffective to describe other systems.

The science of 'disorganized complexity' came next. Together with the rise of steam engines and thermodynamics, humans discovered a science that did not rely on the idea of a trajectory but on that of probability. Probability ensembles do not require tracking the identity or trajectory of the elements involved, and, thus, can be used to describe systems like gases.

But Weaver intuited that our reality goes beyond what we can understand using trajectories and probabilities, so he postulated the emergence of a third science, focused on vast systems for which the identity of the elements involved and their patterns of interaction could not be ignored. He called this the science of 'organized complexity.'

At the time of Weaver's publication, the data and methods needed to quantitatively describe complex system were first emerging. But, since then, increases in fine-grained data, computational capacity and analytical methods have produced an early understanding of complex systems that is starting to honour Weaver's vision. This understanding uses matrices or networks to create representations of complex systems that do not ignore the identity of the elements involved or their interactions. These ideas, which are now prevalent in fields such as machine learning and physics, have begun to make their way into economics under the umbrella of economic complexity.

Two foundational ideas of economic complexity — relatedness and complexity — use network methods to create variables that satisfy Weaver's vision. Relatedness

Box 1 | **Economic complexity and production functions**

Economic complexity is closely related to the idea of production function in economics — a function connecting economic inputs (factors) and outputs. This relation is mediated by singular value decomposition (SVD), a factorization technique used to learn the vectors (factors) that best explain the structure of a matrix. When applied to a specialization matrix, SVD can learn the factors that best explain the geography of multiple economic activities.

Consider an $m \times n$ specialization matrix R (technically, the logarithm of R). Its SVD factorization is:

$$R = U \times S \times V \tag{35}$$

where U is an $m \times m$ orthogonal matrix, V an $n \times n$ orthogonal matrix and S an $m \times n$ diagonal matrix (upper panel of figure). SVD can be used to reconstruct the factorized specialization matrix with an arbitrary degree of precision:

$$R = S_{11}U_1 \otimes V_1^{\mathrm{T}} + S_{22}U_2 \otimes V_2^{\mathrm{T}} + \ldots + S_{nn}U_n \otimes V_n^{\mathrm{T}} \tag{36}$$

where U_n and V_n are the nth column vectors of U (principal components) and V (singular vectors), respectively, and $\otimes$ is the outer product.

Now consider a Cobb–Douglas production function connecting economic output (Y) to productive factors (capital (K) and labour (L)) and productivity (A):

$$Y = AK^{\alpha}L^{\beta} \tag{37}$$

where α and β are elasticities. Generalized to n factors ($F_1, F_2, \ldots, F_n$), the production function takes the form:

$$Y = AF_1^{\alpha_1}F_2^{\alpha_2}F_3^{\alpha_3}\ldots \tag{38}$$

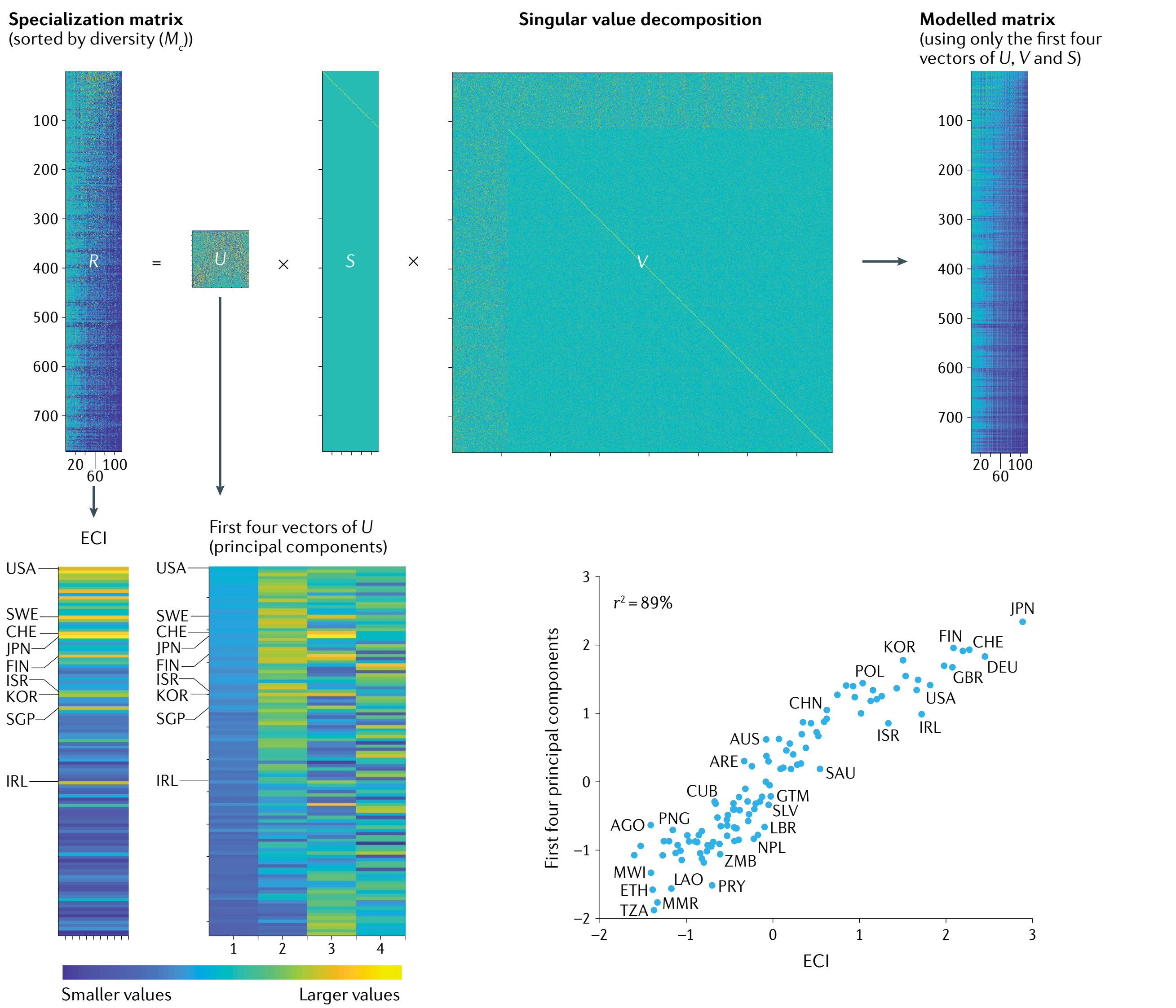
which has a logarithm:

$$y = a + \alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_n f_n \tag{39}$$

with lowercase letters denoting logarithms.

The SVD of a specialization matrix is equivalent to a production function for multiple outputs, where the factors ($U_1, U_2, \ldots, U_n$ for locations and $V_1, V_2, \ldots, V_n$ for activities) and the elasticities ($S_{11}, S_{22}, \ldots$) are all learned from data. Since SVD provides the best way to approximate a matrix with a limited number of vectors, it represents the best possible factor decomposition of a specialization matrix.

The economic complexity index (ECI) is closely related to the leading vectors obtained through SVD ($r^2 = 89\%$ in the example used in the figure), showing that economic complexity summarizes learned vectors or factors that are optimal at explaining the geography of multiple economic activities.



metrics preserve the identity of locations and activities while leveraging information about interactions, such as information about activities in other locations. Metrics of economic complexity measure the presence of multiple factors simultaneously, not by using aggregation (simplicity) or distributions (disorganized complexity) but by using dimensionality reduction techniques that preserve the identity of the elements involved and consider their interactions[9].

A brief history of relatedness

The introduction of network methods invigorated the study of relatedness, but relatedness was not a new idea to economics. Relatedness is tied to the idea of absorptive capacity[97] — the notion that a firm's ability to absorb new knowledge is a function of its prior level of related knowledge. This idea, when applied to knowledge diffusion[82–86,98,99], implies that the success of regions entering an economic activity depends not only on geographical and cultural forms of proximity (such as distance or language) but on the cognitive and technological proximity between the new activity and a region's prior activities[86,100–102].

Earlier studies on relatedness focused on the growth of locations specialized in bundles of related or unrelated activities[103,104]. This question — explored using metrics of agglomeration[98,103,105] or the hierarchy of administrative classifications[104] — was motivated by two competing theories of spillovers (the excess benefits or costs 'spilling over' among economic activities): within-sector or Marshall–Arrow–Romer (MAR) spillovers, and between-sector or Jacobs spillovers. MAR spillovers are considered key for productivity and short-term growth, whereas Jacobs spillovers are considered key for innovation and long-term growth[106].

Relatedness spans the continuum between MAR and Jacobs, because related activities are neither exactly the same nor completely different[104,107]. This distinction is important, because identical activities tend to compete for customers and resources, whereas distant activities provide scant learning opportunities[107]. Relatedness captures the intuition that learning requires interactions among activities that are similar, but not similar enough to be competing. It is about spillovers between different activities that, nevertheless, share combinations of inputs, knowledge and routines. This brings relatedness close to the literature on industrial clusters, which also considers intersectoral links that go beyond shared knowledge[108,109].

The quantitative study of relatedness advanced with the introduction of measures of proximity[7] or coagglomeration[98], which connect pairs of activities, and later with the introduction of methods to estimate the affinity or relatedness between a location and an activity (instead of pairs of activities)[7]. Proximity measures have been used to create maps of similar products[7], industries[11,12,58], technologies[13,14,42], occupations[16,17,46,110], research areas[15,18], sports[111] and music[112]. These networks reveal differences in the topologies of similar activities and show that relatedness predicts the activities that a location will enter or exit in the future[7,12,15,19,41]. Indeed, the 'principle of relatedness'[8] is a statistical law stating that the probability that a location enters or exits an activity is correlated with the presence of related activities.

A brief history of complexity metrics

The second accelerator of the study of economic complexity was the development of metrics of complexity[9]. Unlike the development of relatedness, the discovery of economic complexity represented a more significant departure from the previous literature.

Previous literature focused on the relevance of productive structures had produced a variety of indicators of technological sophistication[113–115]. But quantitative efforts did not rely on iterative or dimensionality reduction methods (an exception is the study reported in REF.[116]) but on indicators that averaged over other indicators, such as data on patents, human capital[117–119] or income[120].

Complexity metrics were originally discovered using international trade data and validated by their ability to predict future economic growth[9,19]. This finding was quickly replicated[20–24], stimulating the exploration of similar metrics[121–126], as well as its application to non-export datasets, such as data on patents by technology for cities in the United States[44,127] or employment by location and industry for Mexico[25] and the United States[45,128].

More recently, scholars have begun using economic complexity to explore additional implications of economic structures, such as geographic differences in income inequality[10,26–29], human development[129,130] and greenhouse gas emissions[30–33,131].

Today, economic complexity indicators are published regularly in online tools, such as the Observatory of Economic Complexity[132], or official government websites, such as Data México from Mexico's Secretary of the Economy.

Current directions

Work in economic complexity has shifted from using relatedness and complexity as explanatory factors to focusing on unpacking these metrics and exploring their factors, causes and consequences. There is now a vibrant literature centred on efforts to unpack relatedness into multiple dimensions[46,58,133–138]. This literature has examined how different forms of relatedness (such as industry-specific and occupation-specific knowledge[58]) affect the diversification, growth and survival of firms, or how policies and institutions modulate the role of relatedness in diversification[135,136]. The idea of relatedness has also been extended to datasets that combine multiple types of activities (such as patents, papers and products)[139,140] and to bilateral trade relationships[134]. There are also efforts focused on exploring the role of relatedness on labour outcomes[141], firm growth[142], sustainability[143–147] and entrepreneurship[148]. Finally, the study of relatedness also involves the development of strategies designed to optimize industrial diversification paths[149] and the evaluation of industrial policy efforts[150,151].

Work on economic complexity has also broadened, and is beginning to consider the consequences of complexity on a variety of outcomes, such as income inequality[10,26–28,152], gender inequality[153,154], human development[129,130,155], output volatility[156], productivity[157], health[158] and greenhouse gas emissions[30,31,131]. This work has also explored factors that contribute to the growth of complexity, such as different modes of taxation[159], intellectual property rights[160], institutions[161], demographics[162,163], transportation[50], digital connectivity[164] and structural reforms[165].

Together, these findings are helping us expand our understanding of the geography of economic activities, and its consequences, in a way that honours Weaver's vision.

Basic definitions

Specialization matrices

We organize data on the geography of activities using matrices that connect locations (such as countries, cities or regions; denoted with the subscript c) and activities (such as products, industries or technologies; denoted with the subscript p). Examples of location–activity matrices include exports by region and product, total payroll by city and industry, or patents by metropolitan area and technology. In this notation, such a matrix is:

$$X_{cp} = \text{volume of activity } p \text{ in location } c \quad (1)$$

where the volume of an activity may refer to exports, sales, total payroll, value added, employment or other quantities.

Because matrices on the geography of activities include units of observation with sizes that are not readily comparable (for instance, China and Uruguay), they need to be normalized into specialization matrices. A specialization matrix (R) is defined by dividing each entry of X_{cp} by the sum of its respective row and column. This metric is known as the location quotient or revealed comparative advantage. Defining the sum of the whole matrix as $X = \sum_{cp} X_{cp}$ and using Einstein notation, so missing indices indicate summed variables (for any matrix A_{ij}, $A_i = \sum_j A_{ij}$), the specialization matrix R_{cp} is:

$$R_{cp} = X_{cp}X/X_cX_p \quad (2)$$

R_{cp} is the ratio between the observed (X_{cp}) and expected (X_cX_p/X) level of economic activity in a location. Locations with $R_{cp}>1$ are considered to be specialized in activity p.

Sometimes, it is useful to normalize matrices using population data[124,125,166]. If P_c is the population of location c and $P=\sum_c P_c$ is the population of the world, then a population-normalized specialization matrix[166] is:

$$R_{cp}^{\text{pop}} = X_{cp}P/P_cX_p \quad (3)$$

Unlike R, which can only be >1 for about half of all activities (because the sum of activities in a same location is in its denominator), R^{pop} can be >1 for all activities. This normalization helps remove noise produced by fluctuations in commodity prices, seasonal employment or currency exchange rates. For instance, when oil prices decrease, the revealed comparative advantage of non-oil activities in oil-producing economies automatically increases. Another common method to reduce noise is to time-average specialization matrices over 3–5 years. Finally, because R and R^{pop} are defined as ratios, they should be log-transformed when used in statistical models.

In addition, we define the binary specialization matrix M as:

$$M_{cp} = \begin{cases} 1 & \text{if} \quad R_{cp} \geq R^* \\ 0 & \text{if} \quad R_{cp} < R^* \end{cases} \quad (4)$$

where $R^*=1$ when using R and $R^*=0.25$ when using R^{pop} (REF.[166]). M helps remove excess variation by focusing only on significant presences ($M_{cp}=1$) and absences ($M_{cp}=0$).

The marginals of M count the number of activities present in a location (diversity) and the number of locations where an activity is present (ubiquity). Formally:

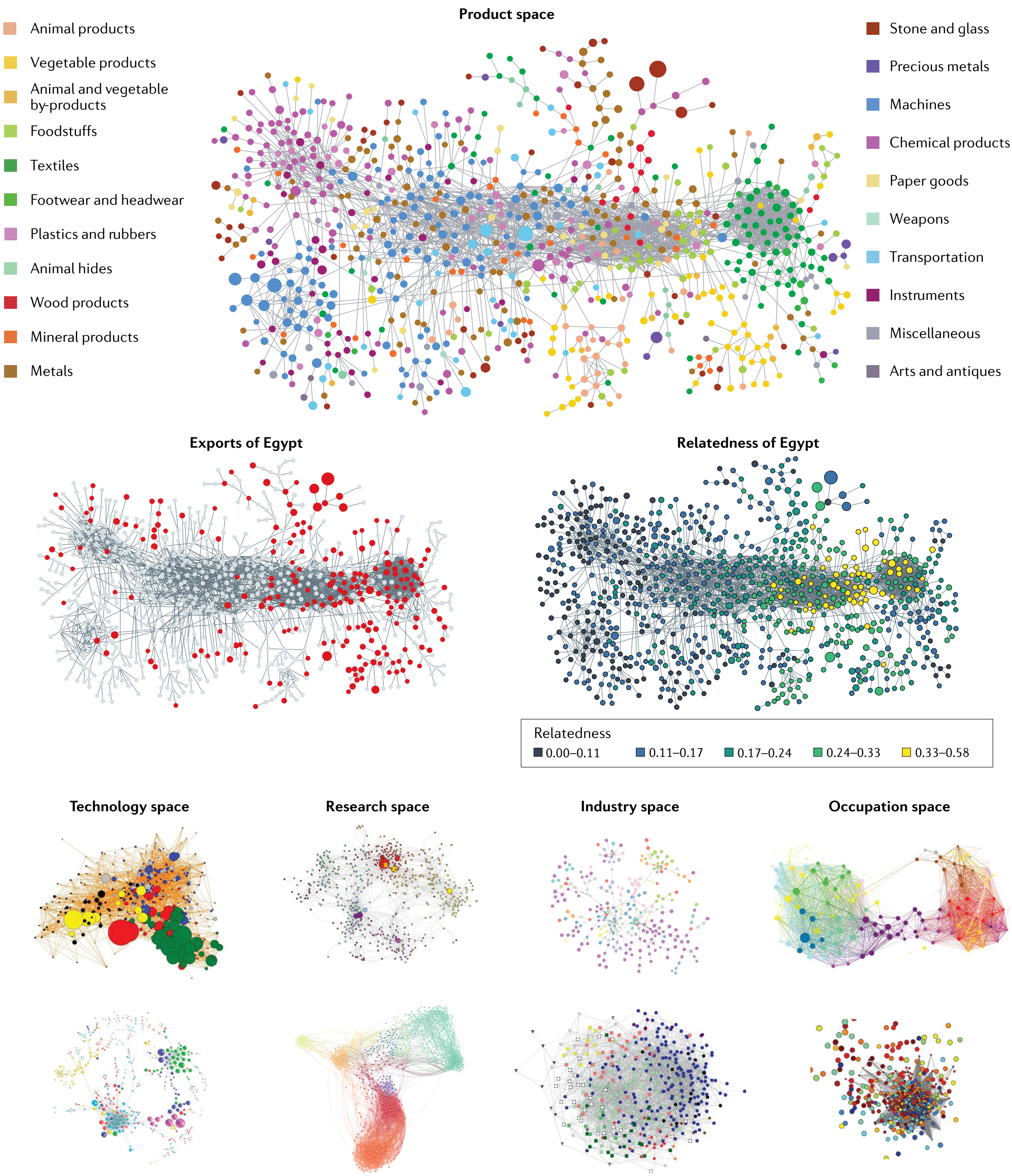
$$M_c = \sum_p M_{cp} = \text{diversity} \quad (5)$$

$$M_p = \sum_c M_{cp} = \text{ubiquity} \quad (6)$$

A property of these geographic matrices is that the average ubiquity of the activities present in a location tends to correlate negatively with that location's diversity[9]. This fact is related to the matrix property known as nestedness[166] and can be seen as evidence that more complex knowledge diffuses with more difficulty[44], and, hence, is only available at a few diverse locations.

It is worth noting that these datasets are limited by the coarsening, frequency and universality of the administrative classifications and geographic boundaries used to define them. Some datasets, such as those on product exports, are extremely fine-grained (>5,000 product categories when using six-digit Harmonized System data), whereas others are notorious for their coarsening and lack of quality (for instance, Service Trade data, which has a few dozen categories). Trade data are commonly used because they use internationally comparable classifications, something that is not true for data on industries or occupations, which use classifications limited to countries or trade blocks (such as the North American Free Trade Agreement (NAFTA) or Mercosur). Additionally, some geographic units may be more meaningful from an economic perspective, such as metropolitan areas, while others are purely administrative (counties, postal codes).

Although these limitations are not particular to work in economic complexity, they need to be considered. In fact, even when data are of relatively good quality, metrics can be noisy when applied naively to matrices that include disparate economic activities and locations. For instance, international trade data include countries as small as Tuvalu, an island nation with <12,000 people, and China, which has a population five orders of magnitude larger. Thus, instead of using all data, it is appropriate to cut the left tails of distributions. For instance, in four-digit international trade data, it makes sense to consider only economies that export more than



one billion USD and have a population of at least one million. Likewise, it is common to disregard products with small export volumes (for instance, <500 million in global exports at the four-digit level). These tricks of the trade are important to perform meaningful comparisons and are used in all data types (including products, patents or industries).

Relatedness

Relatedness measures the affinity between a location and an activity. Yet, since a good measure of affinity is expected to predict changes in specialization patterns, relatedness can be defined as a predictor of changes in specialization that is specific to a location–activity pair and that goes beyond the naive prediction (no change).

◀ Fig. 1 | **Relatedness is constructed using networks that connect similar activities.** These networks have topologies that depend on the type of activity considered. The network of similar products (product space) has well-defined clusters and a periphery composed of primary products. Red nodes represent products exported by Egypt (with revealed comparative advantage >1). The relatedness figure colours nodes according to relatedness density, which measures the affinity between an activity and an economy's current productive structure (Egypt in this case). The other figures show networks constructed using patent data, research publication data and data for industries and occupations. Colours represent categories (technologies for patents, research areas for papers, industrial sectors for industries and occupational categories for occupations). Technology space 1 reprinted with permission from REF.[13], Taylor and Francis. Technology space 2 reprinted with permission from REF.[14], OUP. Research space 1 reprinted with permission from REF.[15], Springer Nature Limited. Research space 2 is reprinted from REF.[18], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Industry space 1 reprinted with permission from REF.[58], PNAS. Industry space 2 reprinted with permission from REF.[12], Wiley. Occupation space 1 reprinted with permission from REF.[17], AAAS. Occupation space 2 is reprinted from REF.[16], CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Formally, relatedness ω_{cp} can be defined as a predictor of a matrix of specialization that satisfies:

$$R_{cp}(t+\mathrm{d}t)=R_{cp}(t)+B\omega_{cp}(t)+\ldots \qquad (7)$$

where B is a positive and significant coefficient. This same idea is readily applied to other specialization matrices (such as M and R^{pop}) and can include data from multiple time periods (for instance, $R_{cp}(t)$, $R_{cp}(t-1)$).

Although there are, in principle, multiple ways to measure relatedness, during the past decade, most work has built on a metric known as relatedness density[7]. Relatedness density looks at the number of similar activities that are present in a location. It can be readily extended to include activities in similar locations[167] or extended to multiple groups of activities (for instance, by combining trade and patent data)[139,140].

To define relatedness density, we first define a measure of proximity[7]. Metrics of proximity connect pairs of activities ($\phi_{pp'}$) or pairs of locations ($\phi_{cc'}$). The first metric yields the idea of 'product space'[7], 'industry space'[11], 'technology space'[13] or 'research space'[15] (FIG. 1). The second metric yields the idea of a 'country space' or 'producer space'[167].

There are multiple ways to define proximity. These include looking at the collocation or coagglomeration[98] of activities, using metrics such as the minimum conditional probability[7]:

$$\phi_{pp'}=\frac{\sum_c M_{cp}M_{cp'}}{\max(M_p, M_{p'})} \qquad (8)$$

or by looking at the correlation between rows or columns of a specialization matrix[167]:

$$\phi_{cc'}=\mathrm{corr}(\log(R_{cp}),\ \log(R_{c'p})) \qquad (9)$$

Proximity has also been measured by looking at industries that produce products in the same plants[11], by mapping labour flows among industries or occupations (skill-relatedness)[12,58,168,169] or by looking at the ratio between observed and predicted collocation patterns[11].

Proximity networks have been constructed for a variety of datasets, revealing differences in the patterns of these networks (FIG. 1). For instance, networks connecting products (such as the product space) tend to be characterized by some well-defined clusters (such as garments, machinery or electronics) and a clear separation between a core and a periphery. Networks connecting research activities or patents tend to have a ring structure[15,44,170]. Networks connecting industries tend to have a dumbbell structure, with a cluster for services and another for manufacturing. Regardless of the structure, these networks are the basis for measuring the relatedness or affinity between activities and locations[7].

Using any measure of proximity, we can define relatedness density as the fraction of related activities present in a location:

$$\omega_{cp}=\frac{\sum_{p'} M_{cp'}\phi_{pp'}}{\sum_{p'}\phi_{pp'}} \text{ or } \omega_{cp}=\frac{\sum_{c'} M_{c'p}\phi_{c'c}}{\sum_{c'}\phi_{c'c}} \qquad (10)$$

This simple form allows variations, such as using the single most similar activity[7] or squaring proximity values, such as using weights equal to $\phi_{pp'}{}^2$ to increase the weight of more proximate activities. For the most part, different measures of proximity produce correlated estimates of related density, making the exact functional forms used for $\phi_{pp'}$ or ω_{cp} a choice made by researchers at the time of implementation.

One important caveat, however, is that relatedness density (Eq. (10)) is highly correlated with diversity (M_c). This correlation is not a problem if relatedness is used to make predictions for a single location or activity. But it can be problematic when comparing across locations or activities. In that case, one should use measures of relative relatedness[171], which can be constructed by taking the Z-score of the values or dividing density by the diversity and ubiquity of a location and activity.

Finally, although the literature has used mostly relatedness density as an indicator, it is possible to estimate relatedness using other methods, such as latent factors for locations and activities in the latent comparative advantage method[172], an Indian buffet process[173] or SVD[174].

Economic complexity

Economic complexity metrics measure economic capacity using methods that are related to dimensionality reduction (SVD or principal component analysis). They also represent generalized dimensionality reduced production functions (BOX 1). Economic complexity metrics can be used to measure the presence of multiple economic factors in a way that is agnostic about what these factors might be.

Formally, the complexity K_c of location c and the complexity K_p of activity p can be defined as a function of each other. This means that they are solutions to a set of coupled equations:

$$K_c=f(M_{cp}, K_p), \qquad (11)$$

$$K_p=g(M_{cp}, K_c). \qquad (12)$$

These equations state that the complexity of a location is a function of the complexity of the activities that

Table 1 | **Rankings of economic complexity**

Rank	Economic complexity rankings		
	US metro areas: payroll by industry (2018)	**US metro areas: patents by technology (2018)**	**Countries/territories: exports (2018)**
1	San Jose–Sunnyvale–Santa Clara, CA	San Jose–Sunnyvale–Santa Clara, CA	Japan
2	San Francisco–Oakland–Hayward, CA	Austin–Round Rock–San Marcos, TX	Switzerland
3	Boston–Cambridge–Newton, MA–NH	San Francisco–Oakland–Fremont, CA	Chinese Taipei
4	Los Angeles–Long Beach–Anaheim, CA	Boise City–Nampa, ID	Germany
5	Seattle–Tacoma–Bellevue, WA	Rochester, MN	South Korea

are present in it, and vice versa. This system is equivalent to a set of decoupled self-consistent equations (one for locations and another one for activities):

$$K_c = f(M_{cp}, g(M_{cp}, K_c)), \tag{13}$$

$$K_p = g(M_{cp}, f(M_{cp}, K_p)). \tag{14}$$

These general equations already eliminate important alternatives. For instance, they rule out metrics of market concentration, like those used in disorganized complexity approaches, such as the Shannon information entropy or the Herfindahl–Hirschman index (HHI). Metrics of diversity or concentration disregard information about the identity of elements involved by failing to couple locations and activities (for instance, HHI takes the form $K_c = f(M_{cp})$).

The idea of measuring complexity using a set of coupled equations was introduced[9] in 2009, using simple averages for *f* and *g*. The resulting metrics are known as the economic complexity index (ECI; K_c) and the product complexity index (PCI; K_p). These are defined by the following system of equations:

$$K_c = \frac{1}{M_c} \sum_p M_{cp} K_p \tag{15}$$

$$K_p = \frac{1}{M_p} \sum_c M_{cp} K_c \tag{16}$$

which is equivalent to the reciprocal average method used in ecology[175,176]. Because both equations are linear, they imply self-consistent equations of the form:

$$K_c = \widetilde{M}_{cc'} K_{c'} \tag{17}$$

$$K_p = \widetilde{M}_{pp'} K_{p'} \tag{18}$$

with:

$$\widetilde{M}_{cc'} = \sum_p \frac{M_{cp} M_{c'p}}{M_c M_p} \tag{19}$$

and:

$$\widetilde{M}_{pp'} = \sum_c \frac{M_{cp} M_{cp'}}{M_c M_p} \tag{20}$$

The solutions to this system are the eigenvectors of $\widetilde{M}_{cc'}$ and $\widetilde{M}_{pp'}$, which are also principal components of M_{cp} (REF.[176]).

Because both $\widetilde{M}_{cc'}$ and $\widetilde{M}_{pp'}$ are technically row-stochastic matrices[176,177] ($\sum_{c'} \widetilde{M}_{cc'} = 1$ and $\sum_{p'} \widetilde{M}_{pp'} = 1$, also known as Markov or transition matrices), their first eigenvector is a vector of 1s, making the second eigenvector the leading metric of economic complexity.

The second eigenvector of a stochastic matrix is the leading correction to the equilibrium distribution and represents a partition of the data. In economic terms, the ECI is the vector that is best at dividing economies into groups based on the activities that are present in them. As discussed in BOX 1, economic complexity is intimately connected to SVD, a matrix factorization technique that provides the best way to explain the structure of a matrix.

To eliminate constant factors, metrics of economic complexity are commonly normalized using a Z-transform (valid for the ECI, which does not follow a heavy-tailed distribution):

$$\mathrm{ECI}_c = (K_c - \mathrm{mean}(K_c))/\mathrm{stdev}(K_c), \tag{21}$$

$$\mathrm{PCI}_p = (K_p - \mathrm{mean}(K_p))/\mathrm{stdev}(K_p), \tag{22}$$

ECI values >0 represent locations with a complexity that is larger than the average location in the dataset (similarly for PCI).

TABLE 1 and FIGS 2,3 show rankings of economic complexity for countries, regions and cities (metropolitan statistical areas, hereafter, MSAs) in the United States. These were calculated using data on payroll by industry, patents by technology and exports by product. At the international level, economic complexity rankings are dominated by technologically advanced economies, such as Japan, Switzerland, Chinese Taipei, Germany and South Korea. In the United States, Silicon Valley (San Jose) leads both rankings, but these, nevertheless, show interesting differences. Payroll data rankings, which are related to employment, are led by large technologically advanced cities (San Francisco, Boston, Seattle and Los Angeles). Complexity rankings based on patent data also include small cities that are well-known enclaves of innovation, such as Boise, Idaho, the home of Micron, the leading manufacturer of computer memory and solid-state technology in the United States, and Rochester, Minnesota, home of 3M, a company well known for its innovations in materials science.

Economic complexity, when defined as iterative averages (ECI), satisfies important properties. First, the complexity of a location only increases when a location adds an activity that is above its current average. This feature avoids a metric of complexity that increases when adding activities that are of low sophistication. Second, because $\widetilde{M}_{cc'}$ and $\widetilde{M}_{pp'}$ are normalized by the sum of rows (M_c) and columns (M_p), they give more weight to unexpected coincidences (coincidences with more information, in the information theory sense). Third, because these complexity metrics come from diagonalizing proximity matrices, they provide similar values for locations that have similar patterns of specialization. Fourth, the ECI also correlates strongly ($r \sim [83–93\%]$) with traditional indicators of technological sophistication[178], such as the indexes proposed in REFS[117–119]. But, unlike these indexes, which are weighted averages of various metrics of patents, technology diffusion and human capital, the ECI does not require aggregation through predefined weights (the weights and the factors to aggregate are determined endogenously by the method)[178]. The ECI also does not need priors about which activities are more technologically sophisticated, because that is inferred directly from the data. Finally, metrics of complexity are connected to the idea of relatedness, because they are derived from the spectrum of proximity matrices ($\widetilde{M}_{cc'}$ and $\widetilde{M}_{pp'}$ are asymmetric proximity matrices).

These properties have made the ECI and the PCI attractive methods to estimate complexity using trade data[9,19], patent data[44], occupation data[128], industry data[45] and cultural consumption data[179].

Since the introduction of the ECI and the PCI[9], variations to the method have been proposed. Examples include the use of Google PageRank[180] on a population-based matrix of specialization (R^{pop})[125], variations in the functional forms for f and g[121,122,181] and methods that expand the calculations of complexity to include data on products and patents, such as the tripartite approach introduced in REFS[123,124] or the innovation-adjusted ECI[126].

These variations yield similar results. For instance, comparing the ECI and log fitness[121] using exports data shows a correlation of $r^2 = 86\%$ (REF.[10]). Using employment data for cities in the United States, the ECI and log fitness correlate with $r^2 = 90.3\%$ (REF.[45]). Using data for

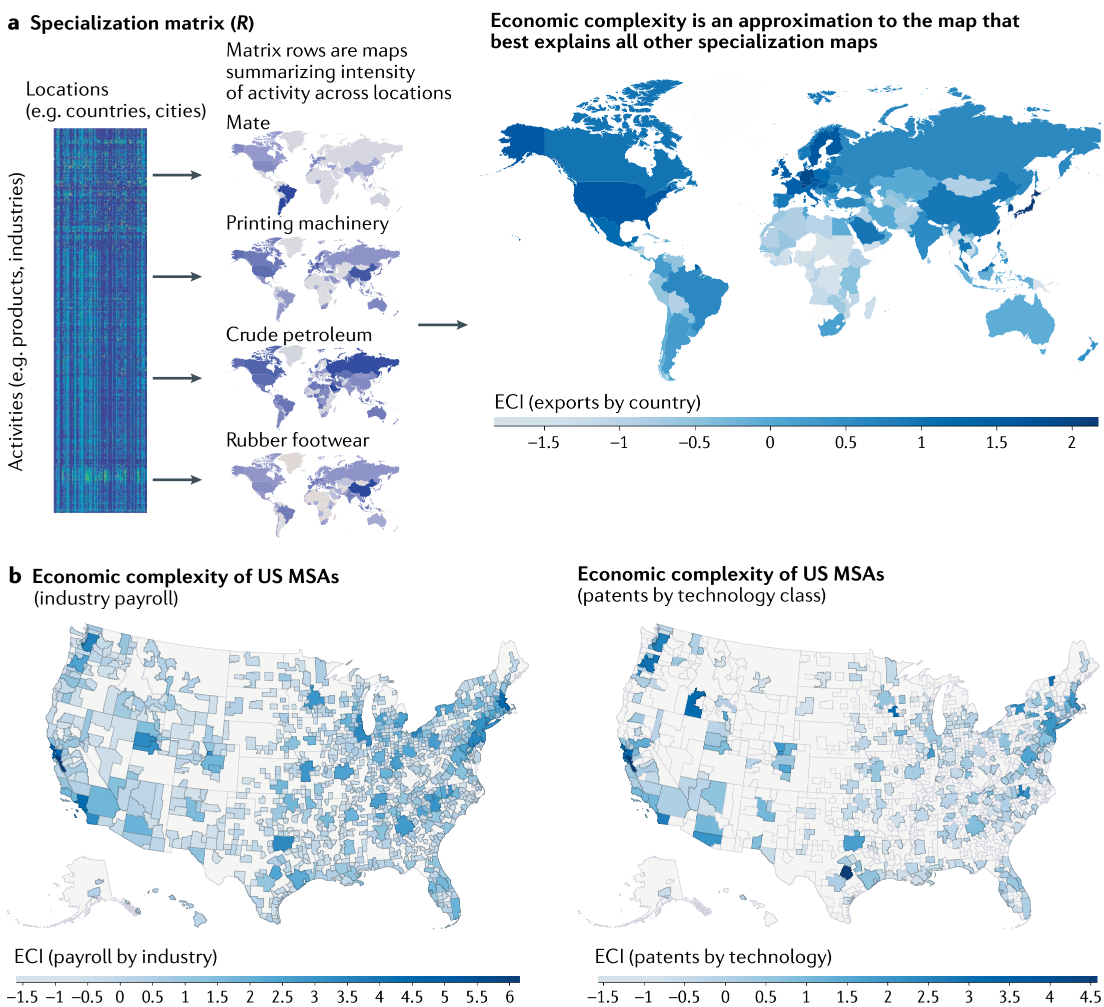


Fig. 2 | **Economic complexity.** Map of the economic complexity index (ECI) estimated using international trade data (panel **a**), as well as maps for the ECI of metropolitan statistical areas (MSAs) in the United States estimated using payroll by industry and patents by technology (panel **b**).

companies that traded in Shanghai and Shenzhen, the ECI and fitness correlate with $r^2 = 76\%$ (REF.[49]). Similarly, both variables have been found to be almost identically predictive of economic growth[20].

Common misconceptions

Because economic complexity is not a simple idea, its understanding has been mired by some common misconceptions.

One misconception is to equate economic complexity to measures of export diversity or concentration. This is wrong on two accounts. First, economic complexity, as measured by the ECI, is orthogonal[176] — or nearly orthogonal — to measures of diversity or concentration. This orthogonality can be seen by comparing a popular measure of concentration, the HHI, with the ECI (FIG. 4).

Second, economic complexity is not about exports or trade (the use of trade data is convenient but not essential). It is a dimensionality reduction technique that summarizes the vectors that best explain the geography of thousands of economic activities and has been applied successfully to data on patents[44], occupations[128], industries[45] and cultural consumption[179]. This means that trying to validate, or invalidate, economic complexity by using results from the export diversity literature is a non sequitur.

At the international level, it is also interesting to see that economic complexity does not correlate with population, meaning that it provides a measure of factors driving the geography of activities that is independent of population (FIG. 4). However, this is not the case at the subnational level (for instance, for US MSAs).

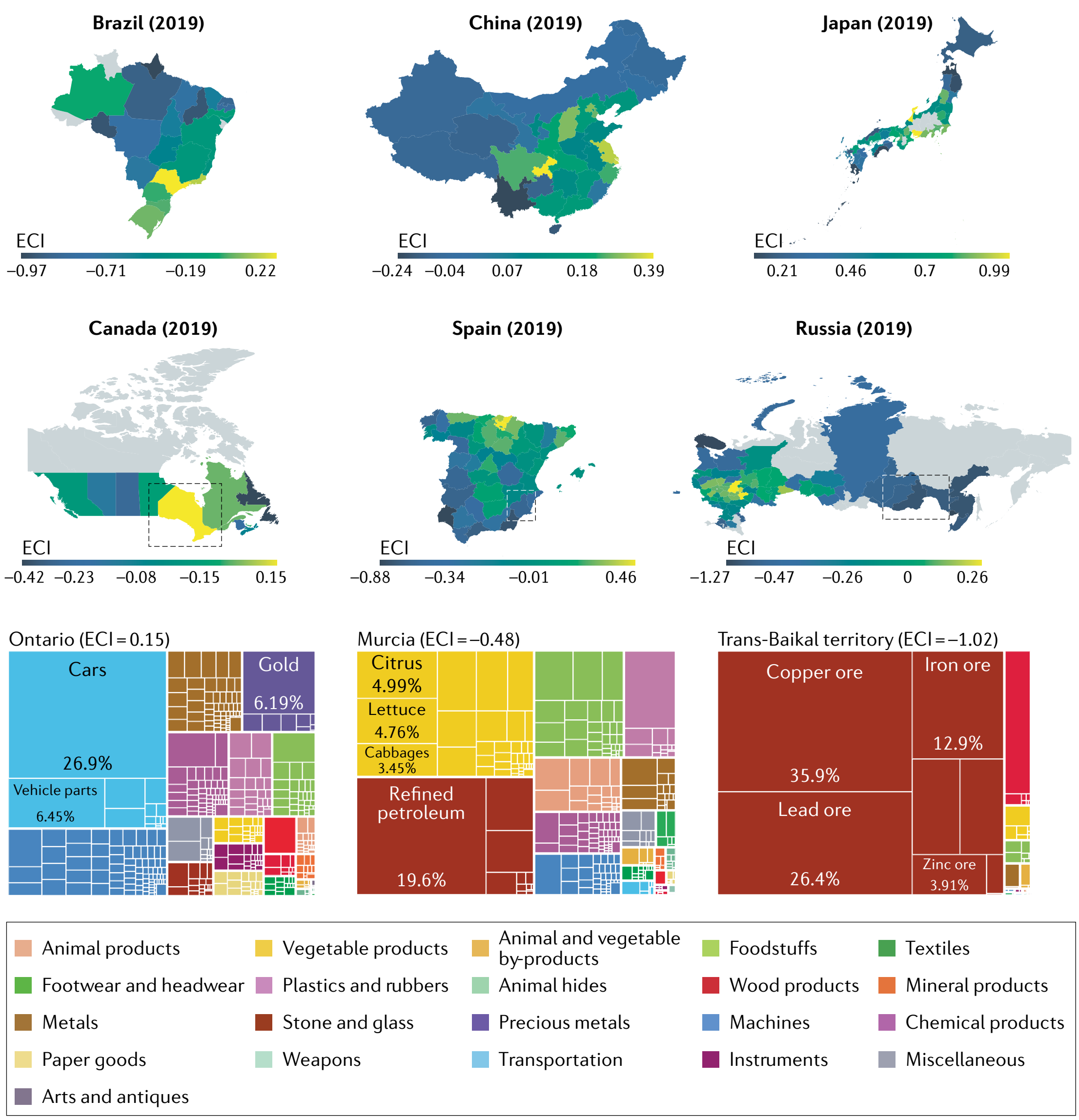


Fig. 3 | **Economic complexity index for regions in Brazil, China, Japan, Canada, Spain and Russia.** These maps use international trade data classified according to the Harmonized System and estimate subnational-level economic complexity indexes (ECIs) (provinces, regions) using product complexity index values calculated at the global level and a binary specialization matrix (M_{cp}) derived from comparing local and global exports for each product ($R_{cp} = (X_{cp}^{\text{local}}/X_c^{\text{local}})/(X_p^{\text{world}}/X^{\text{world}})$), where R_{cp} is the specialization matrix, X_{cp} is the volume of activity p in location c and local variables used subnational data and world variables used international data[132]. Data available at oec.world.

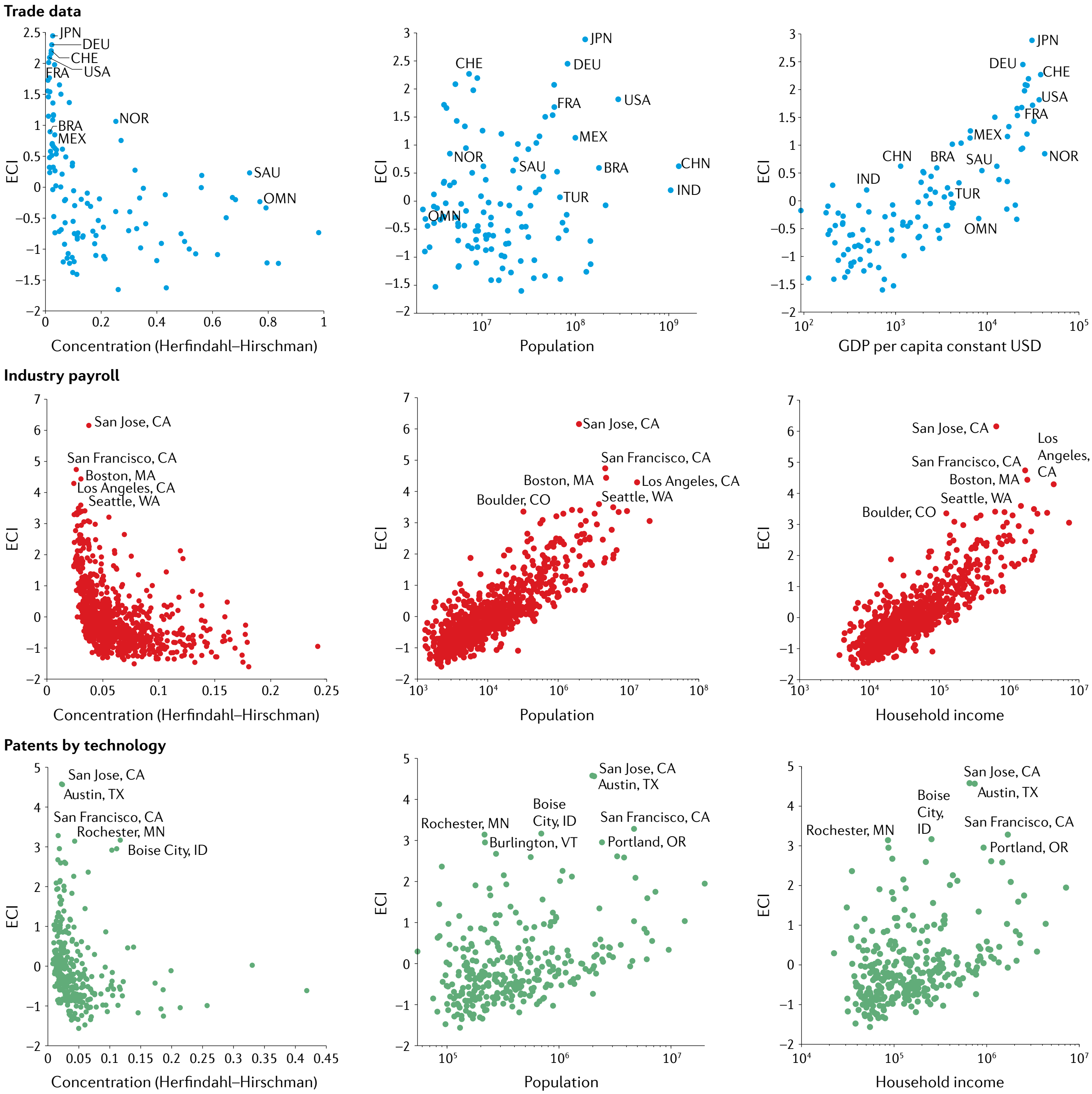


Fig. 4 | **Economic complexity compared with measures of concentration, population and income.** The economic complexity index (ECI) is computed for regions and US metropolitan statistical areas using trade data, payroll by industry, and patents by technology. GDP, gross domestic product.

Application of relatedness

The principle of relatedness

Relatedness is important because it predicts the probability that a location increases or decreases its specialization in an activity. This empirical law, known as the principle of relatedness (FIG. 5), has been shown to hold for multiple types of activities (products[7,19], industries[11,12,50,58,182], patents[13,14], occupations[16,17] and research areas[15,18,183]) and for multiple geographies (such as cities or regions). It can be formally defined by the inequality:

$$\frac{\mathrm{d}}{\mathrm{d}\omega_{cp}}\frac{\mathrm{d}R_{cp}}{\mathrm{d}t}\bigg|_{c,p} > 0 \quad (23)$$

where $|_{c,p}$ indicates that location-specific and activity-specific factors are controlled for.

For practical purposes, it is useful to put the principle of relatedness in a regression form[135,136]:

$$R_{cp}(t+\mathrm{d}t) = R_{cp}(t) + B\omega_{cp}(t) + \boldsymbol{F}_c(t) + \boldsymbol{F}_p(t) + \varepsilon \quad (24)$$

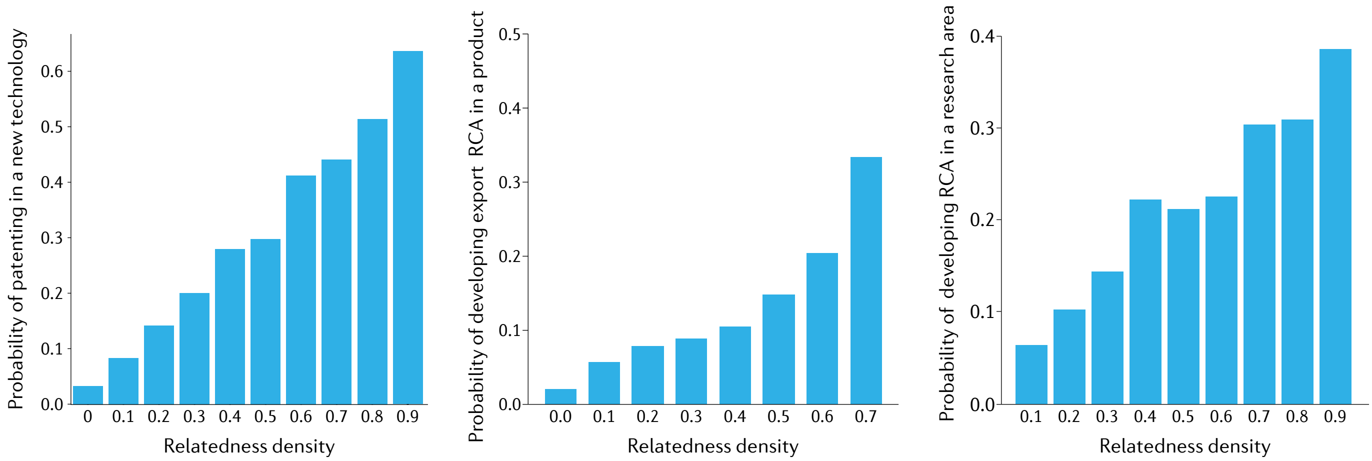


Fig. 5 | **Examples of the principle of relatedness in action.** The principle of relatedness states that the probability that a location will enter (or exit) an economy activity increases (or decreases) with the presence of related activities in that location. The examples use data from technologies, product exports and research areas[15]. RCA, revealed comparative advantage. Reprinted with permission from REF.[8], Springer Nature Limited.

where $\boldsymbol{F}_c$ and $\boldsymbol{F}_p$ are vectors of location-specific and activity-specific factors, respectively, and ε is the error or residual. The principle of relatedness is the idea that the coefficient B is positive and significant.

Unpacking relatedness

Unpacking relatedness into components. The first method to unpack relatedness constructs models that predict changes in patterns of specialization (the activities a location enters or exits) and uses them to compare relatedness metrics derived from different datasets. This method is used to study the contribution of different relatedness channels to changes in specialization[46,58,133,134] (for instance, the relative importance of industry-relatedness versus occupation-relatedness[58]). These efforts use either models of the form:

$$R_{cp}(t+\mathrm{d}t) = R_{cp} + B^1\omega_{cp}^1 + B^2\omega_{cp}^2 + \ldots + \varepsilon \quad (25)$$

where the ω_{cp}^i are different metrics of relatedness, or:

$$R_{cp}(t+\mathrm{d}t) = R_{cp} + B\hat{\omega}_{cp} + \ldots + \varepsilon \quad (26)$$

where $\hat{\omega}_{cp}$ is a 'partialized' metric of relatedness, which is detrended from its overlap with other metrics:

$$\omega_{cp}^1 = B^2\omega_{cp}^2 + B^3\omega_{cp}^3 + \ldots + \varepsilon \quad (27)$$

$$\hat{\omega}_{cp}^1 = \varepsilon \quad (28)$$

These methods have been used to unpack relatedness on a variety of settings.

Using data for the entire formal-sector economy of Brazil, REF.[58] split relatedness into three channels: industry-specific, occupation-specific and location-specific relatedness. Industry-specific and occupation-specific relatedness were based on previous work experience in related industries and occupations (measured using labour flow proximities[12]). Location-specific relatedness asked if individuals had worked in the same location. Data on the birth of firms indicated that industry-specific relatedness, followed by location-specific relatedness, were the most relevant at explaining the entry, growth and survival of new firms[58]. Surprisingly, occupation-specific relatedness did not matter for the entry of pioneer firms (firms operating in an industry that was not previously present in that location).

In a similar work, the entry and exit of occupations in US MSAs was split into three relatedness channels: complementarities (occupations used in the same activity), similarity (occupations described by similar skills) and local synergy (occupations collocating in the same cities)[46]. After controlling for multiple possible confounders, the three relatedness channels were found to be positive and significant predictors of occupational entries (increases in R_{cp} that cross the $R_{cp}=1$ threshold) and exits (decreases in R_{cp} that cross the $R_{cp}=1$ threshold). Interestingly, collocation relatedness (local synergies) was the strongest predictor of entries and exits.

The unpacking of relatedness has also been used to study temporal changes in the relative strength of different sources of collocation. At the beginning of the twentieth century, industries tended to collocate with their value-chain partners, but, today, skill requirements are more important at explaining collocation[138].

Other studies have unpacked relatedness in the context of international trade[133,134], by either using multiple measures of proximity among products[133] or by extending relatedness to bilateral trade dimensions[134,184,185].

Partialized measures of relatedness for technology, labour and supply chains have been used to reveal that relatedness-mediated entries are more likely to occur in upstream links (such as from assembly to parts) than in downstream links (from raw materials to products)[133]. This finding has important development implications, because it implies that diversification from raw materials to intermediate inputs tends to be a less frequent (probably less efficient) diversification path than diversification from final goods to intermediate inputs.

The idea of relatedness has also been extended to bilateral trade data by splitting it into three bilateral relatedness channels: relatedness among products, exporters (origins) and importers (destinations)[134]. This extends the study of relatedness to three-dimensional matrices (from R_{cp} to R_{cpd}, where d stands for destination) and unites efforts that had looked at these forms of relatedness separately (product[7], exporter[167], importer[184,185]). Product relatedness turns out to be the strongest predictor of future trade patterns, especially for technologically sophisticated products[134]. Moreover, the effect of relatedness on diversification is comparable in size with other variables, being approximately half that of sharing a common language and approximately one-third of sharing a border[134]. These are relatively large effects.

Relatedness, path-breaking diversification and institutions. The second method to unpack relatedness involves interacting relatedness with other factors to study how these factors affect the role of relatedness on changes in specialization. This method uses models of the form:

$$R_{cp}(t+\mathrm{d}t) = R_{cp} + B_1\omega_{cp} + B_2F_c\omega_{cp} + B_3F_p\omega_{cp} + \ldots + \varepsilon \quad (29)$$

where F_c and F_p are location-specific and activity-specific factors (for instance, level of education of a location) and B_2 and B_3 are coefficients for the interaction between relatedness and these factors (for instance, relatedness × level of education). Accordingly, the effect of relatedness in the presence of a factor F_c is, technically, $(B_1 + B_2F_c)\omega_{cp}$, and B_2 can be interpreted as a coefficient modifying the effect of relatedness in the presence of F_c.

This method has been used mainly to identify factors that mitigate the effects of relatedness to help identify path-breaking development.

This method has been used to explore how different varieties of capitalism (coordinated market economies versus liberal market economies) enhance or diminish the effect of relatedness on diversification[135]. Relatedness is a stronger driver of diversification for coordinated market economies, making liberal market economies relatively more successful at path-breaking forms of development.

In a similar vein, REF.[136] uses data from China to interact relatedness with foreign direct investment (FDI), imports, research and development (R&D), human capital and other variables. They find the interaction between FDI, R&D and human capital to be negative and significant, meaning that locations with higher levels of FDI, R&D and human capital are less constrained by relatedness (and engage in slightly more path-breaking development). These findings are consistent with results on international trade data[171], which indicate that higher levels of education are associated with more unrelated entries.

Likewise, pollution was interacted with relatedness to find that regions with higher levels of pollution are more likely to enter unrelated activities. This finding was validated using pollution in nearby cities as an instrumental variable[137]. The proposed mechanism is that areas that are more polluted are subject to environmental regulations that promote path-breaking development[137].

But not all factors mitigate the role of relatedness; some enhance it. Interacting density with several forms of social capital reveals that bridging social capital enhances the effects of relatedness[186], meaning that European regions rich in bridging social capital are more likely to enter related economic activities.

Although these studies show that the role of relatedness can be modulated by the presence of local factors, it is worth noting that these interaction terms are, for the most part, small in magnitude. This makes these effects technically second order to the main effect of relatedness.

Relatedness and labour

The intersection between relatedness and labour flows is another area of active research.

Workers are more likely to move among related industries and occupations[12,58,187,188]. But the impact of these flows depends on the degree of relatedness between sources and destinations. For instance, the inflow of related, but not identical, workers was shown to improve the performance of plants and firms in both Sweden[168,189] and Hungary[169]. The flow of related workers also helped speed up the recovery of regions in the Netherlands[190]. Relatedness has also been shown to impact the success of migrant inventors. Inventors that migrate have higher impact when their knowledge is related to that of the incoming region[191].

Relatedness also plays an important role in the fate of displaced workers. Data from establishment closures in Germany indicate that, in regions with a large concentration of the same industry, displaced workers find jobs faster and experience smaller earning losses[141]. Likewise, the presence of related industries in a region protects workers against long-term unemployment, although workers who move to related industries experience larger earning losses than those who find jobs in the same industry[141]. Data from Sweden reveal evidence that displaced migrants who find a job that matches their industry and occupation display the highest earnings compared with all displaced workers[192]. Similar evidence has been found of steep earning losses for workers who switch occupations[193]. All in all, the literature suggests that relatedness helps dampen the effects of labour displacement, by protecting workers against unemployment and large losses in income.

Relatedness and sustainability

Several studies have built on the idea of relatedness, and, in particular, on the product space, to study opportunities for green development. Such studies include efforts to map the potential of economies to produce a target group of green products (such as wind turbines or solar cells) based on their specialization on related products[143–146,194] or green jobs[60]. As expected, green products grow more rapidly in economies that have closely related products[146,194]. The effect of relatedness has also been shown to be stronger than that of political support[194]. The idea that relatedness promotes diversification into green activities has also been confirmed using international data on patents[147].

Applications of economic complexity

During the past decade, metrics of economic complexity have been used to formalize the impact of economic structures in outcomes such as economic growth[9,19–23,25], income inequality[10,26–29,152], greenhouse emissions[30–32], employment[195] and the spatial concentration of economic activities[95].

Economic growth

This literature began with work showing that the ECI predicted future economic growth[9]. More precisely, that an economy's future level of income (such as GDP per capita) was correlated with the ECI after controlling for its initial level of income and other factors (F_c). This finding can be described by a baseline model of the form:

$$\log(\mathrm{GDPPC}_c(t+\mathrm{d}t)) = A\ \log(\mathrm{GDPPC}_c(t)) + B\ \mathrm{ECI}_c(t) + CF_c + \ldots + \varepsilon \tag{30}$$

The fact that the ECI is a significant predictor of long-term economic growth (on the scale of decades) can be interpreted as evidence that the complexity of an economy pegs an equilibrium level of income.

This finding is robust to controlling for numerous factors[19] and has been replicated by multiple studies during the past decade using both international and subnational data[20–25,196,197].

At the international level, the connection between economic complexity and economic growth has been shown to be robust to controlling for natural resource exports, education, export concentration, institutions and competitiveness[19]. The relationship between the ECI and long-term economic growth has also been reproduced using six-digit trade data, while exploring numerous robustness checks[24]. The ECI has been reconstructed using merchandise and service exports data reproducing the positive correlation between economic complexity and growth in GDP per capita, after controlling for time fixed effects, income per capita and increase in natural resource exports during the observation period[20]. There is also evidence in favour of economic complexity predicting economic growth in a panel of 210 territories, and a positive interaction has been found between economic complexity and human capital[198]. Similarly, a positive correlation has been found between an innovation-adjusted ECI and future economic growth[126]. Finally, data from world fairs held in Paris from 1855 to 1900 were used to calculate historical ECIs for dozens of economies[23]. Despite the sparse dataset, it was shown that one standard deviation increases in the ECI contributed ~3–4% to future economic growth, when controlling for initial income and static country and year characteristics (fixed effects).

At the subnational level, the connection between economic complexity and growth has been documented using a panel of 221 Chinese cities and multiple controls (such as human capital, openness, FDI and investment rate)[22], finding that one extra standard deviation in economic complexity contributes ~0.7 percentage points of yearly per-capita economic growth. Similarly, using employment data from Mexico, it was found that one standard deviation in economic complexity is associated with an increase in the rate of economic growth of Mexican states of about 0.4% per year, while controlling for a state's initial GDP per capita, oil production and time fixed effects[25]. Likewise, data on Italian provinces were used to show that one standard deviation in economic complexity was associated with 7–10% growth in GDP per capita in a 3-year interval[199]. Using occupation data for the United States also resulted in evidence of a positive correlation between occupational complexity and future economic growth; splitting complexity between services and manufacturing reveals a more prominent role for service complexity[200]. There is also evidence that economic complexity helped the convergence of lagging Eastern European regions, but also contributes to the increasing gaps between Europe's more and less advanced regions[201]. Finally, a positive correlation between economic complexity and growth has been found using interregional trade data from Spain[197].

Income inequality

Beyond wealth, economic complexity has also been linked to variations in income inequality. Early work showed that comparable regions (in terms of income, education and other factors) exhibit less income inequality when they are more complex[10]. This finding is related to the idea of Kuznets curve[202], the notion that inequality rises and falls during development. However, instead of an inverted U-shape, economic complexity provides a linear (or quasi-linear) relationship because it separates middle-income regions with extractive economies and high levels of income inequality, such as Peru and Chile, from those with complex economies and comparatively lower levels of inequality, such as Malaysia. Following this work, REF.[152] uses a cross-country sample to show a negative correlation between economic complexity and income inequality, but also finds evidence that — within a country — inequality and complexity increase together over time. Moreover, REFS[152,203] use interaction terms to show that the ability of economic complexity to explain inequality is mediated by the presence of high levels of education and good institutions, suggesting that complexity has an equalizing effect only in the presence of good underlying labour market conditions.

The relationship between economic complexity and inequality, however, appears to reverse at the subnational level. State-level data for Brazil indicate a positive correlation between economic complexity and income inequality, with a small but significant quadratic coefficient (suggesting some curvature to the relationship, but not enough to make it an inverted U-shape)[28]. Likewise, a positive correlation between complexity in US counties and income inequality has been found (again, with a small but quadratic coefficient)[27].

Economic complexity has also been connected to reductions in gender inequalities. Using linked employer–employee data for the entire formal sector economy of Brazil, REF.[154] shows that higher complexity industries and occupations exhibit lower gender gaps in wages. Economic complexity has also been found to reduce gender inequalities in education at the tertiary level, regardless of income levels[153].

Other studies on complexity and inequality include REF.[26], which reports that complexity contributes to the reduction of income inequality in urban areas in China, but that urban–rural inequality increases in regions that have more complex export structures. Looking instead at the relationship between income inequality and the complexity of an economy's trading partners shows that trade with more complex economies is correlated with reductions in income inequality[29].

Complexity methods have also been used to estimate expected levels of income inequality for activities, and, consequently, for the locations in which these activities are present. The product Gini index has been defined as the "average level of income inequality of a product's exporters, weighted by the importance of each product in a country's export basket"[10]. Using our notation:

$$\mathrm{PGI}_p = \frac{\sum_c W_{cp}\mathrm{Gini}_c}{\sum_c W_{cp}} \tag{31}$$

where Gini_c is the Gini index, a classic measure of income inequality, and weights are given by the share of an activity in a location:

$$W_{cp} = M_{cp}\frac{X_{cp}}{\sum_{p'} X_{cp'}} \tag{32}$$

The product Gini index can be used to create a counterfactual level of income inequality for an economy, given its portfolio of activities[10].

Sustainability

The ECI has also been linked to sustainability and climate change outcomes, such as greenhouse gas emissions[30–33,59,204] and green jobs[60].

Combining data on the environmental performance index and the ECI reveals a strong and positive relationship between the ECI and environmental performance, but also a negative relationship between the ECI and air quality[32,59]. This finding has been verified using several controls and by using publication and patent data to instrument for economic complexity.

Another popular hypothesis is the environmental Kuznets curve, the idea that pollution grows and then declines with development[31,59,131,205,206]. This hypothesis has been explored using CO_2 emission data for France. After controlling for GDP per capita and its square, higher economic complexity is related to significantly lower CO_2 emissions[31]. Using a sample of 55 countries, environmental degradation has been seen to rise with economic complexity for low-income and middle-income economies, and decline with economic complexity for high-income economies[206]. Similarly, an inverse-U relationship between economic complexity and emissions has been found using a panel of 25 European countries[30]. More recently, this hypothesis was explored using a panel of 118 economies and a dynamic model, finding environmental pollution to increase and then decrease with economic complexity[205]. These findings suggest the existence of an environmental Kuznets curve[31,131], in which emissions grow and then decline with increasing level of development (as measured by economic complexity).

More recently, the effect of the ECI and the lagged ECI on greenhouse gas emission intensity (that is, emissions per unit of GDP) has been studied[33]. After controlling for numerous factors, including GDP per capita, economic openness, urbanization and an economy's share of manufacturing and agriculture, there is evidence for a negative and significant correlation between the lagged ECI and emission intensity.

Following the methods in REF.[10], product-level indices of environmental performance, CO_2 emissions and CO_2 emission intensity have been constructed[32,33]. These indexes enable associating a level of emissions to a product and, consequently, to a productive structure.

Human development and health

Economic complexity has also been connected to human development[129,130] and health indicators[158]. Reference[129] reports a positive relationship between economic complexity and human development, whereas REF.[130] reports no relationship between those variables.

A positive relationship has been found between economic complexity and health outcomes[158] in a study that used the average complexity of neighbouring countries as a mean to address endogeneity concerns.

Causality

Establishing causality with observational data is never easy. Economic complexity efforts, therefore, have been cautious about making causal claims. Nevertheless, there is evidence favouring a causal direction.

The contributions connecting economic complexity to economic growth, inequality and emissions are based on regressions that include multiple controls, fixed effects and robustness checks[19,22,24,25,198]. The ability of economic complexity to predict future economic growth has been shown to be robust to controlling for institutions[19], education[19,22], concentration of economic activities[9,19], openness[22], natural resource exports[19,25] and foreign direct investments[22], and has been reproduced using regional and international data on employment, payroll and trade. These are simple but important forms of identification that help reduce omitted-variable bias and, together, add evidence on a direction of causality. Additionally, these relationships have been tested using instrumental variables and Granger causality methods. For instance, the relationship between complexity and health outcomes has been tested using the average complexity of neighbouring economies as an instrument for complexity[158]. Complexity has been found to Granger-cause GDP per-capita growth, but not the converse[207].

However, because complexity measures are about the factors that best explain the geographic distribution of economic activities, it is important not to neglect common sense. Betting that the arrow of causality points from complexity to economic growth is consistent with the recent growth of China, Singapore and Korea, and with the economic troubles faced by Greece in 2009. Betting that the arrow of causality points from high

GDP to economic complexity is betting that countries with high per-capita GDP and low complexity, such as Qatar, Oman, Libya, Bahrain, Gabon and Kuwait, will increase their complexity in the future. The fact that the latter seems more improbable than the former captures an intuition about economic growth that is consistent with the expectations of economists going back to Adam Smith: the idea that growth is related to the division of knowledge and labour.

Models

Models of economic complexity have usually followed combinatorial approaches[208–211]. These can be described using colourful analogies, such as combining ingredients in a recipe[210], letters in words[19] or Lego blocks in models[9,210].

One such combinatorial model — introduced in 2009 (REF.[9]) and solved analytically in 2011 (REF.[208]) — reproduces the structure of specialization matrices by modelling them as a combination of two matrices. One connects locations and inputs (a) ($C_{ca} \sim [0, 1]$), and the other connects activities to the inputs they require ($P_{pa} \sim [0, 1]$). The model assumes that an activity is present in a location ($M_{cp} = 1$) if that location has all of the inputs required by that activity. That is:

$$M_{cp} = 1 \text{ if } \sum_a C_{ca} = \sum_a C_{ca} P_{pa} \tag{33}$$

Even when assuming that C and P are random binary matrices (with probabilities r and q, respectively), this simple model reproduces some stylized facts, such as the negative correlation between a location's diversity (M_c) and the average ubiquity of its activities (which is related to the matrix property known as nestedness)[9,208]. Formally, the model predicts that, on average:

$$\sum_p M_{cp} M_p = N_c N_p \left(r \left(\frac{M_c}{N_p} \right)^{1/N_a} + (1-q)(1-r) \right)^{N_a} \tag{34}$$

where N_c is the number of locations and N_p is the number of activities in the data, and N_a is the number of inputs in the model. This model also reproduces the empirically observed distribution of coagglomeration proximities[208] and explains differences in the scaling exponents found in the urban scaling laws[212] presented in REF.[213].

Combinatorial models also have interesting theoretical properties. They show increasing returns to the accumulation of inputs, but only when inputs are either highly specific (there are many of them) or hard to acquire (for instance, in a world with limited knowledge diffusion)[208]. Increasing returns imply development traps for locations with few inputs (in the letter–word analogy, they have a few letters but not enough to complete many words), meaning that these models predict economic divergence.

Combinatorial models also reveal unexpected properties. The evolution of a combinatorial model has been studied in the context of two different strategies[210]. One focused on adding inputs that readily combined with others to produce an output; the other strategy chose inputs randomly. The usefulness of inputs (the number of outputs that need them) crosses over, that is, inputs that were relatively useless in the beginning of the innovation process become useful later. This crossover is the consequence of a conservation law hidden in these combinatorial models[210]. The conservation law means that the usefulness of components multiplied by the complexity of outputs (measured as the number of components they require) is fixed throughout the innovation process.

Policy implications

How does economic complexity evolve? And what strategies should we use to promote it? In the last decade, scholars have begun exploring how different strategies, constraints and factors affect changes in the complexity of economies.

A popular approach is to consider the relatedness and complexity of potential activities together in a diagram[19,36]. Relatedness measures how 'easy' it is to enter an activity for a specific location. Complexity gives a measure of that activity's value. When applied to activities that are not present in a location, activities that are high in both relatedness and complexity represent the best 'low-hanging fruit' for diversification, and, hence, they define an efficient diversification frontier (maximizing complexity and the ease of entering into that activity).

This approach is an important way to measure whether diversification efforts tend to build on local capacities[19,36,150]. However, the evidence of whether regions actually follow relatedness is mixed. Evidence using patent data from Europe suggests that regions tend to enter activities at intermediate levels of relatedness[151]. Other efforts find little evidence of a connection between relatedness theory and practice[150].

What is more uncomfortable about this approach, however, is that it may be too specific. By trying to identify specific products, technologies and industries, we may be reading too much into techniques that are relatively new, and true on average, rather than for each data point. This concern has led to other approaches, which, instead of focusing on which activities to pick, focus on when to choose which type of activities.

Mathematical models and numerical simulations have shown that always targeting high-relatedness activities is a suboptimal way to maximize diversification[149,214]. Instead, economies should adapt their strategies by targeting relatively unrelated but connected activities during a window of opportunity that opens at intermediate levels of development. This work helps conceptualize economic diversification policies not as efforts designed to pick activities but as a portfolio allocation problem in which the mix of bets into related and unrelated activity changes with development. It is also an approach that is closer to the idea of leapfrogging — the idea that economies can skip intermediate technologies or stages of development — although leapfrogging involves successfully targeting industries with increasing returns during tight windows of opportunity[215–217].

But since the growth of complexity is related to the growth of knowledge, policy efforts also draw inspiration from works studying knowledge diffusion[82–85], while adding a few important nuances. For instance, complexity has been shown to be negatively correlated with the diffusion of knowledge, meaning that more complex knowledge struggles to diffuse longer geographical distances[44] (patent citations decay with distance more strongly for patents in complex technologies[44]). In addition, activities that are more complex tend to be more concentrated in space, suggesting that the accumulation of complex knowledge tends to require large urban agglomerations[95].

Scholars interested in the factors that affect the growth of complexity have also studied the influences of technology and policy in the growth of economic complexity.

When it comes to technology, the connection between economic complexity and internet usage has been explored using two instrumental variables: number of secure servers per million people and a civil liberty index[164]. This revealed a positive relationship between internet usage and increases in economic complexity, giving weight to arguments for policies designed to increase internet access. There is also evidence that, in China, new high-speed rail connections increased spillovers between the newly connected locations[50], accelerating knowledge diffusion.

The connection between institutional factors and complexity has also been explored. Stronger intellectual property systems have been found to engender higher economic complexity, but only in economies with an above-average level of development and complexity[160]. Studying the link between economic complexity and taxation shows that economies that rely less on capital taxation relative to labour taxation tend to export more sophisticated goods[159]. When it comes to institutions, evidence has been found that institutional quality helps improve economic complexity[161]. The relationship between FDI and economic complexity has been explored in REFS[218–220]. Economic complexity is found to increase with higher stocks of FDI, but mostly for well-educated and financially developed locations. Finally, studies on the effect of International Monetary Fund (IMF) structural adjustment programmes found no evidence of a positive effect of IMF programmes or conditionality requirements on economic complexity[165].

When it comes to cultural and demographic factors, a positive relationship between economic complexity and birthplace diversity has been found[162] using a pseudo-gravity model. Likewise, a positive relationship between LGBTQ+ inclusion and complexity was documented in REF.[163].

Overall, work on the policy implications of economic complexity has explored a variety of techniques and factors, from the use of methods to identify specific industries to theories on the optimal timing for different forms of diversification. The intersection of economic complexity and policy, however, continues to be an area of growing interest.

Outlook

In 1791, Alexander Hamilton's report on manufactures changed the economic history of the United States[221]. By advocating that the United States needed to transcend its agricultural origins and develop more sophisticated economic activities such as manufacturing, Hamilton built a policy foundation on the idea that technological leadership and economic sophistication were key for the subsequent development of the United States.

But Hamilton's intuition was diluted, together with the introduction of mathematical tools built on the idea of aggregation[1]. The use of prices and quantities as universal units of measurement allowed economists to aggregate diverse forms of capital and labour, giving rise to powerful mathematical models, at the cost of washing away the identity and idiosyncrasies of the elements involved.

Economic complexity methods replace the idea of aggregation with that of dimensionality reduction. This provides an additional set of analytical tools that is particularly useful to study differences in specialization and in macroeconomic outcomes, from inequality to growth (FIG. 6). Yet, despite these advances, economic complexity efforts exist amidst a contradiction.

On the one hand, they are built on the idea that knowledge and capabilities are highly specific, and, hence, require methods honouring the idea of organized complexity advanced by Weaver[96]. On the other hand, economic complexity research has been mostly macro, at the level of countries, cities and regions, where those specificities cannot be clearly described.

The increasing availability of even more fine-grained data is starting to change that. Beyond exploring matrices on the geography of activities, scholars can now study the specific inputs that go into the production of individual products. For instance, film credits have been used to study the increasing complexity of the movie industry[222]. This is a significant departure for work that has relied mostly on administrative records, where geographic units, products and inputs are based on standardized categorizations developed for tax and trade purposes. These categorizations do not capture nuances, such as the differences between a director of photography and a film editor. But the idea of studying the complexity of creative industries also speaks of a world in which knowledge-intense cultural activities, such as the production of films, video games and software, are becoming dominant sectors in many economies. It follows that the future of economic complexity research may not lie in the administrative records that fuelled the original contributions, but in online repositories of collaborative work, such as GitHub, LinkedIn or IMDb. This cultural turn is becoming visible in recent work focused on topics such as music[112], sports[111] or the geography of cultural exports[223,224].

Another area that is of growing interest is the connection between economic complexity and industrial policy. While this is a well-trodden territory, economic complexity methods have become attractive to policymakers, even though it is unclear whether they make a difference. Some studies have shown that innovation policies in Europe target activities of intermediate relatedness[151], in line with the optimal diversification theory[149]. Other studies, focused on Europe's Smart Specialization Strategy, find that many regions do not choose paths that are related to their current specialization[150]. The question

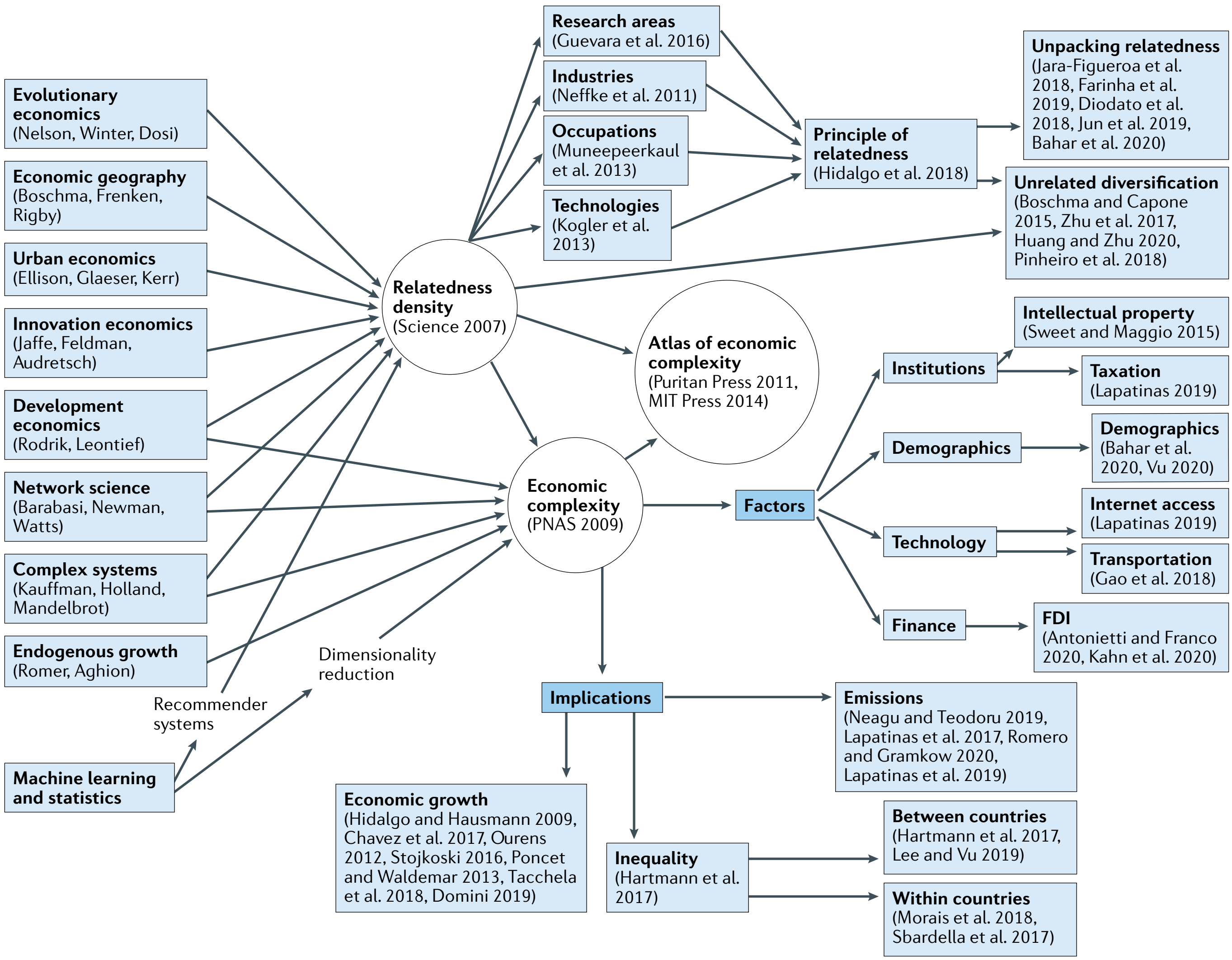


Fig. 6 | **Summary of literature on economic complexity and relatedness.** Due to space constraints, we only include a limited number of topics and papers. FDI, foreign direct investment; MIT, Massachusetts Institute of Technology; PNAS, Proceedings of the National Academy of Sciences of the United States of America.

of when complexity approaches matter for science, innovation and industrial policy, and how best to apply them, is still an open topic.

Still, existing areas of inquiry are far from depleted. For instance, although complex knowledge is known to diffuse more slowly[44], there is much to learn about the speed of knowledge diffusion and its determinants. Similarly, despite early work on the dynamics of complexity and relatedness[225], there has been relatively little work on the dynamics of relatedness and proximity[226,227]. Finally, understanding the interaction and effect of spatial scales, from countries to neighbourhoods[228], is also a relatively open field.

There are also pending issues when it comes to measuring complexity at multiple scales and using multiple data sources. The understanding of complexity metrics has improved in work exploring the linear algebra around it[176]. Yet, despite work estimating complexity using data on employment[25,45,54,58,128] and patents[44], many still confuse complexity with metrics of export diversification, leaving open the question of how to properly measure complexity at the international level by leveraging data on service trade, patents, publications and imports. These data limitations translate into measurement issues. For instance, when using trade data, the Australian economy ranks relatively low in economic complexity — because Australia's exports are dominated by coal, iron ore and petroleum gases. But Australia is also one of the top service exporters in the world. The fact that service export data are coarse and do not add much to the estimation of complexity plays against Australia, but is more a limitation of the data than the methodology. A challenge for future research is the creation of measures of complexity that reduce the dimensionality of multiple data sources simultaneously (such as products, services, patents and research areas).

There is also work to do at the subnational level (regions, cities). Here, complexity measures need data that reflect the intensity of knowledge in a location. This means that measures of employment — for instance — may be inadequate proxies for economic activity when high employment reflects a lack of capital (such as the case of labour-intense non-mechanized farming). In fact, measures built on productivity or value-added data may reflect complexity better than measures based on employment[229].

Still, the most important contribution of economic complexity may not be the discovery of new quantitative methods but the way in which it is helping reconfigure the academic landscape. Only a decade ago, network scientists, economic geographers, innovation economists and development practitioners were part of different academic communities. Today, those working on relatedness, complexity and innovation are closer than ever. Bridging this gap was not easy, but now that this reconfiguration is underway, we are starting to see knowledge being recombined, just as theories of complexity predicted.

Published online: 25 January 2021

1. Beinhocker, E. D. *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics* (Harvard Business School Press, 2006).
2. Holland, J. H. & Miller, J. H. Artificial adaptive agents in economic theory. *Am. Econ. Rev.* **81**, 365–370 (1991).
3. Miller, J. H. & Page, S. E. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton Univ. Press, 2009).
4. Kauffman, S. A. *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, 1993).
5. Krugman, P. Complex landscapes in economic geography. *Am. Econ. Rev.* **84**, 412–416 (1994).
6. Schelling, T. C. *Micromotives and Macrobehavior* (W.W. Norton & Company, 2006).
7. Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482–487 (2007).
8. Hidalgo et al. in *Unifying Themes in Complex Systems IX* (eds Morales, A. J., Gershenson, C., Braha, D., Minai, A. A. & Bar-Yam, Y.) 451–457 (Springer International Publishing, 2018).
9. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl Acad. Sci. USA* **106**, 10570–10575 (2009).
10. Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. A. Linking economic complexity, institutions, and income inequality. *World Dev.* **93**, 75–93 (2017).
11. Neffke, F., Henning, M. & Boschma, R. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Econ. Geogr.* **87**, 237–265 (2011).
12. Neffke, F. & Henning, M. Skill relatedness and firm diversification. *Strategic Manag. J.* **34**, 297–316 (2013).
13. Kogler, D. F., Rigby, D. L. & Tucker, I. Mapping knowledge space and technological relatedness in US cities. *Eur. Plan. Stud.* **21**, 1374–1391 (2013).
14. Boschma, R., Balland, P.-A. & Kogler, D. F. Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Ind. Corp. Change* **24**, 223–250 (2015).
15. Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M. & Hidalgo, C. A. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* **109**, 1695–1709 (2016).
16. Muneepeerakul, R., Lobo, J., Shutters, S. T., Gomēz-Liēvano, A. & Qubbaj, M. R. Urban economies and occupation space: can they get "there" from "here"? *PLoS ONE* **8**, e73676 (2013).
17. Alabdulkareem, A. et al. Unpacking the polarization of workplace skills. *Sci. Adv.* **4**, eaao6030 (2018).
18. Chinazzi, M., Gonçalves, B., Zhang, Q. & Vespignani, A. Mapping the physics research space: a machine learning approach. *EPJ Data Sci.* **8**, 33 (2019).
19. Hausmann, R. et al. *The Atlas of Economic Complexity: Mapping Paths to Prosperity* (MIT Press, 2014).
20. Stojkoski, V., Utkovski, Z. & Kocarev, L. The impact of services on economic complexity: service sophistication as route for economic growth. *PLoS ONE* **11**, e0161633 (2016).
21. Tacchella, A., Mazzilli, D. & Pietronero, L. A dynamical systems approach to gross domestic product forecasting. *Nat. Phys.* **14**, 861–865 (2018).
22. Poncet, S. & de Waldemar, F. S. Economic complexity and growth. *Rev. Écon.* **64**, 495–503 (2013).
23. Domini, G. *Patterns of Specialisation and Economic Complexity Through the Lens of Universal Exhibitions, 1855–1900 LEM Papers Series 2019/20* (Laboratory of Economics and Management (LEM), 2019).
24. Ourens, G. *Can the Method of Reflections Help Predict Future Growth? Documento de Trabajo/FCS-DE; 17/12* (Universidad de la Republica, 2012).
25. Chávez, J. C., Mosqueda, M. T. & Gómez-Zaldívar, M. Economic complexity and regional growth performance: evidence from the Mexican economy. *Rev. Reg. Stud.* **47**, 201–219 (2017).
26. Zhu, S., Yu, C. & He, C. Export structures, income inequality and urban-rural divide in China. *Appl. Geogr.* **115**, 102150 (2020).
27. Sbardella, A., Pugliese, E. & Pietronero, L. Economic development and wage inequality: a complex system analysis. *PLoS ONE* **12**, e0182774 (2017).
28. Bandeira Morais, M., Swart, J. & Jordaan, J. A. *Economic Complexity and Inequality: Does Productive Structure Affect Regional Wage Differentials in Brazil? USE Working Paper Series* (Utrecht University, 2018).
29. Fawaz, F. & Rahnama-Moghadamm, M. Spatial dependence of global income inequality: The role of economic complexity. *Int. Trade J.* **33**, 542–554 (2019).
30. Neagu, O. & Teodoru, M. C. The relationship between economic complexity, energy consumption structure and greenhouse gas emission: heterogeneous panel evidence from the EU countries. *Sustainability* **11**, 497 (2019).
31. Can, M. & Gozgor, G. The impact of economic complexity on carbon emissions: evidence from France. *Environ. Sci. Pollut. Res.* **24**, 16364–16370 (2017).
32. Lapatinas, A., Garas, A., Boleti, E. & Kyriakou, A. Economic complexity and environmental performance: evidence from a world sample. *MDPI Open Access J.* **12**, 1–18 (2019).
33. Romero, J. P. & Gramkow, C. Economic complexity and greenhouse gas emissions. *World Dev.* **139**, 105317 (2021).
34. Mazzucato, M. The entrepreneurial state. *Soundings* **49**, 131–142 (2011).
35. Rodrik, D. Where are we in the economics of industrial policies? *Front. Econ. China* **14**, 329–336 (2019).
36. Balland, P. A., Boschma, R., Crespo, J. & Rigby, D. L. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Reg. Stud.* **53**, 1252–1268 (2018).
37. Foray, D., David, P. A. & Hall, B. Smart specialisation– the concept. *Knowl. Econ. Policy Brief.* **9**, 100 (2009).
38. Montresor, S. & Quatraro, F. Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies. *Reg. Stud.* **54**, 1354–1365 (2019).
39. Zheng, S., Sun, W., Wu, J. & Kahn, M. E. Urban Agglomeration and Local Economic Growth in China: The Role of New Industrial Parks. *Social Science Research Network (SSRN)* https://papers.ssrn.com/abstract=2746711 (2016).
40. Kahn, M. E., Sun, W., Wu, J. & Zheng, S. The Revealed Preference of the Chinese Communist Party Leadership: Investing in Local Economic Development versus Rewarding Social Connections. *National Bureau of Economic Research (NBER)* http://www.nber.org/papers/w24457 (2018).
41. De Waldemar, F. S. & Poncet, S. Product relatedness and firm exports in China. *World Bank Econ. Rev.* **51**, 104Đ118 (2013).
42. Rigby, D. L. Technological relatedness and knowledge space: entry and exit of US cities from patent classes. *Reg. Stud.* **49**, 1922–1937 (2015).
43. Essletzbichler, J. Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Reg. Stud.* **49**, 752–766 (2015).
44. Balland, P.-A. & Rigby, D. The geography of complex knowledge. *Econ. Geogr.* **93**, 1–23 (2017).
45. Fritz, B. S. L. & Manduca, R. A. The economic complexity of US metropolitan areas. Preprint at *arXiv* https://arxiv.org/abs/1901.08112 (2019).
46. Farinha, T., Balland, P.-A., Morrison, A. & Boschma, R. What drives the geography of jobs in the US? unpacking relatedness. *Ind. Innov.* **26**, 988–1022 (2019).
47. Guo, Q. & He, C. Production space and regional industrial evolution in China. *GeoJournal* **82**, 379–396 (2017).
48. He, C. & Zhu, S. *Evolutionary Economic Geography in China* (Springer Singapore, 2019).
49. Gao, J. & Zhou, T. Quantifying China's regional economic complexity. *Phys. A Stat. Mech. Appl.* **492**, 1591–1603 (2018).
50. Gao, J., Jun, B., Pentland, A. "Sandy", Zhou, T. & Hidalgo, C. A. Collective learning in China's regional economic development. Preprint at *arXiv* https://arxiv.org/abs/1703.01369 (2017).
51. Chen, Z., Poncet, S. & Xiong, R. Inter-industry relatedness and industrial-policy efficiency: evidence from China's export processing zones. *J. Comp. Econ.* **45**, 809–826 (2017).
52. Ferrarini, B. & Scaramozzino, P. The product space revisited: China's trade profile. *World Econ.* **38**, 1368–1386 (2015).
53. Pērez Hernāndez, C. C., Salazar Hernāndez, B. C. & Mendoza Moheno, J. Diagnōstico de la complejidad econōmica del estado de Hidalgo: de las capacidades a las oportunidades. *Rev. Mex. Econ. Finanz.* **14**, 261–277 (2019).
54. Zaldīvar, F. G., Molina, E., Flores, M., Zaldīvar, M. & de Jesūs Gōmez Zaldīvar, M. Economic complexity of the special economic zones in Mexico: opportunities for diversification and industrial sophistication. *Ens. Rev. Econ.* **38**, 1–40 (2019).
55. Wang, Y. & Turkina, E. Economic complexity, product space network and Quebec's global competitiveness. *Can. J. Adm. Sci.* **37**, 334–349 (2019).
56. Lyubimov, I. L., Lysyuk, M. V. & Gvozdeva, M. A. Atlas of economic complexity, Russian regional pages. *Vopr. Econ.* https://doi.org/10.32609/0042-8736-2018-6-71-91 (2018).
57. Lyubimov, I., Gvozdeva, M., Kazakova, M. & Nesterova, K. Economic complexity of Russian regions and their potential to diversify. *J. New Econ. Assoc.* **34**, 94–122 (2017).
58. Jara-Figueroa, C., Jun, B., Glaeser, E. L. & Hidalgo, C. A. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proc. Natl Acad Sci. USA* **115**, 12646–12653 (2018).
59. Swart, J. & Brinkmann, L. in *Universities and Sustainable Communities: Meeting the Goals of the Agenda 2030* (eds Filho, L., Tortato, U. & Frankenberger, F.) 3–45 (Springer, 2020).
60. Dordmond, G., de Oliveira, H. C., Silva, I. R. & Swart, J. The complexity of green job creation: an analysis of green job development in Brazil. *Environ. Dev. Sustain.* https://doi.org/10.1007/s10668-020-00605-4 (2020).
61. Britto, G., Romero, J., Freitas, E. & Coelho, C. The great divide: economic complexity and development paths in Brazil and South Korea. *Blucher Eng. Proc.* **3**, 1404–1425 (2016).
62. Gala, P. *Complexidade Economica: Uma Nova Perspectiva Para Entender a Antiga Questao da Riqueza das Nacoes* (Contraponto, 2017).
63. Ferreira-Coimbra, N. & Vaillant, M. Evolución del espacio de productos exportados:¿ está Uruguay en el lugar equivocado? *Rev. Econ.* **16**, 97–146 (2009).
64. Reynolds, C. et al. A sub-national economic complexity analysis of Australia's states and territories. *Reg. Stud.* **52**, 715–726 (2018).
65. Erkan, B. & Yildirimci, E. Economic complexity and export competitiveness: the case of Turkey. *Procedia Soc. Behav. Sci.* **195**, 524–533 (2015).

66. Hartmann, D. in *International Innovation Networks and Knowledge Migration* (eds Pyka, A., Kuştepeli, Y. & Hartmann, D.) 53–71 (Routledge, 2016).
67. Coşkun, N., Lopcu, K. & Tuncer, İ. The economic complexity approach to development policy: where Turkey stands in comparison to OECD plus China? *Proc. Middle East Econ. Assoc.* **20**, 112–124 (2018).
68. Boschma, R., Minondo, A. & Navarro, M. The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. *Econ. Geogr.* **89**, 29–51 (2013).
69. Cicerone, G., McCann, P. & Venhorst, V. A. Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space. *J. Econ. Geogr.* **20**, 293–316 (2020).
70. Basile, R., Cicerone, G. & Iapadre, L. Economic complexity and regional labor productivity distribution: evidence from Italy. *Rev. Reg. Stud.* **47**, 201–219 (2019).
71. Innocenti, N. & Lazzeretti, L. Do the creative industries support growth and innovation in the wider economy? Industry relatedness and employment growth in Italy. *Ind. Innov.* **26**, 1152–1173 (2019).
72. Innocenti, N. & Lazzeretti, L. Related variety and employment growth in Italy. *Sci. Reg.* **16**, 325–350 (2017).
73. Innocenti, N. & Lazzeretti, L. Growth in regions, knowledge bases and relatedness: some insights from the Italian case. *Eur. Plan. Stud.* **27**, 2034–2048 (2019).
74. González, A., Ortigoza, E., Llamosas, C., Blanco, G. & Amarilla, R. Multi-criteria analysis of economic complexity transition in emerging economies: the case of Paraguay. *Socio-Econ. Plan. Sci.* **63**, 100617 (2018).
75. Mealy, P. & Coyle, D. *To Them That Hath: Economic Complexity and Local Industrial Strategy in the UK* (Elsevier, 2019).
76. Bishop, A. & Mateos-Garcia, J. Exploring the link between economic complexity and emergent economic activities. *Natl Inst. Econ. Rev.* **249**, R47–R58 (2019).
77. Maes, P. in *Readings in Human–Computer Interaction* (eds Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S.) 811–821 (Morgan Kaufmann, 1995).
78. Resnick, P. & Varian, H. R. Recommender systems. *Commun. ACM* **40**, 56–58 (1997).
79. Romer, P. M. Endogenous technological change. *J. Polit. Econ.* **98**, S71–S102 (1990).
80. Aghion, P. & Howitt, P. A model of growth through creative destruction. *Econometrica* **60**, 323–351 (1992).
81. Warsh, D. *Knowledge and the Wealth of Nations: A Story of Economic Discovery* (W.W. Norton & Company, 2007).
82. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic localization of knowledge spillovers as evidenced by patent citations. *Q. J. Econ.* **108**, 577–598 (1993).
83. Feldman, M. P. & Florida, R. The geographic sources of innovation: technological infrastructure and product innovation in the United States. *Ann. Assoc. Am. Geogr.* **84**, 210–229 (1994).
84. Audretsch, D. B. & Feldman, M. P. in *Handbook of Regional and Urban Economics* Vol. 4 Ch. 61 (eds Henderson, J. V. & Thisse, J.-F.) 2713–2739 (Elsevier, 2004).
85. Audretsch, D. B. & Feldman, M. P. R&D spillovers and the geography of innovation and production. *Am. Econ. Rev.* **86**, 630–640 (1996).
86. Boschma, R. Proximity and innovation: a critical assessment. *Reg. Stud.* **39**, 61–74 (2005).
87. Breschi, S., Lissoni, F. & Malerba, F. Knowledge-relatedness in firm technological diversification. *Res. Policy* **32**, 69–87 (2003).
88. Breschi, S. & Lissoni, F. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *J. Econ. Geogr.* **9**, 439–468 (2009).
89. Breschi, S. & Lissoni, F. in *Handbook of Quantitative Science and Technology Research* (eds Moed, H. F., Glänzel, W. & Schmoch, U.) 613–643 (Springer, 2004).
90. Singh, J. Collaborative networks as determinants of knowledge diffusion patterns. *Manag. Sci.* **51**, 756–770 (2005).
91. Collins, H. *Tacit and Explicit Knowledge* (Univ. Chicago Press, 2010).
92. Collins, H. M. The TEA set: tacit knowledge and scientific networks. *Sci. Stud.* **4**, 165–185 (1974).
93. Hidalgo, C. *Why Information Grows: The Evolution of Order, from Atoms to Economies* (Basic Books, 2015).
94. Florida, R. The world is spiky. *Atlantic Mon.* **296**, 48–51 (2005).
95. Balland, P.-A. et al. Complex economic activities concentrate in large cities. *Nat. Hum. Behav.* **4**, 248–254 (2020).
96. Weaver, W. Science and complexity. *Am. Sci.* **36**, 536–544 (1948).
97. Cohen, W. M. & Levinthal, D. A. Absorptive capacity: a new perspective on learning and innovation. *Adm. Sci. Q.* **35**, 128–152 (1990).
98. Ellison, G. & Glaeser, E. L. Geographic concentration in US manufacturing industries: a dartboard approach. *J. Polit. Econ.* **105**, 889–927 (1997).
99. Almeida, P. & Kogut, B. Localization of knowledge and the mobility of engineers in regional networks. *Manag. Sci.* **45**, 905–917 (1999).
100. Bellet, M., Colletis, G. & Lung, Y. Économie de proximités. *Rev. Econ. Reg. Urbaine* **3**, 357–608 (1993).
101. Carrincazeaux, C., Lung, Y. & Vicente, J. The scientific trajectory of the French school of proximity: interaction- and institution-based approaches to regional innovation systems. *Eur. Plan. Stud.* **16**, 617–628 (2008).
102. Rallet, A. & Torre, A. Proximité géographique ou proximité organisationnelle? Une analyse spatiale des coopérations technologiques dans les réseaux localisés d'innovation, **54**, 147–171 (HAL INRAE, 2001).
103. Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. & Shleifer, A. Growth in cities. *J. Polit. Econ.* **100**, 1126–1152 (1992).
104. Frenken, K., Oort, F. V. & Verburg, T. Related variety, unrelated variety and regional economic growth. *Reg. Stud.* **41**, 685–697 (2007).
105. Ellison, G. & Glaeser, E. L. The geographic concentration of industry: does natural advantage explain agglomeration? *Am. Econ. Rev.* **89**, 311–316 (1999).
106. Saviotti, P. P. & Frenken, K. Export variety and the economic performance of countries. *J. Evol. Econ.* **18**, 201–218 (2008).
107. Boschma, R. & Frenken, K. in *Beyond Territory. Dynamic Geographies of Knowledge Creation, Diffusion and Innovation* (eds Bathelt, H., Feldman, M. & Kogler, D.) 64–68 (Routledge, 2012).
108. Porter, M. E. *Clusters and the New Economics of Competition* (Harvard Business Review, 1998).
109. Porter, M. E. *The Competitive Advantage of Nations* (Simon and Shuster, 1990).
110. Shutters, S. T., Muneepeerakul, R. & Lobo, J. Constrained pathways to a creative urban economy. *Urban Stud.* **53**, 3439–3454 (2016).
111. Knuepling, L. & Broekel, T. Does relatedness drive the diversification of countries' success in sports? *Eur. Sport Manag. Q.* https://doi.org/10.1080/16184742.2020.1770830 (2020).
112. Klement, B. & Strambach, S. How do new music genres emerge? Diversification processes in symbolic knowledge bases. *Reg. Stud.* **53**, 1447–1458 (2019).
113. Dosi, G. in *Diffusion of Technologies and Social Behavior* (eds Nakićenović, N. & Grübler, A.) 179–208 (Springer, 1991).
114. Dosi, G., Freeman, C., Nelson, R., Silverberg, G. & Soete, L. (eds) *Technical Change and Economic Theory* (Pinter Publishers, 1988).
115. Nelson, R. R. & Winter, S. G. *An Evolutionary Theory of Economic Change* (Belknap Press of Harvard University Press, 1982).
116. Fleming, L. & Sorenson, O. Technology as a complex adaptive system: evidence from patent data. *Res. Policy* **30**, 1019–1039 (2001).
117. Archibugi, D. & Coco, A. A new indicator of technological capabilities for developed and developing countries (ArCo). *World Dev.* **32**, 629–654 (2004).
118. Lall, S. Indicators of the relative importance of IPRs in developing countries. *Res. Policy* **32**, 1657–1680 (2003).
119. Desai, M., Fukuda-Parr, S., Johansson, C. & Sagasti, F. Measuring the technology achievement of nations and the capacity to participate in the network age. *J. Hum. Dev.* **3**, 95–122 (2002).
120. Hausmann, R., Hwang, J. & Rodrik, D. What you export matters. *J. Econ. Growth* **12**, 1–25 (2007).
121. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Sci. Rep.* **2**, srep00723 (2012).
122. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: a metrics for the economic complexity of countries and products. *PLoS ONE* **8**, e70726 (2013).
123. Ivanova, I., Strand, Ø., Kushnir, D. & Leydesdorff, L. Economic and technological complexity: a model study of indicators of knowledge-based innovation systems. *Technol. Forecast. Soc. Change* **120**, 77–89 (2017).
124. Ivanova, I., Smorodinskaya, N. & Leydesdorff, L. On measuring complexity in a post-industrial economy: The ecosystem's approach. *Qual. Quantity* **54**, 197–212 (2020).
125. AlQurtas, A. M. *A New Indicator of Economic Complexity to Guide Industrial Policies* (George Washington University, 2018).
126. Xu, M. & Lybbert, T. J. *Innovation-adjusted Economic Complexity and Growth: Does Product-Specific Patenting Reveal Enhanced Economic Capabilities?* (University of California, 2019).
127. Petralia, S., Balland, P.-A. & Morrison, A. Climbing the ladder of technological development. *Res. Policy* **46**, 956–969 (2017).
128. Wohl, I. *The Method of Reflections and US Occupational Employment* (U.S. International Trade Commission, 2020).
129. Ferraz, D., Moralles, H. F., Campoli, J. S., de Oliveira, F. C. R. & Rebelatto, D. A. Economic complexity and human development: DEA performance measurement in Asia and Latin America. *Gestão Produção* **25**, 839–853 (2018).
130. Lapatinas, A. Economic complexity and human development: a note. *Econ. Bull.* **36**, 1441–1452 (2016).
131. Neagu, O. The link between economic complexity and carbon emissions in the European Union countries: a model based on the environmental Kuznets Curve (EKC) approach. *Sustainability* **11**, 4753 (2019).
132. Simoes, A. J. G. & Hidalgo, C. A. in *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence* (MIT Media Lab, 2011).
133. Bahar, D., Rosenow, S., Stein, E. & Wagner, R. Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Dev.* **117**, 48–60 (2019).
134. Jun, B., Alshamsi, A., Gao, J. & Hidalgo, C. A. Bilateral relatedness: knowledge diffusion and the evolution of bilateral trade. *J. Evol. Econ.* **30**, 247–277 (2019).
135. Boschma, R. & Capone, G. Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Res. Policy* **44**, 1902–1914 (2015).
136. Zhu, S., He, C. & Zhou, Y. How to jump further and catch up? Path-breaking in an uneven industry space. *J. Econ. Geogr.* **17**, 521–545 (2017).
137. Huang, Y. & Zhu, S. Regional industrial dynamics under the environmental pressures in China. *J. Clean. Prod.* **265**, 121917 (2020).
138. Diodato, D., Neffke, F. & O'Clery, N. Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *J. Urban Econ.* **106**, 1–26 (2018).
139. Pugliese, E. et al. Unfolding the innovation system for the development of countries: coevolution of Science, Technology and Production. *Sci. Rep.* **9**, 16440 (2019).
140. Catalán, P., Navarrete, C. & Figueroa, F. The scientific and technological cross-space: is technological diversification driven by scientific endogenous capacity? *Res. Policy* https://doi.org/10.1016/j.respol.2020.104016 (2020).
141. Neffke, F. M., Otto, A. & Hidalgo, C. The mobility of displaced workers: How the local industry mix affects job search. *J. Urban Econ.* **108**, 124–140 (2018).
142. Delgado, M., Porter, M. E. & Stern, S. Clusters, convergence, and economic performance. *Res. Policy* **43**, 1785–1799 (2014).
143. Hamwey, R., Pacini, H. & Assunção, L. Mapping green product spaces of nations. *J. Environ. Dev.* **22**, 155–168 (2013).
144. Mealy, P. & Teytelboym, A. Economic complexity and the green economy. *Res. Policy* https://doi.org/10.1016/j.respol.2020.103948 (2017).
145. Huberty, M. & Zachmann, G. *Green Exports and the Global Product Space: Prospects for EU Industrial Policy* (Bruegel, 2011).
146. Fraccascia, L., Giannoccaro, I. & Albino, V. Green product development: what does the country product space imply? *J. Clean. Prod.* **170**, 1076–1088 (2018).

147. Perruchas, F., Consoli, D. & Barbieri, N. Specialisation, diversification and the ladder of green technology development. *Res. Policy* **49**, 103922 (2020).
148. Delgado, M., Porter, M. E. & Stern, S. Clusters and entrepreneurship. *J. Econ. Geogr.* **10**, 495–518 (2010).
149. Alshamsi, A., Pinheiro, F. L. & Hidalgo, C. A. Optimal diversification strategies in the networks of related products and of related research areas. *Nat. Commun.* **9**, 1328 (2018).
150. Marrocu, E., Paci, R., Rigby, D. & Usai, S. *Smart Specialization Strategy: Any Relatedness Between Theory and Practice?* (University of Cagliari and University of Sassari, 2020).
151. Uhlbach, W.-H., Balland, P.-A. & Scherngell, T. *R&D Policy and Technological Trajectories of Regions: Evidence from the EU Framework Programmes* (Bruegel, 2017).
152. Lee, K.-K. & Vu, T. V. Economic complexity, human capital and income inequality: a cross-country analysis. *Jpn. Econ. Rev.* **71**, 695–718 (2019).
153. Ben Saâd, M. & Assoumou-Ella, G. Economic complexity and gender inequality in education: an empirical study. *Econ. Bull.* **39**, 321–334 (2019).
154. Barza, R., Jara-Figueroa, C., Hidalgo, C. & Viarengo, M. *Knowledge Intensity and Gender Wage Gaps: Evidence from Linked Employer–Employee Data* (University of Munich, 2020).
155. Le Caous, E. & Huarng, F. Economic complexity and the mediating effects of income inequality: reaching sustainable development in developing countries. *Sustainability* **12**, 2089 (2020).
156. Güneri, B. & Yalta, A. Y. Does economic complexity reduce output volatility in developing countries? *Bull. Econ. Res.* https://doi.org/10.1111/boer.12257 (2020).
157. Sweet, C. & Eterovic, D. Do patent rights matter? 40 years of innovation, complexity and productivity. *World Dev.* **115**, 78–93 (2019).
158. Vu, T. V. Economic complexity and health outcomes: a global perspective. *Soc. Sci. Med.* **265**, 113480 (2020).
159. Lapatinas, A., Kyriakou, A. & Garas, A. Taxation and economic sophistication: evidence from OECD countries. *PLoS ONE* **14**, e0213498 (2019).
160. Sweet, C. M. & Eterovic Maggio, D. S. Do stronger intellectual property rights increase innovation? *World Dev.* **66**, 665–677 (2015).
161. Vu, T. V. *Does Institutional Quality Foster Economic Complexity?* (University of Otago, 2019).
162. Bahar, D., Rapoport, H. & Turati, R. Birthplace diversity and economic complexity: Cross-country evidence. *Res. Policy* https://doi.org/10.1016/j.respol.2020.103991 (2020).
163. Vu, T. V. Does LGBT Inclusion Promote National Innovative Capacity? *Social Science Research Network (SSRN)* https://papers.ssrn.com/abstract=3523553 (2020).
164. Lapatinas, A. The effect of the Internet on economic sophistication: An empirical analysis. *Econ. Lett.* **174**, 35–38 (2019).
165. Demir, F. *IMF Conditionality, Trade Structure and Economic Complexity: What Did Structural Adjustment Programs Really Adjust?* (University of Oklahoma, 2019).
166. Bustos, S., Gomez, C., Hausmann, R. & Hidalgo, C. A. The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLoS ONE* **7**, e49393 (2012).
167. Bahar, D., Hausmann, R. & Hidalgo, C. A. Neighbors and the evolution of the comparative advantage of nations: evidence of international knowledge diffusion? *J. Int. Econ.* **92**, 111–123 (2014).
168. Boschma, R., Eriksson, R. & Lindgren, U. How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *J. Econ. Geogr.* **9**, 169–190 (2009).
169. Csáfordi, Z., Lőrincz, L., Lengyel, B. & Kiss, K. M. Productivity spillovers through labor flows: Productivity gap, multinational experience and industry relatedness. *J. Technol. Transf.* **45**, 86–121 (2018).
170. Börner, K. et al. Design and update of a classification system: The UCSD map of science. *PLoS ONE* **7**, e39464 (2012).
171. Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R. & Hidalgo, C. A. Shooting high or low: do countries benefit from entering unrelated activities? Preprint at *arXiv* https://arxiv.org/abs/1801.05352 (2018).
172. Arvis, J.-F. *How Many Dimensions do we Trade in? Product Space Geometry and Latent Comparative Advantage* (The World Bank, 2013).
173. Utkovski, Z., Pradier, M. F., Stojkoski, V., Perez-Cruz, F. & Kocarev, L. Economic complexity unfolded: interpretable model for the productive structure of economies. *PLoS ONE* **13**, e0200822 (2018).
174. Che, N. *Intelligent Export Diversification: an Export Recommendation System with Machine Learning* (International Monetary Fund, 2020).
175. Hill, M. O. Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* **61**, 237–249 (1973).
176. Mealy, P., Farmer, J. D. & Teytelboym, A. Interpreting economic complexity. *Sci. Adv.* **5**, eaau1705 (2019).
177. Caldarelli, G. et al. A network analysis of countries' export flows: firm grounds for the building blocks of the economy. *PLoS ONE* **7**, e47278 (2012).
178. Felipe, J., Kumar, U., Abdon, A. & Bacate, M. Product complexity and economic development. *Struct. Change Econ. Dyn.* **23**, 36–68 (2012).
179. Lizardo, O. The mutual specification of genres and audiences: Reflective two-mode centralities in person-to-culture data. *Poetics* **68**, 52–71 (2018).
180. Page, L., Brin, S., Motwani, R. & Winograd, T. *The Pagerank Citation Ranking: Bringing Order to the Web* (Stanford InfoLab, 1999).
181. Servedio, V. D., Buttà, P., Mazzilli, D., Tacchella, A. & Pietronero, L. A new and stable estimation method of country economic fitness and product complexity. *Entropy* **20**, 783 (2018).
182. Ellison, G., Glaeser, E. L. & Kerr, W. R. What causes industry agglomeration? evidence from coagglomeration patterns. *Am. Econ. Rev.* **100**, 1195–1213 (2010).
183. Boschma, R., Heimeriks, G. & Balland, P.-A. Scientific knowledge dynamics and relatedness in biotech cities. *Res. Policy* **43**, 107–114 (2014).
184. Morales, E., Sheu, G. & Zahler, A. *Gravity and Extended Gravity: Estimating a Structural Model of Export Entry* (University Library of Munich, 2011).
185. Chaney, T. The network structure of international trade. *Am. Econ. Rev.* **104**, 3600–3634 (2014).
186. Cortinovis, N., Xiao, J., Boschma, R. & van Oort, F. G. Quality of government and social capital as drivers of regional diversification in Europe. *J. Econ. Geogr.* **17**, 1179–1208 (2017).
187. Neffke, F. M., Otto, A. & Weyh, A. Inter-industry labor flows. *J. Econ. Behav. Organ.* **142**, 275–292 (2017).
188. Hausmann, R. & Neffke, F. M. H. The workforce of pioneer plants: the role of worker mobility in the diffusion of industries. *Res. Policy* **48**, 628–648 (2019).
189. Borggren, J., Eriksson, R. H. & Lindgren, U. Knowledge flows in high-impact firms: How does relatedness influence survival, acquisition and exit? *J. Econ. Geogr.* **16**, 637–665 (2016).
190. Diodato, D. & Weterings, A. B. The resilience of regional labour markets to economic shocks: Exploring the role of interactions among firms and workers. *J. Econ. Geogr.* **15**, 723–742 (2015).
191. Miguelez, E. & Moreno, R. Relatedness, external linkages and regional innovation in Europe. *Reg. Stud.* **52**, 688–701 (2018).
192. Kekezi, O. & Boschma, A. R. *Returns to Migration After Job Loss – The Importance of Job Match* (Utrecht University, 2020).
193. Nedelkoska, L., Neffke, F. & Wiederhold, S. in *Annual Meeting of the American Economic Association* (American Economic Association, 2015).
194. Santoalha, A. & Boschma, R. Diversifying in green technologies in European regions: does political support matter? *Reg. Stud.* https://doi.org/10.1080/00343404.2020.1744122 (2020).
195. Adam, A., Garas, A. & Lapatinas, A. *Economic Complexity and Jobs: An Empirical Analysis* (University Library of Munich, 2019).
196. Cristelli, M., Tacchella, A. & Pietronero, L. The heterogeneous dynamics of economic complexity. *PLoS ONE* **10**, e0117174 (2015).
197. Balsalobre, S. J. P., Verduras, C. L. & Lanchas, J. D. *Measuring the Economic Complexity at the Sub-National Level Using International and Interregional Trade* (Autonomous University of Madrid, 2017).
198. Zhu, S. & Li, R. Economic complexity, human capital and economic growth: empirical research based on cross-country panel data. *Appl. Econ.* **49**, 3815–3828 (2017).
199. Coniglio, N. D., Lagravinese, R. & Vurchio, D. Production sophisticatedness and growth: evidence from Italian provinces before and during the crisis, 1997–2013. *Camb. J. Reg. Econ. Soc.* **9**, 423–442 (2016).
200. Lo Turco, A. & Maggioni, D. The knowledge and skill content of production complexity. *Res. Policy* https://doi.org/10.1016/j.respol.2020.104059 (2020).
201. Buccellato, T. & Corò, G. in *Capitalism, Global Change and Sustainable Development* (ed. Paganetto, L.) 149–167 (Springer International Publishing, 2020)
202. Kuznets, S. Economic growth and income inequality. *Am. Econ. Rev.* **45**, 1–28 (1955).
203. Chu, L. K. & Hoang, D. P. How does economic complexity influence income inequality? New evidence from international data. *Econ. Anal. Policy* **68**, 44–57 (2020).
204. Dong, Z., Chen, W. & Wang, S. Emission reduction target, complexity and industrial performance. *J. Environ. Manag.* **260**, 110148 (2020).
205. Chu, L. K. Economic structure and environmental Kuznets curve hypothesis: new evidence from economic complexity. *Appl. Econ. Lett.* https://doi.org/10.1080/13504851.2020.1767280 (2020).
206. Doğan, B., Saboori, B. & Can, M. Does economic complexity matter for environmental degradation? An empirical analysis for different stages of development. *Environ. Sci. Pollut. Res.* **26**, 31900–31912 (2019).
207. Güneri, B. Economic complexity and economic performance. *Res. Policy* https://doi.org/10.1016/j.respol.2020.103948 (2020).
208. Hausmann, R. & Hidalgo, C. A. The network structure of economic output. *J. Econ. Growth* **16**, 309–342 (2011).
209. Weitzman, M. L. Recombinant growth. *Q. J. Econ.* **113**, 331–360 (1998).
210. Fink, T. M. A., Reeves, M., Palma, R. & Farr, R. S. Serendipity and strategy in rapid innovation. *Nat. Commun.* **8**, 2002 (2017).
211. Hidalgo, C. A. Economic complexity: from useless to keystone. *Nat. Phys.* **14**, 9–10 (2018).
212. Gomez-Lievano, A., Patterson-Lomba, O. & Hausmann, R. Explaining the prevalence, scaling and variance of urban phenomena. *Nat. Hum. Behav.* **1**, 0012 (2018).
213. Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306 (2007).
214. Waniek, M., Elbassioni, K., Pinheiro, F. L., Hidalgo, C. A. & Alshamsi, A. Computational aspects of optimal strategic network diffusion. *Theor. Comput. Sci.* **814**, 153–168 (2020).
215. Lee, K. & Ki, J. Rise of latecomers and catch-up cycles in the world steel industry. *Res. Policy* **46**, 365–375 (2017).
216. Lee, K. & Malerba, F. Catch-up cycles and changes in industrial leadership: Windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Res. Policy* **46**, 338–351 (2017).
217. Arthur, W. B. Positive feedbacks in the economy. *Sci. Am.* **262**, 92–99 (1990).
218. Antonietti, R. & Franco, C. From FDI to economic complexity: a panel Granger causality analysis. *Struct. Chang. Econ. Dyn.* https://doi.org/10.1016/j.strueco.2020.11.001 (2020).
219. Khan, H., Khan, U. & Khan, M. A. Causal nexus between economic complexity and FDI: empirical evidence from time series analysis. *Chinese Econ.* **53**, 374–394 (2020).
220. Ozsoy, S., Fazlioglu, B. & Esen, S. Do FDI and patents drive sophistication of exports? A panel data approach. *Prague Econ. Pap*. https://doi.org/10.18267/j.pep.755 (2020).
221. Hamilton, A. *Report on Manufactures* (Good Press, 1791).
222. Tinits, P. & Sobchuk, O. Open-ended cumulative cultural evolution of Hollywood film crews. *Evol. Hum. Sci*. https://doi.org/10.1017/ehs.2020.21 (2020).
223. Yu, A. Z., Ronen, S., Hu, K., Lu, T. & Hidalgo, C. A. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Sci. Data* **3**, 1–16 (2016).
224. Ronen, S. et al. Links that speak: the global language network and its association with global fame. *Proc. Natl Acad. Sci. USA* **111**, E5616–E5622 (2014).
225. Hidalgo, C. A. *The Dynamics of Economic Complexity and the Product Space Over a 42 Year Period* (Harvard University, 2009).
226. Kuusk, K. & Martynovich, M. Dynamic nature of relatedness, or what kind of related variety for long-term regional growth. *Tijdschr. Econ. Soc. Geogr.* https://doi.org/10.1111/tesg.12427 (2020).
227. Juhász, S., Broekel, T. & Boschma, R. Explaining the dynamics of relatedness: the role of co-location and complexity. *Pap. Reg. Sci.* https://doi.org/10.1111/pirs.12567 (2020).

228. Hidalgo, C. A., Castañer, E. & Sevtsuk, A. The amenity mix of urban neighborhoods. *Habitat Int.* **106**, 102205 (2020).
229. Koch, P. Economic complexity and growth: can value-added exports better explain the link?. *Econ. Lett.* **198**, 109682 (2021).

Acknowledgements
The author thanks Pierre-Alexandre Balland, Cristian Jara-Figueroa and Mary Kaltenberg for their useful comments. This work was supported by ANITI (ANR-19-PI3A-0004).

Competing interests
César A. Hidalgo is a founder and CEO of Datawheel, a company specialized in the creation of data distribution and visualization systems, including the Observatory of Economic Complexity, Pantheon.World, DataUSA.io and DataMéxico, among other platforms.

Peer review information
Nature Reviews Physics thanks Woo-Sung Jung, Russell Funk, Maryann Feldman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

RELATED LINKS
DataMéxico: https://datamexico.org
The Observatory of Economic Complexity: https://oec.world