# Innovation Capabilities Outlook 2026: Technical Notes


Christian Chacua[*] Shreyas Gadgin Matha[*] Matte Hartog[*]
Ricardo Hausmann[*] Federico Moscatelli[†] Julio Raffo[†]
Muhammed A. Yildirim[*‡]


January 16, 2026

**Disclaimer:** The views expressed in this article are those of the authors and do not necessarily reflect the views of the World Intellectual Property Organization (WIPO) or its Member States.


**Acknowledgments:** We are indebted to the members of the Growth Lab for their comments and suggestions. Any mistake remains ours.


**Author Note:** Some sections of this article have been adapted or reprinted from the following sources: Hausmann et al. (2024b), Hausmann et al. (2024a), and Moscatelli et al. (2024). Adaptations have been made with permission, and original sources are cited where relevant.


[*]The Growth Lab, Center for International Development - Harvard University, Cambridge, MA, United States.
[†]World Intellectual Property Organization, Geneva, Switzerland.
[‡]Department of Economics, College of Administrative Sciences and Economics - Koç University, Istanbul, Turkey.

# Contents

# 1 Introduction

What causes these significant gaps in ideas, income, and productivity? And what could be done to close this gap? Economic complexity methodology elaborates on these questions. When used in the innovation context, it gives us tools on how policies can address current development challenges while considering the natural limits to knowledge creation and diffusion, as well as the complex nature of productive activities. Economic complexity approach posits that effective innovation policies should be place-based and multidimensional, leveraging countries' existing capabilities and addressing countries' current problems. This contrasts policies that lead to economic efficiencies, such as copying other countries' solutions to problems that countries do not currently have.

The economic complexity framework can shed some light on how to address these different motivations by conceiving industrial, science, and innovation policies as part of a process of knowledge diversification. Under the economic complexity framework, the final goal of policy interventions is to maximize economic development by diversifying the set of capabilities (letters) so that economies can create more complex products (longer words). It acknowledges the role of the existing capabilities available in a territory, as knowledge diversification is embedded in a path-dependence process, in which the current knowledge can limit the possibilities to develop new technologies (Hidalgo et al., 2018). However, the role of policy interventions is to help economies identify and develop the capabilities that generate the most binding constraints to growth (Hausmann et al., 2008), even if they are not part of the adjacent set of opportunities or differ from the natural path dependence process. In this way, economies can modify their fate and escape from the traps imposed by their existing set of capabilities (Balland et al., 2019).

Initial applications of the economic complexity framework was built using international trade data (Hausmann & Klinger, 2006; Hidalgo et al., 2007; Hidalgo & Hausmann, 2009; Hausmann et al., 2014). However, innovation capabilities can reveal themselves in other domains such as scientific publications and patents. Scientific publications offer insights into the ideas that mainly originated from scientific research. It captures the nascent ideas from academia and basic research institutions that might underpin future productive innovations. Although not all scientific knowledge leads to productive innovations, as it is not its primary goal, a significant share of productive innovations originated from basic science (Mazzucato, 2015; Gruber & Johnson, 2019).

Patents are an indicator of invention, one of the intermediate steps of the innovation process. It assesses the potential transformation of ideas into market products. An economy that consistently generates patents in a sector likely has productive know-how and capabilities in that sector. Even though we recognize the limitations of using patent data, it remains a valuable source for gauging innovative trajectories.[1]

This methodology paper introduces and applies economic complexity metrics to describe global trends of innovation complexity. It extends concepts from the economic complexity framework on trade to the analysis of scientific and technological progress, measured through scientific publications and patents. Following Hausmann et al. (2014), Hausmann et al. (2024a) and Moscatelli et al. (2024), we discuss measures of capabilities used in these three domains. The definitions here are based on country-level data on scientific publications, patents and international trade but they can easily be extended to the subnational settings.

Here, based on the complexity variables, we develop three types of indices to capture a locations resilience, complexity and potential. These measures can be used by policymakers to assess the performance of their locations in the light of economic complexity framework.

The rest of this paper is organized as follows. In Section 2, we provide some motivation behind the usage of different economic complexity metrics, especially in the policy context. Section 3 reviews the literature. Section 4 describes our data sources. Section 5 gives definitions and formalizes the most used economic complexity variables and Section 6 concludes.

[1] As the use of patents varies by several factors (e.g., industry, firm size, type of innovation, among others) (Mezzanotti & Simcoe, 2023; Cohen et al., 2000; Levin et al., 1987; Harabi, 1995), the statistical inference based on patent data is limited. In addition, patenting practices can differ across countries and may respond to strategic behaviors (Lemley & Shapiro, 2007; Golden, 2007; Henkel, 2022).

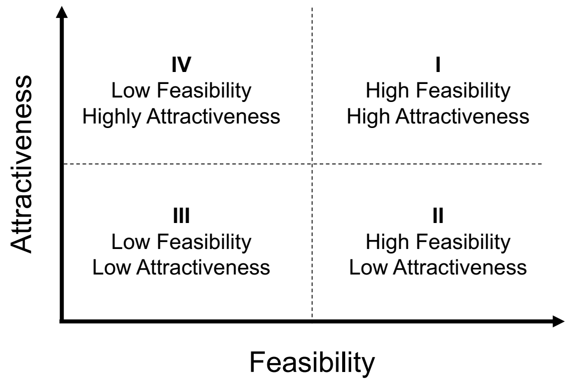
## 2 Motivation

The economic complexity framework requires thinking about transversal policy objectives and designing the learning mechanisms that solve the most binding constraints to growth. Under this approach, coordination of science, innovation, and industrial policy is required to develop the letters (capabilities) that would allow the creation of longer words (more complex products). Considering the rules that govern know-how, which require concentration of activities, and the financial limitations, which determine the bandwidth of simultaneous interventions, countries should design tailored solutions to address the most pressing binding constraints by groups of productive activities. This requires the coordination of actors and institutions from multiple domains but who are part of a limited space and share some common general interests. Moreover, it involves the creation of mechanisms that can identify the most binding constraints, correct them, and learn from the difficulties found in previous interventions. Isolated interventions or copying solutions to problems they do not have is less effective than developing the mechanisms to produce local know-how to correct their own binding constraints.

Technology represents the knowledge that we harness to reshape our physical and social environments. It has grown tremendously over the past centuries, as illustrated by the increasing volumes of books, scientific papers, and patents. Yet our individual capacity to comprehend it remains static. Hence, we increasingly specialize as individuals and distribute knowledge across counterparts. Over time, such knowledge ends up in tools, machines, equipment, and so on (embodied knowledge). At the same time, we codify what we know and convert it into forms that can be shared through documentation, standardization, and classification (codified knowledge). Yet a large part of our knowledge is what Michael Polanyi called 'tacit,' which is much harder to codify.

Figure 1 illustrates a framework where the economic complexity approach could be applied in the design of industrial or innovation policy. Two axes in the figure show the attractiveness and feasibility of innovation domains such as industries, technology classes, or scientific fields in a given location. Feasibility captures the capability or know-how overlap

Figure 1: Attractiveness and Feasibility




Source: Reprinted from *Innovation Policies Under Economic Complexity* (p. 25), by R. Hausmann et al., 2024, Growth Lab Working Paper Series (No. 234). Reprinted with permission.

between the location and the innovation field of interest and is calculated by the share of proximities that are present in the location for the field. This measure is often captured by the relatedness density (Hidalgo et al., 2018). The second axis, attractiveness, could stem from many attributes, such as the complexity of the field or the growth of the field. In the smart specialization approaches, for instance, complexity metrics are often used (Balland et al., 2019). But another attractiveness feature that could be utilized is the complexity outlook gain (COG) measure (Hausmann et al., 2014), which captures how much an innovation field brings other high-complexity entities closer to the location's capability base. We can think of attractiveness measures like PCI and growth as a one-step ahead feature, whereas COG addresses a two-steps ahead dimension because it is about how much closer all other fields become closer.

An industry, a technology class, or a scientific field that a location is not active in would fall into one of the four quadrants in Figure 1. If an entity falls into the first quadrant, it is both highly attractive and feasible. Hence, we expect these industries to appear in the location without in need of much intervention. The second quadrant consists of innovation fields that are feasible but not highly attractive. Generally, locations do not employ industrial or innovation policy to address the fields in this quadrant. The third quadrant has both low attractiveness and low feasibility. This quadrant is not also a part of desirable sets of the innovation policies. The fourth quadrant, on the other hand, harbors the innovation fields that are highly attractive but not feasible. Especially for many emerging market economies, many attractive innovation fields fall into this quadrant. In combination with the absence of many opportunities absent in the first quadrant, the locations need to utilize the break the vicious cycle.

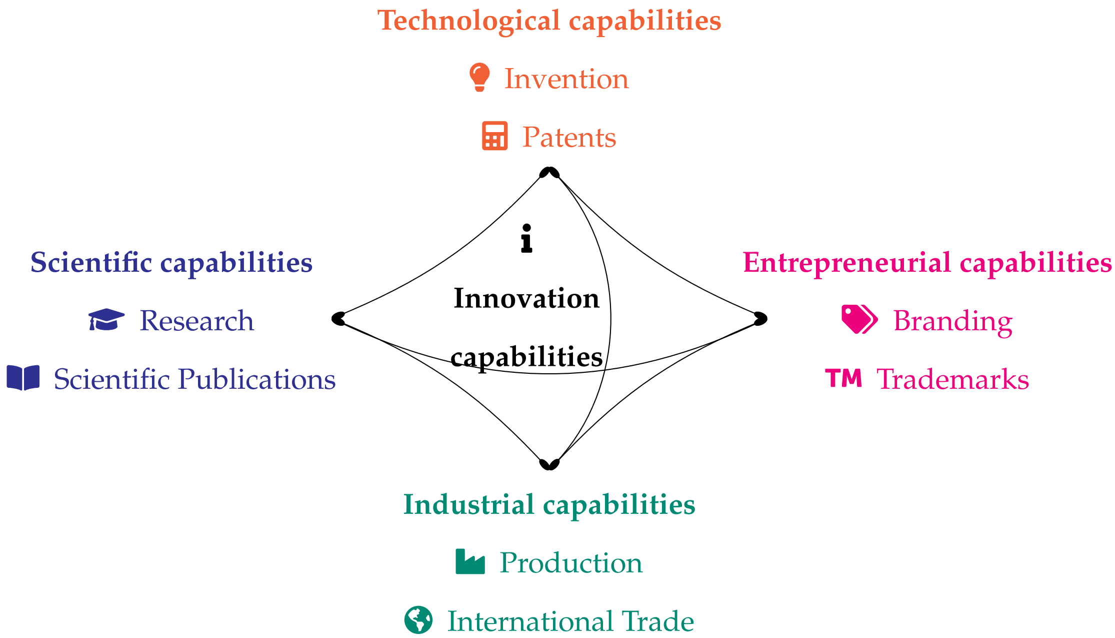
In addition to its prognostic usage, the framework captured in Figure 1 could also be used for diagnostic purposes. In particular, by analyzing the attractiveness-feasibility map of a location in previous years, a policymaker could identify innovation fields that were both feasible and attractive in the previous years but not yet appeared in the location. With this tool in hand, the policymaker could identify market failures that impeded the diversification process. For example, for an innovation field that was in the first quadrant but had not appeared yet, the policymaker could gather information from other innovation fields that share capabilities with the field of interest to uncover where the market failures are and address them through industrial or innovation policy.

We extend the traditional economic complexity measures to multi-dimensions by integrating various innovation domains. Our multidimensional approach builds on three primary sources of information that offer a comprehensive view of the innovation process: scientific publications, patents, and international trade (Figure 2). Scientific publications capture the creation of knowledge that could translate into scientific capabilities. Patents unveil the inventions that could translate into technological capabilities. International trade data reveal the current industrial capabilities of an economy. Together, these three measures provide a more extensive and complementary view of the complex nature of the innovation process, based on the capabilities present in a location.

## 3 Literature Review

Economic complexity literature has grown exponentially in recent years. Here, we will focus on the literature directly related to the concepts that we introduce in this methodology paper. For a more comprehensive review of the field, we would like to refer the readers to Balland et al. (2022) and Hidalgo (2021).

Figure 2: A multidimensional view of innovation capabilities




Source: Adapted from *Global Trends in Innovation Patterns: A Complexity Approach* (p. 5), by R. Hausmann et al., 2024, Growth Lab Working Paper Series (No. 235). Changed with permission. Our updated multidimensional view of innovation capabilities considers four types of capabilities: scientific, technological, entrepreneurial, and industrial capabilities. Hence, we analyze data on four types of activities: research, as measured in scientific publications; invention, as measured in patents; branding as measured in trademarks; and industrial production, as measured in trade data.

Hausmann & Klinger (2006) is the first paper that introduces the concept of proximity to capture relatedness between the products using the export data to define the "product space." This paper also develops the idea of "density" as a measure of distance between capability requirements of a product and capability endowments of a country. Hidalgo et al. (2007) visualized the proximity as a network to map the product space. Hidalgo & Hausmann (2009) describes and iterative process that aims to capture the extent of capability requirements of products and capability endowments and locations that give rise to the product complexity index (PCI) and economic complexity index (ECI) for locations. The iterative process is transformed to an eigenvector problem in Hausmann et al. (2014) complemented with many formal tests of growth and diversification.

The idea of relatedness has since been tested in many different settings and is now called the **principle of relatedness** (Hidalgo et al., 2018; Hausmann et al., 2022; Li & Neffke, 2023): the process of diversification tends to favor activities that are more closely related to each other. Neffke & Henning (2013), for instance, applied this logic to job transitions where every tree is an occupation and a monkey is a person, and the question is how do people move between jobs? People tend to move to jobs that have similar skill requirements.

The metric of the economic complexity index (ECI) was first introduced by Hidalgo & Hausmann (2009), where they applied it using international trade data. Since then, it has been applied to several other domains, including industries, occupations, patents, and scientific publications. Economic complexity calculations have previously been extended to patents (e.g., Balland & Rigby, 2017) and to scientific publications (e.g., Balland & Boschma, 2022), directly applying the economic complexity metric calculations introduced by Hausmann et al. (2014). We compute similar metrics, but we adjust our definition of activity presence to address the pronounced concentration of patents and scientific publications in a few countries. The standard approach is to use Bela Balassa's (Balassa, 1965) concept of revealed comparative advantage (RCA), which he applied to international trade, and say an activity is present when its RCA is greater than 1. When applied to patents or scientific publications this method gives low RCA values to the places that are its main producers. Modifying the standard approach, we consider an activity to be present in a country not only if it has RCA>1, but also if it is one of the $n$ largest producers in that scientific field or technology class, where $n$ is the effective number of producers (i.e. the inverse of the Hirschman-Herfindahl index).

Following Balland & Rigby (2017), several studies have extended economic complexity methods to develop related metrics under various names, for our considered innovation domains (Broekel, 2019; Balland et al., 2019; Sbardella et al., 2022; Stojkoski et al., 2023). For example, Balland & Rigby (2017) use the term "Knowledge Complexity Index" to refer to their complexity metric derived from patent data. In this paper, we denote all these metrics as Economic Complexity Indexes (ECIs), while mentioning their particular innovation dimension (scientific research, invention, and production). Stojkoski et al. (2023) refer to their metric as ECI, apply the ECI algorithm directly to international trade, patents, and scientific publications claiming to employ a multidimensional approach, and use these ECI measures together to explain growth, inequality, and emissions. Inspired by ECI, Tacchella et al. (2012) developed another set of metrics, namely country fitness and product quality, using international trade data by rewarding diversity and heavily penalizing products produced in less diverse locations. Although the methodology is not guaranteed to converge and has biases against small but productive countries, it has been adopted to measure the fitness of locations and quality of technology classes using patent data (See, for example, Straccamore et al., 2023).

The literature introduced several measures to capture capabilities and their relations (Hausmann et al., 2024a; Moscatelli et al., 2024). First, a country's capabilities can be revealed by the concept of Revealed Comparative Advantage (RCA) (Balassa, 1965). The RCA provides information to identify broad trends in the production and geographical distribution of scientific, technological, and industrial outcomes. Second, metrics that capture an economy's degree of sophistication and knowledge accumulation rely on generalized versions of the Economic Complexity Index (Hausmann et al., 2014; Hidalgo & Hausmann, 2009) and are used to examine the relationship between complexity and economic growth. Third, measures that exhibit the temporal evolution of capabilities and the process of scientific and technological diversification. Those measures capture the overlap between the capabilities (i.e., relatedness), generalizing the product space methodology (Hidalgo et al., 2007) developed for products in international trade to scientific fields and technological classes (Balland & Boschma, 2022; Petralia et al., 2017). Finally, the interplay among scientific, technological, and industrial capabilities are explored (Pugliese et al., 2019; Balland & Boschma, 2022; Catalán et al., 2022; Moscatelli et al., 2024; Hausmann et al., 2024a). They are used to explore the potential for countries to achieve complex technologies based on their scientific and industrial capabilities.

Formal tests of related diversification relies on "density regressions," in which density measures the extent to which an activity is surrounded by related activities in a country. The expectation is that the more related activities there are for a particular activity, the more likely that activity is to appear because of similar capabilities. At the same time, unrelated activities are more likely to disappear. These patterns would align with similar density regressions of other studies (Hidalgo et al., 2007; Neffke et al., 2011; Hausmann et al., 2022; Balland et al., 2019; Balland & Boschma, 2022; Li & Neffke, 2023).

Countries are thus specialized in very different areas when it comes to trade, patents, and scientific publications. However, do these areas relate to each other? How do scientific capabilities, for instance, translate into economic capabilities? Moreover, how does scientific output relate to patenting? They may not be directly correlated and may not co-evolve together. Furman et al. (2002) show that patenting activity across countries correlates with scientific publications, but not every publication necessarily leads to patenting. Several previous studies have explored the connections across different domains. For example, using a multi-layered network approach, Pugliese et al. (2019) combine information embedded in international trade, technology classes, and scientific fields based on co-location of these activities in countries with time delays to uncover which activities precede others. Balland & Boschma (2022) and Catalán et al. (2022) build density measures based on relatedness metrics across these domains at the country and regional co-location patterns, respectively, to uncover the patterns of related diversification across domains. Our companion paper, Moscatelli et al. (2024), builds on this literature. Our approach here is complementary to these previous studies, with relatedness measures built on citations between technology classes and scientific publications, or products and patents being produced by the same firms within and across domains.

The development of capabilities of countries is correlated across economic activities, science, and technologies. Existing activities in each area are predictive of future activity in other areas. Our results align with similar findings of technological diversification of countries by Petralia et al. (2017), as well as the relationship at the regional level between science and technology by Balland & Boschma (2022) and Shin et al. (2023). They also reinforce findings of case studies analyzing the co-evolution of industrial and academic domains in regions by Lehmann & Menter (2016) and Kenney & Mowery (2020).

Coniglio et al. (2021) and Pinheiro et al. (2022) highlight the importance of unrelated diversification for faster growth. Especially for countries with a limited number of capabilities, a sustained unrelated diversification to a new innovation field could trigger further diversification opportunities. Nevertheless, the related diversification measure's highly predictive power of disappearances in highlighted in Hausmann et al. (2022) and Hausmann et al. (2024a) show that many unrelated diversification events often become futile. Hence, there is a greater risk of failure associated with unrelated diversification events, and this risk must be carefully assessed by entrepreneurs and policymakers.

As highlighted in (Hausmann et al., 2024b), fostering knowledge infusion from elsewhere is essential to foster diversification. Neffke et al. (2018) found that regional structural change comes not so much from incumbent firms but from entrepreneurs and expanding existing firms, particularly when they come from elsewhere. Similarly, more recently, Miguelez & Morrison (2023) find that immigrant inventors foster technological diversification by developing new technological specializations, which, in turn, are transferred from the home country to the host region. Increasing evidence highlights the importance of the movement of brains - for instance, through migration (Morrison, 2023) - for the structural transformation of economies.

# 4 Data Sources

We rely on data from OpenAlex (Priem et al., 2022) in all our computations related to scientific publications. Although most bibliometric databases, including OpenAlex, claim to have global coverage of scientific publications, we are aware that they do not contain representative data for all countries. Hence, the global trends presented here may reflect biases in coverage, which may disproportionally affect countries in the global south.[2]

Our patent data has been compiled by WIPO, combining data from multiple sources, primarily EPO's Patstat 2023 and WIPO's PatentScope. Trade datasets, from which we extract export data, serve as a benchmark and primarily show the current industrial capabilities of an economy. This information reveals what is feasible for a nation to produce and where it stands in the global economic landscape. Our trade data comes mainly from international data collected by UN COMTRADE from customs offices and further cleaned by Bustos et al.

[2] Most bibliometric databases, including OpenAlex, still have poor coverage of non-English documents, local journals, and journals that are only available in print (Ansorge, 2023). As those characteristics are not randomly distributed across countries (e.g., non-English speaking countries may disproportionally write non-English articles), scientific publication data may not be fully representative of the geographic distribution of scientific knowledge.

(2024). This cleaning procedure tries to account for differences in data reported by exporters and importers, as well as the quality of data reporting by various countries.

In the three considered datasets, we analyze data at the country level for the period 2000-2020. We focus on countries to describe global trends, but we acknowledge that the design of particular innovation policies requires analysis at more disaggregated levels. Moreover, our study period is not large enough to understand the dynamics of all the stages of innovation processes, which in some cases can span multiple decades and require a more detailed assessment of how individual ideas are transformed into final products. However, it allows us to assess the current state of scientific, technological, and industrial capabilities, as well as provide insights into their geographical distribution, degree of sophistication, recent evolution, and potential connections.

As we focus on measuring the capabilities of each country, we assign scientific publications, patents, and exports to the places where they are produced. For scientific publications, we assign papers to countries based on the location of their authors' institutional affiliations. For patents, we rely on the location of the listed inventors in a patent family. For exports, we use the exporter's location from a cleaned version of UN COMTRADE data (Bustos et al., 2024). When measuring the number of scientific publications and patents in a country, we compute fractional counts based on the number of distinct countries, not on the number of different authors or inventors. Furthermore, to make patent data internationally comparable, when we refer to patent counts, we count international patent families (Miguelez et al., 2019), and not individual patent applications. In addition, we apply a series of filters to remove countries for which the data on a particular dimension is not meaningful for statistical analysis.

### 4.1 Patent data

We rely on patent data compiled by WIPO, which builds primarily on EPO's Patstat 2023a and WIPO's PatentScope datasets. Although our patent data has worldwide coverage of patent offices and inventors' locations, the statistical inference based on raw patent data counts is limited for most developing countries. As shown in the accompanying paper (Hausmann et al., 2024b), patent production is highly concentrated in a few countries and limited (or even non-existent) in most of the developing world. These differences in patenting do not necessarily imply that countries with zero patents have no technological capabilities, as patent usage responds to differences in, for instance, industries, firm size, and strategic behaviors (Mezzanotti & Simcoe, 2023). Moreover, there are some concerns about the comparability of patents filed in different patent offices, which limits international comparisons. For these reasons, our analyses:

- Only count international patent families, as defined by Miguelez et al. (2019). That is, "inventions for which the applicant has sought patent protection beyond its home patent office. This definition also includes patent applications by applicants filing only abroad, filing only through the PCT system, or filing only at the EPO" (Miguelez et al., 2019, p. 4).
- Only include countries with over 100 international patent families after the year 2000 (as indicated by their year of earliest filing).
- Only consider countries with a population over 1 million.
- Only include technologies (IPC4 subclasses) that appear in at least ten countries.

### 4.2 Scientific publication data

Scientific publication data comes from OpenAlex (Priem et al., 2022). We rely on a snapshot of OpenAlex from January 2023. We chose OpenAlex for measures of scientific publications and citation counts due to its open availability, comparability with MAG, and relatively high global coverage (Jiao et al., 2023). OpenAlex was launched in January 2022, following the closure of Microsoft Academic Graph (MAG), which had been an important source of openly available data for bibliometric analyses (Wang et al., 2020). Several studies have validated the usefulness of MAG, and compared it to other existing data sources such as Google Scholar, Scopus, and Web of Science (Wang et al., 2019; Harzing & Alakangas, 2017; Thelwall, 2017; Martín-Martín et al., 2021). Some evidence indicates a general consensus between these databases regarding citation counts, but the criteria for inclusion of papers and classification of document types seem to be inconsistent (Scheidsteger et al., 2023; Jiao et al., 2023). While MAG has lower publication and citation coverage compared to Google Scholar across most disciplines, it has higher coverage as compared to Scopus and Web of Science (Martín-Martín et al., 2021; Harzing & Alakangas, 2017; Thelwall, 2017; Huang et al., 2020). OpenAlex was intended to be MAG's immediate successor and has improved upon MAG's coverage, even for overlapping years of coverage (Scheidsteger & Haunschild, 2022).

### 4.3 Clustering

The OpenAlex taxonomy contains 4,516 research topics distributed across 26 scientific fields. To facilitate analysis at a higher level of aggregation, we reduced this to approximately 700 topic clusters while preserving the relative distribution of topics across fields.

**Determining the Number of Clusters.** The target number of clusters per field was calculated proportionally to maintain the original topic distribution.

**Semantic Embedding Generation.** Topic summaries from the OpenAlex taxonomy were converted into numerical vector representations using a large language model.[3] These embeddings position semantically similar topics close together in vector space, enabling quantitative comparison of topic content.

**Clustering.** Topics were grouped within each scientific field using the K-Means algorithm, a standard unsupervised machine learning method that partitions observations into clusters by minimizing within-cluster variance.[4]

**Cluster Labeling.** Each cluster was assigned a descriptive label generated by a large language model.[5] The model analyzed sample topic summaries from each cluster and identified

[3]We used OpenAI's `text-embedding-3-large` model, which produces 3,072-dimensional vectors capturing semantic similarity between texts.

[4]Embedding vectors were L2-normalized prior to clustering to ensure Euclidean distance corresponds to cosine similarity. K-Means was configured with 10 random initializations and 300 maximum iterations. Clustering quality was evaluated using silhouette scores.

[5]GPT-4 was prompted with up to 20 sample topic summaries per cluster and asked to produce a concise label (maximum 8 words) capturing the common theme.

the overarching research theme they share.

### 4.4 Filtering

Due to the automated nature in which OpenAlex collects information on scientific publications, it does not incorporate the data quality filtering mechanisms used in other databases, such as Scopus. However, it offers a potential advantage in providing broader global coverage. To address potential data quality issues, we implement the following procedure for filtering OpenAlex data:

- Only consider countries with over 5000 publications after the year 2000.
- Only include countries with a population over 1 million.
- Only include scientific fields for a given country in RCA calculations if there are over five publications in that field from that country.
- Only include scientific fields with over ten total publications in 2020.

Additionally, to calculate ECI values, we add a condition that a paper must have received at least five citations to be included in country counts.

It is important to note that we do not apply these filters when analyzing distributions and inequality patterns, as certain countries may have limited publication output, and excluding them could skew our analyses.

### 4.5 Trade data

We use data on international trade from Bustos et al. (2024), who clean trade data from UN COMTRADE. This cleaning approach improves data quality by accounting for mismatches in reporting by exporters and importers. Furthermore, we only consider countries listed in the Atlas of Economic Complexity and products included in the product space (Hausmann et al., 2014; Harvard's Growth Lab, 2023).

### 4.6 Trademark data

We use international trademark filings from the WIPO Global Brand Database, covering granted applications across multiple jurisdictions. To categorize trademark descriptions into meaningful thematic groups, we employed a multi-stage clustering approach that combines text embeddings with unsupervised machine learning. Each trademark description was first converted into a numerical vector re presentation using a neural language model.[6] These high-dimensional embeddings capture semantic similarities between descriptions, allowing trademarks with related meanings to be positioned closer together in the embedding space.

The clustering was performed separately within each NICE class to preserve the structure of the international trademark classification system while identifying finer-grained thematic distinctions within each class.[7] For each NICE class, we applied K-Means clustering, yielding approximately 780 distinct clusters across all 45 NICE classes, with broader classes such as Class 42 (scientific and technological services) naturally producing more sub-clusters than narrower classes.

To generate interpretable labels for each cluster, we employed a large language model to analyze representative trademark descriptions from each group and produce concise thematic labels.[8] This resulted in cluster labels such as "Natural and Synthetic Filling Fibers", "Insect and Pest Control Products" and "Electronic Control and Monitoring Devices", enabling intuitive interpretation of the clustering results.

### 4.7 Other data sources

We use the World Bank's World Development Indicators (WDI) database to source data on population, and GDP.

[6]We used OpenAI's text-embedding-ada-002 model, which produces 1,536-dimensional vectors. Embeddings were L2-normalized prior to clustering.

[7]The NICE Classification is the international system for classifying goods and services for trademark registration, comprising 45 classes (34 for goods, 11 for services).

[8]We used GPT-4 to generate labels of up to six words based on samples of 20–30 trademark descriptions per cluster.

# 5 Complexity Variables

Here, we introduce various concepts frequently used in the study of economic complexity. We present our definitions and interpretations within the context of international trade, while also highlighting their applicability to technology and scientific outputs.

Besides international trade and products, we apply these concepts to scientific and technological ideas. We create scientific field spaces based on data from scientific publications and technology spaces based on patenting data. We explore various definitions of relatedness. Further, we quantify the sophistication involved in innovating in various scientific fields and technologies and chart paths of least resistance for countries to move towards more complex scientific fields and technologies.

All our observations are at the country level, denoted by $c$. We start with three types of country-level data, namely trade, invention, and scientific outputs. For the trade data, we denote the value of trade by country $c$ in industry (product) $i$ with $X_{ci}$. For the invention data, $P_{ct}$ represents the number of patents in technology class $t$ by country $c$. For the scientific output, $S_{cf}$ is the number of publications by country $c$ in scientific field $f$. Finally, for trademarks, we will use $E_{cn}$ is the number of trademarks by country $c$ in trademark field $n$. Whenever needed, to illustrate the time dimension, we will add $y$ as an index to show the year.

Below, we will use the international trade data to describe the relevant variables, and these definitions can easily be extended to other spaces. For our derived variables, such as the Economic complexity, we will use superscripts of $X$, $P$, $S$ or $E$ to reflect the source of the data.

## 5.1 Who makes what

### 5.1.1 Smoothing the data

In our calculations, to minimize the year to year variations, we use a rolling window of 5 years.

### 5.1.2 Comparative Advantage

Originally, many complexity variables relied on the binary presence/absence of industries in locations. These binary variables are generated via comparing the country's production level to a benchmark, which gives comparative advantage measures. RCA calculations are all done within each dimension (i.e., we calculate RCA values separately for Science, Technology, Entrepreneurship and Production (trade) dimensions).

**Revealed Comparative Advantage (RCA)** (Balassa, 1965) is a metric that determines a country's relative advantage or disadvantage in producing a specific product compared to other nations. Essentially, if a country has an RCA greater than one for a product, that country has a comparative advantage in producing that product.

The comparative advantage measures, including RCA, have the following common structure:

$$R_{ci} = X_{ci}/\hat{X}_{ci},$$

where $\hat{X}_{ci}$ is the expected level of trade value in the country. RCA assumes that the expected level of trade of country $c$ in industry $i$ should be proportional to the share of country $c$ in world exports. We denote the total exports of country $c$ as $X_c$, total exports of product $i$ as $X_i$ and total trade in the world with $X_W$, Mathematically, these can be written as:

$$X_c \equiv \sum_i X_{ci}, \qquad X_i \equiv \sum_c X_{ci} \quad \text{and} \quad X_W \equiv \sum_c \sum_i X_{ci}.$$

With these in hand, we can write the expected value of the exports of country $c$ in industry $i$ as:

$$\hat{X}_{ci} = X_i \frac{X_c}{X_W}.$$

Therefore, the RCA measure is defined as:

$$\text{RCA}_{ci} \equiv \frac{X_{ci}/X_i}{X_c/X_W} = \frac{X_{ci}/X_c}{X_i/X_W} = \frac{X_{ci}X_W}{X_cX_i}.$$

We normalize the RCA values by a transformation that takes a value between 0 and 1.

Therefore, we use:

$$\widehat{\mathrm{RCA}}_{ci} = \frac{\mathrm{RCA}_{ci}}{\mathrm{RCA}_{ci}+1}.$$

Another comparative advantage value uses the population share of a country $c$ while calculating the expected trade. Mathematically, this accounts to:

$$\hat{X}_{ci} = X_i \frac{\mathrm{pop}_c}{\mathrm{pop}_W},$$

where $\mathrm{pop}_c$ captures the population of country $c$ and $\mathrm{pop}_W$ captures that of the world. The Revealed per capita advantage (RpCA) is defined as:

$$\mathrm{RpCA}_{ci} \equiv \frac{X_{ci}/X_i}{\mathrm{pop}_c/\mathrm{pop}_W}.$$

Similar capping to the RpCA can be applied using the population shares of countries and trade shares of the products.

We use the normalized RCA and RpCA to create an index that takes advantage of both measures. In particular, we define:

$$\widetilde{\widehat{\mathrm{RCA}}}_{ci} \equiv \left(\widehat{\mathrm{RCA}}_{ci}\right)^{\alpha} \left(\widehat{\mathrm{RpCA}}_{ci}\right)^{1-\alpha},$$

where $\alpha \in [0,1]$ is the weight of the measure. We use $\alpha = 0.5$ in our calculations.

When we binarize the data, we use RCA. But in many instances, we revert to a normalized version of RCA to take advantage of its continuous nature. In particular, economic complexity measures are built with the normalized values.

### 5.1.3 Binary presence/absence matrix

Many complexity calculations require defining a presence/absence matrix. Comparative advantage measures, such as RCA, provide a good basis for such calculations. In particular, we assume that the country has all the necessary capabilities (letters) for making a product if it competitively exports the product with a comparative advantage value larger than 1. Mathematically, we define the country presence/absence matrix, $M$, as follows:

$$M_{c,i} = \begin{cases} 1 & \text{if} \quad \text{RCA}_{ci} \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

This simple definition has a caveat, especially for invention and scientific output datasets. For example, a country like the United States is active in almost all technology classes or scientific fields, but because of the nature of the RCA variable, some of them would be considered as absent because of the threshold of 1 even though the US could be among the top inventors or publishers in the field. To circumvent this issue, we add an additional criterion allowing us to assign a presence value if the country is among the top-ranked countries in a technology class, a scientific field or a trademark field. The effective number of countries active in a field could be defined as the inverse Herfindahl–Hirschman Index (HHI):

$$\frac{1}{n_t} \equiv \text{HHI}_t \equiv \sum_c \left( \frac{P_{ct}}{P_t} \right)^2,$$

where $P_t$ is the total number of patents in the technology class globally. Let's define $\text{rank}_{ct}$ as the rank of country $c$ in technology class $t$. Then, the updated $M^P$ matrix is defined as:

$$M^P_{c,t} = \begin{cases} 1 & \text{if} \quad (\text{RCA}_{ct} \geq 1 \text{ and } \text{rank}_{ct} \leq 3 * n_t) \text{ or } (\text{rank}_{ct} \leq n_t) \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We make sure that the countries are among time $3 * n_t$ to eliminate very small numbers of output in a field dependent manner. A similar modification is done for the presence/absence matrix for scientific output, $M^S$.

### 5.2 Diversity and Ubiquity

With binary presence/absence matrix -$M$- in hand, we can define the economic complexity metrics. Two fundamental concepts, **diversity** and **ubiquity**, intertwine closely with the presence / absence matrix. While diversity assesses the range of products a country produces with a comparative advantage, ubiquity evaluates how widespread a product is among all countries. Both offer insights into a nation's productive capabilities and global demand.

First, we define diversity as the number of products that a country has a presence in. Mathematically, we define diversity, $k_{c,0}$, as:

$$k_{c,0} \equiv \sum_{i} M_{ci}.$$

Similarly, the ubiquity of a product is defined as the number of countries making the product:

$$k_{i,0} \equiv \sum_{c} M_{ci}.$$

Highly complex countries have relatively high diversity, and highly complex products are made by fewer countries exhibiting lower ubiquity levels. These measures could be considered the zeroth order approximation to the complexity levels (hence the zero in their subscript).

Since we have multiple dimensions of Science, Technology, Entrepreneurship and Production (trade), when we compare across fields, we normalize the measures. In particular:

$$\hat{k}_{c,0} \equiv \frac{k_{c,0}}{N_i},$$

where $N_i$ is the number of fields.

On ICO 2026, diversity and ubiquity are featured, for instance, in Figures 2.3 and Table 2.2.

### 5.3 Linking Dimensions and Time

We create normalized RCA and $M$ matrices within each dimension for Science, Technology, Entrepreneurship and Production (trade). For the complexity calculations below, we combine them along the country dimension to create a one "capability" matrix for whole innovation ecosystems. With some abuse of notation, we will call these matrices $RCA$ and $M$ matrices below. We can write these matrices as:

$$M = [M^S\ M^T\ M^E\ M^P] \qquad \text{and} \qquad \mathrm{RCA} = [\widetilde{\mathrm{RCA}}^S\ \widetilde{\mathrm{RCA}}^T\ \widetilde{\mathrm{RCA}}^E\ \widetilde{\mathrm{RCA}}^P].$$

Note that each RCA and $M$ matrix has dimension of $N_C \times N_F$ where $N_F$ is the total number of fields in Science, Technology, Entrepreneurship and Production (trade) dimensions.

When we compare the complexity variables over time, we of then fix the complexity of the fields but allow countries complexity to change over time. To achieve this, we stack the yearly $\mathrm{RCA}_y$ and $M_y$ matrices to create a one giant matrix that has country, year in the rows and fields for all dimensions in the columns.

$$\mathrm{gRCA} = \begin{bmatrix} \mathrm{RCA}_1 \\ \vdots \\ \mathrm{RCA}_Y, \end{bmatrix} \qquad gM = \begin{bmatrix} M_1 \\ \vdots \\ M_Y \end{bmatrix}$$

where $Y$ is the total number of years. Hence, gRCA and $gM$ matrices are $(N_C Y) \times N_F$ dimensional.

### 5.4 Economic Complexity and Product Complexity Indexes

The **Economic Complexity Index (ECI)** measures the sophistication and knowledge accumulation of an economy. It captures an economy's ability to produce a wide variety of complex products. Higher values of ECI signify that an economy produces diverse products that are less commonly manufactured globally, revealing depth in knowledge and capabilities.

The Economic Complexity Index (ECI) and the Product Complexity Index (PCI) are calculated as refinements to diversity and ubiquity measures. Hidalgo & Hausmann (2009) introduce these measures as an iterative process with the following:

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_i M_{ci} k_{i,n-1} \quad \text{and} \quad k_{i,n} = \frac{1}{k_{c,0}} \sum_c M_{ci} k_{c,n-1}.$$

This iterative process has a trivial solution where all complexity variables are equal to each other. If we focus on the component driving the difference between countries, this process results in the following specification that gives us ECI and PCI:

$$\text{ECI}_c = \frac{\gamma}{k_{c,0}} \sum_i M_{ci} \text{PCI}_i \quad \text{and} \quad \text{PCI}_i = \frac{\gamma}{k_{i,0}} \sum_c M_{ci} \text{ECI}_c,$$

where $\gamma$ is a constant that will be determined below. We can write these equations in terms of matrix equations. First, we define the matrix $D$ ($U$) as the diagonal matrix whose diagonal elements correspond to the diversity (ubiquity). Hence:

$$\text{ECI} = \gamma\, D^{-1}\, M\, \text{PCI} \quad \text{and} \quad \text{PCI} = \gamma\, U^{-1}\, M^{\dagger}\, \text{ECI}.$$

Combining these two equations yields:

$$\text{ECI} = \gamma^2 \underbrace{D^{-1}\, M\, U^{-1}\, M^{\dagger}}_{\equiv \widetilde{M}} \text{ECI}.$$

Mathematically, we define ECI as the eigenvector corresponding to the second largest eigenvalue ($\lambda_2$) of the matrix $\widetilde{M}$ with $\gamma = (\lambda_2)^{-1/2}$.

With these normalization, ECI of a country could be compared to the PCI of a product. In particular, these normalizations allow us to write:

$$\text{ECI}_c = \frac{1}{\sqrt{\lambda_2} k_{c,0}} \sum_i M_{ci} \text{PCI}_i \qquad \text{and} \qquad \text{PCI}_i = \frac{1}{\sqrt{\lambda_2} k_{i,0}} \sum_c M_{ci} \text{ECI}_c.$$

For example, with these normalization, we can compare ECI of a country like Argentina to the PCI of a scientific publication such as artificial intelligence.

Another diversion from Hausmann et al. (2014) is that we do not standardize (i.e., subtract mean and divide by the standard deviation) ECI and PCI. But ECI and PCI are robust to a multiplicative factor.

Above, we explain the complexity calculations with $M$ matrix. But all these calculations could also be done with RCA, gRCA and $gM$ matrices. In our baseline classification we often use RCA. When we need to compare the complexity values over time, we rely on gRCA matrix.

On ICO 2026, ECIs and PCIs are presented as ecosystem complexity and capability complexity, respectively in several figures and tables (Figure 1.3, Figure 2.4, 2.7, Table 3.1, etc.).

## 5.5 Inferring Capability Overlaps

### 5.5.1 Relatedness

**Relatedness** examines the shared capabilities, skills, and know-how necessary for producing two different products. Two products are considered more related if countries that competitively produce one product are also frequently competitive in the other.

Relatedness or proximity measure captures the capability overlap between products, between technology classes, or between scientific fields. In the international trade data, we infer the capability overlaps through the co-location of exports. In particular, we measure the probability of a country exporting a product $i$ given that the country already exports $i'$. To minimize the error, we take the minimum of these conditional probabilities between these products. Mathematically, the proximity between to products, $i$ and $i'$ is:

$$\phi_{ii'} = \frac{\sum_c M_{ci} M_{ci'}}{\max(k_{i,0}\,,\, k_{i',0})}. \tag{2}$$

We can filter the overlaps that are not different from an expectation that stems from a Bernoulli distribution. For example, if field $i$ is made by 50% of the countries, and field $j$ is also made by 50% of the countries, then we would expect 25% of the countries to overlap if they were distributed randomly. We posit that the random overlaps do not add information. That is why we devised a mechanism to eliminate the overlaps that are not significantly different with 95% confidence from a Bernoulli expectation. In particular:

- Overlap between entities $i$ and $j$ is:

$$\tilde{O}_{ij} = \sum_c M_{ci} M_{cj}$$

- To determine if the overlap is significant, we can define the expected overlap:

$$\hat{O}_{ij} = \frac{k_i k_j}{N_C}$$

and the standard deviation of the overlap:

$$\sigma_{ij} = \sqrt{\frac{k_i k_j (N_C^2 - k_i k_j)}{N_C}}$$

With these can filter out non-significant overlaps:

$$O_{ij} \equiv \begin{cases} \sum_c M_{ci} M_{cj} & \text{if } \tilde{O}_{ij} > \hat{O}_{ij} + 1.96\sigma_{ij} \\ 0 & \text{otherwise} \end{cases}$$

- Proximity between entries can be defined from the filtered overlaps as:

$$\phi_{ij} \equiv \frac{O_{ij}}{\max(k_i, k_j)}$$

The relatedness of products can be visualized on a network called the **Product Space**. In the Product Space, products are represented as nodes, and the proximity between products indicates how often they co-occur in countries' export baskets, suggesting shared capabilities in production. On ICO 2026, this is presented as the innovation capability space in Figure 1.2.

### 5.5.2 Density

We infer the overlap between capabilities present in a location and capabilities required for a product, a technology class, or a scientific field through a measure called density. The density measure is calculated as the share of "relatedness" present in a location around a product.

Mathematically, we can write the density as defined in Hidalgo et al. (2007):

$$d_{ci} = \frac{\sum_{i'} M_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}. \tag{3}$$

The numerator captures the portion of the related products that are present in the country. If the country makes products in industries with high relatedness, we expect the density value to be large. One caveat of the density measure is that if the country makes many products, the density also increases.

We can also write a density for the continuous RCA-like measures, which Hausmann et al. (2022) define as implied comparative advantage. This measure is calculated as the weighted average of RCAs of a country in related products. Mathematically, the implied comparative advantage measure can be written as:

$$\hat{R}_{ci} = \frac{\sum_{i'} \mathrm{RCA}_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}.$$

Density values are presented on ICO 2026 as "relatedness" on Table 3.1 and 3.2.

We can write the density metric as a sum of contributions coming from different dimensions. In particular:

$$d_{ci} = \underbrace{\frac{\sum_{i' \in S} M_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}}_{\substack{\text{Contribution from} \\ \text{Science}}} + \underbrace{\frac{\sum_{i' \in T} M_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}}_{\substack{\text{Contribution from} \\ \text{Technology}}} + \underbrace{\frac{\sum_{i' \in E} M_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}}_{\substack{\text{Contribution from} \\ \text{Entrepreneurship}}} + \underbrace{\frac{\sum_{i' \in P} M_{ci'} \phi_{ii'}}{\sum_{i'} \phi_{ii'}}}_{\substack{\text{Contribution from} \\ \text{Production}}}.$$

Here, $i$ could be from any dimension. We use this decomposition in our report extensively.

## 5.6 Calculating Expected Outcome (Cross-fertilization)

Calculating expected outcome in any give field of innovation follows the method proposed by Moscatelli et al. (2026). Based on what they produce on related outputs, economies can estimate their expected number of innovation outputs in any field. For instance, the number of patents given the production structure involves the following steps:

- First, we calculate a normalization factor for each technology class:

$$\Gamma_t^X = \frac{T_t}{\sum_p X_p \phi_{pt}}$$

- With this normalization, the expected output in technology class $t$ in country $c$ is:

$$T_t^X = \Gamma_t^X \sum_p X_{cp}\phi_{pt} = T_t \frac{\sum_p X_{cp}\phi_{pt}}{\sum_p X_p \phi_{pt}}$$

- Similarly, for scientific fields:

$$\Gamma_t^F = \frac{T_t}{\sum_f F_f \phi_{ft}}$$

  and

$$T_t^F = \Gamma_t^F \sum_f F_{cf}\phi_{ft} = T_t \frac{\sum_f F_{cf}\phi_{ft}}{\sum_f F_f \phi_{ft}}$$

In ICO 2026, this indicator appears as innovation potential in Figures 3.4, Figure 3.5, Table 3.3, Table 3.4 and Table 3.5.

## 5.7 Growth, Entries and Exits

To measure the growth of comparative advantage of a country $c$ in product $i$, we use the following growth measure:

$$\text{Growth}_{ci,y-y'} = \frac{X_{ci,y'}}{X_{ci,y}} - 1$$

We define the appearance of a product $i$ in country $c$ as:

$$\text{Appearance}_{ci,y-y'} = \begin{cases} 1 & \text{if} \quad M_{ci,y} = 0 \text{ and } M_{ci,y'} = 1 \\ 0 & \text{if} \quad M_{ci,y} = 0 \text{ and } M_{ci,y'} = 0 \\ \text{undefined} & \text{otherwise} \end{cases}$$

and disappearance as:

$$\text{Disappearance}_{ci,y-y'} = \begin{cases} 1 & \text{if} \quad M_{ci,y} = 1 \text{ and } M_{ci,y'} = 0 \\ 0 & \text{if} \quad M_{ci,y} = 1 \text{ and } M_{ci,y'} = 1 \\ \text{undefined} & \text{otherwise} \end{cases}$$

## 5.8 Other Measures

### 5.8.1 Growth Potential

Another measure that we introduce is the weighted growth of the outputs of a location to capture the sustainability aspect. Formally, we define the growth sustainability as:

$$\text{Sustainability}_c = \sum_p \frac{X_{cp}}{X_c} \underbrace{\frac{X_{p,t+\Delta}}{X_{p,t}}}_{g_{p,\Delta}},$$

where $g_{p,\Delta}$ is the growth rate of field $p$ in the world.

### 5.8.2 Complexity Optimization

ECI measure gives equal weight to the fields that the location is active in. We can also capture whether the location gives higher weights to complex activities. Formally, the complexity optimization is given by the PCI values weighted by the output shares

$$\text{Optimization}_c = \sum_p \frac{X_{cp}}{X_c} \times \text{PCI}_p.$$

The differences between ECI and the Optimization values might highlight discrepancies that a country can focus on.

In ICO 2026, this indicator appears as capability management in Figure 2.10 and Table 2.5.

### 5.8.3 Complexity Outlook Index

This is a measure that captures whether a country is close by to high complexity products/ technologies / fields (Hausmann et al., 2014). Mathematically, it can be defined as:

$$\mathrm{COI}_c = \sum_p (1 - M_{cp}) d_{cp} \mathrm{PCI}_p.$$

In ICO 2026, this indicator appears first as the ease to diversify in complex fields in Figure 3.1 and Figure 3.2. Additionally, a variation is added to define ease to diffuse into complex ecosystems in Figure 3.3.

$$\mathrm{COI}_p = \sum_c (1 - M_{cp}) d_{cp} \mathrm{ECI}_c.$$

### 5.8.4 Complexity Outlook Gain

This measure is also introduced by Hausmann et al. (2014). It captures the distribution of different locations on the product space vis-a-vis complex products. Complexity Outlook Gain is defined as the change in complexity outlook if location $c$ starts makes entity $p$. We can define the complexity outlook for location $c$ after it makes entity $p$ with $\mathrm{COI}_{c,p}$. Then the complexity outlook gain is:

$$\mathrm{COG}_{cp} = \mathrm{COI}_{c,p} - \mathrm{COI}_c.$$

### 5.8.5 Growth Outlook Index

This is a variant of Complexity Outlook Index with product attractiveness measure replaced with its growth rate instead of PCI:

$$\mathrm{GOI}_c = \sum_p (1 - M_{cp}) d_{cp} g_p,$$

where $g_p$ is the global output of this field in last 5 ($\Delta$) years.

In ICO 2026, this indicator appears as the ease to diversify in fast growing fields in Figure 3.1 and Figure 3.2. Additionally, a variation is added to define ease to diffuse into fast growing ecosystems in Figure 3.3, with growth for each location $c$ considered for the dimension $d$ thta corresponds to each field $p$.

$$\mathrm{GOI}_p = \sum_c (1 - M_{cp}) d_{cp} g_c^d,$$

# 6 Conclusion

The proposed complexity measure could help countries (including WIPO member states) orient through the complex ecosystem while designing innovation policies. As we discuss in Section 2 above, these metrics can be used to define attractiveness / feasibility dimensions in multitudes of ways. In addition, the current capability levels of countries revealed through different metrics could also be captured with the complexity metrics.

To summarize, the complexity metrics can be used to:

i Assess the current level of capabilities of a location in each innovation domain by assessing:

  (a) Resilience of capabilities.

  (b) Complexity Level.

  (c) Overall potential for further diversification.

  We have developed several measures in Section **??** that could be applied to each domain.

ii Assess the sophistication and other attractiveness measures for each entity in innovation domains.

iii Assess the feasibility of each entity in each location through the relatedness metrics.

iv Analyze potential / feasibility trade-offs using cross-domain density measures.

v Integrate attractiveness and feasibility dimensions to capture further diversification opportunities.

Atlas of Economic Complexity[9] has some of these dimensions at the country level under the "Country Profiles" section using international trade data. This could be the starting point for country profiles for innovation domains.

Policymakers must consider the co-evolution of different cognitive domains and their interdependencies. Promoting, for instance, the acceleration of certain scientific domains through public funding may induce positive externalities that increase the likelihood of new technological capabilities emerging related to those scientific domains. Similarly, the emergence of a technological capability is correlated with countries' past exports in related domains. Hence, developing complex technologies can leverage existing scientific and industrial capabilities. Disentangling how these capabilities interact in a given location and identifying their most binding constraints to grow may boost the presence of positive externalities and the development of a resilient innovation ecosystem.

[9] `https://atlas.hks.harvard.edu/`

## References

Ansorge, L. (2023). Hidden limitations of analyses via alternative bibliometric services. *Scientometrics*, *128*(3), 2031–2033. doi: 10.1007/s11192-022-04626-w

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School*, *33*(2), 99–123.

Balland, P.-A., & Boschma, R. (2022). Do scientific capabilities in specific domains matter for technological diversification in European regions? *Research Policy*, *51*(10), 104594. doi: 10.1016/j.respol.2022.104594

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268.

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450.

Balland, P.-A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23.

Broekel, T. (2019). Using structural diversity to measure the complexity of technologies. *PloS one*, *14*(5), e0216856.

Bustos, S., Jackson, E., Torun, D., Leonard, B., Tuzcu, N., Lukaszuk, P., … Yildirim, M. A. (2024). *Tackling discrepancies in trade data: The harvard growth lab international trade datasets* (Growth Lab Working Paper Series No. 251). Harvard Kennedy School.

Catalán, P., Navarrete, C., & Figueroa, F. (2022). The scientific and technological cross-space: is technological diversification driven by scientific endogenous capacity? *Research Policy*, *51*(8), 104016.

Cohen, W. M., Nelson, R., & Walsh, J. P. (2000). *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (Working Paper No. 7552). Cambridge, MA, USA: National Bureau of Economic Research.

Coniglio, N. D., Vurchio, D., Cantore, N., & Clara, M. (2021). On the evolution of comparative advantage: path-dependent versus path-defying changes. *Journal of International Economics*, *133*, 103522.

Furman, J. L., Porter, M. E., & Stern, S. (2002). The determinants of national innovative capacity. *Research policy*, *31*(6), 899–933.

Golden, J. M. (2007). "Patent trolls" and patent remedies. *Texas Law Review*, *85*(7), 2111.

Gruber, J., & Johnson, S. (2019). *Jump-starting america: How breakthrough science can revive economic growth and the american dream*. Hachette UK.

Harabi, N. (1995). Appropriability of technical innovations an empirical analysis. *Research Policy*, *24*(6), 981–992.

Harvard's Growth Lab. (2023). *The Atlas of Economic Complexity.* Retrieved 2023-09-26, from `https://atlas.cid.harvard.edu/`

Harzing, A.-W., & Alakangas, S. (2017). Microsoft Academic: Is the Phoenix Getting Wings? *Scientometrics*, *110*(1), 371–383. doi: 10.1007/s11192-016-2185-x

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yıldırım, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. The MIT Press.

Hausmann, R., & Klinger, B. (2006). *Structural transformation and patterns of comparative advantage in the product space* (CID Working Paper No. RWP06-041). Harvard Kennedy School.

Hausmann, R., Rodrik, D., & Velasco, A. (2008). Growth diagnostics. *The Washington consensus reconsidered: Towards a new global governance*, *2008*, 324–355.

Hausmann, R., Stock, D. P., & Yıldırım, M. A. (2022). Implied comparative advantage. *Research Policy*, *51*(8), 104143.

Hausmann, R., Yildirim, M. A., Chacua, C., Hartog, M., & Matha, S. G. (2024a). *Global trends in innovation patterns: A complexity approach* (Growth Lab Working Paper Series No. 235). Harvard Kennedy School.

Hausmann, R., Yildirim, M. A., Chacua, C., Hartog, M., & Matha, S. G. (2024b). *Innovation policies under economic complexity* (Growth Lab Working Paper Series No. 234). Harvard Kennedy School.

Henkel, J. (2022). Licensing standard-essential patents in the IoT – A value chain perspective on the markets for technology. *Research policy*, *51*(10), 104600.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, *3*(2), 92–113.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., ... others (2018). The principle of relatedness. In *Unifying themes in complex systems ix: Proceedings of the ninth international conference on complex systems 9* (pp. 451–457).

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487.

Huang, C.-K. K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, *1*(2), 445–478. doi: 10.1162/qss_a_00031

Jiao, C., Li, K., & Fang, Z. (2023). *How are exclusively data journals indexed in major scholarly databases? An examination of the Web of Science, Scopus, Dimensions, and OpenAlex* (No. arXiv:2307.09704). arXiv. doi: 10.48550/arXiv.2307.09704

Kenney, M., & Mowery, D. C. (2020). *Public universities and regional growth: Insights from the University of California*. Stanford University Press.

Lehmann, E. E., & Menter, M. (2016). University–industry collaboration and regional wealth. *The Journal of Technology Transfer*, *41*, 1284–1307.

Lemley, M. A., & Shapiro, C. (2007). Patent holdup and royalty stacking. *Texas law review*, *85*(7), 1991.

Levin, R. C., Klevorick, A. K., Nelson, R. R., Winter, S. G., Gilbert, R., & Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brookings papers on economic activity*, *1987*(3), 783–831.

Li, Y., & Neffke, F. (2023). *Evaluating the principle of relatedness: Estimation, drivers and implications for policy* (CID Research Fellows and Graduate Student Working Paper Series No. 146). Cambridge, MA: Center for International Development at Harvard University.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*(1), 871–906. doi: 10.1007/s11192-020-03690-4

Mazzucato, M. (2015). *The entrepreneurial state: debunking public vs. private sector myths* (Revised edition. ed.). New York, NY: PublicAffairs.

Mezzanotti, F., & Simcoe, T. (2023). *Innovation and appropriability: Revisiting the role of intellectual property* (Working Paper No. 31428). Cambridge, MA, USA: National Bureau of Economic Research.

Miguelez, E., & Morrison, A. (2023). Migrant inventors as agents of technological change. *The Journal of Technology Transfer*, *48*(2), 669–692.

Miguelez, E., Raffo, J., Chacua, C., Coda-Zabetta, M., Yin, D., Lissoni, F., & Tarasconi, G. (2019). Tied in: The global network of local innovation. *WIPO Economic Research Working Paper*, *58*.

Morrison, A. (2023). Towards an evolutionary economic geography research agenda to study migration and innovation. *Cambridge Journal of Regions, Economy and Society*, rsad013.

Moscatelli, F., Chacua, C., Matha, S. G., Hartog, M., Hernandez Rodriguez, E., & Yildirim, M. A. (2024). *Can we map innovation capabilities?* (Economic Research Working Paper Series No. 81). World Intellectual Property Organization (WIPO).

Moscatelli, F., Raffo, J., Gadgin Matha, S., Chacua, C., Hartog, M., Hernandez Rodriguez, E., & Yildirim, M. A. (2026). *The technological potential of innovation ecosystems* (Economic Research Working Paper Series No. 90). World Intellectual Property Organization (WIPO). Retrieved from `http://tind.wipo.int/record/59226` doi: https://doi.org/10.34667/tind.59226

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, *94*(1), 23–48.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, *34*(3), 297–316.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic geography*, *87*(3), 237–265.

Petralia, S., Balland, P.-A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5), 956–969.

Pinheiro, F. L., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2022). The time and frequency of unrelated diversification. *Research Policy*, *51*(8), 104323.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., & Gabrielli, A. (2019). Unfolding the innovation system for the development of countries: coevolution of science, technology and production. *Scientific reports*, *9*(1), 16440.

Sbardella, A., Barbieri, N., Consoli, D., Napolitano, L., Perruchas, F., & Pugliese, E. (2022). *The regional green potential of the european innovation system.* (mimeo, Joint Research Centre (Seville site))

Scheidsteger, T., & Haunschild, R. (2022). Comparison of metadata with relevance for bibliometrics between Microsoft Academic Graph and OpenAlex until 2020. In *26th International Conference on Science and Technology Indicators.* Granada, Spain. doi: 10.5281/zenodo.6975102

Scheidsteger, T., Haunschild, R., & Bornmann, L. (2023). How similar are field-normalized scores from different free or commercial databases calculated for large German universities? In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023).* International Conference on Science, Technology and Innovation Indicators. doi: 10.55835/6441118c643beb0d90fc543f

Shin, H., Kogler, D. F., & Kim, K. (2023). The relevance of scientific knowledge externalities for technological change and resulting inventions across European metropolitan areas. *Review of Regional Research*, 1–17.

Stojkoski, V., Koch, P., & Hidalgo, C. A. (2023). Multidimensional economic complexity and inclusive green growth. *Communications Earth & Environment*, *4*(1), 130.

Straccamore, M., Bruno, M., Monechi, B., & Loreto, V. (2023). Urban economic fitness and complexity from patent data. *Scientific Reports*, *13*(1), 3655.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., & Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific reports*, *2*(1), 723.

Thelwall, M. (2017). Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals. *Journal of Informetrics*, *11*(4), 1201–1212. doi: 10.1016/j.joi.2017.10.006

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413. doi: 10.1162/qss_a_00021

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., ... Rogahn, R. (2019). A Review of Microsoft Academic Services for Science of Science Studies. *Frontiers in Big Data*, *2*.