# The universal structure of national scientific development

Lili Miao[1], Dakota Murray[1], Woo-Sung Jung[2,3], Vincent Larivière[4,5], Cassidy R. Sugimoto[6], Yong-Yeol Ahn[*] [1,7,8]

[1]*Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing and Engineering, Indiana University Bloomington, IN 47408, USA.*

[2]*Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, Republic of Korea.*

[3]*Department of Physics, Pohang University of Science and Technology, Pohang 37673, Republic of Korea.*

[4]*École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, Québec H2L 2C4, Canada.*

[5]*Observatoire des sciences et des technologies, Université du Québec à Montréal, Montréal, Québec H2L 2C4, Canada.*

[6]*Luddy School of Informatics, Computing and Engineering, Indiana University Bloomington, IN 47408, USA.*

[7]*Network Science Institute, Indiana University, Bloomington (IUNI), IN 47408, USA*

[8]*Connection Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

## Abstract

Science is considered essential to innovation and economic prosperity[1–5]. Understanding how nations develop scientific capacity is therefore crucial for reducing global inequities and addressing the wealth gap[2]. Although studies have shown that a national scientific development is affected by its geographic, historic, and economic factors[6–10], it remains unclear whether there are universal structures and trajectories behind national scientific development that can inform decision making and forecasting. By examining countries' *scientific exportation*—the number of publications that are indexed in an international bibliographic database, we reveal a three-cluster structure in the relatedness network of disciplines that underpin national scientific development and the organization of global science. Tracing the evolution of national research portfolios reveals that while nations are proceeding to more diverse

[*] yyahn@iu.edu

research profiles, scientific production is increasingly specialized in global science. We further demonstrate that the diversity of a nation's scientific exports is more predictive of economic growth than Economic Complexity Index defined by the exportation of products. Our results suggest that low scientific diversity contributes to the development trap faced by low-income countries. By uncovering the underlying structure of scientific development, our results may offer a new perspective to study national scientific development and its relationships with economic development.

## Introduction

Is there a universal trajectory of national scientific development? Several approaches have been taken to describe the evolution of science as well as the universal patterns of scientific development. Comte[11] argued—albeit not considering national context—that science develops along an innate trajectory from high-consensus physical sciences, towards more complex, low-consensus social sciences. Basalla[12] took a colonial perspective, arguing that scientific development in non-Western countries began with colonial exportation—providing natural resources to Western countries—and then developed their scientific capacity within the Western tradition. In contrast to this "western recipe" of national scientific development, studies have examined how the interplay between geography[10], history[9], existing scientific strengths[13], and economic conditions[14] influence development. Chile exemplifies the influence of geopolitical opportunities and constraints on national knowledge production: despite relatively low scientific investments[15], Chile's unique mountainous and remote terrain made it ideal for astronomical observatories, a comparative advantage that allowed the nation to become an international hub for astronomy and astrophysics[16,17]. By contrast, South Korea, with its heavy investment in science[18,19], has experienced diversified scientific expansion, developing into a science and innovation powerhouse[20].

With the increasing scientific capacity across countries, global science has been experiencing rapid transformation. During Cold War, the USSR and the United States competed in science; the collapse of the USSR in the late 1990s and the concurrent rise of China on the international stage significantly altered the power dynamics in science. Whereas China only accounted for 5% of the world's scientific publications in 2000, it became the most productive country in the world by 2018, surpassing US production[21]. The increase in scientific growth was also coupled with Asia's economic take-off: for example, the rapid expansion and intense industrialization of the "Four Asian Tigers"— Hong Kong, Singapore, South Korea, and Taiwan, also occurred during this time. These contemporary shifts prompt the question of the relationship between economic and scientific development and the potential universality of the trajectory of scientific development.

The study of economic output with respect to the "product space"—i.e., the network of relatedness between exported products—revealed that the networked structure of industrial advantage is critical to understanding the economic development of nations[22]. We apply this framework to examine national science production by considering scientific disciplines as types of "products" that are exported by countries. That is, we investigate nation's scientific development through their *scientific exports*, in which research articles produced by a country and indexed in international bibliographic databases represent the exported scientific "products" of the nation[23,24]. This is an important operationalization for our study, whereas there is a significant amount of scientific production that may happen within a country which —e.g., in non-English languages, in grey literature, or governmental reports—we argue that it is those works that are made visible through indexation that are the best proxies for *exportation*. This is not to diminish the scientific activity within a country, but rather to create a measurement that approximates economic exportation.

## Results

We employ Revealed Comparative Advantage (RCA)[12]—a common measure for quantifying the economic and production advantages of countries[22]—to assess each nation's relative disciplinary strengths based on publications indexed by the Web of Science database (see Data & Methods). If country $c$ produces a greater share of its publications in field $i$ compared to the world average share in the discipline, then $\mathrm{RCA}_{c,i} > 1$ and country $c$ is considered to have a revealed comparative advantage (or specialization) in discipline $i$.

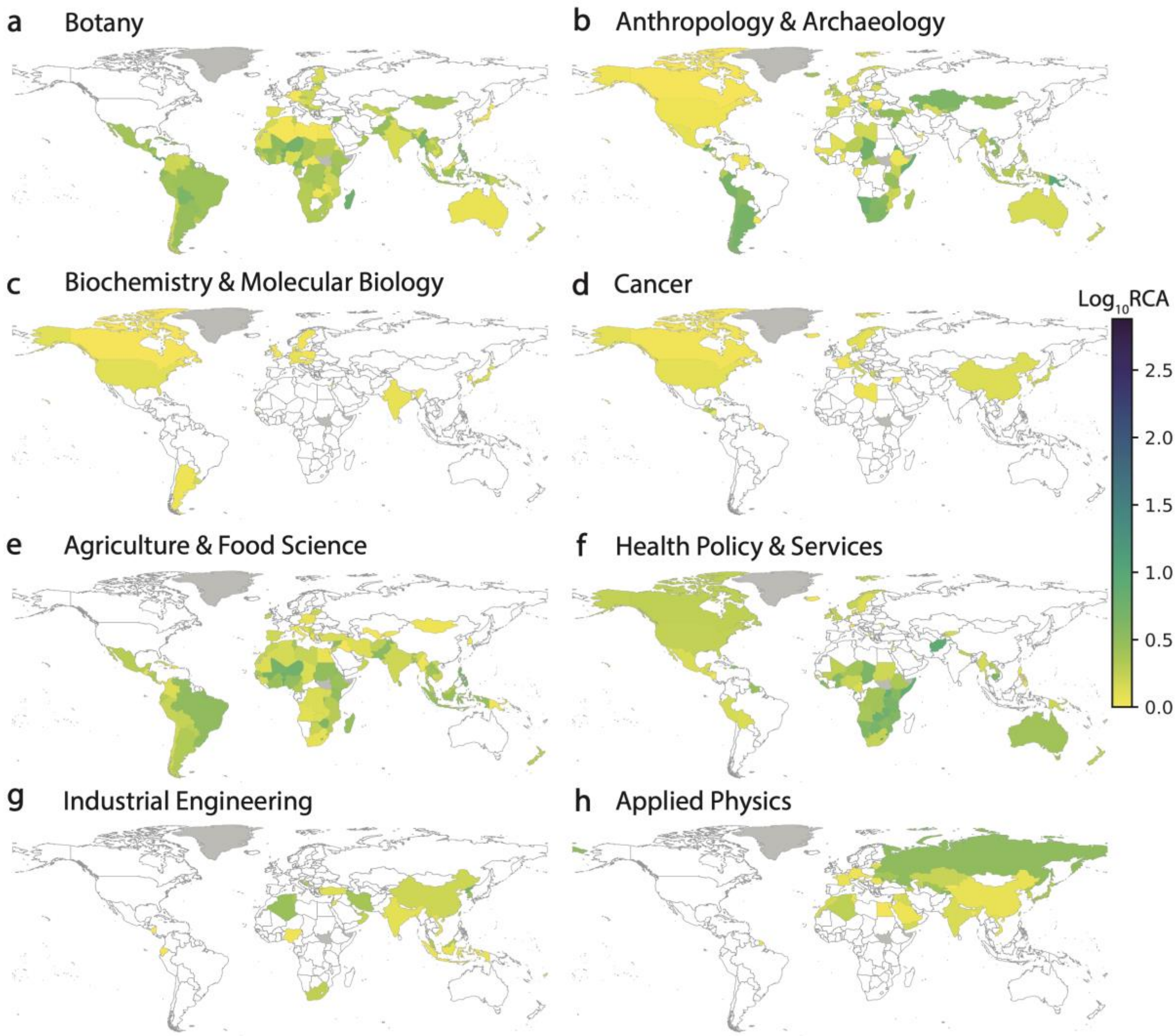


***Fig 1. Disciplinary specialization reflects geographical, historical and economic factors.*** *Eight examples illustrate the distribution of disciplinary specializations. Discipline specialization is measured by Revealed Comparative Advantage (RCA). Color represents the Logarithm of RCA; a nation is only colored if its* $\log_{10} \mathrm{RCA}_{c,i} > 0$. *Grey corresponds to nations that were not represented in our dataset. Botany, Anthropology, and Archaeology reflect the presence and access to natural and anthropological resources in a country. Economic inequality underpin specialization in resource-intensive disciplines like Biochemistry & Molecular Biology and Cancer. Local issues also drive research, as can be seen from the distribution of Agricultural & Food Science and Health Policy & Services. The distribution of Industrial Engineering and Applied Physics likely reflects national economic priorities and policies.*

We calculated the RCA for all combinations of 143 disciplines and all countries in our dataset. As expected, the patterns of relative advantage reflect a range of historical, geographical, and cultural factors (see Fig. 1). For instance, countries with relative strength in *Botany* are located in tropical areas rich in botanical resources; *Anthropology and Archaeology* features both wealthy and developing nations, reflecting the remnants of colonial science and alluding Basalla's postulation that the science in colonial and post-colonial countries began with Western countries' exploitation of natural resources[12]. By contrast, far fewer countries—mostly in North America and Europe—specialize in *Biochemistry & Molecular Biology*, a discipline which requires sufficient funding and sophisticated technologies. Similarly, *Cancer* research is largely concentrated in countries with high cancer mortality (which is associated with longer lifespans)  as well as advanced countries with the capacity to invest in clinical research[25]. That research and innovation emerge as a response to local issues and threats can also be observed in other contexts. For example, *Agricultural & Food Science* and *Health Policy and Services* are prominent in nations across the global south, where infectious disease[26] and food security[27] are pressing issues. Large emerging economies like China and India are specialized in fields such as *Industrial Engineering* and *Applied Physics* that contribute to industrially-relevant research[28]. Similarly, the relative strength of Russia, Ukraine and Kazakhstan in Applied Physics may be explained as a remnant of the Soviet Union's research priorities[9].

The distribution of disciplinary specialization suggests scientific exportation is affected by geographic, historical, social, and economic factors. Do these idiosyncratic factors dominate the course of scientific development of a nation? Or is there an underlying structure that governs the scientific development of nations?

Inspired by the relatedness network of economic product exports that underpins national economic development[22], we construct a discipline relatedness network, in which the proximity between disciplines is defined by the minimum conditional probability that two disciplines are co-specialized in a country (see Data & Methods). The network builds on the idea that disciplines that are co-specialized are likely to require similar knowledge, skills, methods, or equipment. To reveal its most salient structure, we apply the multi-scale backbone extraction method[29]. This "backbone" reveals three clusters—which we confirm in the full network with Leiden Algorithm[30] (see Data & Methods)—which we call *Natural*, *Physical*, and *Societal* clusters (see Fig. 2a). These clusters—while resembling previous observations[31]—do not conform to the common high-level classifications of disciplines. None of the clusters exclusively coincide with major classifications such as natural sciences, engineering, or medical sciences. The high-level disciplinary classifications appearing in *Natural* cluster (left) are primarily Natural and Medical Sciences. Most disciplines are dependent upon natural resources (e.g., Geology, Entomology, and Agriculture & Food Science), or concern the prevalent medical concerns in low-income areas (e.g., Nutrition & Dietetic and Parasitology). The *Physical* cluster (right) contains primarily physical sciences and engineering, which are commonly considered as foundations for industry-based economic growth (e.g., Chemistry and Applied Physics) and those that require technological investment (e.g., Civil Engineering, Astronomy & Astrophysics, and Aerospace Technology); this cluster suggests the intimate relationships between basic physical science and engineering. The *Societal* cluster (top) is formed by human-centric disciplines that are focused on improving societal welfare, including Medical Sciences (e.g., Psychiatry, Nursing, and Cancer) as well as Social Sciences and Arts & Humanities (e.g., Education, Sociology, and International Relations).

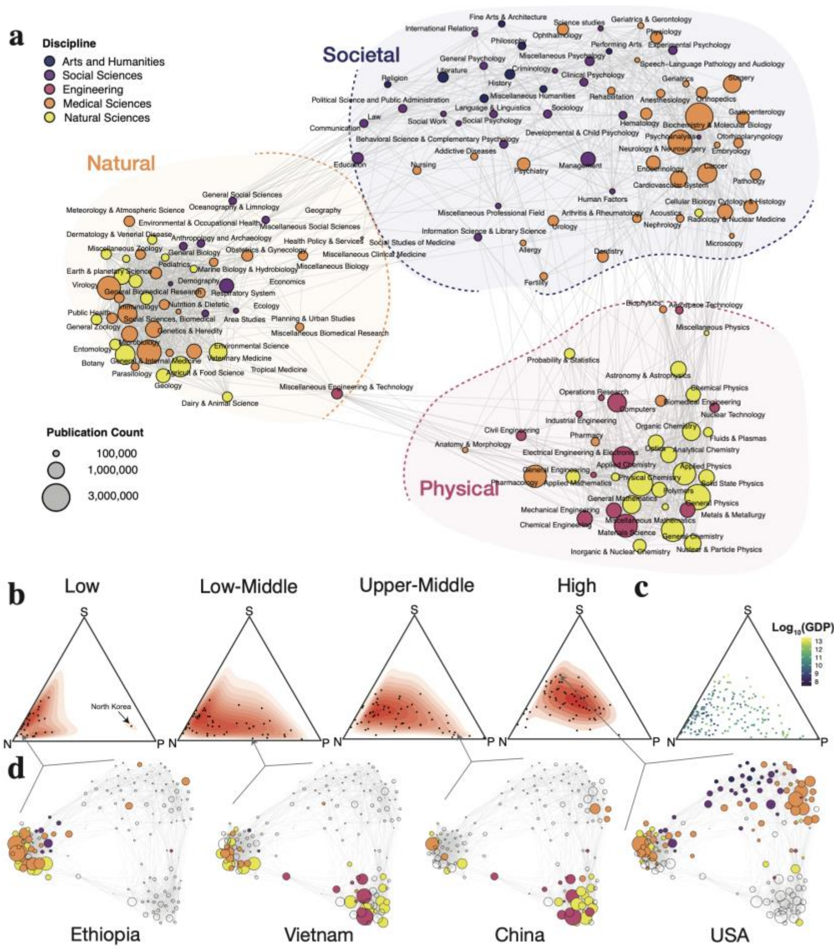


***Fig 2. The Structure of the disciplinary proximity network and national development***. *(a) The backbone of the disciplinary relatedness network reveals three clusters, which we call Natural, Physical, and Societal. Each node corresponds to a discipline and the weight of an edge captures the minimum conditional probability of co-specialization (see Data & Methods). The area of a node is proportional to the number of total publications indexed in that discipline. Node color maps to five broad disciplinary categories. (b) Nations are classified into four groups by their income level: Low, Low-Middle, Upper-Middle and High (from left to right). Dots correspond to nations, and a nation's position inside the simplex is calculated as the fraction of advantaged disciplines in each cluster normalized by its total number of advantaged disciplines. The density estimate of each income group is shown in red. (c) National research profile snapshots (2013-2017) and GDP. Points are colored according to the nation's log-transformed GDP. (d) Four example countries, Ethiopia, Vietnam, China, and the United States (USA) at 2013-2017. Only the discipline with an advantage (*$\log_{10}\mathrm{RCA} > 0$*) are colored. Node colors are the same as in (a).*

These clusters offer a concise representation of each country's research portfolio. Namely, each country's scientific portfolio can be represented as a point in the simplex of the three clusters (see Fig. 2b; see Data & Methods). Aggregating countries based on their income-level classification[32] reveals that niches are largely related with national wealth (see Fig. 2b-d). Low-income countries (e.g., Afghanistan, Ethiopia, and South Sudan) tend to be confined to the *Natural* cluster; some of the low-middle countries extend towards the *Physical* disciplines whereas upper-middle income countries are located closer to the center. High-income countries (e.g., United States, France, and Japan) tend to occupy the center and the space between *Natural* and *Societal*, suggesting balanced exportation. This pattern suggests there might be a universal tendency that as a nation's economic power increases, their scientific exports move towards a more balanced portfolio.

To understand the temporal evolution of national research portfolios, we first examine whether the development of a new revealed comparative advantage (or the loss) follows the "*law of proximity*"[13], which predicts that countries are more likely to develop a new advantage in a discipline that are close to their existing advantages (see Fig. 2). By examining the (de)activation of advantages across each subsequent time steps (see Data & Methods), we show the law of proximity indeed holds (see Fig. 3a-b). The probability of a new activation increases with the density of proximate specialized disciplines; the probability of deactivation follows the opposite pattern.

Moreover, if we disaggregate countries based on income groups, we further discover that low-income countries are more strongly constrained by the law of proximity than others (See Fig 3c-d). In other words, it is more difficult for low-income countries to develop a new relative advantage if it is not in the vicinity of already existing advantage, while wealthy countries are more flexible to develop new advantage, likely related with the already existing diverse and complex research portfolio with broader and higher disciplinary production, allowing them to make less portfolio-dependent choices when building scientific capacity.

The law of proximity shapes scientific development, but to what extent? We compare the actual trajectories with a null model that is solely based on the law of proximity (see Data & Methods). As shown in Fig 3e, the predicted research profiles converge towards the center of the simplex; in other words, even with the constraining effect of the law of proximity, the connections across clusters are strong enough to attract countries towards a balanced research portfolio. By contrast, the aggregated actual trajectories display much weaker attraction towards the center, suggesting scientific development is not entirely dictated by the law of proximity but may also be conditioned by the three clusters (see Fig 3e and SI). The difference is particularly stark for countries specialized in *Natural* cluster, suggesting that low-income countries may face a heavy hurdle breaking into other disciplines.

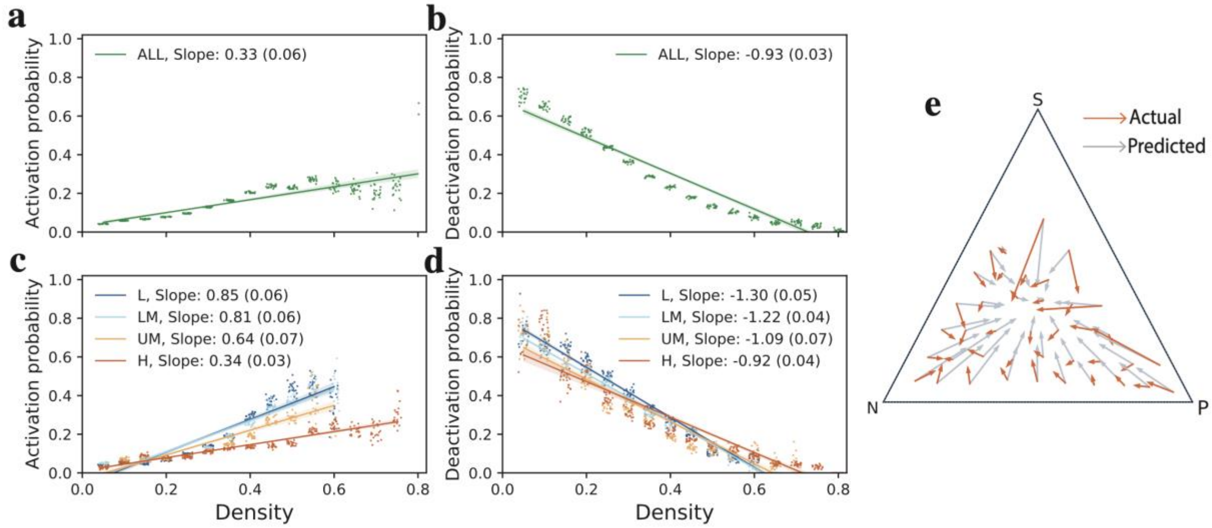


***Fig 3. Disciplinary proximity dictates the development and loss of competencies*** *(a) Probability of a new relative advantage in the next time period given the density of existing advantages surrounding the discipline. Dots represent the estimated probabilities from Bootstrapping and solid lines are the estimated regression line from bootstrapped samples. (b) Probability that an advantaged discipline will lose its advantage in the next period given the density of existing advantages. (c)-(d) Same plots where countries are grouped based on their income class. (e) We show the predicted and actual evolution in the simplex. Arrows point to the average simplex position of countries in the next period. Red arrows represent the empirical movement while gray arrows represent the movement predicted from the null model based only on the law of proximity.*

This "trapping" of low-income countries can also be observed in the patterns of specialization in global science. It has been observed that the global economy exhibits a hierarchical[22,33,34] (or *nested)* structure, where rich countries can export a wide range of products—especially those that exported by only a few countries—whereas poor countries can only export a small number of products that can be exported by many[22,35,36]. This pattern contrasts a more classical theory of specialization, where countries specialize and form a 'modular' structure. Inspired by these two ideas, we measure the nestedness and modularity of the scientific exports over time (see Data & Methods and Fig. 4). In contrast to the case of economic products, we do not observe strong evidence of nestedness; instead, we find that the modularity of the network has been increasing, which suggests countries are increasingly specialized in one of the three clusters. This rise of modularity and specialization in the global science may indicate widening gaps for low-income countries to develop scientific disciplines in Physical and Societal clusters.

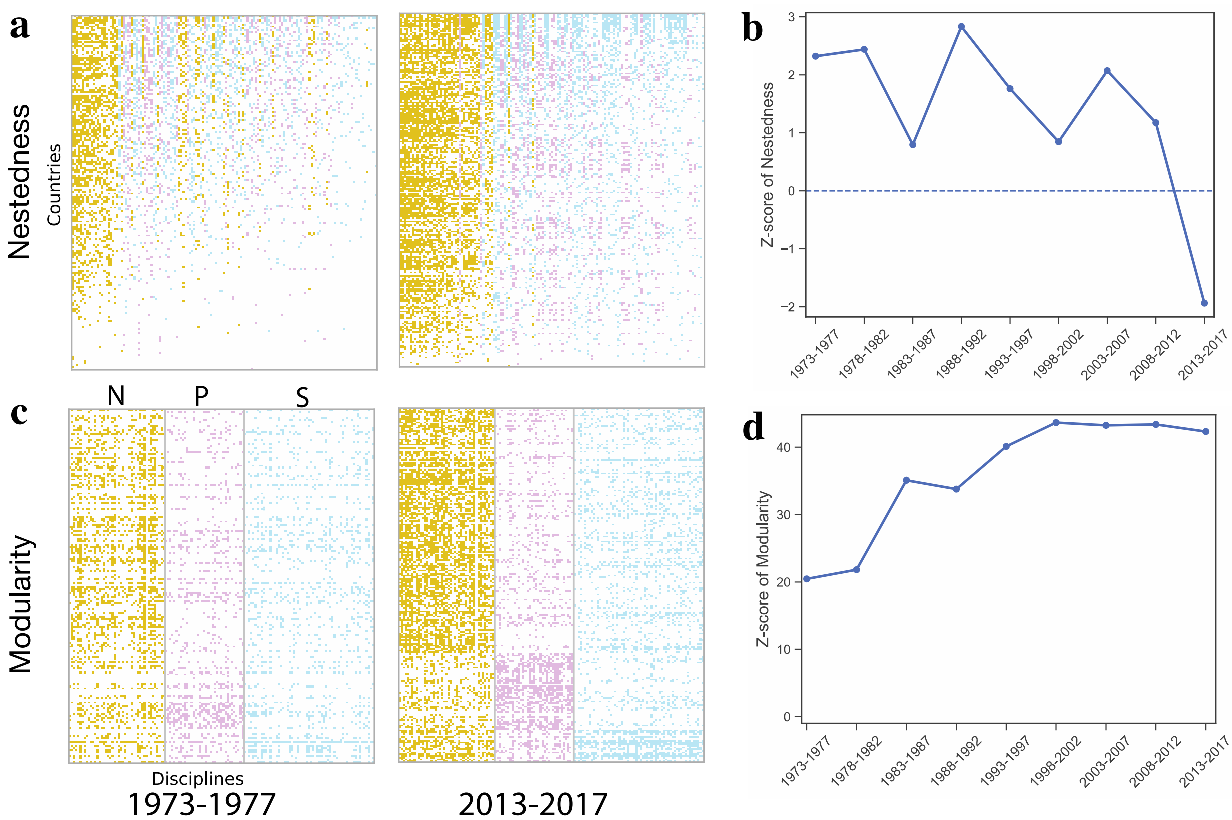


***Fig 4. Nestedness and modularity of global science*** *(a) The country-discipline RCA matrices of the earliest and the most recent periods where rows and columns are arranged in a descending order of number of advantages. (b) The z-score of nestedness over time which is calculated through Fixed-Fixed null model. (c) The country-discipline RCA matrices of the earliest and the most recent time periods where disciplines(columns) are arranged by its classification into the three clusters. (d) The z-score of modularity over time which shares the same null matrix as used for calculating nestedness.*

Although we observe the lack of nested structure and increasing modularity over time, the wealthiest nations still tend to have the most diverse portfolios of scientific exports (see Fig. 2b-c), which raises the question about the connection between scientific diversity and economic prosperity. To address this question, we study the association between the diversity of a scientific portfolio and economic wealth by measuring the diversity of a nation's scientific enterprises with the Gini index of disciplinary RCA values (see Data & Methods): For convenience, we define the *scientific diversity* of a nation as one minus their Gini index. High scientific diversity corresponds to a more balanced and diversified portfolio, whereas low scientific diversity indicates more skewed and specialized exportation. The nations of the industrialized West (e.g., USA, Australia and France) tend to have the most diversified research portfolios (Fig. 5a). By contrast, nations across the Global South tend to have highly unbalanced portfolios. The rapidly industrializing BRIC nations, Brazil, Russia, India, and China, fall between these two extremes.

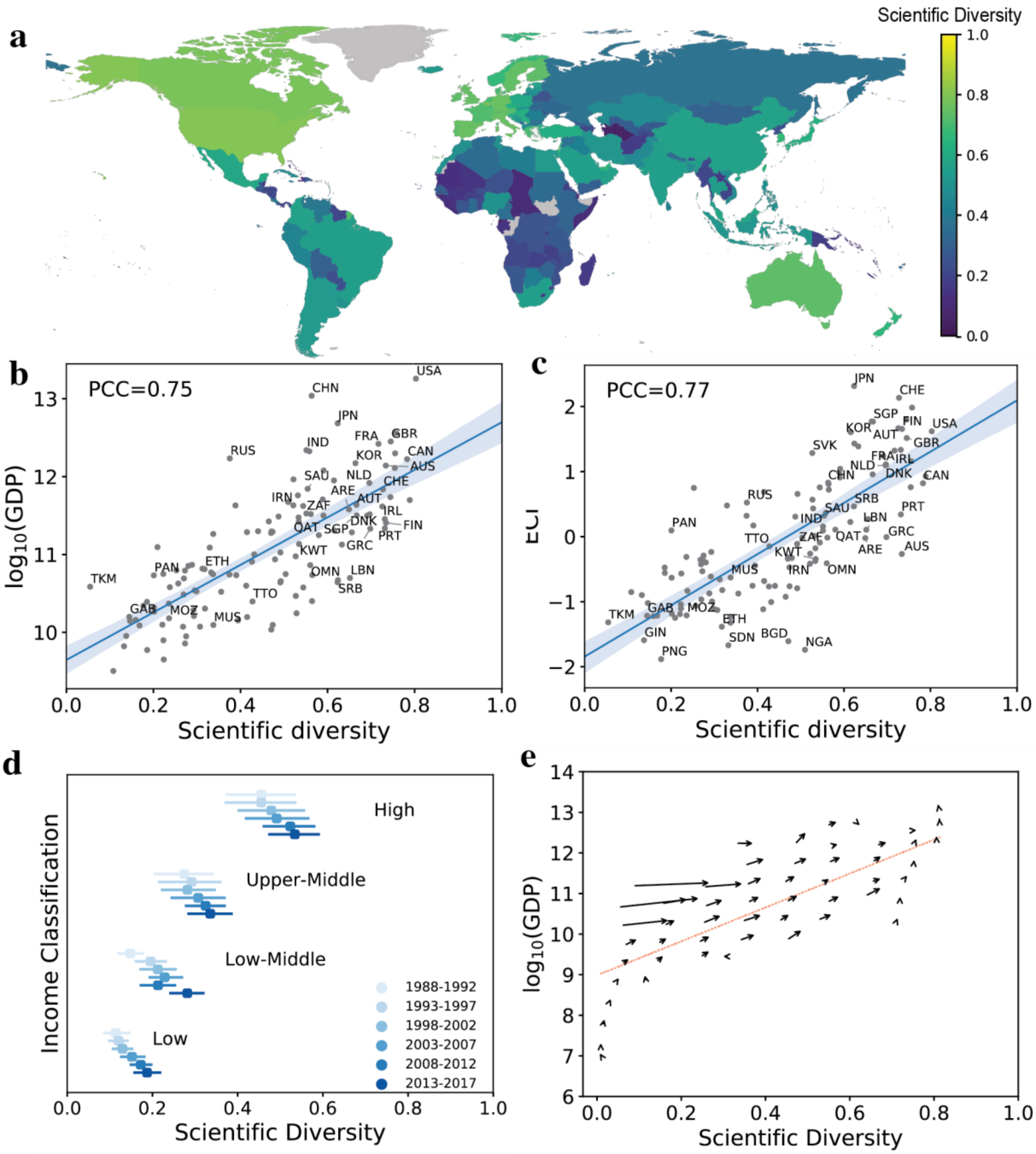


***Fig5. Scientific diversity is correlated with national development indicators.*** *(a) Map of scientific diversity (defined as one minus the GINI index of the RCA values of a country). High diversity (brighter color) indicates a more even distribution of RCA values. Grey corresponds to a lack of data for that country. (b-c) The relationship between scientific diversity and nation's GDP (b) and their economic complexity index (c). (d) The temporal development of the diversity of scientific*

*production by income group. Error bars represent the confidence interval drawn from bootstrapping. (e) The developmental trajectory of nations in the scientific diversity-income (national GDP) plane. While the two are positively correlated, low-income countries grow slowly in both science and economy.*

The gradient of scientific diversity closely reflects national wealth. In fact, national scientific diversity strongly correlates with national GDP (Fig. 5b), as well as the Economic Complexity Indicator (ECI)[35], which captures the complexity of economic systems and is used to predict future GDP growth (Fig. 5c). We also observe that over the past 40 years the scientific diversity of nations has been steadily increasing across all income groups (Fig. 5d). This global diversification of science cannot be explained by the increase in number of publications alone (see Fig. S9). However, this steady diversification is not enough to close the gap between income groups, with the diversity of high-income countries doubling or even tripling that of low-income countries in the most recent time period.

To examine the interplay between income and scientific diversity, we project the development trajectory of scientific diversity onto the income-diversity plane. Overall, they are correlated, developing together. There are, however, notable deviations. First, countries with high scientific diversity and high income tend to develop slowly in both. Second, countries with weak economic strength and low scientific diversity display little progress in both, suggesting a potential scientific development trap. There exist a group of countries that rapidly increased their scientific diversity (e.g., Republic of Serbia, Czech Republic, Slovakia, Slovenia, and China; See Fig. 5e). This surge in development may stem from data issues following their independence (see Fig. S10). We also observe that China makes tremendous progress in diversifying their research portfolio.

Finally, we show that scientific diversity is more predictive of future economic growth than Economic Complexity Index after controlling for the heterogenous effect across countries (see Table 1; see Data & Methods and SI for other models). Under the same model used to show the predictive power of ECI[35], scientific diversity has better predictive power than ECI in predicting future economic growth over the past 45 years. Every 0.1 increase in scientific diversity is associated with about 6% growth in national GDP, although we do not claim a direct causal relationship. The predictive power of scientific diversity is robust across the number of countries and time periods included in the analysis, except in the two-way fixed effects model where linear additive effects of both country and time heterogeneity are assumed[37] (see Table S4-6). In contrast to the belief that developing industry-relevant disciplines (*Physical* cluster) contributes to countries' economic growth, we observe no statistically significant relationship (see Table 1. And SI) between the number of new activated disciplines across clusters with economic growth. The regression results add further evidence for the existence of a scientific developmental trap faced by low-income countries. Lacking scientific diversity, the scientific development of low-income countries is more path-dependent and constrained, potentially hindering future scientific and economic development.

*Table 1. The regression results of the country fixed effect 5-year GDP growth panel linear*

*regression models.*

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | GDP growth (log-ratio) | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| GDP | -0.098$^{***}$ | -0.151$^{***}$ | -0.144$^{***}$ | -0.113$^{***}$ | -0.150$^{***}$ |
| | (-0.127, -0.070) | (-0.187, -0.115) | (-0.191, -0.097) | (-0.137, -0.090) | (-0.198, -0.103) |
| ECI | 0.012 | | 0.008 | | 0.009 |
| | (-0.029, 0.052) | | (-0.032, 0.049) | | (-0.032, 0.050) |
| Diversity | | 0.266$^{***}$ | 0.262$^{**}$ | | 0.274$^{**}$ |
| | | (0.093, 0.440) | (0.049, 0.474) | | (0.053, 0.494) |
| Natural | | | | 0.0003 | -0.001 |
| | | | | (-0.002, 0.003) | (-0.004, 0.002) |
| Physical | | | | -0.001 | 0.0002 |
| | | | | (-0.005, 0.002) | (-0.004, 0.004) |
| Societal | | | | 0.002 | 0.002 |
| | | | | (-0.001, 0.005) | (-0.001, 0.005) |
| Observations | 642 | 936 | 642 | 936 | 642 |
| $R^2$ | 0.077 | 0.104 | 0.087 | 0.098 | 0.091 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | | | |

## Conclusion

It is widely believed that scientific development holds the key for a nation's future prosperity[1,38]. Yet, whether there are universal patterns of scientific development at the national level has been an open question. By analyzing over 30 million scientific publications across 217 countries spanning the period 1973-2017, we provide a large-scale temporal analysis of national science development. We find that the disciplinary proximity network constructed from these publications exhibits three clusters of disciplines which roughly capture the relative advantages of countries across the spectrum of economic wealth. Although each country's position in the network is shaped by various historical, geographical, social, and economic factors, the three-cluster structure still conditions their scientific development.

Our results are reminiscent of classical theories. The clusters and the niches that are occupied by nations show some semblance of Comte's "Hierarchy of the Sciences" (1855) hypothesis—that science progresses from natural sciences that require readily-available, simple subjects, towards social sciences that deal with more complex subjects. At the same time, the prominence of *Natural* disciplines in low-income countries and less resource-driven fields in the *Physical* and *Societal* clusters resonates with Basalla's "Spread of Western Science" (Basalla, n.d.) whereby post-colonial nations transition from colonial exploitation towards independence.

This study is subject to limitations. First, our study relies on a bibliographic database created and maintained by a western scientific enterprise. Therefore, it overestimates the research from western countries and the publications in English while underestimating the production in other nations and languages (see SI Data). Still, we argue that our operationalization is reasonable under the analogy to products exportation[22,35] and the status of English as the de-facto lingua franca[39] of science. Second, many analyses considered the RCA matrix as a bipartite network. This approximation is not strictly valid because the edges are not independent from each other. Finally, we also note that reliable causal inference with country-level data is often infeasible and our results do not necessarily imply a direct causal relationship between national scientific development (diversity) and economic growth. A more likely explanation would be that there exist unobserved hidden confounders or complex feedback mechanisms between scientific and economic development.

Even with these limitations, our empirical framework may provide a useful perspective to study the structure and evolution of national scientific portfolios and their relationships to economic development. Our results call for attention to the barriers faced by low-income countries in building their scientific capacity, and the potential consequences on future scientific capacity and economic growth. We hope our analysis opens a new avenue towards the understanding of the mechanisms of scientific development as well as its relationship to economic prosperity.

## Data & Methods

**Data.** The dataset was drawn from the Clarivate Analytics' Web of Science database hosted and managed by the *Observatoire des Sciences et des Technologies* at the University of Montreal. The Web of Science database contains three main citation indices: The Science Citation Index Expanded, the Social Science Citation Index, and the Arts and Humanities Citation Index. We used all indexed publication records listed as being published between 1973 to 2017, which included 37,479,532 papers published across 20,252 scholarly journals. To examine temporal patterns, we split the data into nine five-year snapshots. We limited this set to only journal articles, review articles, and notes (discontinued in 1991 but included in articles). We also excluded any publication that did not list any institutional address, and publications that could not be assigned a disciplinary category according to the steps below. After these filters, the dataset contained 35,793,320 papers published across 20,123 scholarly journals (See Fig. S2).

Discipline classification of publications is based on the National Science Foundation typology of journals, which categorizes papers into a hierarchy of disciplines. The high-level and granular classification was further complemented with an in-house classification of the Arts and Humanities[40]. The resulting classification scheme contains 144 granular categories. After removing “Unknown” from the 144 granular categories, we manually classified each of the 143 categories into one of five broad categories: “Natural Science”, “Medical Science”, “Engineering”, “Social Science”, and “Arts and Humanities”; this scheme is used to color nodes in Figure 2.

Publications are associated with nations using the institutional addresses listed by the authors. We assign a full unit credit of a publication to every country of affiliation represented on the paper’s author byline (“full counting”). For example, a paper listing five authors—two with affiliations in the United States, two in Canada, and one in the Netherlands—would count as one paper to all three countries. Full counting method assumes each author contributes equally to the publication. Different counting methods are highly correlated at the macro level[41]. See SI data section for more details.

We use data on national GDP from the World Bank[32,42] to approximate the economic wealth of each country. The dataset covers 264 countries from 1960 to 2019. Income classification comes from World bank database[32] which contains 224 countries between 1987 and 2018. We convert the annual classification to a time snapshot classification by assigning each country to its most frequent income group during each period. See SI data section for more details.

**Revealed Comparative Advantage.** The revealed comparative advantage (RCA) of country $c$ in discipline $i$ is defined as:

$$\mathrm{RCA}_{c,i} = \frac{\mathcal{P}(c,i)/\sum_i \mathcal{P}\,(c,i)}{\sum_c \mathcal{P}\,(c,i)/\sum_{c,i} \mathcal{P}\,(c,i)}$$

where $\mathcal{P}(c,i)$ is the number of publications produced and "exported"—the number of publications which is indexed in the Web of Science—by country $c$ in discipline $i$, $\sum_i \mathcal{P}\,(c,i)$ is the total number of publications produced by country $c$, $\sum_c \mathcal{P}\,(c,i)$ is the total number of publications produced in a discipline globally, and $\sum_{c,i} \mathcal{P}\,(c,i)$ is the total number of publications across all countries and disciplines.

**Disciplinary Proximity.** The proximity between disciplines $i$ and $j$ is defined as the minimum of the pairwise conditional probabilities of a country having an advantage ($\mathrm{RCA} > 1$) in one discipline given an advantage in another:

$$\phi_{ij} = \min\{P(\mathrm{RCA}_i > 1|\mathrm{RCA}_j > 1), P(\mathrm{RCA}_j > 1|\mathrm{RCA}_i > 1)\}$$

$\phi$ is a $143 \times 143$ matrix that captures the proximity between pairs of disciplines (see SI Fig. S4).

**Identifying the disciplinary clusters.** The relatedness network is constructed from the disciplinary proximity matrix derived from aggregating data across all years (from 1973 to

2017). The network is fixed over the analysis. Although the network structure changes over time, networks derived from a snapshot of data closely resemble with the aggregated network (see SI Fig. S5). The multi-scale backbone extraction method[29] exposes three visual clusters when laid out with a force-directed layout algorithm (Gephi's ForceAtlas2). We formally confirm the structure by applying a community detection algorithm (the Leiden algorithm[30]) to the full network. We use modularity as the quality function and ran the model 50 times to obtain consensus. Other methods produce similar results although some partitions the network into smaller communities (see SI Disciplinary Relatedness Network section).

**Position within the simplex.** Position within the simplex measure countries' cluster-level specialization concentration. We first calculate $C_i = n_i / N_i$, where $n_i$ is the number of disciplines in cluster $i$ with RCA > 1, $N_i$ is the total number of disciplines in cluster $i$. Then we normalize $C_i$ so that $\sum_i C_i = 1$.

**The Density of existing advantages and the null development model.** The density of existing advantages around a given discipline is defined as follows:

$$\omega_j^k = \frac{\sum_i x_i \phi_{ij}}{\sum_i \phi_{ij}}$$

where $\phi_{ij}$ is the proximity between discipline $i$ and $j$, and $x_i = 1$ if $RCA_{ki} > 1$ else $x_i = 0$, and the density of existing advantages, $\omega_j^k$, is the proximity-weighted sum of all disciplines that are connected to $j$ with $RCA_{ki} > 1$. We bin the density values and aggregate across countries and time periods to calculate the probability of activation and deactivation, given the density. We also perform a bootstrap sampling with 20 samples to estimate the uncertainty of the slope and report the mean and standard deviation of the slopes across bootstrapped samples. A linear regression model (OLS) is fit by pooling all bootstrap samples to obtain the parameters (intercept and slope) for the null model. The null model works as following: for every inactive (RCA < 1) discipline, we assign a probability the discipline will be activated (RCA > 1) in the subsequent time period based on its current density using the intercept and slope obtained from the pooled regression model that include all countries. We use the same procedure for the deactivation. For each time period and each country, the newly activated and deactivated disciplines are sampled using the null model while preserving the number of new activation and deactivation in the next time period. We repeat this procedure 100 times. When visualizing the actual profile and the predicted profile on the simplex, to reduce the influence of extreme cases, we remove data points located on the boundary of the simplex. To smooth out the noise, we aggregate data points within each rhombus with the side length of 0.1 that tessellates across the simplex. We observe that the difference between actual trajectory and the predicted trajectory is robust against the direction of rhombus.

**Modularity and Nestedness.** We use the country-discipline bipartite network to represent knowledge exportation. Country $c$ is connected to discipline $i$ if $RCA_{c,i} > 1$. Modularity[43] of the country-discipline bipartite network is defined as:

$$Q = \frac{1}{m}\sum_{i=1}^{p}\sum_{j=1}^{q}(A_{ij} - P_{ij})\delta(g_i, h_j)$$

Where $m$ is the number of links, $A_{ij}$ equals to 1 if there is a link from node $i$ to node $j$, $P_{ij}$ is the probability the edge between $i$ and $j$ exists under the null model, $g_i$ and $h_j$ are communities that the country and discipline belong to. The community of a country is decided by its largest cluster level revealed comparative advantage; for example, China is classified to *Physical* cluster since it has highest cluster-level RCA value in *Physical* cluster. The community of disciplines is defined by the Leiden algorithm. Although the elements of the RCA matrix are not strictly independent from each other, we use $P_{ij} = \frac{k_i d_j}{m}$ (where $k_i$ and $d_j$ are the degree of node $i$ and $j$ respectively) as an approximation. Larger modularity means countries tend to be specialized in one of the three clusters rather than having advantages spread across multiple clusters.

Nestedness is measured by the overlap and decreasing fill (NODF)[44] method. NODF measures the degree of overlapping between row pairs and column pairs in the adjacency matrix. The metric is defined as

$$NODF = \frac{\sum N_{paired}}{\left[\frac{n(n-1)}{2}\right] + \left[\frac{m(m-1)}{2}\right]}$$

Where $\sum N_{paired}$ is the averaged degree of nestedness for each pair of row and column based on the principles of decreasing fill and paired overlap[44], $n$ and $m$ are the number of rows and columns.

We use a null model to test whether modularity and nestedness are significant. We construct the null model of the bipartite network by swapping edges between node pairs while constraining the degree of each node which we refer to as the Fixed-Fixed null model.

**Scientific Diversity.** The Gini index of a nation's RCA values across disciplines is used to capture the scientific diversity of a nation. For convenience, we use 1 minus the Gini index as a measure of scientific diversity. If all disciplines have the same RCA value in the country, the diversity value would be 1. If a country only produces scientific publications in one discipline, then the diversity value would be 0. To investigate the dynamic relationship between scientific diversity and economic power, we project countries evolution into the diversity-GDP plane. To smooth out noise, we averaged the trajectory in each grid with width equals to 0.1 and height equals to 0.5. The starting point of arrow represents average of all displacements whose starting points were in the grid. The direction and length of arrows are computed by averaging the subsequent displacements of all countries within a grid.

**Predicting Economic Growth.** We use country fixed-effect panel regression models to investigate the relationship between countries' economic growth and scientific development as following:

$$\log\left(\frac{GDP_{c,t+1}}{GDP_{c,t}}\right) = \beta_0 + \beta_1 \log\left(GDP_{c,t}\right) + \beta_2 ECI_{c,t} + \beta_3 Diversity_{c,t} + \beta_2 T_{N,(t,t+1)} + \beta_3 T_{A,(t,t+1)} + \beta_4 T_{S,(t,t+1)} + \alpha_c$$

Where $c$ denotes countries, $t$ denotes time periods, $GDP_{c,t}$ is the averaged GDP value of country $c$ during time period $t$, $Diversity_{c,t}$ is the scientific diversity value of country $c$ during time period $t$, $T_N$, $T_A$ and $T_s$ are the number of increased advantage disciplines in *Natural*, *Physical* and *Societal* clusters between *t* and *t*+1, $\alpha_c$ is the country-specific intercept that capture the heterogeneity across countries. Countries with incomplete temporal data are excluded in the analysis. However, the performance of scientific diversity is robust even if we use imbalanced panels (see SI Table S5).

**Data availability.** Data will be available at https://figshare.com/account/projects/96980/articles/13623035.

**Code availability.** The code used for data processing and analysis will be available https://github.com/yy/national-science-exports.

## References


1. Krammer, S. M. S. Science, technology, and innovation for economic competitiveness: The role of smart specialization in less-developed countries. *Technol. Forecast. Soc. Change* **123**, 95–107 (2017).
2. Wagner, C. S. *The New Invisible College: Science for Development*. (Brookings Institution Press, 2009).
3. Rosenberg, N. Science, Invention and Economic Growth. *Econ. J.* **84**, 90–108 (1974).
4. Salter, A. J. & Martin, B. R. The economic benefits of publicly funded basic research: a critical review. *Res. Policy* **30**, 509–532 (2001).
5. Kumar, R. R., Stauvermann, P. J. & Patel, A. Exploring the link between research and economic growth: an empirical study of China and USA. *Qual. Quant.* **50**, 1073–1091 (2016).
6. Adam, E. 'Governments base performance-based funding on global rankings indicators': A global trend in higher education finance or a global rankings literature fiction? A comparative analysis of four performance-based funding programs. *Int. J. Educ. Dev.* **76**, 102197 (2020).
7. Demeter, M. The World-Systemic Dynamics of Knowledge Production: The Distribution of Transnational Academic Capital in the Social Sciences. *J. World-Syst. Res.* **25**, 111–144 (2019).
8. Schofer, E. Cross-national Differences in the Expansion of Science, 1970-1990. *Soc. Forces* **83**, 215–248 (2004).
9. Kozlowski, J., Radosevic, S. & Ircha, D. History matters: The inherited disciplinary structure of the post-communist science in countries of central and eastern Europe and its restructuring. *Scientometrics* **45**, 137–166

(1999).

10. Livingstone, D. N. *Putting Science in Its Place: Geographies of Scientific Knowledge*. (University of Chicago Press, 2010).
11. Comte, A. *The Positive Philosophy of Auguste Comte*. (C. Blanchard, 1855).
12. Basalla, G. The Spread of Western Science. *Science* **156**, 611–622 (1967).
13. Hidalgo, C. A. *et al.* The Principle of Relatedness. 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). doi:10.1007/978-3-319-96661-8_46.
14. Paula, S. *How Economics Shapes Science*. (Harvard University Press, 2012).
15. Hornyak, T. Chilean research grows despite poor investment. (2016).
16. Bronfman, L. A panorama of Chilean astronomy. *The Messenger* **107**, 14–18 (2002).
17. Bajak, A. Chile's chance to embrace science for the twenty-first century. *Nature* **552**, S53–S55 (2017).
18. Yeom, H. W. South Korean science needs restructuring. *Nature* **558**, 511–513 (2018).
19. National Science Board. *Science and Engineering Indicators 2018*. (2018).
20. Noorden, R. V. Science in East Asia — by the numbers. *Nature* **558**, 500–501 (2018).
21. National Science Board. *Publications Output: U.S. Trends and International Comparisons*. (2019).
22. Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. The Product Space Conditions the Development of Nations. *Science* **317**, 482–487 (2007).
23. King, D. A. The scientific impact of nations. *Nature* **430**, 311–316 (2004).
24. May, R. M. The Scientific Wealth of Nations. *Science* **275**, 793–796 (1997).
25. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
26. Evans, J. A., Shim, J.-M. & Ioannidis, J. P. A. Attention to Local Health Burden and the Global Disparity of Health Research. *PLoS ONE* **9**, e90147 (2014).
27. World Health Organization. *World health statistics 2018: monitoring health for the SDGs*. (2018).
28. Zhou, P. & Leydesdorff, L. The emergence of China as a leading nation in science. *Res. Policy* **35**, 83–104 (2006).

29. Serrano, M. A., Boguna, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.* **106**, 6483–6488 (2009).

30. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

31. Moya-Anegón, F. & Herrero-Solana, V. Worldwide Topology of the Scientific Subject Profile: A Macro Approach in the Country Level. *PLoS ONE* **8**, e83222 (2013).

32. Fantom, N. & Serajuddin, U. *The World Bank's Classification of Countries by Income*. (The World Bank, 2016).

33. Bustos, S., Gomez, C., Hausmann, R. & Hidalgo, C. A. The Dynamics of Nestedness Predicts the Evolution of Industrial Ecosystems. *PLoS ONE* **7**, e49393 (2012).

34. Gao, J., Zhang, Y.-C. & Zhou, T. Computational Socioeconomics. *Phys. Rep.* **817**, 1–104 (2019).

35. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl. Acad. Sci.* **106**, 10570–10575 (2009).

36. Morrison, G. *et al.* On Economic Complexity and the Fitness of Nations. *Sci. Rep.* **7**, 15332 (2017).

37. Imai, K. & Kim, I. S. On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. *Polit. Anal.* 1–11 (2020).

38. Powell, W. W. & Snellman, K. The Knowledge Economy. *Annu. Rev. Sociol.* **30**, 199–220 (2004).

39. Gordin, M. D. *Scientific Babel: How science was done before and after global English*. (University of Chicago Press, 2015).

40. Éric, A., Étienne, V.-G., Grégoire, C., Vincent, L. & Gingrasb, Y. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics* **68**, 329–342 (2006).

41. Sugimoto, C. R. & Vincent, L. *Measuring research: What everyone needs to know*. (Oxford University Press, 2018).

42. World Bank. World Development Indicators. (2019). GDP (current US$). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=1W.

43. Barber, M. J. Modularity and community detection in bipartite networks. *Phys. Rev. E* **76**, 066102 (2007).

44. Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227–1239

(2008).

## Acknowledgements

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-19-1-0391. We thank Staša Milojević, Jisung Yoon, Sadamori Kojaku, Byungkyu Lee, Bruce Weinberg, Paula Stephan, Santiago Schnell, Tao Zhuo, Junming Huang, Jian Gao, Yi Bu, and Nima Dehmamy for helpful discussion and comments.

## Author Contributions

L.M. and D.M. conceived the study; All authors contributed to the design of the study; V.L. prepared the primary datasets; L.M., D.M., V.L., Y.Y.A performed analysis; All authors contributed to the interpretation of the results and writing of the manuscript.