nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-022-01367-x

Check for updates

The latent structure of global scientific development

Lili Miao[1], Dakota Murray[1], Woo-Sung Jung[2,3], Vincent Larivière[4,5,6], Cassidy R. Sugimoto[7] and Yong-Yeol Ahn[1,8,9] ✉

Science is essential to innovation and economic prosperity. Although studies have shown that national scientific development is affected by geographic, historic and economic factors, it remains unclear whether there are universal structures and trajectories of national scientific development that can inform forecasting and policy-making. Here, by examining the scientific 'exports'—publications that are indexed in international databases—of countries, 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 research profiles individually, scientific production is increasingly specialized in global science over the past decades. By uncovering the underlying structure of scientific development and connecting it with economic development, our results may offer a new perspective on the evolution of global science.

Science has experienced rapid transformation due to increasing scientific capacity across countries. During the 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 power dynamics in science. Whereas China only accounted for 5% of scientific publications in international indexes in 2000, it became the most productive country in the world by 2018, surpassing US scientific production[1–3]. The increase in scientific capacity was also coupled with Asia's economic acceleration: for example, the rapid expansion and intense industrialization of the 'Four Asian Tigers'— Hong Kong, Singapore, South Korea and Taiwan—occurred during this time[4,5]. These rapid transformations provide an opportunity to examine the relationship between economic and scientific development and to test theories of universality in this relationship.

Studies have examined how the interplay between geography[6,7], history[8], existing scientific strengths[9–12] and economic conditions[13–15] influence scientific development. Chile, for example, exemplifies the influence of geographical opportunities on national knowledge production: despite relatively low scientific investments[16], 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 in the field[17,18]. By contrast, South Korea, with its heavy investment in science and technology[19,20], has experienced diversified scientific expansion, developing into a science and innovation powerhouse[4]. Institutional organization and investment are also potential factors. For example, May[21] compared research organizations in France and Germany—where research institutes such as the CNRS and Max-Planck play a central role—to the United States and the United Kingdom—which centralizes basic research in universities and engages students—and concluded that the former structure negatively affects research activity.

In contrast to localized explanations of national scientific development, several scholars have attempted to develop universal frameworks. For example, Comte argued that science develops along a natural trajectory from high-consensus physical sciences towards more complex, low-consensus social sciences[22]. Basalla took a colonial perspective, arguing that scientific development of non-Western countries generally undergoes three phases: countries first provide resources for Western scientists, then transition to a replication model—in which science develops following the institutions and traditions from scientifically established nations—and culminate with scientific independence, often obtained with mixed success[23]. In Basalla's model, phases of development also affect the research specialization of nations. For instance, in the first phase, disciplines are descriptive in nature and strongly tied to natural resources and exploration, the second phase has a stronger focus on experimental domains. Despite acknowledging that his model was an 'heuristic device' and that the environment in which research is conducted should be taken into account[24,25], Basalla's model remained criticized for being Eurocentric and insensitive to cultural factors[25].

Despite criticisms, some alignment between these theories and the actual phases of development has been observed. Moya-Anegón and Herrero-Solana[26] classified countries into three groups on the basis of their research specialization and showed that countries with high gross domestic product (GDP) specialize in biomedicine, formerly communist countries specialize in basic science and engineering and less developed countries specialize in agriculture. Cimini et al.[27] arrived at a different conclusion, showing that—rather than

[1]Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA. [2]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea. [3]Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea. [4]École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, Québec, Canada. [5]Observatoire des sciences et des technologies, Université du Québec à Montréal, Montréal, Québec, Canada. [6]Department of Science and Innovation—National Research Foundation Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa. [7]School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA. [8]Network Science Institute, Indiana University, Bloomington, IN, USA. [9]Connection Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: yyahn@iu.edu

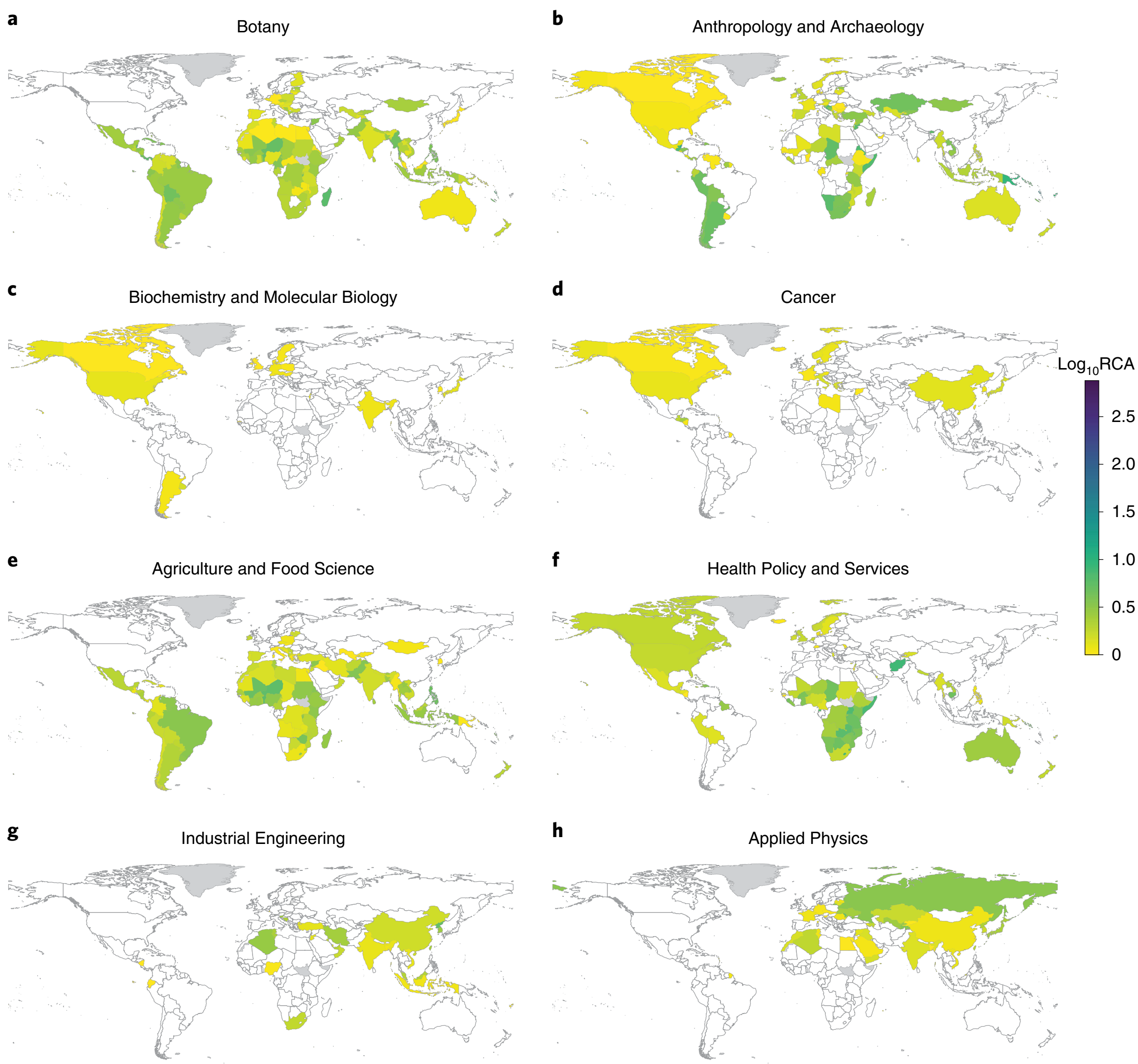

Fig. 1 | Disciplinary specialization reflects geographical, historical and economic factors. a–h, Eight examples illustrate the distribution of disciplinary specializations. Discipline specialization is measured by RCA. Colour represents the logarithm of RCA; a nation is only coloured if $\log_{10} \mathrm{RCA}_{c,i} > 0$. Grey corresponds to nations that were not represented in our dataset. Botany and Anthropology and Archaeology reflect the presence and access to natural and anthropological resources in a country. Economic inequality underpins specialization in resource-intensive disciplines like Biochemistry and Molecular Biology and Cancer. Local issues also drive research, as can be seen from the distribution of Agricultural and Food Science and Health Policy and Services. The distribution of Industrial Engineering and Applied Physics probably reflects national economic priorities and policies. The map data were downloaded from Natural Earth. Free vector and raster map data are at naturalearthdata.com.

specializing—technologically leading countries have been active in a diversity of scientific domains.

Economic complexity framework provides a useful lens to evaluate economic and scientific transformations from a global perspective[28]. This approach uses quantitative methods to predict and explain economic trajectories of geographic regions[29–31], often using measures of relatedness, which inform changes in specialization and explain particular outcomes based on existing specializations[28]. These methods of complexity, relatedness, and application of dimensionality reduction are well-served to examine the production and exportation of knowledge. The economic complexity approach is closely aligned with contemporary perspectives in economic geography, which focus on issues such as path dependence, lock-in and proximity[32,33]. Developments toward evolutionary economic geography[34] recognize the relationship between macro- and micro-level perspectives: in the words of Boschma and Frenken[33], how 'spatial structure of the economy emerge from and are transformed by the micro-behaviour of individual and collective agents and why and how these processes of change are themselves path- and place-dependent'.

These approaches can be used to examine the degree to which scientific development follows universal patterns conditioned on existing research specializations. For example, Boschma and colleagues[12] applied the principle of relatedness to understand the research topic evolution of cities and showed that the emergence of new research topics and the disappearance of existing topics in cities are dependent on their degree of relatedness with existing topics in cities. Guevara and colleagues[11] constructed a research space by using career trajectories of scientists and demonstrated that the research space could predict research evolution of individuals, organizations and countries. Chinazzi et al.[10] used an embedding method to predict research evolution in urban areas, providing evidence that the average knowledge density in physics is correlated with scientific and economic development in a country.

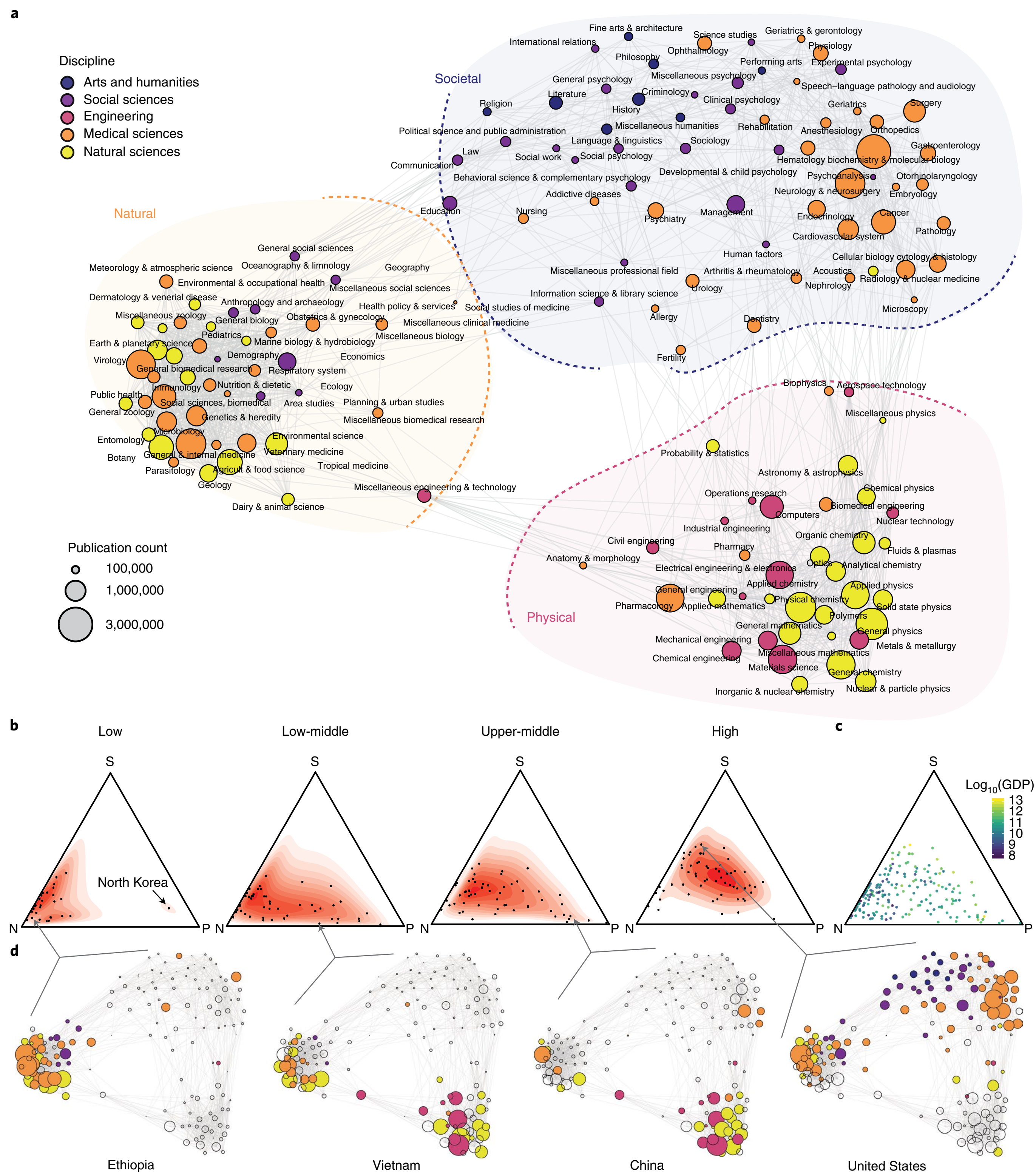

Fig. 2 | The structure of the disciplinary proximity network and national development. a, The backbone of the disciplinary relatedness network reveals three clusters, which we call Natural, Physical, and Societal. Each node corresponds to a discipline and the weight of an edge captures the minimum conditional probability of co-specialization (Methods). The area of a node is proportional to the number of total publications indexed in that discipline. Node colour 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. N, Natural; P, Physical; S, Societal. **c**, National research profile snapshots (2013–2017) and GDP. Points are coloured according to the nation's log-transformed GDP. **d**, Four example countries, Ethiopia, Vietnam, China and the United States, during 2013–2017. Only the disciplines with an advantage ($\log_{10}$RCA > 0) are coloured. Node colours are the same as in **a**.

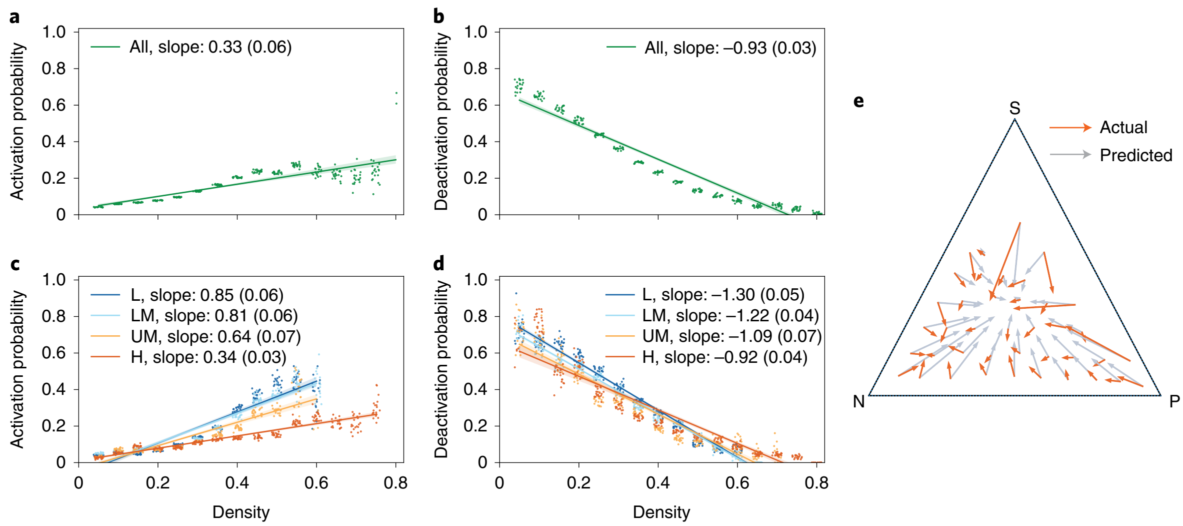

Fig. 3 | The principle of relatedness dictates the development and loss of competencies. a, Probability of a new relative advantage in the next time period given the density of existing advantages surrounding the discipline. Dots represent the estimated probabilities from bootstrapping and solid lines are the estimated regression line from bootstrapped samples. Translucent band lines and the number in the parentheses describe bootstrapped 95% confidence interval and the standard deviation of the estimated slope. The 211 countries are involved in the analysis and we performed 20 times of bootstrapping. **b**, Probability that an advantaged discipline will lose its advantage in the next period given the density of existing advantages. **c**,**d**, Same plots as **a** and **b** but where countries are grouped on the basis of their income class. L, low-income group contains 72 countries; LM, low-middle group contains 104 countries; UM, upper-middle group contains 74 countries; and H, high-income contains 56 countries. Because countries transit to different income groups during different time periods, countries are double-counted in the abovementioned numbers. **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 grey arrows represent the movement predicted from the null model based only on the principle of relatedness.

The study of economic output with respect to the 'product space'—that is, the network of relatedness between exported products—argues that the networked structure of industrial advantage is critical to understanding the economic development of nations[31]. Following the well-trodden path of Adam Smith, economists argue that division of labour—specialization—is related to economic efficiency; therefore, development is associated with increased capacity and complexity[28,35]. This argument has been explored with regard to economic exports[35]; however, less attention has been paid to the relationship between the complexity of scientific exports and economic development, as well as potential universal aspects of these relationships[36–39].

In this study, we apply maps of science[30–33] and revealed comparative advantage (RCA)[23]—a common measure for quantifying the economic and production advantages of countries[31]—to examine national science production, by considering scientific disciplines as types of 'products' that are exported by countries. That is, we investigate a nation's scientific development through scientific exports, in which research articles produced by a country and indexed in international bibliographic databases represent the exported scientific 'products' of the nation[21,40] (Methods and Supplementary Information section on Data). This is an important operationalization for our study; whereas a significant amount of scientific production occurs 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 localized scientific activity but to create a measurement that approximates economic exportation. This allows us to examine how knowledge is constructed within and flows across countries and can be used to inform science policy.

Results

Geography of revealed comparative advantage. We use the RCA[23] to assess the relative disciplinary strengths of each nation on the basis of publications indexed by the Web of Science database (Methods). If country c produces a greater share of its publications in field i compared to the world average share in the discipline, then $RCA_{c,i} > 1$ and country c is considered to have an RCA (or specialization) in discipline i.

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 (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 to Basalla's assumption that the science in colonial and post-colonial countries began with Western countries' exploitation of natural resources[23]. By contrast, far fewer countries—mostly in North America and Europe—specialize in Biochemistry and Molecular Biology, a discipline that 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[41]. That research and innovation emerge as a response to local issues and threats can also be observed in other contexts. For example, Agricultural and Food Science and Health Policy and Services are prominent in nations across the global south, where infectious disease[42] and food security[43] 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[3]. 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[8].

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?

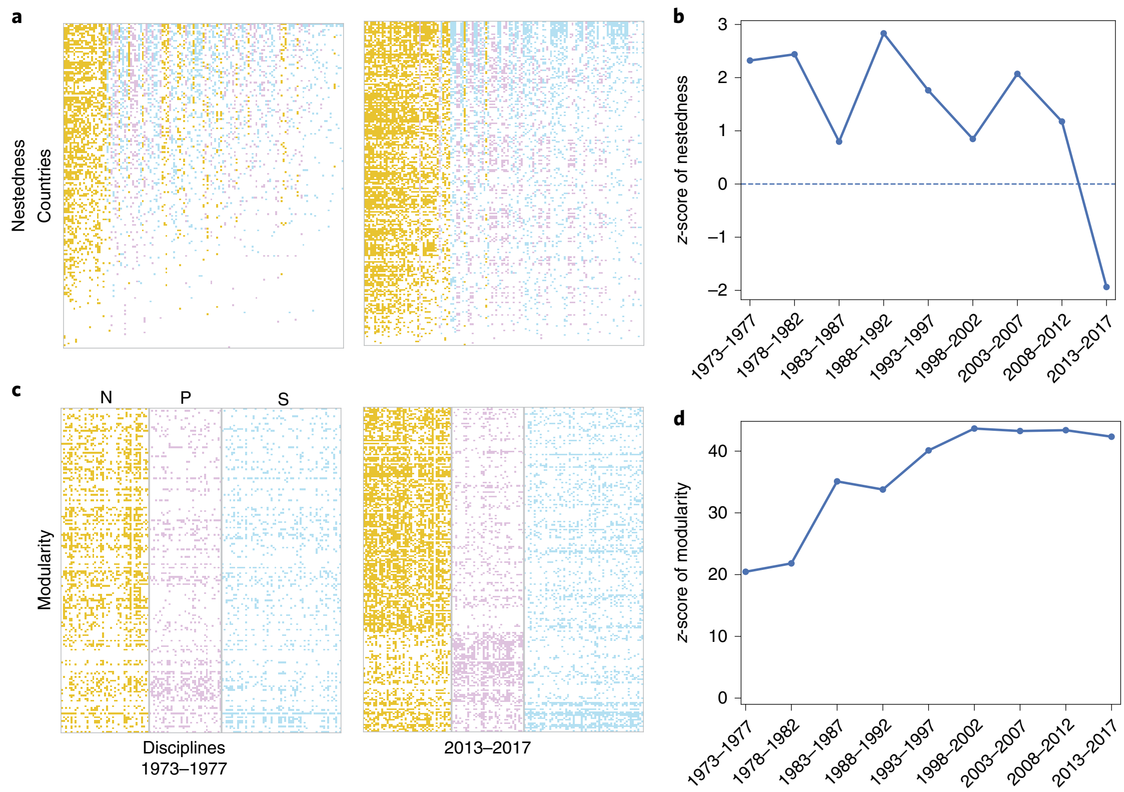

Fig. 4 | Nestedness and modularity of global science. a, The country–discipline RCA matrices of the earliest and the most recent periods where rows and columns are arranged in a descending order of number of advantages. **b**, The *z*-score of nestedness over time which is calculated through fixed–fixed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

Discipline relatedness network. Inspired by the relatedness network of economic product exports that underpins national economic development[31] as well as by the studies on scientific research space[10–12], 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 (Methods; Supplementary Figs. 4 and 5). The network builds on the idea that disciplines that are co-specialized are likely to require similar knowledge, skills, methods or equipment. To show its most salient structure, we apply the multiscale backbone extraction method[44]. This ‘backbone’ reveals three clusters, which are then formally defined by applying the Leiden community detection algorithm[45] with 100 iterations. Multiple runs of the algorithm did not alter the community memberships. Here, on the basis of the composition, we call these clusters Natural, Physical, and Societal clusters (Fig. 2a). The three clusters—Natural, Physical, and Societal—contain 45 disciplines, 37 disciplines and 61 disciplines, respectively. These clusters—while resembling previous observations[26,46]—do not conform to the common high-level classifications of disciplines. None of the clusters exclusively coincides with major classifications such as natural sciences, engineering or medical sciences. The high-level disciplinary classifications appearing in the Natural cluster (left) are primarily Natural and Medical Sciences. Most disciplines are dependent upon natural resources (for example, Geology, Entomology and Agriculture and Food Science) or concern the prevalent medical concerns in low-income areas (for example, Nutrition and Dietetic and Parasitology). The Physical cluster (right) contains primarily physical sciences and engineering, which are commonly considered as foundations for industry-based economic growth (for example, Chemistry and Applied Physics) and those that require technological investment (for example, Civil Engineering, Astronomy and 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 (for example, Psychiatry, Nursing and Cancer) as well as Social Sciences and Arts and Humanities (for example, Education, Sociology and International Relations).

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 (Fig. 2b; Methods). Aggregating countries on the basis of their income-level classification[47] reveals that niches are largely related to national wealth (Fig. 2b–d). Low-income countries (for example, 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 centre. High-income countries (for example, the United States, France and Japan) tend to occupy the centre and the space between Natural and Societal, suggesting balanced exportation. This pattern suggests that there might be a universal tendency that as a nation's economic power increases, their scientific exports move towards a more balanced portfolio.

The principle of relatedness. To understand the temporal evolution of national research portfolios, we first examine whether the development (or the loss) of RCA follows the principle of relatedness[9–12], which predicts that countries are more likely to develop a new advantage in a discipline that is close to their existing advantages (Fig. 2). By examining the entry (exit) of advantages across each subsequent time step (Methods), we show that the principle of

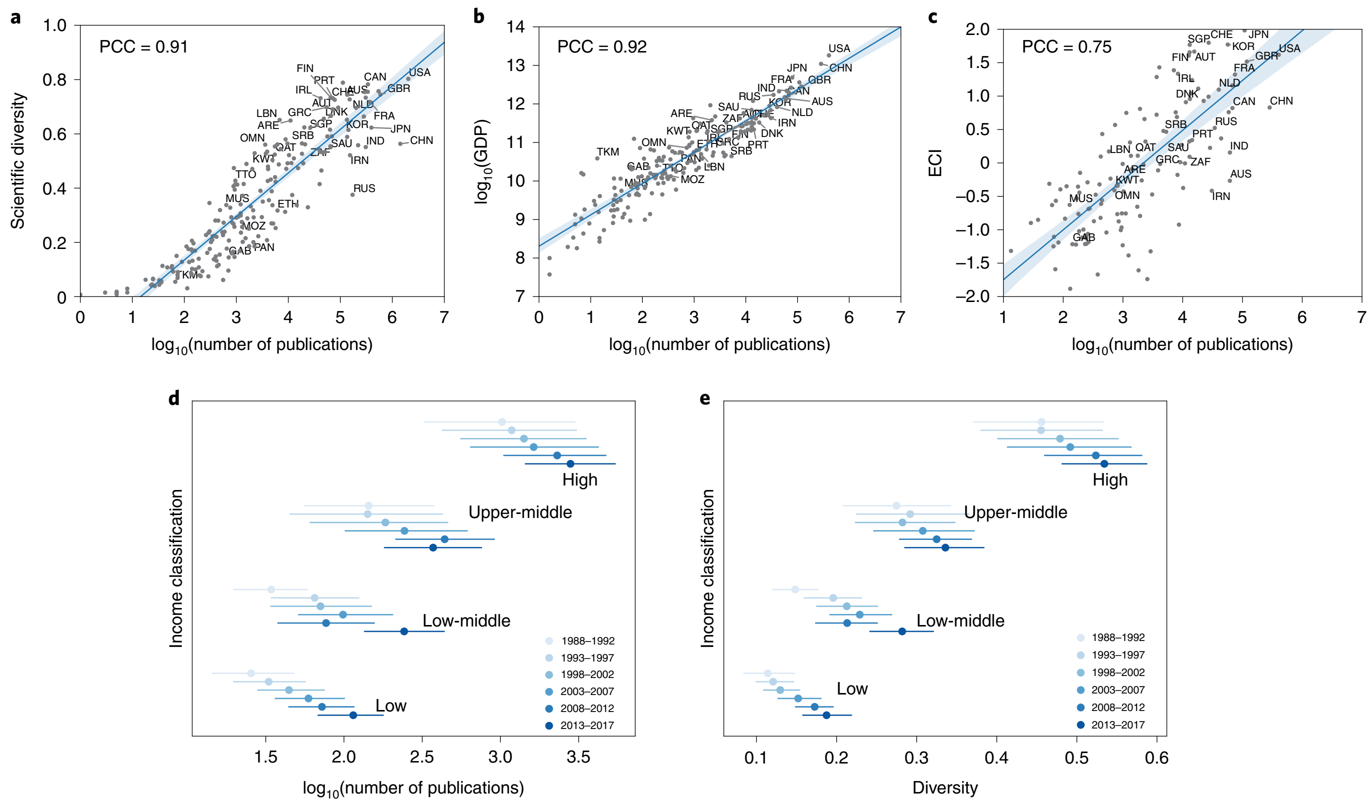

Fig. 5 | Scientific production is correlated with national development indicators. a, Number of publications is strongly correlated with scientific diversity (defined as one minus the GINI index of the RCA values of a country). Lines represent a linear regression model fit with *x* axis variable as the independent variable and *y* axis variable as dependent variable. Translucent band lines describe bootstrapped 95% confidence interval. PCC, Pearson correlation coefficient. ISO 3166-1 alpha-3 codes are used to denote each country. **b**,**c**, The relationship between scientific publication volume and nations GDP (**b**) and their ECI (**c**). **d**,**e**, The temporal development of the number of publications (**d**) and scientific diversity (**e**) by income group. The point shows the mean value of each group. Error bars represent the 95% confidence interval of the mean value drawn from bootstrapping. The number of countries in income group during period is presented in Supplementary Table 3. A total 1,000 iterations are used to compute the confidence interval. Point represents the mean value. Error bars represent the 95% confidence interval drawn from bootstrapping. The number of countries in each time period is presented in Supplementary Table 2.

relatedness indeed holds (Fig. 3a,b). The probability of a discipline's entry increases with the density of proximate specialized disciplines ($\beta = 0.33$, 95% CI $= [0.30, 0.36]$); the probability of a discipline's exit follows the opposite pattern ($\beta = -0.93$, 95% CI $= [-0.94, -0.92]$).

Moreover, if we aggregate countries on the basis of income groups, we further discover that low-income countries are more strongly constrained by the principle of relatedness than others (one-tailed *t*-test $t_{19} = 34$, $P < 0.001$) (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 likely to develop new advantage, thanks to 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. Our result aligns with the recently reported pattern that low-complexity economies experience stronger pull from relatedness and are less likely to enter unrelated activities[48].

The principle of relatedness shapes scientific development but to what extent? We compare the actual trajectories with a null model that is solely based on the principle of relatedness (Methods). As shown in Fig. 3e, the predicted research profiles converge towards the centre of the simplex; in other words, even with the constraining effect of the principle of relatedness, 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 centre, suggesting that scientific development is not entirely dictated by the principle of relatedness but may also be conditioned by the three clusters (Fig. 3e and Supplementary Figs. 6–8). The difference is particularly stark for countries specialized in the Natural cluster, suggesting that low-income countries may face a heavy hurdle breaking into other disciplines (Supplementary Figs. 6–8).

Structure of global science. Meanwhile, global science has been moving from a nested structure to a modular structure. It has been observed that the global economy exhibits a hierarchical[31,49,50] (or nested) structure, where rich countries can export a wide range of products—especially those that are exported by only a few countries—whereas poor countries can only export a small number of products that can be exported by many[31,35,51]. This pattern contrasts with a more classical theory of specialization, where countries would specialize and form a 'modular' structure. Inspired by the tension between these two ideas, we measure the nestedness and modularity of the scientific exports over time (Methods; 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 is probably associated with the trapping of low-income countries in the Natural cluster (Fig. 3) and the heavy investment and emphasis of applied sciences in rising economies such as China.

Table 1 | Regression results of predicting growth rate of GDP

	Dependent variable:						
	GDP growth (log ratio)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log GDP	−0.42***	−0.44***	−0.43***	−0.43***	−0.42***	−0.43***	−0.44***
	(−0.48, −0.36)	(−0.50, −0.38)	(−0.49, −0.37)	(−0.49, −0.37)	(−0.48, −0.36)	(−0.49, −0.37)	(−0.50, −0.38)
	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$
ECI	0.01	0.003	0.004	0.01	0.01	0.01	0.002
	(−0.02, 0.04)	(−0.03, 0.03)	(−0.03, 0.03)	(−0.03, 0.04)	(−0.02, 0.04)	(−0.03, 0.04)	(−0.03, 0.03)
	$P=0.64$	$P=0.86$	$P=0.79$	$P=0.73$	$P=0.65$	$P=0.72$	$P=0.92$
Log Population	0.15**	0.11*	0.12**	0.13**	0.15**	0.13**	
	(0.04, 0.27)	(−0.01, 0.23)	(0.002, 0.25)	(0.003, 0.25)	(0.02, 0.27)	(0.004, 0.25)	
	$P=0.02$	$P=0.09$	$P=0.05$	$P=0.05$	$P=0.03$	$P=0.05$	
Log no. Publications		0.04***					0.06***
		(0.01, 0.07)					(0.02, 0.09)
		$P=0.01$					$P=0.002$
Log no. Natural			0.03**			−0.01	
			(0.004, 0.07)			(−0.05, 0.04)	
			$P=0.03$			$P=0.84$	
Log no. Physical				0.04***		0.04**	
				(0.01, 0.07)		(0.001, 0.09)	
				$P=0.005$		$P=0.05$	
Log no. Societal					0.03*		
					(−0.003, 0.06)		
					$P=0.09$		
Diversity							−0.08
							(−0.27, 0.11)
							$P=0.44$
Observations	836	836	836	828	827	828	836
R^2	0.21	0.22	0.22	0.22	0.21	0.22	0.22
Adjusted R^2	0.07	0.08	0.07	0.07	0.06	0.07	0.07
F statistic	63.57*** (d.f. = 3; 705)	50.00*** (d.f. = 4; 704)	49.19*** (d.f. = 4; 704)	48.32*** (d.f. = 4; 696)	46.04*** (d.f. = 4; 695)	38.61*** (d.f. = 5; 695)	49.23*** (d.f. = 4; 704)

Note: (1)–(7) correspond to the regression models each of which employs a different set of independent variables shown. The P value is derived from a two-sided t-test. *$P<0.1$; **$P<0.05$; ***$P<0.01$.

Relationships between scientific activities and economic growth. Motivated by the connection between the economic wealth and their scientific niches and the increasing diversity shown above, we investigate the relationships among scientific diversity, publication volume and economic performance. We measure the diversity of a scientific portfolio with the Gini index of disciplinary RCA values (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.

We find that the number of publications, scientific diversity, GDP and ECI (Economic Complexity Indicator) are all strongly correlated with each other (publication and diversity, Pearson's correlation $r=0.91$; publication and GDP, Pearson's $r=0.92$; and publication and ECI, Pearson's $r=0.75$) (Fig. 5). Over the past 40 years, the average number of publications as well as the average scientific diversity of all income groups have been steadily increasing (see Fig. 5d,e). However, this steady growth is not enough to close the gap between income groups; the gap between high-income countries and low-income countries remains wide. Although scientific diversity is correlated with the number of publications, the diversification of research portfolio cannot be explained by the increase in the number of publications alone ($t_{1561}=17.02$, $P<.001$) (see Supplementary Fig. 9 for details).

Our results with two-way fixed effects panel regression models corroborate a mutual-influence relationship between publication volume and economic development[14,52,53] ($t_{704}=2.7$, $P=0.01$, effect size $=0.04$, 95% CI $=[0.01,0.07]$) (Table 1) and ($t_{705}=6.8$, $P<0.001$, effect size $=0.29$, 95% CI $=[0.21,0.38]$) (Supplementary Table 4). As indicated in the model 2 in Table 1, a 10% increase in number of publications is associated with 0.4% relative increase in economic growth rate (Table 1). The results are robust across different models with GDP and GDP per capita as independent variables, respectively ($t_{687}=2.29$, $P=0.03$, effect size $=0.02$, 95% CI $=[0.003,0.04]$) (Table 2). It further shows that, if the publications are divided into the three clusters we identified, the number of publications in the Physical cluster predicts GDP growth ($t_{695}=1.98$, $P=0.05$, effect size $=0.04$,

Table 2 | Regression results of predicting growth rate of GDP per capita

	Dependent variable:						
	GDP per capita growth (log ratio)						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log GDP per capita	−0.25***	−0.27***	−0.26***	−0.27***	−0.26***	−0.27***	−0.22***
	(−0.29, −0.21)	(−0.32, −0.22)	(−0.31, −0.22)	(−0.32, −0.23)	(−0.31, −0.21)	(−0.32, −0.23)	(−0.26, −0.18)
	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$	$P=0.00$
ECI	0.01*	0.01	0.01	0.01	0.01	0.01	0.01*
	(−0.002, 0.03)	(−0.01, 0.03)	(−0.004, 0.03)	(−0.005, 0.03)	(−0.003, 0.03)	(−0.005, 0.03)	(−0.002, 0.03)
	$P=0.10$	$P=0.19$	$P=0.15$	$P=0.16$	$P=0.11$	$P=0.16$	$P=0.10$
Log Population	−0.11***	−0.14***	−0.13***	−0.15***	−0.12***	−0.15***	
	(−0.18, −0.04)	(−0.22, −0.07)	(−0.21, −0.06)	(−0.22, −0.07)	(−0.20, −0.04)	(−0.22, −0.07)	
	$P=0.003$	$P=0.0003$	$P=0.001$	$P=0.0002$	$P=0.003$	$P=0.0002$	
Log no. Publications		0.02**					0.02**
		(0.003, 0.04)					(0.001, 0.04)
		$P=0.03$					$P=0.04$
Log no. Natural			0.02*			−0.003	
			(−0.001, 0.03)			(−0.03, 0.02)	
			$P=0.08$			$P=0.81$	
Log no. Physical				0.02**		0.02*	
				(0.003, 0.03)		(−0.002, 0.04)	
				$P=0.03$		$P=0.09$	
Log no. Societal					0.01		
					(−0.01, 0.02)		
					$P=0.44$		
Diversity							−0.12**
							(−0.22, −0.02)
							$P=0.03$
Observations	818	818	818	811	810	811	818
R^2	0.16	0.17	0.17	0.17	0.16	0.17	0.16
Adjusted R^2	0.01	0.01	0.01	0.01	0.005	0.01	0.001
F statistic	44.60*** (d.f. = 3; 688)	34.97*** (d.f. = 4; 687)	34.35*** (d.f. = 4; 687)	35.45*** (d.f. = 4; 680)	33.43*** (d.f. = 4; 679)	28.33*** (d.f. = 5; 679)	32.61*** (d.f. = 4; 687)

Note: The *P* value is derived from a two-sided *t*-test. *$P<0.1$; **$P<0.05$; ***$P<0.01$.

95% CI = [0.001,0.09]) (Table 1). However, this result is driven by the countries like China that simultaneously exhibited scientific investment focused on the Physical cluster and strong economic growth. Removing China from the model makes the coefficient of number of publications in the Physical cluster statistically insignificant ($t_{688}=0.77$, $P=0.45$, effect size = 0.02, 95% CI = [−0.03, 0.06]) (Supplementary Tables 5 and 6). However, the difference between the coefficient of the number of publications in the Physical cluster with all observations and the coefficient of the corresponding parameter excluding China is not significant ($\beta_{\text{with China}}=0.04\pm0.02$, $\beta_{\text{without China}}=0.02\pm0.02$) (ref. [54]). Although we cannot exclude the possibility that China's focused investment in the Physical cluster might have played an important role in its economic growth, our models do not provide enough evidence to demonstrate whether the strength in the Physical cluster is associated with the future economic growth across countries. The relationship between scientific development and economic development may be contingent on countries and complex.

In contrast to the theory that diversity and complexity of economy are closely linked to the economic growth, we could not find strong associations between the diversity of scientific portfolio and economic growth. Scientific diversity could neither predict GDP growth nor predict the growth of publication ($t_{704}=-0.78$, $P=0.44$, effect size = −0.08, 95% CI = [−0.27,0.11]) (Table 1) and ($t_{704}=-0.77$, $P=0.44$, effect size = −0.10, 95% CI = [−0.36,0.16]) (Supplementary Table 4), whereas GDP is associated with the growth rate of scientific diversity ($t_{705}=2.71$, $P=0.007$, effect size = 0.43, 95% CI = [0.12,0.74]) (Supplementary Table 7). Our results suggest that a balanced research portfolio may be a result, rather than the cause, of economic development. However, scientific diversity is negatively associated with the similarity between newly entered disciplines and the existing advantages, suggesting that balanced research profile may be associated with more flexibility to develop research areas ($t_{703}=-1.71$, $P=0.09$, effect size = −0.60, 95% CI = [−1.29,0.09]) (Table 3). Countries with high diversity tend to develop more easily beyond their current research advantages.

Table 3 | Regression results of predicting average similarity of new entered disciplines

	Dependent variable:	
	Similarity	
	(1)	(2)
Log GDP	−0.11	−0.09
	(−0.34, 0.11)	(−0.31, 0.14)
	$P=0.34$	$P=0.45$
ECI	−0.02	−0.02
	(−0.13, 0.09)	(−0.13, 0.09)
	$P=0.70$	$P=0.76$
Log no. Publications	−0.02	0.04
	(−0.13, 0.09)	(−0.09, 0.17)
	$P=0.70$	$P=0.57$
Diversity		−0.60*
		(−1.29, 0.09)
		$P=0.09$
Observations	835	835
R^2	0.002	0.01
Adjusted R^2	−0.18	−0.18
F statistic	0.54 (d.f. = 3; 704)	1.14 (d.f. = 4; 703)

Note: The P value is derived from a two-sided t-test. $^{*}P<0.1$; $^{**}P<0.05$; $^{***}P<0.01$.

Discussion

It is widely believed that scientific development holds the key to the future prosperity of a nation[55,56]. Yet, whether there are universal structural patterns of scientific development at the national level remains an open question. By analysing more than 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. We further reveal that, although individual country is moving towards a more balanced research profile, global science is becoming more modular. We also confirm that economic wealth and scientific publication are mutually predictive of each other, suggesting a strong feedback loop. Finally, we find evidence that economic growth leads to higher scientific diversity and countries with diverse research portfolios are more flexible regards developing new research areas.

Our results in part reaffirm the general patterns observed in previous studies on the structure of knowledge space and the principle of relatedness[10–12,22,23,26,46,57]. 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 resonates with Basalla's 'Spread of Western Science' (Basalla), pointing to the colonial exploitation of natural resources. Klavans and Boyack[46] and Moya-Anegón and Herrero-Solana[26] identified similar disciplinary clusters that condition on national scientific development. Furthermore, studies have demonstrated the principle of relatedness in knowledge domain[9–12]. Building on this literature, by taking a country-level approach and network analysis, our work quantitatively identifies three disciplinary clusters where a country's niche in the knowledge space is associated with its economic development. We further reveal that the principle of relatedness alone is not enough to explain the dynamic evolution of research areas in countries.

This study has several limitations. First, it relies on a bibliographic database created and maintained by a Western scientific enterprise. Therefore, it overestimates research from Western countries and publications in English while underestimating the production in other nations and languages (Supplementary Information section on Data). Still, we argue that our operationalization is reasonable under the analogy to product exportation[31,35] and the status of English as the de facto lingua franca of science[58]. Second, many of our analyses considered the RCA matrix as a bipartite network. This approximation is not strictly valid because the edges are not independent of each other. Finally, our regression models are not free from multicollinearity issues as there exists significant correlation between GDP and the number of publications. Moreover, we show that the inferred relationships between scientific enterprise and economic growth can be easily driven by a small number of countries by demonstrating that the association between the publications in the Physical cluster and economic growth is driven by China, which have achieved strong growth in both scientific production in the Physical cluster and economy. Due to the complexity of economic and scientific development, we note that reliable causal inference with country-level data is often infeasible and our results do not necessarily confirm nor reject a direct causal relationship between national scientific development and economic growth. There exist many unobserved hidden confounders and complex feedback mechanisms between scientific and economic development.

Despite those limitations, our empirical framework may provide a useful perspective to study the structure and evolution of national scientific portfolios and the relationship 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 for future scientific capacity and economic growth. Our results also highlight the importance of considering scientific capacity in the study of economic development. 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.

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 5-yr 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 (Supplementary Fig. 2).

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[59]. 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 colour nodes in Fig. 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 that each author's country contributes equally to the publication. Fractional counting and counting based on corresponding authorship are another two widely used counting methods. These counting methods are highly correlated at the macro level[60]. However, Web of Science has a highly inaccurate coverage on the corresponding author information before 2008, where corresponding author is, in most cases, simply assigned to the first author/institution. Given the diachronic nature of our analysis, we were unable to use a counting method based on corresponding authorship. The discipline network constructed from fractional counting shares high similarity with the network constructed using full counting method. See Supplementary Information section on Data for more details.

We use data on national GDP from the World Bank[47,61] to approximate the economic wealth of each country. The dataset covers 264 countries from 1960 to 2019. Income classification comes from the World Bank database[47] 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 Supplementary Information section on Data for more details.

Revealed comparative advantage. The 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 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:

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

ϕ is a 143×143 matrix that captures the proximity between pairs of disciplines (Supplementary Fig. 4).

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 the aggregated network (Supplementary Fig. 5). The multiscale backbone extraction method[44] exposes three visual clusters when laid out with a force-directed layout algorithm (Gephi's ForceAtlas2). We then apply the Leiden algorithm[45] to the full network with 100 iterations to define the membership of each discipline in the three communities. Multiple runs of the algorithm produced exactly the same results. Other methods produce similar results, although some methods partition the network into smaller communities (Supplementary Information section on the Disciplinary relatedness network).

Position within the simplex. Position within the simplex measures cluster-level specialization concentration of a country. We first calculate $C_i = n_i/N_i$, where n_i is the number of disciplines in cluster i with $\mathrm{RCA} > 1$ and 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 ϕ_{ij} is the proximity between discipline i and j and $x_i = 1$ if $\mathrm{RCA}_{ki} > 1$ else $x_i = 0$ and the density of existing advantages, ω_j^k, is the proximity-weighted sum of all disciplines that are connected to j with $\mathrm{RCA}_{ki} > 1$. We bin the density values and aggregate across countries and time periods to calculate the probability of entry and exit, 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 (ordinary least squares) 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 ($\mathrm{RCA} < 1$) discipline, we assign a probability that the discipline will be entered ($\mathrm{RCA} > 1$) in the subsequent time period on the basis of 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 exit. For each time period and each country, the new entered and exited disciplines are sampled using the null model while preserving the number of new entered and exited disciplines 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 datapoints located on the boundary of the simplex. To smooth out the noise, we aggregate datapoints 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 $\mathrm{RCA}_{c,i} > 1$. Modularity[62] 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 RCA; for example, China is classified to the Physical cluster since it has highest cluster-level RCA value in the 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)[63] method. NODF measures the degree of overlapping between row pairs and column pairs in the adjacency matrix. The metric is defined as:

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

where $\sum^{N}_{\mathrm{paired}}$ is the averaged degree of nestedness for each pair of row and column based on the principles of decreasing fill and paired overlap[63] and 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 one 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 equal to 0.1 and height equal to 0.5. The starting point of an 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.

Regression analysis. We use a fixed-effect panel regression model to investigate the relationship between economic growth and scientific development. The model is written as follows:

$$Y_{c,t} = \beta_0 + \beta_1 X_{1,ct} + \beta_2 X_{2,ct} + \ldots + \beta_k X_{k,ct} + \alpha_c + \alpha_t$$

where c denotes countries, t denotes time periods, $Y_{c,t}$ is the investigated dependent variable and α_c and α_t are the country-specific and time-specific intercepts that capture the heterogeneity across countries and across time periods. The dependent variables involved in our analysis are log ratio of GDP growth rate, log ratio of publication growth rate, scientific diversity growth rate and averaged disciplinary similarity. Growth rate is measured as $N_{c,t+1}/N_{c,t}$ where $t+1$ represents the next time period which is the following 5 yr. The included controlled variables are averaged GDP value, averaged ECI value and averaged number of populations. The investigated independent variables are the total number of publications, scientific diversity and the number of publications in Natural, Physical and Societal clusters. We apply log-transformation with base 10 to GDP, GDP per capita, the number of publications, the number of populations and the growth rate.

Averaged disciplinary similarity measures how similar the new entered disciplines are compared with the current existing advantages. The averaged disciplinary similarity is defined as:

$$\rho_c = \sum_i^n \frac{\omega_i^c}{n}$$

where n is the number of new entered disciplines and ω_i^c is the normalized density of new entered disciplines. Normalized density is measured as the z-score of raw disciplinary density of all disadvantaged disciplines. High averaged disciplinary similarity indicates that the new entered disciplines have higher similarity with existing advantaged disciplines compared with average similarity between advantaged disciplines and inactivated disciplines.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data used in this study are available at https://figshare.com/articles/journal_contribution/Untitled_Item/13623035/3

Code availability

The code used for data processing and analysis is available at https://github.com/yy/national-science-exports

Received: 5 May 2021; Accepted: 27 April 2022;
Published online: 02 June 2022

References

1. *Publications Output: U.S. Trends and International Comparisons* (National Science Board, 2019).
2. Tollefson, J. China declared world's largest producer of scientific articles. *Nature* **553**, 390–390 (2018).
3. Zhou, P. & Leydesdorff, L. The emergence of China as a leading nation in science. *Res. Policy* **35**, 83–104 (2006).
4. Noorden, R. V. Science in East Asia—by the numbers. *Nature* **558**, 500–501 (2018).
5. Li, K.-W. *Capitalist Development and Economism in East Asia: The Rise of Hong Kong, Singapore, Taiwan and South Korea* (Routledge, 2002).
6. Livingstone, D. N. *Putting Science in its Place: Geographies of Scientific Knowledge* (Univ. Chicago Press, 2010).
7. Seth, S. Putting knowledge in its place: science, colonialism, and the postcolonial. *Postcolonial Stud.* **12**, 373–388 (2009).
8. 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).
9. Hidalgo, C. A. et al. in *Unifying Themes in Complex Systems IX Proc. Ninth International Conference on Complex Systems* (eds Morales, A. J. et al.) 451–457 (Springer International, 2018).
10. Chinazzi, M., Gonçalves, B., Zhang, Q. & Vespignani, A. Mapping the physics research space: a machine learning approach. *EPJ Data Sci.* **8**, 33 (2019).
11. Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M. & Hidalgo, C. A. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* **109**, 1695–1709 (2016).
12. Boschma, R., Heimeriks, G. & Balland, P.-A. Scientific knowledge dynamics and relatedness in biotech cities. *Res. Policy* **43**, 107–114 (2014).
13. Stephan, P. *How Economics Shapes Science* (Harvard Univ. Press, 2012).
14. Lee, L.-C., Lin, P.-H., Chuang, Y.-W. & Lee, Y.-Y. Research output and economic productivity: a Granger causality test. *Scientometrics* **89**, 465–478 (2011).
15. 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).
16. Hornyak, T. Chilean research grows despite poor investment. *Nat. Index* (2016).
17. Bronfman, L. A panorama of Chilean astronomy. *Messenger* **107**, 14–18 (2002).
18. Bajak, A. Chile's chance to embrace science for the twenty-first century. *Nature* **552**, S53–S55 (2017).
19. Yeom, H. W. South Korean science needs restructuring. *Nature* **558**, 511–513 (2018).
20. *Science and Engineering Indicators 2018* (National Science Board, 2018).
21. May, R. M. The Scientific Wealth of Nations. *Science* **275**, 793–796 (1997).
22. Comte, A. *The Positive Philosophy of Auguste Comte* (C. Blanchard, 1855).
23. Basalla, G. The spread of western science. *Science* **156**, 611–622 (1967).
24. Anderson, W. Remembering the spread of western science. *Hist. Rec. Aust. Sci.* **29**, 73 (2018).
25. Raina, D. From west to non-west? Basalla's three-stage model revisited. *Sci. Cult.* **8**, 497–516 (1999).
26. 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).
27. Cimini, G., Gabrielli, A. & Labini, F. S. The scientific competitiveness of nations. *PLoS ONE* **9**, e113470 (2014).
28. Hidalgo, C. A. Economic complexity theory and applications. *Nat. Rev. Phys.* **3**, 92–113 (2021).
29. Hong, I., Frank, M. R., Rahwan, I., Jung, W.-S. & Youn, H. The universal pathway to innovative urban economies. *Sci. Adv.* **6**, eaba4934 (2020).
30. Alabdulkareem, A. et al. Unpacking the polarization of workplace skills. *Sci. Adv.* **4**, eaao6030 (2018).
31. Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482–487 (2007).
32. Boschma, R. A. & Koen, F. Evolutionary economics and industry location. *Rev. Reg. Res.* **23**, 183–200 (2003).
33. Boschma, R. A. & Frenken, K. Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *J. Econ. Geogr.* **6**, 273–302 (2006).
34. Boschma, R. & Koen, F. in *The New Oxford Handbook of Economic Geography* (eds Clark, G. L. et al.) 213–229 (Oxford Univ. Press, 2018).
35. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl Acad. Sci. USA* **106**, 10570–10575 (2009).
36. Boyack, K. W. & Klavans, R. Creation of a highly detailed, dynamic, global model and map of science. *J. Assoc. Inf. Sci. Technol.* **65**, 670–685 (2014).
37. Boyack, K. W. & Klavans, R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? *J. Am. Soc. Inf. Sci. Technol.* **61**, 2389–2404 (2010).
38. Boyack, K. W., Klavans, R. & Börner, K. Mapping the backbone of science. *Scientometrics* **64**, 351–374 (2005).
39. Boyack, K. W., Börner, K. & Klavans, R. Mapping the structure and evolution of chemistry research. *Scientometrics* **79**, 45–60 (2009).
40. King, D. A. The scientific impact of nations. *Nature* **430**, 311–316 (2004).
41. 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).
42. 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).
43. *World Health Statistics 2018: Monitoring Health for the SDGs* (World Health Organization, 2018).
44. Serrano, M. A., Boguna, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl Acad. Sci. USA* **106**, 6483–6488 (2009).
45. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
46. Klavans, R. & Boyack, K. W. The research focus of nations: economic vs. altruistic motivations. *PLoS ONE* **12**, e0169383 (2017).
47. Fantom, N. & Serajuddin, U. *The World Bank's Classification of Countries by Income* (The World Bank, 2016).
48. Pinheiro, F. L., Hartmann, D., Boschma, R. & Hidalgo, C. A. The time and frequency of unrelated diversification. *Res. Policy* https://doi.org/10.1016/j.respol.2021.104323 (2021).
49. Bustos, S., Gomez, C., Hausmann, R. & Hidalgo, C. A. The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLoS ONE* **7**, e49393 (2012).
50. Gao, J., Zhang, Y.-C. & Zhou, T. Computational socioeconomics. *Phys. Rep.* **817**, 1–104 (2019).
51. Morrison, G. et al. On economic complexity and the fitness of nations. *Sci. Rep.* **7**, 15332 (2017).
52. Inglesi-Lotz, R., Balcilar, M. & Gupta, R. Time-varying causality between research output and economic growth in US. *Scientometrics* **100**, 203–216 (2014).
53. Vinkler, P. Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries. *Scientometrics* **74**, 237–254 (2008).
54. Gelman, A. & Stern, H. The difference between "significant" and "not significant" is not itself statistically significant. *Am. Stat.* **60**, 328–331 (2006).
55. 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).
56. Powell, W. W. & Snellman, K. The knowledge economy. *Annu. Rev. Sociol.* **30**, 199–220 (2004).
57. Glänzel, W. National characteristics in international scientific co-authorship relations. *Scientometrics* **51**, 69–115 (2001).
58. Gordin, M. D. *Scientific Babel: How Science was Done Before and After Global English* (Univ. Chicago Press, 2015).
59. 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).
60. Sugimoto, C. R. & Vincent, L. *Measuring Research: What Everyone Needs to Know* (Oxford Univ. Press, 2018).
61. *GDP (current US$)* World Development Indicators (World Bank, 2019); https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=1W
62. Barber, M. J. Modularity and community detection in bipartite networks. *Phys. Rev. E* **76**, 066102 (2007).

63. 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
L.M., D.M. and Y.Y.A. acknowledges funding support from the Air Force Office of Scientific Research under award no. FA9550-19-1-0391. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank S. Milojević, J. Yoon, S. Kojaku, B.-K. Lee, B. Weinberg, P. Stephan, S. Schnell, T. Zhou, J. Huang, J. Gao, Y. Bu, Q. Zhang, B. Uzzi, H. Youn and N. 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. and Y.Y.A. performed analysis. All authors contributed to the interpretation of the results and writing of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-022-01367-x.

Correspondence and requests for materials should be addressed to Yong-Yeol Ahn.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature research

Corresponding author(s): Yong-Yeol Ahn

Last updated by author(s): Mar 27, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No primary data were collected.
Data analysis	Python 3.7, added packages: leidenalg 0.8.3; nestedness_calculator: https://github.com/tsakim/nestedness/blob/master/nestedness_calculator.py Gephi 0.9.2 (ForceAtlas2); R 3.6.2, added packages: plm 2.4-1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GDP data at country level from World Bank (https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)
Income group classification at country level (https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups)
Economic complexity indicator (https://legacy.oec.world/en/)

Compiled publication record at country level and above mentioned datasets are available at (https://figshare.com/articles/journal_contribution/Untitled_Item/13623035/3)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We study the universal patterns behind national scientific development. The study is quantitative, combining network analysis and regression analysis.
Research sample	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 dataset covers publication records of 217 countries from 1973 to 2017.
Sampling strategy	We used the full set of publication record of countries. No sampling strategy is used.
Data collection	No primary data is collected.
Timing	1973-2017
Data exclusions	We used the complete publication records of countries. No country is excluded.
Non-participation	Our study is based on the secondary dataset. Data exclusion is described above.
Randomization	Our study doesn't involve random-control design.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging