Institute for Scientific Information

Global Research Report
Subject diversity in research portfolios: what it is, how to index it and its role in innovation

Jonathan Adams, David Pendlebury and Martin Szomszor

Clarivate™

Author biographies

Jonathan Adams is Chief Scientist at the Institute for Scientific Information (ISI)™. He is also a Visiting Professor at King's College London, Policy Institute. In 2017 he was awarded an Honorary D.Sc. by the University of Exeter, for his work in higher education and research policy. ORCiD: https://orcid.org/0000-0002-0325-4431.

David Pendlebury is Head of Research Analysis at the Institute for Scientific Information (ISI). Since 1983 he has used the Web of Science™ data to study the structure and dynamics of research. He worked for many years with ISI founder Eugene Garfield. With Henry Small, David developed the Web of Science Essential Science Indicators™. ORCiD: https://orcid.org/0000-0001-5074-1593.

Dr Martin Szomszor is Director at the Institute for Scientific Information. He has a PhD in Computer Science, with expertise in knowledge engineering, machine learning and natural language processing. He was named a 2015 top-50 U.K. Information Age data leader for his work in creating the REF2014 impact case studies database for the Higher Education Funding Council for England (HEFCE). ORCiD: https://orcid.org/0000-0003-0347-3527.

Foundational past, visionary future

About the Institute for Scientific Information

The Institute for Scientific Information (ISI) at Clarivate has pioneered the organization of the world's research information for more than half a century. Today it remains committed to promoting integrity in research whilst improving the retrieval, interpretation and utility of scientific information. It maintains the knowledge corpus upon which the Web of Science™ index and related information and analytical content and services are built.

It disseminates that knowledge externally through events, conferences and publications whilst conducting primary research to sustain, extend and improve the knowledge base. For more information, please visit https://clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/

ISBN 978-1-8382799-3-6

Cover image: Feather star and colorful Gorgonian Sea Fans

Coral reefs are among the most diverse habitats in the world with reefs in the Philippines supporting as many as 3,000 fish species as well as crustacea, annelids, echinoderms and the corals themselves. Bleaching due to rising sea temperature and damage due to poor fisheries management is threatening these extraordinary places.

Executive summary

- **Diversity matters. Organizations have long recognized that cultural diversity in all its forms is a competitive differentiator that correlates to economic success.** Organizations with high levels of cultural diversity are better equipped to adapt to new ideas, technologies, social and economic challenges. Research diversity, in all its forms, is an engine of economic progress and stimulates innovation. Diversity of research contributes to stability, resilience and innovation in ecological and economic systems. It should be an equally valuable part of management planning in research systems, reflected in the subject spread of countries and institutions.

- **Diversity analysis provides a forward-looking view of the opportunities for intellectual, technological and scientific evolution, in contrast to retrospective publication citation analysis, which looks back to achievement.** The association between diversity and response to challenge and innovation points to a new source of information for research managers that will form an important tool in support of strategic investment planning.

- **Analysis of research diversity has suffered from confusion with interdisciplinarity, mathematical complexity and lack of academic consensus on indicators and methodologies.** The Web of Science provides a practical and stable global reference baseline for subject diversity that reduces calculations to one of evenness of publication counts across journal-based subject categories, using the Gini index.

- **The global baseline is stable over forty years as the index grew from 8,000 to 20,000 journals and more than fourfold to nearly three million articles and reviews (original papers).** The distribution of publication counts between categories is innately uneven, as are raw national counts. Indexing (normalizing) counts against the baseline enables comparison across years and countries. (Figure 1)

- **The United States and United Kingdom had the most even portfolios in 1981.** Other G7 countries' publications were more specialist and the BRICK nations were most uneven. Research portfolios became progressively more even, with the G7 countries converging on a similar Gini index. Similarity analysis shows that the G7 converged in subject coverage but the BRICKs diverged to a technology focus. (Figures 2 and 3)

- **COVID-19 presented an unforeseen global research challenge.** Countries with a diverse research base responded with a rapid and comprehensive range of innovative research – but most specialist countries did not. An exception, Brazil, has a narrow research base but one pre-adapted to this challenge. (Figure 4)

- **Institutional change in Australia and the United Kingdom shows universal diversity increase.** Rates were similar across institutions, with greater evenness associated with leading institutions. Australian institutions converged more than in the United Kingdom but the most even portfolios settled at a similar index level in both countries. (Figures 5 and 6)

Diversity analysis provides a forward-looking view of the opportunities for intellectual, technological and scientific evolution.

Introduction

An oak wood is a complex network of plants and animals whose diversity promotes ecosystem services and provides a robust response to natural events like fire and storms (Gamfeldt et al 2013). A specialist economy that depends on a single production sector is vulnerable to changes in supplies and consumer demand (European Central Bank 2021). Diversity “offers a means to promote innovation, hedge ignorance, mitigate lock-in and accommodate pluralism,” and offers a strategy “for achieving qualities of precaution, resilience and robustness that are central to sustainability” (Stirling 2007, p 715). Similarly, the agility of nations or institutions to respond to challenges or opportunities in science, medicine, technology and social sciences may well depend on a diversity of research activities and competencies (Rousseau et al 2019a).

What is diversity? It is not interdisciplinarity, though the concepts are clearly related. A research unit may be interdisciplinary without being diverse. Substantial organizational diversity may, however, facilitate the capacity for dynamic interdisciplinary research (Rafols and Meyer 2010, Wang et al 2015, Zhang et al 2016, Leydesdorff 2018, Finardi and Lamberti 2021, Hackett et al 2021, Huang et al 2021, Zhang and Leydesdorff 2021). We will show that national responses to the COVID-19 challenge benefit from a diverse research base.

Diversity is a property of the organization, not a measure of its research outcomes. When we look at research impact we are always looking back, which may be interesting but does not necessarily inform decision-making. If good research outcomes are linked to research diversity, then we can plan and manage. The research diversity of organizations has not been widely studied, however, so the relationship between diversity and research outcomes is only weakly informed (Janavi et al 2020). For many policy initiatives the mantra of ‘selectivity and excellence’ was long promoted (Aston and Shutt 2009), although there was little evidence that concentration or scale produced more significant research (Adams and Smith 2003). Instead, evidence pointed to the agility and resilience of generalist universities over specialist institutes (May 1997).

In this Global Research Report, we extend our previous work examining the changing diversity of national research portfolios by focusing on the balance or evenness component (Adams et al 2020a,b). This is a pragmatic approach, aimed at a simple, practical index of portfolio diversity to inform and be used by research managers.

Diversity and interdisciplinarity

Diversity and interdisciplinarity are different but the research literature frequently mixes them without definition. Diversity is the co-occurrence of several distinct topics or disciplines. Interdisciplinarity is when those strands are brought together in innovative research that makes a new thread. The lack of clear metrics and consensus methodologies makes this confusing for the observer. Many analyses use metadata, such as the mix of journal categories covered by a paper's reference list, to index the interdisciplinarity of the paper using the same variables discussed in this report and may also refer to it as diversity (Wang et al 2015, Yegros-Yegros et al 2015). Such an analysis does not necessarily, in practice, measure either of these things. (See also Hackett et al 2021, Thijs et al 2021).

What have researchers said about research diversity?

Diversity, an established concept in theoretical and practical work in economics and ecology (e.g. Leinster and Cobbold 2012), received increasing attention in scientometrics after the publication of a systematic framework by Professor Andy Stirling (Science Policy Research Unit, University of Sussex, U.K.) Stirling (2007) points out, first, that systems are neither diverse nor specialist: they are more or less diverse. He describes three key properties: variety, disparity and balance. Variety is the number of categories used to group things (species, economic sectors or research topics); disparity is the distance or difference between these categories; and balance (or evenness) is the distribution of things across these categories. These properties offer a route to measurement and exploration.

Well known formulae for indexing the components of diversity include Shannon entropy, the Simpson index, the Gini coefficient and Rao-Stirling diversity (Rao 1982, Stirling 2007). More recently, so-called True RS (Rao-Stirling) diversity (Zhang et al 2016), DIV (Leydesdorff et al 2019a) and DIV* (Rousseau 2019b, Leydesdorff et al 2019b) have been suggested as improvements or alternatives. These recent studies on the meaning and measurement of diversity in research portfolios have stimulated debate and innovation of approach. But they have not yet provided a clear method that would be straightforward for managers to operationalize, interpret and then use for decision-making.

Informative analysis needs reference points, because diversity is not an absolute, and it needs to draw on relevant comparisons – with peers or over time – made on a like-for-like basis. We have developed a simple approach, informed by prior research, that cuts across the complexity of measurement by using a simple and powerful categorical reference system that reduces variety and disparity to constants. We focus only on evenness, or balance, in the distribution of research publications at discipline level. Such data are already available in institutional and global databases. Because this approach favors simple and intuitive comparisons, it bypasses an array of theoretical choices in measurement and application that, however interesting, are of more academic than managerial or policy interest.

Global indexing for evenness and diversity

As a place to start our assessment of diversity, we turn to the Web of Science Core Collection™, a selective database of publications in a global set of around 20,000 academic journals, with sustained coverage over many decades, and curated to achieve a reasonable representation of research activity across disciplines and regions. This database provides a global background that shows what the systemic balance has been and now is between research areas and also provides a consistent reference point or benchmark against which we can explore difference and similarity at national levels.

We first quantify secular trends in the overall Web of Science global background in research balance and then, second, describe and compare broad trends at the national level.

Our analysis uses the three principal Web of Science citation indices: Science Citation Index Expanded (SCIE)™; Social Sciences Citation Index (SSCI)™; and Arts & Humanities Citation Index (AHCI)™. The Emerging Sources Citation Index (ESCI)™ was not included. Web of Science journal categories have a sufficient reasonable granularity that reflects shifts in research activity across 254 research disciplines including all subject areas. A criterion for inclusion is that the journal should publish English titles and abstracts and this language constraint may mean that the more 'national' areas of the arts, humanities and much of social science do not provide global comparability. However, reducing the set to 194 science and technology categories produced results little different to using 248 categories (Dance, Poetry and four regional literature categories were always omitted). We therefore used the larger set.

The choice of a classification scheme is critical to a good diversity index and the Web of Science categories have been tried and tested by researchers over decades.

Diversity may be calculated in terms of variety, disparity and balance or evenness. For the Web of Science, the variety of journal-based categories is constant across the analytical period and is the same for all sub-samples. The choice of a classification scheme is critical to a good diversity index (Thijs et al 2021) and the Web of Science categories have been tried and tested by researchers over decades.

While the disparity between categories may change as journals are added and deleted, it does so slowly and in a uniform manner for all entities. We can therefore simplify the analysis and focus solely on the relative frequency (the evenness of the count) of publications in each of the 248 categories. For this analysis we used only substantive, original academic publications, which are documents identified as articles and reviews and which we refer to collectively as 'papers'.

The study period covers the 38 years from 1981-2018, during which there was global growth in research publication and indexing, so all global and national counts of papers were converted to proportions of the respective annual total. Note that we apply no fractional attribution of addresses or journals. Each paper is assigned to a country's tally if there is at least one author address for that country and assigned as a whole count to a category tally if the journal was assigned to that category.

The specific numerical value of an index of diversity (or its components) for a country or institution carries no practical information. Useful management policy information comes from comparisons, across time or between countries, about the trajectory of an index.

Our preferred index was guided by Henk Moed (Moed 2006) who used Pratt's Index (Pratt 1997; Egghe 1987) to compare publication portfolios among universities. The Pratt Index is a variant of the Gini coefficient and has a similar arithmetic value (Carpenter 1979). Both index disciplinary specialization, indicating whether a publication portfolio (of a country or an institution) is highly specialized (high index values indicating specialism and a lack of balance and evenness) or evenly spread (low index values indicating an even distribution) across categories.

The Gini coefficient

The Gini coefficient, sometimes called the Gini index or Gini ratio, is a measure of statistical dispersion originally intended to represent income or wealth inequality. It was developed by the Italian, Corrado Gini (Gini, 1909). The Gini coefficient measures inequality among values of a frequency distribution (for example, numbers of publications in journal categories). A Gini coefficient of zero expresses perfect equality, where all values are the same (for example, where every category has the same count or proportion of papers). A Gini coefficient of one (or 100%) expresses maximal inequality among values. Where many papers are in a few categories and all others have only a few papers, the Gini coefficient will be nearly one.

Web of Science categorical publication counts are innately uneven (discipline categories naturally vary in size), so it is informative to 'normalize' absolute counts for any entity against the global baseline. The research diversity of any country thus has two components that may carry useful information:

- Research choices in line with the consensus of the global research base, computed from the proportion of original publication counts across categories.
- Research choices specific to that country, dictated by resource and policy factors, computed after normalization of original publication proportions by comparison with global proportions.

To compute the global balance of papers, the total annual numbers of papers (articles and reviews) are counted by category and by year. Papers are dispersed across categories according to the assignment of journals and the numbers of papers published in each journal. We do not, consequently, expect that there will be an 'even' spread of papers. The question is whether a particular sample has a distribution that is more or less concentrated than the global 'null model'. The relationship between the publication balance for an individual country (or institution or other entity) and the global baseline is computed by calculating the ratio between the national and global papers in each category and dividing this by the ratio of total national and global papers for the same year.

The methodology for management of the annual, categorical paper counts, for normalization against the global baseline and for the subsequent calculation of the Gini coefficient is detailed in Adams et al (2020a) and will not be repeated here.

Indexing a global diversity baseline through category evenness

In 1981 there were 614,608 articles and reviews (academic research papers) in the Web of Science. This annual count of papers rose to exceed one million for 1994, two million for 2011 and reached 2,622,355 for 2020. The database expanded in Engineering and technology, from 8.4% in 1981 to 17.1% (the biggest relative growth is in Nanosciences), Environmental (5.1 to 8.6%) and Mathematical sciences (3.2 to 4.1%). There were relative falls in Clinical sciences (down from 22.3 % to 18.3%, notably in General & internal medicine) and Biological sciences (20.7% to 14.8%). We can visualize changing annual diversity with a Lorenz curve (shown in Adams et al 2020a: Figure 3) but this valuable 'first impression' provides no analytical power and its illustrative value breaks down with large numbers of years and/or countries. A summary index such as the Gini coefficient is therefore necessary for any large scale or complex comparative analysis.

Change indexed via the Gini coefficient is small, as might be expected for counts of millions of papers per year spread across 248 categories. Global evenness increases slightly from 1981 to 2008 and then plateaus. It may recently have shifted back towards specialization, although the significance of this inflexion will need re-evaluation later. This stability – during a more than four-fold expansion – confirms the balance and value of the Web of Science database as the reference point for diversity analysis. This is shown as a reference baseline in Figure 1.

The overall global pattern is, as noted, the summed outcome of many national changes, themselves a synthesis of local and institutional decisions, driven by change in disciplinary opportunities and priorities. Furthermore, the index of evenness is not an index of dynamic change: the same index value can be produced by many combinations of counts across categories. When we examine individual entities, we do so in a generic global context and, while we might expect that all countries and institutions share a common component towards less specialized portfolios, we do not in practice know how much the peaks and troughs of the underlying distribution may have changed.

Normalizing national data against a global benchmark

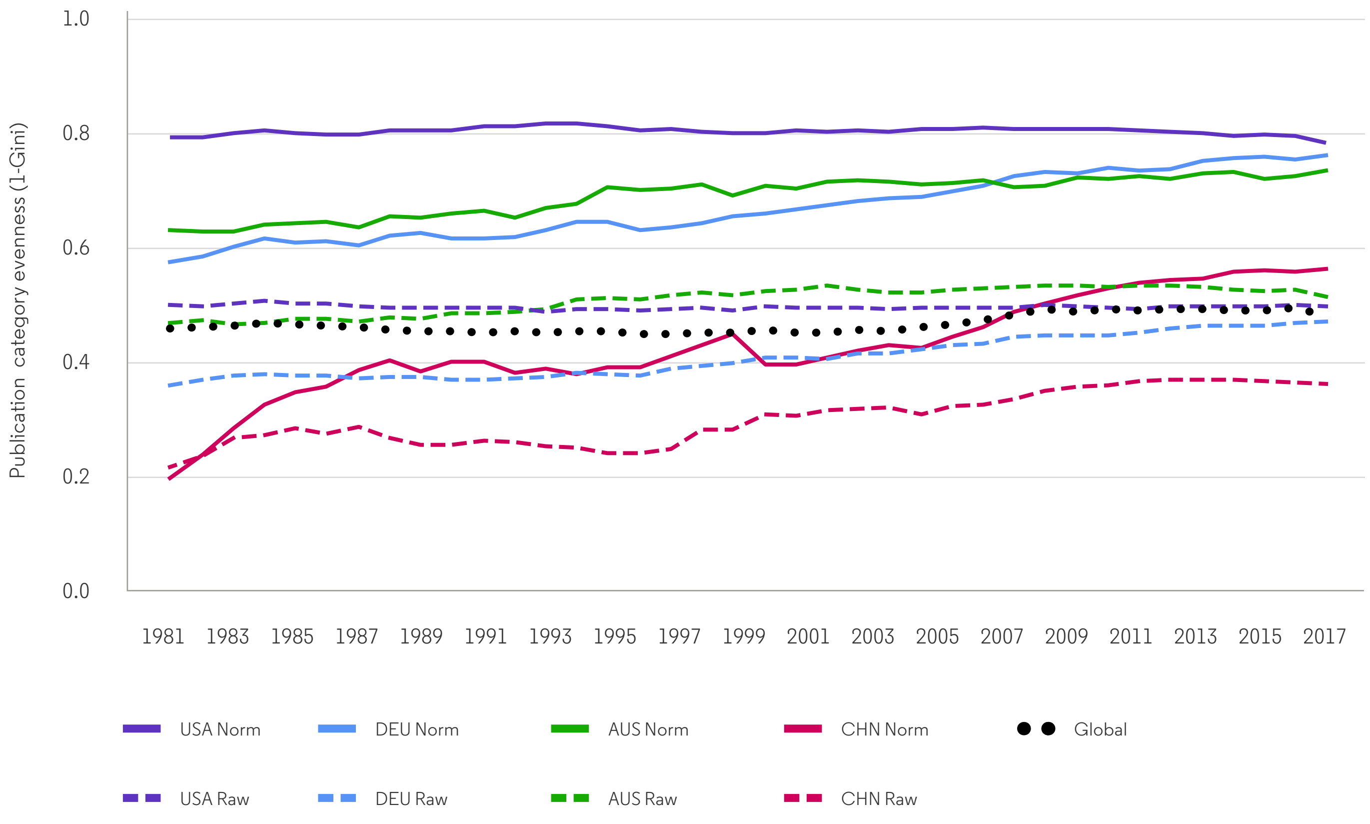
We can now compute the evenness of national publication portfolios on a standardized basis using the Web of Science global baseline and the constant set of categories for every year from 1981. We can compare the relative paper counts for each country with that baseline, by 'normalizing' the annual proportion of a country's paper counts in each category against the relevant proportion for the world.

To understand both the difference between global and national patterns and the effect of normalization, we start by comparing the results of analyses using the original, raw paper counts with the results of the analysis after these raw counts have been normalized.

We selected a global spread of four countries: the United States, as the dominant publisher in 1981; Germany, as part of the E.U. region; Australia, as an established economy with existing global links; and Mainland China, as a powerful research economy emerging over the period of analysis. (Figure 1)

The normalized data reveal greater publication evenness for every country than an index based on the raw paper counts would suggest.

Figure 1.
Evenness of publication portfolios for the United States (USA), Germany (DEU), Australia (AUS) and Mainland China (CHN) for 1981 – 2018. Data are indexed by the Gini coefficient across 248 Web of Science journal categories and displayed as (1-Gini). Comparison may be made between the global baseline and the curves of index values using raw paper counts by category (dashed lines) and the values when the raw counts are normalized against the global baseline (solid lines).



*ISO 3166-1 alpha-3 codes are used for country/region abbreviations in all figures

This is because of the innate variance in the scope and size of the different journal categories. The Gini index for the original data on U.S. paper counts follows a trend and value that is similar to the global average, reflecting the dominance and influence of the U.S. research system. However, when the data are normalized it becomes apparent, first, that the United States is more even in its portfolio than the global average and, second, that it has become slightly more specialized over the period (publication evenness declines after 1990).

German research output, as indexed in the Web of Science, becomes progressively more even throughout the period reaching a similar index value to the United States by 2015. It should be noted that whereas its research portfolio is less even than Australia on raw data, it becomes more even on the normalized data. Australia's evenness appears to change slowly on the raw data but the normalized index reveals a clear step in the early 1990s, which is likely to be a response to the Dawkins Higher Education reforms (Dawkins 1988). The raw and normalized curves for Mainland China follow similar profiles, but the normalized data reveal a much greater shift than the raw data might suggest and by the end of the period its evenness approaches that of the other three.

We have compared the evenness index values obtained from raw and normalized paper counts to show the importance of using a reference baseline, but for practical purposes it is variance from the global base that is likely to be most useful as a guide to relative specialism and evenness. The rest of this analysis is therefore based on national and institutional publication data after normalization against the same set of global category counts.

International trends in research evenness, diversity and similarity

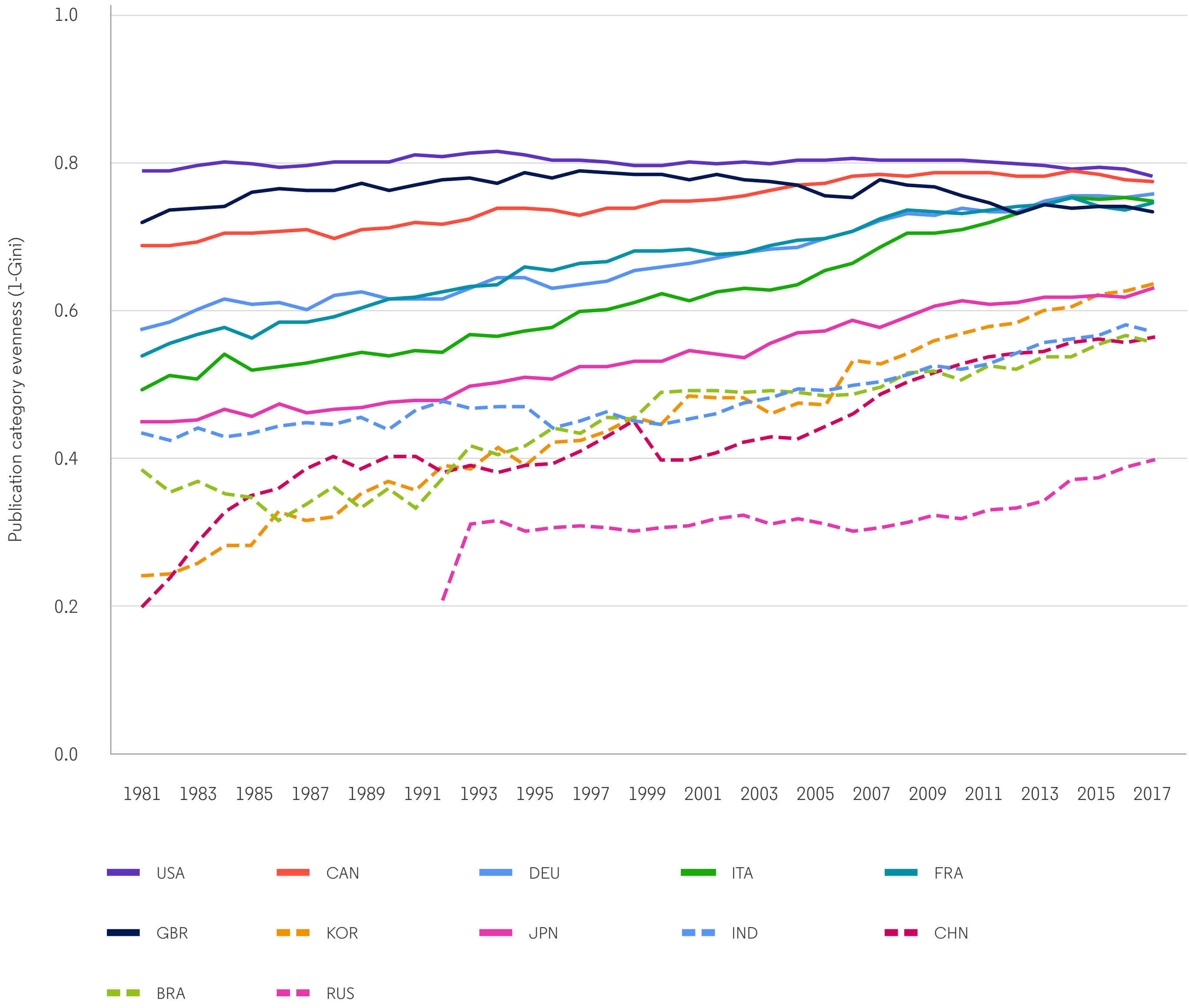
Our analysis draws on data for two groups of countries: the G7 and the BRICKs. In 1981, the global research base was dominated by the G7 group of advanced research economies (Canada, France, Germany, Italy, Japan, the United Kingdom and the United States), which accounted for around 70% of all outputs then indexed by the Web of Science. By 2018, this was no longer the case and these countries accounted for less than half of indexed outputs. The United States was by far the largest economy and had enjoyed a relatively high level of sustained R&D investment since 1945, but its share of global output fell from a dominant 36% in 1981 to a little over 25% in 2018.

Much of the change in global diversity (not volume) might be attributable to the rise of the BRICK economies (Brazil, Russia, India, Mainland China and South Korea) and other Asian nations (Adams and Wilsdon 2006; Bound et al 2013). For the others, their public-sector research economy in the 1980s was far smaller and less advanced than the G7 but Mainland China is now set to overtake the U.S. as the most prolific research publisher. The BRICKs therefore provide an informative contrast to the G7 in their research growth and management over the last four decades. (Figure 2)

Unquestionably, the near-universal trend is towards greater national evenness in the spread of papers across subject categories. The analysis shows that the Anglophone G7 (United States, United Kingdom and Canada) had the most evenly distributed publication portfolios in the 1980s and that the research output of their E.U. partners (France, Germany and Italy) has changed to match them.

Unquestionably, the near-universal trend is towards greater national evenness in the spread of papers across subject categories.

Figure 2.
Evenness of publication portfolios for the G7 (solid lines) and BRICK (dashed lines) groups of countries (1981 – 2018) indexed by the Gini coefficient for publication data. The index is calculated on national paper counts across 248 Web of Science journal categories where raw counts are normalized against the global baseline for year and category and displayed as (1-Gini).



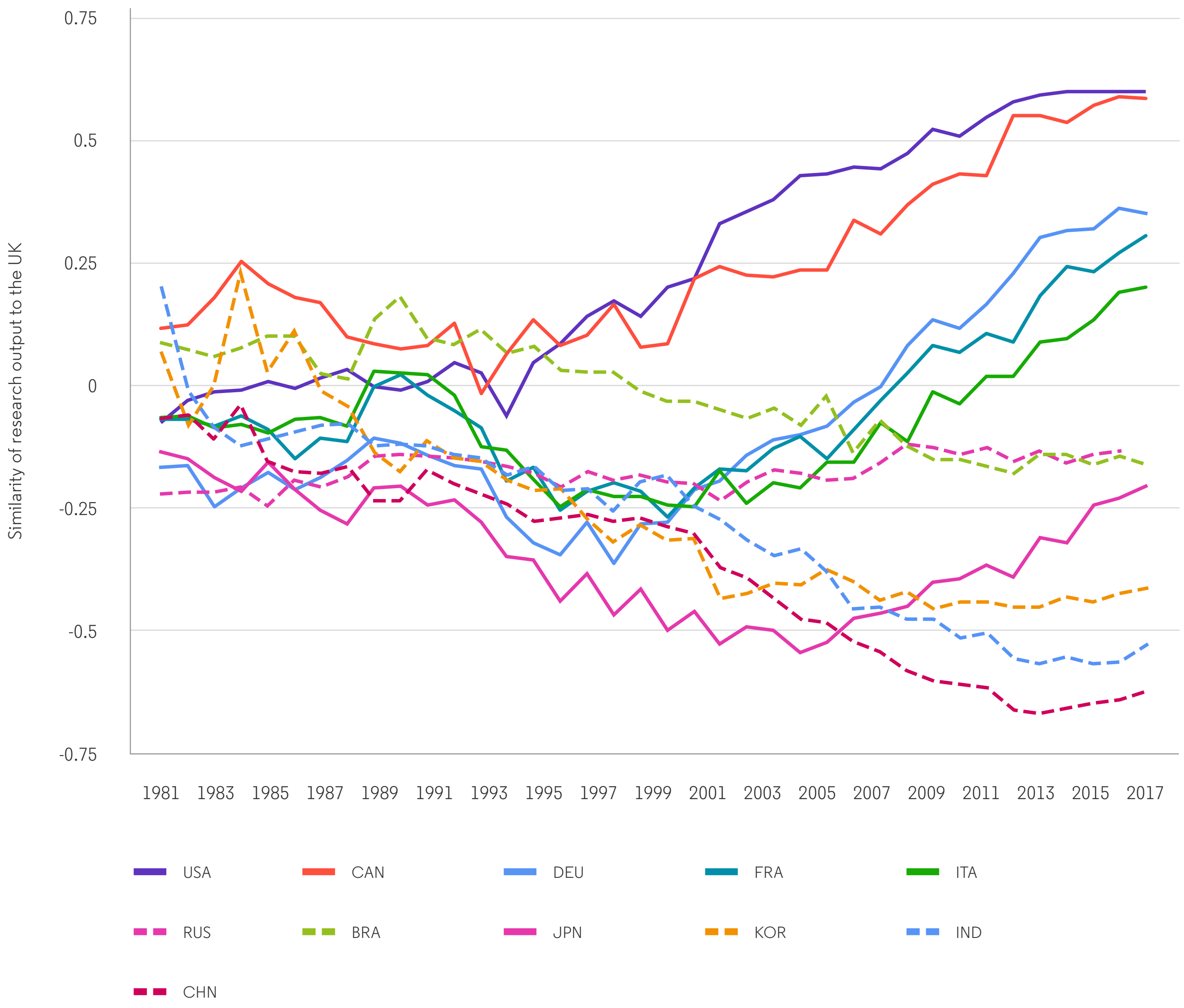
This is likely to reflect the shift of national research publication into 'international' Anglophone journals and a general increase in international research collaboration. The only countries that show any increase in selectivity are the United States and the United Kingdom, which were the most even in 1981, so the group appears to be converging on and possibly plateauing at a similar level of portfolio evenness (around 1-Gini=0.8), excepting Japan which has lagged this trend. The United Kingdom dynamics stand out as a saw-tooth of dwindling diversity, which may be an effect of its assessment system.

The countries with the least diverse research portfolios in 1981 are Mainland China and South Korea and their subsequent shift to greater evenness is similar. South Korea has overtaken Japan in this regard and seems likely to shift towards the G7 profile. India was similar in evenness to Japan but has in fact evolved very little over 40 years, whereas Brazil showed great dynamism up until 2000 and rather less change since. Brazil, Mainland China and India converge on similar levels of diversity, but at a more specialist level than the G7 (around 1-Gini=0.5).

The same value of an index such as the Gini coefficient can be produced by more than one set of data points, so two countries with the same indexed evenness may not be even in the same way. We can, however, test the similarity of these distributions and see how they have converged over time. Since displaying these would inevitably produce many lines (each being a pairwise track of similarity over time) we have simplified the presentation here to a single example, and we have chosen the United Kingdom as the focus of the graph. The analysis is simply an annual correlation between the paper count by Web of Science journal category for the United Kingdom and that of another country. (Figure 3)

The research portfolios of the United Kingdom and its major G7 partners have become increasingly similar since 1981, but only after a dip in U.K. similarity to its European Union partners prior to 2000. Its similarity to the BRICK group has however, declined over the same period which suggests that the difference in evenness of their research output (Figure 3) reflects increasing divergence in their research foci.

Figure 3.
Correlation between the distribution across Web of Science journal categories of the proportion of the United Kingdom's annual output and that of other countries in the G7 group and in the BRICK group.



Diversity and research response

The countries with more even research bases, especially the U.S., Germany and the U.K., tend to support a response across a wider range of COVID-19 topics.

The evidence from economics and ecology is that diversity provides benefits, notably in resilience and responsiveness. The global pandemic caused by the novel-coronavirus COVID-19 must represent one of the most severe challenges to a national research base and its capacity to respond to citizen needs. The question is therefore whether the data already available throw any light on the relationship between research response and prior research diversity.

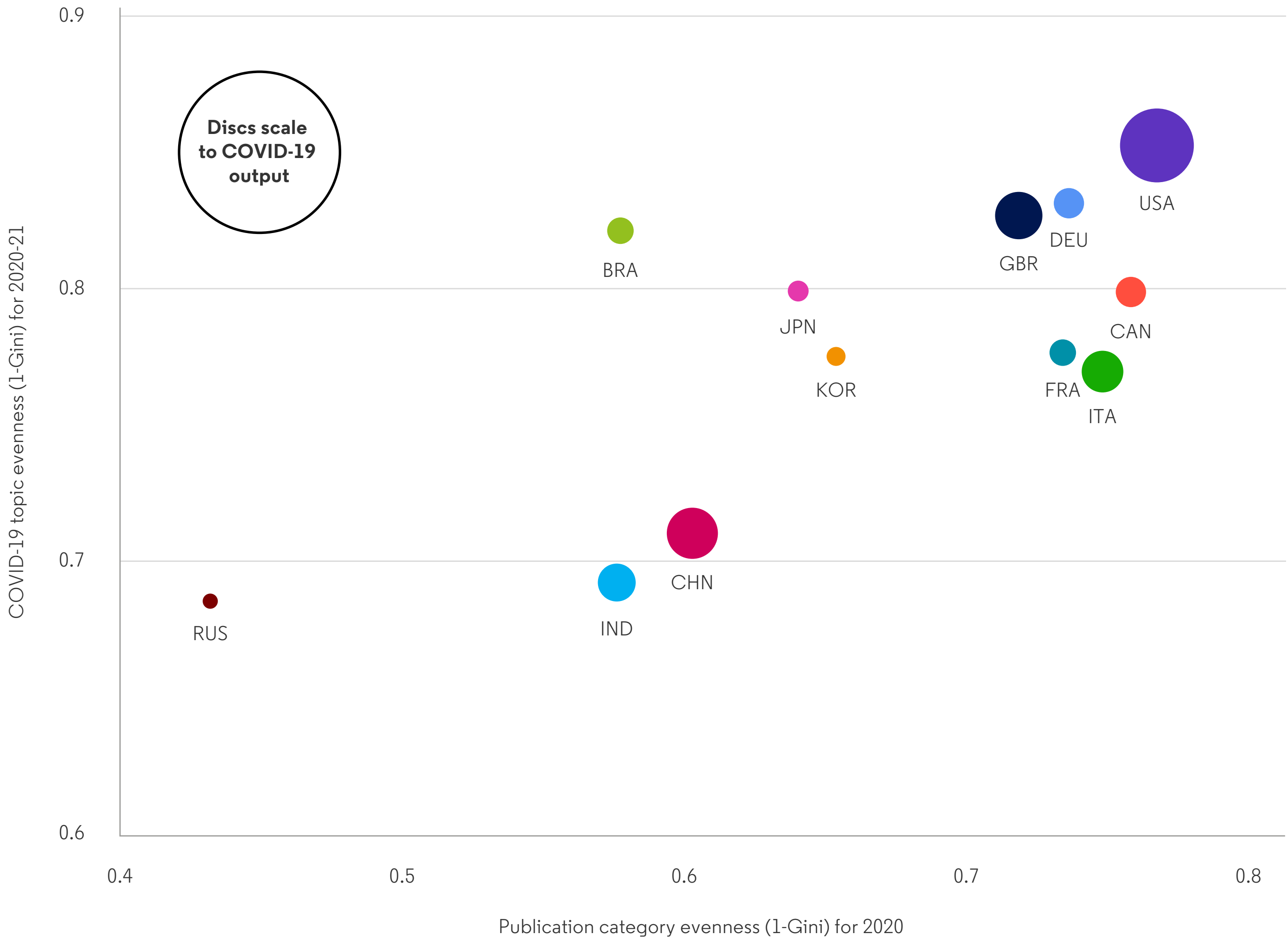
Our methodology for measuring subject diversity uses the established Web of Science journal categories – but this is not suitable for analyzing research relating to COVID-19. During the pandemic, innovative research topics emerged to tackle societal, economic and health issues that often draw on contributions from a broad spectrum of academic disciplines and are not aligned with conventional subject categories. Consequently, we turn to topic modelling to create a bespoke classification system of COVID-19 research that can distinguish the research topics addressing these themes (see also Adams et al., 2020b).

We identified 67,756 papers (articles or reviews) indexed in the Web of Science and published in 2020 – 21 that are related to COVID-19. We did this by searching titles, abstracts and keywords for terms such as 2019-nCoV, COVID-19, SARS-COV-2, novel coronavirus, etc.

Text from the titles and abstracts of these papers was used to create a topic model with 40 topics. These covered areas of clinical practice, molecular biology, virology, immunology, epidemiology, virtual learning, mental health, food security, economics, crisis management, environmental impact, and so on. This model was then used to profile nations according to the number of papers produced in each topic, revealing the range of the response and providing a mechanism for comparison. For each country, the relative paper count for each topic (i.e., the number of papers in each topic with an author from the country divided by the total number of papers in the topic) was used to calculate a Gini coefficient, and so provide a measure of relative evenness (1-Gini).

We can now compare the evenness of each national research base in 2020 (from Figure 2) with the evenness of that country's research on COVID-19 (Figure 4). The countries with more even research bases, especially the U.S., Germany and the U.K., tend to support a response across a wider range of COVID-19 topics. Brazil has a relatively high evenness for COVID-19 papers when compared to India and Mainland China although these three have a similar level of evenness in their research portfolio. While all three had a substantial output in the core clinical topics, Brazil also published COVID-19 research in areas that were less prominent in Mainland China and India's COVID-19 portfolios, such as online learning, economics and digital media. Russia, with a significantly less even research portfolio, nonetheless achieved a similar topic evenness in COVID-19 outputs to India and Mainland China.

Figure 4.
Comparison between the evenness of national publication portfolios in 2020 and that of COVID-19 publications for the G7 and BRICK countries shown in Figure 2. Points are scaled according to the volume of COVID-19 output (e.g., USA 18,578, Mainland China 9,171, Brazil 2,397, South Korea 1,213).



A preliminary examination of the data on research responses to the current COVID-19 pandemic suggests that a diverse research base is indeed of potential benefit in enabling a more comprehensive response.

Institutional trends in research evenness and diversity

The most informative analysis is one that compares the institution with its peers over time.

Web of Science data provide an effective and stable benchmark and source from which publication data for any country can be collated and analyzed to assess and compare – across time or across borders – the evenness of their research portfolio as an indicator of their research diversity (Figure 2). A preliminary examination of the data on research responses to the current COVID-19 pandemic suggests that a diverse research base is indeed of potential benefit in enabling a more comprehensive response (Figure 4).

If research diversity is beneficial then research managers will want to know how their institution's research portfolio shapes up. Possible points of comparison include: the global benchmark, a national 'average', comparison with peer competitors and the organization's own timeline.

The global benchmark is essential for standardizing normalization but is uninformative for interpretation. As Andy Stirling (2007) noted, diversity is more or less, not yes or no, so there is no threshold for good or poor diversity.

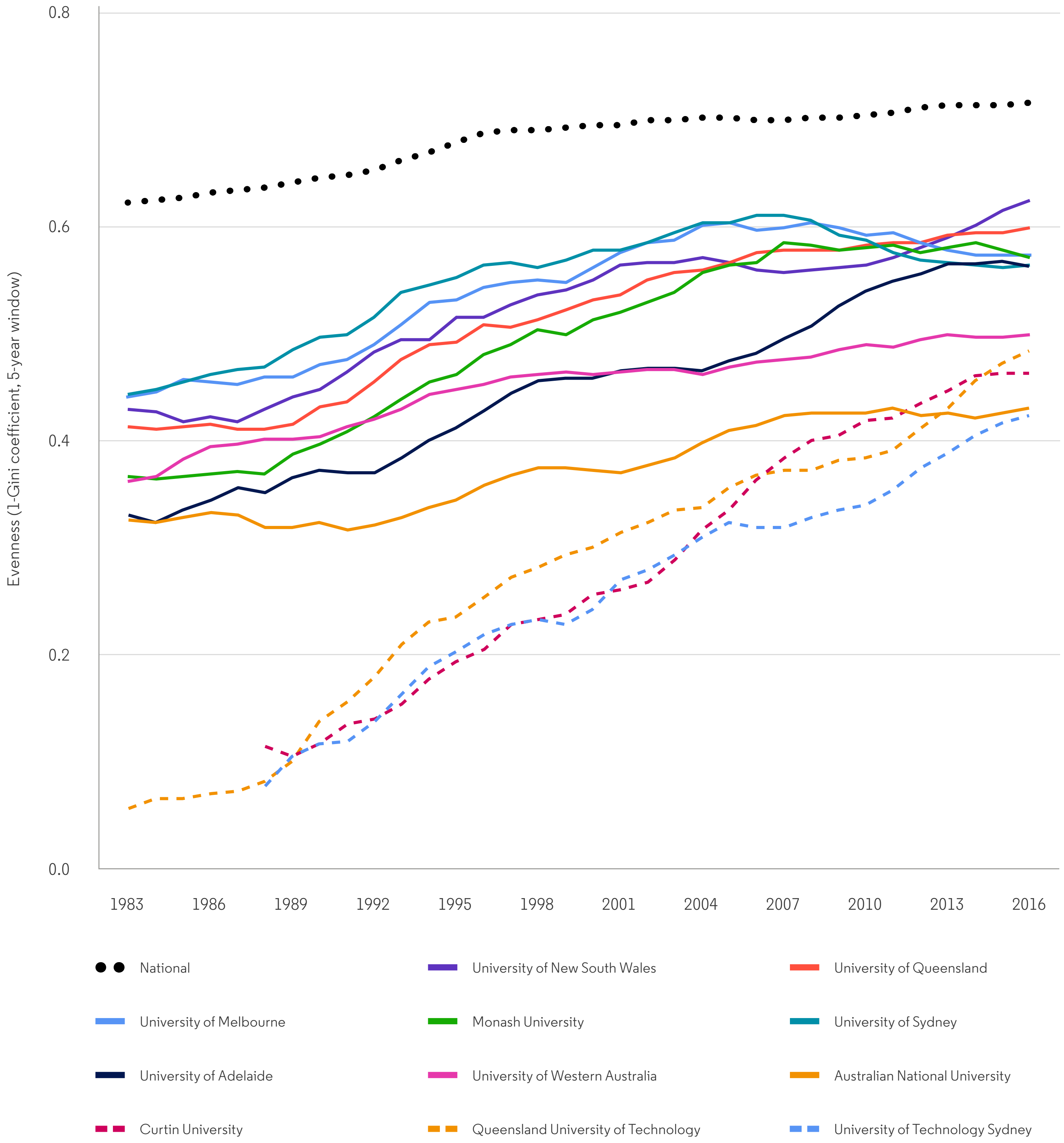
National profiles are not necessarily the right reference point either. The national profiles described in the earlier sections are not averages in the institutional context. Each is a product of bringing together many different institutional portfolios to synthesize a combined national data set. That means the national profile makes an interesting reference point, but each institution is likely to be more specialist and less even than the overall national pot.

The most informative analysis is therefore one that compares the institution with its peers over time. We analyzed publication data for a spread of universities in Australia (Figure 7) and the U.K. (Figure 8). The universities shown in these graphs were selected because they represent a spread of histories, investment and mission while most are relatively multi-faculty. We have omitted, for example, specialist arts institutions that would be better compared with one another. In each graph, we have included the overall national profile to confirm the point that it is not an average of the component parts.

The Australia sample includes the Group of 8 (Go8), already well established in 1981, and three relatively new institutions. Most of the Go8 tracked upwards in diversity (from 1-Gini around 0.4) into the 2000s, but the most diverse members appear to shift towards more specialized portfolios in the last decade. There also appears to be a plateau in evenness where Gini reaches 0.4 (1-Gini somewhat less than 0.6), which mimics the pattern but not the value seen at country level (Figure 4). Adelaide has a steeper rise from a more specialized position in the 1980s and plateaus as it reaches the diversity of the main group.

Exceptions to this trajectory are the University of Western Australia, which parallels the national profile and is relatively unchanged after the shifts of the 1990s, and the Australian National University, which has a stable and much lower level of diversity than the rest of the Go8. (Figure 5)

Figure 5.
Research publication diversity (indexed as evenness via a Gini coefficient of publication counts across Web of Science categories) for higher education institutions in Australia.



The emergence and development of the new generation of institutions is clear from the three selected as examples in this analysis. From a much more concentrated portfolio they increase in diversity at about the same rate and up to a similar level by the late 2010s.

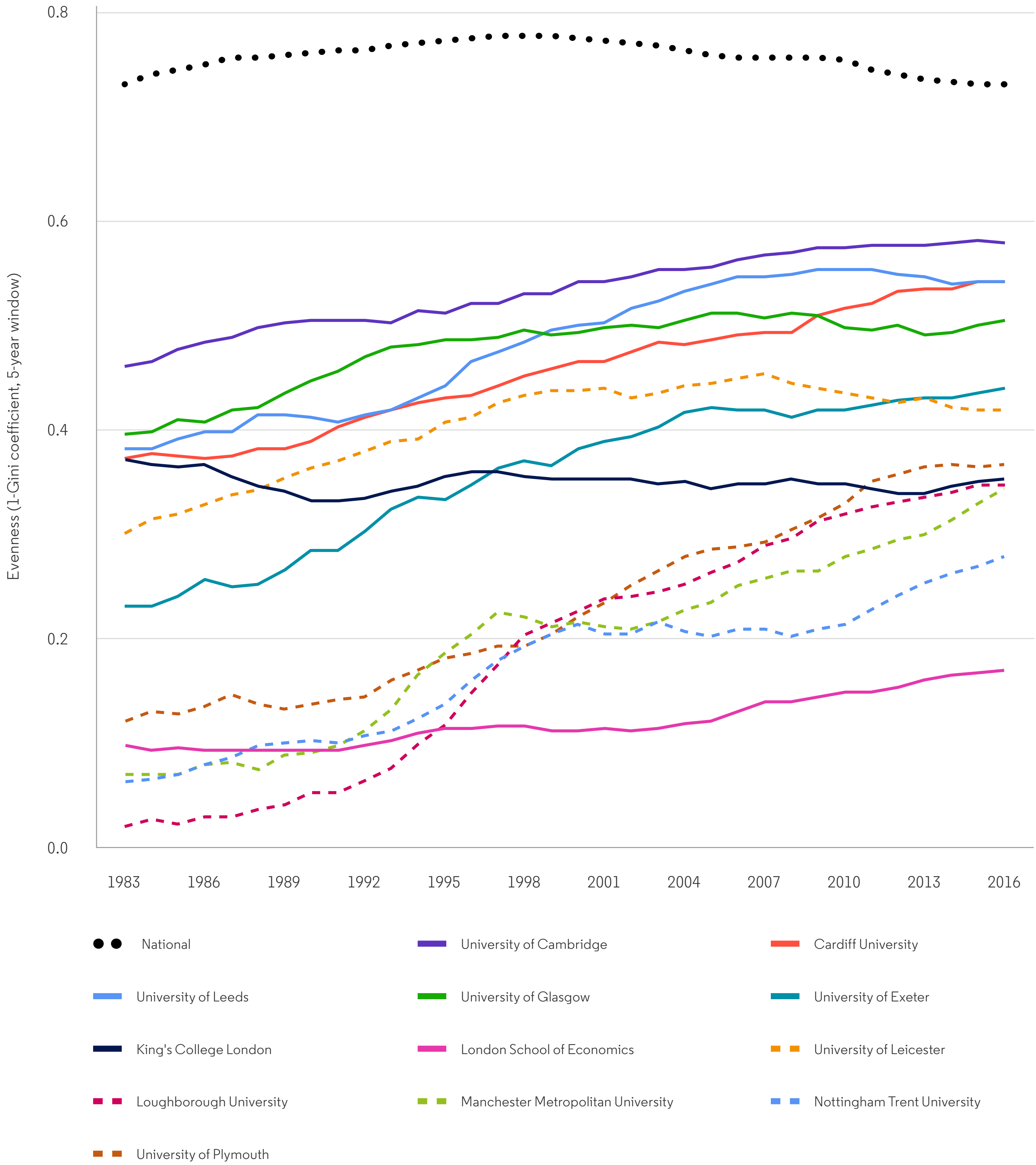
The evolution of the U.K. system has much in common with the Australian picture. The ranked hierarchy remains much the same with, again, some notable exceptions and the newly emerging institutions follow a steep rise in diversity. There again seems to be a plateau in diversity (1-Gini=0.6) and some retrenchment in the last few years. (Figure 6)

Institutions of particular interest include King's College London, which has a small drift downwards in diversity. However, the College absorbed several large London medical establishments in the 1990s and this likely shifted the balance of its activity. The relative change of the University of Exeter, which overtakes the University of Leicester in its rising diversity, is also a reflection of strategic change as this institution focused on diversity through international partnerships that enabled it to both broaden and strengthen its portfolio and develop innovative interdisciplinary capacity. Loughborough University's significant increase in diversity during the 1990s is a further example of change led by insightful management response to national policy on research funding. The London School of Economics's position also confirms the sense of the index, in that this well-known specialist institution has an appropriately specialized research portfolio.

At the end of the period there remains a greater separation and spread of diversity across institutions than for Australia with some relatively large U.K. universities retaining a lower subject diversity than the Australian examples. There are thus common trends and distinct national flavours to the outcome.

The evolution of the U.K. system has much in common with the Australian picture. The ranked hierarchy remains much the same with some notable exceptions and the newly emerging institutions follow a steep rise in diversity.

Figure 6.
Research publication diversity (indexed as evenness via a Gini coefficient of publication counts across Web of Science categories) for higher education institutions in the U.K.



Discussion and conclusions

We have shown how a simple index of the evenness of activity across a research portfolio can be developed from readily available data by reference to the stable, comprehensive, global database available in the Web of Science. Evenness is a key component of diversity and in this database the other key components (variety and disparity) are constant. The diversity of publication data is a familiar and intuitive signal of diversity for most researchers and managers, corresponding to information they see and use daily.

The index, relying on the familiar Gini coefficient, rapidly provides national and institutional information. It enables rapid visual comparisons to be made over time and between entities. These comparisons are immediately informative, make sense in terms of other information about these entities, and reveal new facets of change as well as confirming other assumptions. The same data that generate the index of evenness can be repurposed to compare research similarity and reveal further information about evolving portfolios.

However, indexing diversity would only be of esoteric interest if it were not also related to other aspects of research performance. The COVID-19 topic analysis shows us how nations with a more diverse research base respond more comprehensively to an unprecedented scientific challenge. Diversity provides agility and the scope for recombining knowledge in unexpected interdisciplinary situations.

Many recent studies have asserted a link between the capacity for interdisciplinary research and the quality of research outcomes. If diversity does help to foster interdisciplinarity, and interdisciplinarity appears to be a feature of many innovative research programs, then the trends and comparisons should immediately be of interest to the institutions we analyzed here, and to others. If diversity is linked to responsiveness, then that is also an important strategic consideration at national policy level. And there are evidently more facets to explore in the comparisons between the national systems, in their similarities and differences.

The reference data underpinning our diversity calculations will be made available through our consultancy teams, who can also advise on the best way of developing informative comparisons. Bibliometric data have been thoroughly explored as a tool for tracking past performance and outcomes. Understanding and assessing diversity may prove not only to be a useful forward-looking index for research organizations but also to be a critical tool for managers in anticipating and preparing for the unexpected and then being able to deploy an effective response.

Diversity provides agility and the scope for recombining knowledge in unexpected interdisciplinary situations.

References and background reading

Adams, J., Rogers, G., Smart, W., & Szomszor, M. (2020a). Longitudinal variation in national research publication portfolios: Steps required to index balance and evenness. Quantitative Science Studies, 1(3), 1182–1202. https://doi.org/10.1162/qss_a_00073

Adams, J., Rogers, G., & Szomszor, M (2020b). The Annual G20 Scorecard – Research Performance 2020. Clarivate Global Research Report. ISBN 978-1-8382799-0-5

Adams J., & Smith D. (2003). Funding research diversity. A report to Universities UK, pp 114. ISBN 1 84036 103 4. http://www.universitiesuk.ac.uk/Publications/Pages/Publication-196.aspx

Aston, I., & Shutt, E. (2009). Concentration and diversity: understanding the relationship between excellence, concentration and critical mass in UK research. University Alliance, research paper 2009/01. ISBN 978-1-908190-04-8. https://www.unialliance.ac.uk/wp-content/uploads/2011/05/Publication_Research_Concentration_and_Diversity.pdf

Carpenter, M. P. (1979). Similarity of Pratt's measure of class concentration to the Gini index. Journal of the American Society for Information Science, 30(2), 108-110. https://doi.org/10.1002/asi.4630300210

Dawkins, J. S. (1988) Higher Education: a policy statement, 170 pp. Canberra, Australian Government Publishing Service. ISBN 0 644 08300 X

European Central Bank. (2021). Economic diversity. https://www.ecb.europa.eu/mopo/eaec/diversity/html/index.en.html

Finardi, U., & Lamberti, C. (2021). Diversity and interdisciplinarity in nanoscience and nanotechnology: a time-related analysis of the subject category. Journal of Nanoparticle Research, 23(1): article number 7. https://doi.org/10.1007/s11051-020-05119-8

Gamfeldt, L., Snäll, T., Bagchi, R. et al. (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. Nature Communications, 4, 1340. https://doi.org/10.1038/ncomms2328

Gini, C. (1909). Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza. Giornale degli Economisti, 11, 37.

Hackett, E.J., Leahey, E., Parker, J.N., Rafols, I., Hampton, S.E., Corte, U., Chavarro, D., Drake, J.M., Penders, B., Sheble, L., Vermeulen,N. and Vision,N.J. (2021). Do synthesis centers synthesize? A semantic analysis of topical diversity in research. Research Policy, 50(1), 104069. https://doi.org/10.1016/j.respol.2020.104069.

Huang, Y., Glänzel, W., Thijs, B., Porter, A. L., & Zhang, L. (2021). The comparison of various similarity measurement approaches on interdisciplinary indicators. Working Papers of the Department of Management, Strategy and Innovation, Leuven 670612, KU Leuven, Faculty of Economics and Business (FEB), Department of Management, Strategy and Innovation, Leuven. https://lirias.kuleuven.be/3394973?limo=0

Janavi, E., Mansourzadeh, M.J., & Samandar Ali Eshtehardi, M. (2020). A methodology for developing scientific diversification strategy of countries. Scientometrics, 125, 2229–2264. https://doi.org/10.1007/s11192-020-03685-1

Leinster, T. and Cobbold, C.A. (2012). Measuring diversity: the importance of species similarity. Ecology, 93, 477-489.https://doi.org/10.1890/10-2402.1

Leydesdorff, L. (2018). Diversity and interdisciplinarity: How can one distinguish and recombine disparity, variety, and balance? Scientometrics, 116(3): 2113-2121. https://doi.org/10.1007/s11192-018-2810-y

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019a). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. Journal of Informetrics, 13(1), 255-269. https://doi.org/10.1016/j.joi.2018.12.006

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019b). Diversity measurement: Steps towards the measurement of interdisciplinarity? Journal of Informetrics, 13(3), 904-905. https://doi.org/10.1016/j.joi.2019.03.016

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. Quarterly Publications of the American Statistical Association, 9(70), 209-219.

May, R.M. (1997). The scientific wealth of nations. Science, 275, 793-796.

Moed, H. F. (2006). Bibliometric ranking of world universities. CWTS Report 2006-01, pages 1-36.

Nijssen, D., Rousseau, R. and Hecke, P. van. (1998). The Lorenz curve: a graphical representation of evenness. Coenoses, 13(1), 33-38.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. Scientometrics, 82(2), 263-287. https://doi.org/10.1007/s11192-009-0041-y

Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. Sankhyā: The Indian Journal of Statistics, Series A, 44(1), 1-22.

Rousseau, R., Van Hecke, P., Nijssen, D., & Bogaert, J. (1999). The relationship between diversity profiles, evenness and species richness based on partial ordering. Environmental and Ecological Statistics, 6(2), 211-223. https://doi.org/10.1023/A:1009626406418

Rousseau, R., Zhang, L., & Hu. X. (2019a). Knowledge integration: Its meaning and measurement. In Glänzel, W., Moed, H. F., Schmoch, U., & Thelwall, M. (eds.) Springer Handbook of Science and Technology Indicators. Cham, Switzerland: Springer, pp. 69-94.

Rousseau, R. (2019b). On the Leydesdorf-Wagner-Bornmann proposal for diversity measurement. Journal of Informetrics, 13(3), 906-907. https://doi.org/10.1016/j.joi.2019.03.015

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. Interface, a Journal of the Royal Society, 4, 707–719. doi: 10.1098/rsif.2007.0213

Thijs, B., Huang, Y., & Glänzel, W. (2021). Comparing different implementations of similarity for disparity measures in studies on interdisciplinarity. Working Papers of Department of Management, Strategy and Innovation, Leuven 670614, KU Leuven, Faculty of Economics and Business (FEB), Department of Management, Strategy and Innovation, Leuven. https://ideas.repec.org/p/ete/msiper/670614.html

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. PloS ONE, 10(5), e0127298. https://dx.doi.org/10.1371%2Fjournal.pone.0127298

Yegros-Yegros, A., Rafols, I. and D'Este, P. (2015) Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. PLoS ONE 10(8): e0135095. doi:10.1371/journal.pone.0135095

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. Journal of the Association for Information Science and Technology, 67(5), 1257-1265. https://doi.org/10.1002/asi.23487

Zhang, L., & Leydesdorff, L. (2021). The scientometric measurement of interdisciplinarity and diversity in research portfolios of Chinese universities. Preprint. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3798519

About the Global Research Report series from the Institute for Scientific Information (ISI)

Our Global Research Reports draw on our unique industry insights to offer analysis, ideas and commentary to enlighten and stimulate debate.

Each one demonstrates the huge potential of research data to inform management issues in research assessment and research policy and to accelerate development of the global research base.

Advice on the use of the standard methodology and information about comparative institutional analyses used in this report is available.

e: ISI@clarivate.com

Previous reports include:

Profiles not metrics: Beyond single point metrics

Research integrity: Understanding our shared responsibility for a sustainable scholarly ecosystem

Identifying Research Fronts in the Web of Science: From metrics to meaning

Navigating the structure of research on Sustainable Development Goals

Download here:

www.webofsciencegroup.com/isi

About Clarivate

Clarivate™ is a global leader in providing solutions to accelerate the lifecycle of innovation. Our bold mission is to help customers solve some of the world's most complex problems by providing actionable information and insights that reduce the time from new ideas to life-changing inventions in the areas of science and intellectual property. We help customers discover, protect and commercialize their inventions using our trusted subscription and technology-based solutions coupled with deep domain expertise. For more information, please visit clarivate.com.

The Web of Science™ organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the world's largest publisher-neutral citation index and research intelligence platform.

webofsciencegroup.com/isi

XBU699116795 / 01