The MIT Press

RESEARCH ARTICLE

A half-century of global collaboration in science and the "Shrinking World"

Keisuke Okamura[1,2] iD

[1]The University of Tokyo
[2]National Graduate Institute for Policy Studies (GRIPS)

ABSTRACT

Recent decades have witnessed a dramatic shift in the cross-border collaboration mode of researchers, with countries increasingly cooperating and competing with one another. It is crucial for leaders in academia and policy to understand the full extent of international research collaboration, their country's position within it, and its evolution over time. However, evidence for such world-scale dynamism is still scarce. This paper provides unique evidence of how international collaboration clusters have formed and evolved over the past 50 years across various scientific publications, using data from OpenAlex, a large-scale open bibliometrics platform launched in 2022. I first examine how the global presence of top-tier countries has changed in 15 natural science disciplines over time, as measured by publication volumes and international collaboration rates. Notably, I observe that the United States and China have been rapidly moving closer together for decades but began moving apart after 2019. I then perform a hierarchical clustering to analyze and visualize the international collaboration clusters for each discipline and period. Finally, I provide quantitative evidence of a "Shrinking World" of research collaboration at a global scale over the past half-century. My results provide valuable insights into the big picture of past, present, and future international collaboration.

1. INTRODUCTION: LANDSCAPE OF INTERNATIONAL COLLABORATION REVISITED

With the rapid evolution of advanced digital communication platforms, the world has entered a new era of intense competition for knowledge. The volume of information being produced, the degree to which it is integrated and utilized, and the range and depth of the community involved at the interface between open science and highly digitized society are all growing exponentially (Beck, Bergenholtz et al., 2022; Burgelman, Pascu et al., 2019; Dong, Ma et al., 2017; Miedema, 2022; Wittenburg, 2021). There is no doubt that the power of science and technology (S&T) to explore the knowledge frontier is the key to innovation, driving national growth and international competitiveness. As a result, policymakers in many countries have begun to pay significant attention to the "science of science" (Fortunato, Bergstrom et al., 2018), which quantifies and investigates all activities associated with S&T, providing valuable insights for policymaking. Bibliometricians with expertise in describing and evaluating the research communities' activities through S&T-related metrics are increasingly collaborating with policymakers and institutional practitioners, influencing the policymaking process (Cabezas-Clavijo & Torres-Salinas, 2021; Doria Arrieta, Pammolli, & Petersen, 2017; Hicks,

Wouters et al., 2015; Institutes of Science and Development of the Chinese Academy of Sciences, n.d.; Ismail, Nason et al., 2012; National Science Board & National Science Foundation, 2021; OECD, n.d.; OECD & SCImago Research Group, 2016; Wagner & Jonkers, 2017; Wilsdon, Allen et al., 2015).

The bibliometric method provides a systematic, quantitative, and objective overview of information about researchers, research institutions, and venues of research results publication (such as journals, conferences, institutional repositories, preprint servers, and book chapters). It also provides numerous metrics and evaluation indicators derived from the information. Indeed, bibliometric approaches have various conceptual and methodological difficulties and limitations, and their use requires careful attention in practical application (Haustein & Larivière, 2015; Hicks et al., 2015; Waltman, 2016; Wilsdon et al., 2015). However, to understand things at some macrolevel, bibliometrics can provide a unique lens through which people can see the lively ecosystem of scholarly communications. This observation is even more true today, when accessing and utilizing big data through various open platforms is increasingly possible.

Policymakers in many countries recognize the significance of bibliometrics in making decisions on research and development (R&D) investment portfolios at the state or institutional level. An increasingly important issue of global interest in the policy arena is how to address international research collaboration (Adams, 2012; Chen, Zhang & Fu, 2019; Dusdal & Powell, 2021; Kwiek, 2023; Sloan & Alper, 2018; Wagner & Jonkers, 2017). On the one hand, collaboration across countries is essential for large-scale academic R&D projects, such as accelerator science and earth and planetary science, which require extended R&D costs and periods. It is also indispensable in addressing world-scale issues such as the United Nations Sustainable Development Goals or responding to global crises such as the COVID-19 pandemic (Maher & Van Noorden, 2021; National Science Foundation, 2020). On the other hand, geopolitical aspects, such as economic security and defense-related R&D, set other primary boundary conditions for proceeding with international collaboration. Therefore, the policy environment surrounding the frontiers of international research collaboration has become increasingly complex and challenging in recent years, both scientifically and geopolitically.

Amid these multilayered dimensions of cooperation and competition in international research collaboration, policymakers are keen to understand their country's collaboration partners and the clusters of research collaborations around the world (Adams & Gurney, 2018; Adams, Gurney, & Marshall, 2007; He, 2009; Kwiek, 2023; Mattsson, Laget et al., 2008; OECD, 2017; OECD & SCImago Research Group, 2016; Yuan, Hao et al., 2018). Knowing how the composition of international collaboration clusters of interest has changed over time and which policies are effective drivers of that change, whether for better or worse, is undoubtedly advantageous for a country. However, it is a challenging task for various reasons. A significant challenge from the policymaker's perspective is data availability and accessibility, which involves at least three aspects. First, data coverage matters. Commercial databases that focus on journal articles often do not cover enough data in today's rapidly changing, highly digitized R&D world, where considerable scholarly communication occurs outside journals (Supplementary Figure S2 in Okamura, 2022). Furthermore, there is a field-dependent bias concerning data coverage. For instance, in computer science, conferences and preprint servers have been more common venues for publishing research results than journals (Kim, 2019). If there are substantial biases or inconsistencies in data coverage, the credibility of the results of analyses based on such data will be flawed.

The second challenge concerns the timing of data availability. For policymakers and practitioners, planning and executing R&D investments at the right time is of utmost importance.

However, the results of bibliometric analyses based on a database that only covers journal articles may provide information too late for policymakers to reflect on their policies. It is known that, for many publications, it takes several years from when research results are generated to when they are published in peer-reviewed journals (Aman, 2013; Larivière, Sugimoto et al., 2014; Okamura, 2022). Consequently, if metrics based on such databases are generated, evaluated, and then used to inform the policymaking process, by the time those metrics are used to make decisions on R&D investments, the situation and trends in R&D may have already changed by the time decisions on R&D investments are made.

The third hurdle regarding accessibility, related to the first and second points, is data acquisition and utilization autonomy. In many cases, data on academic publications are held by major publishers or companies providing subscription-based services, such as Elsevier (with Scopus) and Clarivate Analytics (with Web of Science), who provide various commercial services for a fee. The added value, including the data quality, must be fully appreciated (Visser, van Eck, & Waltman, 2021). The data they provide are also valuable for bibliometricians who wish to conduct detailed quantitative analyses, where precision and comprehensiveness of data are essential. However, not all policymakers and institutional practitioners necessarily require such precision at a fee and with restrictions for usage every time. More critical for them could be to access data as and when they need it, even if at a different level of quality than commercial services. Here, comparing the quality of commercial and open services in itself is an interesting issue that requires validation; as different perspectives and degrees of quality are required for different types of use, it may not always be assumed that open service is inferior to commercial data.

Metrics used in policymaking must enable practical and transparent policy accountability (Hicks et al., 2015; Wilsdon et al., 2015). Therefore, for bibliometric analysis to be timely and valuable for policy considerations, the bibliometrics platform should be open, accessible, large scale, systematic, and continuously updated and operated. OpenAlex (Priem, Piwowar, & Orr, 2022) is a promising candidate for such an open bibliometrics platform. This paper employs data from OpenAlex to present the results of my preliminary analyses of how international collaboration clusters have formed and evolved over the past half-century for a broad set of scientific publications. These results reflect the underlying trend of purely academia-driven or/and various state-driven cooperations, providing valuable insights for all stakeholders involved in international research collaboration.

In summary, this paper stands out for utilizing OpenAlex's open data instead of commercial data to investigate 50 years of research across up to 15 distinct natural science disciplines. The research encompasses various types of publications, including journal articles and nonjournal outputs. Additionally, the paper employs a hierarchical clustering technique to visualize and clarify the international collaborative relationships between countries worldwide. This approach provides a more nuanced perspective than a simplistic network structure analysis that focuses solely on the number of coauthored papers.

The rest of this paper is organized as follows. In Section 2, I describe the OpenAlex data used in this study and the R&D disciplines focused on. Section 3 presents the results of my analyses for each discipline, including changes over time in publication volume and international collaboration rate for each country. I highlight that the United States and China have rapidly moved closer together over the decades but started moving apart after 2019. Furthermore, we analyze and visualize the international collaboration clusters of top-tier countries in each discipline and their evolution over time. I also provide quantitative evidence for the "Shrinking World" phenomenon of the past half-century's research collaboration. Finally,

Section 4 is devoted to summary, discussion, and concluding remarks. Supplementary material A provides the technical details of the data analysis conducted in this paper. Supplementary material B presents the analysis results for the disciplines not fully presented in the main body of the paper, along with other complementary analysis results.

2. METHODS: THROUGH THE LENS OF OPEN BIBLIOMETRICS

This section provides an overview of the data used in this study; Supplementary material A.1 provides additional information on the data analysis and visualization platforms. First, I explain the method used to acquire the data and describe the data's characteristics and practical applications. Subsequently, I clarify the concept of the "nationality" of a scientific publication adopted in this paper. In addition, I provide a description of the R&D disciplines focused on in this paper.

2.1. The OpenAlex Data

The data utilized in this paper were obtained through the OpenAlex API[1], which is a fully open catalogue of global research systems (Priem et al., 2022). It was launched to replace Microsoft Academic Graph (MAG) (Sinha, Shen et al., 2015), which retired at the beginning of 2022. OpenAlex collects information on scientific publications, including journal articles, nonjournal articles, preprints, conference papers, books, and data sets—hereinafter collectively referred to as "works"—from various sources such as Crossref, ORCID (Open Researcher and Contributor ID), ROR (Research Organization Registry), and PubMed; preprint servers such as arXiv; and institutional or disciplinary repositories such as Zenodo. OpenAlex indexes about 239 million works, with approximately 50,000 new works added daily (Priem et al., 2022). The preprint version of this study, submitted to arXiv on November 8, 2022, is based on data obtained on October 25, 29, and 30 and November 7, 2022, containing data published until 2021. This study's speed was made possible by several open science/bibliometrics platforms, including OpenAlex as an open data source, OpenAlex API as an open standard API, and arXiv as an open preprint server. In addition, R and Python, open-source programming languages updated and enhanced daily by the open data community, were used for data acquisition, analysis, and visualization. Furthermore, the data sets generated and/or analyzed during this study can be found on Zenodo (see the "Data availability" statement), an open dissemination research data repository.

The advantages of using OpenAlex data as a data source are summarized below, with particular emphasis on its usefulness for my analysis. First, it provides extensive, if not exhaustive, coverage of meta-information on works, including those not published in journals. This feature is advantageous because it can more accurately supplement the volume of R&D activities and their associated outputs without underestimating it, even in disciplines where journals are not the primary venue for publishing research results, such as computer science. This approach also enables the capture of outputs in preprint format, which may exist for a certain period ranging from months to years, or indefinitely, without ever becoming journal articles, as well as other data formats. This is particularly significant given the increasing importance of such outputs in certain disciplines in recent years (Larivière et al., 2014; Okamura, 2022); see Supplementary material B.1 and Figure S1. Therefore, my approach provides a more comprehensive measure of scholarly outputs produced by each country during a certain period,

---

[1] OpenAlex API, https://docs.openalex.org/api/ (accessed October 31, 2022).

including those beyond journal articles. Although some preprint servers provide their house APIs (such as arXiv API, bioRxiv API, and medRxiv API), there have been no other freely accessible platforms than OpenAlex that cover all disciplines, from natural sciences to humanities and social sciences, on such a large scale.

It is important to acknowledge that there are potential disadvantages to not distinguishing research outputs in different formats, such as treating an article the same way as a data set or a book when counting outputs. However, this issue is not unique to this study's approach. Equating different journal articles could also have the same issue when counting outputs due to differences in content, length, and quality, even within the same discipline. The primary aim of this paper is to quantify the "momentum" of scholarly knowledge production outputs by different countries, regardless of format, and the potential disadvantages mentioned above are not the primary concern. Properly identifying disciplines, isolating fields with a homogeneous publishing culture, and comparing countries within that homogeneity can mitigate these potential disadvantages, as this study does.

The second advantage of using OpenAlex is that all data are organized at the microlevel, allowing users to selectively acquire and reorganize data according to their needs with a relatively high degree of flexibility. For instance, users can selectively extract metadata about journal articles, as also demonstrated in the present study. Third, OpenAlex is entirely open to the public and freely accessible, allowing a wide range of individuals, including data scientists, bibliometricians, and other interested parties, to utilize the data and ensure the transparency and reproducibility of analysis results. These advantages establish OpenAlex as one of the standard infrastructures supporting bibliometrics, in line with the growing momentum of open science and data science (Dong et al., 2017; Wittenburg, 2021), which I refer to as *open bibliometrics* in this paper.

2.2. R&D Disciplines

Policy documents that discuss international research collaboration often provide an overall assessment of trends across all R&D fields, sometimes with a field weighting. However, significant differences exist across fields regarding their characteristics, including the resources required for R&D, time scale and collaboration methods. Consequently, such a generalized or averaged picture of the R&D field is often of limited practical use. Therefore, it is crucial to identify and adopt an appropriate classification scheme for various R&D fields to derive meaningful policy implications for international research collaboration. In this regard, OpenAlex has an attribution called *concept* assigned to each work, equivalent to a well-defined set of R&D fields. More than 87% of the works on OpenAlex have been associated with one or more concepts (i.e., specific research areas or technologies) (Priem et al., 2022). The concepts have various levels of granularity, with 19 concepts at the coarsest (primitive) level 0; 284 at a slightly more specific level 1; followed by levels 3, 4, and 5, for 65,026 concepts at six different levels. OpenAlex's concept tree is a version of that used in MAG (Shen, Ma, & Wang, 2018; Sinha et al., 2015), improved with a new algorithm unique to the OpenAlex API.

In this study, I specifically focus on 15 level 1 concepts from the OpenAlex classification: Artificial Intelligence, Quantum Science, Biotechnology, Nanotechnology, Agricultural Engineering, Particle Physics, Aerospace Engineering, Nuclear Engineering, Marine Engineering, Neuroscience, Condensed Matter Physics, Environmental Engineering, Earth Science, Astronomy, and Pure Mathematics. I leverage the fact that OpenAlex assigns accompanying "related concepts" to each concept, which can be more refined or coarser than the concept's level. For

example, the level 1 concept of Artificial Intelligence is associated with level 2 subconcepts such as Artificial Neural Network and Deep Learning, as well as level 0 concepts such as Computer Science and Mathematics. To construct an enhanced notion of R&D discipline, I include all associated subconcepts of level 2 or higher for each of the above 15 level 1 concepts. For instance, my defined discipline of Artificial Intelligence includes OpenAlex's level 2 concepts of Artificial Neural Network and Deep Learning, but not the level 0 concepts of Computer Science or Mathematics.

Although I construct my R&D disciplines based on level-1 concepts, disciplines constructed based on higher concept levels could provide even more practical suggestions and implications depending on the situation. For example, using Quantum Computer (level 3) instead of Quantum Science or Biopharmaceuticals (level 2) instead of Biotechnology could provide more concrete implications for a country's R&D activities or international collaborations. I will defer such specific analyses to future work and focus on analyzing the broad level 1 discipline in this paper, where the volume of work can accumulate to the scale of hundreds of thousands to millions.

2.3. "Nationality" of Works and the Counting Method

Despite various intrinsic difficulties with bibliometric methods, one of the most challenging aspects to capture for individual works is information about the research institutions to which the contributors belong (Lammey, 2020). To investigate the R&D activities' status at the state or international level, we need information about the countries where the research institutions are located at the time of publication. Indeed, the literature has demonstrated that internationally coauthored publications are a reliable proxy for research collaboration (Glänzel, 2001; Glänzel & Schubert, 2004; Leydesdorff & Wagner, 2008; Luukkonen, Persson, & Sivertsen, 1992; Melin & Persson, 1996). However, in many cases, such information is unknown or unavailable in a database. Even if it is available, accurately analyzing the metadata can be difficult due to the identification or aggregation of institution names. This problem is typical of any bibliographic database, including commercial databases, preprint servers and repositories[2].

To alleviate, if not resolve, this issue, I employed information recorded in OpenAlex's data on "institutions." OpenAlex indexes about 109,000 institutions, around 94% of which have the ROR ID (Lammey, 2020) as the canonical identifier (Priem et al., 2022). The data about institutions are derived from metadata found in Crossref, PubMed, ROR, MAG, and publisher websites, linked to individual works through its unique algorithm. By retrieving necessary data filtered on the OpenAlex API according to the appropriate conditions, data on the number of works matching the conditions and associated metadata can be obtained in a format broken down by country. For simplicity, I refer to the work produced by contributors from institutions in Country X as a *work of nationality X.* Work produced through the collaboration of two contributors, one from an institute in Country X and the other from an institute in Country Y, has dual nationality of X and Y, counted both as a work of nationality X and a work of nationality Y. According to the terminology introduced here, work generally has multiple nationalities. If the country of the institution to which all contributors belong is unknown, the work is called a *work of unknown nationality.*

---

[2] As an illustration, arXiv does not require institutional affiliation information on a submitted eprint, resulting in a situation where obtaining the corresponding metadata via the arXiv API is challenging.

The results of the analysis on the nationality status of works by discipline and year are presented in Figure S2 in the Supplementary material. Despite yearly fluctuations, the percentage of works with unknown nationality consistently decreased in all disciplines over the past 50 years. Agricultural Engineering had the highest percentage of works with unknown nationality, at about 60–80% over the past few decades. In contrast, Nanotechnology and Condensed Matter Physics had a lower percentage in recent years, at about 30–40%. I excluded works with unknown nationality from the analysis due to data unavailability, even though the proportion of such works is large. I conducted, visualized, and interpreted the analysis assuming that the excluded works have trends similar to those of works with a known nationality.

It is worth noting how my method for counting works. As I aim to analyze the international presence of countries and how it has changed over time, I quantify the presence of each country by using a binary indicator (0 or 1) based on whether the country's name appears in the affiliation of one of the authors, indicating their involvement. Let us consider an example of an article coauthored by three individuals, two of whom are affiliated with an institute in Country X and one of whom is affiliated with an institute in Country Y. If the standard full counting method based on authorship is used, each author is assigned a weight of 1, and the sum of the weights is the number of coauthors, which in this case is 3. If the fractional counting method based on authorship is used, each author is assigned a weight of one-third, and the sum of the weights is always 1, regardless of the number of coauthors. In contrast, the binary counting method based on nationality used in this paper assigns a weight of 1 to each country for each article, regardless of the number of authors from that country. Therefore, in the example above, the article is counted with a weight of 1 for each Country X and Y, and Country X is not assigned a weight of 2, but rather a weight of 1, because as long as there are nonzero authors from a country, a weight of 1 is assigned to that country.

Let us take another example to see how this specific counting method would work better for the purpose of the present study. Consider an article with 10 coauthors from each of 10 different countries, making 100 coauthors. Indeed, it is not uncommon for the number of coauthors to exceed 100 (or even 1,000) in the case of extensive collaborative studies (Chawla, 2019; Nogrady, 2023). If the full counting method based on authorship were adopted, this article would assign a weight of 10 to each country. However, this could lead to an overestimation of each country's contribution to a single scientific result or a significant variation in the value implied by the weights for each article. On the other hand, if the fractional counting method based on authorship were adopted, the weights assigned to each country would be 0.1 each. However, this could lead to an underestimation of the presence of each country and would only be given a minimum weight in terms of nationality. These situations are not ideal for quantifying international presence in every scientific work. The counting method adopted in this paper (i.e., the binary counting method based on nationality presence/absence) assigns a weight of 1 to each country, mitigating the effects of bias that arise from the full and fractional counting methods based on authorship. In other words, the involvement or noninvolvement of each country in each work can be assessed more appropriately.

2.4. Clustering of Countries

Identifying international collaboration clusters requires grouping countries that cooperate closely. However, before I can do this, I must first find a way to quantify the distance between two countries in a reasonable manner. This requires careful consideration, as simply conceptualizing proximity between two countries as the number of collaborative works produced

would result in an ill-defined notion of distance. I require that the distance used for clustering satisfies the triangle inequality, which means that if Country X and Country Y are close and Country Y and Country Z are close, then Country X and Country Z must also be close to each other. Notice that even if there are many collaborative works in X and Y and many in Y and Z, this does not necessarily mean that the number of collaborative works in X and Z is large. This clearly illustrates that the number of collaborative works between two countries cannot simply be associated with the proximity between them.

To ensure conceptual soundness and mathematical well-definedness, I first establish the *affinity* between X and Y, denoted as $A_{X,Y}$. I define this as the number of works with nationalities of both X and Y divided by the total number of works with nationalities of at least one of X and Y. Mathematically, this is expressed as:

$$A_{X,Y} := \frac{|S_X \cap S_Y|}{|S_X \cup S_Y|} = \frac{n_{X,Y}}{n_X + n_Y - n_{X,Y}},$$

where S_X denotes the set of works of nationality X in a given period, $n_X := |S_X|$ denotes its size, and the same for S_Y and n_Y. Subsequently, I define the *distance* between X and Y as $D_{X,Y} := 1 - A_{X,Y}$. See Supplementary material A.2 for further details. The resulting distance measure ranges between 0 and 1 for arbitrary country pair {X,Y}. If $D_{X,Y} = 0$, X and Y always collaborate on all works, indicating that no works of nationality Y are without nationality X, and vice versa. If $D_{X,Y} = 1$, then X and Y never collaborate on works, meaning that there are no works with both X and Y nationalities. This set-theoretic distance metric satisfies the triangle inequality and can be used to calculate the distance between arbitrary country pairs in each period, as well as how the international collaborative clusters have changed over time.

Finally, a hierarchical cluster analysis (HCA) can be performed for the above-defined distance matrix D; see Supplementary material A.3 for the technical details. HCA is a widely used family of unsupervised statistical methods for classifying a set of items into some hierarchy of clusters (groups) according to the distances among the items. This method can provide a new way of looking at the international collaboration sphere when applied to the current context. Specifically, it informs us of which countries are close to each other and to what extent, and as a result, which countries can be considered to form an international research collaboration cluster at what threshold for the closeness.

3. RESULTS: KALEIDOSCOPES OF INTERNATIONAL COLLABORATION

This section presents my main results on work production and the state of international collaboration as viewed through the lens of open bibliometrics using the OpenAlex data.

3.1. Number of Works

I begin by presenting the results of my analysis on the number of works produced over the past half-century. To count works for each country, I adopted the binary counting method based on nationality, as introduced in the previous section. To reiterate the rule, if a work has nationalities of X and Y, it is counted as one work output for each country. Even if there are multiple contributors from X, the work is only counted as one in the production volume for X. If no country information is known for all the contributors to a given work, the work is counted as "unknown."

The left-hand side graphs of Figures 1 and S3 show the trend in the number of works for the top 10 countries in work production in each discipline in 2001–2020. To save space, I present the results for three disciplines—Artificial Intelligence, Quantum Science and
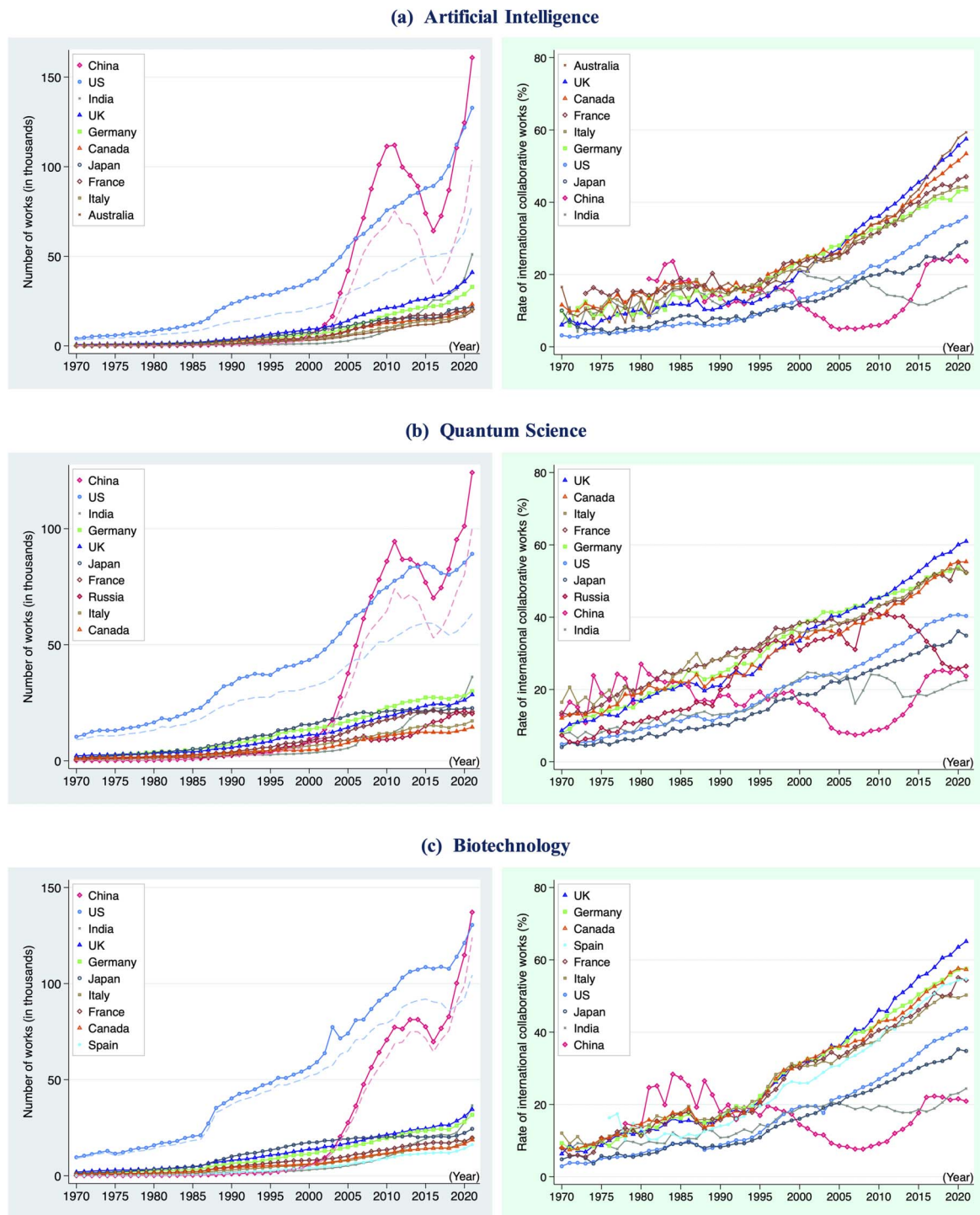
(a) Artificial Intelligence

(b) Quantum Science

(c) Biotechnology

Figure 1. Trends in the number of works (scientific publications) (left) and the international collaboration rate (right). The number of works is displayed in thousands.

Biotechnology—in Figure 1 and those for the other 12 disciplines in Figure S3 in Supplementary material B. In all of these disciplines, it is noticeable that China has shown dramatic growth over the past two decades (He, 2009; Institutes of Science and Development of the Chinese Academy of Sciences, n.d.; National Science Board & National Science

Foundation, 2021; Yuan et al., 2018). For example, in Artificial Intelligence, China surpassed the United States around 2020, producing more than 150,000 works in 2021 (Figure 1(a)). It also overtook the United States in Quantum Science (Figure 1(b)) and Biotechnology (Figure 1(c)) by 2021. Thus, what was an era of the United States as a single power a few decades ago has transitioned to a new era of two powerhouses, the United States and China. For reference, the graphs for the United States and China are also shown for the case where only journal papers are counted (dashed lines in lighter colors). It can be seen that the trends over time are generally similar between the case where all works are counted and the case where only journal papers are counted. However, the volume of works differs remarkably depending on the discipline, indicating how significant the contributions of works in forms other than journal papers can be for some disciplines (Kim, 2019; Larivière et al., 2014; Okamura, 2022)[3].

Although I do not delve into detailed analyses of individual curve profiles in this paper, it is important to consider the underlying reasons for each observation. For instance, in China, many disciplines exhibit an "N-shaped" curve with a peak around 2011, followed by a sharp decline, and then a sharp rise from around 2016. A major factor contributing to this trend could be the number of researchers. From 2000 to 2020, the total number of researchers in China has been increasing overall, but there was a significant drop in 2008 and 2009. According to data from the OECD (2023), the number of researchers per 1,000 employed decreased from 2.11 in 2008 to 1.52 in 2009. Similarly, data from the UNESCO Institute for Statistics (2023) shows a decrease in researchers in R&D (per million people) from 1,176 in 2008 to 847 in 2009. Although the reason for this decline in the number of researchers is still unclear, it is possible that its impact is reflected in a decrease in the number of scientific works after about 2 years. I will revisit this and other individual curves in future studies.

3.2. International Collaboration Rate

Next, the right-hand graphs of Figures 1 and S3 depict the trend of the international collaboration rate by discipline and country over the past half-century. The international collaboration rate for a specific year and country is the yearly number of international collaborative works divided by the total number of works produced. Only cases in which the yearly number of works produced is 100 or more are shown for each discipline and country. As a result, data are missing around 1970–1990 for some disciplines and countries. The countries selected for display are the top 10 countries in work production in each discipline during 2001–2020, which are the same as the corresponding left-hand side graphs[4].

In all disciplines, there has been a steadily increasing trend in the international collaboration rate, in line with the previous studies' findings based on commercial data (Kwiek, 2023; Leydesdorff & Wagner, 2008)[5]. Over the past 2 decades, the United Kingdom has been among the highest in many disciplines, and European countries such as Germany, France, and Italy have maintained high levels across the board. By contrast, despite the United States' rising

---

[3] See Okamura (2023) for the results for all the top 30 countries and the 15 disciplines, where the same conclusion can be confirmed.

[4] This approach does not reveal countries that have high international cooperation rates but are outside the top 10 in terms of work production. To complement the analysis results, Figure S4 presents an additional set of diagrams.

[5] This trend does not have a direct causal relationship with the upward trend in the rate of works with unknown nationality seen in Figure S2 because the international collaboration rate discussed here only refers to works with a known nationality.

trend (Adams & Gurney, 2018; National Science Board & National Science Foundation, 2021), its international collaboration rate is generally lower than the top-tier countries mentioned above in all 15 disciplines. For example, in 2021, it is around 40% for the three disciplines displayed in Figure 1. The international collaboration rate observed for China and India is notably lower (National Science Board & National Science Foundation, 2021; OECD, 2017; OECD & SCImago Research Group, 2016), which is consistent with prior studies based on commercial data. It appears that a relatively high percentage of work is completed only with R&D resources within their own countries, although it is unclear whether this is due to the policies of R&D institutions or a natural consequence of their large researcher populations. Moreover, some disciplines, such as Particle Physics, Aerospace Engineering, Nuclear Engineering and Astronomy (Figure S3c, d, e, and k, respectively), have a marked downward trend in Russia's international collaboration rate over the past decade (c.f., Kwiek, 2023). In recent years, other eye-catching features include Canada's high rate of international collaboration in many disciplines, including Artificial Intelligence, Quantum Science, Biotechnology, Neuroscience and Earth Science (Figures 1a, b, c, S3g, and j, respectively), Germany's and Australia's in Agricultural Engineering (Figure S3b), Spain's in Particle Physics and Astronomy (Figure S3c and k, respectively), France's and Canada's in Aerospace Engineering (Figure S3d), Italy's in Nuclear Engineering (Figure S3e), and Australia's in Environmental Engineering (Figure S3i)[6].

Additionally, comparisons across disciplines are also illustrative, as shown in Figure S5. International collaboration is particularly indispensable in large-scale academic R&D disciplines, such as Particle Physics and Astronomy, which usually require extended time and high cost, gathering many contributors from many countries. The international collaboration rates have been consistently high throughout the past half-century. Other disciplines, such as Condensed Matter Physics and Earth Science, have also shown relatively high rates of international collaboration in recent decades, and Quantum Science has recently shown a comparatively high rate as well.

3.3. Bilateral Collaborative Relationships

Many previous studies have reported on trends and patterns of international collaboration by country (Adams, 2012; Adams & Gurney, 2018; Adams et al., 2007; Glänzel, 2001; He, 2009; Kwiek, 2023; Luukkonen et al., 1992; Mattsson et al., 2008; National Science Board & National Science Foundation, 2021; OECD, 2017; OECD & SCImago Research Group, 2016; Yuan et al., 2018). Below I present the results of my analysis of the breakdown of specific collaborative partner countries based on the OpenAlex data for each of the 15 disciplines. Figures 2 and S6 divide the half-century from 1971 to 2020 into four periods: Period I (1971–1990), Period II (1991–2000), Period III (2001–2010), and Period IV (2011–2020)[7]. The chord diagrams visualize the status of bilateral collaborative relationships for each discipline and period. The countries selected for display are the top 10 countries in work production in each discipline and period. The scale along the circumference edge indicates

---

[6] Switzerland is ranked 12th in works production in Particle Physics and is not shown in Figure S3c. However, if it were, it would consistently be ranked the highest in terms of the international collaboration rate, reflecting the influence of CERN, the European Organization for Nuclear Research (as shown in Figure S4).

[7] The four periods defined here are simply a mechanical division of a 50-year time span into four for the sake of calendar convenience. They do not take into account any economic, social or geopolitical changes that may have occurred during this time. Therefore, it should be noted that trends such as the "N-shaped" trend in the number of works in China noted in Section 3.1 would be masked when illustrated as snapshots (Figure 2) in integral values for each period. This equally applies to the later Section 3.4 (Figure 4).
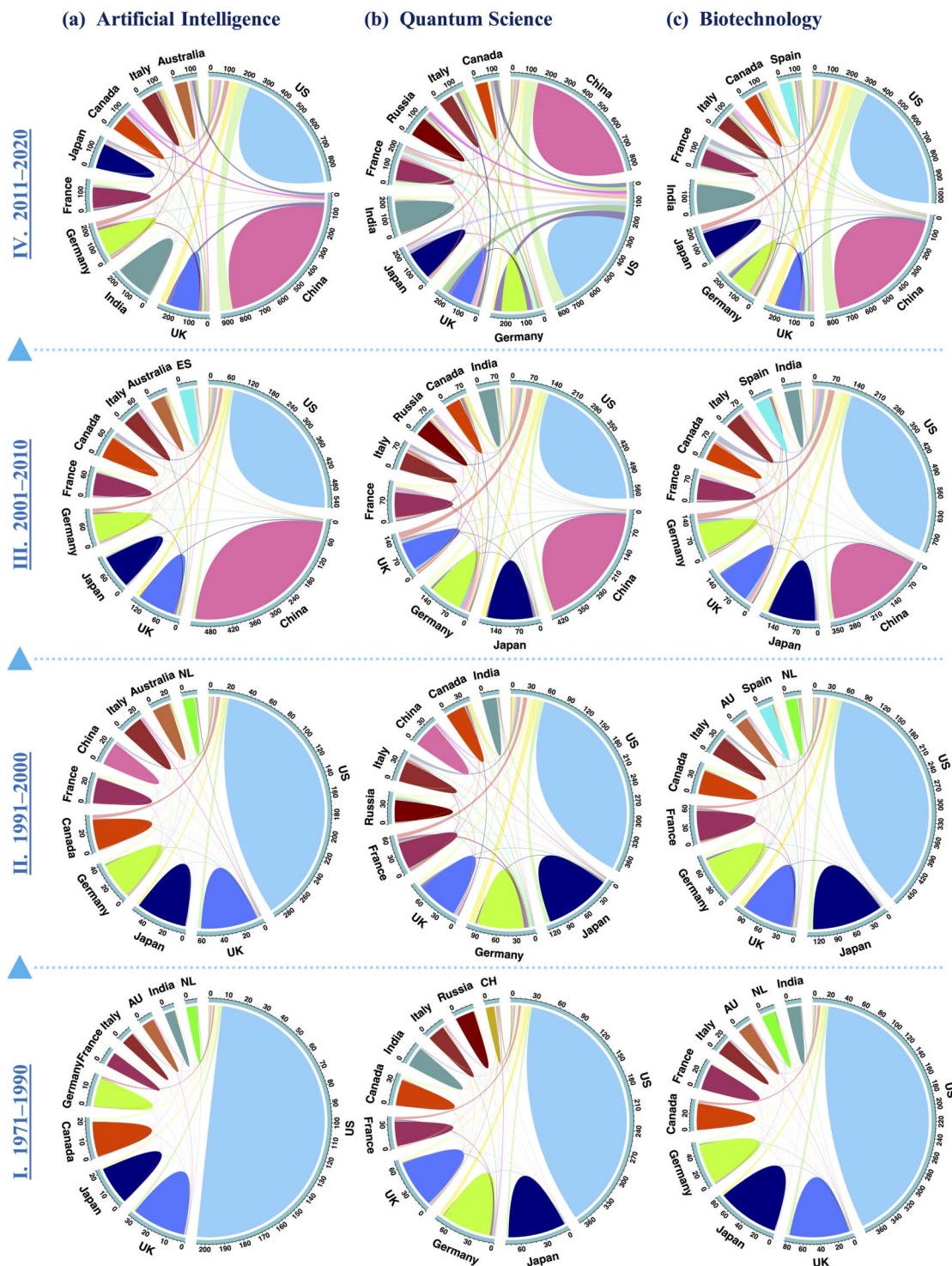
Figure 2. Changes in bilateral relationships over time. The number of works is displayed in thousands.

the number of produced works (in thousands), and the width of the band connecting the two country's arcs is proportional to the number of works collaboratively produced by them during each period. Some country names are abbreviated by two-letter country codes (ISO 3166-1 alpha-2) to make the diagrams easier to read.

A common trend among the disciplines is, again, China's remarkable progress that began this century (He, 2009; Institutes of Science and Development of the Chinese Academy of Sciences, n.d.; Yuan et al., 2018), accompanied by a decline in the relative positions of the United States and other major countries. For example, in Biotechnology (Figure 2(c)), the United States accounted for just under half of global output in Period I, and its international collaboration rate was low, at around 5%. Over time, the U.S. presence has declined significantly in relative terms. By Period IV, its presence on the chord diagram of these top 10 countries had dropped to about a quarter of the total. The reason for this is China's major breakthrough since Period III. When viewed on a 10-year period-integrated basis, the United States still holds the top position in Period IV, but when viewed on an annual basis, China has already overtaken the United States in the top position by 2021 (see Figure 1(c)).

Also evident is the revitalization of diverse international collaboration. The growing mutual presence of the United States and China can be seen from the expanding width of the band connecting the two countries. The international collaboration rate in the United States has been on an upward trend with China and other countries, which is consistent with the findings from previous studies based on commercial data (Adams & Gurney, 2018; National Science Board & National Science Foundation, 2021), resulting in the declining share of solely produced works (a hump-shaped part) on the U.S. arc, from approximately 95% to 70%. These chord diagrams indicate that over the past half-century, many disciplines have moved away from an era of single power (i.e., the United States) and towards an era of collaboration among a diverse range of countries. This feature is particularly evident in Particle Physics and Astronomy (Figure S6c and k), where the chord diagram becomes more colorful and balanced as we move towards Period IV.

In the following, I particularly focus on the collaborative relationship between the United States and China. As noted, the two countries have indeed deepened the relationship as an overall trend over the past half-century. However, given the strained U.S.–China relationship in recent years in the policy arena, I aim to examine whether geopolitical aspects have impacted international research collaboration through my bibliometric analysis. Looking solely at the increase or decrease in the number of coauthored papers between the two countries is inadequate in providing a complete picture of this impact. This is because an increase in the number of coauthored papers does not necessarily indicate a deepening relationship if the number of papers is increasing worldwide as a global trend. To effectively measure the degree of bilateral collaborative relationships, I must scale the absolute number of works per the trend of the times. As an appropriate indicator for this purpose, I adopt the affinity measure introduced in Section 2.4.

Figure 3(a) depicts the affinity between the United States and China over the past two decades, considering only data in which the yearly number of collaborative works between the two countries is 100 or more. For clarity, the affinity measure is rescaled; specifically, the rescaling is achieved as $A \mapsto \widetilde{A} := \ln(A/\epsilon)$ with $\epsilon = 0.001$. Note that although the absolute value of this rescaled affinity measure ($\widetilde{A}$) does not have direct physical significance, its relative comparison across disciplines or periods and increase/decrease over time do. Regarding the comparison across disciplines, on the one hand, the affinity between the two countries has been relatively small in massive and heavy R&D fields such as Aerospace Engineering, Nuclear Engineering, and Marine Engineering in recent years. On the other hand, a relatively large affinity between the two countries is observed for Condensed Matter Physics and Particle Physics. It is noteworthy that the affinity between the United States and China has been growing remarkably in all disciplines in common. As a reference, Figure 3(b) uses the same analytical method to show the affinity between the United States and Japan, also located in Asia. Despite
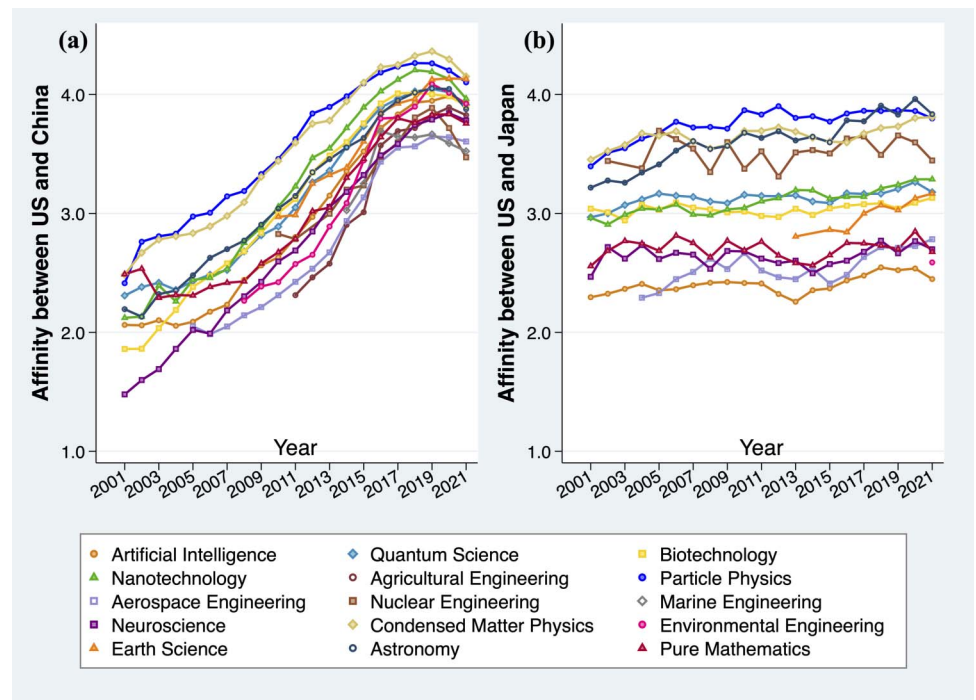
Figure 3. Trends in the affinity between the United States and China (a) and the United States and Japan (b).

the fact that the number of coauthorships between the United States and Japan has increased during the past decade, the affinity between the two countries has remained almost the same in all disciplines, with slight fluctuations, albeit at different levels per discipline. Therefore, it cannot be concluded that the relationship between the two countries has significantly deepened during the period. By contrast, the overall increasing affinity trend in Figure 3(a) suggests that the relationship between the United States and China is actually deepening.

Another noteworthy point about the U.S.–China relationship in Figure 3(a) is that the affinity between the two countries started decreasing again around 2019 for most disciplines. Although the implications of this phenomenon require careful examination, it could reflect the "chilling effect" stemming from the measures taken in the United States around 2018 to prevent technology outflows to China. The observed recent repulsive trend aligns with earlier research based on commercial data, which found a sharp decrease in the number of researchers with affiliations in both the United States and China in 2021 (Van Noorden, 2022). It is also consistent with the recent report that Chinese-origin scientists conducting research in the United States have been distancing themselves from the United States (Xie, Lin et al., 2022).

3.4. International Research Collaboration Clusters

Figures 4 and S7 show the analysis results for the formation of international research collaboration clusters in each discipline for the same four periods (I–IV) as before, spanning the half-century from 1971 to 2020. The visualization is based on a series of circular dendrograms that represent the results of HCA for the distance matrixes defined in Section 2.4. A dendrogram is a branching diagram based on the distances among a group of entities. In the case of the circular dendrogram employed here, the countries or clusters closer to each other are combined
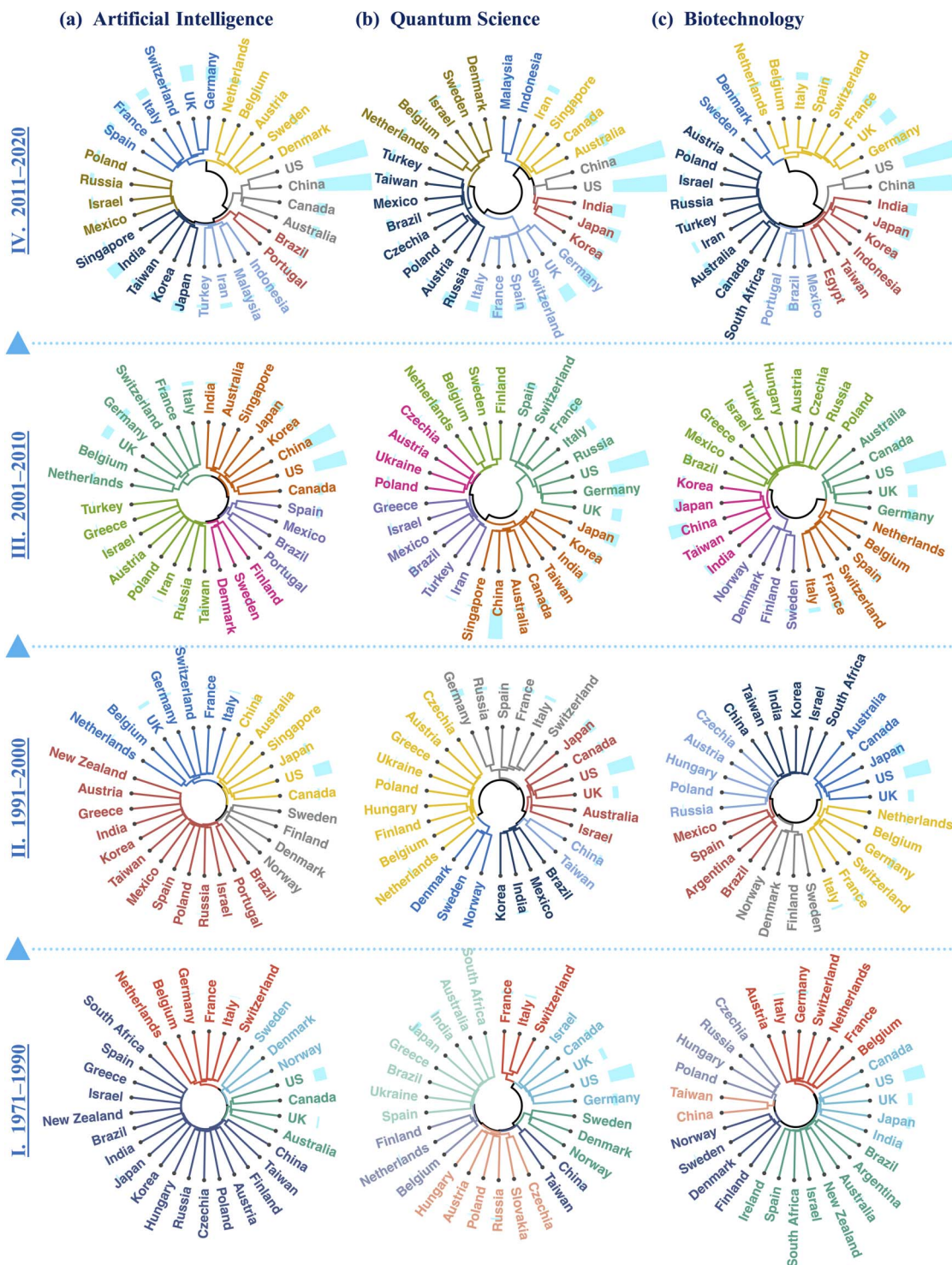
Figure 4. Evolution of international research collaboration clusters.

earlier as one moves from the outer edge of the circle towards its center. The height of the branching points, measured from the circumference and referred to as the coupling height, indicates how distant the countries or clusters of countries are from each other; the greater the coupling height, the farther away their relationships are.

The countries selected for display are the top 30 countries in work production in each discipline and period. The circular bar graph in the outer region of the circular dendrogram shows the number of works produced by each country during each period[8]. The number of clusters that are color-coded was calculated based on the preset threshold value for the coupling height. As Table S1 shows, the number of clusters does not necessarily increase or decrease with time; it highly depends on the discipline-wise situations and the preset coupling height threshold (Supplementary material A.3). In this regard, the distribution of coupling heights is more useful than the number of clusters to characterize the structure of international research collaboration clusters and compare it across disciplines and periods. If all the coupling heights are maximally high, then the circular dendrogram would look like a shape in which lines run parallel from equally spaced points on the circumference to the center of the circle, only to join together at a certain minimal radius all at once. If the coupling heights are all relatively low, the branches soon couple with each other as they move from a point on the circumference to the center of the circle, quickly forming clusters, with the so-formed clusters also coupling one after another well before the minimal radius.

It is useful first to have a big picture of overall trends. For all R&D disciplines illustrated in Figures 4 and S7, we can observe that over time, the open space in the center of the circle tends to expand, like a tightly closed bud opening, and the structure of the branching tree becomes easier to see. This observation suggests that countries worldwide are increasing their collaborative tendencies. Within this overall trend, different trends can be observed in the status of connections within the dendrogram (i.e., the formation of clusters), depending on the period and the disciplines. For example, in the three disciplines shown in Figure 4 (and also in the other disciplines indicated in Figure S7), the United Kingdom and Germany have always been the first pair to connect since this century. To that pair, France, Italy, Switzerland, and Spain have attached to form a European subcluster. Belgium, the Netherlands, Denmark, and Sweden often connect first and tend to form another set of subclusters. The distance between these subclusters varies according to disciplines and period, even within the same Europe. Furthermore, in many disciplines, the United Kingdom in the last century was more deeply tied to the United States, Canada, and Japan than European countries. Thus, various combinations and recombinations over time have formed a snapshot of the international research collaboration clusters of the time.

There may be various circumstances behind the formation of a particular cluster and its change over time (Hou, Pan, & Zhu, 2021; Luukkonen et al., 1992; Vieira, Cerdeira, & Teixeira, 2022). As discussed, the geographical proximity of the countries involved, such as in Europe or Asia, is often the most significant factor affecting the international collaboration status (Doria Arrieta et al., 2017; Fitzgerald, Ojanperä, & O'Clery, 2021; Katz, 1994). For example, the European cluster can be identified as the direction from 10 to 2 o'clock of the Artificial Intelligence diagram (Figure 4(a)) in Period IV. It is noteworthy that as time passes, in many R&D disciplines, China has moved out of what may be regarded as the Asian circle and moved into the prominent top-tier group per the works produced. Alternatively, the formation of clusters may be related to geopolitical and historical perspectives (Luukkonen et al., 1992; Maher & Van Noorden, 2021). There may also have been movements among universities or research institutions where researcher exchanges flourish due to policy support under state-led scientific agreements or economic cooperation. Thus, the background to forming the international research collaboration clusters involves a variety of national and international policies and changes in the R&D environment surrounding the academic arena, resulting in each cluster snapshot. In other words, they are relatively determined rather than determined by the policies of a single country, and

---

[8] The circular bar graphs are comparable within the same discipline across different periods, but not between different disciplines.
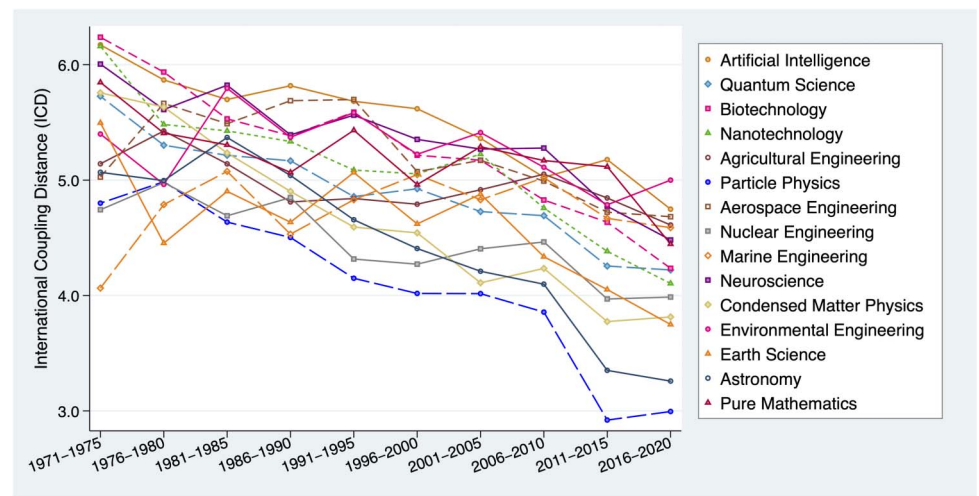
Figure 5.   "Shrinking World": Change in the level of International Coupling Distance (ICD).

it is challenging to decipher them convincingly. Although this paper will not go into contextual interpretations of individual observations, a deeper contextual discussion, complementing expert knowledge that cannot be obtained from the bibliometric approach, will provide more implications for the structure and the formation dynamics of collaboration clusters.

3.5.   "Shrinking World"

To analyze the trend of increasing collaboration discussed earlier in a more quantitative manner, I divided the half-century from 1971 to 2020 into 10 periods of 5 years each and performed cluster analysis in each discipline and period. I then rescaled the set of coupling heights with an appropriate monotonically increasing function to make the graph easier to read (Supplementary material A.3). By comparing the mean of the rescaled coupling height distribution—hereinafter referred to as the (mean) *International Coupling Distance (ICD)*[9]— across periods, I can determine whether countries are getting closer or further away from each other on average over time.

Figure 5 displays the trends in the ICD index calculated for each discipline over time, showing variations in ICDs among disciplines. Artificial Intelligence exhibits a relatively high level of ICD, possibly because individual researchers in the field tend to work independently, without needing to cross borders. The low level of ICD for Particle Physics and Astronomy is a good reflection of reality; single countries cannot accomplish the mission in such large-scale academic disciplines, and international collaboration is indispensable, consistent with the previous observation in Figure S5. The level of ICD is also low for Nuclear Engineering, which may reflect that it is a broad engineering discipline that spans many fields, from nuclear physics to materials science and applied chemistry, requiring combining technical knowledge and expertise from various countries and sectors.

Despite having its ups and downs at different times, the overall trend is that ICD has fallen over the past half-century, indicating a "Shrinking World" of research collaboration. Figure S8 shows the result of the kernel density estimation of ICD for each period. As time passes, the density curve's peak position shifts to a smaller value of ICD, supporting the observation at the

---

9   The suggested view of the "Shrinking World" of research collaboration would remain the same even if I used the median, instead of the mean, of the rescaled coupling height distribution.

discipline-aggregate level. These results suggest that the S&T world has been consistently getting smaller, and research collaboration has been more active across borders to better address expanding knowledge. This phenomenon might be attributed to the fact that the scale of social issues targeted by S&T has expanded to the global scale, and technological advances have made it possible to address such global-scale issues with improved international connectivity among researchers. Additionally, policymakers, aware that S&T is the source of national strength and industrial competitiveness and key to economic security, have launched strategic international collaborative projects from the political arena[10].

4. SUMMARY AND DISCUSSION: ACROSS BORDERS, DISCIPLINES, AND GENERATIONS

The past few decades have witnessed the development and implementation of various digital technologies in society, drastically changing the relationship between people, S&T and society. The development of digital communication tools and platforms in the advanced information society has significantly updated how scientists and engineers interact and transfer knowledge, resulting in a smaller world (Figure 5). Driven by the rapid and irreversible movement of open science (Burgelman et al., 2019; Miedema, 2022), all forms of scientific publication, including nonjournal articles, preprints, databases, and social networking services, have become indispensable tools and platforms in scholarly communication today. It would be fair to say that a new initiative of open bibliometrics is replacing the traditional bibliometrics that relies solely on commercial databases centered on journal articles. This initiative will undoubtedly play an essential role in forming future S&T and innovation policies, evaluating and publishing them, and accelerating interdisciplinary approaches (Ledford, 2015; Mol & Hardon, 2020; Okamura, 2019; Yanai & Lercher, 2020).

With this philosophy in mind, the present study provided unique evidence of how international collaboration clusters have formed and evolved over the past half-century for a broad set of scientific publications based on the OpenAlex data set. I first reviewed the global presence change of top-tier countries for each research discipline, as measured by publication volumes and international collaboration rates. Notably, the United States and China were shown to have rapidly moved closer together for decades but started moving apart after 2019. Subsequently, I analyzed and visualized the international collaboration clusters for each discipline and period based on a hierarchical clustering method. Finally, I provided global-scale quantitative evidence for a "Shrinking World" of the past half-century's research collaboration. These results provide valuable insights into the big picture of past, present, and future international collaboration.

Several methodological innovations were developed and demonstrated in this study, making these dynamic quantitative analyses possible. Specifically, the first and most novel device was formulating the distance between two countries as a simple set-theoretic distance, the so-called Jaccard distance. This approach highlighted the dynamic distance relationship between countries and groups of countries, which had not been possible before. The second device was applying the Jaccard distance function to the hierarchal clustering of countries with Ward's method to identify international research collaboration clusters each period for each R&D discipline. Furthermore, the visualization method is also novel; the circular dendrogram is frequently used in papers on phylogenetics, but this is the first time it has been utilized in scientometrics/bibliometrics.

---

[10] The evidence obtained for the repelling force between the United States and China at the end of Section 3.3, along with the observation in Section 3.5 of a shrinking trend in research collaboration, suggests a picture of a "Shrinking-and-Polarizing World," provided these trends continue.

There are several directions in which the results of this study can be further developed. One important area for future research is the contextual understanding of cluster formation dynamics. A more in-depth discussion of national and international S&T-related policies that impact global R&D collaboration would provide additional implications for the study's findings. Instead of simply investigating bilateral relationships using the "symmetric" measure of distance ($D_{X,Y}$) or affinity ($A_{X,Y}$) utilized in this paper, analysis based on an "asymmetric" measure that distinguishes between the mobility of researchers from Country X to Country Y and vice versa would also be informative. Exploring this approach, in combination with the study's results, would offer a complementary understanding of and deeper insights into the dynamics of international collaboration. Moreover, although this study focused on the top 10 (Figures 1, 2, S3, S4, and S6) or top 30 (Figures 4 and S7) countries in work production in each discipline and period, it is also relevant to consider countries outside the top 10 or top 30 for certain policy purposes. Additionally, it would be interesting to analyze not only the absolute value of the number of works but also the relative value divided by other R&D-related indicators defined at the state level, including researcher population, total R&D budget, and GDP.

Again, we must be fully aware of the limitations of bibliometrics or, more broadly, scientometrics. It should always be kept in mind that due to various methodological difficulties in scientometrics, their policy implications are inherently limited (Hicks et al., 2015; Waltman, 2016; Wilsdon et al., 2015); see also Okamura (2019, 2022). The method proposed and implemented in this paper is no exception. In addition to the various limitations discussed in the previous sections, including the metadata availability of OpenAlex, my results are also likely highly dependent on the R&D field classification scheme. Different specifications of disciplines could have been applied, leading to quantitatively different implications for each country's international presence and the landscape of international research collaborations. Furthermore, it should be noted that a considerable volume of R&D outputs is still not published in papers or data, including classified research results related to national security and defense, which are not recorded in open databases. The insights from bibliometric analysis, including those of the present paper, can only represent a part of the unclassified, open world.

Despite its limitation, this paper provides valuable knowledge and new insights into the macro trends in international research collaboration and its current situation. When looked at through the same bibliometric lens several years from now, the results will show a new landscape that reflects the integrated impact of all major global issues underway, including the COVID-19 pandemic, various geopolitical issues, and highly digitized and diversified scholarly communication modes. The new landscape offered by the "science of science" (Fortunato et al., 2018) will continue to expand, where scientometric methods will be used in increasingly sophisticated and exciting ways. A new generation of scientometricians will create new values for the times exposed to new data platforms and a highly digitized society. They can change the angle to see the world, adjust the resolution, transcend across disciplines and gain unique perspectives on the S&T ecosystem. Hopefully, these new generations will bring practical hints and actions that resonate and sympathize with many people on how to proceed with our international research collaboration for a better society.

"Science has no borders, but scientists have their homelands," said Louis Pasteur. How many scientists and engineers over the decades and generations have had their minds blown by these words, only to be confronted with the gap between their ideals and reality? We know that, by definition, science has no borders. At the same time, we must accept that *accessible* science has borders in reality. Now that the power and use of S&T determine the course of the world, all stakeholders must revisit what the borders and homelands mean and redefine them in the contemporary context of responsible R&D. How can we embody the value of S&T

literally without borders? To this end, how can we create a policy environment that maximizes the social value of borderless academia, and how can we pass that on to the next generation? The challenge to answer these questions confronts all policymakers and stakeholders of S&T today. Amid this unprecedentedly complex and unpredictable international situation, I hope this paper sheds light on some essential nature of global R&D cooperation for those seeking to open up new horizons at the interface of S&T and society.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

COMPETING INTERESTS

FUNDING INFORMATION

DATA AVAILABILITY

The data sets and figures generated and/or analyzed during this study can be found in the Zenodo repository at https://doi.org/10.5281/zenodo.7297122. For reference for interested readers, the package also contains data and figures based on bibliometric data from only journal articles.

REFERENCES

Adams, J. (2012). The rise of research networks. *Nature, 490*(7420), 335–336. https://doi.org/10.1038/490335a, PubMed: 23075965

Adams, J., Gurney, K., & Marshall, S. (2007). Patterns of international collaboration for the UK and leading partners: A report commissioned by the UK Office of Science and Innovation. https://image.guardian.co.uk/sys-files/Education/documents/2007/07/13/OSICollaborationSummaryRepo.pdf

Adams, J., & Gurney, K. A. (2018). Bilateral and multilateral coauthorship and citation impact: Patterns in UK and US international collaboration. *Frontiers in Research Metrics and Analytics, 3*, 12. https://doi.org/10.3389/frma.2018.00012

Aman, V. (2013). *The potential of preprints to accelerate scholarly communication: A bibliometric analysis based on selected journals.* Masters Thesis, School of Library and Information Science, Humboldt University of Berlin. https://doi.org/10.48550/arXiv.1306.4856

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.-M., Conradsen, M. L., … Xu, S. M. (2022). The Open Innovation in Science research field: A collaborative conceptualisation approach. *Industry and Innovation, 29*(2), 136–185. https://doi.org/10.1080/13662716.2020.1792274

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., … Schouppe, M. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data, 2*, 43. https://doi.org/10.3389/fdata.2019.00043, PubMed: 33693366

Cabezas-Clavijo, A., & Torres-Salinas, D. (2021). Bibliometric reports for institutions: Best practices in a responsible metrics scenario. *Frontiers in Research Metrics and Analytics, 6*, 696470. https://doi.org/10.3389/frma.2021.696470, PubMed: 34278205

Chawla, D. S. (2019). Hyperauthorship: Global projects spark surge in thousand-author papers. *Nature*, December 13. https://doi.org/10.1038/d41586-019-03862-0, PubMed: 33311624

Chen, K., Zhang, Y., & Fu, X. (2019). International research collaboration: An emerging domain of innovation studies? *Research Policy, 48*(1), 149–168. https://doi.org/10.1016/j.respol.2018.08.005

Dong, Y., Ma, H., Shen, Z., & Wang, K. (2017). A century of science: Globalization of scientific collaborations, citations, and innovations. In *Proceedings of the 23rd ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD 2017)*

(pp. 1437–1446). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3097983.3098016

Doria Arrieta, O. A., Pammolli, F., & Petersen, A. M. (2017). Quantifying the negative impact of brain drain on the integration of European science. *Science Advances, 3*(4), e1602232. https://doi.org/10.1126/sciadv.1602232, PubMed: 28439544

Dusdal, J., & Powell, J. J. W (2021). Benefits, motivations, and challenges of international collaborative research: A sociology of science case study. *Science and Public Policy, 48*(2), 235–245. https://doi.org/10.1093/scipol/scab010

Fitzgerald, J., Ojanperä, S., & O'Clery, N. (2021). Is academia becoming more localised? The growth of regional knowledge networks within international research collaboration. *Applied Network Science, 6*(1), 38. https://doi.org/10.1007/s41109-021-00371-w

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., … Barabási, A.-L. (2018). Science of science. *Science, 359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185, PubMed: 29496846

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics, 51*(1), 69–115. https://doi.org/10.1023/A:1010512628145

Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 257–276). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2755-9_12

Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. In I. M. Welpe, J. Wollersheim, S. Ringelhan, & M. Osterloh (Eds.), *Incentives and performance: Governance of research organizations* (pp. 121–139). Cham: Springer. https://doi.org/10.1007/978-3-319-09785-5_8

He, T. (2009). International scientific collaboration of China with the G7 countries. *Scientometrics, 80*(3), 571–582. https://doi.org/10.1007/s11192-007-2043-y

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–431. https://doi.org/10.1038/520429a, PubMed: 25903611

Hou, L., Pan, Y., & Zhu, J. (2021). Impact of scientific, economic, geopolitical, and cultural factors on international research collaboration. *Journal of Informetrics, 15*(3), 101194. https://doi.org/10.1016/j.joi.2021.101194

Institutes of Science and Development of the Chinese Academy of Sciences (CASISD). (n.d.). Research Fronts 2021 and other publications. https://english.casisd.cn/research/rp/

Ismail, S., Nason, E., Marjanovic, S., & Grant, J. (2012). *Bibliometrics as a tool for supporting prospective R&D decision-making in the health sciences.* Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/technical_reports/TR685.html

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics, 31*(1), 31–43. https://doi.org/10.1007/BF02018100

Kim, J. (2019). Author-based analysis of conference versus journal publication in computer science. *Journal of the Association for Information Science and Technology, 70*(1), 71–82. https://doi.org/10.1002/asi.24079

Kwiek, M. (2023). The globalization of science: The increasing power of individual scientists. In P. Mattei, X. Dumay, E. Mangez, & J. Behrend (Eds.), *The Oxford handbook of education and globalization* (pp. 1–35). Oxford: Oxford University Press. https://ssrn.com/abstract=3851630

Lammey, R. (2020). Solutions for identification problems: A look at the Research Organization Registry. *Science Editing, 7*(1), 65–69. https://doi.org/10.6087/kcse.192

Larivière, V., Sugimoto, C., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv e-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology, 65*(6), 1157–1169. https://doi.org/10.1002/asi.23044

Ledford, H. (2015). How to solve the world's biggest problems. *Nature, 525*(7569), 308–311. https://doi.org/10.1038/525308a, PubMed: 26381968

Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics, 2*(4), 317–325. https://doi.org/10.1016/j.joi.2008.07.003

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values, 17*(1), 101–126. https://doi.org/10.1177/016224399201700106

Maher, B., & Van Noorden, R. (2021). How the COVID pandemic is changing global science collaborations. *Nature, 594*(7863), 316–319. https://doi.org/10.1038/d41586-021-01570-2, PubMed: 34135528

Mattsson, P., Laget, P., Nilsson, A., & Sundberg, C.-J. (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics, 75*(3), 555–574 https://doi.org/10.1007/s11192-007-1793-x

Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics, 36*(3), 363–377. https://doi.org/10.1007/BF02129600

Miedema, F. (2022). *Open science: The very idea.* Dordrecht: Springer. https://doi.org/10.1007/978-94-024-2115-6

Mol, A., & Hardon, A. (2020). What COVID-19 may teach us about interdisciplinarity. *BMJ Global Health, 5*(12), e004375. https://doi.org/10.1136/bmjgh-2020-004375, PubMed: 33328203

National Science Board & National Science Foundation. (2021). *Publications output: U.S. trends and international comparisons. Science and engineering indicators 2022.* https://ncses.nsf.gov/pubs/nsb20214/

National Science Foundation. (2020). *Dear Colleague Letter: Future of international research collaboration post COVID-19.* October 1. https://www.nsf.gov/pubs/2020/nsf20132/nsf20132.jsp

Nogrady, B. (2023). Hyperauthorship: The publishing challenges for "big team" science. *Nature, 615*(7950), 175–177. https://doi.org/10.1038/d41586-023-00575-3, PubMed: 36849643

OECD. (2017). *OECD science, technology and industry scoreboard 2017: The digital transformation.* Paris: OECD Publishing. https://doi.org/10.1787/9789264268821-en

OECD. (2023). Researchers (indicator). https://doi.org/10.1787/20ddfb0f-en. (accessed April 23, 2023).

OECD. (n.d.). Scientometrics. https://www.oecd.org/sti/inno/scientometrics.htm

OECD and SCImago Research Group (CSIC). (2016). *Compendium of bibliometric science indicators.* Paris: OECD Publishing. https://www.oecd.org/sti/inno/Bibliometrics-Compendium.pdf

Okamura, K. (2019). Interdisciplinarity revisited: Evidence for research impact and dynamism. *Palgrave Communications, 5*(1), 141. https://doi.org/10.1057/s41599-019-0352-4

Okamura, K. (2022). Scientometric engineering: Exploring citation dynamics via arXiv eprints. *Quantitative Science Studies, 3*(1), 122–146. https://doi.org/10.1162/qss_a_00174

Okamura, K. (2023). Data and figures for "A half-century of global collaboration in science and the 'Shrinking World'". *Zenodo.* https://doi.org/10.5281/zenodo.7297122

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv.* https://doi.org/10.48550/arXiv.2205.01833

Shen, Z., Ma, H., & Wang, K. (2018). A web-scale system for scientific knowledge exploration. In *Proceedings of ACL 2018, System*

Demonstrations (pp. 87–92). Melbourne: Association for Computational Linguistics. https://aclanthology.org/P18-4015

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., ... Wang, K. (2015). An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web* (WWW 2015 Companion, pp. 243–246). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2740908.2742839

Sloan, S. S., & Alper, J. (Eds.). (2018). *Data matters: Ethics, data, and international research collaboration in a changing world: Proceedings of a workshop.* Washington, DC: National Academies Press. https://doi.org/10.17226/25214, PubMed: 30601606

UNESCO Institute for Statistics. (2023). Researchers in R&D (per million people) – China. https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6?locations=CN (accessed April 23, 2023).

Van Noorden, R. (2022). The number of researchers with dual US-China affiliations is falling. *Nature*, *606*(7913), 235–236. https://doi.org/10.1038/d41586-022-01492-7, PubMed: 35637342

Vieira, E. S., Cerdeira, J., & Teixeira, A. A. (2022). Which distance dimensions matter in international research collaboration? A cross-country analysis by scientific domain. *Journal of Informetrics*, *16*(2), 101259. https://doi.org/10.1016/j.joi.2022.101259

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Wagner, C. S., & Jonkers, K. (2017). Open countries have strong science. *Nature*, *550*(7674), 32–33. https://doi.org/10.1038/550032a, PubMed: 28980660

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., ... Johnson, B. (2015). *The Metric Tide: Report of the independent review of the role of metrics in research assessment and management.* SAGE Publications Ltd. https://doi.org/10.4135/9781473978782

Wittenburg, P. (2021). Open science and data science. *Data Intelligence*, *3*(1), 95–105. https://doi.org/10.1162/dint_a_00082

Xie, Y., Lin, X., Li, J., He, Q., & Huang, J. (2022). Caught in the crossfire: Fears of Chinese-American scientists. *arXiv*. https://doi.org/10.48550/arXiv.2209.10642

Yanai, I., & Lercher, M. (2020). Renaissance minds in 21st century science. *Genome Biology*, *21*(1), 67. https://doi.org/10.1186/s13059-020-01985-6, PubMed: 32169112

Yuan, L., Hao, Y., Li, M., Bao, C., Li, J., & Wu, D. (2018). Who are the international research collaboration partners for China? A novel data perspective based on NSFC grants. *Scientometrics*, *116*(1), 401–422. https://doi.org/10.1007/s11192-018-2753-3