ARL-TR-10057 • FEB 2025

Global Collaboration Dynamics in Scientific Research: How Often Does It Occur?

by Mark A. Tschopp

NOTICES

Disclaimers

ARL-TR-10057 • FEB 2025

Global Collaboration Dynamics in Scientific Research: How Often Does It Occur?

by Mark A. Tschopp
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
February 2025	Technical Report	01/01/2023	12/31/2024

4. TITLE AND SUBTITLE

Global Collaboration Dynamics in Scientific Research: How Often Does It Occur?

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)

Mark A. Tschopp

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-I 2800 Powder Mill Road, Adelphi, MD 20783-1138	ARL-TR-10057

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

ORCID: 0000-0001-8471-5035; email: <mark.a.tschopp.civ@army.mil>.

14. ABSTRACT

This technical report aims to understand the role of collaboration in the scientific landscape, with a focus on international collaborations that transcend geographical boundaries. The objective of this research is to explore how often this occurs by examining the temporal evolution of scientific articles, the dynamics of global scientific collaboration, and the asymmetries in global collaboration for countries, U.S. institutions, U.S. funding agencies, and research areas. The research methodology is grounded in the interdisciplinary discipline of Science of Science, utilizing the OpenAlex database to extract insights into the collaborative landscape from millions of scientific articles. The report presents new methodologies for analyzing and visualizing global collaboration to show the changes in collaboration through the decades for top countries, with a particular focus on the United States' collaborations and the evolving nature of China in global collaboration. The findings provide insights into the global dynamics of collaboration on scientific progress and ultimately aim to guide decisions pertaining to international partnerships in advancing science.

15. SUBJECT TERMS

Science of Science, data science, collaboration, global partnerships, Military Information Science

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	72

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Mark A. Tschopp	301-938-5389

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

Scientific curiosity, the fundamental pursuit of "why," has propelled the technological landscape witnessed today. Across diverse domains, each new scientific discovery does not halt progress; instead, it unfurls deeper questions and prompts contemplation on how to translate these findings into broader impact. This perpetual cycle fuels a continuous influx of ideas, driven by the exponential growth of rich data, innovative techniques, advanced tools, and the enhanced skill sets of countless scientists. Collectively, they shape the shared understanding of the world. Science assumes a pivotal role in the current era marked by unprecedented global collaboration and technological strides, prompting researchers to scrutinize the intricate workings of the complex scientific ecosystem.

The inquiry of how the scientific enterprise works has given rise to a thriving transdisciplinary field of study known as the Science of Science.[1,2] This interdisciplinary domain employs computational methods to dissect and comprehend the network structure, dynamics, and impact of scientific research. Rooted in the analysis of patterns governing the production, dissemination, and consumption of scientific knowledge, the Science of Science integrates insights from diverse disciplines including sociology, information science, network science, data analytics, and, more recently, advances in ML and AI. Researchers within this field aim to unveil the underlying mechanisms steering scientific discovery and innovation. From investigating scientists' careers and collaboration dynamics to dissecting citation patterns and investigating the roles of funding agencies and policy, the Science of Science offers a holistic lens with which to view science from a 30,000-ft view. It provides researchers with a comprehensive view of the scientific enterprise, shedding light on what does and does not work. As this discipline grows, it not only enriches the comprehension of knowledge generation and dissemination but also provides a high-altitude perspective for policymakers, funding agencies, and scientists alike, fostering a more informed and collaborative scientific landscape.

Some of the biggest advances in driving this new field forward is the development of large fully-open datasets that allow researchers to finally answer questions about the science ecosystem at scale through easy-to-use interfaces. Microsoft Academic Graph was perhaps the first open heterogeneous graph containing scientific records, citation relationships, authors, institutions, journals, conferences and field,[3,4] and this database has been carried on with OpenAlex.[5,6] Needless to say, with a dataset of millions of grants, articles, authors, and billions of citations, geographical boundaries no longer constrain the pursuit of knowledge into the vast scientific enterprise. This report aims to use this dataset to better understand some of the complexities of global collaboration—defined as scientific works that hail from authors from multiple countries—on the scientific enterprise.

Is collaboration a catalyst for scientific progress? The diminishing prevalence of solo-authored papers over time suggests a shifting trend towards collaborative efforts in scientific research.[7] This surge in collaboration is driven by the recognition that team-based endeavors often yield more impactful outcomes than solo-authored works. Wuchty et al.'s extensive analysis of nearly 20 million papers and over 2 million patents spanning five decades demonstrates the ascendancy of teams in knowledge production across diverse research domains.[7] Moreover, their findings reveal that articles produced collaboratively garner more frequent citations, signifying greater impact within the research community. The mounting evidence underscores the positive correlation between collaboration, knowledge production, and the heightened impact of scientific contributions.

Yet, as collaboration becomes more ingrained in the fabric of scientific inquiry, the question arises: what about collaboration that transcends the geographical boundaries of countries? The advent of the Internet has revolutionized the exchange of information globally, epitomized by the phrase "death of distance" in scientific discourse. The coronavirus disease 2019 (COVID-19) pandemic has only accelerated this trend, pushing scientific interactions into the virtual realm. The so-called death of distance prompts inquiries into the frequency of papers featuring collaborations between multiple countries. Existing studies indicate a steady increase in the percentage of papers involving international collaborations.[8] However, delving deeper, questions emerge: with whom do the top countries collaborate? How does this trend manifest for institutions and funding agencies? In which research areas do countries forge collaborations, and how has this landscape evolved over time? The growing impact of science's globalization, coupled with concerns about the openness of scientific endeavors, fuels curiosity within the scientific community, as well as among policymakers and funding agencies, making this area of research a focal point for exploration and understanding.

Building on the foundations laid by team science and recognizing the collaborative shift in the scientific landscape, this study aligns itself within the domain of the Science of Science. It explores of the evolutionary trajectory of scientific articles, concurrently delving into the dynamics of global scientific collaboration. The impetus for this inquiry stems from persistent questions surrounding collaborative endeavors among countries, the detailed frameworks of emerging open databases such as OpenAlex, global institutional collaborations, and the intersecting realms of interdisciplinary interactions on a global scale.

In a world where intellectual property effortlessly crosses national borders and knowledge operates on a global stage, understanding the complexities of collaboration is vital. This study illuminates global collaboration patterns, specifically examining interactions between top (elite) countries, further delving into the landscape of collaboration within national laboratories and universities, and potentially extending to individual researcher collaborations on a global scale. The utilization of OpenAlex serves as a lens to extract insights into the collaborative landscape and assess data quality within this extensive repository.

The implications of this study extend across policy, practice, and future research, providing valuable insights into intellectual discoveries that transcend geographic boundaries. Unraveling the intricate pathways through which scientific knowledge circulates globally, these findings aim to guide decisions pertaining to international partnerships in advancing science. Designed for those captivated by global collaboration across diverse domains, this study serves as a resource and starting point for understanding its impact on scientific progress and potentially safeguarding discoveries within the realm of intellectual property. Rooted in intellectual curiosity, this pursuit not only satiates personal interest but also establishes a groundwork for subsequent investigations, benefitting various stakeholders and communities.

2. Methodology

2.1 A Brief Introduction to OpenAlex

OpenAlex is an openly accessible catalog of the global research system. The analysis of global collaboration in this study relies on the OpenAlex database, which has been reported by the present author.[9] At its core, the OpenAlex dataset describes various scholarly entities and their interconnected relationships, forming an extensive network or, more technically, a heterogeneous directed graph. This dataset includes entities such as works, authors, sources, institutions, concepts, publishers, and funders, creating a data representation of the global research landscape, as illustrated in Figure 1.

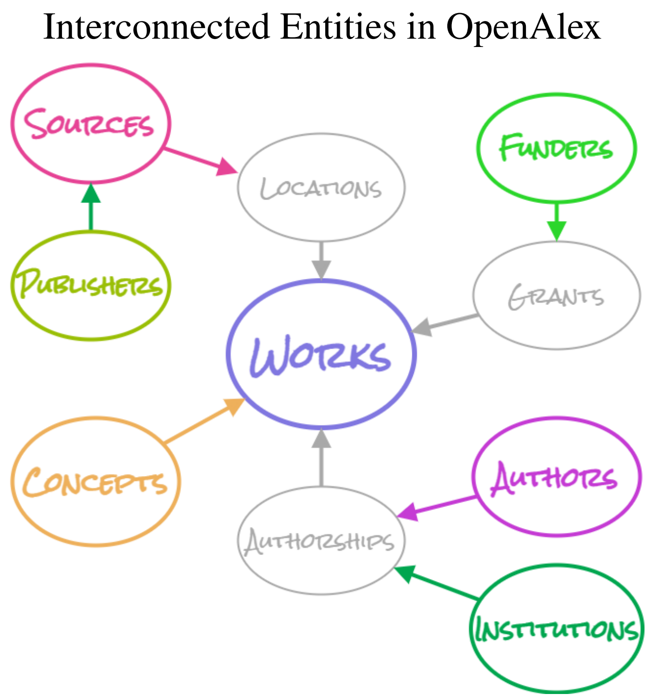


Figure 1. The intricate web of connections among various entities within the OpenAlex dataset, including works (scholarly documents), authors (creators of works), sources (hosting venues for works), institutions (affiliated universities and organizations), concepts (topics assigned to works), publishers (organizations distributing works), and funders (organizations funding research).

As of the January 29, 2025 update, OpenAlex boasts impressive coverage* encompassing 263 million works; 60 million open-access works; 33 million works from the Global South;† 8 million datasets; 102 million authors (8 million with Open Re-

*https://openalex.org/stats

†The Global South is a term used to identify regions within Latin America, Asia, Africa, and Oceania.

searcher and Contributor IDs [ORCIDs]); 16 million authors from the Global South; 261,000 sources (45,000 of which are open access); 11,000 publishers; 32,000 funders; 112,000 institutions; and 65,000 concepts.

Table 1 lists the various entities along with URL web links to both an example and the application programming interface (API), which is a set of rules and protocols for building and interacting with software applications. Each entity in OpenAlex is assigned a unique OpenAlex ID, allowing for stability and identification of specific resources within the database. The OpenAlex ID comprises two parts: the base, which is always https://openalex.org/, and the key, a unique primary identifier (first letter) corresponding to the entity's type (e.g., work, author, source, institution, concept, publisher, funder). Canonical external IDs, such as digital object identifier (DOI) for works, ORCID for authors, linking international standard serial number (ISSN-L) for sources, Research Organization Registry (ROR) ID for institutions and funders, and Wikidata ID for concepts and publishers, provide standardized identifiers across various systems. The OpenAlex ID can also be tied to IDs within other databases, such as these examples for works: DOI (doi), Microsoft Academic Graph (mag), and PubMed (pmid).

Table 1. The various accessible entities within OpenAlex along with a URL to an example record and to the API. The following examples were chosen: "The Limitations of Deep Learning in Adversarial Settings"[10] (works), Mark A. Tschopp (authors), DEVCOM Army Research Laboratory (institutions), Army Research Office (funders), Science (sources), American Association for the Advancement of Science (publisher), and Computer Science (concept).

Entity	URL for example entity	URL to access API
Works	https://openalex.org/W2180612164	https://api.openalex.org/works
Authors	https://openalex.org/A5073812740	https://api.openalex.org/authors
Institutions	https://openalex.org/I166416128	https://api.openalex.org/institutions
Funders	https://openalex.org/F4320338281	https://api.openalex.org/funders
Sources	https://openalex.org/S3880285	https://api.openalex.org/sources
Publishers	https://openalex.org/P4310315823	https://api.openalex.org/publishers
Concepts	https://openalex.org/C41008148	https://api.openalex.org/concepts

OpenAlex not only offers a wealth of data but now also presents a user-friendly web interface for exploring and analyzing the dataset, both through a user interface/dashboard* and an API for algorithmic analysis.† This resource proves a valuable database for Science of Science to institutions, researchers, governments, publish-

*https://openalex.org/
†https://api.openalex.org/

ers, funders, and anyone interested in global research and scholarly communication. Its open nature positions OpenAlex as an alternative to paywalled services like Elsevier's Scopus and Clarivate's Web of Science, emphasizing inclusivity, affordability, and availability. The complete dataset is freely available under the Creative Commons Zero license.

2.2 Navigating the JavaScript Object Notation (JSON) Format

Utilizing the OpenAlex API enables research and information retrieval from the dataset. It is essential to understand the nature of information provided by the API. To illustrate, focus on the entity `"works"` that encompasses articles, proceedings, books, datasets, and more. An excerpt from the JSON response of an API call is presented in Figure 2, specifically highlighting the `"meta"` key and its associated contents. JSON is a lightweight data-interchange format that's easy for humans to read and write, and easy for machines to parse and generate. Nested within this key is a dictionary comprising various keys and values. Noteworthy components include the total count of entities returned from the query (247,597,644 records), the response time in milliseconds (54 ms), the page number (pertinent for multiple pages), the count of records returned per page (25), and the count of groups returned (applicable when utilizing a group_by function). The `"page"` and `"per_page"` keys play a pivotal role in navigating through multiple JSON files, particularly when aggregating extensive records surpassing the 200-per-page limit.

JSON format for OpenAlex: `"meta"`


```
"meta": {
    "count": 247597644,
    "db_response_time_ms": 54,
    "page": 1,
    "per_page": 25,
    "groups_count": null
},
"results": [
    {
```


Figure 2. The JSON format for OpenAlex within the key `"meta"` that describes the total number of entities returned by the API (`"count"`), the response time (`"db_response_time_ms"`), and other information.

Direct your attention to the subsequent key presented in the JSON response—namely, `"results"`. In Figure 2, the `"results"` key, positioned at the bottom, houses a list of entries (25 in this instance, not shown). Each entry follows a dictionary format, featuring various keys that can be leveraged to extract pertinent information associated with each entry. Figure 3 provides an initial glimpse into the first entry within "results". The default is to sort the entries by citation count, so this first entry is the Lowry et al.[11] 1951 journal article entitled *Protein Measurement with the Folin Phenol Reagent*, which was cited 304,009 times. This entry furnishes details such as the DOI (10.1016/s0021-9258(19)52451-6), OpenAlex ID (W1775749144),

JSON format for OpenAlex: "results"


```
"id": "https://openalex.org/W1775749144",
"doi": "https://doi.org/10.1016/s0021-9258(19)52451-6",
"title": "PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT",
"display_name": "PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT",
"publication_year": 1951,
"publication_date": "1951-11-01",
"ids": {
    "openalex": "https://openalex.org/W1775749144",
    "doi": "https://doi.org/10.1016/s0021-9258(19)52451-6",
    "mag": "1775749144",
    "pmid": "https://pubmed.ncbi.nlm.nih.gov/14907713"
},
"language": "en",
"primary_location": {
    "is_oa": true,
    "landing_page_url": "https://doi.org/10.1016/s0021-9258(19)52451-6",
    "pdf_url": null,
    "source": {
        "id": "https://openalex.org/S140251998",
        "display_name": "Journal of Biological Chemistry",
        "issn_l": "0021-9258",
        "issn": [
            "1083-351X",
            "0021-9258",
            "1067-8816"
        ],
        "is_oa": true,
        "is_in_doaj": true,
        "host_organization": "https://openalex.org/P4310320990",
        "host_organization_name": "Elsevier BV",
        "host_organization_lineage": [
            "https://openalex.org/P4310320990"
        ],
        "host_organization_lineage_names": [
            "Elsevier BV"
        ],
        "type": "journal"
    },
    "license": "cc-by",
    "version": "publishedVersion",
    "is_accepted": true,
    "is_published": true
},
"type": "article",
"type_crossref": "journal-article",
"indexed_in": [
    "crossref"
],
"open_access": {
    "is_oa": true,
    "oa_status": "hybrid",
    "oa_url": "https://doi.org/10.1016/s0021-9258(19)52451-6",
    "any_repository_has_fulltext": true
},
```


Figure 3. The JSON format for OpenAlex within the key "results" that contains information about OpenAlex ID, DOI, title, publication year, location of the published work, whether it is open access, etc.

title, publication year, language, publication location, work type, and open-access status (`true`). It is imperative to recognize that this overview only captures a fraction of the comprehensive information encapsulated within an entry. Additional data encompasses authorship details, countries involved, affiliated institutions, publishing charges, citation counts and rankings, associated concepts, work locations, optimal open-access repositories, alignment with sustainable development goals, references, abstract content, links to the list of citations, and citation counts per year (since 2012). Accessing this wealth of information programmatically enables researchers to swiftly address queries, aggregate relevant data, and conduct diverse analyses, thereby enhancing their understanding of the available information.

2.3 API Calls within OpenAlex

Navigating from a dataset of 263 million works to a refined understanding of a specific area of interest requires leveraging the flexibility offered by additional parameters in the API URL. The OpenAlex API allows users to tailor their queries by filtering, sorting, grouping, and conducting keyword searches within the dataset—explained within the technical documentation of OpenAlex.[12] To illustrate this, Table 2 provides a step-by-step example of constructing a URL for API access. In the initial step, initiate with the API address and specify the entity type that is to be accessed (works), bearing in mind that filters may vary depending on the entity type. Subsequent steps showcase how to progressively filter the dataset. For instance, the second step demonstrates filtering for a specific type of work, such as articles (inclusive of both journal articles and conference proceedings). The third step narrows down the dataset to records with publication years from 2019 to 2024 (inclusive). Further refinements follow, such as filtering for works associated with institutions in the U.S., limiting the dataset to publications featuring global collaboration (>1 institution), specifying a minimum citation threshold (>100), and finally, focusing on publications associated with the level 1 concept of AI (using the OpenAlex ID C154945302). This particular example highlights AI to demonstrate the API's capability to narrow down to a highly relevant and rapidly evolving field. This particular query results in 5,153 publications, offering a focused dataset for more in-depth exploration across various dimensions within the JSON information, but this can alternatively be applied to sort through and understand multiple factors to examine the changing nature of global collaboration.

Table 2. An example of accessing the OpenAlex API in successive steps: 1) initial URL, 2) only journal articles and conference proceedings, 3) for publication years from 2019 to 2024, 4) only when U.S. is one of the author countries, 5) when there are multiple countries involved, 6) when citations are above 100, and 7) when one of the concepts is AI (accessed 2025-01-21).

Step	URL to access API	No. of records
1	https://api.openalex.org/works	263,328,007
2	https://api.openalex.org/works?filter=type:article	198,275,445
3	https://api.openalex.org/works?filter=type:article,publication_year:2019-2024	38,211,899
4	https://api.openalex.org/works?filter=type:article,publication_year:2019-2024,institutions.country_code:US	4,723,179
5	https://api.openalex.org/works?filter=type:article,publication_year:2019-2024,institutions.country_code:US,countries_distinct_count:>1	1,717,707
6	https://api.openalex.org/works?filter=type:article,publication_year:2019-2024,institutions.country_code:US,countries_distinct_count:>1,cited_by_count:>100	35,638
7	https://api.openalex.org/works?filter=type:article,publication_year:2019-2024,institutions.country_code:US,countries_distinct_count:>1,cited_by_count:>100,concepts.id:C154945302	5,153

3. Results and Discussion

3.1 Temporal Evolution of Global Science

Before addressing questions about global collaboration, it is crucial to understand some fundamental aspects of the OpenAlex database, particularly the evolution of the number of articles over time. Figure 4 illustrates the annual production of articles since 1900, encompassing both peer-reviewed journal articles and conference proceedings. The plot reveals a nearly monotonic increase in the number of articles per year, with notable interruptions during World War I and World War II, as indicated in the plot. Intriguingly, the peak in the maximum number of articles per year occurs in 2016 and then another peak around 2020, followed by a decline, notably influenced by the global COVID-19 pandemic. The red line serves as a visual guide, signaling that the number of articles doubles approximately every 14 years. Notably, the logarithmic scale on the y axis captures the nonlinear growth in publications over time. The OpenAlex database currently hosts an extensive collection of over 200 million articles. As an illustration of its dynamic nature, the database witnesses a continuous influx of new content; to put this in perspective, the past few years have had about 7 million articles added to OpenAlex annually, resulting in the addition of $\approx$ 140,000 articles each week. Notably, the curation of the database in terms of these articles is performed through scripting rather than manual curation, so this relies upon good scripting tools that extract metadata at some point in the process. The quick response (QR) code provides direct access to the most recent data used for this plot.

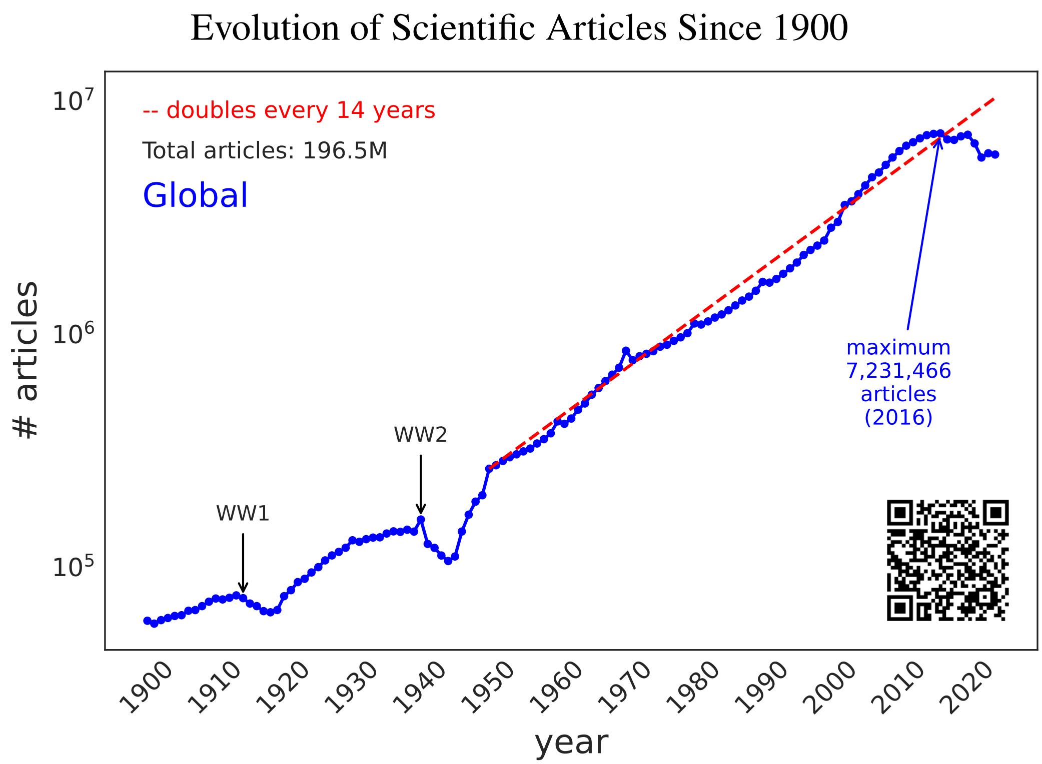


Figure 4. The evolution of the number of scientific articles produced globally per year since 1900. The "article" type refers to peer-reviewed journal articles and conference proceedings. The beginning of the first two world wars are identified as is the maximum in the produced articles per year (peaks in 2016). The red line shows a doubling of articles every 14 years as a guide to the eye. In this figure and subsequent figures, the QR code in the lower right-hand corner is a link to the OpenAlex data source.

Figure 5 delves into the production of book chapters over time, mirroring the format of Figure 4 with key distinctions. The OpenAlex dataset contains over 20 million book chapters, $\approx 10\%$ of the total articles. Unlike the decreasing trend in articles post-2020, the number of book chapters steadily rises, with the maximum observed in 2023. Noteworthy is the historical fluctuation in book chapter production, especially during world wars, giving way to a post-World War II era marked by a consistent increase. This evolution sees the number of book chapters doubling approximately every 10 years.

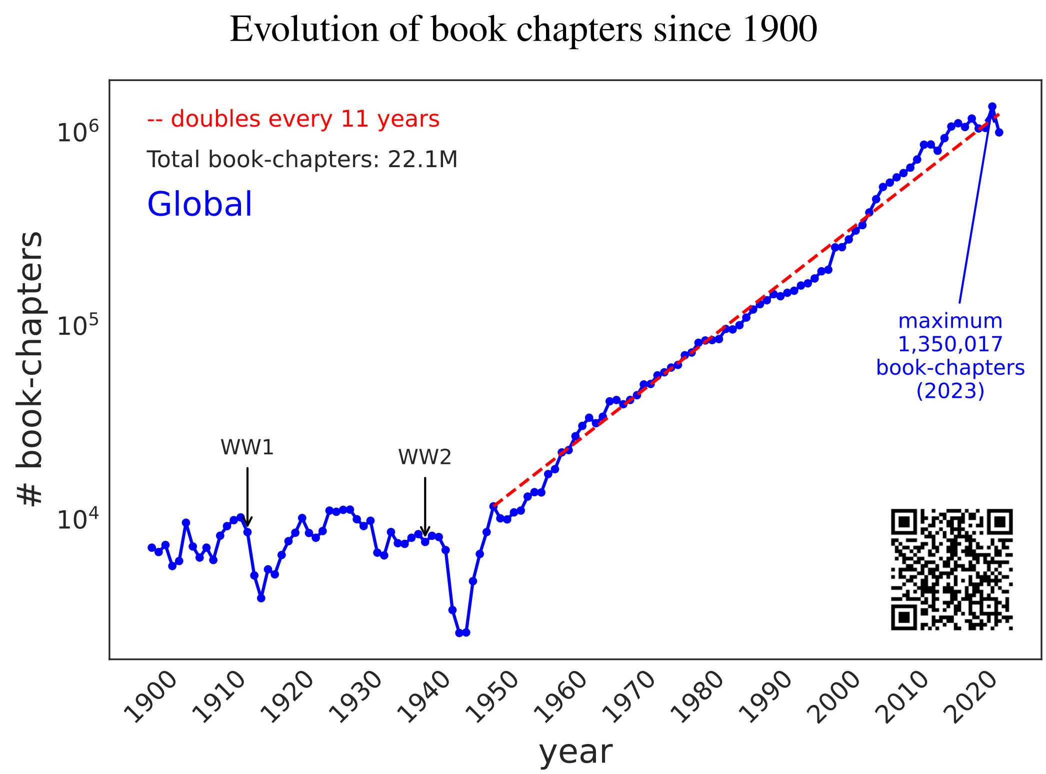


Figure 5. The evolution of the number of book chapters produced globally per year since 1900. The beginning of the first two world wars are identified as is the maximum in the produced book chapters per year (peak is in 2023). The red line shows a doubling of book chapters every 10 years as a guide to the eye. The total number of book chapters is ≈ 10% of the total scientific articles.

Turning attention to another medium for disseminating scientific knowledge, Figure 6 explores the annual production of books over time. The format aligns with Figures 4 and 5, revealing interesting trends. The maximum number of books occurs in 2011, with a substantial decline in subsequent years. Despite a historical trend of monotonic increase since around 1950, this pattern shifts in the last decade. While multiple factors may contribute to this decline, the popularization of the h-index as a metric for assessing scientific productivity and impact could play a significant role in the incentive system for scientists; i.e., one book takes significantly more effort to produce than one journal article, but each can increase the h-index by one. The logarithmic y axis accentuates the magnitude of this decrease over the past decade.

Patterns within the global distribution of articles can be better understood by examining the countries listed in the authors' affiliations. Table 3 provides a comprehensive breakdown of countries listed within the author affiliations for articles in the OpenAlex database. The initial two rows distinguish whether the country is

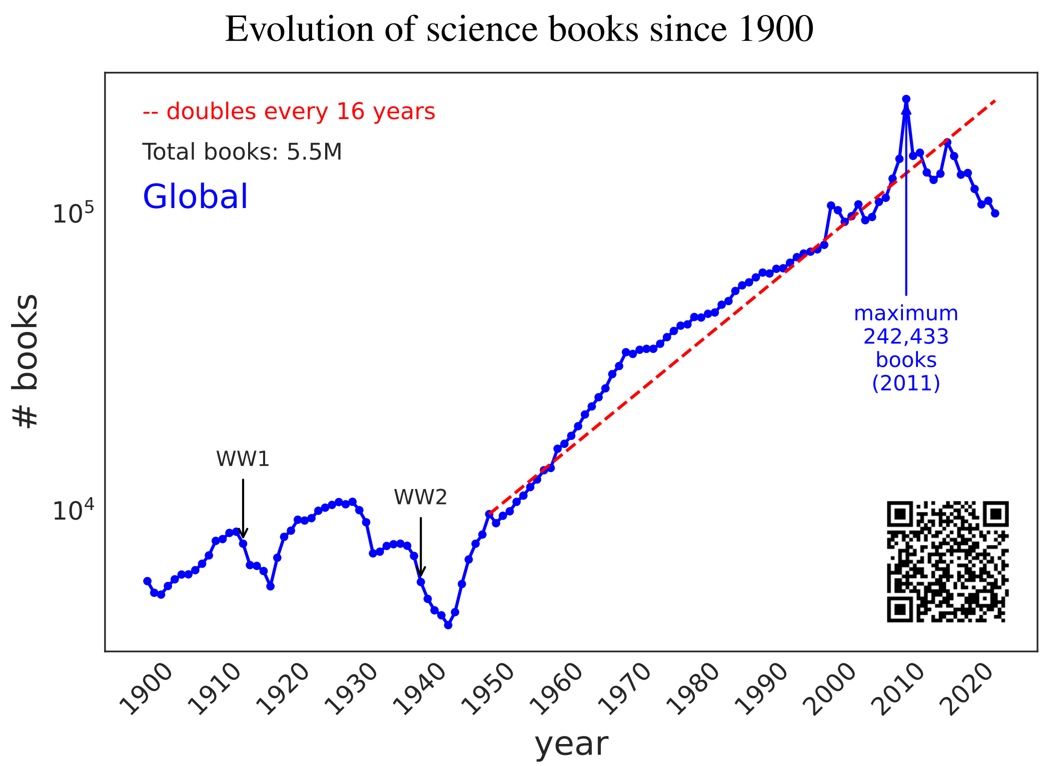


Figure 6. The evolution of the number of science books produced globally per year since 1900. The beginning of the first two world wars are identified as is the maximum in the produced books per year (peaks in 2011). The red line shows a doubling of books every 16 years as a guide to the eye. The total number of books is ≈ 2.5% of the total scientific articles.

mentioned in the article. Country designations in this report follow the ISO 3166-1 alpha-2 standard, as used in the OpenAlex dataset. For example, "GB" refers to the United Kingdom of Great Britain and Northern Ireland (shortened to United Kingdom within this report), following its official ISO code, even though "UK" is also commonly used. This convention is retained throughout to ensure consistency with the dataset. Notably, ≈ 55% of articles lack a specified country. Further investigation into a subset of these articles reveals instances where institutional information is also absent. Various factors contribute to this, ranging from articles featuring only author names without affiliations to algorithmic limitations in extracting such details. Despite the dataset's imperfections, more than 88 million articles include author affiliations with country information, leading to over 100 million country-paper pairs.* It is assumed that trends within this subset are representative of the full sample population.

*106,897,485 country–paper pairs representing 208 countries

Table 3. The breakdown of countries within author affiliations for all articles in OpenAlex. The first two rows highlight the number of unknown and known country affiliations within OpenAlex and their corresponding percentages. For the country information, "% of total" refers to the percent of articles that include at least one author with an affiliation from that country. The column "% collaboration" is the percentage of articles with that country listed that have other countries as well (accessed 2025-01-21).

Country code	Country	No. of publications	Total (%)	Collaboration (%)
unknown	unknown	109,141,394	55.0	NA
known	known	89,133,985	45.0	16.2
US	United States	23,830,224	26.7	23.4
CN	China	12,923,605	14.5	17.0
GB	United Kingdom	6,172,271	6.9	40.1
JP	Japan	5,129,404	5.8	20.2
DE	Germany	4,925,157	5.5	39.0
FR	France	3,580,269	4.0	40.4
IN	India	3,389,068	3.8	18.9
CA	Canada	3,121,236	3.5	39.7
BR	Brazil	2,970,222	3.3	20.8
IT	Italy	2,660,691	3.0	40.0
ES	Spain	2,417,293	2.7	38.5
AU	Australia	2,327,297	2.6	43.7
ID	Indonesia	2,269,917	2.5	7.1
RU	Russian Federation	2,209,201	2.5	24.3
KR	Korea	1,835,459	2.1	26.6
NL	Netherlands	1,571,392	1.8	48.9
PL	Poland	1,496,466	1.7	23.7
TR	Turkey	1,306,877	1.5	18.7
CH	Switzerland	1,151,954	1.3	57.0
SE	Sweden	1,101,376	1.2	48.1
IR	Iran	1,085,372	1.2	21.3
BE	Belgium	1,004,377	1.1	49.6
TW	Taiwan	986,823	1.1	28.5
MX	Mexico	833,061	0.9	32.2
DK	Denmark	687,731	0.8	48.7
IL	Israel	649,184	0.7	39.6
AT	Austria	621,661	0.7	52.2
EG	Egypt	586,119	0.7	37.7
MY	Malaysia	574,637	0.6	38.5
CZ	Czechia	570,220	0.6	49.6
PT	Portugal	569,846	0.6	46.7
ZA	South Africa	553,178	0.6	39.6
FI	Finland	539,316	0.6	45.9
NO	Norway	526,111	0.6	51.0
SA	Saudi Arabia	485,156	0.5	65.9
UA	Ukraine	483,617	0.5	24.4
AR	Argentina	470,584	0.5	31.9
GR	Greece	459,960	0.5	41.6
SG	Singapore	449,225	0.5	53.9
PK	Pakistan	429,360	0.5	49.2
CO	Colombia	414,689	0.5	29.3

Figure 7 presents a bubble plot illustrating the contribution of various countries to articles in the OpenAlex database. Each bubble's size corresponds to the frequency with which a particular country is mentioned in the affiliations of articles. This dynamic visualization provides a snapshot of the cumulative contributions of countries across all articles, emphasizing the prevalence of each country in scholarly publications. The bubble plot aligns with the insights presented in Table 3 but adds a visual dimension to showcase the relative impact of different countries on a global scale.

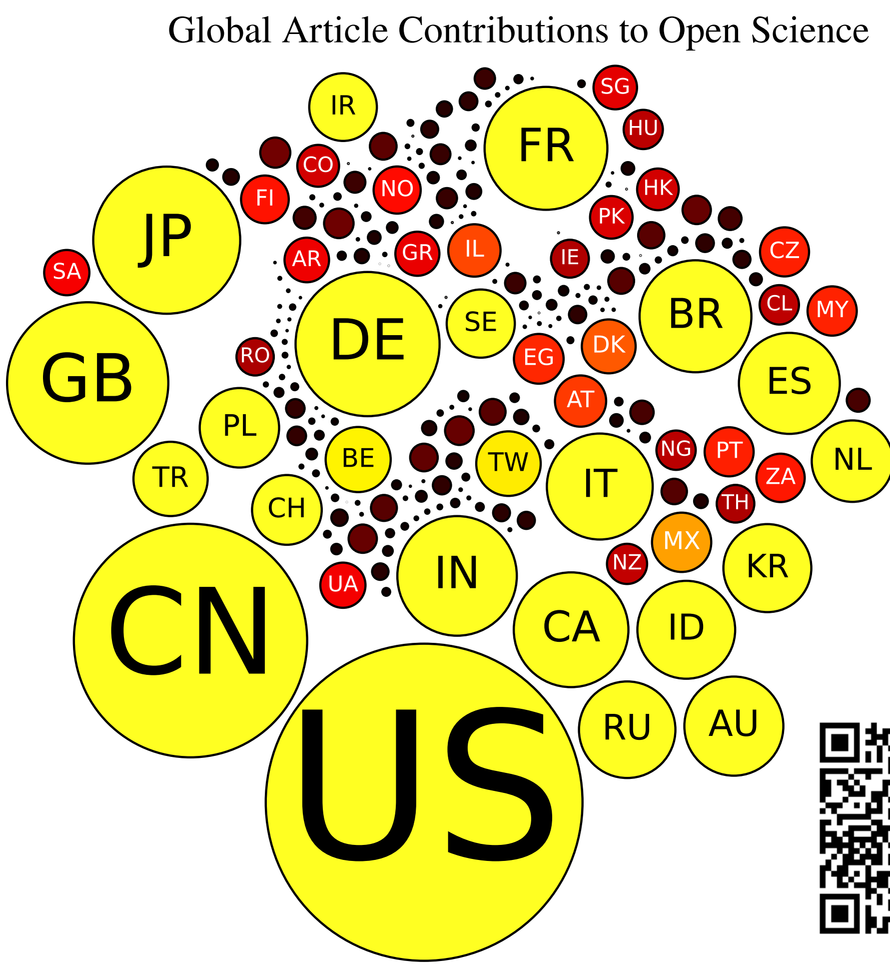


Figure 7. A bubble chart illustrating the number of scientific articles contributed by each country. The size of each circle is proportional to the total number of articles listing affiliations from that specific country. For the full names of countries, refer to Table 3.

It is crucial not to interpret the percentages in Table 3 and the bubble sizes in Figure 7 as the proportion of authors from the listed countries. For example, a single publication may enumerate multiple countries in the affiliation section, counting once for each distinct country. Additionally, a publication with 10 authors from one country contributes only once to that country. The “% total” represents the proportion of total articles that have that country listed; i.e., at least one of the authors is from an institution that is located within that country. The sum of all these column precentages exceeds 100% since multiple countries can be listed for an article. The “% collaboration” column takes the subset of all articles for each particular country and lists the proportion of those articles that have multiple countries listed, i.e., global collaboration. This reveals how often other countries are listed for those publications. While this offers a snapshot of global science and publications, it does not capture the temporal evolution of articles, changes in research concepts over time, or the evolution of collaboration between countries. Subsequent analyses will delve into these aspects.

3.2 Temporal Evolution of Scientific Productivity for Top Countries

Exploring the evolution of scientific articles over time provides intriguing insights into the scientific output of different countries. In Figure 8, the continuous growth in the number of articles per year with affiliations from the U.S. is evident, with occasional dips during the first and second world wars. Remarkably, the peak in articles authored by the U.S. occurred in 2020, coinciding with the global impact of the COVID-19 pandemic.

Figure 9 shifts the focus to China. The rapid pace of article production, highlighted by the steep trajectory on the y axis, sets China apart from other top-contributing countries. Examining the period stretching from the 1900s to the 1970s, China’s output remained below 1,000 articles per year, reaching the milestone of 1,000 articles annually with affiliations from China around 1980. Notably, China’s scientific productivity during the world wars appears less affected, possibly due to lower article numbers or the country’s relative neutrality, which could have shielded it from the direct impact experienced by nations deeply involved in the conflicts. Also of note is the rapid ascent in articles published around the 2000s, which experienced a sizeable dip in the 2010s, and is now on a path where the rate of scientific output doubles every 5 years. Approximately 90% of China’s articles have been produced in the last 20 years, and 99% of China’s articles have been produced in the last 30 years.

Figures 10–12 extend this analysis to the United Kingdom, Japan, and Germany, respectively. Following a trajectory parallel to the U.S. trend, these countries exhibit

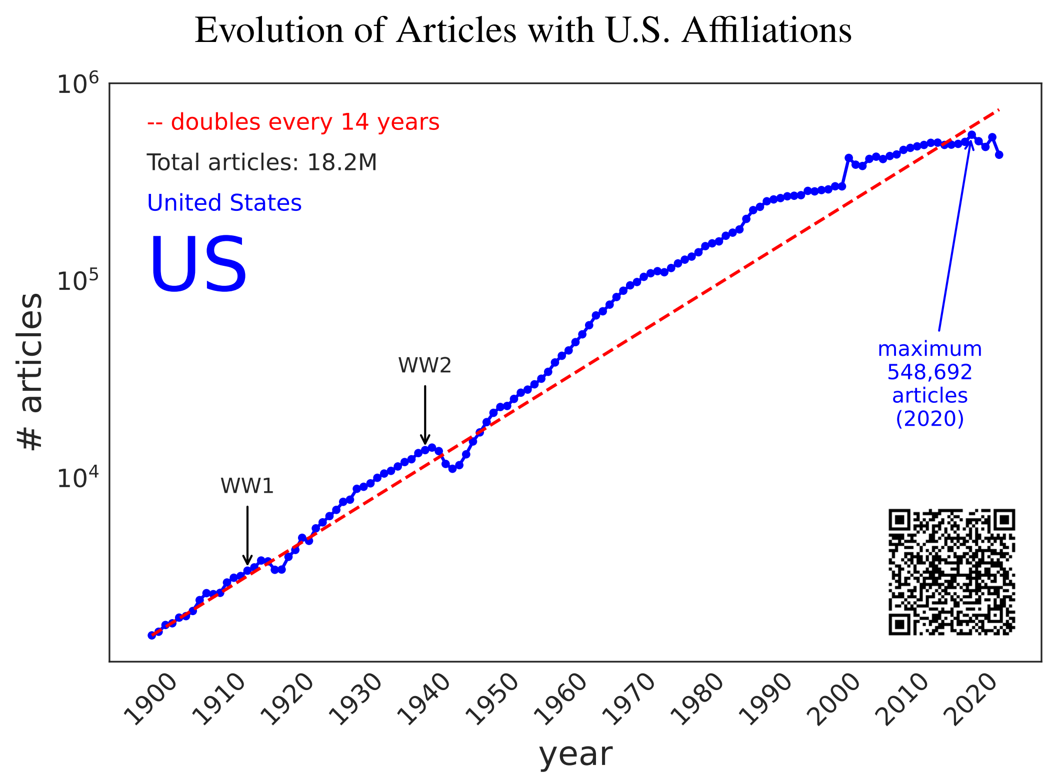


Figure 8. The evolution of the number of articles produced globally per year since 1900 that list an author affiliation located within the U.S.

Figure 9. The evolution of the number of articles produced globally per year since 1900 that list an author affiliation located within China.

overall growth, punctuated by dips during wartime. Japan, for instance, experiences a notable productivity dip during World War II. Subsequently, it resumes the upward trend typical of scientific article output until the mid-to-late 20th century. However, Japan's output has stabilized over the last few decades. This persistent but plateauing productivity prompts intriguing questions about the dynamics within a world where scientific output is on a continuous ascent. Unlike other nations that demonstrate sustained growth, Japan's consistent but unchanging productivity over recent decades may suggest a unique position or challenges within the evolving landscape of global scientific contributions.

Germany, on the other hand, faces the most substantial loss in scientific output during both world wars. However, Germany's recovery following World War II is remarkable, with a resurgence along the original steady growth trend similar to other countries. Germany's ability to recover and rejoin the scientific growth trend underscores its resilience and adaptability, contributing to the complex interplay of factors shaping the scientific output of different nations.

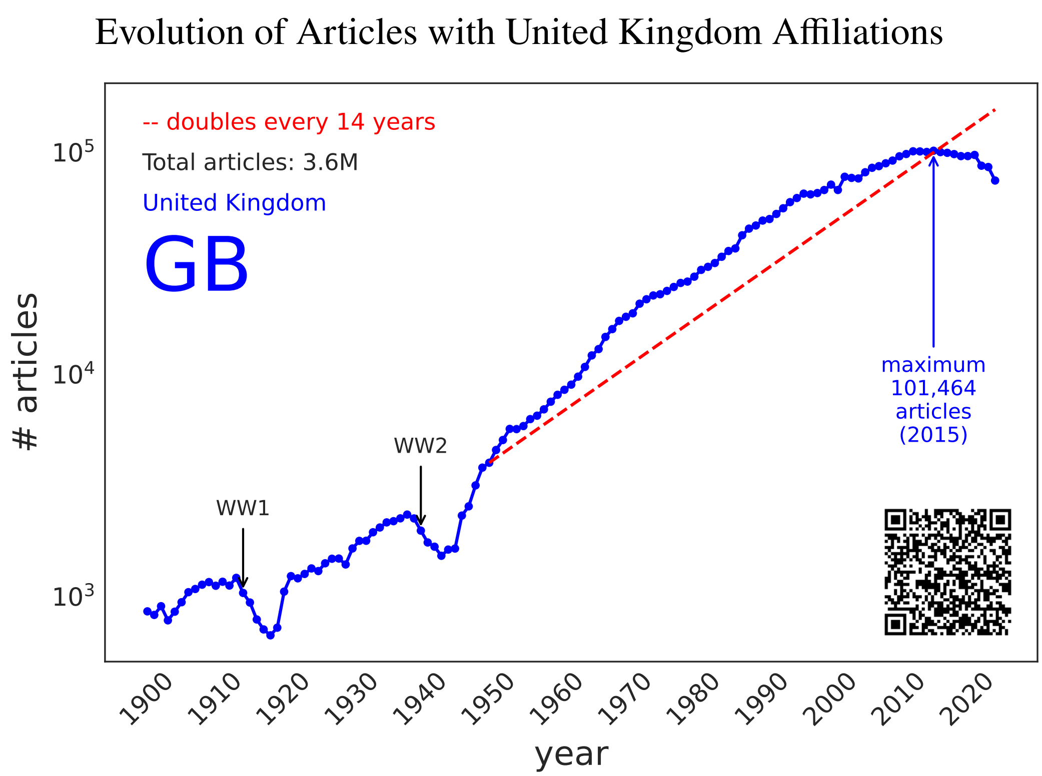


Figure 10. The evolution of the number of articles produced globally per year since 1900 that list an author affiliation located within the United Kingdom of Great Britain and Northern Ireland.

Evolution of Articles with Japan Affiliations

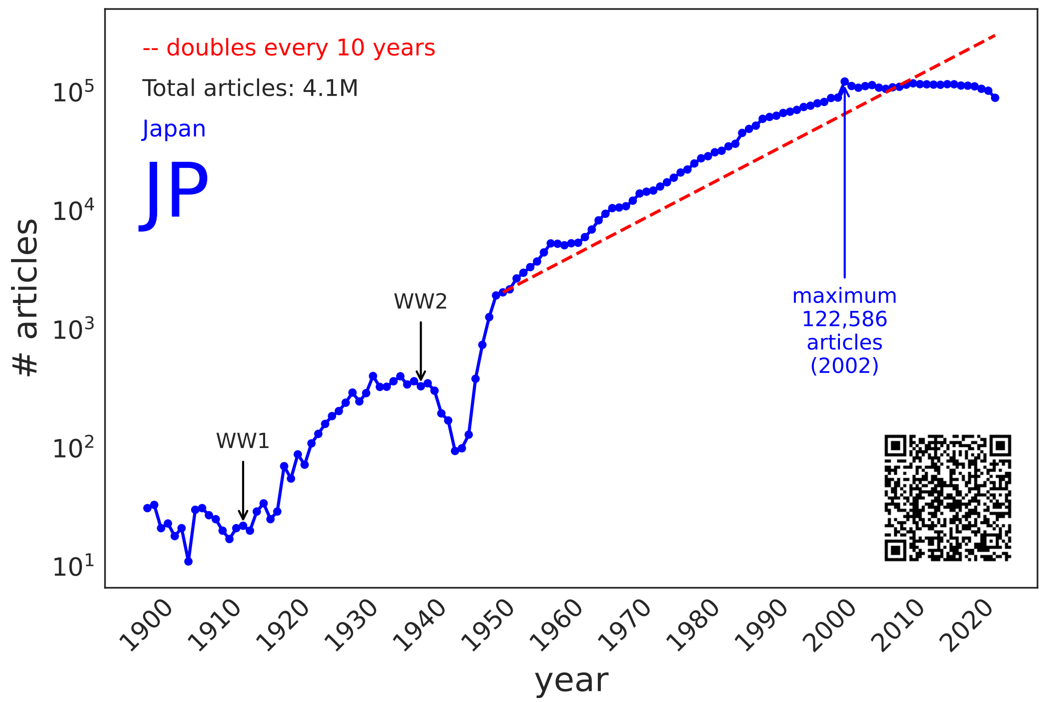


Figure 11. The evolution of the number of articles produced globally per year since 1900 that list an author affiliation located within Japan.

Evolution of Articles with Germany Affiliations

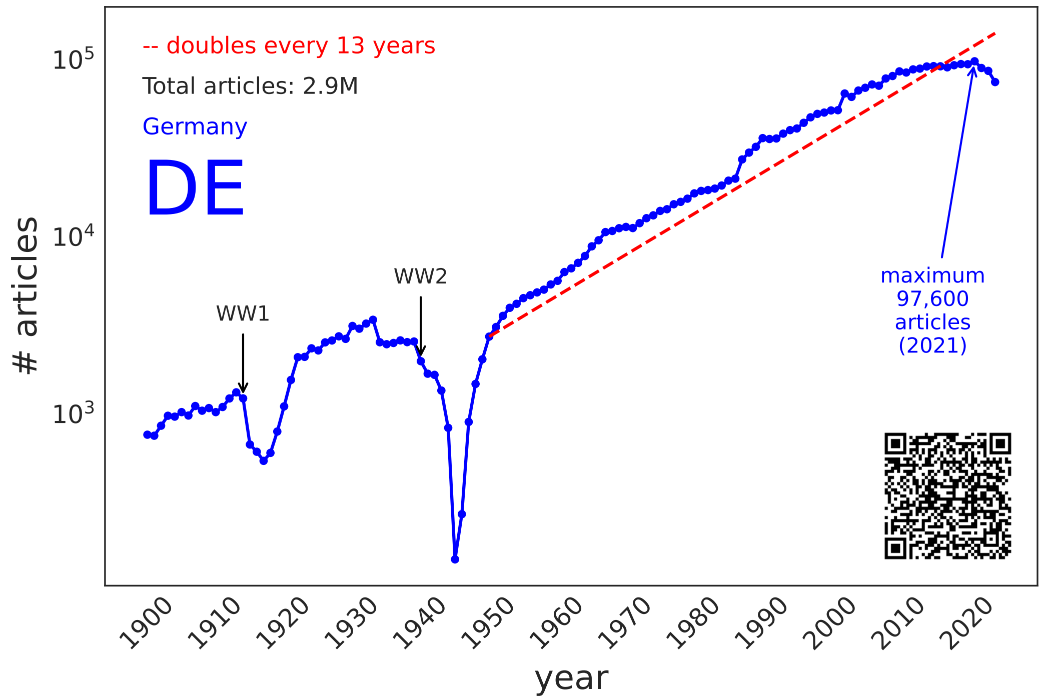


Figure 12. The evolution of the number of articles produced globally per year since 1900 that list an author affiliation located within Germany.

This exploration underscores the global dynamics of scientific productivity, capturing unique trajectories for each country and revealing the impact of historical events on their scientific output. Further examination in subsequent figures delves into specific aspects, providing a comprehensive understanding of the evolving landscape of scientific contributions.

As the evolution of scientific collaboration is examined, the question arises: what proportion of articles, over time, involves contributions from multiple countries? Figure 13 illustrates this trend, showcasing the percentage of collaborative articles since 1900. Leveraging OpenAlex, articles with distinct affiliations from more than one country were identified and then grouped by publication year. Intriguingly, the percentage of collaborative articles remained relatively constant in the first half of the 20th century but exhibited a steady increase thereafter, surpassing 20% in recent years. Indeed, the maximum number of collaborative articles was in 2024, the last full publication year in this dataset. This trend prompts contemplation about the factors driving such collaboration. Is it fueled by incentives for scientists, such as access to added capabilities, instrumentation, expertise, and impact? Or does it reflect the increasing trend of joint affiliations by scientists, whereby they hold appointments with institutions in multiple countries, enabling them to draw additional resources towards research? Or could it be fueled by the mobility of graduate students and postdoctoral researchers, prior to finding more permanent roles? While not explored in this study, some of these questions can also be answered via large databases like OpenAlex.

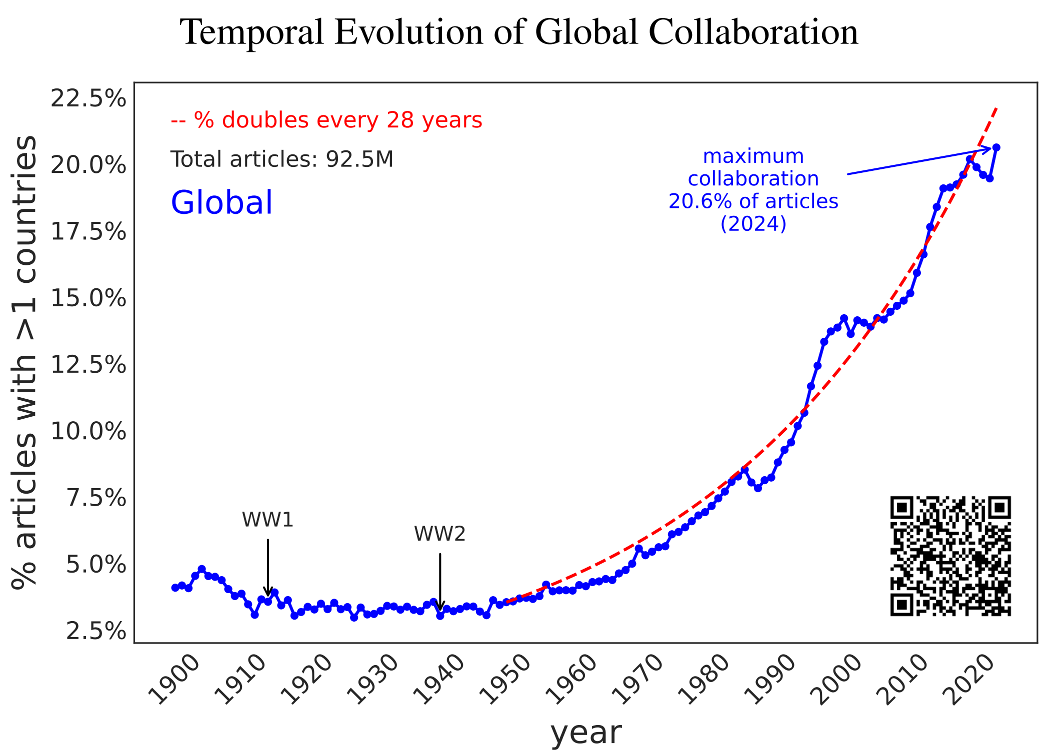


Figure 13. The trend in the number of articles with authors from multiple countries since 1900.

3.3 Frequency of Countries, Institutions, Authors, and Concepts

Given the established base rate of collaboration exceeding 20%, the exploration now delves into the distribution of distinct countries involved in these collaborative publications. Figure 14 delineates the percentage of articles against the count of distinct countries within author affiliations. In the plot's lower left corner, the substantial presence of articles without any specified country—totaling 105 million out of the 198 million articles in the dataset—becomes evident, constituting $\approx 53\%$ of the dataset. Beyond this, $\approx 40\%$ of articles feature one distinct country, with a subsequent monotonic decrease in percentages as the count of distinct countries increases. To accentuate these percentages, the plot incorporates annotations listing them beside the respective data points, while the count of distinct countries is highlighted below each point in red. The logarithmic scale on the y axis effectively reveals the continual decrease, even for small percentages that may span several orders of magnitude, providing a detailed perspective on articles with different numbers of

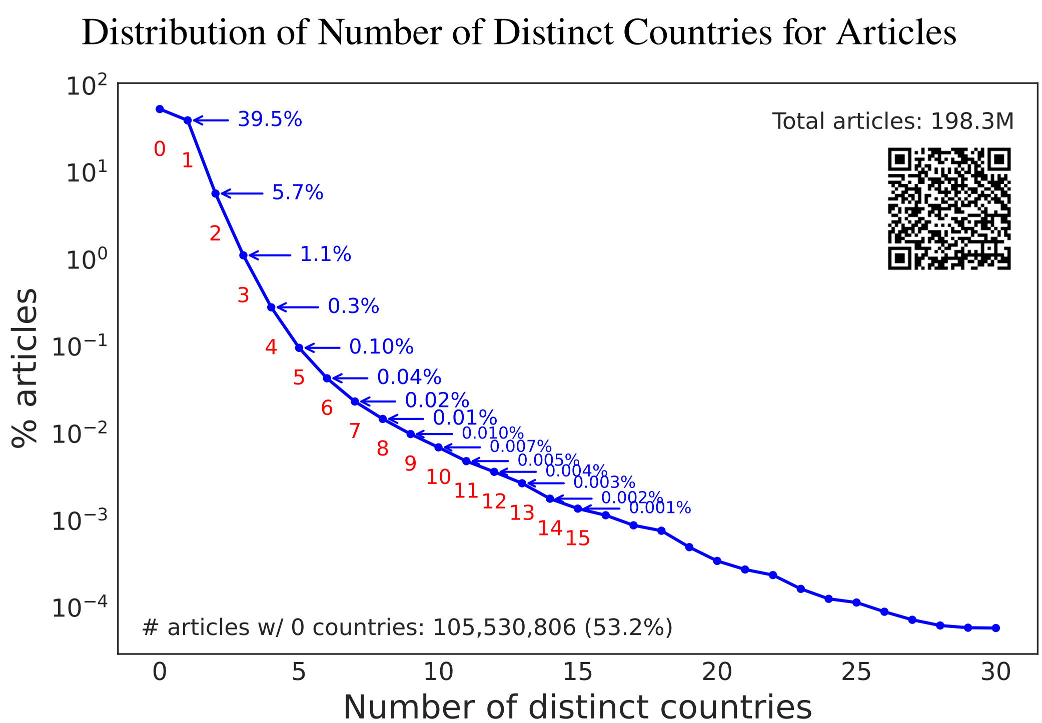


Figure 14. The distribution of the percentage of articles involving various numbers of distinct countries within the author affiliations. The percentages represent the proportion of articles relative to the total number of articles in the database. The y axis employs a logarithmic scale, emphasizing the continual monotonic decrease in percentage with an increasing number of distinct countries collaborating on a publication.

distinct countries.

Given the prevalence of articles lacking country information, the inquiry shifts to the presence of institutions listed for each article. Figure 15 illustrates the percentage of articles across various counts of distinct institutions. Notably, a substantial fraction of the database comprises articles with zero institutions listed—exceeding 108 million articles or 55%. This considerable proportion closely aligns with the absence of information about distinct countries; specifically, 105,296,698 entries with zero countries also exhibit zero institutions (out of 105,530,648 or 99.8%). Furthermore, a breakdown reveals that 233,943 entries possess institution information but lack country details, while 3,599,661 entries include country information but lack institution specifics. Within these entries, the number of authors emerges as a pivotal factor; single-author articles constitute over 53%, with two- and three-author papers comprising a significant portion of the remaining entries (16% and 9%, respectively). Another important factor is the open-access status of these ar-

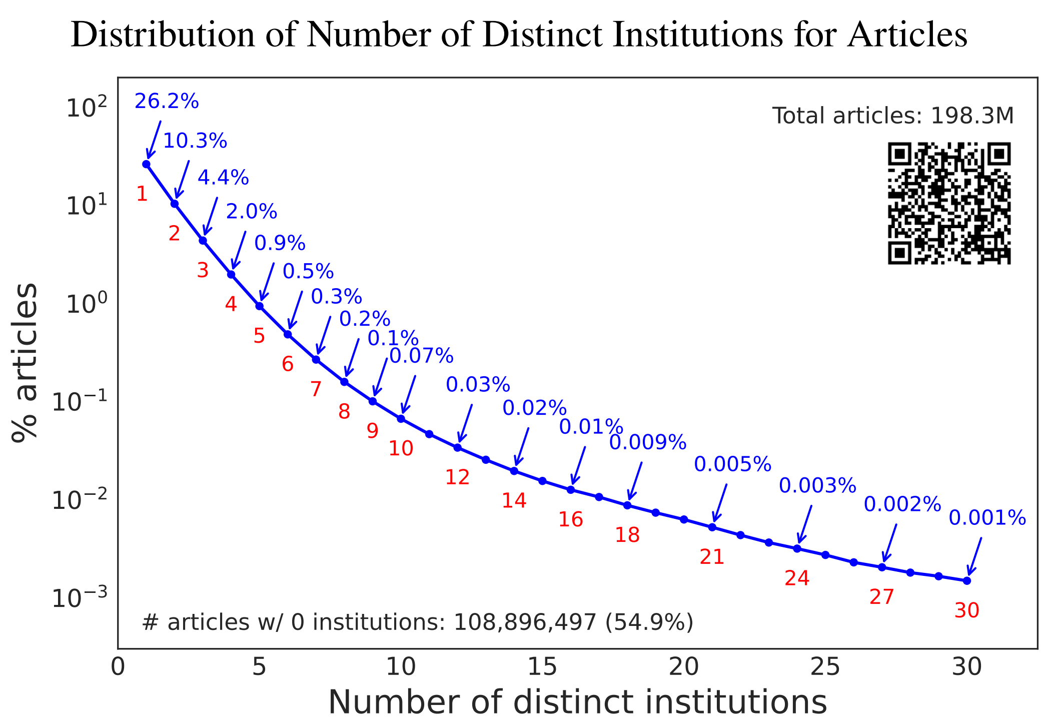


Figure 15. The distribution of the percentages of articles involving different numbers of distinct institutions within the author affiliations. The percentages represent the proportion of articles relative to the total number of articles in the database. The y axis employs a logarithmic scale, emphasizing the continual monotonic decrease in percentage with an increasing number of distinct institutions collaborating on a publication.

ticles; nearly 90% of articles with zero countries are not open-access, presenting possible problems with extracting key information such as author affiliations and countries. This interplay underscores the faceted nature of authorship, open-access status, institutions, and country affiliations within the dataset.

Exploring the composition of articles within the database, examining authorship reveals the distribution of articles across varying author counts. Figure 16 depicts the percentage of articles relative to the number of distinct authors per article. Notably, articles with zero authors constitute just under 9 million entries, roughly 4.5% of the total 198 million articles. Among the remaining articles, the majority feature a single author, followed by two-author publications as the next highest percentage. Subsequently, three-author articles, four-author articles, and so forth, demonstrate a diminishing percentage as the number of authors increases. A noteworthy trend emerges, where the percentage of articles decreases monotonically with each incremental increase in the number of authors, extending up to 30 authors, as illustrated in Figure 16. This decreasing relationship is accentuated by employing a logarithmic

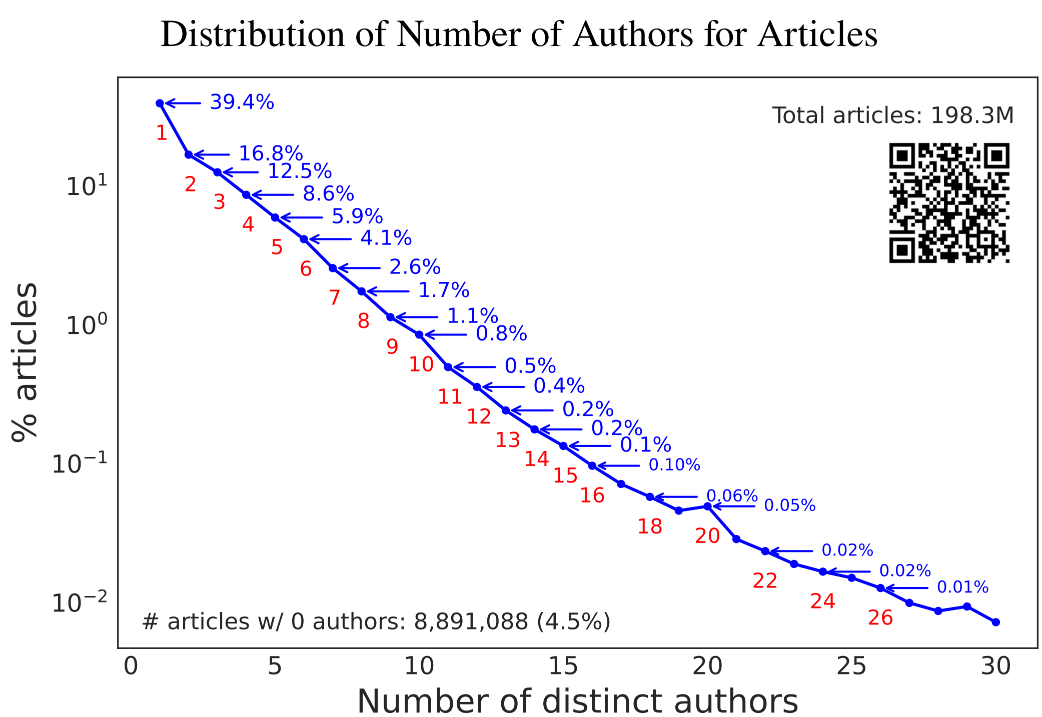


Figure 16. The distribution of percentages of articles involving different numbers of authors. The percentages represent the proportion of articles relative to the total number of articles in the database. The y axis employs a logarithmic scale, emphasizing the continual monotonic decrease in percentage with an increasing number of distinct authors collaborating on a publication.

scale for the y axis to visualizing the authorship distribution within the dataset.

Turning the attention to the diversity of research concepts across articles, Figure 17 presents the percentage of articles corresponding to the number of distinct concepts. It is important to highlight that these concepts result from the ML classification of article titles, abstracts, and journal titles, thereby organizing articles into a taxonomy of research topics. A striking observation emerges: a mere fraction, $\approx$ 12,695 articles, out of the vast 198 million articles in the database, lack any assigned concepts. For the majority of articles, one research concept is predominant, followed by articles with three concepts constituting the second-highest percentage. The distribution continues, encompassing two research concepts and beyond. Intriguingly, the percentage of articles with two to 20 concepts maintains a consistent pattern before monotonically decreasing for counts surpassing 20 concepts. This trend, outlined in Figure 17, is accented by using a logarithmic scale for the y axis to show the distribution of research concepts across the dataset, even for much smaller

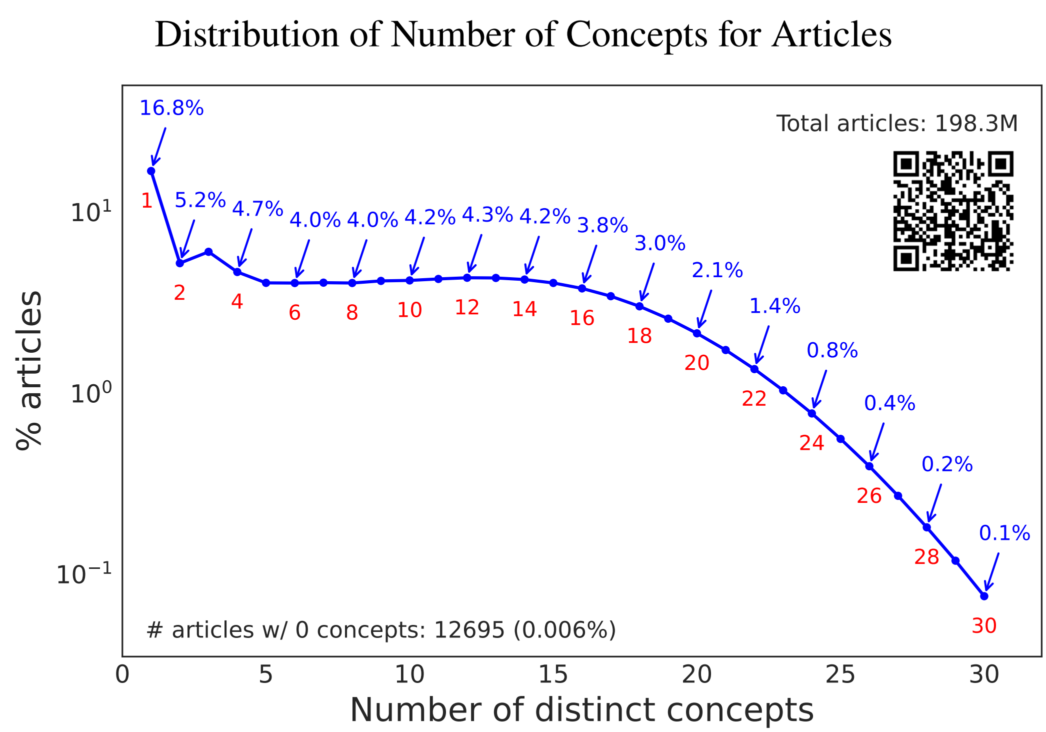


Figure 17. The distribution of percentage of articles involving different numbers of research concepts tagged within OpenAlex. The percentages represent the proportion of articles relative to the total number of articles in the database. The y axis employs a logarithmic scale to highlight the wide range of percentages for different numbers of concepts.

percentages.

3.4 Global Collaboration between Leading Countries

Having delved into the distinct countries and institutions represented in the database, the exploration now turns to the frequency of collaboration between countries. Figure 18 presents a heat map illustrating global collaboration among the top 12 countries in terms of scientific article production from 2019 to 2024.

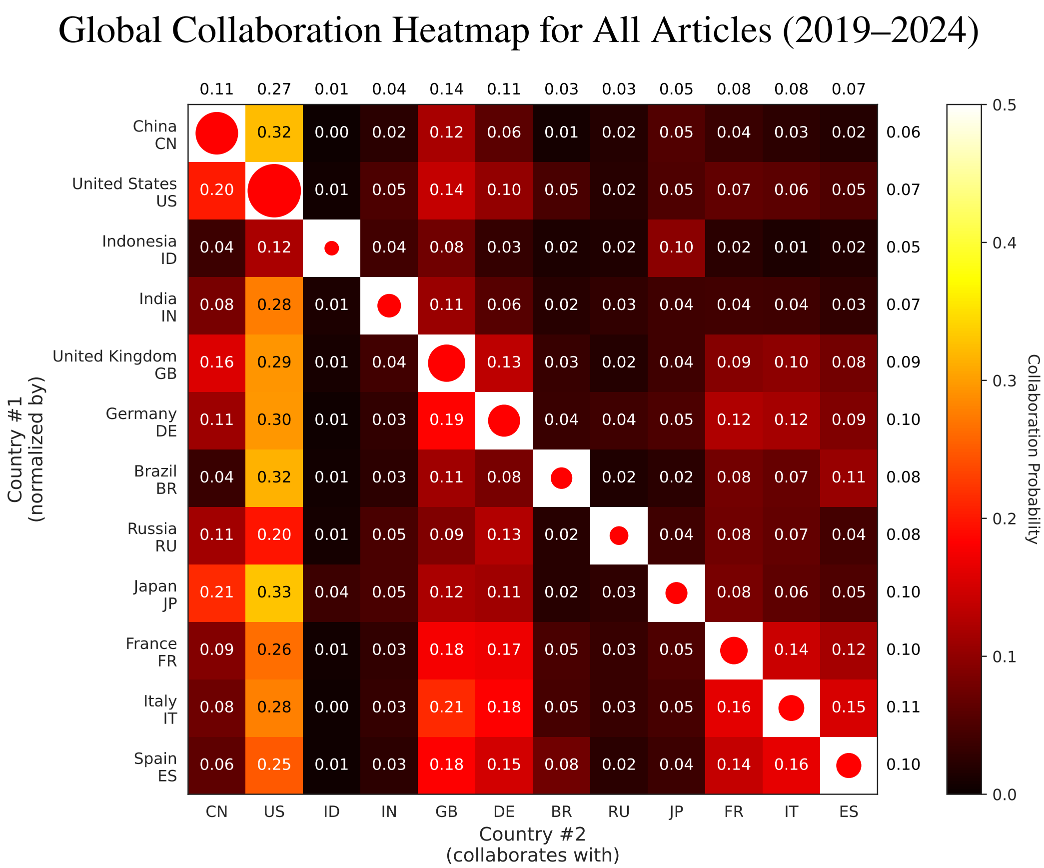


Figure 18. A heat map illustrating global collaboration among the top 12 most prolific countries from 2019 to 2024. Countries are arranged based on their scientific article productivity, with the y axis representing the top-down order and the x axis depicting the left-to-right order. The circular regions (highlighted in red) along the diagonal are proportional to the number of articles featuring multiple countries for Country 1. Each row is normalized by this count, converting each element within the heat map into a probability indicating how often a multi-country article for Country 1 (left) also contains Country 2 (below). The egocentric collaboration index is displayed to the right and the collaborative role index is displayed above the heat map.

First, the order of the countries on the left reflects their total article production over the specified time span from top to bottom (i.e., CN is largest), encompassing both articles with exclusive affiliation to that country and articles featuring multiple countries in their affiliations. This informative heat map format will be progressively unraveled, aiding in the comprehension of collaborative patterns.

Second, the focus shifts within the heat map to articles where the affiliations involve more than one distinct country. The red circles along the diagonal represent the number of articles featuring multiple countries that include Country 1. Specifically, the size of the red circle at each country's position signifies the number of collaborative articles involving that country. For instance, observing this plot reveals that China holds the top position for the overall number of articles for this time period but the U.S. boasts the highest number of collaborative multi-country articles, evident from the size of its corresponding red circle. Intriguingly, while Indonesia ranks third in total article production globally, its collaborative output with other countries is relatively modest.

Third, attention turns to the values and corresponding colors within individual elements of the heat map. These values denote the probability that Country 2 appears on a multi-country article that also includes Country 1. Importantly, this probability has been normalized by the total number of collaborative articles that Country 1 produces, introducing an asymmetry of colors within the heat map. To illustrate, consider the comparison between the U.S. and China. While both countries have a similar frequency of occurrence in each other's collaborative articles, the normalization by the total output of articles with multiple countries reveals complex patterns. For instance, China is involved in 20% of collaborative articles from the U.S., whereas the U.S. is present in 32% of China's collaborative articles. This asymmetry underscores that some countries have a higher profile for international collaboration that attracts authorship from other countries; i.e., partnerships involving the U.S., China, the United Kingdom, and Germany are common as observed by the large collaboration probabilities in their columns. However, regional collaborative partnerships are also evident; European countries, such as the United Kingdom, Germany, France, Italy, and Spain, tend to engage in more collaborative endeavors within their region. Further analysis, clustering countries based on continents or geographical proximity, would likely unveil additional collaborative patterns.

The heat map introduces two indices aimed at quantifying collaborative patterns among top-producing countries. First, the Egocentric Collaboration Index (ECI) is displayed on the right, measuring the extent of collaboration from the perspective of each nation with other leading countries. This index is derived by averaging the probabilities within the corresponding row. Notably, the ECI exhibits less variation, offering a glimpse into countries that frequently engage in collaborations with other top nations, such as France, Japan, Italy, and Canada, while pointing out those with lower collaboration rates, like Indonesia (more details on this will follow). Second, the Collaborative Role Index (CRI) is displayed above the columns, providing insights into the collaborative role of each country for other leading nations. Similarly, the CRI is derived by averaging the probabilities within the corresponding column. The index above reveals that the U.S., for instance, is a prominent collaborative partner for other leading countries, contrasting with Indonesia, which engages less frequently in collaborations with other top nations. Instead, Indonesia fosters partnerships with regional counterparts like Malaysia, Japan, Australia, Taiwan, China, Thailand, India, Korea, and others (not shown). The CRI is an average probability that that country is on multi-country publications from other leading countries; i.e., when other leading countries publish a multi-country scientific article, the U.S. is on that paper 27% (0.27) of the time for the time frame given. Despite being averages of probabilities within the heat map, care must be taken when interpreting the subtle meanings of these indices. For instance, while the CRI can be interpreted as an average probability, the ECI is distorted by articles that contain more than one other leading country; i.e., multiplying the average probabilities by the number of top countries (11, excluding itself) can exceed 100%. If the probabilities from all other countries were used for the ECI, it would be proportional to the average number of distinct countries on its multi-country articles; herein, the ECI includes a subset of only the leading countries shown in the heat map.

Returning to Indonesia, despite ranking third in total publications over this time span, it surprisingly has the fewest collaborations with other countries among the top 12 producing nations. One potential factor contributing to this trend is the substantial number of scientific articles written in Indonesian. Notably, from 2019 to 2024, Indonesia has produced the highest count of non-English scientific articles (658,647 articles, predominantly in Indonesian). Brazil follows as the second-highest contributor to non-English publications (549,710 articles, mainly in Portuguese), with a significant drop in non-English articles for other nations (France ranks third with 212,654 articles in French). Both Indonesia and Brazil exhibit low CRI index values (columns), other countries such as Russia and India are also low, even though the latter two primarily publish articles in English (9th and 31st in the world in non-

English articles, respectively). When Indonesia and Brazil engage in collaborations with other countries, the article is written in English 77% of the time. In contrast, when the article features only these countries in the affiliations (i.e., only one distinct country), the language switches to English only 53% of the time. Figure 19 presents the heat map with a focus on articles written exclusively in English. Consequently, Indonesia shifts from the third to the sixth position in total articles produced, while Brazil is replaced in the top 12 by Australia. For the remainder of this study, the analysis concentrates on articles written in English.

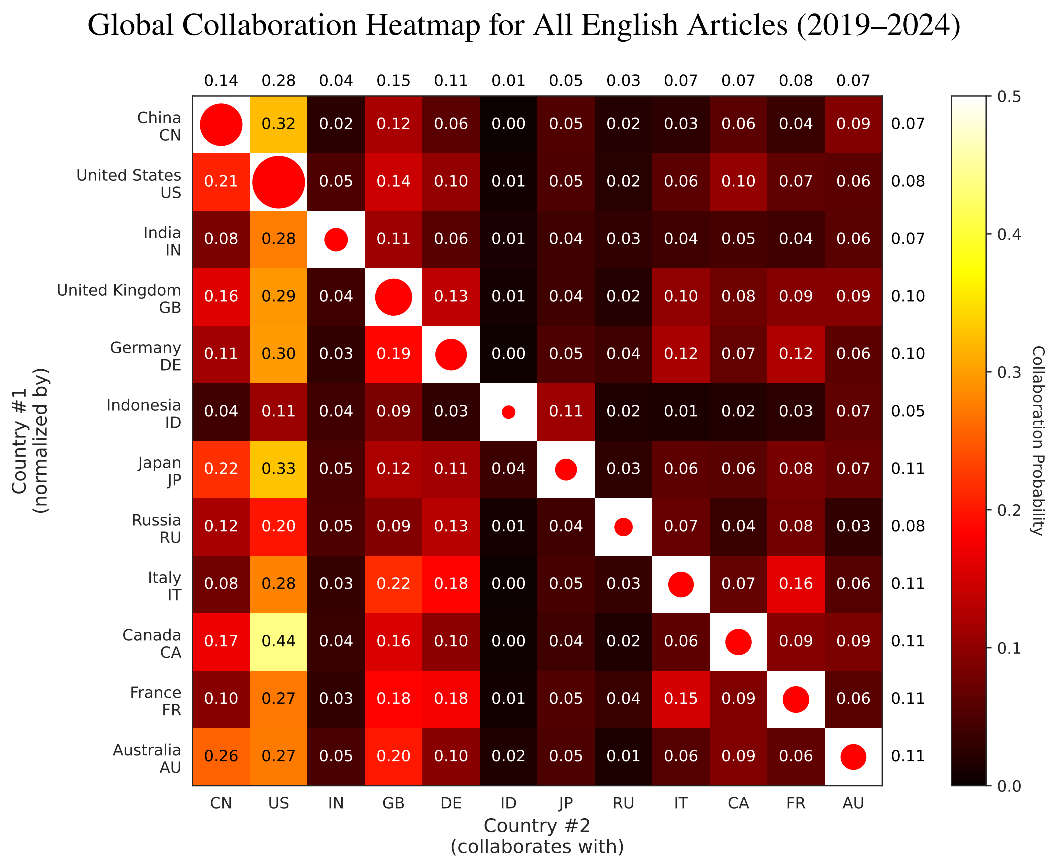
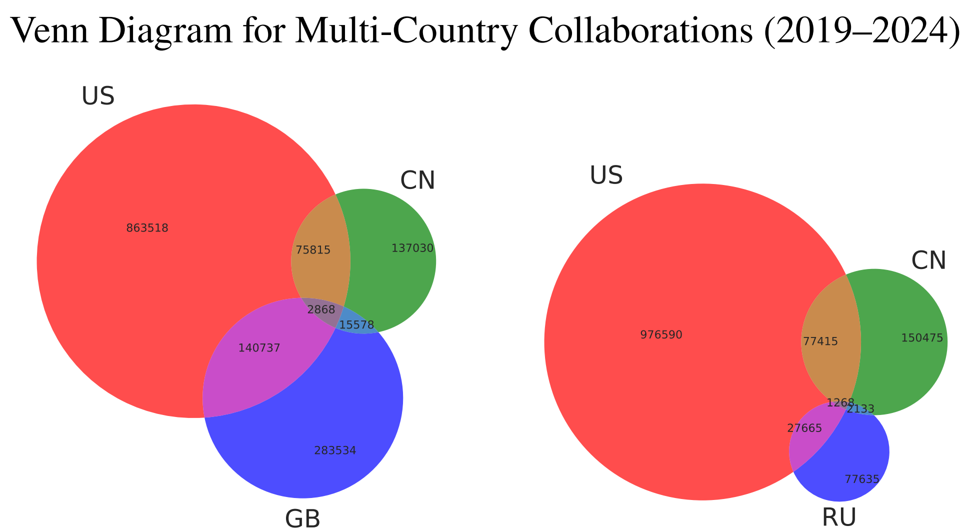


Figure 19. A heat map illustrating global collaboration among the top 12 most prolific countries from 2019 to 2024. The formatting is similar to Figure 18 with the only difference being that only articles written in English are used to calculate the top 12 countries and the global collaborations.

To delve deeper into the ECI and the CRI, Figure 20 illustrates Venn diagrams representing the collaborative relationships of the U.S. with China and the United Kingdom, and with China and Russia. Each colored region in the Venn diagram signifies a specific subset of multi-country publications, with the number indicating the total count for that subset. For instance, the large red section in the U.S. circle denotes the number of multi-country publications from 2019 to 2024 that lack China or the United Kingdom as an institutional affiliation of a co-author (in Figure 20a). Similarly, shaded regions indicating overlap between two countries exclude the third country, while the intersection of all three countries represents multi-country publications where affiliations from each country exist for the co-authors.



(a) U.S.–CN–GB (2019–2024) (b) U.S.–CN–RU (2019–2024)

Figure 20. A Venn diagram showing the numbers of multi-country publications from 2019 to 2024 for (a) U.S.–CN–GB and (b) U.S.–CN–RU. The numbers represent the total number of multi-country articles in each colored region, e.g., the U.S. has 863,518 multi-country publications without China (CN) or the United Kingdom as a co-author, the U.S. has 140,737 multi-country publications with the United Kingdom and without China, and the U.S. has 1,368 multi-country publications with both China and Russia (RU).

Regarding the ECI for the U.S., it entails the sum of the overlap subsets between the U.S. and every other country, divided by the total number of multi-country publications produced by the U.S. Notably, in the example Venn diagram, the region where all three countries are listed on a publication is counted twice for the ECI exhibited for these three countries. However, when all leading countries are included, the complexity of numerous overlapping regions would render a 2D Venn diagram impractical (but the overlap subsets between more than two countries exists and is counted multiple times for the ECI).

On the other hand, the CRI for the U.S. is computed from the overlap subsets between the U.S. and every other leading country, with each subset divided by the total number of multi-country publications for the other countries. These two indices can differ significantly due to variations in the total number of publications for each country. For instance, in Figure 20b, collaborative publications between the U.S. and Russia are approximately an order of magnitude lower than those with the U.S. and China. From the U.S.'s (and China's) egocentric perspective, collaboration with Russia appears minimal, while from Russia's egocentric viewpoint, collaboration with the U.S. (and China) represents a more substantial fraction of their total multi-country publications. Thus, the average of this probability from the perspective of all collaborating countries, as represented by the CRI, offers insights into how often these other countries collaborate with the country at hand. With a comprehensive grasp of the ECI and CRI, depicted in the accompanying Venn diagrams, the focus now returns to the examination of various heat maps. These visualizations continue to unravel and analyze global collaboration patterns, offering a detailed view on collaborative relationships among leading nations.

3.5 Global Collaboration between Leading Countries over Time

Figures 21–23 present collaboration heat maps spanning the last three decades: the 1990s, the 2000s, and the 2010s, respectively. Throughout this period, collaboration among the top countries has shown a modest increase. Notably, the U.S. and the United Kingdom consistently maintained leading positions in scientific article production. A pivotal shift occurred with China's ascent, transitioning from ninth globally in the 1990s to the second position in the 2010s. Beyond the sheer increase in article production, China's engagement in articles involving multiple nations saw a substantial rise, evident from the seven-fold increase in the CRI (above the heat map, shifting from 0.02 in the 1990s to 0.14 in 2019–2024). To contrast, the United Kingdom's CRI has slightly increased (0.13 in 1990s to 0.15 in 2019–2024) and the U.S.'s CRI has slightly decreased through the decades (0.34 in 1990s to 0.28 in 2019–2024).

In the 1990s, several countries in the bottom half of the top 12 had relatively fewer collaborations within this group. However, scientific collaboration has witnessed a general uptrend over the decades, with the exception of India, which experienced a stagnation in collaborations despite an increase in scientific articles. The dynamics of the Top 12 have also seen countries moving in and out, influenced by closely matched scientific outputs. One should keep in mind that the chosen years are somewhat arbitrary, and countries at the border of the Top 12 might have shifted under different year ranges.

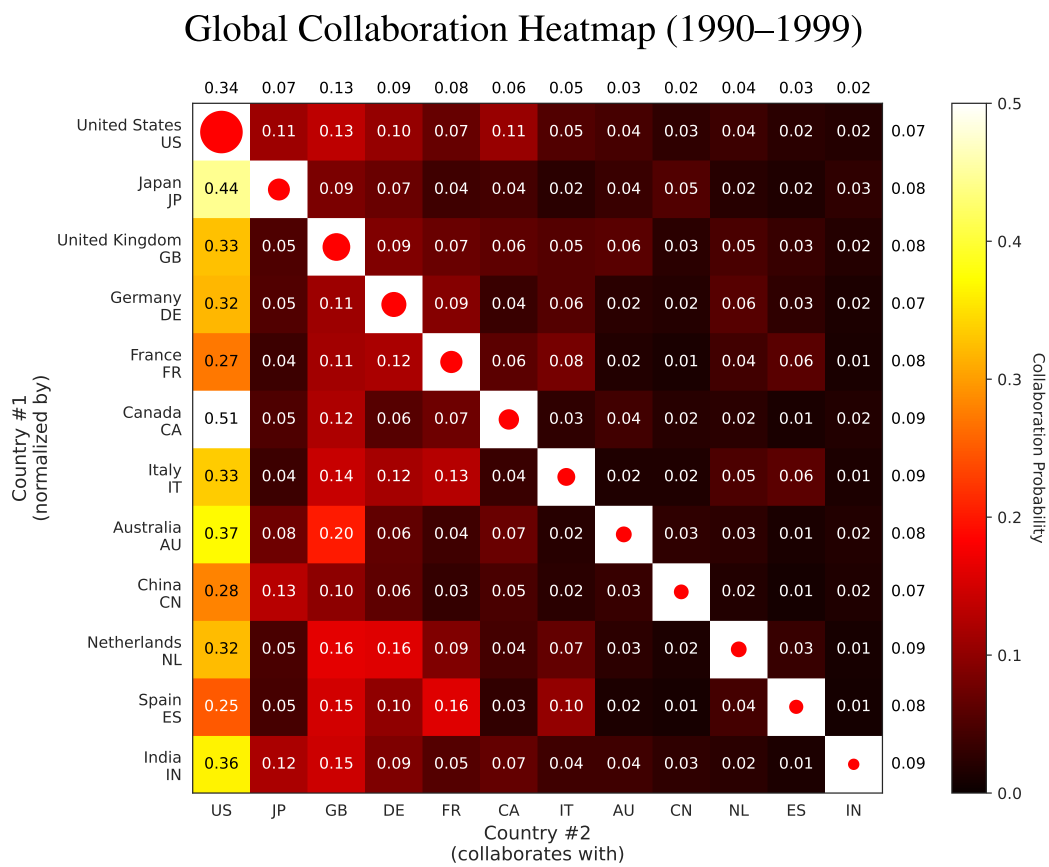


Figure 21. A heat map illustrating global collaboration among the top 12 most prolific countries in the 1990s. The formatting is similar to Figures 18 and 19.

Across the decades, certain countries consistently rank among the world's top contributors to scientific output: the U.S., China, the United Kingdom, Germany, Japan, France, Canada, Italy, Australia, and India. The recent emergence of Indonesia, as discussed earlier, adds an intriguing dimension, suggesting a rapid increase in the scientific presence of Southeast Asian countries.

This heat map serves as a dynamic lens into the evolving collaboration between country pairs. An intriguing example lies in the asymmetry within the relationship between the U.S. and China. In the 1990s, collaborative articles involving the U.S. and China accounted for 3% and 28%, respectively. This asymmetry persisted into the 2000s, with percentages at 7% (U.S. articles with China) and 33% (China articles with the U.S.). The trend continued into the 2010s, with percentages growing to 18% (U.S. articles with China) and 41% (China articles with the U.S.). Notably, in Figure 19 (2019 to 2024), these percentages further changed to 21% and 32%, respectively.

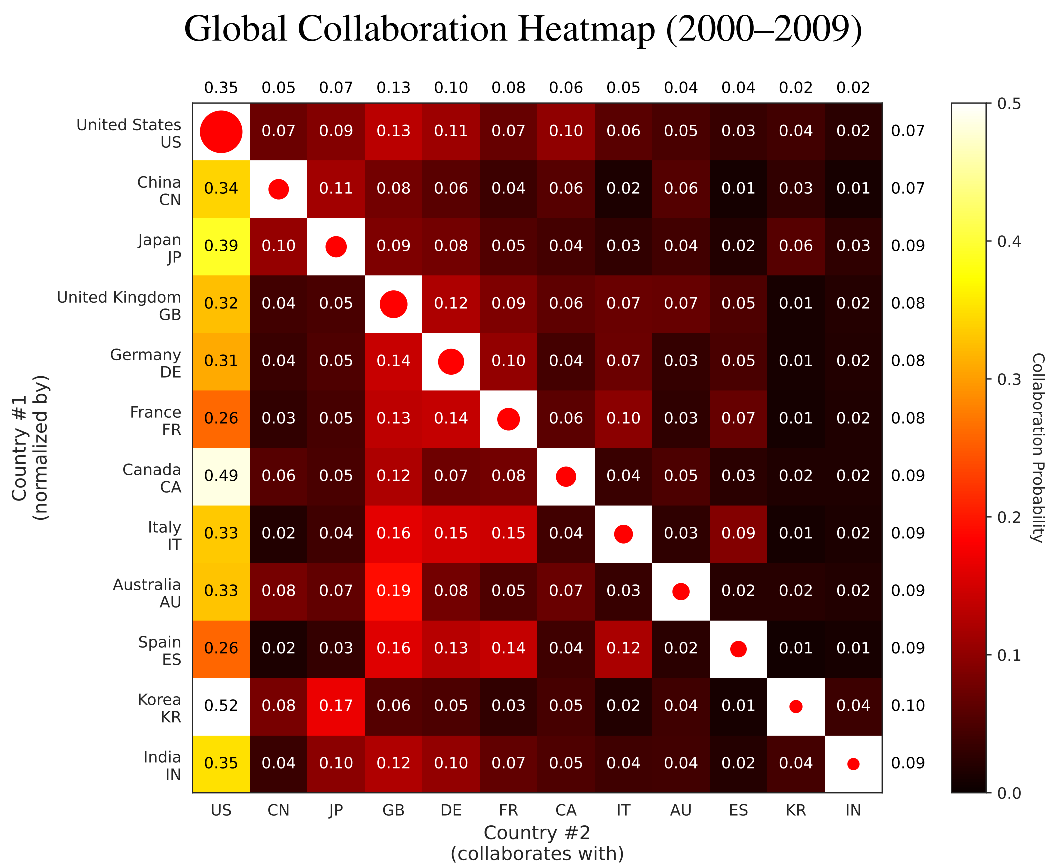


Figure 22. A heat map illustrating global collaboration among the top 12 most prolific countries in the 2000s. The formatting is similar to Figures 18 and 19.

While the focus here was on the U.S.–China relationship and its asymmetry, a broader pattern emerges across various examples: the greater the discrepancy in scientific productivity between two countries, the more pronounced the asymmetry in the percentages of collaborative articles. It is essential to acknowledge that higher percentages could also be influenced by the collaborative output of countries with higher overall article production. The Matthew effect, also known as "the rich get richer," seems to apply to the total scientific productivity of countries in its global collaboration with other countries. This does not necessarily mean that high-producing countries have more collaborative scientists (by percentage of total), just more of them. This greater number of scientists leads to a higher probability that scientists from other countries will interact with them at conferences or the like, leading to more multi-country collaborations and publications. The global shift to virtual conferences, meetings and seminars following COVID-19 enables even more opportunities for scientists to engage with scientists around the world without the requirement of physical presence.

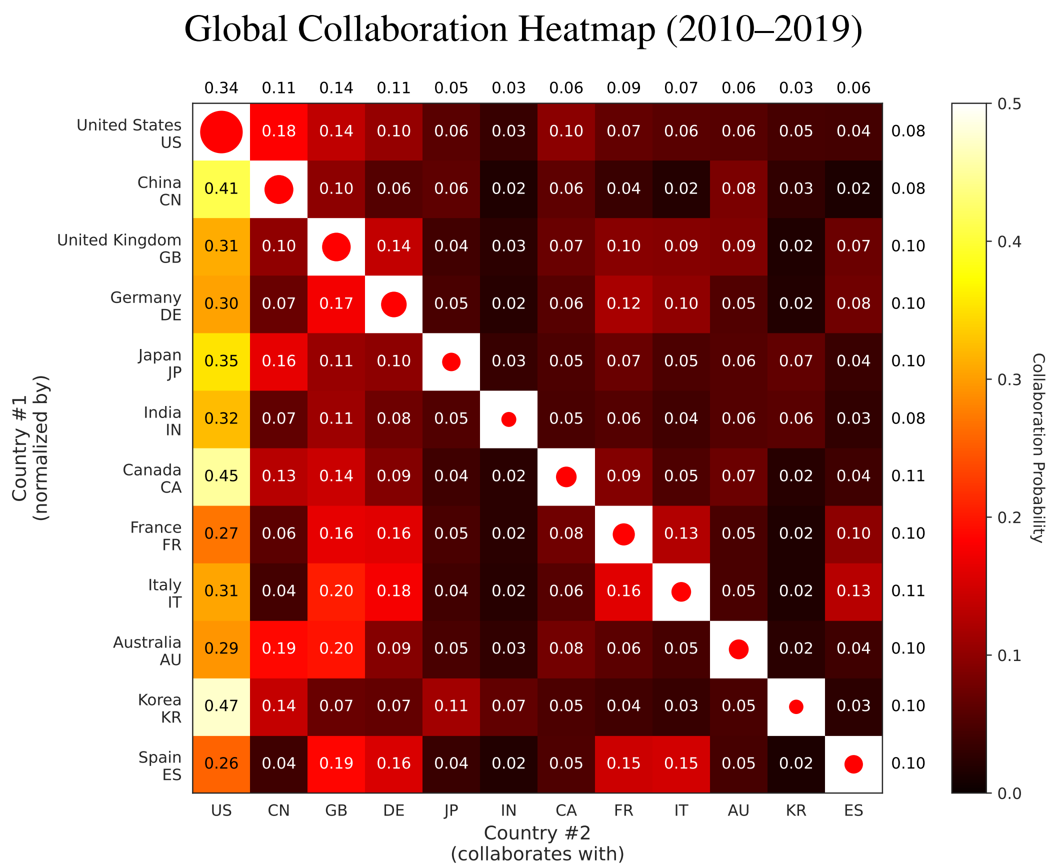


Figure 23. A heat map illustrating global collaboration among the top 12 most prolific countries in the 2010s. The formatting is similar to Figures 18 and 19.

3.6 U.S. Institution Collaboration with Leading Countries

Turning the focus to a more granular scale, this exploration delves into global collaboration at the institutional level, as illustrated in Figure 24. The y axis of this heat map features a myriad of institutions, predominantly federal laboratories, federally funded R&D centers, and university affiliated research centers within the U.S., maintaining similarity with previous heat maps. Each element within the heat map represents a probability, with the denominator being the total number of publications by institutions whose authors list more than one distinct country in their affiliations. The numerator signifies the count of publications listing the country corresponding to the bottom of the heat map within at least one author's affiliation.

For instance, 40% of the National Energy Technology Laboratory's (NETL) multi-country papers involve China, while 30% of Johns Hopkins University Applied

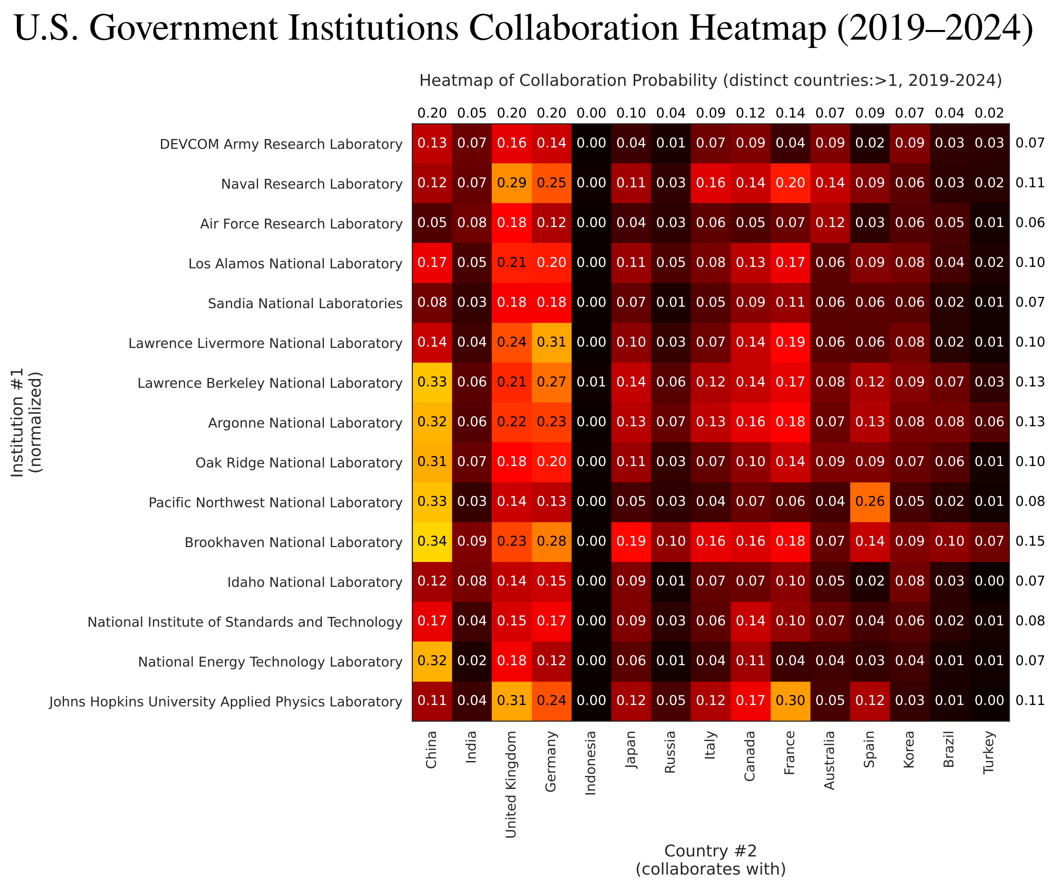


Figure 24. A heat map illustrating U.S. institutional collaboration with the top 15 most prolific countries between 2019 and 2024. The formatting is akin to Figures 18 and 19, with the y axis now featuring various government agencies, federally-funded research and development centers, and university affiliated research centers. Probabilities in the heat map signify the percentage of papers with more than one distinct country listing an author affiliated with an institution from the respective country. The color bar (removed herein) is similar to past heat maps, ranging from 0 to 0.50.

Physics Laboratory's such papers are in collaboration with the United Kingdom. The CRI index (above) the heat map maintains a similar interpretation as earlier, gauging the extent of collaborative papers with the country at the bottom for all institutions listed. It reveals that a substantial portion of collaboration involves countries such as China, the United Kingdom, Germany, Japan, Italy, France, Canada, Australia, Spain, and Korea. The ECI (right) has a similar interpretation as well, except now it is for the institutions at the left rather than for countries.

In terms of scale, this heat map, expressed in probability values rather than frequencies, prompts a question about the total volume of collaborative articles involving more than one country and these U.S. institutions over the given timeframe. The answer: 54,586 articles. Among these, NETL has the smallest number of collaborative articles (360). Notably, out of the 54,586 articles, only 3,592 are published without an academic institution as a collaborator. This highlights that most publications within these institutions, involving collaborations with other countries, also feature an academic collaborator, either in the U.S. or abroad.

This prompts an inquiry into how these collaborations compare to those involving academic institutions. In Figure 25, a heat map depicting global collaboration is presented with the y axis featuring the top-producing academic institutions from 2019 to 2024. The range of articles published, listing more than one distinct country during this period, spans from 17,924 articles (University of Wisconsin–Madison) to 78,257 articles (Harvard University), aggregating to 393,230 articles across all these institutions.

Comparing this to Figure 24, the collaboration targets for academia closely align with those in the national laboratory ecosystem. These academic institutions, operating across diverse fields, consistently engage in global collaborations, notably with China, the United Kingdom, Germany, Japan, Italy, France, Canada, Australia, Spain, Korea, and Brazil. It is essential to acknowledge that these institutions represent the top academic producers.

However, considering the broader academic landscape in the U.S., Figure 26 unveils a heat map of top institutions as defined by the ECI values on the right of the plot. Here, these academic institutions demonstrate high probabilities of collaboration with multiple countries, in particular China, the United Kingdom, Germany, and Canada. It is crucial to clarify that the institutions listed in Figure 26 are not ranked based on the frequency of publications with more than one distinct country. For instance, the University of Michigan (34,164 articles) and University of Arizona (16,347 articles) publish over an order of magnitude more articles from 2019 to 2024 than Northern Illinois University (1,592 articles) or Brandeis University (1,667 articles).

An intriguing aspect to explore within this dataset is the role of funding agencies that underpin the research leading to global collaborations. Figure 27 presents a heat map showcasing funding agencies acknowledged in publications with affiliations across different countries. This encompasses notable entities such as the National Science Foundation (NSF), the National Institutes of Health, the Department of

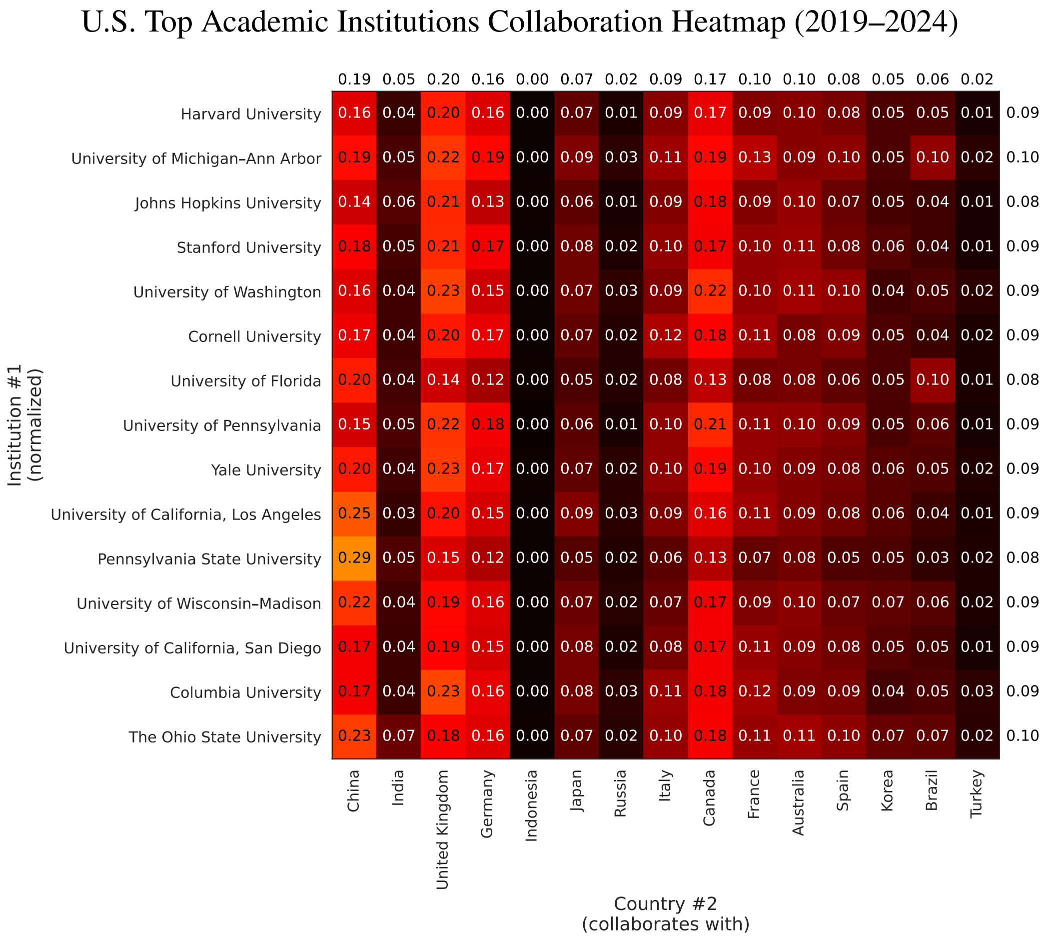


Figure 25. A heat map illustrating U.S. academic institutional collaboration with the top 15 most prolific countries between 2019 and 2024. The formatting is similar to Figure 24, except now the y axis features the top producing academic institutions over this time period (ordered top to bottom). The color bar (removed herein) is similar to past heat maps, ranging from 0 to 0.50.

Energy (DOE), various defense agencies (Army Research Office [ARO], Office of Naval Research [ONR], Air Force Office of Scientific Research [AFOSR]), and various government laboratories supporting both basic and applied research and development.

Distinguishing itself from previous heat maps, this visualization considers the process where principal investigators (PIs) submit proposals to federal agencies, which are then reviewed and funded (or declined). Following approval, the PI carries out the research outlined in the proposal, often engaging with researchers from other countries to enhance the project. Consequently, the funded work contributes to publications, with authors frequently acknowledging the supporting funding agencies upon publication.

Remarkably, funding agencies primarily devoted to basic research, such as NSF, DOE and DOE Basic Energy Science, AFOSR, ONR, and ARO, exhibit substantial percentages of globally collaborative publications, particularly with countries like China, the United Kingdom, Germany, France, or Canada. The first column of the heat map illustrates collaborations with China. Notably, China is less frequently listed as a collaborator for federal agencies like National Aeronautics and Space Administration (NASA), Defense Advanced Research Projects Agency, National Institute of Standards and Technology, the National Nuclear Security Administration, and agreements originating directly from federal defense laboratories.

U.S. Top Globally Collaborative Academic Institutions Heatmap (2019–2024)

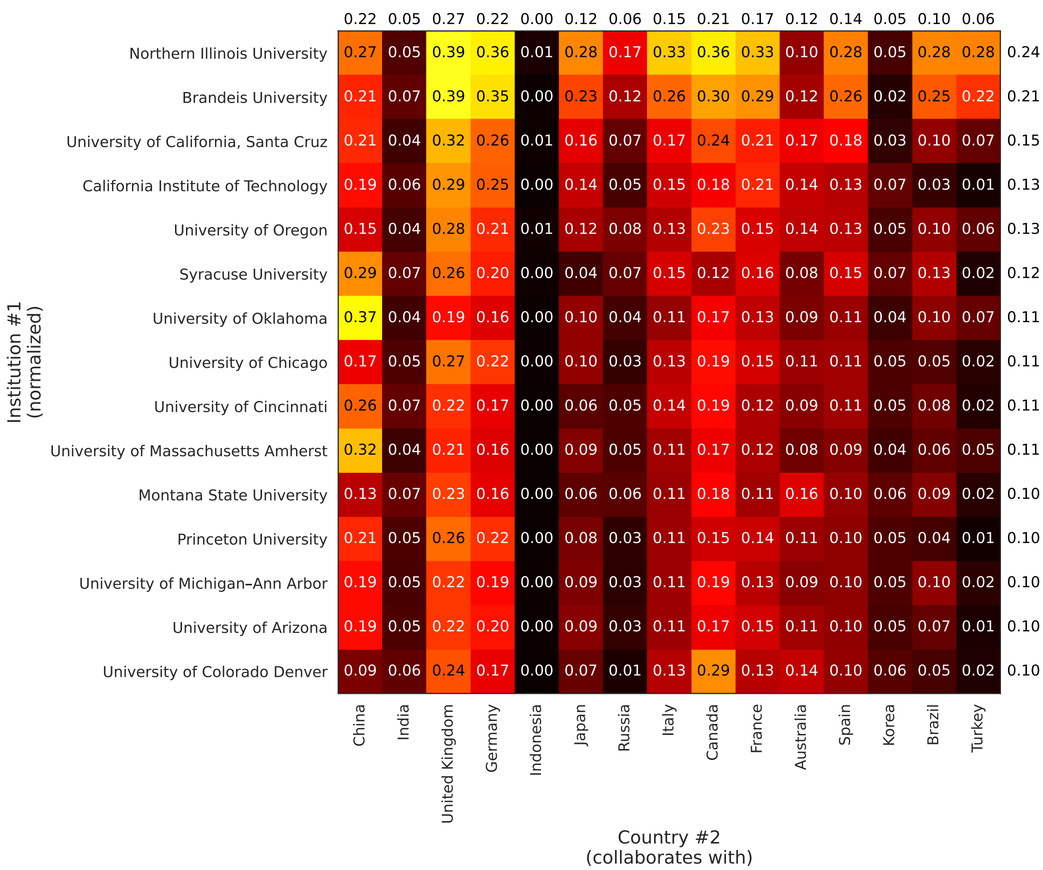


Figure 26. A heat map illustrating U.S. top collaborative academic institution collaboration with the top 15 most prolific countries between 2019 and 2024. The formatting is similar to Figures 24 and 25, except now the y axis features the academic institutions with the highest collaboration index (right-hand side) over this time period (ordered top to bottom).

NASA stands out with the highest collaboration rates between its PIs and researchers from other countries (highest ECI value), especially with significant percentages involving the United Kingdom, Germany, and France. This heightened collaboration may be attributed, in part, to the nature of research conducted, particularly in fields like astronomy and astrophysics where hyper-authorship has become more prevalent, resulting in author lists comprising hundreds or even thousands of contributors from various countries.

Exploring global collaboration at the institutional level reveals even more nuance.

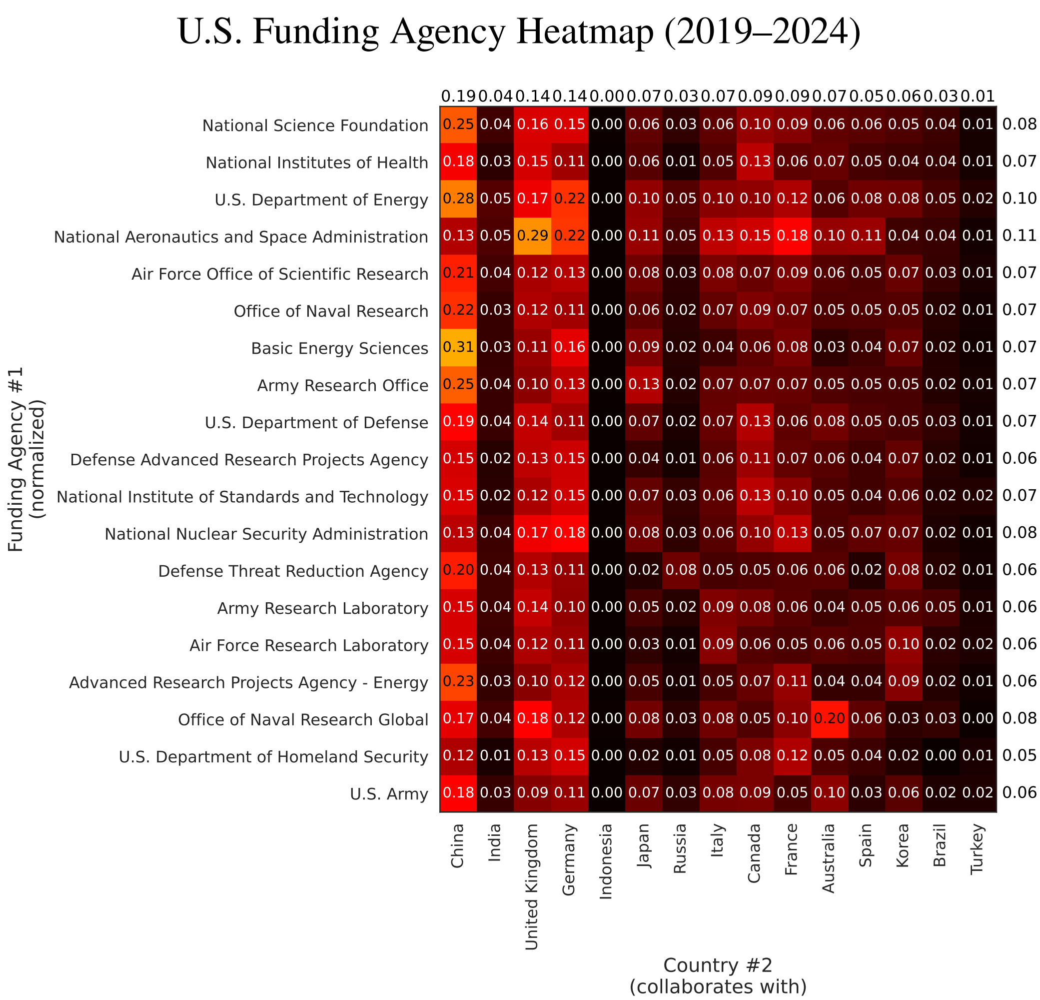


Figure 27. A heat map illustrating U.S. agency funding of collaborative articles with the top 15 most prolific countries between 2019 and 2024. The formatting is similar to Figure 18, except now the y axis features multiple U.S. Government agencies who are listed within the acknowledgments section as partially funding the research. The probabilities within the heat map represent the percentage of papers (funded by the agency) with more than one distinct country that list an author with an institution from the country on the x axis.

For instance, national laboratories frequently collaborate with academic institutions, and both entities demonstrate substantial partnerships with other countries. Consequently, collaborations between institutions within the U.S. emerge as a confounding factor when attempting to pinpoint the primary driving force behind global collaborations. This complexity echoes the challenges of credit allocation for individual authors within multi-author publications, extending the difficulty to attributing the predominant factor for collaborations with other countries within multi-institution publications.

The intricacies persist when considering funding agencies, where attributing contribution of authors to global collaborations remains challenging. Many publications involve numerous authors from various institutions, each acknowledging funding from different agencies. Adding to this complexity, different funding agencies allocate resources across a portfolio of research areas, some of which naturally involve more global collaboration. This intricate web of interactions raises a follow-up question, which will be explored in the next section: How do research areas contribute to multi-country collaborative publications?

3.7 Global Collaboration for Research Concepts

In this section, the impact of research areas, or research concepts, on global collaboration is explored. The research area can affect various factors such as the total number of published papers, citation rates, authorship distribution, and even the collaborative dynamics within a specific research field. Before exploring this effect, first look into the process of mapping papers to research concepts within OpenAlex.

Concepts, in this context, represent abstract ideas tied to publications. These were first called "fields of study" in Microsoft Academic Graph to represent the semantic contents of the document.[3] They are organized in a hierarchical structure across five levels. Level 0 encompasses broad concepts like materials science, engineering, computer science, and sociology (totaling 19 concepts), while level 5 delves into much more specific areas such as pandemic, coronavirus, black hole (networking), chirality (physics), and rectifier (neural networks). Employing an automated classifier, OpenAlex assigns concepts to works—over 65,000 concepts in total—based on titles, abstracts, and the interconnections among concepts.

This ML algorithm predicts a similarity score between the article's text and each concept. After obtaining scores for all concepts, a threshold is applied to finalize concept assignments, as depicted in Figure 17. Remarkably, over 99.9% of articles have at least one concept assigned. Each assigned concept is presented as a list accompanied by details such as a concept score, name, identifier, ancestors, related concepts, and a link to the corresponding Wikidata item.

This section primarily centers on level 0 concepts, yet its applicability extends seamlessly to a diverse array of concepts within the expansive hierarchical network. Before probing the details, it is essential to underscore the hierarchical structure of the concept graph. To illustrate this, examine a selection of level 0 concepts and their interconnections with concepts at the immediately lower level, level 1.

Figures 28 and 29 present Sankey plots that visually capture the relationships between a level 0 concept and its level 1 counterparts. For this illustration, two exemplars were chosen from the top level 0 concepts: material science and computer science. A striking observation is the substantial number of level 1 concepts associated with each level 0 concept. This pattern holds true for all 19 level 0 concepts, with a slightly reduced count for some of the social science and humanities concepts at level 1.

The width of the lines in these Sankey plots reflects the predicted score derived from the title and abstract, signifying the similarity between different concepts. It is noteworthy that a high score for a lower-level concept can influence the scores and inclusion of the concepts above it. Consequently, these upper-level concepts are often listed, even if their individual scores fall below the inclusion threshold. Interestingly, the Sankey plot can also be plotted with line widths relative to the number of citations associated with works from the different level 0 and level 1 concepts. Figure 30 shows this for computer science. As can be seen from this plot, AI as a level 1 concept under computer science has the highest total number of citations within this database, followed by programming language and statistics.

Now equipped with insights into the realm of research concepts within the database and the mapping of articles to these concepts, the interplay between these concepts and global collaboration can be dived into. In Figure 31, a heat map unveils the landscape of collaboration across articles associated with level 0 concepts. The concepts are arranged alphabetically, while the countries along the bottom are strategically ordered based on the production of multi-country articles spanning from 2019 to 2024.

The values within the heat map maintain a familiar format, aligning with prior encounters with such visualizations. To illustrate, consider the level 0 concept “en-

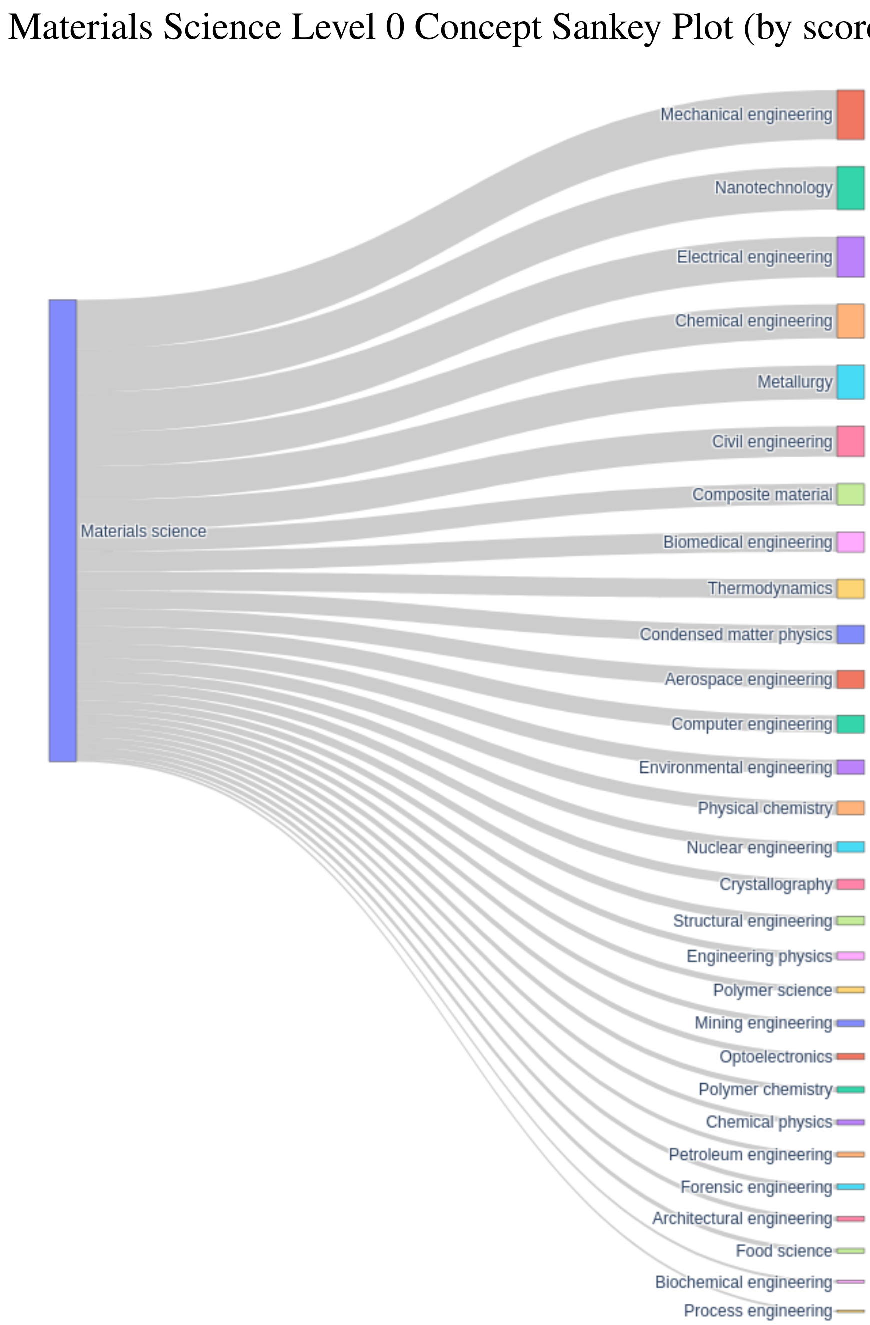


Figure 28. A Sankey plot illustrating the connections from the level 0 concept “material science” (on the left) to its diverse level 1 counterpart concepts (on the right). The thickness of the connecting lines in this network is determined by the predicted score within OpenAlex, serving as a metric that indicates the similarity between concepts.

gineering." Within this domain, 27% (or 0.27) of all articles with multiple countries affiliated are attributed to the U.S., while China claims a comparable share of 29% (or 0.29). Shifting focus to the domain of medicine, the U.S. emerges as a dominant

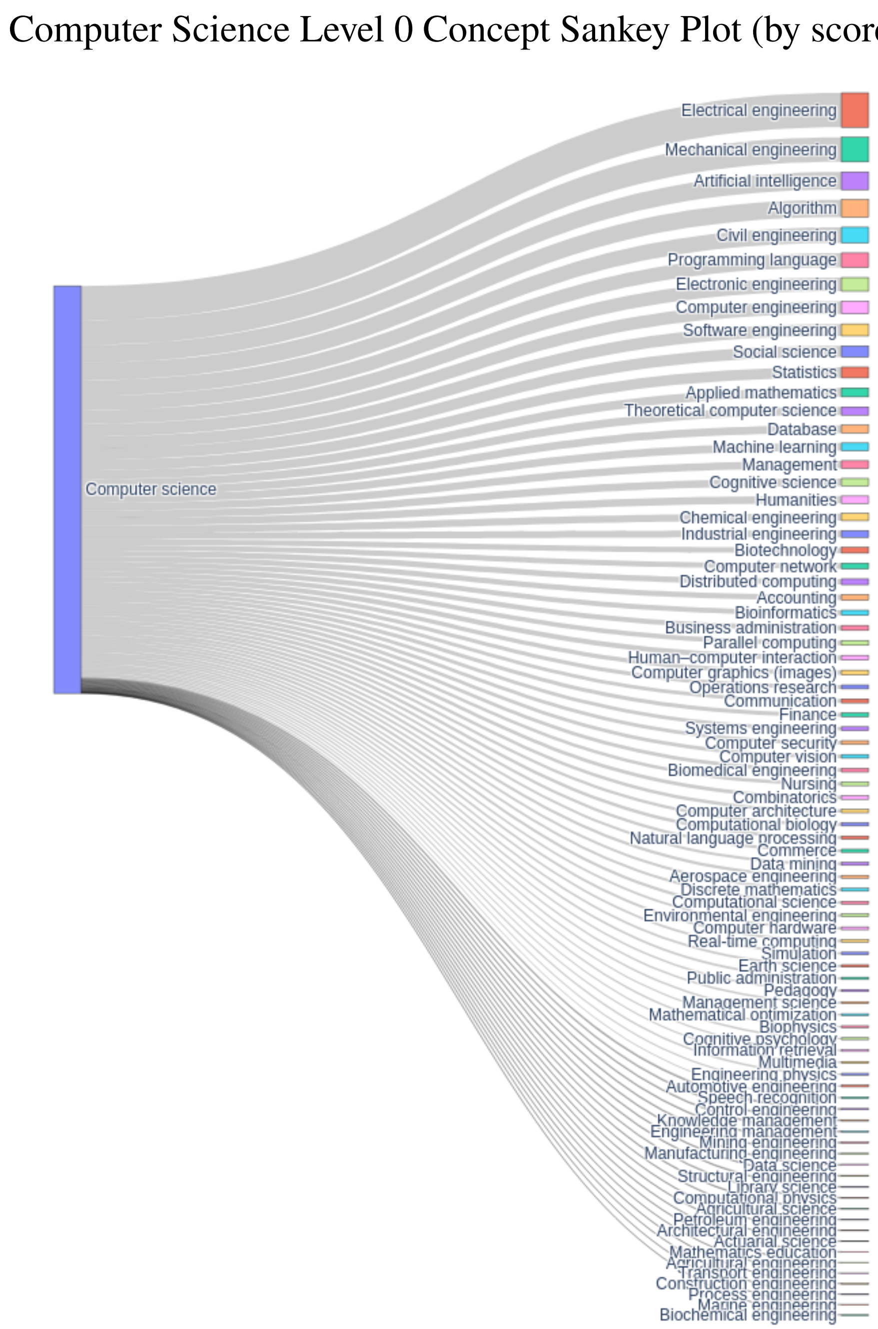


Figure 29. A Sankey plot depicting the connections originating from the level 0 concept "computer science" (positioned on the left) to its various level 1 counterpart concepts (displayed on the right). The line widths in this network are determined by the predicted scores within OpenAlex, reflecting the similarity between the associated concepts.

force, contributing to 43% of all globally collaborative articles, surpassing other nations significantly (with the United Kingdom second at 19%).

The CRI value above the heat map is similar to past measures, except it is now

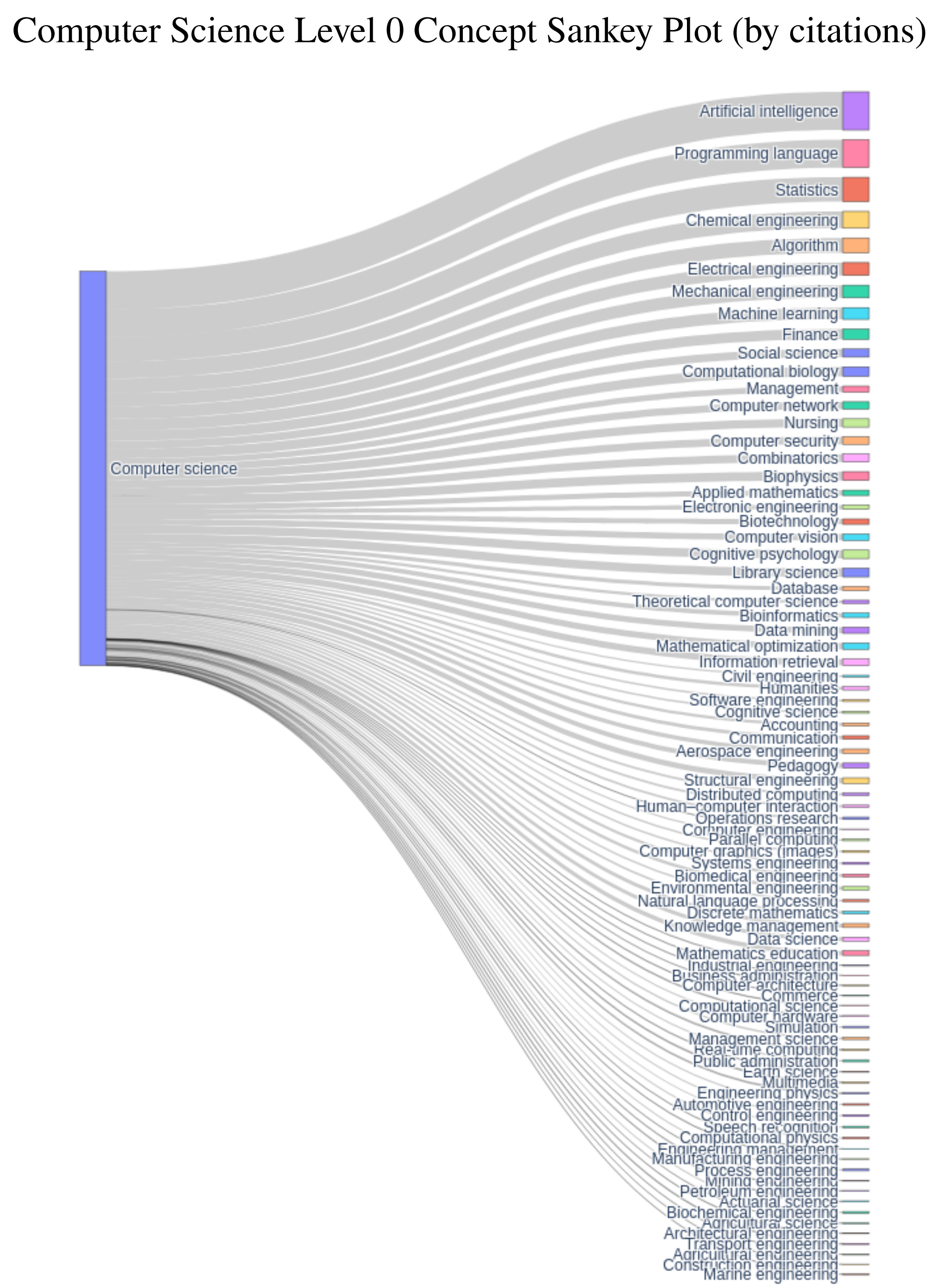


Figure 30. A Sankey plot depicting the connections originating from the level 0 concept "computer science" (positioned on the left) to its various level 1 counterpart concepts (displayed on the right). The line widths in this network are determined by the citations, reflecting the relative impact of the associated concepts.

quantifying the propensity for global collaboration across all concept areas, with the U.S. boasting the highest index (0.32). Meanwhile, the ECI adorning the right side of the heat map sheds light on the frequency with which top countries engage in

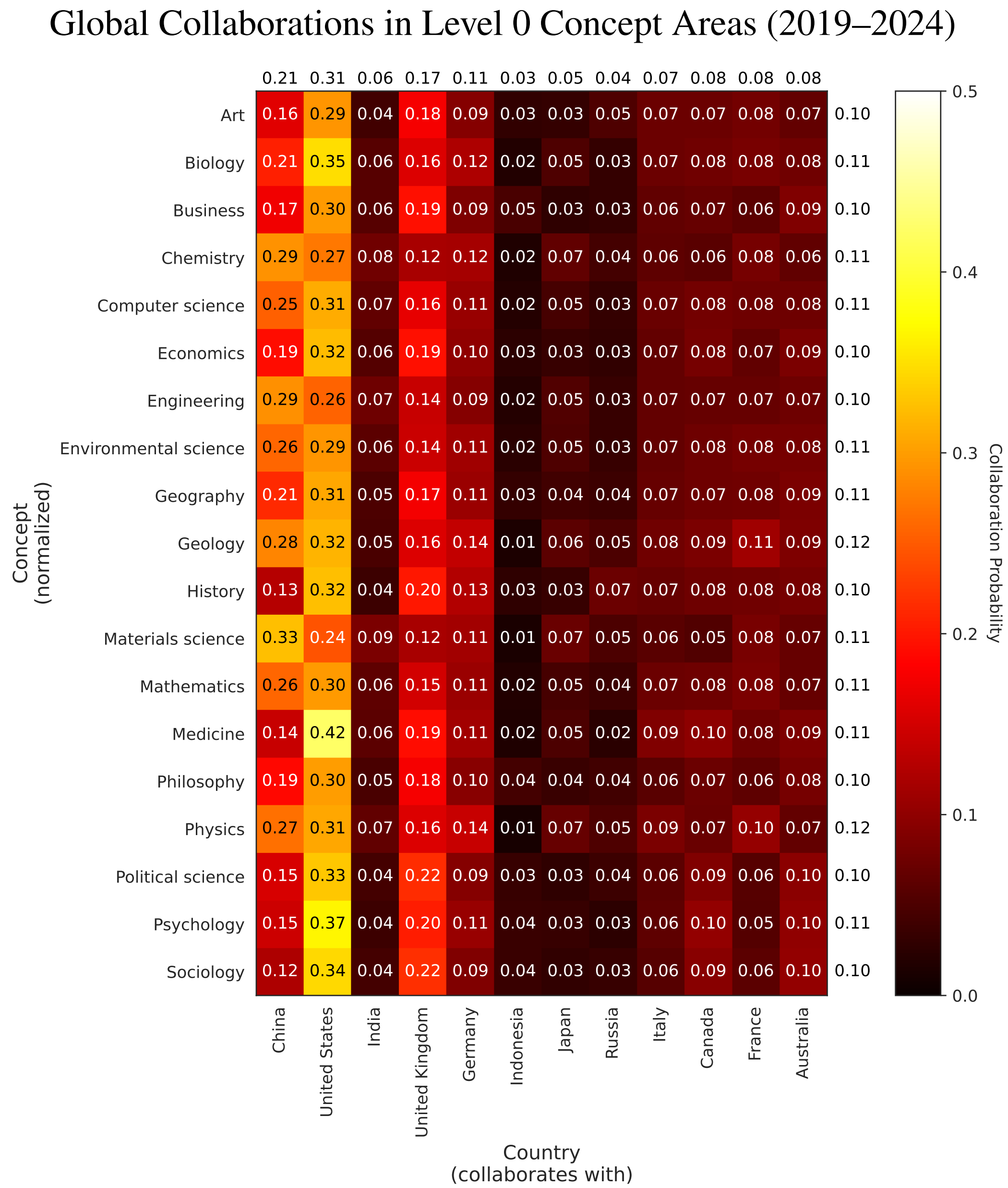


Figure 31. A heat map highlighting patterns in global collaboration from 2019 to 2024 for level 0 research concepts associated with articles within OpenAlex. The y axis showcases concept names arranged alphabetically, while the x axis reveals the leading countries in producing multi-country articles. Values within the heat map denote the probability that the country at the bottom is affiliated with a multi-country publication linked to the left-listed concept. The summed index above each column reveals the degree of each country's entwinement with publications in these concept domains, while the right-side index compiles row probabilities, delineating the top countries' association with publications tagged with these concept areas.

multi-country publications for each concept. This index exhibits minor fluctuations, indicating that the relationship between concepts and countries is similar from the self-centered concept perspective.

In Figure 32, the U.S.'s global collaborations within distinct level 0 concept areas are explored. The heat map unveils that the U.S. engages in substantial collaboration, particularly with China, followed by robust associations with the United Kingdom, Germany, and Canada. Notably, science, technology, engineering, and mathematics (STEM) fields, including material science, physics, engineering, chemistry, mathematics, and biology, exhibit the highest collaboration probabilities in U.S.–China partnerships. Intriguingly, U.S.–United Kingdom collaborations showcase elevated probabilities in non-STEM areas like history, political science, sociology, economics, and business.

Figure 33 delves into China's collaborative landscape across level 0 concept areas, revealing distinctive patterns compared to the U.S. China demonstrates a predilection for strong collaborations with fewer countries across diverse areas. Foremost among its collaborators is the U.S., with heightened probabilities in fields such as medicine and biology. While the United Kingdom and Australia also emerge as significant collaborators, their contribution remains a fraction of the collaborative output seen with the U.S.

The collaborative picture of the United Kingdom, illustrated in Figure 34, mirrors some aspects of both the U.S. and China. Similar to the U.S., the United Kingdom showcases elevated probabilities in collaborations with the U.S. and China, but the heat map also highlights robust partnerships with European counterparts, Canada, and Australia. Intriguingly, United Kingdom–U.S. collaborations lean towards medicine and biology, aligning with China's preferences, while United Kingdom–China collaborations predominantly gravitate towards material science and engineering, surpassing collaborations even with the U.S.

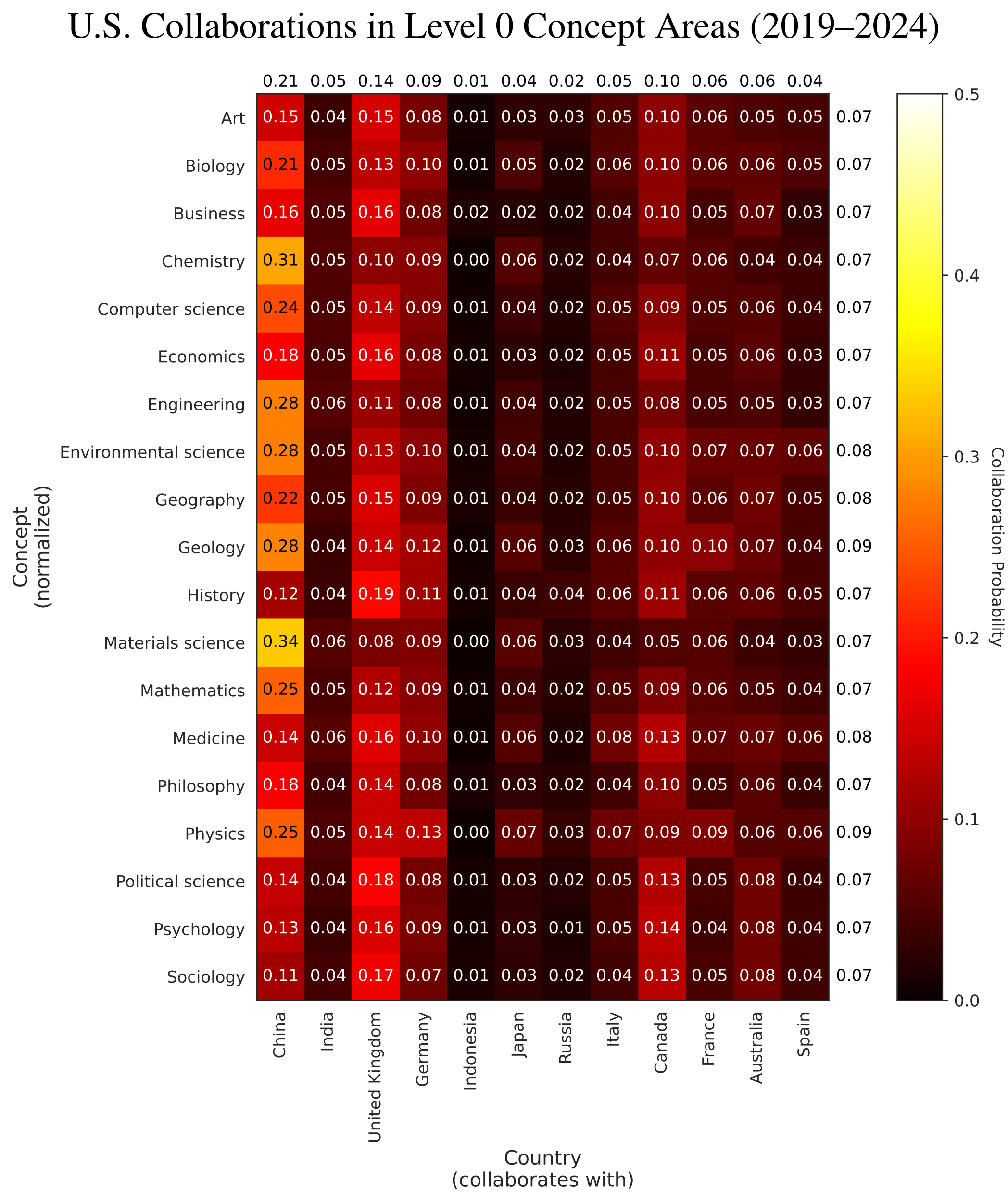


Figure 32. A heat map uncovering the dynamics of global collaboration from 2019 to 2024 for articles associated with the U.S. The y axis presents concept names alphabetically, while the x axis reveals the leading countries in producing multi-country articles. Values within the heat map represent the probability of the U.S. being affiliated with a multi-country publication related to the left-listed concept. The summed index above each column illuminates the extent of the U.S.'s involvement in publications across these concept domains.

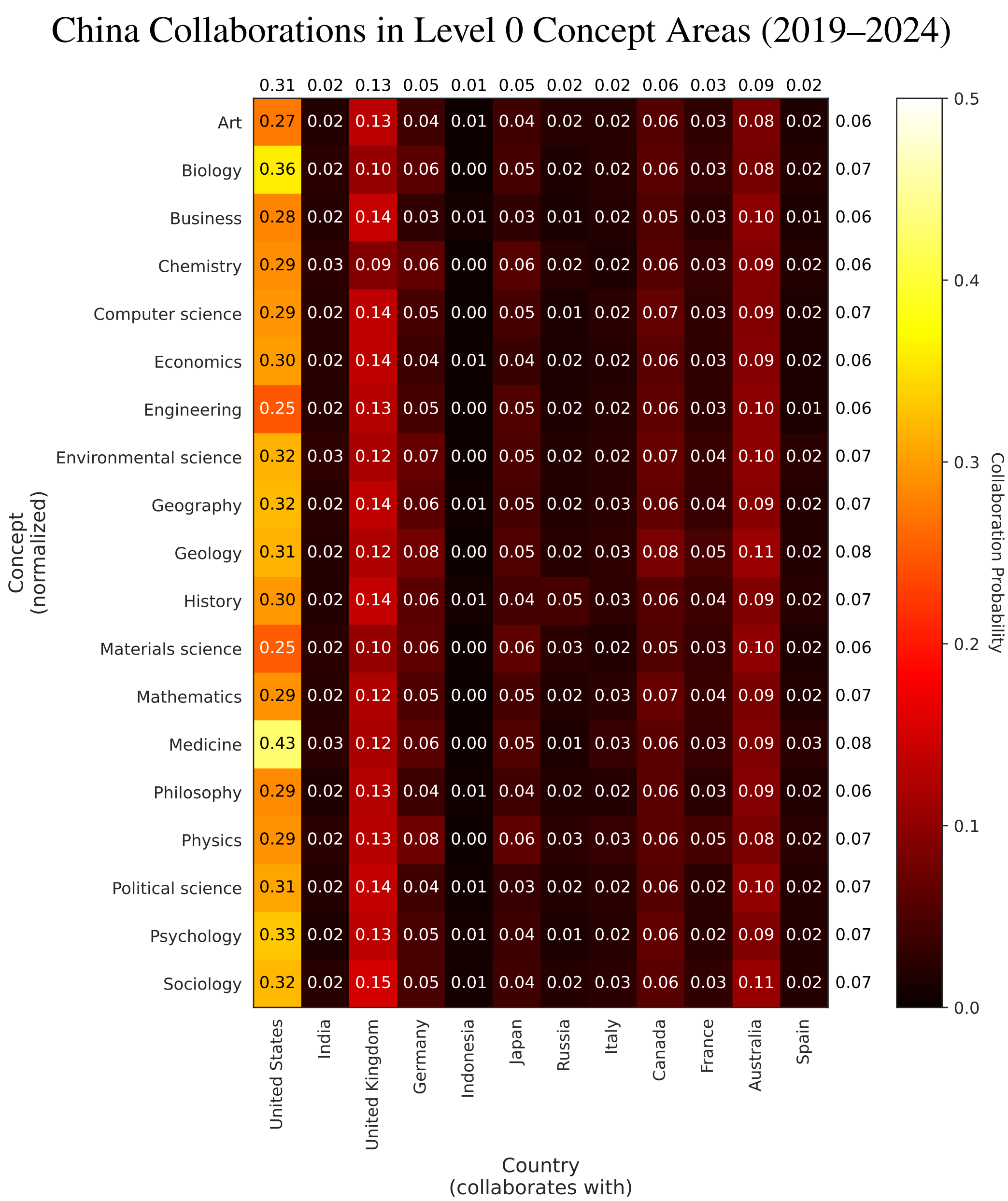


Figure 33. A heat map delving into global collaboration from 2019 to 2024 for articles associated with China. The heat map is formatted similarly to Figure 32.

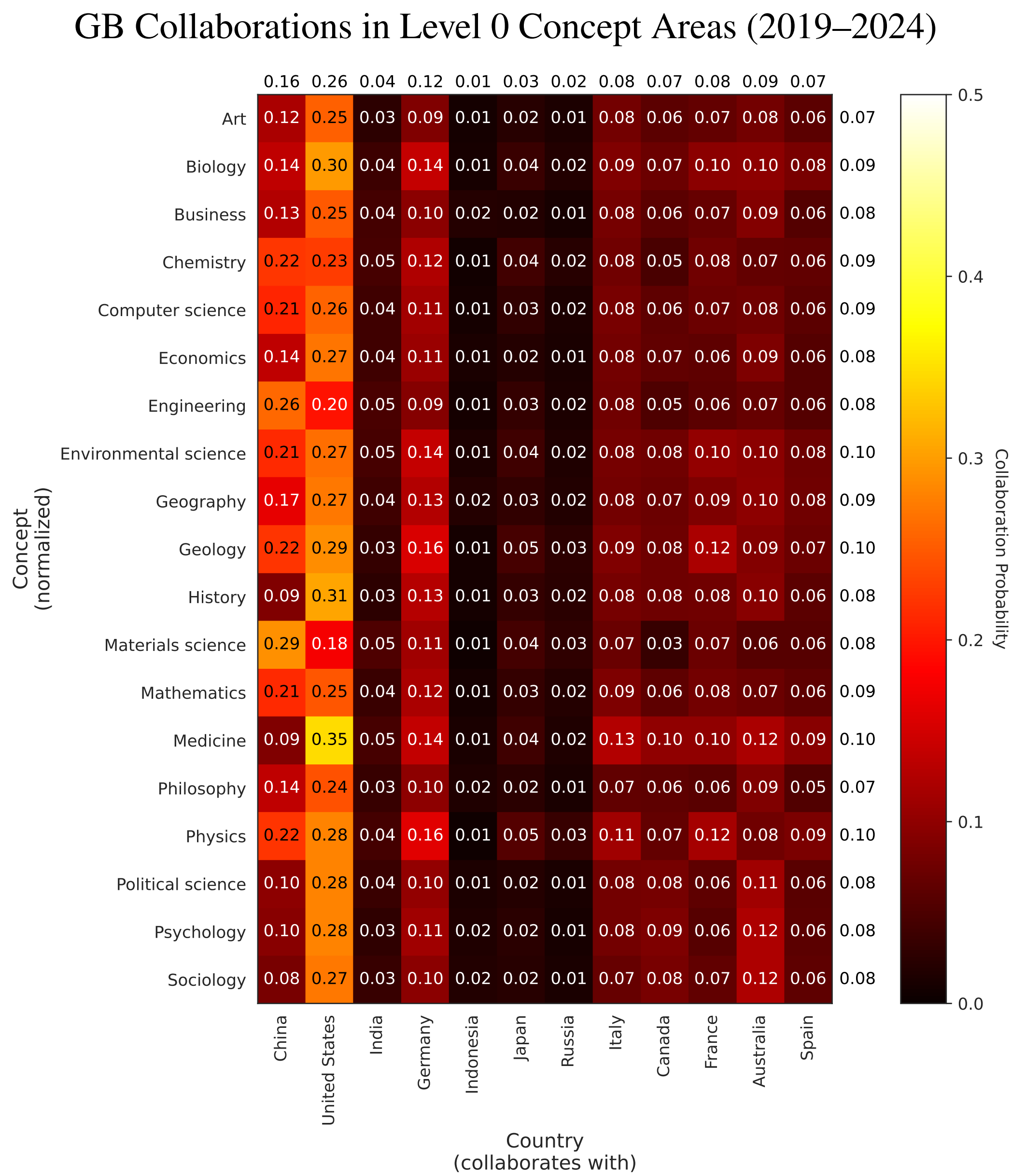


Figure 34. A heat map elucidating the patterns of global collaboration from 2019 to 2024 for articles associated with the United Kingdom. The heat map is formatted similarly to Figure 32.

3.8 Global Collaboration between Leading Countries within Specific Research Areas

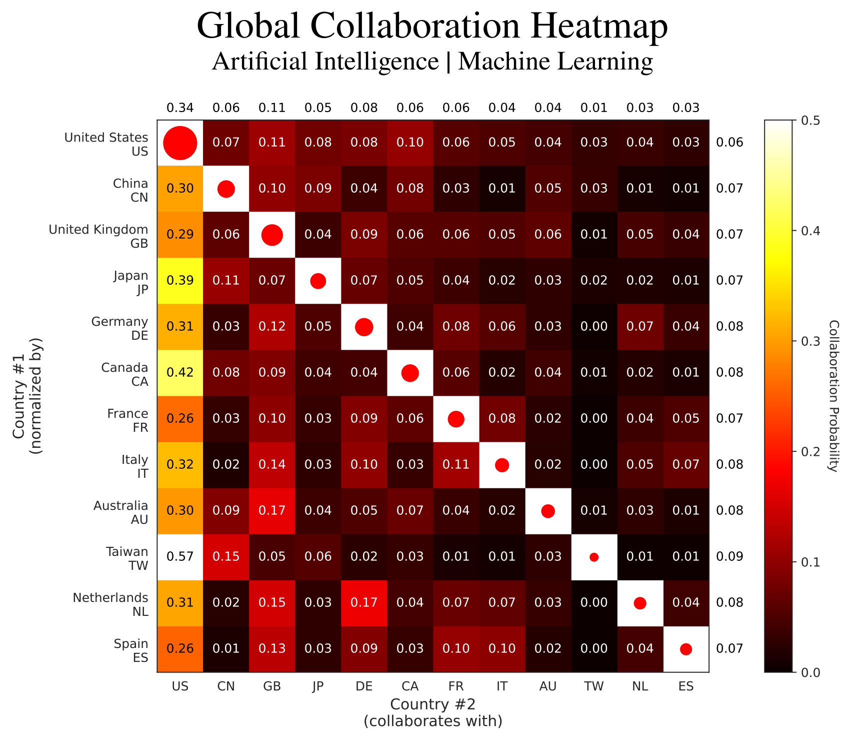
With 284 concepts at level 1 and a multitude of concepts at subsequent levels, approaching them in the same manner as the 19 level 0 concepts poses a challenge. Consequently, this subsection will narrow its focus to specific research areas, showcasing the capability to explore global collaboration within distinct topics, such as AI and ML, nanotechnology, and quantum science. It is important to note that the selected concept names may not cover all aspects of these areas, but the objective is to offer a snapshot of collaboration trends and their evolution over time.

Figure 35 presents the global collaboration heat map for the concept areas of AI and ML, showcasing trends both a) before and b) after 2010. Pre-2010, the U.S. led in AI and ML publications, producing the highest number of multi-country articles. Collaborations in this domain predominantly involved the U.S., with other countries often jointly publishing papers. Post-2010, China, the U.S., and the United Kingdom emerged as prominent leaders. While the U.S. maintained collaborations, there was a notable increase in partnerships with China and the United Kingdom. The index above the heat map reflects a decline in the U.S. index and a corresponding rise in China's index. Some cases even show similar percentages of multi-country articles between the U.S. and China. Australia, notably, increased collaborative publications with China. Examining the size of the red circles reveals a rise in multi-country publications for China and the United Kingdom, closing the gap with the U.S. Shifts in the top countries in this field, with Taiwan, the Netherlands, and Spain falling from the Top 12, and increased contributions from India, Korea, and Indonesia, are additional factors influencing these dynamics.

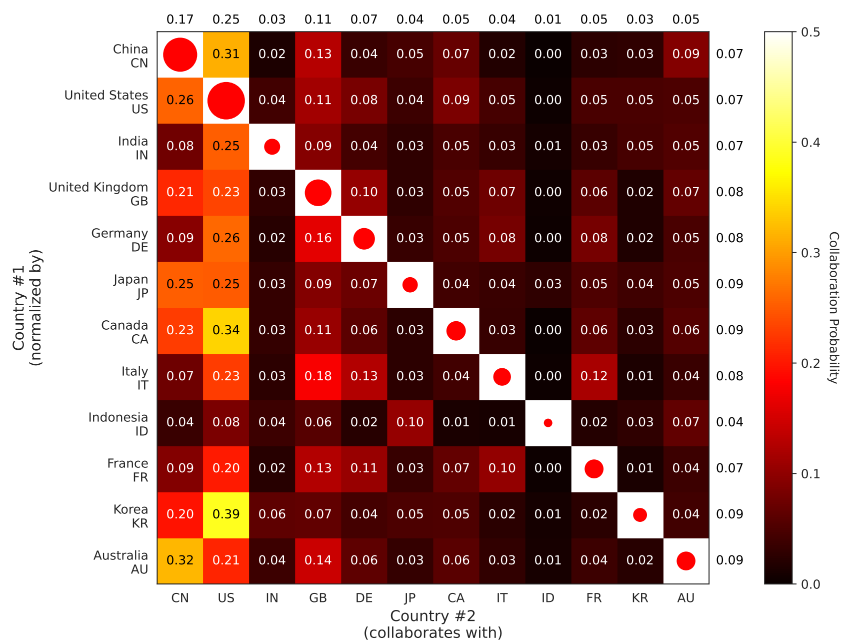
Figure 36 presents a heat map for nanotechnology, encompassing research concepts like nanotechnology, nanomaterials, nanostructure, nanocrystal, nanocrystalline material, nanoparticle, nanocomposite, nanowire, nanorod, and nanofiber. This map exhibits similarities to the AI and ML heat map. Pre-2010, the U.S. dominated nanotechnology publications, producing the most multi-country articles and often collaborating with other top countries. Post-2010, a shift occurred, with China taking the lead in nanotechnology publications. The red circle, indicating the number of multi-country articles, shows China and the U.S. in close collaboration, with each being a significant partner in the other's multi-country publications. The collaboration dynamics have evolved significantly, as evidenced by the index above the heat map, highlighting China's increased presence and the narrowing gap with the U.S. Interestingly, while the U.S. remains a probable collaborator for other top countries, China has closed this gap and become the leading contributor to nanotechnology research worldwide.

Figure 37 displays a heat map for various aspects of quantum science, incorporating research concepts like quantum tunneling, quantum entanglement, quantum dot, quantum sensor, and quantum electrodynamics. While these concepts do not cover the entirety of quantum science, they offer insights into the broader landscape. The heat maps, both before and after 2010, reveal several trends. Pre-2010, the U.S. and Germany were the top publishers in this field, with several other countries frequently collaborating on multi-country papers with them. China, although ranking fourth in publications, had a lower collaboration index. Post-2010, the global collaboration landscape has evolved significantly, with China now surpassing Germany and Japan to become the second-largest contributor to quantum science. The collaboration index indicates that China is now a more frequent collaborator with other nations in this field. Importantly, this increase in collaboration by China does not seem to come at the expense of other leading countries. The U.S. and Germany maintain similar collaboration rates, and the United Kingdom has also shown improvement. Overall, the heat map underscores the highly collaborative nature of research in quantum science, involving multiple countries in various capacities.

Figure 38 displays a heat map for traumatic brain injury and biomechanics, shifting to more medically-focused research concepts. Here, the U.S. has been the top producing country in these areas both before and after 2010, commanding a majority of the top-producing country's attention when it comes to multi-country articles. While China is second in terms of the number of articles in these concepts since 2010, the United Kingdom attracts many more partnerships with many of the top countries than China within these areas; i.e., global collaboration is more than just sheer number of articles produced. Proximity appears to play a role as well; many European nations collaborate with each other, while Japan and Korea collaborate more with China than with many of the top European countries.

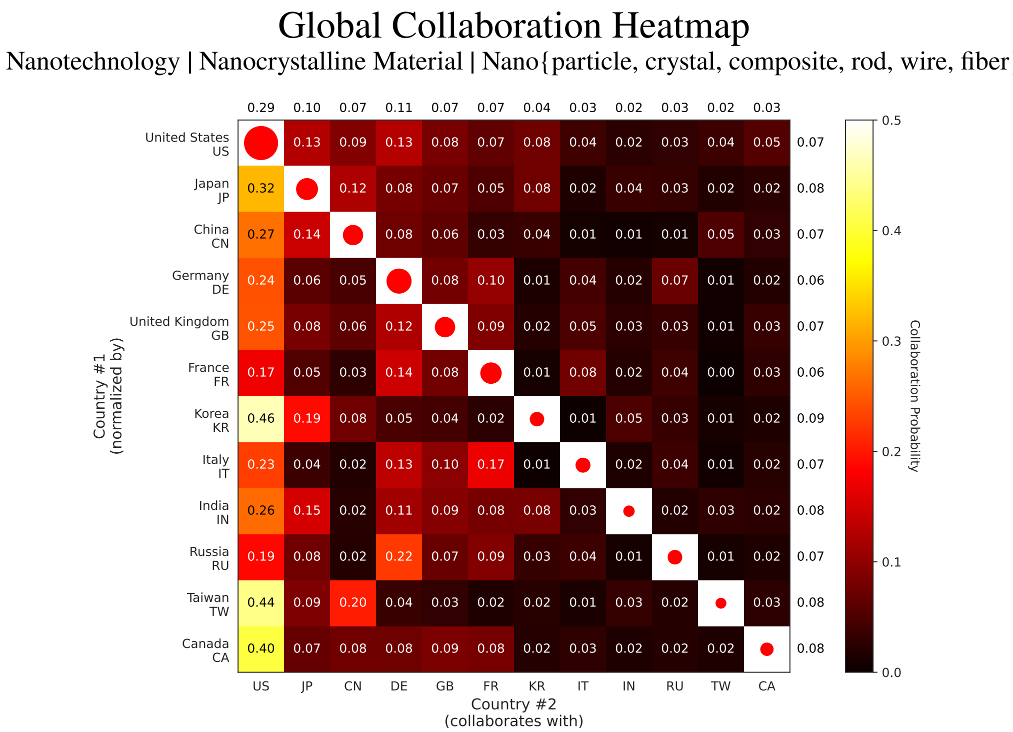


(a) >1 Country, Before 2010

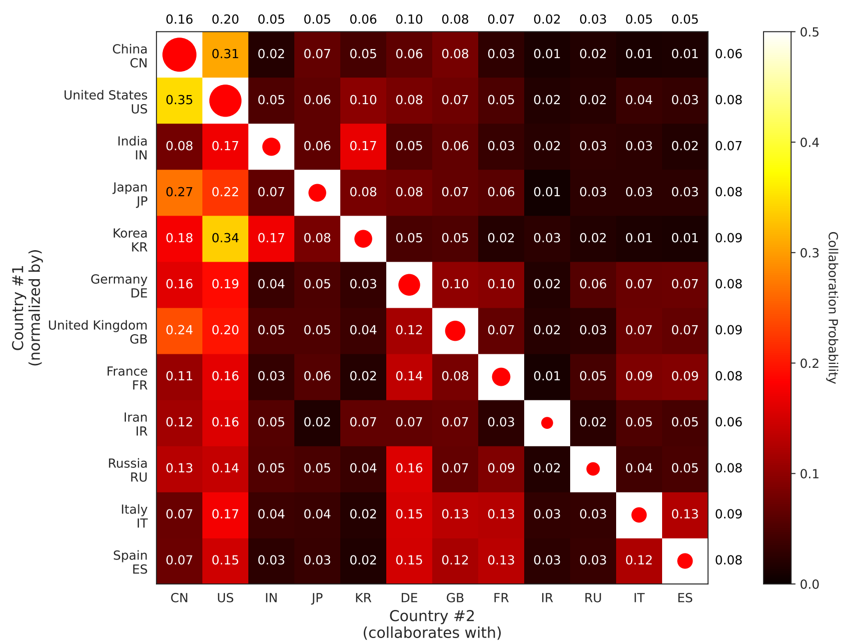


(b) >1 Country, 2010–2024

Figure 35. A heat map illustrating global collaboration among the top 12 most prolific countries in AI and ML (a) before 2010, and (b) after 2010. Countries are arranged based on their scientific article productivity in these concept areas, with the y axis representing the top-down order and the x axis depicting the left-to-right order. The formatting is similar to prior figures such as Figures 18 and 19.

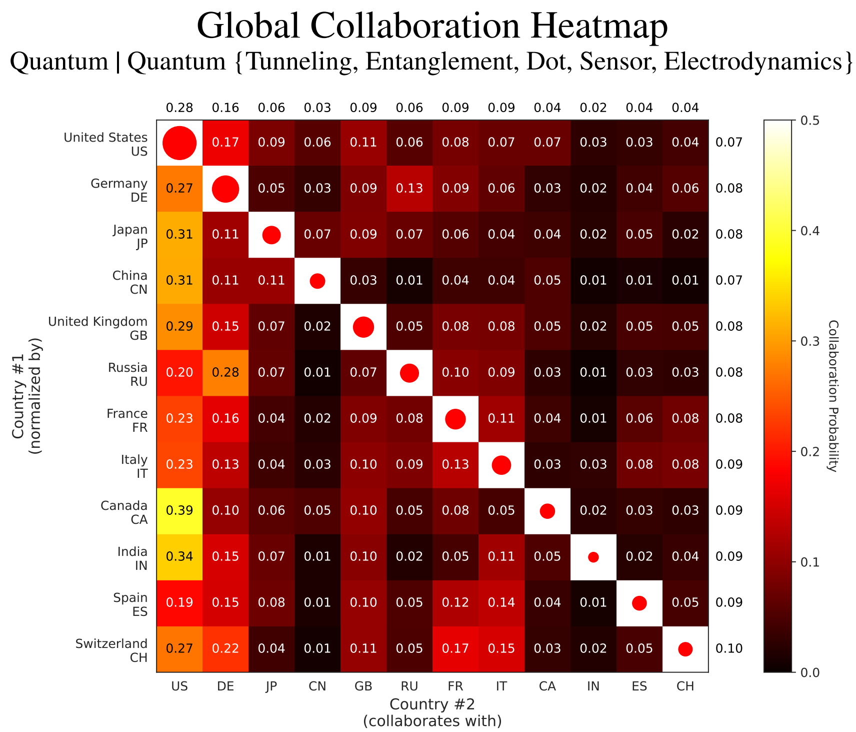


(a) >1 Country, Before 2010

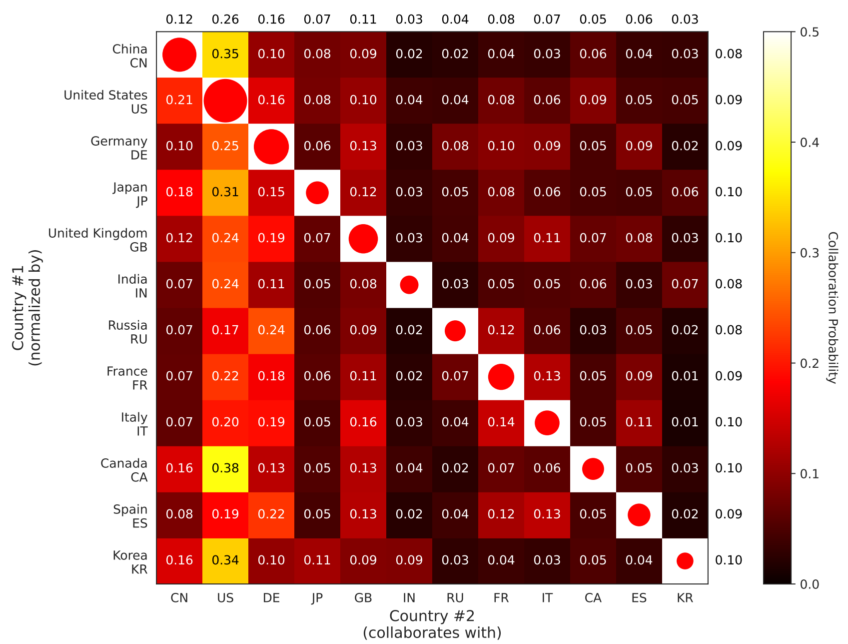


(b) >1 Country, 2010–2024

Figure 36. A heat map illustrating global collaboration among the top 12 most prolific countries in a number of nanotechnology concept areas (a) before 2010, and (b) after 2010. Countries are arranged based on their scientific article productivity in these concept areas, with the y axis representing the top-down order and the x axis depicting the left-to-right order. The formatting is similar to prior figures such as Figures 18 and 19.

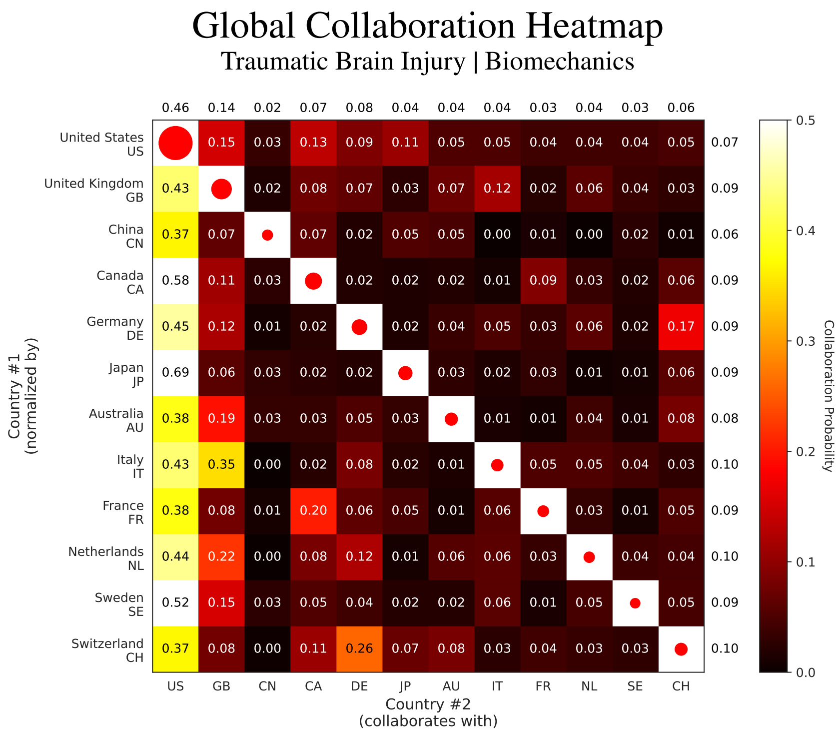


(a) >1 Country, Before 2010

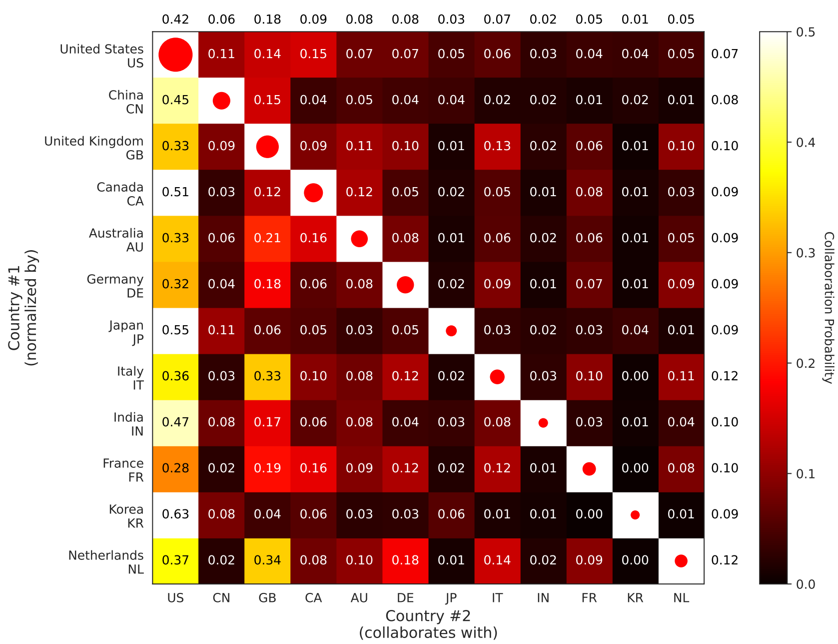


(b) >1 Country, 2010–2024

Figure 37. A heat map illustrating global collaboration among the top 12 most prolific countries in a number of quantum concept areas (a) before 2010, and (b) after 2010. Countries are arranged based on their scientific article productivity in these concept areas, with the y axis representing the top-down order and the x axis depicting the left-to-right order. The formatting is similar to prior figures such as Figures 18 and 19.



(a) >1 Country, Before 2010



(b) >1 Country, 2010–2024

Figure 38. A heat map illustrating global collaboration among the top 12 most prolific countries in traumatic brain injury and biomechanics concept areas (a) before 2010, and (b) after 2010. Countries are arranged based on their scientific article productivity in these concept areas, with the y axis representing the top-down order and the x axis depicting the left-to-right order. The formatting is similar to prior figures such as Figures 18 and 19.

3.9 Discussion: Beyond the Data

This report examines the frequency of global collaboration in scientific research, offering insights with potential implications for policymakers, funding agencies, research institutions, and individual scientists. By examining how collaboration patterns have evolved across countries, institutions, and research domains through the decades, these findings provide a foundation for decision-making on the future of international scientific cooperation.

One of the most striking trends is the rapid growth of China's research output and its increasing centrality in global collaborations. Historically, international research partnerships were primarily anchored in Western nations such as the United States, the United Kingdom, and Germany. However, over the past decade, China has emerged as a dominant player, significantly reshaping global scientific networks. This shift suggests a move toward a more multipolar research ecosystem, reducing Western dominance and fostering a global landscape with more distributed research hubs (i.e., countries). The diversification of research hubs has introduced new theoretical frameworks, methodologies, and problem-solving approaches, enriching scientific discourse and expanding access to funding and infrastructure beyond traditionally dominant regions.

While these changes present new opportunities, they also raise strategic considerations for nations seeking to maintain their research leadership. If China's investment trajectory continues to attract an increasing share of global collaborations, the research network could evolve from a multipolar structure to one in which a single nation exerts disproportionate influence. This could shape global research priorities, direct major funding flows, and influence the direction of scientific inquiry. Overconcentration in one region may also limit equitable access to research opportunities, making it harder for smaller nations or underfunded institutions to establish international partnerships. Because scientific influence is closely tied to economic and geopolitical factors, an imbalance in research collaboration could further affect intellectual property flows, technological development, and global innovation ecosystems.

This shift also has broader geopolitical parallels. Historically, integrating China into global trade and financial networks was a strategic decision by Western nations, intended to foster interdependence and exert influence through economic ties. However, just as China has leveraged its role in global supply chains to exert economic and strategic pressure, a similar dynamic may be unfolding in scientific research. The ability to shape research priorities, control key collaborations, and concentrate intellectual resources carries strategic weight, particularly in fields with technological and defense implications. As research networks become increasingly intertwined with geopolitical strategy, understanding the balance of influence within these networks is critical.

As global collaboration expands, a key question arises: Is this inherently beneficial? International partnerships are often associated with enhanced research quality, as they bring together diverse expertise, expand access to resources, and foster knowledge exchange. However, the true impact of these collaborations remains an open question, particularly in terms of how they shape research influence and scientific progress.

One important factor is the role of global collaboration in the citation economy—the system in which citation counts serve as a form of academic currency, influencing reputations, funding opportunities, and institutional rankings. Preliminary findings from a related study[13] suggest that partnerships with certain nations correlate with increased citation counts, a metric often linked to research quality and prestige. This trend may reflect greater visibility and dissemination of research outputs when authors collaborate with well-established institutions in influential scientific networks. However, citation counts do not always align with true scientific advancement, as they are shaped by network effects, regional citation practices, and journal impact dynamics. If collaborations primarily serve as a means to boost citations rather than advance knowledge, this raises critical questions about the incentives driving international partnerships. The very question of how citation incentives shape global collaboration is explored in this complementary study examining nation-state collaboration structures.[13]

Moreover, shifting collaboration networks create the potential for certain nations to exert disproportionate influence over research agendas. If citation advantages drive scientists to prioritize partnerships based on expected visibility rather than scientific necessity, research diversity could suffer. Ensuring that international collaborations are guided by genuine scientific value rather than strategic positioning within the citation economy is crucial for maintaining a balanced and resilient research landscape.

Ultimately, while the expansion of global collaboration has transformed scientific networks, its long-term implications remain complex. Balancing broad international partnerships with meaningful scientific contributions requires careful evaluation. Future research should continue to explore how these shifting collaboration patterns influence knowledge production and dissemination on a global scale.

This study is not intended to advocate for or against any particular policy but rather to provide a data-driven analysis of how global collaboration patterns have evolved over time. The findings highlight the changing dynamics of international scientific networks and the potential implications of these shifts. Policymakers may consider various levers to incentivize or disincentivize collaboration with certain nations if the global research landscape begins to exhibit structural imbalances—whether intentional or emergent. While global scientific cooperation can foster innovation and progress, it is also intertwined with broader economic and strategic considerations, making it essential for decision-makers to evaluate these trends with a nuanced understanding of their potential benefits and challenges.

4. Conclusion

Based on the analysis of the OpenAlex database, this study provides insights into the role of collaboration in the scientific landscape, with a focus on international collaborations that transcend geographical boundaries. The findings of this study reveal the following:

1. **Evolution of Science and Global Collaboration:** While the number of articles, book chapters, books, and articles in the U.S., China, the United Kingdom, Japan, and Germany has increased since 1900 (Figures 4–6, 8–12), so has the percentage of articles with authors from more than one country (Figure 13). The analysis revealed a shift in global collaboration trends over time, showcasing the steady increase in the number of collaborative publications over time.

2. **Novel Method for Analysis and Visualization:** The heat maps (e.g., Figure 18) assembled give a rapid indication of asymmetries and trends in global collaboration pertaining to the frequency of different countries' collaborations

with other top countries. The individual squares within the maps highlight the probabilities of collaboration between different countries. Additionally, new indices were created to evaluate the extent of a country's collaboration with other top countries (rows) and its role in the collaborations of others (columns), i.e., the egocentric collaboration index ECI and the collaborative role index CRI, respectively.

3. **Asymmetry in Global Collaboration:** There are asymmetries in global collaboration for countries, U.S. institutions, U.S. funding agencies, and research areas, with some countries and institutions being more dominant than others within the international collaboration landscape. The use of asymmetric co-occurrence probabilities in heat maps facilitated a more realistic representation of directed collaborations, acknowledging the directionality in research relationships. For instance, in the decades spanning 1990–2019, the U.S. was a leader in the global collaboration landscape, with an author on a majority of multi-country articles for all top publishing countries at the time (1990s, 2000s, and 2010s, shown in Figures 21, 22, and 23, respectively). While the U.S. remains a leader in global collaboration of science in the recent past (since 2019), the global collaborations are more diffuse than observed in prior decades.

4. **Global Collaboration of U.S. Institutions and Funding Agencies:** Many U.S. national laboratories have a significant fraction of multi-country articles with China, the United Kingdom, and Germany (Figure 24). Perhaps the most observable trend is that national laboratories with major focuses on energy research, development, and engineering tended to have more collaborative publications with China than other U.S. national laboratories. However, another trend emerges. Of these multi-country articles, over 90% have an academic collaborator as an author on the article as well. U.S. universities have many collaborative articles with other countries, but in most cases no more than the U.S. institutions (Figures 25 and 26). U.S. funding agencies often fund PIs who, in the course of their research, collaborate on articles with authors from other countries; similar to other U.S. institutions, many of these collaborations are with China, the United Kingdom, and Germany (Figure 27).

5. **Global Collaboration of Research Concepts:** This study explored collabo-

ration patterns in specific research areas (concepts) at a high level (e.g., computer science, materials science) and more detailed level (AI, nanotechnology, quantum science, biomechanics), shedding light on global collaboration in these domains. In high-level research concepts (i.e., level 0 concepts), the U.S., China, and the United Kingdom (and Germany to a lesser extent) lead in global collaboration over all concepts (Figure 31). In examining more detailed concepts, while the U.S. has led in many areas prior to 2010, the U.S., China, and the United Kingdom have emerged as prominent leaders in global collaboration post-2010. China has made significant strides in the number of articles in most areas, which has shown an increase in multi-country collaborations in some areas (AI/ML, nanotechnology, and quantum science in Figures 35, 36, and 37, respectively) but not all (e.g., traumatic brain injury/biomechanics in Figure 38).

In conclusion, this study highlights collaboration in the scientific landscape with a particular focus on the dynamics of international collaborations. The findings offer interesting insight into the role of collaboration beyond borders in advancing scientific progress and can serve as practical guidance for policymakers, funding agencies, and researchers aiming to understand how to unpack some of the complexity of various factors involved in global collaboration on the scale of millions of scientific articles.

While this study focuses on how often do countries collaborate, it does not delve into why this collaborate occurs between different countries; i.e., what is the incentive for this increasing behavior? The next study dives into the citation economy of global collaboration and how an analysis of that may provide some answers.[13]

5. References

1. Wang D, Barabási AL. The science of science. Cambridge University Press; 2021.

2. Fortunato S et al. Science of science. Science. 2018;359(6379).

3. Sinha A et al. An overview of Microsoft Academic Service (MAS) and applications. In: Proceedings of the 24th International Conference on World Wide Web; ACM; 2015.

4. Wang K et al. A review of microsoft academic services for science of science studies. Front Big Data. 2019;2.

5. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv. 2022 May 4. arXiv:2205.01833.

6. OpenAlex. [accessed 2025-01-29]. https://openalex.org/

7. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science. 2007;316(5827):1036–1039.

8. Adams J. The fourth age of research. Nature. 2013;497:557–560.

9. Tschopp MA. From scientific institutions to research concepts to global collaboration and beyond: Uncovering patterns in scientific networks. DEVCOM Army Research Laboratory (US); 2024 Feb. Report No.: ARL-SR-0493.

10. Papernot N et al. The limitations of deep learning in adversarial settings. 2016 IEEE European Symposium on Security and Privacy (EuroS&P); 2016; Saarbrücken, Germany. 372–387.

11. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the folin phenol reagent. J Biol Chem. 1951;193(1):265–275.

12. OpenAlex API technical documentation. [accessed 2024 Oct 24]. https://docs.openalex.org.

13. Tschopp M.A. Global collaboration dynamics in scientific research: Why does it occur? Examining the citation economy. DEVCOM Army Research Laboratory (US); 2025 Feb. Report No.: ARL-TR-10058.

List of Symbols, Abbreviations, and Acronyms

AGENCY ACRONYMS:

AFOSR	Air Force Office of Scientific Research
ARL	Army Research Laboratory
ARO	Army Research Office
DEVCOM	U.S. Army Combat Capabilities Development Command
DOE	Department of Energy
NASA	National Aeronautics and Space Administration
NETL	National Energy Technology Laboratory
NSF	National Science Foundation
ONR	Office of Naval Research

OTHER ACRONYMS:

2D	two-dimensional
AI	artificial intelligence
API	application programming interface
COVID-19	coronavirus disease 2019
CRI	collaborative role index
DOI	digital object identifier
ECI	egocentric collaboration index
ID	identifier
ISSN	international standard serial number
ISSN-L	linking international standard serial number

JSON	JavaScript object notation
ML	machine learning
ORCID	Open Researcher and Contributor ID
PI	principal investigator
QR	quick response
ROR	Research Organization Registry
STEM	science, technology, engineering, and mathematics
URL	uniform resource locator (web address)