Journal of Informetrics 19 (2025) 101665

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Research Paper

A large-scale temporal analysis of scientific production across disciplines and countries ☆

Irene Finocchi [a], Andrea Ribichini [b], Marco Schaerf [b,*]

[a] Department of AI, Data and Decision Sciences, Luiss Guido Carli University, Viale Romania 32, 00197 Rome, Italy
[b] Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy

ARTICLE INFO

Keywords:
Scientific productivity
Temporal evolution
Research fields
Citations
Publications
Gender gap

ABSTRACT

In this article, we undertake a comprehensive large-scale analysis of the evolution of scientific communities across different disciplines and countries, spanning the period 1991-2020. Our analysis uses data obtained from Scopus and involves a total of 15,756,144 authors, 74,847,508 publications, and 1,501,206,153 citations. Besides the overall research production, we investigate multiple disciplines at various levels of aggregation (namely, scientific sectors as defined by the European Research Council and Scopus research categories). The geographical focus of our analysis takes into account first the worldwide scientific production and then addresses the 19 countries that are members of the G20 group (thus excluding the EU).

Research production generally increases with time (in terms of authors, publications, and citations), both on a global scale and specifically in each country. The growth is not only in terms of raw numbers but also relative to population and gross domestic product. The gender gap appears to be narrowing, albeit at a slower pace for STEM disciplines than others. Although the United States started out as the dominant country in all research fields, its primacy has eroded constantly with the passage of time. The fastest growing emerging country, China, recently managed to overtake the United States, at least in STEM disciplines.

1. Introduction

Analyzing research productivity provides an opportunity to enhance the efficiency, effectiveness, and impact of academic research, which benefits both individual researchers and institutions as a whole. Examining the temporal progression of research communities is crucial, offering valuable insight into the advancement of various knowledge production domains. Furthermore, it contributes to the exploration of bibliometric indicators, increasingly employed – although with some controversy – in various countries to guide career advancements (see, e.g., Smith et al., 2013; Demetrescu et al., 2020) and allocation of research funds (see, e.g., Stuart, 2015; Franceschini & Maisano, 2017; Kulczycki, 2017; Lupia et al., 2018; Demetrescu et al., 2019, 2020).

In this article, we present the results of a large-scale study that analyzes the characteristics of scientific productivity across three different dimensions: temporal, geographical, and disciplinary. Along the temporal dimension, we span a period of thirty years, from

☆ This work has been partially supported by MUR, the Italian Ministry of University and Research, under PRIN Project n. 2022ME9Z78 "NextGRAAL— Next-generation algorithms for constrained GRAph visuALization".

* Corresponding author.
E-mail addresses: finocchi@luiss.it (I. Finocchi), ribichini@diag.uniroma1.it (A. Ribichini), marco.schaerf@uniroma1.it (M. Schaerf).

https://doi.org/10.1016/j.joi.2025.101665
Received 9 January 2024; Received in revised form 5 March 2025; Accepted 9 April 2025

1991 to 2020. We examine research disciplines at three aggregation levels: overall scientific production, scientific production divided into three main sectors (Life Sciences, Physical Sciences and Engineering, and Social Sciences and Humanities) as defined by the European Research Council, and scientific production classified into 27 research categories as defined by Scopus. In addition to a worldwide view, we also address selected countries, with a focus on G20 member nations whose population and gross domestic product (GDP) are also considered.

By combining these three orthogonal and mutually independent dimensions, we aim to shed light on various inquiries related to the evolution of research over time and to the temporal changes of subcommunities in terms of authors, publications, and citations. We explore variations in size at both macro-sector and individual research category levels, also investigating shifts in female representation and assessing whether the research focus has changed during our 30-year observation period. Furthermore, we examine whether global patterns align with individual countries or if there are distinct country-specific behaviors. More systematically:

- We conduct a thorough analysis of the scientific community, focusing on its size and composition. In addition to revealing how the number of researchers has changed across different disciplinary areas and countries, we examine women's participation in the academic community. Our analysis quantifies the improvements over time (or lack thereof) with respect to the presence of women, especially highlighting which areas and countries exhibit the most substantial progress.
- We perform a comprehensive analysis of scientific production, examining research outputs both from a productivity perspective, i.e., based on the number of publications, and in terms of impact, i.e., considering metrics such as the number of citations. We highlight trends at different levels of detail, from global worldwide production down to ERC sectors and specific research areas.
- We correlate changes in the scientific community with changes in scientific production. Our main goal is to understand whether individual researchers have become more productive over time or if the number of publications is just growing due to an increase in the size of the scientific community. By analyzing these aspects across research areas and countries, we aim to uncover significant shifts and trends in the relevance of the various disciplines in different countries and to pinpoint which countries are more focused on specific topics and which focus on a broader set of disciplines in the development of research.

To the best of our knowledge, this is the first comprehensive, large-scale quantitative examination of the global scientific output, covering multiple dimensions and offering a wide overview that ranges from the worldwide scientific production to details of each single discipline in each single country.

The analysis is based on the analysis of data sourced from Elsevier's Scopus,[1] a widely recognized global bibliographic repository, and encompasses a comprehensive dataset featuring 15,756,144 authors, 74,847,508 publications, and 1,501,206,153 citations.

2. Related work

The temporal evolution of scientific production has been a central theme in a variety of studies over the past decades. Previous investigations often examine how patterns of production and collaboration evolve over time, encompassing diverse time spans, countries, and disciplines. While many studies concentrate on specific countries or disciplines, a few others adopt a more comprehensive approach to capture broader trends. What follows is a concise overview of the literature, focusing exclusively on works that give prominence to the temporal dimension.

A variety of works explore the connection between scientific productivity and economic wealth, establishing correlations between scientific output, GDP, and investments in research and development. In King (2004) the top performing nations, in terms of knowledge production, are analyzed over several research fields, taking into account population and gross domestic product as scale factors. It is shown that the US maintain a larger footprint mostly due to their dominance in life sciences. In contrast, EU15 countries perform better in physical sciences and engineering. Significant disparities in citation rankings among countries are also highlighted, with the top eight nations producing approximately 84.5% of the most cited publications. More recently, Jaffe et al. (2020) highlight that research productivity is closely linked to a nation's intellectual and economic wealth, uncovering shared trends in the distribution of research productivity between disciplines: high-income nations exhibit similar patterns in research discipline importance and form a stable cluster characterized by consistent relative productivity dynamics over time. In contrast, lower-income countries form smaller, more independent clusters with significant differences in their research, and appear to be also influenced by geographical proximity.

Courtioux et al. (2019) establish a model linking public expenditure in scientific research with world share of publications and citations for each nation. Their model provides a justification for the steep rise of China, which has allocated more funds than its competition to research in recent decades, at least in specific fields. China is indeed the subject of numerous works that focus on specific countries and their scientific production. Its progressive emergence over other nations has been investigated, for instance, in Zhou and Leydesdorff (2006), highlighting significant policy shifts that have facilitated its rise: the authors note that since the late 20^{th} century China has invested heavily in science and technology, moving from a position of relative insignificance in global science to becoming one of the top contributors to scientific literature. The quantity vs. quality issue is also discussed. Along the same line, Wang (2016) emphasize how hard sciences have progressed more rapidly than soft sciences in China.

Leydesdorff and Wagner (2009) focus instead on the position of the United States within the global science landscape, analyzing the erosion of US preeminence in scientific research. While the US continues to lead in total publications and citations, there is evidence suggesting that other countries, particularly in Asia and Europe, are increasing their research outputs at a faster rate. The

[1] This work uses Scopus data provided by Elsevier through ICSR Lab.

authors argue that, in a more competitive global environment, maintaining leadership in science requires adaptation to emerging global trends and fostering international collaborations: collaborative networks can enhance research quality and impact, allowing countries to leverage each other's strengths.

Several other articles target instead specific disciplines or their subfields. Liang et al. (2015) investigate the field of arthroscopy between 1999 and 2013, noting a positive correlation between scientific output and population/gross domestic product (GDP). Quantity and quality of endocrinology and diabetology research also appear to be associated with GDP, expenditure for R&D, and number of researchers per million inhabitants (Mantovani et al., 2020). There are however some notable exceptions, most prominently Italy, whose research output in this field is higher than expected. Das et al. (2013) examine research productivity in the field of economics between 1985 and 2005, reporting a strong correlation between research output and GDP per capita. They also highlight how, in this discipline, articles about the US are more likely to be published in top-tier journals than articles about any other country. Guan and Ma (2004) compare the research performance in the field of computer science (CS) of China, US, Germany, UK, Japan, and India over the time period 1993-2002. The authors divide CS into eight subfields and analyze both productivity and impact of research, clearly identifying a catch-up trend for China relative to the other nations. Computer science is also the focus of Demetrescu et al. (2022), who examine G20 member countries over the years 1991-2020. The area is divided into 12 subfields and shifting trends are observed between them. Additional articles also target more specific fields. For instance, the global scientific production of geographic information systems (i.e., computerized systems for geographically referenced data) over a ten-year time period is provided in Tian et al. (2008), also identifying the most productive countries in this area. Finally, moving from the observation that interdisciplinary research is becoming more and more popular, the overlapping across different fields over time is studied in Sun et al. (2016) using a network-based approach.

In this paper we contribute by applying a temporal lens on a more global scale, providing a distinct and broader perspective compared to the aforementioned studies. With a similar goal, the study by Bongioanni et al. (2014) presents a quantitative framework for analyzing and comparing the dynamics of research productivity across different disciplines over time. The analysis reveals significant differences in disciplinary focuses between countries: nations like India and China are increasing their contributions to international research while maintaining distinct disciplinary strengths, suggesting a dynamic interplay between globalization in science and national specialization. Bongioanni et al. (2014) advocate for further empirical studies to validate their findings and to explore the implications for national and regional research strategies. Our work contributes along this direction.

3. Datasets and approach

In this section, we provide details on the datasets used in our study (Section 3.1) and on our data-analytics methodology to extract from them information concerning the time evolution of scientific production (Section 3.2). Considerations and nuances inherent in our approach are discussed in Section 3.3.

3.1. Datasets

Our study is based on Scopus data provided by Elsevier's ICSR Lab. The main focus is on publications indexed by Scopus between January 1^{st}, 1991, and December 31^{st}, 2020, totaling 74,847,508 publications. The information for each article includes its publication year, citations received by March 1^{st}, 2021, Scopus-assigned subject categories, and a list of unique author IDs together with their affiliations, as reported in the article itself, in the form of $\langle institution, country \rangle$ pairs. We also consider 15,756,144 authors, maintaining their unique ID, frequency vector of subject categories, and gender information as inferred by Scopus through a machine learning-based classifier. The classifier takes as input an author's first name, last name, and country of first affiliation. We only took into consideration gender attributions provided with a confidence level of at least 85%. Despite potential inaccuracies at the individual level, the classifier's output can be used to indicate general trends, as reported in the ICSR Lab documentation.

3.2. Approach

In order to assign disciplines, at different levels of granularity, to authors and publications, we resort to Scopus' own division of knowledge production into 27 subject categories and to the European Research Council's classification into 3 sectors. Table 1 lists all Scopus subject categories and ERC sectors, together with their respective abbreviations (extensively used throughout the article) and with our mapping between the two classification systems.

For the purpose of our analysis, we have divided our 30-year observation period (1991-2020) into six consecutive quinquennial time frames (e.g., 1991-1995, 1996-2000, and so forth). By joining the publications and authors datasets based on matching author ids, we retrieved the list of articles of each author. Publications and authors are assigned one or more Scopus categories (or ERC sectors), and associated with one or more countries, based on the following criteria.

Publications. We assign a Scopus category to a publication if that category is in the set of subject categories assigned by Scopus to that publication. A publication is assigned an ERC sector if its subject category set contains at least one Scopus category that maps to that sector. A publication is attributed to a country if at least one of its authors is affiliated, as reported in the publication record itself, with at least one institution located in that country. Finally, a publication is assigned to a time frame tf if its release date falls within the boundaries of tf.

Authors. An author a is assigned a Scopus subject category sj if the entry in a's frequency vector corresponding to sj is ≥ 5, and it is at least 50% of the maximum in the entire vector. An author is considered active in a given time frame if they have contributed to at

Table 1
Full lists of Scopus subject categories and ERC sectors, together with their abbreviations. Our mapping between the two classification systems is also displayed.

Scopus category	Scopus category abbrev.	ERC sector	ERC sector abbrev.
Agricultural and Biological Sciences	AGRI	Life Sciences	LS
Biochemistry, Genetics and Molecular Biology	BIOC		
Dentistry	DENT		
Health Professions	HEAL		
Immunology and Microbiology	IMMU		
Medicine	MEDI		
Neuroscience	NEUR		
Nursing	NURS		
Pharmacology, Toxicology and Pharmaceutics	PHAR		
Veterinary	VETE		
Chemical Engineering	CENG	Physical Sciences and Engineering	PE
Chemistry	CHEM		
Computer Science	COMP		
Earth and Planetary Sciences	EART		
Energy	ENER		
Engineering	ENGI		
Environmental Science	ENVI		
Materials Science	MATE		
Mathematics	MATH		
Physics and Astronomy	PHYS		
Arts and Humanities	ARTS	Social Sciences and Humanities	SH
Business, Management and Accounting	BUSI		
Decision Sciences	DECI		
Economics, Econometrics and Finance	ECON		
Psychology	PSYC		
Social Sciences	SOCI		
Multidisciplinary	MULT	--	--

least one publication in that time frame. An author, active in time frame tf, is associated with country c for the duration of tf if they have c as affiliation country in at least one publication to which they have contributed during tf.

We coded our data preprocessing, aggregation and extraction tasks in the Databricks environment, provided by Scopus, using PySpark.

3.3. Threats to validity

Our study is based on a large population of authors and publications, thus minimizing the potential impact of author disambiguation related issues, such as multiple author profiles or misattribution of citations within Scopus, arguably one of the most accurate bibliographic databases available (see, e.g., Baas et al., 2020; Pranckutė, 2021).

Authors' affiliations in different Scopus publication records can still present discrepancies in affiliation formats. Although Scopus has an entity resolution algorithm to address these discrepancies, differing affiliation wording may lead to incorrect associations. The focus on affiliation countries rather than specific institutions makes this issue a minor concern.

For all publications, citations are recorded up to a specific date (March 1^{st}, 2021), irrespective of their release year. This approach could potentially disadvantage recently released publications due to the shorter time span available for accumulating citations. However, the bias is mitigated, especially for older publications, by the fact that many publications garner significant citations within a few years of release, followed by a decline phase (see, e.g., Walters, 2011; Pradhan et al., 2019). In this article, whenever we report the absolute number of citations, we will limit our scope to publications released no later than 2010, giving each publication a time window of at least 10 years in which to accrue citations. When dealing with world shares of citations among different research

Table 2
Average gender unknown authors per country over all time frames of interest.

Country	*Average gender unknown authors over all time frames*
Argentina	32.75%
Australia	38.71%
Brazil	39.84%
Canada	39.22%
China	71.38%
France	36.76%
Germany	30.96%
India	51.92%
Indonesia	64.70%
Italy	30.75%
Japan	35.31%
Mexico	47.49%
Russian Fed.	66.06%
Saudi Arabia	43.05%
South Africa	43.15%
South Korea	67.07%
Turkey	49.66%
United Kingdom	35.75%
United States	38.77%

areas, we will extend our analyses up to the year 2020, assuming that all areas incur a comparable citation bias for their most recent publications.

A single article may in general be included in more than one research category, or even more than one ERC sector, based on the set of subject categories that Scopus has assigned to it. Moreover, we assign a publication to a country if at least one of its authors reports that country as part of their affiliation. We do not use fractional counting in any of these cases.

A single author may also be assigned more than one research category, or even ERC sector, if they meet multiple times the requirements specified in Section 3.2. In a given time frame, an author is assigned to all countries they are affiliated to in articles published during that time frame. Again, no fractional counting is used in these cases.

As noted in Section 3.1, we rely on Scopus' own classifier for gender attribution to authors, considering valid only attributions with a confidence level of at least 85%. Not all authors are thus assigned a gender. "Gender unknown" percentages vary slightly between the six time frames, with significant differences between countries, as reported in Table 2 for the G20 group. The global average of gender unknown authors over all time frames is 45.01%. Values range from as low as 30.75% for Italy to as high as 71.38% for China. We emphasize that, whenever we compute gender statistics, we only consider the subset of authors with a known gender, and not the entire population. This may potentially lead to biased statistics, in the hypothesis that the gender unknown author set suffers itself from such a bias.

4. Scientific growth across ERC sectors and research areas

This section describes the temporal evolution of global scientific production (Section 4.1) as well as that of individual ERC sectors (Section 4.2). We then zoom in to the level of individual research categories, analyzing authors and research productivity for each area (Section 4.3 and Section 4.4, respectively).

4.1. An expanding scientific community with increased productivity

Fig. 1 provides information on the temporal evolution of world scientific production from 1991 to 2020, without distinctions between research areas or countries. Subfigures (a), (b) and (c) report, for each time frame of interest, the number of authors (male and female authors are also individually displayed), publications, and citations received cumulatively by all publications in that time frame at the date of March 1^{st}, 2021 (possible limitations inherent in this approach are discussed in Section 3.3), respectively. We remark that male and female authors do not sum to total authors because of a non-negligible number of authors to which Scopus' inference mechanism did not attribute a gender (see Section 3).

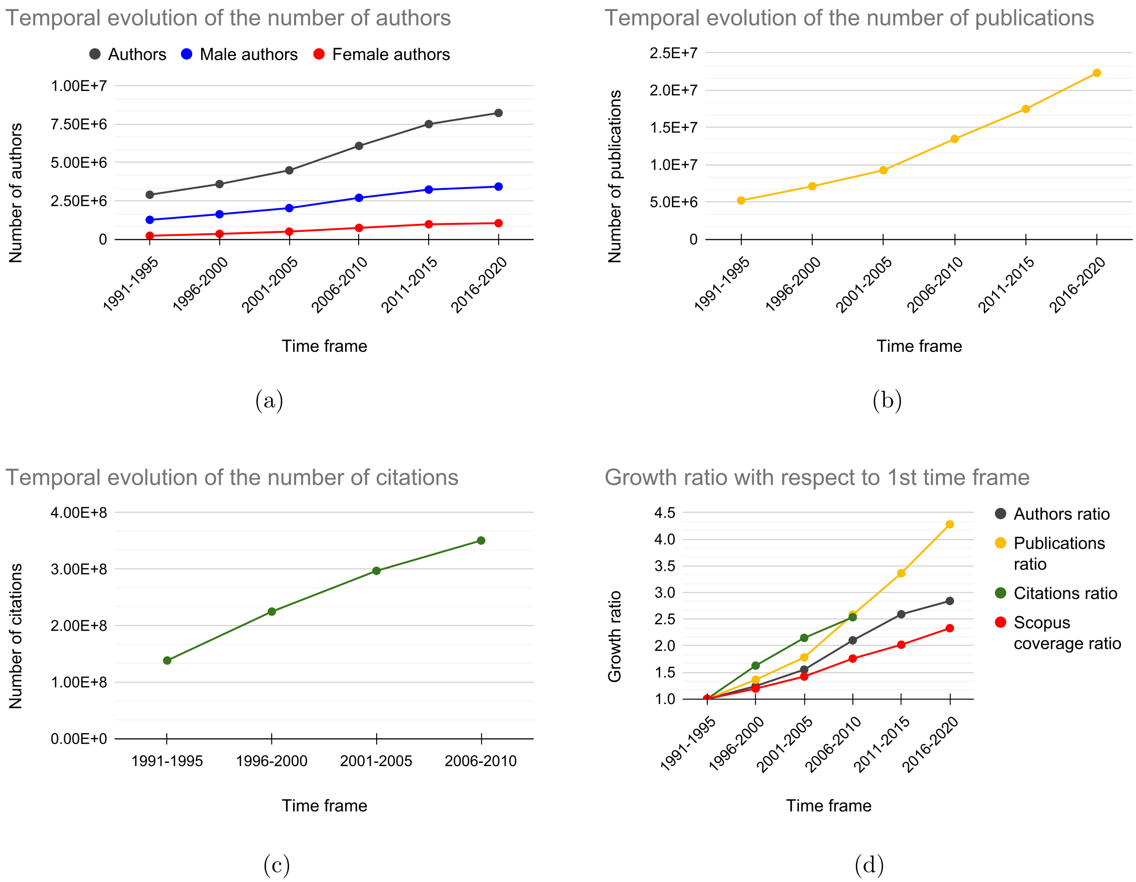

Fig. 1. Temporal evolution of world's scientific production.

Fig. 1(d) shows, for each of the above quantities, its ratio compared to the 1991-1995 value. We observe that publications are increasing more rapidly than authors, suggesting an increase in the productivity of individual researchers. The plots concerning citations stop in the 2006-2010 time frame because of the observations in Section 3.3.

The increase in citation numbers can also be influenced by the increase of open access publications. This issue, however, has not yet been settled in the relevant literature (see, for example, the works of Langham-Putrow et al. (2021); Plume (2024); Clark et al. (2024)).

It may be the case that publications have a steeper growth rate than authors due to wider coverage of the Scopus database over time. In order to analyze this issue, we provide ratios also for Scopus coverage, expressed as the number of sources (e.g., journals, series, periodicals) actively indexed in each time frame. To assess scale, we remark that the absolute number of indexed sources in 1991-1995 was 12,351. Scopus coverage increased significantly less than the number of authors and publications over the observation period. Indexed sources grew by a factor of 2.33, the number of authors by 2.84, and most notably the number of publications by 4.29. This strongly suggests that improved Scopus coverage should not be regarded as the only cause behind the detected phenomenon.

The observed increase in the number of citations (see, e.g., Fig. 1(c)) is in line with the findings in Nicolaisen and Frandsen (2021), which show that the average number of references in an article has steadily increased in recent years.

4.2. Delving into ERC sectors

Fig. 2 depicts the temporal evolution, over the period 1991-2020, of the three macrosectors defined by the European Research Council. In particular, subfigures (a), (b), and (c) report the evolution of number of authors, publications, and citations, while subfigures (d), (e), and (f) provide the ratios of these quantities relative to their values in the most remote time frame (i.e., 1991-1995). The above graphs also show trends for the MULT category, which corresponds to multidisciplinary research and is the only Scopus category not assigned to any ERC sector (see Section 3.2). The graphs show increasing trends for all observed quantities.

As shown in Fig. 2(a), LS authors remain the most numerous, with PE authors closely following and SH authors farther away. MULT authors increased from 1,180 to 7,502 and – not surprisingly – represent a much smaller community than any of the ERC sectors. However, when considering the authors' growth rate in Fig. 2(d), we interestingly observe an opposite scenario: MULT has a very steep curve, followed by that of SH. PE, in turn, exhibits a growth rate greater than that of LS. In the realm of publications, it seems noteworthy that PE managed to overtake LS in the 2001-2005 period, increasing the gap ever since, as shown in Fig. 2(b).

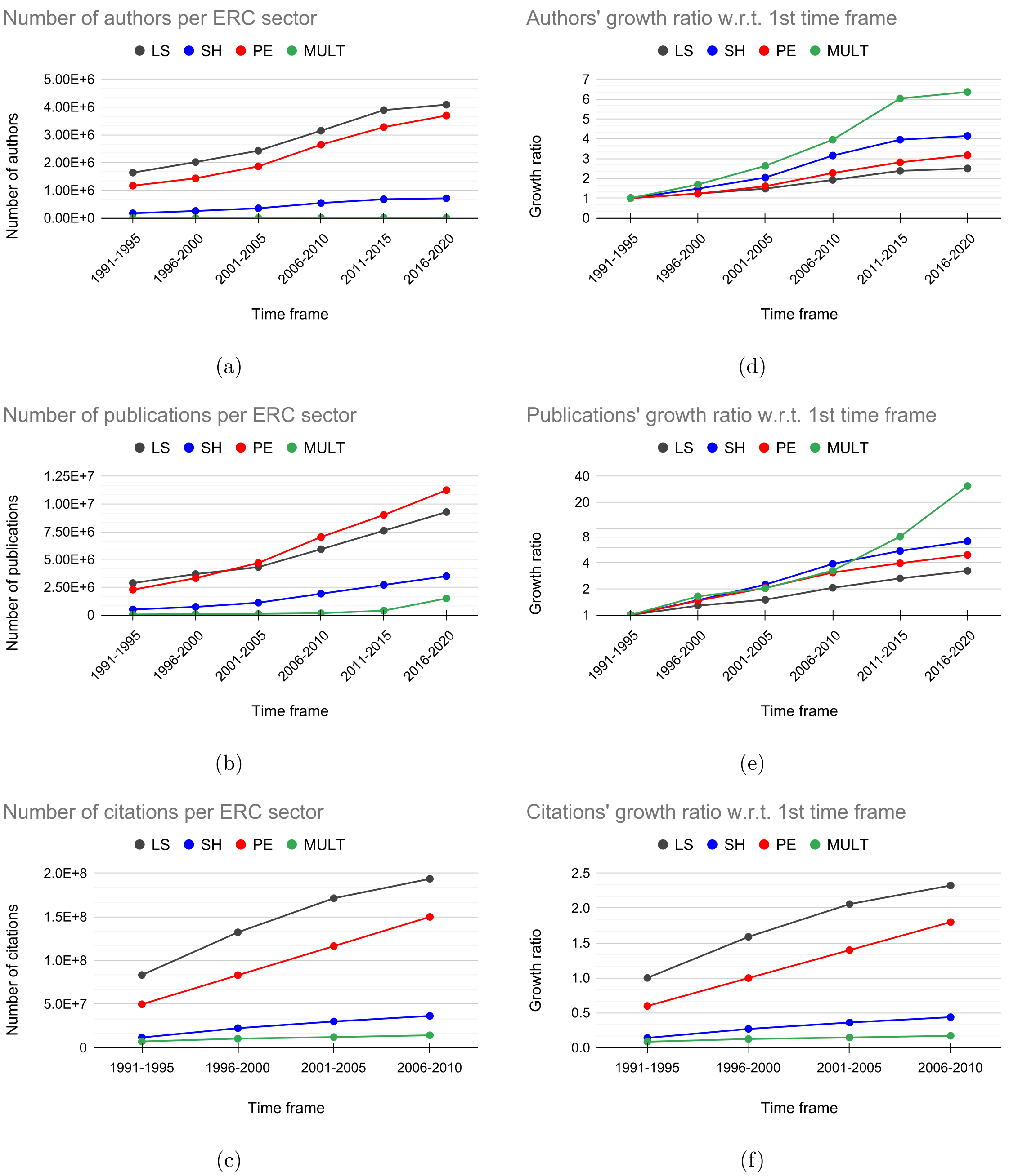

Fig. 2. Raw numbers and growth ratios of authors, publications and citations for ERC sectors and MULT.

The significant increase in MULT publications in the two most recent time frames shown in Fig. 2(e) also appears remarkable, and it is likely due to the emergence of many multidisciplinary journals in recent years.

Fig. 3 complements our analysis by reporting, in addition to raw numbers and growth ratios of Fig. 2, the world shares of authors, publications, and citations. This extended analysis reveals slightly decreasing world shares of publications for both LS and PE, as SH and MULT expand (while remaining smaller communities). Also, in the most recent time frame, PE appears to have finally managed to overtake LS in terms of world share of citations, albeit by a very narrow margin. When considering publications, PE overtook LS much earlier (2001-2005), while the number of authors remains smaller even in the most recent time frame.

Fig. 4 reports the temporal evolution of both author to publication ratios (a) and average citations per publication (b), for all ERC sectors plus MULT. Author to publication ratios show declining trends for all sectors, suggesting that the growth of the community in terms of authors is outpaced by the parallel increase in publication numbers. The reported author to publication ratios for MULT are significantly lower than those relative to the ERC sectors. This suggests that many authors contributing to multidisciplinary

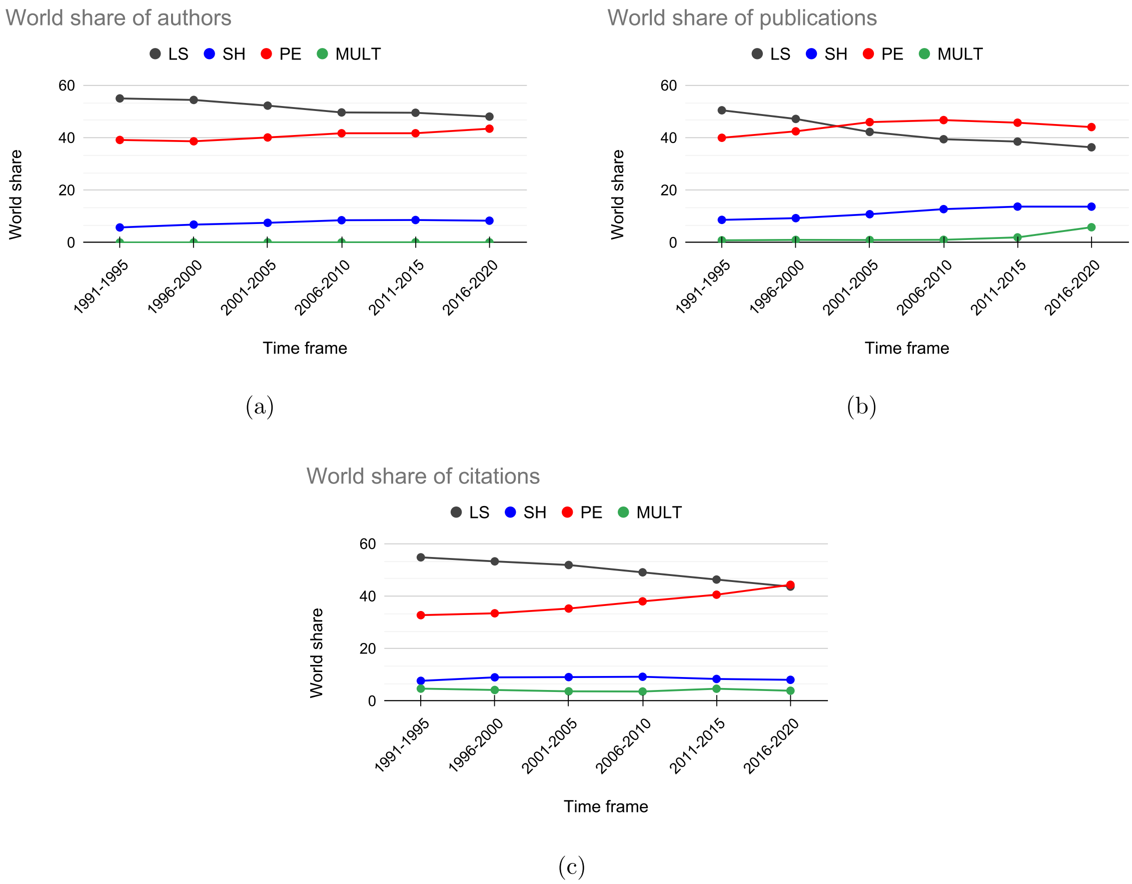

Fig. 3. World shares of authors, publications and citations for ERC sectors and MULT category.

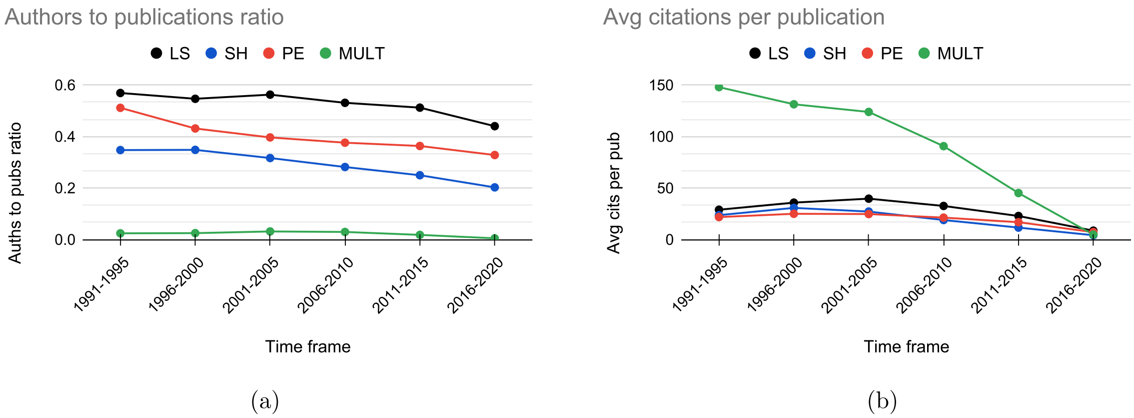

Fig. 4. Authors to publications ratio (a) and average citations per publication (b), for each ERC sector plus MULT.

publications are in fact primarily focused on other areas, and only in a minority of cases get classified as full MULT authors by our classification scheme (see Section 3.2 for details on how we classify authors and publications). Average citations per publication for individual ERC sectors show increasing trends in early time frames, followed by a declining phase. This decline may be partly attributed to the fact that more recent publications did not have enough time to accumulate citations (see Section 3.3 for a discussion concerning this issue), and partly to the consideration that even the increased number of references per article, often mentioned in the literature (see, e.g., Nicolaisen and Frandsen (2021)), may not be enough to compensate for the increased volume of publications in recent years. Average citations per publication decrease throughout the observation period for MULT. Besides the above observations, in this case an additional motivation may be found in the considerable increase in number of multidisciplinary journals, of significantly

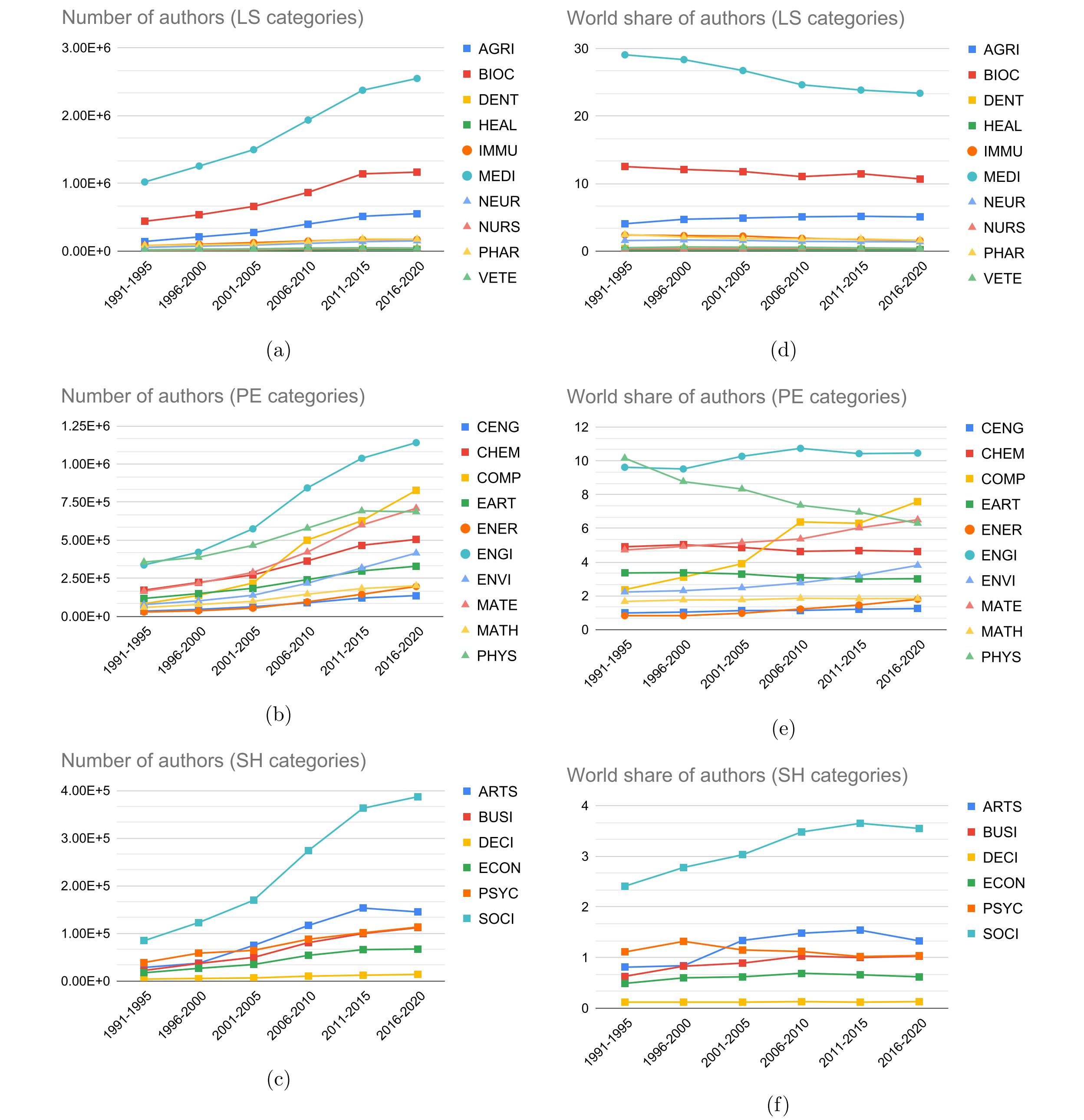

Fig. 5. Number of authors and their world share, for each research category.

varying impact factors, over the observation period, compared to the few, high impact, multidisciplinary publication venues present in the early 1990s.

4.3. Scientific areas gaining and losing traction

Fig. 5 shows, for each research category (excluding MULT, already analyzed in Section 4.2), the absolute number of authors and the respective world share, for each time frame. In the charts, research categories are grouped based on the ERC sector they fall into. However, we remark that world shares for any single time frame are computed over all categories.

The absolute numbers (subfigures (a), (b), and (c)) show increasing trends for almost all categories, confirming a growing scientific community in all disciplines. The only two exceptions are PHYS (sector PE) and ARTS (sector SH), for the most recent time frame. While LS and PE have not too distant numbers of authors, a closer look highlights some differences: the vast majority of LS authors seems confined to a limited number of categories (namely, MEDI, BIOC and AGRI), while the distribution of PE researchers is more

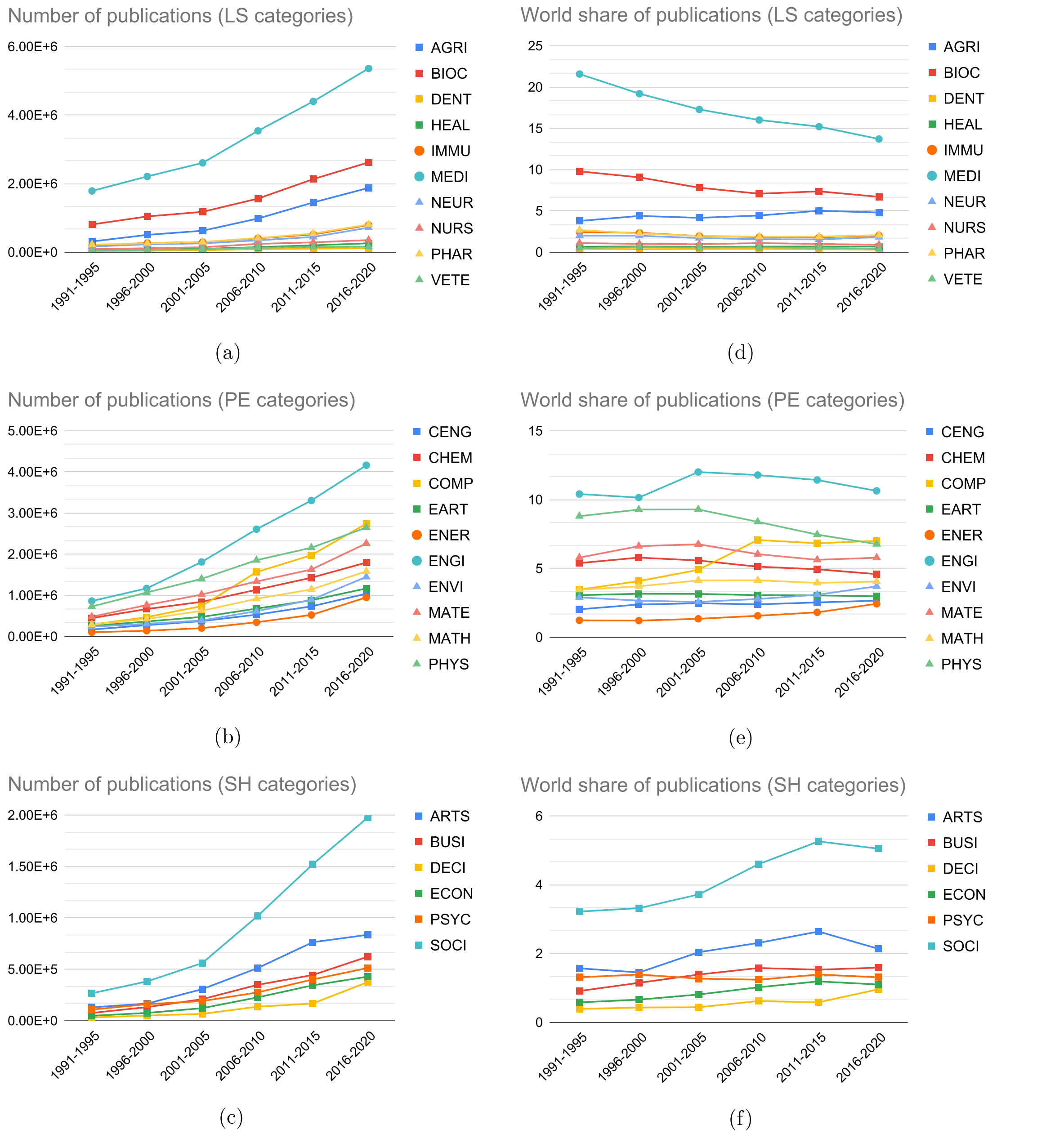

Fig. 6. Number of publications and their world share, for each research category.

balanced and all categories provide significant contributions to the overall author total. The SH sector is clearly dominated by the SOCI category.

To get an idea of trends in the evolution of research communities, we now move to world shares (subfigures (d), (e), and (f)). Declining trends can be observed for the two major LS categories (namely, MEDI and BIOC). In the PE sector, the growth of computer science from 2% to almost 8% in terms of world share is remarkable: overall COMP moved from the 6th to the 2nd position, within the PE sector, in just 30 years. Energy, as well as material and environmental sciences, are also gaining traction. The slow growth of PHYS, which moved from the 1st to the 4th position, is also notable. In the SH sector, SOCI, the dominant category, also shows an increasing world share trend up to the 2011-2015 time frame.

4.4. *Research areas and productivity*

Fig. 6 shows the number of publications, and respective world share, for each research category, with patterns very similar to those observed for authors: a focus on MEDI, BIOC, and AGRI in LS, a more distributed production in PE, and a dominance of SOCI

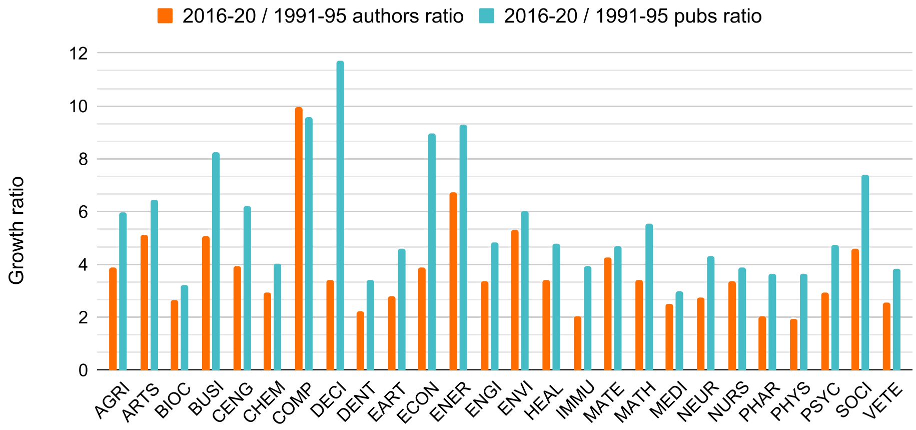

Fig. 7. Authors and publications growth ratios in each research area.

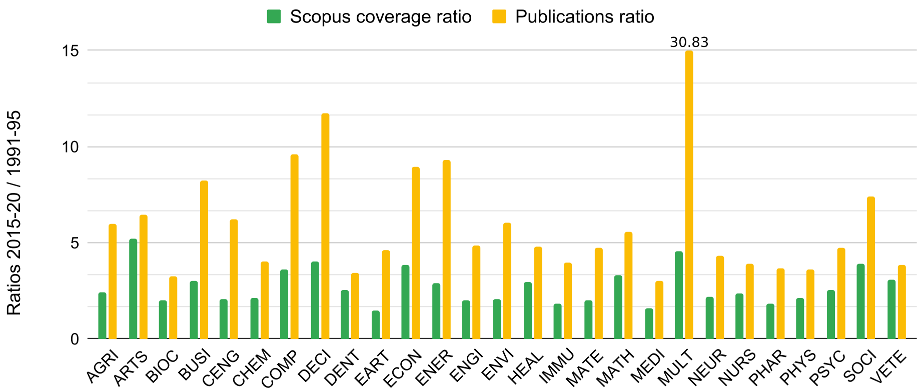

Fig. 8. Scopus coverage and publications ratios by subject category.

in SH. In terms of world share, MEDI, BIOC and PHYS are losing traction also in terms of scientific production. In contrast, COMP, ENVI, and ENER appear to be on the rise.

In order to better assess the significant growth of knowledge production over the 1991-2020 time period, Fig. 7 shows, for all categories, authors and publications ratios between the 2016-2020 and the 1991-1995 time frames. All ratios are above 1, indicating growth over the whole period of observation, in the realms of both authors and publications. COMP is the fastest growing in terms of authors, with an increase of almost a factor of 10 in thirty years. It is also second in terms of publications growth, surpassed only by DECI, which is however a much smaller community. We remark that COMP is the only category for which the growth factor of publications is smaller (albeit slightly) than that of authors. The second fastest growing community, in terms of authors, is ENER, which is also third in terms of publications growth.

Similarly to the analysis in Section 4.1 for global data, Fig. 8 reports, for each category, the growth ratios of indexed Scopus sources and publications between the most remote and the most recent time frames. This is because a second plausible cause for the observed growth may be improved Scopus coverage. In all cases, however, the publications growth ratio is higher than the Scopus coverage growth ratio, and often by a very large margin. This indicates that, although improved Scopus coverage is a contributing factor, it does not fully explain the observed increase in productivity.

Fig. 9 shows the number of citations, and corresponding world share, for each research category. Absolute values (subfigures (a), (b) and (c)) are displayed only up to the 2006-2010 time frame (see Section 3.3 for a discussion of this limitation), but still provide a sense of scale. World shares (subfigures (d), (e), and (f)) reveal patterns mostly in line with those highlighted in Section 4.3 and Section 4.4, when considering the dominant and fastest growing areas in each sector. However, the performance of individual categories sometimes differs from that observed for authors and publications. This is the case, for example, of PHYS, whose decline in terms of the world share of citations is not as marked as that for authors or publications. At the same time, the ascent of COMP appears not as steep as previously observed, while CHEM ranks surprisingly high in terms of citations. This is in line with well-known different citation practices in various areas.

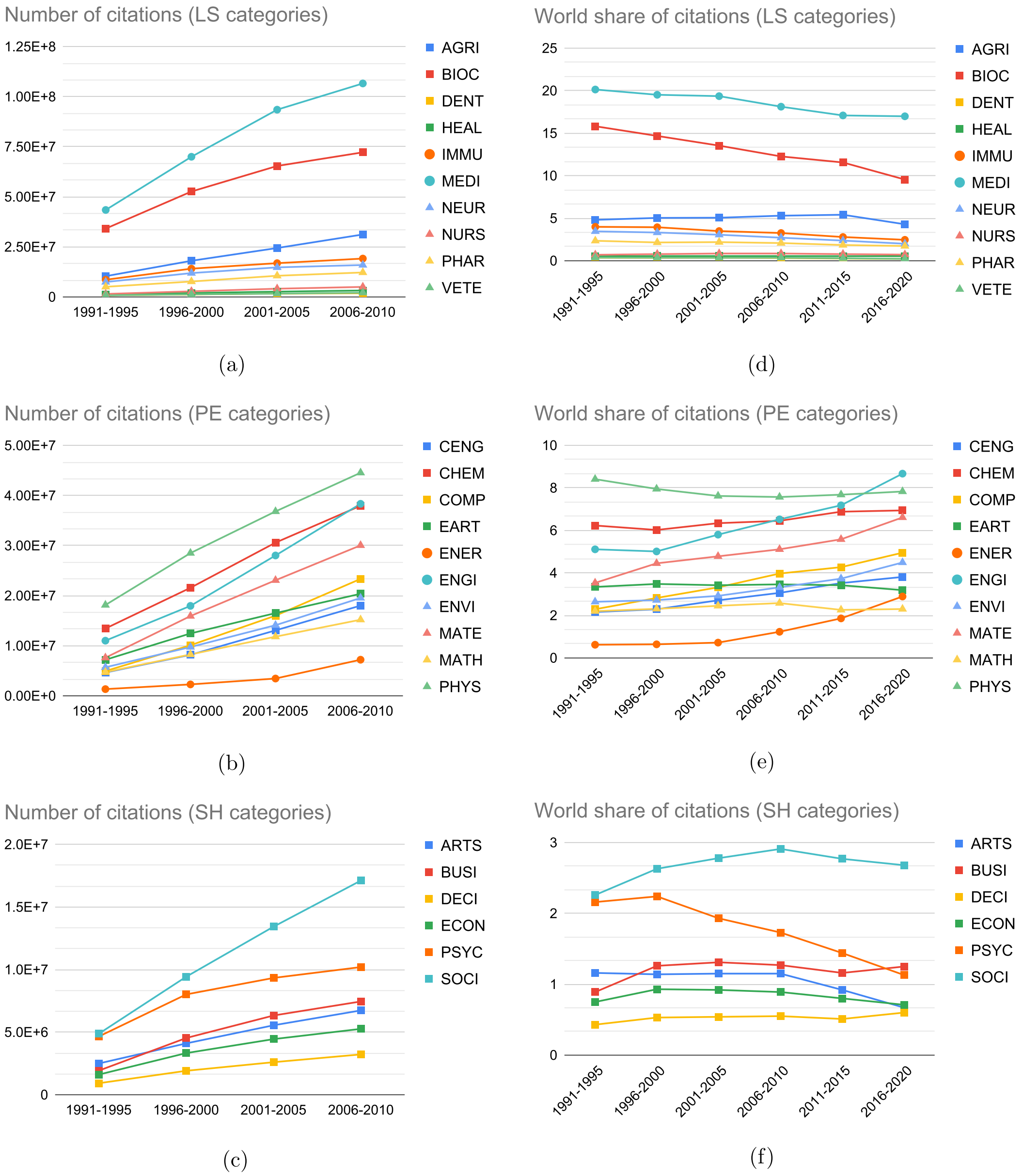

Fig. 9. Number of citations and their world share, for each research category.

5. A country-based analysis of scientific communities

In this section we analyze the size of the scientific community of the 19 member nations of the G20 group over time. We use the number of authors as the main indicator, as it appears to be more stable and reliable than the number of publications and citations among the various disciplines and countries. Besides a global view (Section 5.1), in Section 5.2 and Section 5.3 we examine country participation in ERC sectors and in specific research areas, respectively.

5.1. Scientific community size over time

Fig. 10 shows, for each G20 country, the size over time of its scientific community with respect to million inhabitants (subfigure (a)) and to gross domestic product (GDP) expressed in billions of 2017 International dollars (subfigure (b)). Population and GDP data, extracted from The World Bank,[2] refer to the most recent year of each time frame.

[2] www.worldbank.org.

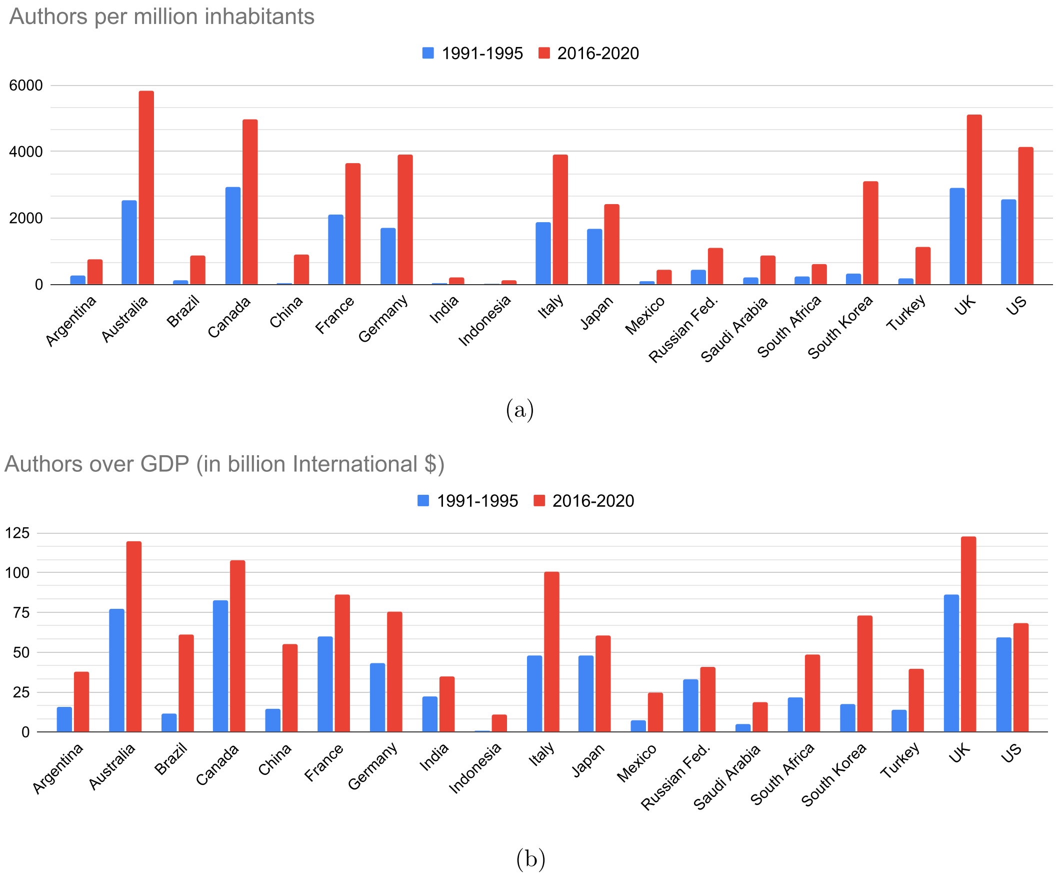

Fig. 10. Temporal evolution of the number of authors, for each G20 country, (a) over million inhabitants and (b) over GDP (expressed in billions of 2017 International dollars).

When considering the number of authors per million inhabitants, a general increasing trend can be observed for all countries. North American and European countries, together with Australia, South Korea and Japan exhibit the highest values, with Australia firmly in first position. We can observe a similar scenario when considering numbers of authors over GDP. The highest values, in the most recent time frame, are those of Australia and UK, with Canada and Italy in third and fourth positions, respectively. Overall, the data convey a rather positive message: the scientific community is expanding, with increasing efforts to establish research communities in all countries.

Table 3 reports, for each country, the average, standard deviation, and relative standard deviation of the world share of authors in all research categories, for the two time frames of interest. The latter metric, defined as the ratio of the standard deviation to the mean, better shows the extent of variability in relation to the mean itself. The percentage variation of the average between the two time frames is also shown. Countries that were already established in 1991-1995, such as Canada, France, Japan, the UK and the US, show decreases in their averages (with Germany as an exception). In contrast, emerging countries, such as Brazil, China, and India show major increases. It is also interesting to note that emerging countries in general exhibit much larger relative standard variations compared to more established nations. This suggests a more uneven development, or possibly an intentionally designed catch-up strategy that prioritizes specific research disciplines over others.

5.2. Country involvement in ERC sectors

Fig. 11 complements our analysis by showing, for each ERC sector and G20 country, its world share of authors, i.e., the percentage of authors, over all authors of that sector in the entire world, that have affiliation in that country. Analogous investigations for publications and citations reveal similar trends and the corresponding charts are omitted. We consider the first and last time frames.

In all sectors, the US was by far the dominant power in the earliest time frame. While this preeminence has constantly been eroded over time, the US is still, and by quite a margin, the leading nation in both the LS and SH sectors in 2016–2020. However, in PE, China has now overtaken the US. Data not reported in the chart show that this happened since 2011-2015, and that, since the same time period, this is also true in terms of world share of publications (China's world share of citations has also overtaken that of the US, but only more recently).

Table 3
Average world share of authors, standard deviation and percentage relative standard deviation, computed over all categories, for each G20 country in the 1991-1995 and 2016-2020 time frames. The last column reports the percentage variation between the 2016-20 and the 1991-95 averages.

Country	Avg. 1991-95	Std. dev. 1991-95	%RSD 1991-95	Avg. 2016-20	Std. dev. 2016-20	%RSD 2016-20	% var. btw avgs.
Argentina	0.27	0.25	93%	0.43	0.34	79%	59%
Australia	1.94	1.01	52%	2.26	1.28	57%	16%
Brazil	0.58	0.39	67%	3.03	2.89	95%	422%
Canada	3.52	1.33	38%	2.49	0.96	39%	-29%
China	1.13	1.20	106%	13.53	9.85	73%	1097%
France	3.29	1.90	58%	2.94	0.95	32%	-11%
Germany	3.98	1.89	47%	4.13	1.16	28%	4%
India	1.46	0.97	66%	3.61	2.38	66%	147%
Indonesia	0.03	0.04	133%	0.50	0.39	78%	1567%
Italy	2.33	1.48	64%	2.77	0.96	35%	19%
Japan	5.37	3.99	74%	2.99	1.52	51%	-44%
Mexico	0.25	0.19	76%	0.64	0.34	53%	156%
Russian Fed.	1.84	2.43	132%	2.02	1.42	70%	10%
Saudi Arabia	0.13	0.13	100%	0.42	0.34	81%	223%
South Africa	0.35	0.30	86%	0.52	0.29	56%	49%
South Korea	0.47	0.44	94%	2.03	0.99	49%	332%
Turkey	0.28	0.25	89%	1.08	0.87	81%	286%
UK	6.67	2.57	39%	4.88	2.07	42%	-27%
US	30.12	9.56	32%	18.67	6.52	35%	-38%

China is on the rise in the LS sector, too, and other growing nations in both LS and PE include Brazil, India, and South Korea. Knowledge production in the SH sector seems mostly concentrated in the US and in the UK, with fast growing nations in LS or PE, such as China, India, or South Korea, lagging behind in this respect.

5.3. Country involvement in research areas

A more detailed analysis is reported in the heat maps of Fig. 12, showing as a percentage the world share of authors of each research category for all G20 countries. Darker squares denote higher percentages.

Subfigure (a) refers to 1991-1995, when the US and the UK, albeit at different scales, were the only two nations to have relevant presence in all disciplines. They were followed by Canada, with a few weaker areas, and then by Japan and some major European countries (Germany, France, Italy). The situation significantly changes in 2016-2020 (subfigure (b)), where the surprising development of China, mostly at the expense of the UK and the US, is evident. This development, however, is somewhat uneven, with some disciplines left way behind others. Other countries, such as Australia, Brazil, India, Russian Federation, and South Korea, all have significantly broadened their coverage of disciplines. In general, a comparison of subfigures (a) and (b) reveals a much more widespread knowledge production, at least for the G20 group, in 2016-2020 than in 1991-1995.

We remark that US preeminence in science, its erosion, and the extremely rapid rise of China has been the primary focus of other previous studies. For instance, Leydesdorff and Wagner (2009) derive data from WoS and investigate the time period 1991-2006, concluding that the US, although losing ground to China, other Asian countries and European countries, is still the leading nation in science during their observation period. Similarly, King (2004), using data from the ISI database over the time period 1993-2002, observes the US to be still in top position as far as knowledge production goes, followed by the UK, Germany and Japan. Our Scopus-based analysis is consistent with those results, in spite of the different data sources, and shows the trend continuation into more recent years, with China finally overtaking the US, at least for physical sciences and engineering.

6. Gender dynamics

As already observed in Section 4.1, both male and female authors are increasing. In this section we investigate the changes in women participation across time: we focus on ERC sectors and research areas in Section 6.1 and provide a country-based analysis in Section 6.2.

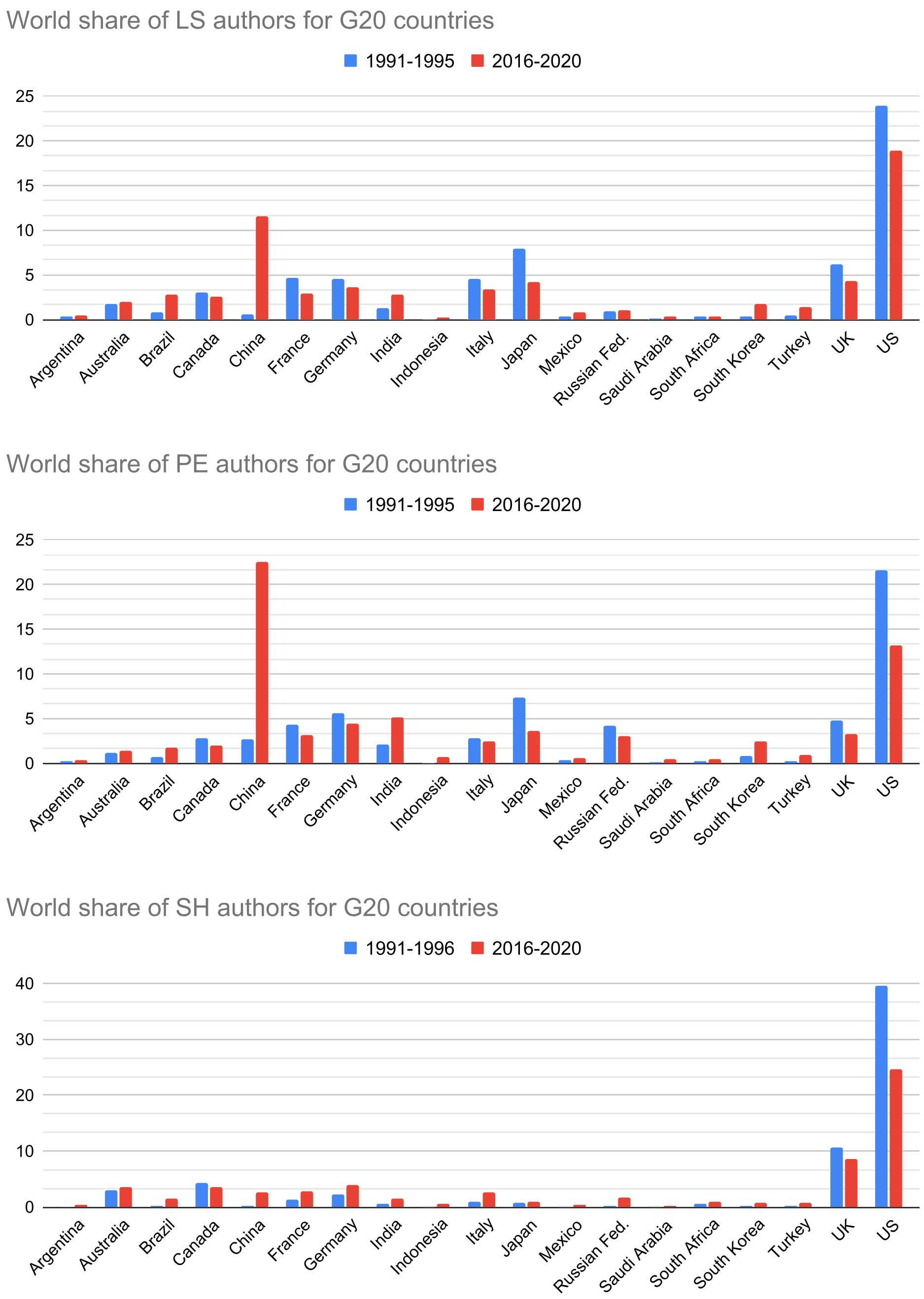

Fig. 11. Temporal evolution of the world share of authors in each ERC sector in G20 countries.

6.1. *Women participation across ERC sectors and disciplines*

Fig. 13 reports the female-to-male ratio, focusing on authors to which a gender could be attributed (see Section 3.3). Considering the global trend, the ratio increases from 0.18 in 1991-1995 to 0.31 in 2016-2020, indicating a narrowing gender gap. Data plotted in Fig. 13 also show that the female-to-male ratio increases for all ERC sectors. In spite of the increasing female representation, the trends seem to suggest that the process is slowing down in recent years. It should be noted that both LS and SH exhibit ratios that are higher than those concerning global scientific production, while PE displays the lowest ratios in the chart. Also, all values are far below 1, highlighting a significant underrepresentation of female researchers in all sectors.

Fig. 14(a) reports absolute numbers of authors per gender and ERC sectors over time. LS and PE are dominated by male researchers followed, from a long distance, by LS female researchers. SH male and PE female author populations are roughly the same size, with

	AGRI	ARTS	BIOC	BUSI	CENG	CHEM	COMP	DECI	DENT	EART	ECON	ENER	ENGI	ENVI	HEAL	IMMU	MATE	MATH	MEDI	MULT	NEUR	NURS	PHAR	PHYS	PSYC	SOCI	VETE
Argentina	0.8	0.1	0.5	0.0	0.7	0.5	0.1	0.1	0.1	0.4	0.1	0.2	0.1	0.3	0.0	0.8	0.3	0.3	0.3	0.1	0.5	0.0	0.3	0.3	0.1	0.1	0.6
Australia	4.3	2.6	1.7	2.7	1.0	1.3	1.5	1.2	1.6	2.3	3.0	1.1	0.8	2.3	3.3	2.4	0.9	1.7	1.6	0.8	1.6	1.5	0.8	0.8	2.3	3.6	3.9
Brazil	1.0	0.1	0.7	0.1	0.6	0.7	0.5	0.6	1.1	0.6	0.2	0.5	0.4	0.5	0.2	1.6	0.6	0.9	0.8	0.0	1.1	0.1	0.9	1.1	0.1	0.2	0.7
Canada	5.1	3.8	3.7	3.4	3.3	2.7	3.7	6.7	1.9	3.8	4.3	3.0	2.6	4.4	5.3	3.3	1.8	4.0	2.6	0.3	4.6	3.0	2.2	2.4	5.7	3.7	4.0
China	0.6	0.1	0.6	0.4	3.2	2.5	1.7	0.8	0.2	1.7	0.1	2.1	3.0	0.9	0.1	0.3	4.1	2.5	0.5	0.3	0.3	0.0	1.3	3.0	0.1	0.1	0.1
France	4.5	1.3	5.9	0.8	4.3	5.6	4.7	1.8	0.7	4.3	1.8	2.0	3.1	2.8	2.6	5.9	5.3	5.1	4.6	1.3	5.4	0.1	4.7	5.6	1.7	0.9	2.1
Germany	4.0	4.1	5.3	1.3	4.5	7.6	4.6	3.4	1.3	4.1	2.1	3.2	3.9	3.9	4.1	5.0	6.9	6.5	4.4	1.5	5.1	0.2	4.8	7.6	2.6	1.6	4.0
India	3.0	0.3	1.4	1.0	2.3	3.1	1.2	2.1	0.3	2.0	0.5	2.3	1.4	2.4	0.2	1.4	3.2	1.8	1.0	1.1	0.3	0.1	2.5	2.0	0.2	0.6	1.9
Indonesia	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
Italy	3.0	0.9	3.9	0.7	2.3	3.4	2.9	2.3	1.2	2.5	1.6	1.8	1.9	2.2	0.8	3.5	2.0	3.8	5.6	0.2	5.0	0.1	4.3	3.8	1.3	0.6	1.1
Japan	5.4	0.3	9.3	0.6	7.5	9.3	6.8	1.8	5.7	3.3	1.1	9.3	9.9	2.5	2.9	6.5	10.6	3.9	8.2	1.1	8.0	0.7	15.8	7.8	1.1	0.6	5.2
Mexico	0.8	0.1	0.3	0.1	0.3	0.3	0.1	0.1	0.1	0.3	0.1	0.3	0.2	0.5	0.0	0.5	0.3	0.3	0.4	0.0	0.5	0.0	0.4	0.4	0.1	0.1	0.3
Russian Fed.	0.5	0.2	1.6	0.1	7.2	4.3	1.0	0.3	0.0	2.4	0.1	8.6	3.2	1.0	0.1	1.4	6.2	3.1	0.6	0.3	0.9	0.0	1.1	5.3	0.1	0.2	0.0
Saudi Arabia	0.2	0.0	0.0	0.1	0.2	0.1	0.1	0.4	0.3	0.1	0.0	0.6	0.1	0.2	0.1	0.1	0.1	0.2	0.2	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.2
South Africa	1.1	0.7	0.2	0.2	0.2	0.3	0.2	0.2	0.4	0.6	0.4	0.1	0.2	0.6	0.2	0.3	0.2	0.4	0.3	0.1	0.0	0.0	0.2	0.2	0.3	0.7	1.1
South Korea	0.2	0.0	0.4	0.3	1.2	0.8	1.2	1.5	0.1	0.2	0.3	0.6	1.0	0.2	0.2	0.4	1.3	0.6	0.3	0.0	0.1	0.1	0.8	0.8	0.0	0.1	0.1
Turkey	0.2	0.1	0.1	0.1	0.4	0.4	0.3	1.0	0.8	0.3	0.1	0.5	0.3	0.4	0.1	0.1	0.3	0.3	0.6	0.0	0.2	0.0	0.6	0.2	0.1	0.1	0.2
UK	6.8	14.7	6.4	11.6	3.8	5.6	5.4	6.3	5.8	5.4	7.8	3.6	4.4	5.8	8.2	7.9	4.6	4.7	6.1	8.0	5.8	4.7	6.4	4.5	6.6	11.9	7.6
US	24.8	31.9	30.2	41.9	20.1	20.1	30.5	39.6	25.6	19.8	39.6	20.5	26.4	26.0	42.5	27.0	16.9	24.7	21.9	41.1	33.3	48.4	24.1	19.3	51.6	37.8	27.7

(a)

	AGRI	ARTS	BIOC	BUSI	CENG	CHEM	COMP	DECI	DENT	EART	ECON	ENER	ENGI	ENVI	HEAL	IMMU	MATE	MATH	MEDI	MULT	NEUR	NURS	PHAR	PHYS	PSYC	SOCI	VETE
Argentina	1.3	0.6	0.5	0.1	0.6	0.4	0.2	0.1	0.1	0.7	0.3	0.2	0.1	0.5	0.3	1.1	0.2	0.4	0.4	0.1	0.6	0.0	0.4	0.4	0.3	0.4	1.3
Australia	2.6	2.5	1.6	3.6	1.3	1.2	1.4	1.7	1.3	2.4	3.3	1.5	1.2	2.5	6.1	2.1	1.2	1.4	2.1	1.1	2.3	5.0	1.0	1.0	3.3	4.0	2.1
Brazil	6.9	0.7	2.0	2.3	2.1	2.1	1.6	3.0	12.6	1.8	1.0	1.7	1.4	2.7	3.4	3.9	1.4	1.9	2.1	0.9	2.5	5.2	3.3	1.4	1.2	1.5	11.2
Canada	2.5	3.1	2.3	2.6	2.0	1.6	1.9	3.8	1.3	2.6	2.8	2.0	1.8	2.7	4.0	2.2	1.4	2.2	2.6	1.2	4.1	3.3	1.1	1.6	4.8	3.6	2.3
China	15.2	0.8	20.1	4.5	27.6	25.9	20.0	11.1	2.9	18.0	4.2	28.6	26.1	24.1	0.8	13.1	31.0	17.0	9.0	9.4	9.2	1.7	21.2	16.7	2.2	2.4	2.8
France	2.9	3.8	3.2	2.7	2.6	3.2	2.9	3.9	1.3	4.3	4.1	2.0	2.7	2.5	1.7	4.1	3.0	4.9	3.0	1.7	3.5	1.6	2.1	4.5	2.7	2.4	2.3
Germany	3.0	4.3	4.4	3.7	4.2	4.8	4.0	4.5	3.6	4.8	5.4	3.6	4.1	3.1	3.0	4.4	4.7	5.2	3.4	5.0	6.6	1.3	2.4	6.6	4.7	3.4	3.1
India	4.4	0.4	3.0	3.3	4.0	5.5	8.4	4.0	7.7	2.4	1.9	4.0	5.1	3.6	0.6	2.9	5.0	3.9	2.1	3.1	0.8	0.3	8.5	3.8	0.4	1.4	7.2
Indonesia	0.7	0.2	0.1	1.6	0.3	0.1	1.2	0.7	0.7	0.9	0.7	0.4	0.7	1.2	0.0	0.2	0.3	0.2	0.2	0.1	0.0	0.7	0.7	0.6	0.0	0.6	0.5
Italy	2.8	3.4	2.9	3.0	1.7	2.4	2.0	3.8	4.3	3.3	4.0	2.2	2.2	2.4	1.5	2.5	1.5	3.3	3.9	1.3	4.1	0.9	2.6	3.7	2.8	2.2	4.2
Japan	2.8	0.5	4.6	0.7	3.0	4.0	3.3	1.3	6.3	2.8	1.8	2.6	4.1	1.7	3.4	3.3	4.1	2.5	4.6	3.7	4.2	1.2	4.8	5.1	1.0	0.8	3.0
Mexico	1.8	0.3	0.6	0.3	0.9	0.8	0.5	0.5	0.2	0.6	0.3	0.6	0.6	1.0	0.2	1.2	0.7	0.9	0.6	0.2	0.7	0.4	0.7	0.8	0.4	0.5	1.0
Russian Fed.	1.2	2.6	1.3	2.2	2.6	3.2	1.7	2.1	0.1	5.2	2.3	3.4	2.6	1.6	2.3	1.3	3.4	4.0	0.9	1.0	0.6	0.3	1.1	5.5	0.5	1.9	0.1
Saudi Arabia	0.3	0.1	0.3	0.4	0.7	0.6	0.6	0.4	1.7	0.4	0.3	0.9	0.4	0.4	0.2	0.3	0.5	0.8	0.3	0.3	0.1	0.2	0.9	0.3	0.0	0.1	0.3
South Africa	0.8	1.0	0.2	1.0	0.4	0.3	0.3	0.6	0.1	0.5	0.8	0.4	0.2	0.7	0.8	0.5	0.2	0.4	0.4	0.3	0.1	0.5	0.4	0.7	0.5	1.3	0.7
South Korea	1.6	0.3	2.4	1.5	4.2	2.9	2.4	1.2	2.3	1.0	1.3	2.9	3.1	1.6	2.6	2.5	4.0	1.3	1.6	1.5	1.4	3.0	3.0	2.4	0.4	0.8	1.8
Turkey	1.2	0.4	0.5	0.9	0.7	0.9	1.0	1.7	4.1	0.7	1.0	1.1	0.9	1.1	0.6	0.3	0.9	1.4	1.7	0.1	0.4	1.2	1.0	0.8	0.5	1.0	3.4
UK	2.9	10.6	4.1	7.3	2.8	2.9	3.1	4.8	4.5	4.7	7.6	2.9	2.9	3.3	6.2	4.6	2.6	3.7	4.7	6.0	5.4	6.1	2.9	3.8	7.3	9.0	5.2
US	13.2	21.2	20.9	19.5	13.2	11.6	12.7	17.2	13.0	16.9	23.0	12.0	13.0	14.1	27.1	21.5	11.2	14.2	19.1	28.8	26.3	30.4	15.6	14.2	34.4	25.2	14.7

(b)

Fig. 12. World share of authors of each research category for all G20 countries: (a) 1991-1995; (b) 2016-2020.

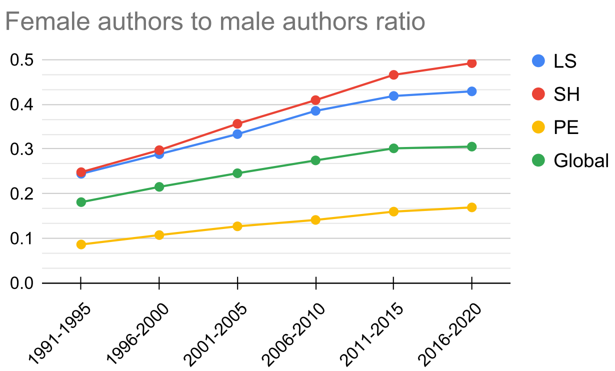

Fig. 13. Temporal evolution of female to male author ratios, for global scientific production and separately for each ERC sector.

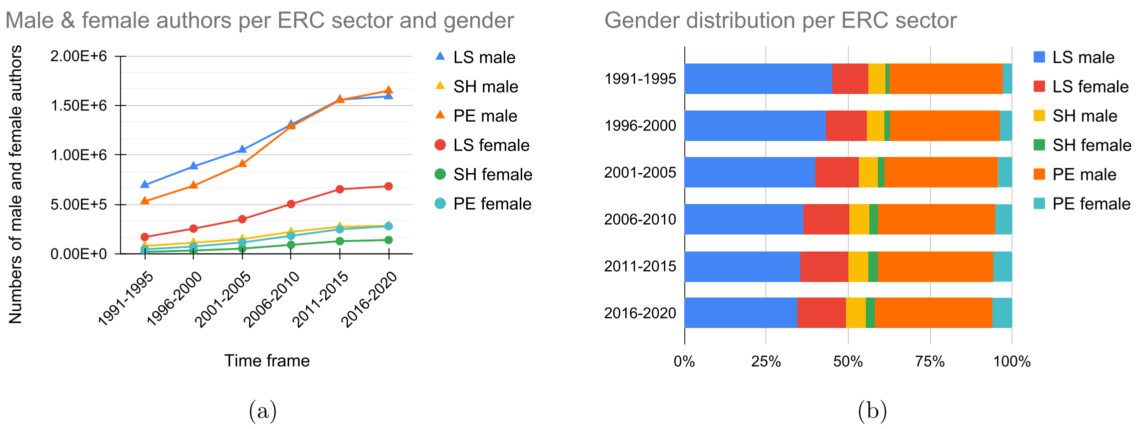

Fig. 14. Gender dynamics in the three ERC sectors.

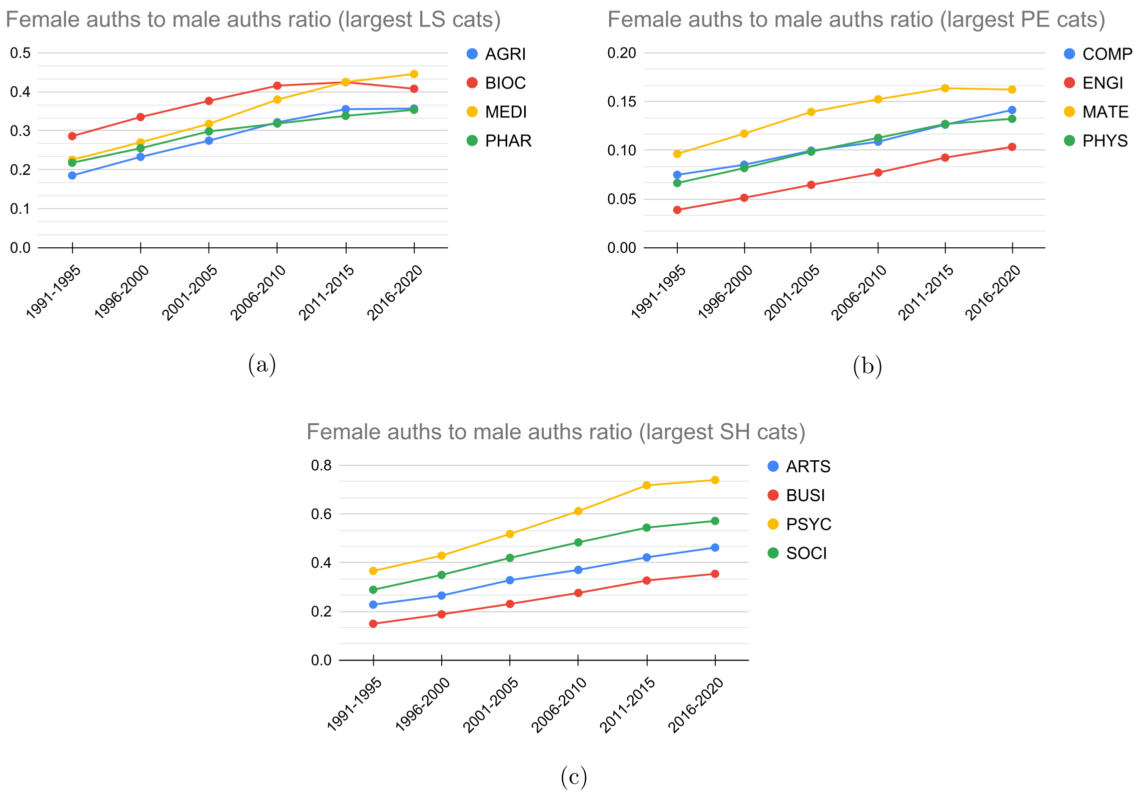

Fig. 15. Female to male author ratios for the largest categories in each ERC sector.

SH female authors being the smallest group. Moving to percentages, the relative female representation increases with time in all sectors as can be seen in Fig. 14(b), which indicates the proportion of authors by gender relative to the entire population.

Fig. 15 focuses on the four most popular disciplines in each ERC sector, showing a significant gap for all categories considered. The gap is greatest for PE (0.16 female to male ratio for MATE in the most recent time frame) and smaller for LS and SH categories (0.45 ratio for MEDI and 0.74 for PSYC in 2016–2020, respectively). It appears to be narrowing, albeit slowly, with some exceptions for the most recent time frame (e.g., BIOC and MATE).

Fig. 16 reports a full picture of gender trends by discipline, considering only the first and last decades. Female representation increases in all areas, with the only exception of NURS, where female researchers decrease from 73.33% in 1991-1995 to 66% in 2016-2020. This is however the only category in which female researchers outnumber their male colleagues.

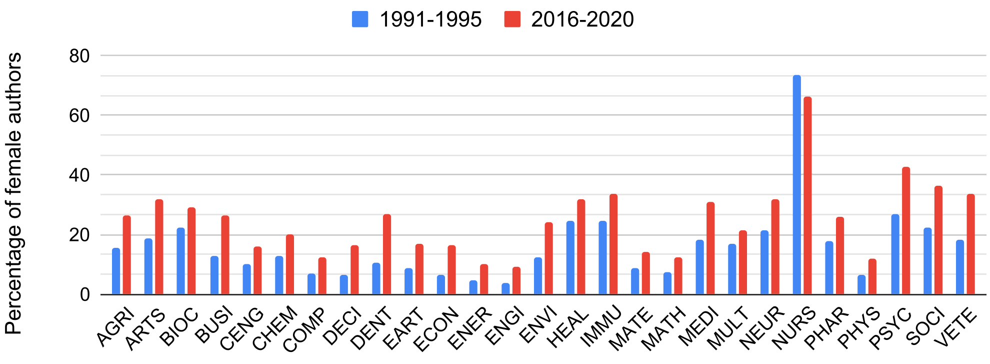

Fig. 16. Percentage variation of female representation across research categories.

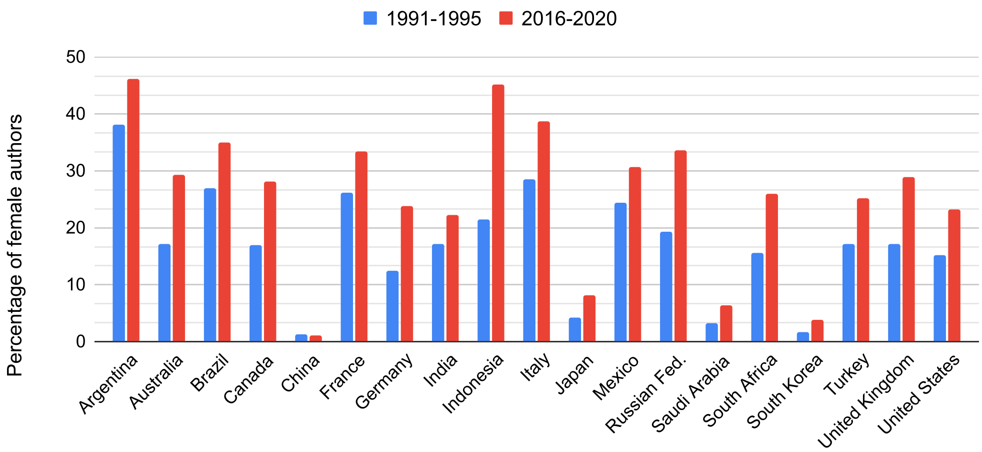

Fig. 17. Female research participation across G20 nations.

6.2. Gender trends per country

Women participation in the scientific community is not uniform across countries, which exhibit large differences. Fig. 17 shows the temporal evolution of the percentage of female authors (over male plus female authors) for each nation in the G20 group. The percentages are in all cases well below 50%, although generally increasing with time. A negative record of China, and particularly low values for Japan, Saudi Arabia, and South Korea seem also noteworthy. We remark, however, that the number of authors with unknown gender in some of these countries is very high, as shown in Table 2.

7. Concluding remarks

In this article we have presented a large-scale analysis of the temporal evolution of scientific production across three different dimensions: temporal, geographical, and disciplinary. The breadth and coverage of our analysis make our study far more comprehensive than other works. By incorporating data across multiple geographical regions and disciplines, over a time period of thirty years, the study ensures a high level of representativeness and reliability. The analysis and different levels of detail allow for nuanced insights into a variety of aspects of the scientific population and productivity. In this section we first discuss our main findings, and then we present directions for further research.

7.1. Summary and discussion

Our main conclusions can be summarized as follows.

Growth in research production Research production generally increases with time, in terms of authors, publications and citations. This is true both on a global scale and at the level of ERC sectors.

At ERC sector level, Life sciences (LS) started as the dominant sector, but it has been constantly losing ground to Physical sciences and engineering (PE): PE has overtaken LS in terms of world share of publications already in 2001-2005, and in terms of citations in 2016-2020, with the gap between number of authors constantly reducing. These phenomena bear a quantitative witness to the increasing relevance of technology, and in turn of STEM disciplines, as we get closer to the present. Social sciences and humanities (SH), while exhibiting a higher growth ratio than LS and PE, still remains by far the smallest of the three sectors.

When zooming in to individual research categories, the bulk of the LS sector appears to be concentrated into three major categories: AGRI, BIOC, and MEDI. PE's output, on the other hand, seems more evenly spread between its different subdisciplines, the major areas being ENGI – the largest by all measures considered – followed by COMP, which is the fastest growing one. COMP managed to move from sixth position in 1991-1995 to second in 2016-2020 in terms of world share of authors, and from fifth to second in terms of world share of publications. PHYS, while losing ground in recent time frames, also remains a major player in this sector. Social sciences are largely dominated by SOCI, according to all indicators.

Multidisciplinarity seems more and more important. In this area, we found the highest growth ratios, a phenomenon also closely tied to the recent appearance of several journals devoted to multidisciplinary works.

Geographic trends Research production generally increases with time, in terms of authors, publications and citations also for the 19 countries in the G20 group, which we have examined more closely. For the above countries, research efforts relative to population and gross domestic product also appear to be on the rise, demonstrating global development and the ever-increasing relevance of innovation in modern societies.

The US started out, in the early time frames of our observation period, as the most relevant country in all sectors. This primacy, however, has constantly been eroded with the passage of time. China, on the other hand, rising from very low values for all its indicators, has proven to be the fastest growing emerging country, managing to climb, in recent years, to the second position in LS (right behind the US), and to first position in PE, overtaking the US. These advances have not been paralleled by an analogous development of SH, suggesting targeted research policies.

Gender representation Female participation tends to increase with time, albeit with a few exceptions and some negative records. The gender gap appears to be closing at a slower pace for PE disciplines, compared to LS and SH. Moreover, the increase appears to slow down in the most recent time frame (2016-2020), and the gap between genders remains quite large, with overall female representation at 23.39%. All engineering disciplines (notably CENG, COMP, ENER, ENGI, and MATE) show particularly low values. Our gender analysis also reveals general increasing trends for all countries, with occasional slightly declining phases for some nations. A negative record is observed for China and particularly low values for Japan, Saudi Arabia, and South Korea.

Our findings indicate a steep increase in knowledge production over the 1991-2020 time period, an erosion of US predominance in all fields in favor of emerging nations, and an increased female participation, although this trend seems to be slowing down in recent years. We believe these findings may be useful to research policy makers in established and emerging countries alike. We are also convinced that the above aspects should be continuously monitored, so that research policies, as well as research assessment practices, can be properly adjusted to the current situation and desired objectives.

7.2. Future research

While our work helps shed light on some interesting aspects of scientific research and its temporal evolution, it also leaves open several interesting research avenues. In particular, we have analyzed knowledge production at various levels of aggregations (e.g., ERC sectors, Scopus Research categories), but we have not delved into the interplay between these different domains. It is natural to assume that results in certain fields provide tools that enable research in others. We believe it would be quite interesting to study what the strongest connections are (for example, in terms of citations), and their relative intensities.

As far as gender issues are concerned, previous research efforts, differently from this work, often quantify female representation in terms of authorships rather than active authors. When active authors are counted, for example in Huang et al. (2020) (who, notably, use data derived from WoS rather than Scopus), the reported results are mostly in line with ours, both country wise and discipline wise. Several articles focus specifically on aspects such as the productivity gap and/or the research impact gap (measured mostly in terms of citations) between male and female authors (see, e.g., (Larivière et al., 2013; Bendels et al., 2018; Huang et al., 2020)). While this kind of analysis lies outside the scope of the present work, we plan to tackle these issues in future research that will focus specifically on gender while maintaining the same broad perspective as this article in terms of time periods, disciplines and countries investigated.

We also believe that studying the mobility patterns of researchers in the various scientific fields, and their temporal evolution, would add an interesting new dimension to our comparative analysis. In fact, the majority view in the existing literature is that geographical mobility of researchers has a relevant effect on scientific productivity and impact (see, e.g., (Franzoni et al., 2014; Halevi et al., 2016; Netz et al., 2020; Finocchi et al., 2023)). Last but not least, applying predictive analytics techniques to research productivity, in the spirit of Tian et al. (2023), would be very interesting and could give valuable strategic planning insights for academic institutions, funding agencies, and policymakers.

CRediT authorship contribution statement

Irene Finocchi: Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Andrea Ribichini:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Schaerf:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

References

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a_00019.

Bendels, M. H. K., Müller, R., Brueggmann, D., & Groneberg, D. A. (2018). Gender disparities in high-quality research revealed by Nature Index journals. *PLoS ONE*, *13*(1), 1–21. https://doi.org/10.1371/journal.pone.0189136.

Bongioanni, I., Daraio, C., & Ruocco, G. (2014). A quantitative measure to compare the disciplinary profiles of research systems and their evolution over time. *Journal of Informetrics*, *8*(3), 710–727. https://doi.org/10.1016/j.joi.2014.06.006.

Clark, A. D., Myers, T. C., Steury, T. D., Krzton, A., Yanes, J., Barber, A., Barry, J., Barua, S., Eaton, K., Gosavi, D., Nance, R., Pervaiz, Z., Ugochukwu, C., Hartman, P., & Stevison, L. S. (2024). Does it pay to pay? A comparison of the benefits of open-access publishing across various sub-fields in biology. *PeerJ*, *12*. https://doi.org/10.7717/peerj.16824.

Courtioux, P., Métivier, F., & Reberioux, A. (2019). Scientific competition between countries: Did China get what it paid for? https://halshs.archives-ouvertes.fr/halshs-02307534. Documents de travail du Centre d'Economie de la Sorbonne 2019.13 - ISSN: 1955-611X.

Das, J., Do, Q.-T., Shaines, K., & Srikant, S. (2013). U.S. and them: The geography of academic research. *Journal of Development Economics*, *105*, 112–130. https://doi.org/10.1016/j.jdeveco.2013.07.010.

Demetrescu, C., Finocchi, I., Ribichini, A., & Schaerf, M. (2020). On bibliometrics in academic promotions: A case study in computer science and engineering in Italy. *Scientometrics*, *124*, 2207–2228. https://doi.org/10.1007/s11192-020-03548-9.

Demetrescu, C., Finocchi, I., Ribichini, A., & Schaerf, M. (2022). On computer science research and its temporal evolution. *Scientometrics*, *127*(8), 4913–4938. https://doi.org/10.1007/s11192-022-04445-z.

Demetrescu, C., Lupia, F., Mendicelli, A., Ribichini, A., Scarcello, F., & Schaerf, M. (2019). On the Shapley value and its application to the Italian VQR research assessment exercise. *Journal of Informetrics*, *13*(1), 87–104. https://doi.org/10.1016/j.joi.2018.11.008.

Demetrescu, C., Ribichini, A., & Schaerf, M. (2020). Are Italian research assessment exercises size-biased? *Scientometrics*, *125*(1), 533–549. https://doi.org/10.1007/s11192-020-03643-x.

Finocchi, I., Ribichini, A., & Schaerf, M. (2023). An analysis of international mobility and research productivity in computer science. *Scientometrics*, *128*, 6147–6175. https://doi.org/10.1007/s11192-023-04837-9.

Franceschini, F., & Maisano, D. (2017). Critical remarks on the Italian research assessment exercise VQR 2011-2014. *Journal of Informetrics*, *11*(2), 337–357. https://doi.org/10.1016/j.joi.2017.02.005.

Franzoni, C., Giuseppe, S., & Stephan, P. (2014). The mover's advantage: The superior performance of migrant scientists. *Economics Letters*, *122*, 89–93. https://doi.org/10.1016/j.econlet.2013.10.040.

Guan, J., & Ma, N. (2004). A comparative study of research performance in computer science. *Scientometrics*, *61*(3), 339–359. https://doi.org/10.1023/b:scie.0000045114.85737.1b.

Halevi, G., Bar-Ilan, J., & Moed, H. (2016). Researchers' mobility, productivity and impact: Case of top producing authors in seven disciplines. *Publishing Research Quarterly*, *32*. https://doi.org/10.1007/s12109-015-9437-0.

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, *117*(9), 4609–4616. https://doi.org/10.1073/pnas.1914221117.

Jaffe, K., ter Horst, E., Gunn, L. H., Zambrano, J. D., & Molina, G. (2020). A network analysis of research productivity by country, discipline, and wealth. *PLoS ONE*, *15*(5). https://doi.org/10.1371/journal.pone.0232458.

King, D. A. (2004). The scientific impact of nations. *Nature*, *430*(6997), 311–316. https://doi.org/10.1038/430311a.

Kulczycki, E. (2017). Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation*, *26*(1), 41–52. https://doi.org/10.1093/reseval/rvw023.

Langham-Putrow, A., Bakker, C., & Riegelman, A. (2021). Is the open access citation advantage real? A systematic review of the citation of open access and subscription-based articles. *PLoS ONE*, *16*. https://doi.org/10.1371/journal.pone.0253129. All Open Access, Gold Open Access, Green Open Access.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, *504*(7479), 211–213. https://doi.org/10.1038/504211a.

Leydesdorff, L., & Wagner, C. (2009). Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics*, *78*. https://doi.org/10.1007/s11192-008-1830-4.

Liang, Z., Luo, X., Gong, F., Bao, H., Qian, H., Jia, Z., & Li, G. (2015). Worldwide research productivity in the field of arthroscopy: A bibliometric analysis. *Arthroscopy*, *31*. https://doi.org/10.1016/j.arthro.2015.03.009.

Lupia, F., Mendicelli, A., Ribichini, A., Scarcello, F., & Schaerf, M. (2018). Computing the Shapley value in allocation problems: Approximations and bounds, with an application to the Italian VQR research assessment program. *Journal of Experimental and Theoretical Artificial Intelligence*, *30*(4), 505–524. https://doi.org/10.1080/0952813X.2018.1456791.

Mantovani, A., Rinaldi, E., & Zusi, C. (2020). Country rankings on the scientific production in endocrinology and diabetology. *Exploration of Medicine*, *1*. https://doi.org/10.37349/emed.2020.00020.

Netz, N., Hampel, S., & Aman, V. (2020). What effects does international mobility have on scientists' careers? A systematic review. *Research Evaluation*, *29*(3), 327–351. https://doi.org/10.1093/reseval/rvaa007.

Nicolaisen, J., & Frandsen, T. F. (2021). Number of references: A large-scale study of interval ratios. *Scientometrics*, *126*(1), 259–285. https://doi.org/10.1007/s11192-020-03764-3.

Plume, A. (2024). Open-access publishing: Citation advantage is unproven. *Nature*, *626*(7999), 480. https://doi.org/10.1038/d41586-024-00405-0. Cited by: 1.

Pradhan, D. K., Chakraborty, J., & Nandi, S. (2019). Applications of machine learning in analysis of citation network. In *Proceedings of the ACM India joint international conference on data science and management of data (Kolkata, India) (CODS-COMAD '19)* (pp. 330–333). New York, NY, USA: Association for Computing Machinery.

Pranckutė, R. (2021). Web of science (WoS) and scopus: The titans of bibliographic information in Today's Academic World. *Publications*, *9*(1). https://doi.org/10.3390/publications9010012.

Smith, K. M., Crookes, E., & Crookes, P. A. (2013). Measuring research 'impact' for academic promotion: Issues from the literature. *Journal of Higher Education Policy and Management*, *35*(4), 410–420. https://doi.org/10.1080/1360080X.2013.812173.

Stuart, D. (2015). Finding "good enough" metrics for the UK's research excellence framework. *Online Information Review*, *39*(2), 265–269. https://doi.org/10.1108/OIR-01-2015-0021.

Sun, X., Ding, K., & Lin, Y. (2016). Mapping the evolution of scientific fields based on cross-field authors. *Journal of Informetrics*, *10*(3), 750–761. https://doi.org/10.1016/j.joi.2016.04.016.

Tian, Y., Li, G., & Mao, J. (2023). Predicting the evolution of scientific communities by interpretable machine learning approaches. *Journal of Informetrics*, *17*(2), 101399. https://doi.org/10.1016/j.joi.2023.101399.

Tian, Y., Wen, C., & Hong, S. (2008). Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. *Journal of Informetrics*, *2*(1), 65–74. https://doi.org/10.1016/j.joi.2007.10.001.

Walters, G. D. (2011). The citation life cycle of articles published in 13 American psychological association journals: A 25-year longitudinal analysis. *Journal of the American Society for Information Science and Technology*, *62*(8), 1629–1636. https://doi.org/10.1002/asi.21560.

Wang, L. (2016). The structure and comparative advantages of China's scientific research: Quantitative and qualitative perspectives. *Scientometrics*, *106*, 435–452. https://doi.org/10.1007/s11192-015-1650-2.

Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, *35*(1), 83–104. https://doi.org/10.1016/j.respol.2005.08.006.