Article
Differences in citation patterns across areas, article types and age groups of researchers

Marcelo Mendoza [iD] **1***

1   Department of Informatics, Universidad Técnica Federico Santa María, Valparaíso, Chile; marcelo.mendoza@usm.cl

*   Correspondence: marcelo.mendoza@usm.cl; Tel.: +56-23037213

Abstract: The evaluation of research proposals and academic careers is subject to indicators of scientific productivity. Citations are critical signs of impact for researchers, and many indicators are based on these data. The literature shows that there are differences in citation patterns between areas. The scope and depth that these differences may have to motivate the extension of these studies considering types of articles and age groups of researchers. In this work, we conducted an exploratory study to elucidate what evidence there is about the existence of these differences in citation patterns. To perform this study, we collected historical data from Scopus. Analyzing these data, we evaluate if there are measurable differences in citation patterns. This study shows that there are evident differences in citation patterns between areas, types of publications, and age groups of researchers that may be relevant when carrying out researchers' academic evaluation.

Keywords: Citation patterns; Age groups of researchers; Types of publications.

1. Introduction

A seminal contribution to the measure of scientific impact is the one proposed by Eugene Garfield, who noted that the best way to follow the life cycle of a scientific article is through its citations. He claims that a relevant article for a community is frequently cited, so by tracking its citations, we can measure its impact [1]. Garfield extended the notion of impact of an article to the impact of a journal, introducing the journal impact factor (IF). Despite their wide use in research evaluation, metrics based on citations have criticisms and limitations. Moed highlighted the importance of the use of more complex measures that consider differences in citation patterns and dynamics between disciplines [2]. The problem with the use of a unique index that will not consider the specificities of each discipline produces distortions at different levels of research evaluation, both for projects, people, and even institutions [3].

When evaluating individuals, citation-based indices at the researchers level take an even greater role [4]. Hirsch [5] introduced the H-index. An H-index equal to *h* indicates that the researcher authored *h* papers with at least *h* citations. The H-index avoids the evaluation of long lists of uncited articles, determining which of these papers have had an impact on the scientific community. However, since citation patterns between disciplines are different, in some disciplines, the evolution of citations is much slower than in others [6]. Accordingly, to make a fair comparison, citation-based indices as the H-index [5] should be observed in very long time windows. These delays in citation patterns discouraged their use in the evaluation of young researchers.

One of the crucial points in the criticism of the scientific impact indicators is if they have predictive capacity [7]. Robien [8] showed that several of the most relevant works in Chemistry did not come from journals with the highest impact factor. The same observation was made in entrepreneurship journals [9]. Finardi [10] also observed something similar in a study that covered more disciplines. Vanclay [11] showed that the IF is based on a sample that may represent half the whole-of-life citations to some journals and a small fraction of the citations accruing to other journals. Since citation patterns across disciplines have different dynamics, the IF negatively impacts disciplines with slow dynamics. Other factors, such as the impact of self-citations, have also shown that some journals show

abrupt changes in their IF [12]. In addition, journals that publish reviews tend to be highly cited. In contrast, journals that publish in very specific topics often have low citation impact [13].

Variants of the IF or H-index have been proposed to overcome their limitations. For example, Braun et al [14] proposed weighting the effect of citations on the H-index computation, incorporating the journal's IF, giving high weight to citations that come from journals with a more significant impact. Balaban [15] proposed the opposite, giving high weight to citations that come from low IF journals since there would be more exceptional merit in receiving citations. Bergstrom [16] introduced the eigenfactor score, which homologates the concept of prestige with the stationary probability of a random walk in a citation graph. An extension of the eigenfactor is the Article Influence Score (AIS), which calculates the average impact of articles in a given journal, measuring impact in terms of eigenfactors. To infer prestige, agnostically at the scale of a community, González-Pereira et al [17] introduced the Scimago Journal Research indicator (SJR), which ranks journals based on citation weighting schemes and eigenfactor centrality. Moed [18] introduced SNIP (Source Normalized Impact per Paper), which is a ratio of a journal's citation impact and the citation potential in its subject field. Liu and Fang [19] detected the differences between citation-based indicators, comparing IF with 2-year windows (IF2) with IF over five years (IF5). The research shows that IF2 has much more significant changes in terms of quartiles than IF5. IF5 was also found to have a strong correlation with AIS, having a better predictive ability of paper influence than IF2.

Whatever the metric used to evaluate, it is essential to get citations [20]. The need to have a recognized track record is a crucial element to access competitive funds and promote scholars [21]. The scientific environment is becoming increasingly competitive. In this context, delays in a publication, measured as the time elapsed between reviews, acceptance, and definitive publication, affect research evaluation [22].

We introduce a citation-based study that elucidates differences between areas, types of publications, and researchers' age groups. This article expands the scope of previous studies, measuring the cross effect of citations patterns between areas and types of publications. In addition, we study differences in citation patters in terms of age groups of researchers. By extending the domain of analysis, we will find new citation patterns that have not been clearly evidenced in the related work. To achieve this goal, we conducted a data collection process in Scopus, including different types of publications by researchers from a broad spectrum of areas. The study will elucidate the existence of differences in citation patterns between areas, conditioned by the type of media considered for publication. We will show that these patterns have clear differences according to the age of the researchers. The conclusions of the study will indicate the need for indicators that consider different media of publication incorporating the age factor when evaluating the scientific outputs of a researcher.

The contributions of this article are:

– We study differences in citation patterns considering a large volume of data that covers different areas, types of publications and age group of researchers.
– We detect differences in citation patterns conditioned on areas and types of publications.
– We detect differences in citation patterns between areas conditioned on age groups of researchers.

The article is organized as follows. Related work is discussed in Section 2. In Section 3, we present the materials and methods used in this study. Section 4 presents differences in citations patterns between areas and types of publications. In Section 5, we study differences in citation patters across areas and age groups of researchers. Finally, Section 6 highlights findings and future work.

2. Related work

Lariviere et al [23] evaluated the concentration of papers and funding resources at the researchers' level, showing the existence of differences between areas. The study highlights a higher concentration of research funds, publications, and citations in Social Sciences and Humanities than in other fields. Nederhof et al [24] detected that the publications in Space-Life Physical Sciences are cited initially much lower than similar ones of ground-based research, showing that the use of short-citation window (1–4 years) is detrimental in impact assessment. The study states that bibliometric indicators should be calibrated to provide optimal monitoring of the impact. Prins et al [25] detected that some areas had a greater diversity of media in which they produced scientific outputs. In these areas, the use of Google Scholar could be beneficial to improve the coverage of publications, offering advantages over databases such as WoS, which offer a lower diversity of publication media.

Studies of differences in citation patterns according to type of publication have strongly focused on measuring the impact of conferences in Computer Science. Freyne et al [26] shows that Computer Science emphasis on publishing in conference proceedings hurt research evaluation when this evaluation is based on WoS indices. Vrettas and Sanderson [27] corroborates that Computer Scientists value conferences as a publication venue more than any other area. The analysis also shows that only a few conferences concentrate on many citations, while the mean rate per paper is higher in articles than in conferences. Glänzel et al [28] shows that conference proceedings are a valuable source of information for research evaluation in Computer Science. Furthermore, the study shows interesting international collaboration patterns triggered by these types of events not observed in other areas. Meho [29] assesses the quality of Computer Science conferences matching the citescore ranges of the top quantile journals. The study shows that top conferences make up 30 % of top-quartile publications. Thelwall [30] compares Scopus citations with Mendeley reader counts for conference papers and journal articles in eleven computing subfields showing high correlations between both counts. Kochetkov et al [31] applies a methodology similar to that of SJR to rank conferences in Computer Science. The study shows that the top conferences have similar SJR values to those reached by the top journals in the area. Li et al [32] study the impact of conference rankings in Computer Science using regression analysis. The study determines how these rankings have generated changes in the citation patterns of Australia and China. The study shows that the researchers align themselves with the rankings used in their country, modifying their publication patterns to fit these rankings better. However, Qian et al [33] shows that journals and conferences' relative status varies significantly between different areas of Computer Science. Recently, Yang and Qi [34] show that book publications are more relevant than conference publications in most areas. They also show that productivity indicators based on citations fail to capture these types of publications' relevance.

The skewness of citation distributions between areas has also received much attention from the community. Ruiz-Castillo and Costas [35] study the skewness of citations per area, comparing 27 different areas. The study shows that skewness is very similar across areas, and only a small part of the authors of each area are responsible for the production in this area. The most prolific authors dominate citation skewness. Radicchi et al [36] study that the probability of citing an article c times varies between areas. However, it shows that by rescaling the distributions by the average number of citations per article, a universal distribution of citations with high skewness emerges. Based on this index, the study introduces a variant of the h index that allows scientists from different areas to be compared. Albarrán et al [37] show that citation distributions based on 5-year observation windows are highly skewed. Rescaling these distributions, the level of skewness across areas is quite similar, showing that 10% of the most cited articles concentrate more than 40% of the total citations. By reproducing the study incorporating subfields, Crespo et al [38] shows that the differences between areas increase. At a finer granularity, more differences are detected between areas in terms of citation distributions. Bensman et al [39] studies the mean citation rate per article in Mathematics. The study shows that the citation rate is

highly skewed. Much of the publications that explain this phenomenon are due to many citations concentrated in Reviews. The study also analyzes the bibliometric structure of the area, detecting a weak central core of journals. The study hypothesizes that these differences with other areas would explain why methods based on citations do not apply to this particular area.

3. Materials and methods

To build the data repository we used for this study, we accessed information provided by Scopus. Scopus provides access to researcher profiles, in which lists of publications are displayed. To have a significant number of profiles of researchers, we built a profile crawler, using the rscopus library [1]. This library makes available functionalities that allow using the Scopus API, from which data can be accessed. Our crawler was built on top of rscopus, which invokes the Scopus API to get the data. The crawler uses a set of seed profiles from which it retrieves its collaborators, which in turn are used as seeds by the crawler. Iteratively, the crawler builds a repository of researcher profiles along with its publication lists.

Our crawler stores the profiles in a relational database, on which it asks if a profile is already stored. If it is not stored, it indexes it. It does the same with publications, avoiding duplicates. Scopus marks the areas of knowledge of the publishing media. A method of imputation maintains a system of classification of areas at the researchers' level. Our crawler downloaded all this data.

Each author's list includes descriptive fields such as the title, author names, name of the media, doi, publication date, and some descriptive keywords. An author id is stored for each author, provided by Scopus. The identifier of each publication is the Scopus id. A crucial method used in the study is the one that allows us to retrieve the references of each publication. This method establishes the relationship between the Scopus id of the publication and that of each of its references. We use these data records to count citations between the papers of the repository.

We conducted the data collection process from a set of profiles for each of the disciplines considered in the taxonomy of areas maintained by Scopus. This taxonomy considers 27 areas of knowledge. In each of them, we fed the crawler with a list of top cited authors. To determine the seeds of each area, we looked for the most reputable journals by field, according to the impact factor, and in them, we looked for the most cited papers. From these works, we selected the top ten most-cited authors. The crawler ran for about three months collecting profiles from 111813 researchers. We stored a total of 4,504,779 citable document records in our repository. Our collection spans the years 1970 to 2018. A summary of our collection's basic statistics is shown in the Appendix.

Each publication media declares a list of related areas, in which Scopus maps 27 areas of knowledge. Then, Scopus classifies authors in knowledge areas, according to the journals and conferences in which they publish. These areas are attributed to researchers, accounting for the proportion of publications in each field. Accordingly, each investigator is classified in the area in which he/she records more publications. We discarded for this study areas with too few authors. We picked the areas with at least 3000 authors in our repository. Accordingly, the dataset used for this study comprises data records for ten areas from 95668 different authors, as it is shown in Table 1.

Scopus classifies the documents into 15 different types of publications. Four publication types (Conference Review, Report, Abstract Report, and Business Article) have too few documents in our dataset, totaling less than 600 documents. We discarded these publication types from the study. Accordingly, the resulting dataset has 4102168 documents classified into 11 types of publications. We show the number of documents per publication type in Table 2.

---

[1] https://cran.r-project.org/web/packages/rscopus/index.html

Table 1: Distributions of authors across knowledge areas covered in this study.

Area	# authors
Medicine	34974
Physics and Astronomy	11308
Computer Science	9052
Biochemistry, Genetics and Molecular Biology	8418
Mathematics	7520
Engineering	7367
Psychology	6145
Agricultural and Biological Sciences	3886
Social Sciences	3691
Chemistry	3307
Total	95668

Table 2: Documents per publication type.

Publication type	# documents
Article	2978819
Conference Paper	561048
Review	237186
Letter	95650
Editorial	66501
Chapter	45191
Note	42988
Article in Press	23912
Short Survey	23743
Erratum	21727
Book	5403
Total	4102168

Citations are verified relationships of references between citable documents in our collection. It should be note that this track of citations is a sample of the total citations of Scopus since it corresponds to citations between documents in our repository. Our collection totals more than eight and a half million citations among citable documents. However, we discarded citations from or to areas not considered in this study. In total, we discarded 976349 citations, focusing the analysis into the 7534371 citations recorded between documents of the top-ten most salient areas of our repository.

4. Differences in citations patterns across areas and publication types.

An essential factor in this study is to conduct an analysis that distinguishes between the type of media in which the publications are made. We consider eleven types of publications, among which articles and conference papers are the ones with the most documents (see Table 2). Several of the other types of publications are related to journals. Review, letter, editorial, note, short survey, and erratum are types of publications included within journals. We decided to report them separately to distinguish the differences between these specific types of publications and regular articles. Something similar happens with chapter and book, two related types since chapter is a type of publication included within a book. We decided to report these types separately to distinguish the differences between them.

The number of documents according to each type of publication varies between areas. Table 3 shows that Reviews and Letters are more frequent in Medicine than in other areas, while Chapters and Books are more frequent in Social Sciences. A significant difference is observed in Computer Science, where conference papers are much more frequent than in other areas, totaling more than 60% of the publications in that area.

Table 3 shows significant differences between areas. Articles are a majority in all areas except Computer Science, in which they cover 31% of the sample. On the other hand, conference papers cover 61% of Computer Science production, surpassing the coverage shown in other areas. Conference papers are also relevant in other areas, with 38% in Engineering and 19% in Physics. However, in the other areas its incidence is very low, with around 3% or even less.

Citations according to each type of publication vary between areas. Table 4 shows that Reviews, Letters, and Editorials receive more citations in Medicine than in other areas, while Chapters and Books are more cited in Social Sciences. This fact is explained because these types of publications are more frequent in these areas than in other areas. Articles receive more citations than the other types of publications, which is valid in all areas, even in Computer Science, where these publications are not the majority.

The volume of citations captured by articles is majority in all areas. For example, in Mathematics, articles capture 95% of citations, while in Physics, they capture 92%. Something similar occurs in Biochemistry, and Agricultural and Biological Sciences. Psychology

Table 3: Types of publications by area. Some types of publications are much more frequent in some areas than in others. While Review and Letters are frequent in Medicine and Book and Chapters are in Social Sciences, conference papers are very relevant in Computer Science. The quantities are shown as percentages of the amount of documents in each area, to facilitate comparison between areas.

Area	Documents	Art.	Conf.	Rev.	Let.	Edit.	Chap.	Note	in Pr.	Surv.	Err.	Book
Medicine	2156501	0,7536	0,0361	**0,0941**	**0,0448**	0,0220	0,0089	**0,0178**	0,0068	0,0094	0,0057	0,0008
Physics.	532602	0,7549	0,1998	0,0165	0,0067	0,0032	0,0037	0,0028	0,0011	0,0030	0,0076	0,0007
Comp. Sci.	411796	0,3191	**0,6123**	0,0109	0,0010	**0,0314**	0,0151	0,0009	0,0059	0,0003	0,0008	0,0023
Biochemistry	295112	0,8508	0,0272	0,0619	0,0085	0,0070	0,0120	0,0079	0,0030	**0,0102**	**0,0106**	0,0008
Mathematics	211292	**0,9150**	0,0328	0,0097	0,0009	0,0087	0,0140	0,0034	0,0083	0,0004	0,0047	0,0022
Engineering	332754	0,5833	0,3842	0,0111	0,0014	0,0055	0,0062	0,0013	0,0043	0,0002	0,0010	0,0013
Psychology	145775	0,8326	0,0091	0,0556	0,0024	0,0152	0,0467	0,0130	0,0149	0,0019	0,0033	0,0052
Agr. and Bio.	112747	0,8812	0,0263	0,0457	0,0052	0,0059	0,0119	0,0072	0,0086	0,0025	0,0043	0,0011
Social Sci.	43080	0,7326	0,0157	0,0750	0,0042	0,0305	**0,0934**	0,0096	**0,0166**	0,0009	0,0024	**0,0191**
Chemistry	185624	0,9109	0,0307	0,0259	0,0062	0,0048	0,0059	0,0039	0,0027	0,0017	0,0064	0,0007

Table 4: Citations across types of publications per area. Articles receive more citations than the other types of publications, which is valid in all areas. The quantities are shown as percentages of the amount of citations in each area, to facilitate comparison between areas.

Area	Citations	Art.	Conf.	Rev.	Let.	Edit.	Chap.	Note	in Pr.	Surv.	Err.	Book
Medicine	3329158	0,8348	0,0203	**0,1177**	**0,0115**	**0,0048**	0,0006	**0,0046**	0,0002	0,0050	0,0001	0,0003
Physics	1819721	0,9203	0,0192	0,0492	0,0066	0,0005	0,0001	0,0007	0,0000	0,0011	**0,0016**	0,0005
Comp. Sci.	396829	0,5152	**0,4377**	0,0360	0,0011	0,0020	0,0034	0,0003	**0,0005**	0,0005	0,0000	0,0034
Biochemistry	943584	0,8816	0,0135	0,0857	0,0030	0,0006	0,0005	0,0027	0,0001	**0,0118**	0,0003	0,0002
Mathematics	188730	**0,9567**	0,0132	0,0229	0,0003	0,0006	0,0018	0,0012	0,0001	0,0003	0,0006	0,0024
Engineering	252274	0,7088	0,2578	0,0207	0,0017	0,0007	0,0013	0,0006	0,0003	0,0000	0,0001	0,0080
Psychology	283727	0,8991	0,0071	0,0749	0,0003	0,0021	0,0105	0,0025	0,0001	0,0006	0,0001	0,0029
Agr. and Bio.	33321	0,8383	0,0380	0,1137	0,0005	0,0014	0,0020	0,0028	0,0001	0,0016	0,0001	0,0016
Social Sci.	26305	0,7920	0,0174	0,1088	0,0002	0,0043	**0,0288**	0,0032	0,0004	0,0001	0,0000	**0,0447**
Chemistry	410929	0,9097	0,0130	0,0579	0,0107	0,0014	0,0013	0,0021	0,0000	0,0020	0,0004	0,0015

shows a similar pattern, with 89% of citations captured in articles. In Engineering, the incidence of citations in articles is 70%, showing that 25% of citations go to conference papers. This difference is most important in Computer Science. In this area, 51% of citations go to articles and 43% to conference papers.

Table 5 shows citations across areas. The quantities are shown as a fraction of the total amount of citations produced by a given area. For instance, the entry 0.12 from Medicine to Biochemistry indicates that 12% of the citations produced by Medicine go to Biochemistry. Table 5 shows that citations within the same area are a majority. We show the diagonal with a red color to indicate a higher prevalence of within citations and blue a less prevalence of this pattern. Second majorities, indicated with underlines, are essential to understand which areas are strongly related to others. Medicine and Biochemistry are firmly related, with the volume of citations from Biochemistry to Medicine being higher than the other way. Other strong relationships occur between Psychology and Medicine and between Agricultural and Biological Sciences and Psychology. The most substantial connection is from Social Sciences to Psychology, which delivers 35% of its citations to publications in this area. Curiously, this relationship is not symmetric since only 2% of Psychology citations go to Social Sciences.

Table 6 shows the average number of citations per paper according to the type of publication and area. These results show that the expected number of citations per paper vary across areas. While the largest amount is observed in Physics Reviews, other types of

Table 5: Citations across areas. The quantities are shown as a fraction of the total amount of citations produced by a given area, to facilitate comparison between areas. Red and blue entries show the diagonal, indicating the prevalence of within citations. Underlined entries show the second majority for each area, suggesting connections between areas.

From/To	Medicine	Physics	Comp. Sci.	Biochemistry	Mathematics	Engineering	Psychology	Agr. and Bio.	Social Sci.	Chemistry
Medicine	0,8304	0,0245	0,0024	0,1200	0,0016	0,0008	0,0148	0,0023	0,0010	0,0022
Physics	0,0081	0,9557	0,0036	0,0144	0,0057	0,0037	0,0002	0,0003	0,0001	0,0082
Comp. Sci.	0,0386	0,0280	0,7999	0,0243	0,0242	0,0537	0,0225	0,0018	0,0051	0,0019
Biochemistry	0,3163	0,0640	0,0031	0,5779	0,0026	0,0009	0,0032	0,0062	0,0003	0,0256
Mathematics	0,0220	0,0636	0,0287	0,0176	0,8111	0,0443	0,0020	0,0072	0,0008	0,0026
Engineering	0,0254	0,0401	0,0939	0,0137	0,0480	0,7610	0,0023	0,0019	0,0014	0,0123
Psychology	0,1470	0,0076	0,0068	0,0284	0,0015	0,0007	0,7805	0,0078	0,0197	0,0000
Agr. and Bio.	0,2096	0,0414	0,0111	0,2346	0,0237	0,0109	0,0472	0,3868	0,0044	0,0303
Social Sci.	0,1242	0,0081	0,0349	0,0237	0,0105	0,0096	0,3499	0,0068	0,4290	0,0033
Chemistry	0,0268	0,0666	0,0012	0,0737	0,0009	0,0032	0,0001	0,0027	0,0001	0,8247

Table 6: Average number of citations per type of publication and area. Quantities are reported over the amount of documents per area, to facilitate the comparison between areas.

	Art.	Conf.	Rev.	Let.	Edit.	Chap.	Note	in Pr.	Surv.	Err.	Book
Medicine	1,8586	0,9456	2,0977	0,4290	0,3646	0,1112	0,4357	0,0534	0,8985	0,0386	0,6337
Physics	**4,2375**	0,3340	**10,3465**	3,4167	0,5580	0,1401	0,8387	**0,1394**	1,3014	**0,7485**	2,7389
Comp. Sci.	1,6274	0,7205	3,3217	1,0918	0,0651	0,2278	0,3439	0,0617	1,5210	0,0572	1,5045
Biochemistry	3,4465	**1,6526**	4,6012	1,1762	0,3010	0,1393	1,1428	0,0852	**3,8305**	0,0911	0,6609
Mathematics	0,9667	0,3723	2,1897	0,3005	0,0618	0,1199	0,3127	0,0108	0,5854	0,1238	1,0067
Engineering	0,9291	0,5131	1,4225	0,8994	0,0915	0,1581	0,3670	0,0519	0,0395	0,0590	4,7176
Psychology	2,2745	1,6489	2,8388	0,2663	0,2951	0,4715	0,4002	0,0091	0,6230	0,0320	1,1642
Agr. and Bio.	0,2914	0,4431	0,7624	0,0267	0,0704	0,0505	0,1171	0,0043	0,1956	0,0043	0,4500
Social Sci.	0,7179	0,7361	0,9632	0,0258	0,0941	0,2048	0,2210	0,0165	0,0625	0,0114	1,5587
Chemistry	2,2390	0,9446	5,0129	**3,8876**	**0,6759**	0,4874	**1,1848**	0,0102	2,6484	**0,1420**	**4,7200**

publications show great relevance for other areas. For example, Letters have an expected number of citations per paper close to 4 in Chemistry, followed by Physics, with 3.4. This type of publication has a much lower quantity in other areas. Other types of publications with a high number of citations per document are Books in Chemistry and Engineering. The amounts of publications of this type are very low, as shown in Table 3. Despite this fact, this type of publication captures a great deal of attention in these areas, reflected in a high average number of citations. The average number of citations per article varies between areas, obtaining its highest value in Physics, with 4.2 average citations, followed by Biochemistry with 3.4. These numbers are very low in Agricultural and Biological Sciences, with only 0.29 citations per paper. In this area, the expected number of citations is higher in reviews. In general, the expected number of citations per paper in Agricultural and Biological Sciences is low, which is explained because many citations go to other areas.

Interestingly, in Computer Science, the expected number of citations in conference papers reaches only 0.72, less than half of what the index for articles shows. This finding suggests that many conference papers receive very few citations, a fact that can be corroborated from the Zipfians that show the distributions of citations per paper, as was illustrated in Brzezinski [40], Moreira et al [41]. In this area, reviews also have a high average number of citations, with 3.32 citations per review. In general, the results show that the attention provided to each type of publication strongly depends on each area. In general, reviews receive more citations than other types of publications, and both conference papers and articles have very different indicators between areas. It is also striking that in Biochemistry, the expected number of citations in conference papers is very high, with 1.65 per paper, which is the highest in the repository. This fact occurs even though there are very few conference papers in Biochemistry (just over 1% of the total number of Biochemistry
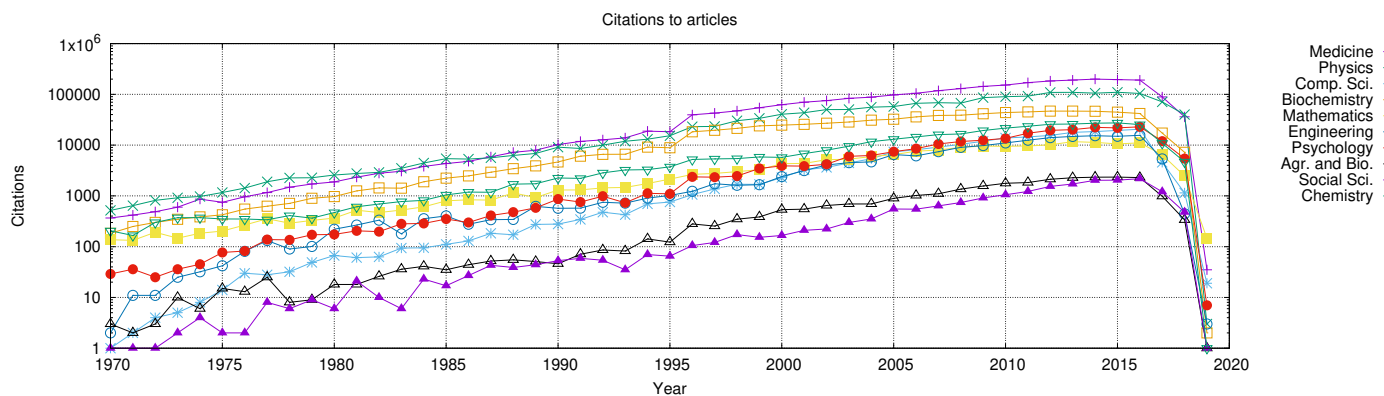
Figure 1. Number of citations to articles per year and area. Each point represents the number of citations recorded by documents published in that year that belongs to the indicated area.
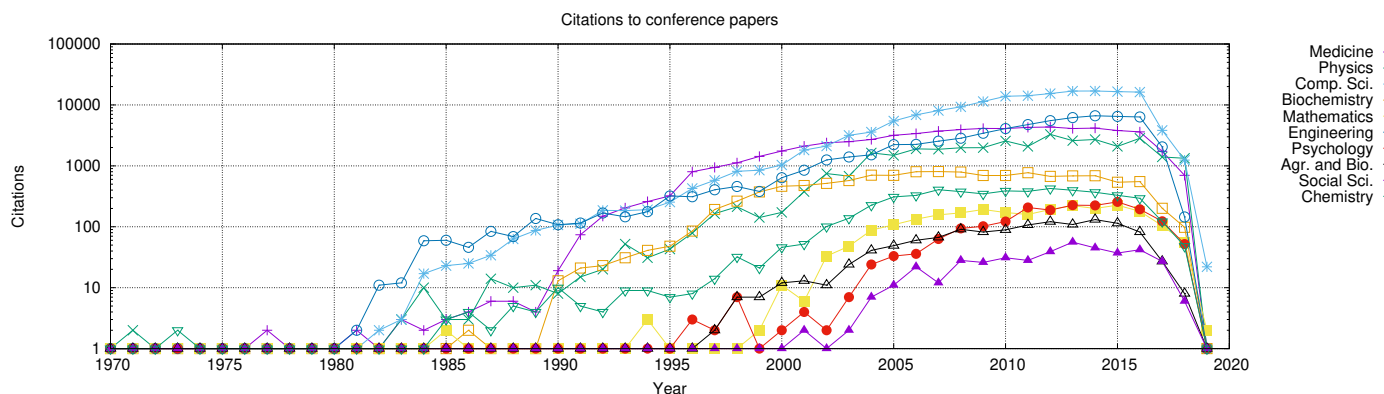

Figure 2. Number of citations to conference papers per year and area. Each point represents the number of citations recorded by documents published in that year that belongs to the indicated area.

papers, see Table 3), which indicates that although conference papers' production is low, this attracts much attention from the community.

Our repository records documents between the years 1970 to 2018. In this period, the volume of citations to articles has grown in all areas, as shown in Figure 1. However, the volume of citations between areas is very different. While Physics, Medicine, and Biochemistry show explosive growth in citations since the '80s, growth in Social Sciences and Agricultural and Biological Sciences has been slower. The differences in volumes are noticeable. The gap between the most cited areas and the rest is in one to two orders of magnitude (note that the figure uses a logarithmic scale). We can see that Figure 1 shows a decrease in citation growth from the year 2013 onwards. This fact is due to the delay produced by the dynamics of citations. As the sample ended in 2018, citation growth is captured up to five years earlier. This observation indicates that many papers' life span are more extensive than five years, on average, corroborating the recent finding of Liu and Fang [19].

Figure 2 shows that citations to conference papers have grown since the '80s. This growth begins in Engineering and Computer Science. Other areas show slow growth in the volume of citations to conference papers. For example, Medicine started growth in the late 1980s, as did Biochemistry and Physics. Even later is the growth in this type of citation in Chemistry (the '90s). In the other areas, this growth has been prolonged or, for all practical purposes, it is non-existent.

The growths on the curves in Figure 3 represent different events. While in Engineering and Computer Science, this dynamic represents a significant volume of the total citations, in the other areas, these quantities are much lower. We observe this fact in Figure 3, where
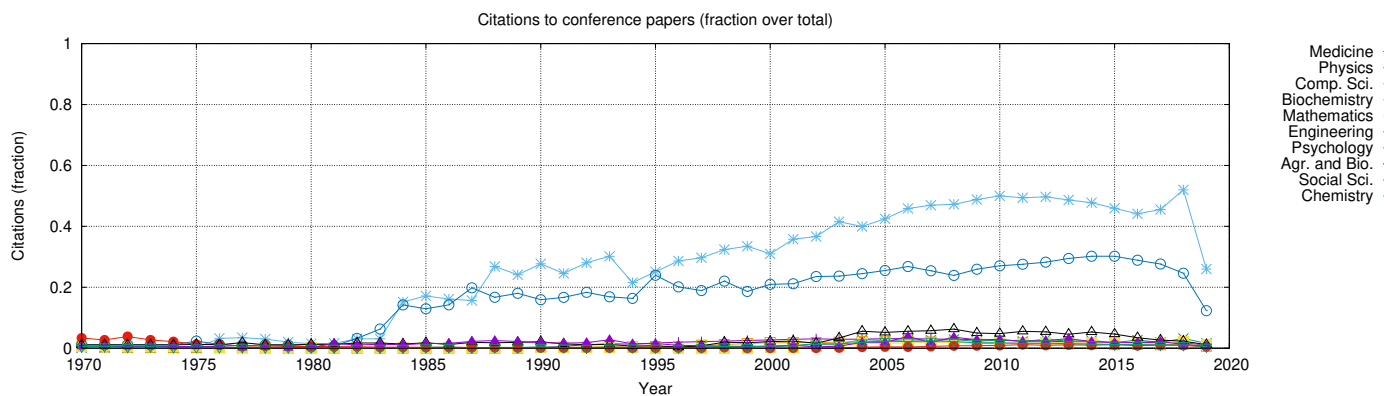
Citations to conference papers (fraction over total)

Figure 3. Fraction of citations to conference papers per year and area. Each point represents the number of citations recorded by documents published in that year that belongs to the indicated area.
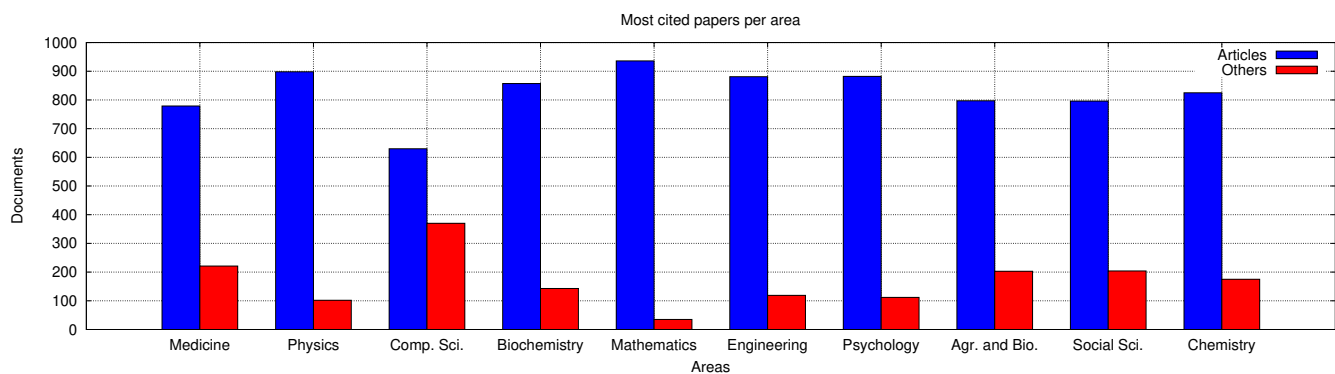
Most cited papers per area

Figure 4. Most cited papers per area. Even though many highly cited papers are articles, there are highly cited papers in the category others. Computer Science records nearly 400 highly cited papers in the category others.

we show the fraction of citations to conference papers over the total number of citations per area and year. The figure clearly shows that citations' dynamics to conference papers are part of a phenomenon in Engineering and Computer Science. In contrast, in the other areas, this phenomenon is practically non-existent. The curves show slower growth in Engineering than in Computer Science. In this area, citations to conference papers reached almost half of the total citations in 2010.

We analyze citation patterns in the top-1000 most cited papers per area. Figure 4 shows that many of these papers are articles. The figure also shows that some of these papers belong to other types of publications. For example, more than 200 highly cited papers in Medicine belong to other types of publication. Something similar appears in other areas such as Physics and Chemistry. This fact is explained by the high prevalence in these areas of citations to reviews and letters. Agricultural and Biological Sciences and Social Sciences also show highly cited papers in other types of publications, which is explained by the high prevalence of citations to letters, books, and chapters. The area with the least highly cited papers in other types of publications is Mathematics, where more than 900 highly cited papers are articles. On the other hand, Computer Science shows that several highly cited papers are conference papers, reaching almost 400 top papers. This quantity is the largest amount in this category in the entire repository. Curiously, the Engineering conference papers that are not articles are very few, indicating that even though there is a growth in the number of citations to conference papers, the most cited papers in Engineering continue to belong to journals.

The top-1000 papers per area span many years of the timeline. Figures 5 and 6 shows that in Medicine, many of the top-cited articles were published in the '90s and' 00s, with
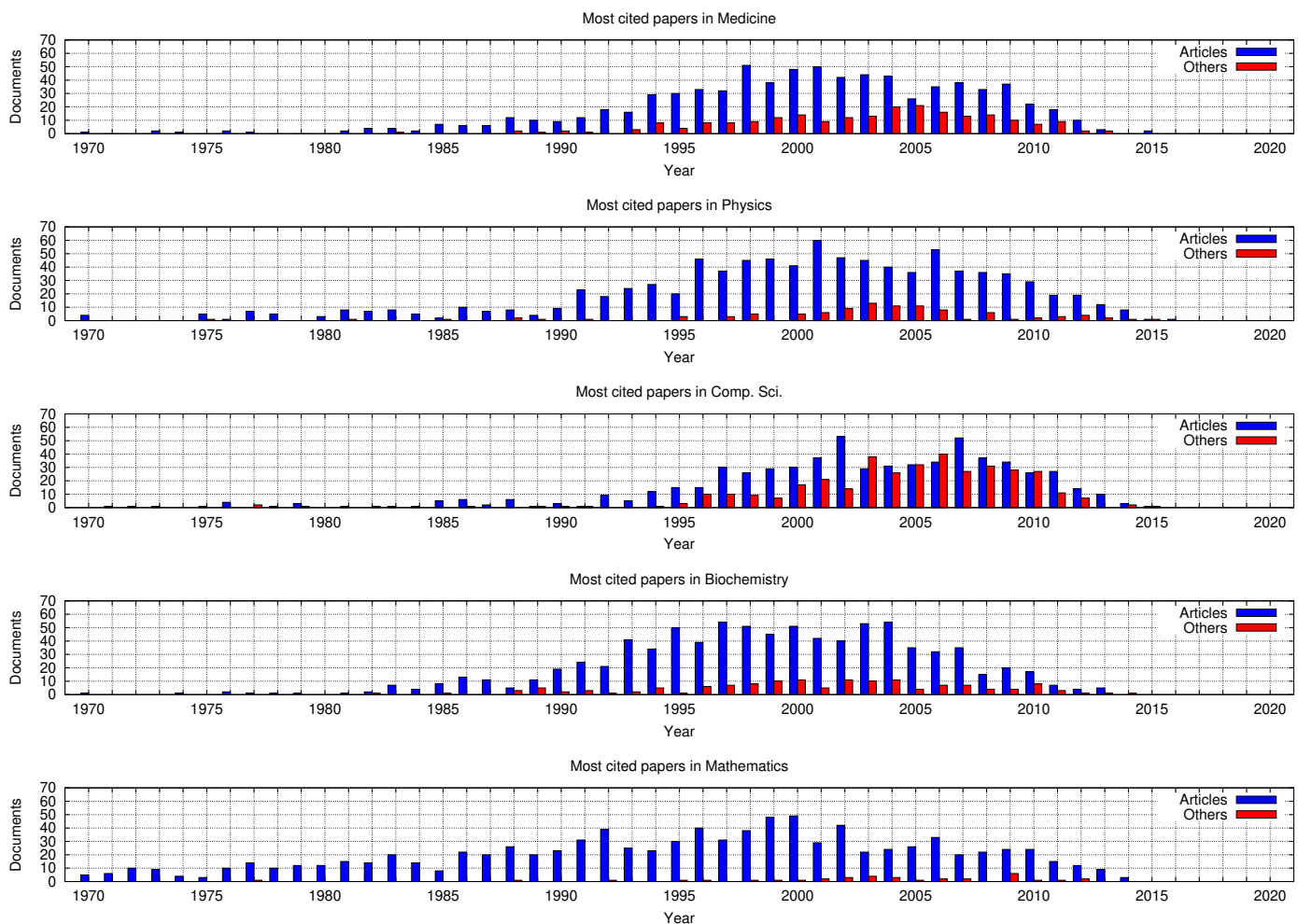
Figure 5. Most cited papers per area along time (Medicine, Physics, Comp. Sci., Biochemistry, and Mathematics).

a distribution very similar to that exhibited by Physics, Biochemistry, and Chemistry. Mathematics and Psychology papers span along more years, ranging from the '70s to' 00s. Computer Science is the area in which top papers are located in a shorter timeline. The bulk of these papers were published from the late '90s and until the early 10s. This distribution is similar to Engineering, even though its first most cited papers are young. Psychology shows two periods with many highly cited papers, around the years 96 and 2000. The distributions of Social Sciences and Agricultural and Biological Sciences are similar. As for the highly cited papers of another type, such as conference papers, letters, and reviews, these show to be relevant in Medicine and to a lesser extent in Physics and Biochemistry. In Engineering, the impact of this type of paper is much lower. It is striking that these papers acquired a little more relevance in Social Sciences in the mid-'00s and to a lesser extent around the same time in Chemistry and Agricultural and Biological Sciences. The area where conference papers have had the most impact is Computer Science. In this area, its relevance from 2005 onwards is significant, surpassing even the articles. This trend is only observed in Computer Science, illustrating that this reality is unique to this area.

We also analyze if there are differences between areas regarding the papers' life span, according to their activity in terms of citations. To do this analysis, we counted the number of citations received for each paper in our repository since its year of publication. Our repository spans almost 50 years. In Figure 7, we show the average life span of all the papers in our repository. These results show that the average life span of a paper begins in the same year in which it is published and acquires its most significant relevance in its
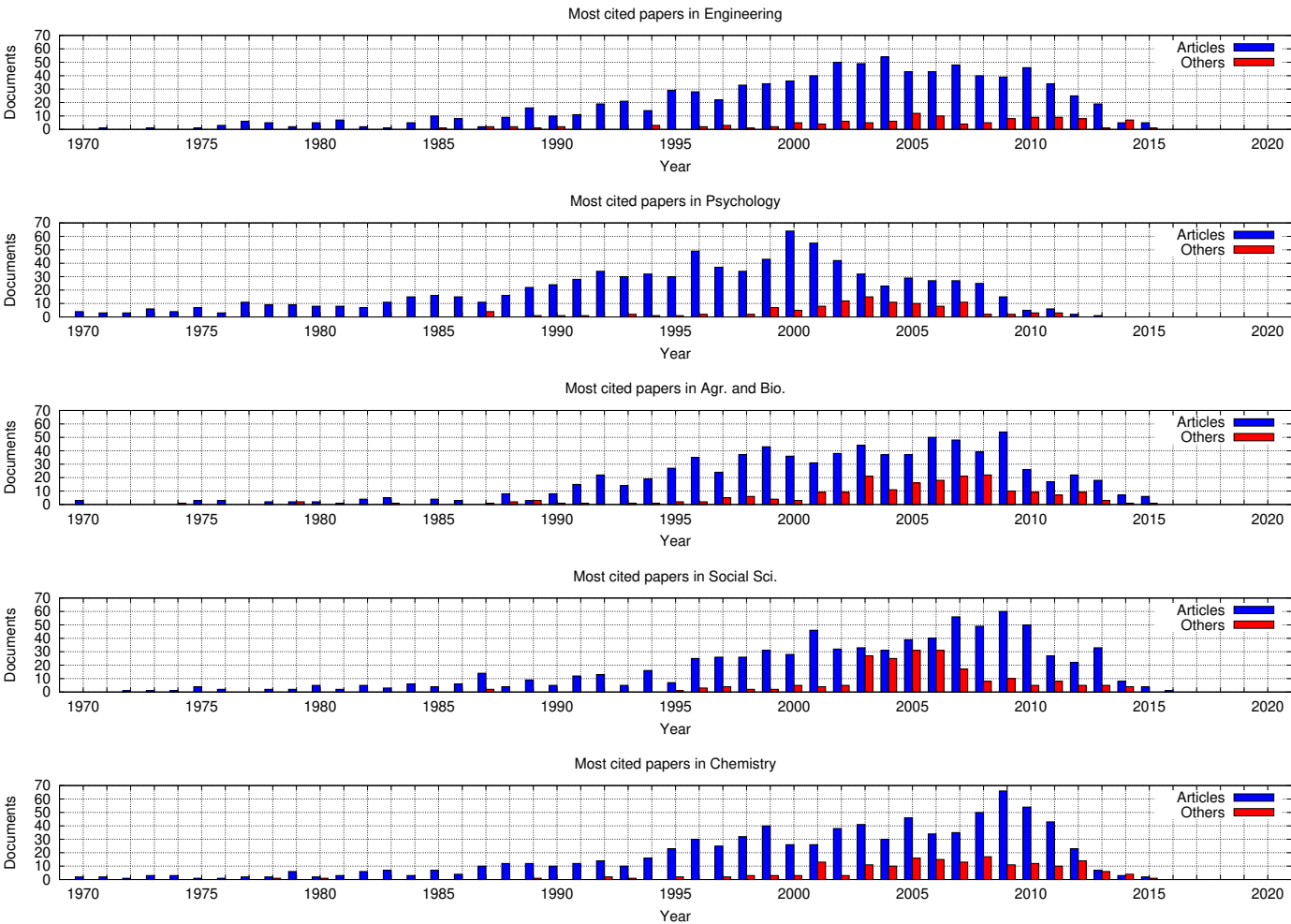
Figure 6. Most cited papers per area along time (Engineering, Psychology, Agr. and Bio, Social Sciences and Chemistry).

third year of life. Then, its impact progressively wanes. Very few papers have a life span of more than 20 years, and as time progresses, the citations that most papers receive diminish until they are almost entirely extinct.
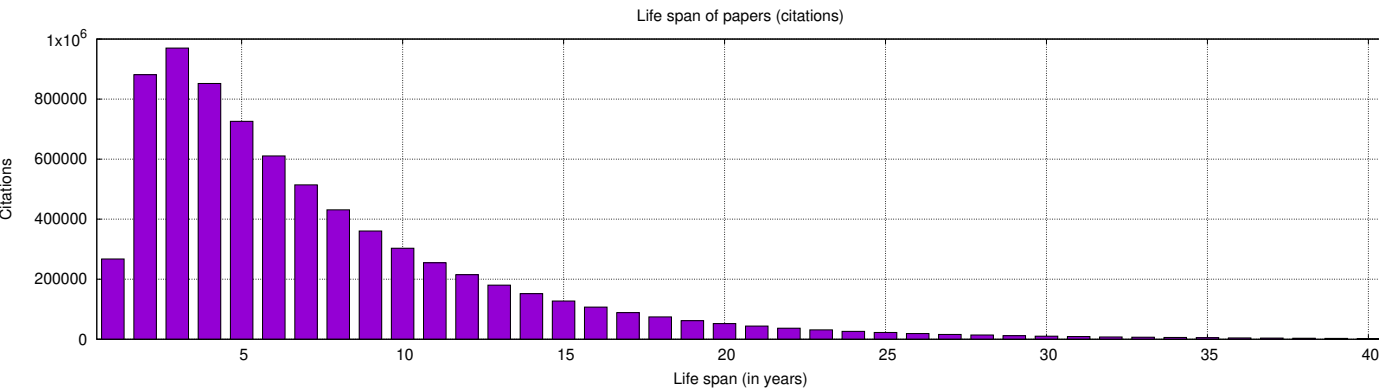

Figure 7. Life span of papers in our repository.

To study the differences in life span according to citations between areas, we split the timeline into six five-year periods to compare papers published in close years in each

plot. These results are shown in Figure 8, which shows the 20-year life spans of papers published in 6 different five-year periods, from the '80s to the end of the '00s.
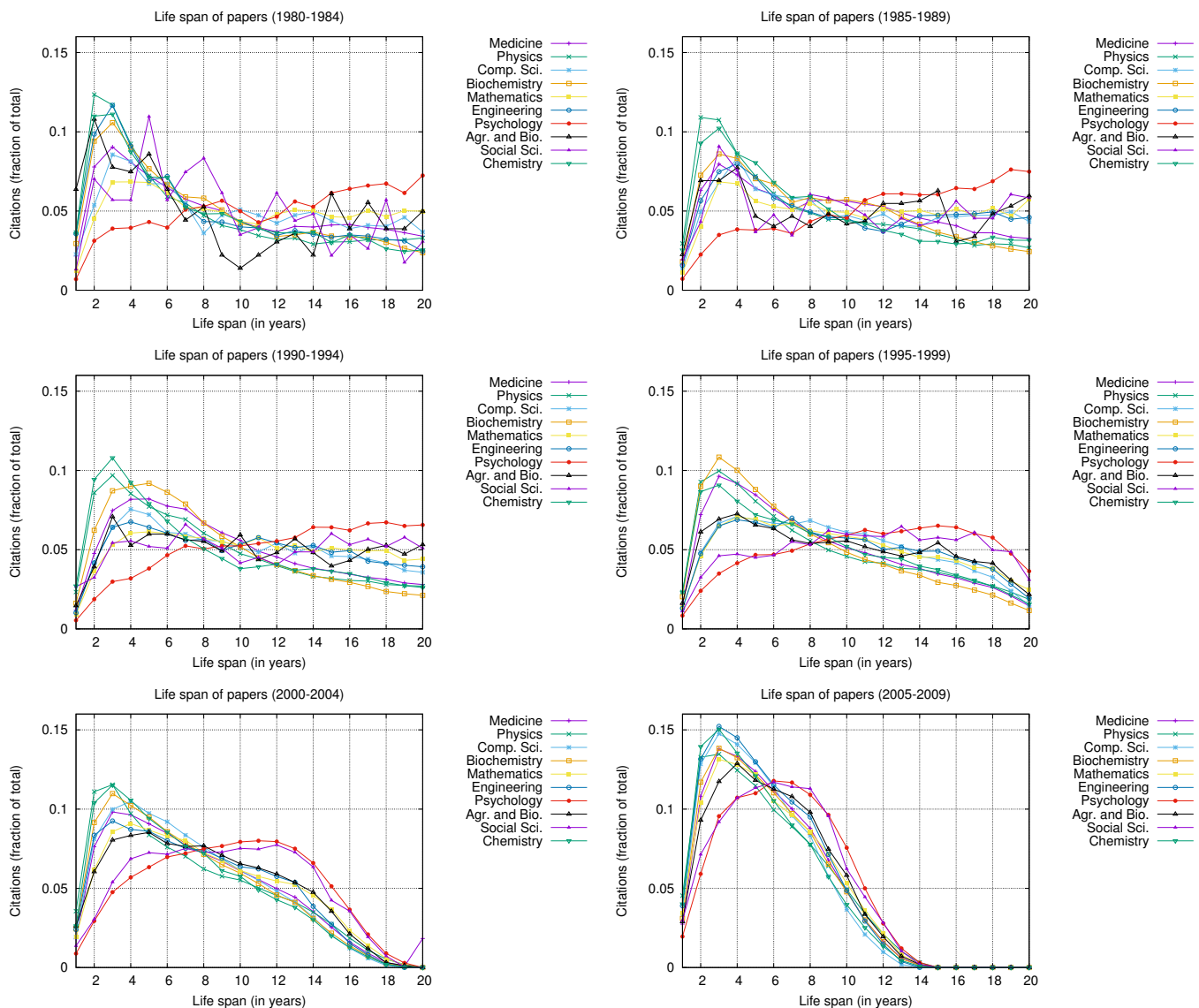

Figure 8. Life span of papers across areas. The figure shows the 20-year life spans of papers published in 6 different five-year periods, from the '80s to the end of the '00s.

Figure 8 shows some interesting dynamics. While various areas in the 1980s followed a similar life span to Figure 7, showing a performance peak in the third year, Psychology and Social Sciences show a much slower growth curve. This effect becomes more noticeable in the '90s, where the curves of these two areas show growth during the 20 years of the life span while the rest of the areas' curves begin to decline after the third year of life. This finding suggests that the papers cited in Psychology and Social Sciences have a longer life span than other areas. This difference is also noticeable in the '00s. Most of the areas show their peak of citation in the third year, followed by a substantial decline, while Psychology and Social Sciences have a slower growth but a long life.

5. Differences in citation patterns across areas and age groups of researchers

We analyze the differences in citation patterns by comparing groups of researchers of different ages. For this purpose, we calculate each researcher's academic birth, defined as the year in which the author published their first work. Authors span many years of the
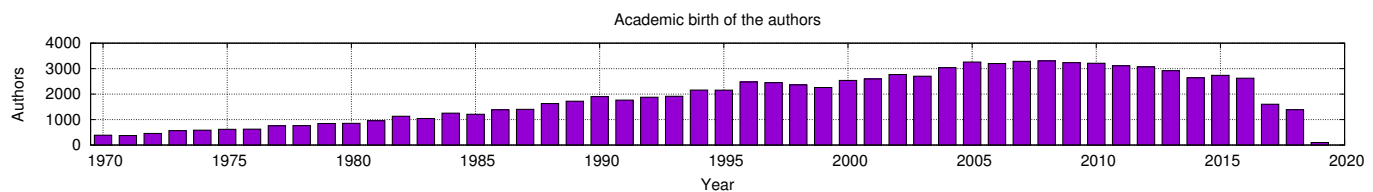
Academic birth of the authors

Figure 9. The academic birth of the authors studied in this paper. Academic births are firmly located in the '90s and '00s.

timeline, beginning in 1970. Figure 9 shows that many authors recorded their academic birth after the '90s. Most of the academic births are located in the decade of the '00s.

We separate the authors into three groups according to their academic birth. Seniors are those researchers whose birth is before the year 2000. Mid-age researchers are those born between the years 2000 and 2009. Young researchers are those born after 2009. The sizes of these groups do not have significant differences, which facilitates the comparison between them. The senior group has 42417 (44%) researchers; the mid-age group has 29923 (31%) and the young group 23321 (25%).

For each of these groups, we compute authors' life span, defined as the track of citations recorded over time. Each citation track records the total number of citations received by an author per year. We compute a life span of ten years for young researchers since these authors were academically born after 2009. The other two groups are analyzed using 20-year life spans.

Figures **??** and **??** show the life span per area and age group. The solid line shows the mean life span, while the colored area indicates the standard deviation around the mean. The axes use the same scales for each age group to facilitate comparison between areas. The only area in which it was not possible to scale was Physics, which receives many more citations per author than the rest of the groups.

Authors' life spans vary widely between areas and age groups. While the young and mid-age life spans decrease towards the end of the timeline, the seniors show growth towards the year 20. This fact is attributable to the cut-off year in data acquisition. Some comparisons between groups are very interesting. Physics shows rapid growth in authors' life span, with an average of citations in young authors of more than 20 citations by the third year of life. This effect is observed in the other age groups, with a peak in mean citations in the tenth year of life in mid-age researchers and a sustained growth that exceeds 50 citations on average by year 20 in seniors. Physics is out of scale in this comparison because the volumes of citations in the area records are much higher than in the other areas. On the other hand, the other areas record comparable amounts of citations, which allows the comparison between areas on the same scale. It can be seen that the axes used in these figures are the same for each age group (except Physics).

Mathematics shows a strong fit to the mean, with a low standard deviation (see Figure **??**). This effect is maintained in the three age groups, indicating that very few authors have rapid growth in citations in this area. It is also noteworthy that the mean life span in Mathematics is very low, which indicates that most authors concentrate few citations. This fact is due to two factors. First, the volume of citations in Mathematics is low (see Table 4). Furthermore, the average number of citations per paper is also very low (see Table 6). Figure **??** shows that the life spans of Medicine and Biochemistry are similar, which may be because these two areas have a high volume of cross-citations. The Computer Science standard deviation indicates that some authors receive many citations and others receive very few. In this area, the fit to the mean is more significant than in Medicine and Biochemistry, indicating that authors with rapid growth in citations are scarcer than in these other areas.

Figure **??** shows that Agricultural and Biological Sciences and Social Sciences also show a strong fit to the mean, with many authors receiving very few citations. The authors' low mean life span is because the volume of citations in these areas is very low (see Table 4)

Table 7: Spearman's correlation between consecutive citation quinquenniums in each age group. The correlation in young researchers is measured between the first and second quinquennium; in mid-age, it is measured between the second and third quinquennium and seniors between the third and fourth quinquennium.

	Medicine	Physics	Comp. Sci.	Biochemistry	Mathematics	Engineering	Psychology	Agr. and Bio.	Social Sci.	Chemistry
young	0,585	0,626	0,557	0,613	0,316	0,428	0,611	0,671	0,733	0,719
mid age	0,611	0,491	0,503	0,692	0,679	0,743	0,584	0,464	0,584	0,621
senior	0,863	0,831	0,811	0,838	0,799	0,817	0,879	0,805	0,781	0,857

and because many of their citations go to other areas (see Table 5). In the case of Agricultural and Social Sciences, only 0.3868 of its citations remain in the area, while in Social Sciences, the fraction of within citations is 0.429. The other areas show a weaker fit to the mean. Both Engineering, Psychology, and Chemistry show the existence of separate author tracks from the media. Chemistry shows more careers with a high volume of citations, with consolidation in mid-age around year 10 of the life span. In young researchers, the peak of citations is observed at year 4. The mean of citations in Engineering has a slower growth, which indicates that many authors receive few citations in this area. Psychology has a higher growth curve than Engineering, with a peak of citations around year 10 for mid-age researchers

A key question in citation-based research is whether citations do indeed have predictive power. The ground on which citation-based evaluation is based is that citations received in the past are a good indicator of future citations. A question for this study is whether the citation data has the same predictive capacity in different areas. One possible scenario is that citation data has better predictive capabilities in some areas than in others. If this is the case, we may conclude that evaluation based on citations is more informative in some areas than in others.

We analyze the correlation of citations in each age group. A strong correlation indicates that the volume of past citations is a good indicator of future citations. To do this, and following the guidelines discussed by Liu and Fang [19], we add each author's total citations for five years. In the group of young researchers, we compare the correlation between the first and second five-year periods. We measure the correlation between the second and third five years of the life span for mid-age researchers. Finally, in the senior group, we measure the correlation between the third five-year and the fourth five-year period. Spearman's correlation coefficients are shown in Table 7.

Table 7 shows that the correlation increases as the authors become more experienced. For the senior group, all areas show a correlation factor close to 0.8. The strong correlation in this group indicates that the careers of these researchers are already consolidated. Therefore, authors with low productivity will have low productivity in the future, and authors with high productivity will maintain their productivity. The academic assessment based on citations in this segment is very informative. This correlation decreases in the mid-age group in all areas. It is striking that this segment is where the correlations are lower, even lower than in young researchers. A plausible explanation is that many authors stop researching between the second and third five years of the life span, which causes the correlation between past and future citation to decline. In young researchers, the correlation between the first five years and the second five years of the life span is more significant than in the mid-age. In some areas, the correlations are much lower than in others. For example, in young researchers at Mathematics, the correlation is only 0.31, the lowest in the entire study. This fact indicates that the first five years of the academic life of mathematicians is very uninformative. In other areas, this life span period provides much more information. For example, in Social Sciences and Chemistry, the correlation exceeds 0.7. The reasons why these areas reach this value are different. While in Social Sciences, young researchers have a slow growth (cold start phenomenon), which explains the high correlation; in Chemistry, the first years of activity show rapid growth with a peak of citations between years 4 and 6.

To analyze how sensitive the evaluation is concerning the year of the life span in which it is carried out, we repeated the previous analysis using sliding quinquenniums. If the evaluation is carried out in year i, the five-year sliding period is set five years before and five years after i. Then we measure the correlation between the sum of citations for both five-year periods. In this experiment, we span the evaluation year through the first ten years of the life span. The results are disaggregated by age group in Figure **??**.

Figure **??** shows that there are differences between areas. Concerning young researchers, Physics, Biochemistry, Engineering, Psychology, and Medicine show that between years 3 and 4, the correlation between past and future citations reaches its maximum value. Computer Science shows that the first two years of the life span are very uninformative for an evaluation, but in the third year, the correlation rises to 0.6. Both Social Sciences and Agricultural and Biological Sciences need a longer time to make a fair evaluation, requiring an observation window of at least four years for a young researcher. In Social Sciences, the fourth year is very informative, with a correlation of 0.8. Chemistry is the area in which the evaluation times could be shorter since high correlations are already observed between past and future citations from the second year. Mathematics is the most challenging area for young researchers, with the lowest correlation values between citations. When analyzing the mid-age and senior groups, all areas show an improvement in the correlation between past and future citations. In fact, in most areas, the correlation increases with the year of evaluation.

6. Conclusion

We have introduced a citation-based study that elucidates differences between areas, types of publications, and researchers' age groups. The main findings of this study are:

– Articles are a majority in all areas except Computer Science, corroborating the finding of Vrettas and Sanderson [27]. Conference papers cover 61% of Computer Science production, surpassing the coverage shown in other areas. The volume of citations captured by articles is a majority in all areas.

– Citations between areas show strong symmetric relationships between some specific areas (e.g., Biochemistry and Medicine) and a robust asymmetric citation pattern from some areas to other (e.g., Social Sciences to Psychology). Asymmetric relationships reduce citations received by articles in the same area.

– The average number of citations per paper has significant variations between areas and types of publications, corroborating previous studies [35]. The lowest amounts are seen in Social Sciences and Agricultural and Biological Sciences, contradicting the findings of Lariviere et al [23]. Reviews attract many citations, its performance in Physics being notable. Computer Science conference papers have a lower average number of citations than journal articles in this area. The citations to conference papers are only relevant in Computer Science, and only in this area an important fraction of the top-cited papers belong to this category.

– The life span of a paper, measured from the track of citations it receives over time, reaches its maximum value in the third year of the publication's life. Psychology and Social Sciences' life span has slower growth than in the rest of the areas, suggesting that these papers have a longer useful life.

– The life span of authors is very different between areas. While most areas show a weak fit to the mean number of citations, this indicator is more informative in some specific areas (e.g., Mathematics). The predictive capacity of citations, measured from the productivity correlation between consecutive five-year periods, is very informative in seniors and less informative in mid-age researchers.

– The evaluation of researchers based on citations has different informative levels depending on the year in which the observation is carried out. If the evaluation is

done in the early years of academic life, this evaluation is more informative in some areas than in others. As researchers gain experience, their citation-based evaluation is more informative in all areas.

This study has allowed us to identify some critical differences between areas, types of publication, and age groups of researchers that may be relevant when carrying out researchers' academic evaluations. Each area has its specificities, which must be taken into account when comparing scientific productivity. It is expected that the findings of this article will motivate further research that requires the development of more and better methodologies and metrics for the evaluation of scientific productivity, which should include the differences in citation patterns between areas, types of publication, and age groups of researchers.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Garfield E (1955) Citation index for science: a new dimension in documentation through association of ideas. Science 122(3159):108–111.
2. Moed HF (2002) The impact-factors debate: The isi's uses and limits. Nature 415(6873):731–732.
3. Glanzel W, Moed HF (2002) Journal impact measures in bibliometric research. Scientometrics 53(2):171–193.
4. Waltman L (2016) A review of the literature on citation impact indicators. J Informetrics 10(2):365–391
5. Hirsch J (2005) An index to quantify an individual's scientific research output. PNAS 102(46):16,569–16,572.
6. Finardi U (2014) On the time evolution of received citations in different scientific fields: An empirical study. J Informetrics 8(1):13–24.
7. Smith DR (2012) Impact factors, scientometrics and the history of citation-based research. Scientometrics 92(2):419–427.
8. Robien W (2009) Do high-quality 13c-nmr spectral data really come from journals with high impact factors? TrAC - Trends in Analytical Chemistry 28(7):914–922.
9. Stewart A, Cotton J (2013) Making sense of entrepreneurship journals: Journal rankings and policy choices. International Journal of Entrepreneurial Behaviour and Research 19(3):303–323.
10. Finardi U (2013) Correlation between journal impact factor and citation performance: An experimental study. J Informetrics 7(2):357–370.
11. Vanclay JK (2009) Bias in the journal impact factor. Scientometrics 78(1):3–12.
12. Campanario JM (2011) Large increases and decreases in journal impact factors in only one year: The effect of journal self-citations. J Assoc Inf Sci Technology 62(2):230–235.
13. Walters WH (2017) Citation-based journal rankings: Key questions, metrics, and data sources. IEEE Access 5:22,036–22, 053.
14. Braun T, Glanzel W, Schubert A (2006) A hirsch-type index for journals. Scientometrics 69(1):169–173.
15. Balaban AT (2012) Positive and negative aspects of citation indices and journal impact factors. Scientometrics 92(2):241–247.
16. Bergstrom C (2007) Eigenfactor: Measuring the value and prestige of scholarly journals. College & Research Libraries News 68(5):314–316.
17. González-Pereira B, Bote VPG, de Moya Anegón F (2010) A new approach to the metric of journals' scientific prestige: The sjr indicator. J Informetrics 4(3):379–391
18. Moed HF (2010) Measuring contextual citation impact of scientific journals. J Informetrics 4(3):265–277.

19. Liu X, Fang H (2020) A comparison among citation-based journal indicators and their relative changes with time. J Informetrics 14:1.
20. Pyke GH (2013) Struggling scientists: Please cite our papers! Current Science 105(8):1061–1066.
21. Liu M, Hu X, Wang Y, Shi D (2018) Survive or perish: Investigating the life cycle of academic journals from 1950 to 2013 using survival analysis methods. J Informetrics 12(1):344–364.
22. Shi D, Rousseau R, Yang L, Jiang Li A (2017) journal's impact factor is influenced by changes in publication delays of citing journals. J Assoc Inf Sci Technology 68(3):780–789
23. Lariviere V, Macaluso B, Archambault E, Gingras Y (2010) Which scientific elites? on the concentration of research funds, publications and citations. Research Evaluation 19(1):45–53.
24. Nederhof A, Van Leeuwen T, Clancy P (2012) Calibration of bibliometric indicators in space exploration research: A comparison of citation impact measurement of the space and ground-based life and physical sciences. Research Evaluation 21(1):79–85.
25. Prins AA, Costas R, van Leeuwen TN, Wouters PF (2016) Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. Research Evaluation 25(3):264–270.
26. Freyne J, Coyle L, Smyth B, Cunningham P (2010) Relative status of journal and conference publications in computer science. Commun ACM 53(11):124–132.
27. Vrettas G, Sanderson M (2015) Conferences versus journals in computer science. J Assoc Inf Sci Technol 66(12):2674–2684.
28. Glänzel W, Schlemmer B, Schubert A, Thijs B (2006) Proceedings literature as additional data source for bibliometric analysis. Scientometrics 68(3):457–473.
29. Meho LI (2019) Using scopus's citescore for assessing the quality of computer science conferences. J Informetrics 13(1):419–433.
30. Thelwall M (2020) Mendeley reader counts for US computer science conference papers and journal articles. Quant Sci Stud 1(1):347–359.
31. Kochetkov DM, Birukou A, Ermolayeva A (2020) The importance of conference proceedings in research evaluation: a methodology for assessing conference impact. CoRR abs/2010.01540.
32. Li X, Rong W, Shi H, Tang J, Xiong Z (2018) The impact of conference ranking systems in computer science: a comparative regression analysis. Scientometrics 116(2):879–907.
33. Qian Y, Rong W, Jiang N, Tang J, Xiong Z (2017) Citation regression analysis of computer science publications in different ranking categories and subfields. Scientometrics 110(3):1351–1374.
34. Yang S, Qi F (2020) Multidisciplinary comparison of proceedings papers and academic books based on altmetrics and citation. In: iConference 2020 Proceedings.
35. Ruiz-Castillo J, Costas R (2018) Individual and field citation distributions in 29 broad scientific fields. J Informetrics 12(3):868–892
36. Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences 105(45):17,268–17,272
37. Albarrán P, Crespo JA, Ortuño I, Ruiz-Castillo J (2011) The skewness of science in 219 sub-fields and a number of aggregates. Scientometrics 88(2):385–397
38. Crespo JA, Herranz N, Li Y, Ruiz-Castillo J (2014) The effect on citation inequality of differences in citation practices at the web of science subject category level. J Assoc Inf Sci Technol 65(6):1244–1256
39. Bensman SJ, Smolinsky LJ, Pudovkin AI (2010) Mean citation rate per article in mathematics journals: Differences from the scientific model. J Assoc Inf Sci Technol 61(7):1440–1463
40. Brzezinski M (2015) Power laws in citation distributions: evidence from scopus. Scientometrics 103(1):213–228
41. Moreira J, Zeng X, Nunes Amaral L (2015) The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. PLoS ONE 10(11)

7. Appendix

Table 8: Basic statistics of our repository. In articles we group up regular papers, articles in press, surveys, reviews, letters, erratum, reports, business articles, abstract reports, editorials, and notes that were included in any number of an indexed journal. Conference papers group up conference reviews and papers included in indexed proceedings.

Statistics	Value
# citable documents	4,504,779
# articles	3,849,858
# conference papers	595,122
# books and chapters	59,799
# authors	111,813
# citations	8,510,720
# subjects	27

Table 9: Distributions of authors, documents, and citations across Scopus subjects. Scopus covers 27 subject areas.

Scopus subject	# authors	# documents	# citations
Medicine	34974	1931976	3367647
Physics and Astronomy	11308	517723	1830612
Computer Science	9052	389411	398904
Biochemistry, Genetics and Molecular Biology	8418	278396	980208
Mathematics	7520	202670	190678
Engineering	7367	326294	256990
Psychology	6145	132199	320362
Agricultural and Biological Sciences	3886	107716	39276
Social Sciences	3691	36614	30053
Chemistry	3307	179808	440192
Neuroscience	2375	53251	243522
Materials Science	2331	122090	73575
Pharmacology and Toxicology	2076	45877	80618
Earth and Planetary Sciences	1778	39962	21390
Business and Management	1380	25396	65818
Economics	1286	32753	120277
Environmental Science	1132	32638	26989
Energy	969	8129	1961
Immunology and Microbiology	744	3195	2324
Nursing	542	5676	860
Chemical Engineering	465	18366	10682
Arts and Humanities	373	6573	5873
Veterinary	257	6875	1834
Multidisciplinary	241	35	9
Decision Sciences	80	295	90
Health professions	76	14	2
Dentistry	40	620	74