Publication Practices in the Argentinian Computer Science Community: A Bibliometric Perspective

Daniela Godoy · **Alejandro Zunino** · **Cristian Mateos**

Abstract The Computer Science (CS) community has been discussing, for some time now, the role of conferences as publication venues. In this regard, computer scientists claim to have a long-standing tradition in publishing their research results in conferences, which are also recognized as being different to events in other disciplines. This practice, however, contrasts with journal driven publication practices which are the prevailing academic standard. Consequently, the assessment of the quality of CS conferences with respect to journals is a recurrent topic of discussion within evaluation boards in charge of assessing researchers' performance. Even when agreements are feasible inside the discipline, they are often subject to the scrutiny in the context of multi-disciplinary evaluation boards –usually ruled by standard bibliometrics– in which CS researchers compete for obtaining scholarships, positions and funding. The Argentinian CS community is not an exception in this respect. In this paper, we present a study of the publication practices of the Argentinian CS community, their evolution over time and, more importantly, the impact they achieved in terms of citations. The findings of this study are good basis for understanding the publishing practices of our community, promoting future discussions as well as supporting the community positions regarding these issues.

Keywords Computer science · publication practices · conferences · journals.

1 Introduction

A common practice within the Computer Science (CS) community is to publish papers in conferences. These events allow researchers not only to exchange new ideas and socialize, but unlike other disciplines, they serve as venues to publish (sometimes mature) research results. A recent study [21] shows that in average 35% of the papers in CS are published in conferences. Furthermore, many CS conferences usually publish fully refereed papers that are often perceived within the CS community to be of equal quality and visibility as

ISISTAN Research Institute - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), University Campus, Paraje Arroyo Seco (BBO7001B), Tandil, Buenos Aires, Argentina
E-mail: daniela.godoy@isistan.unicen.edu.ar

those from journals. The rationale behind conference publication in CS is that it ensures fast dissemination and high impact[1]. Besides, papers in leading CS conferences seem to match the impact of papers in mid-level journals [7].

This has, however, given rise to controversies when it comes to assessing researchers' performance [19,10,9,8,4]. Although within the CS community the value of conference papers is recognized, in most other disciplines journal papers are the main publication venue. This causes many problems when evaluating CS researchers, especially in multi-disciplinary evaluation boards [10,9]. First, establishing special criteria for CS is often difficult or impossible, because CS publication practices are not shared by the vast majority of scientists across all disciplines. Even different subareas of computer science have significantly different publication practices [21]. Second, while raw bibliometric evaluation has its problems [13], it is an unavoidable fact that its use as an indicator of the impact or quality of research over the scientific community is widespread [14]. Some examples of popular bibliometrics are the Impact Factor (IF), h-index and number of citations. Moreover, bibliometry is usually done using the standard journal databases, where there are a few CS conferences. As a consequence, conference-driven publication practices of CS are severely punished by this evaluation criteria [19,9].

Particularly, in South America, the Argentinian and the Chilean national research councils (CONICET[2] and CONICYT[3], respectively) ask their CS researchers to publish their results preferably in journals indexed in the well-known Science Citation Index or Science Citation Index Expanded from Thomson-Reuters. Conference papers, on the other hand, are given much less importance or, depending on the venue, they are even not considered at all when evaluating researchers. In an attempt to better assess conference papers, some South American councils have adopted systems to stratify publications in these venues, such as Brazil's Qualis[4], which is similar to the Australian's research ranking for journals [18]. Qualis classifies 1,650 conferences in six categories based on their h-index calculated from Google Scholar citations. Another typical indicator to judge the quality of conferences has been their rejection rates, which however have a low coefficient of correlation with papers impact in terms of citations [7]. Therefore, it is difficult to judge conference paper quality based on conference rejection rates.

In light of this problem, some researchers have reacted, pleading for changing the way in which results are published in CS [4,19,10]. Although there is a clear consensus on the social value of conferences, many serious concerns have been raised about the role of conferences as venues for publishing research. These concerns include inbreeding around specific publication venues [20], incentive for producing least-publishable units in hot topics instead of breakthroughs [4], papers with hypotheses that cannot be empirically confirmed and because of that would not be considered eligible for publication in a journal [15], few and low quality evaluations caused by the strict and short conference deadlines [10,9,11] as well as overloaded program committees [2], reviewers' criteria variability [1] and space limitation hindering reproducibility [3].

Some reactions include that of Vardi [19], who states that CS researchers should consider journals as primary publications forums, and conferences as secondary ones. He exemplifies this recommendation by drawing an analogy with a joke where a driver (i.e., the CS

---

[1] Evaluating computer scientists and engineers for promotion and tenure: `http://www.cra.org/reports/tenure_review.html`.

[2] `http://www.conicet.gov.ar`

[3] `http://www.conicyt.cl`

[4] `http://www.capes.gov.br/avaliacao/qualis`

community) returning from home after having too many drinks, hears a warning over the radio about a car going in the wrong way on the highway (i.e., publishing only in conferences) and thinks the rest of the cars (i.e., other disciplines) are actually going in the wrong direction. Indeed, in other disciplines such as Physics, Biology and Chemistry, journals are the main, and sometimes the only, accepted way of disseminating research results. A more extreme position is that of [4], which proposes to de-emphasize the publication role of conferences to let them play a social role exclusively. In his view, events should not publish proceedings anymore, moving publication results to specialized journals. A more intermediate position in the line of [19] is the Proceedings of the VLDB Endowment (PVLDB) journal[5], the flagship publication of the Very Large Database (VLDB) community. Papers submitted to PVLDB undergo a regular, rigorous review process, and accepted papers are allowed to be presented at the next edition of the VLDB Conference. A similar approach is proposed by other researchers and adopted by an increasing number of conferences [10, 9]. This approach essentially inverts the publication process followed in many areas of CS –enforced by the idea of *journal special issues*– where journal publications are extensions of papers previously published in conferences by the same authors. For example, in Computer Vision research, it was determined that 30% of the journal papers are based on previous results published in conferences papers, called "priors" [3].

In the middle of this heated debate and changing publication practices, CS researchers, referees and evaluation boards have little factual and unbiased data to make decisions. Motivated by these problems, we conducted an analysis of research outputs by focusing on CS researchers from CONICET, Argentina's main national scientific research council. In South America, Argentina has a long tradition in hard sciences –it concentrates 3 out of 4 Nobel prizes awarded in Chemistry and Physiology or Medicine– and a relentlessly growing CS community due to nationwide promotion programs. The goal of this paper is to quantitatively determine the CS publication practices in Argentina, their evolution over time and their relative impact. Having a clear picture of these issues is crucial as CONICET is a multi-disciplinary organization, where research funding, grants and positions are open for all scientific disciplines so that multi-disciplinary evaluation boards, competition and bibliometry are the rule.

In the rest of the paper we try to answer the following questions regarding the Argentinian CS community publication practices, which at the same time will guide the bibliometric analysis:

– Where do CONICET CS researchers publish? In conferences or journals? Considering that CONICET categorizes researchers according to their performance: does the publication practices vary according to the category? how have these practices evolved over time?
– Following the common idea that conferences in CS provide more visibility than journals: have conference papers published by this community been more cited than journal papers? can this fact be verified with the publications of CONICET CS researchers?

For this analysis, we gathered data from two well-known journal and conference publications repositories, namely Scopus[6] and Google Scholar[7]. Section 2 describes the methodology used to obtain bibliographic information and statistics of the data collected. Section 3 focuses on a bibliometric analysis of the CONICET CS community and a citation analysis

---

[5] http://www.vldb.org/pvldb/pvldb-faq.html

[6] http://www.scopus.com/home.url

[7] http://scholar.google.com/

of scientific production of this community. Finally, conclusions drawn from this study are stated in Section 4.

2 Data

Since the assessment of the importance of conferences and journals is a common concern of CS evaluation boards, the analysis was carried out considering CONICET researchers, whose performance is regularly subject to evaluation. Even though Ph.D. students with fellowship grants from CONICET were not included in the analysis as they have still very few articles published, either in conferences or in journals, their scientific production and impact in terms of citations are indirectly considered through their advisors' production, i.e., papers in which the PhD students play the role of co-authors. First, in Section 2.1, we introduce CONICET organization, the aim of each researcher category and the group of researchers analyzed in this study. Section 2.2 summarizes the process of bibliographic data collection for the selected researchers.

2.1 CONICET Researchers

CONICET[8], the National Scientific and Technical Research Council, is the main organization in charge of the promotion of Science and Technology in Argentina. Founded in 1958, the goal of this governmental agency is to boost and implement scientific and technical activities in all fields of study throughout the country.

In order to promote full-time and permanent commitment of researchers to scientific and technological work, CONICET implemented a Scientific and Technological Research Career, whose members carry out their activities in Universities and other academic, scientific and technological organizations nationwide. This career consists of fives stages or categories, and promotions are achieved through rigorous performance evaluations. Usually, researchers start in the career as research assistants after they have finished a period of postdoctoral studies. The requirements and aim of each category of researcher, as defined by CONICET, are:

Research Assistant: requires having done scientific research, any technological development or creative work, usually in the context of a Ph.D. thesis, demonstrating skills to carry out research under the guidance or supervision of others, and possessing the necessary technical skills to perform autonomous problem solving;

Research Associate: requires having acquired the ability to plan and execute research or development themselves as well as to work effectively in teams;

Independent Researcher: requires having made significant original works in scientific research or development, and to be able to choose the topics, plan and carry out investigations independently or having distinguished themselves as a member of a team of recognized competence;

Principal Researcher: requires having done extensive scientific or technological development of high originality backed by publications and influence of the researcher's work in the progress of his field. Researchers in this category must also possess the ability to train disciples and lead research groups.

---

[8] http://www.conicet.gov.ar/

Superior Researcher: requires having done extensive work in scientific research or technological development, so that the researcher is among the core of internationally recognized specialists in his field. Researchers in this category should also have demonstrated mentoring and direction of research centers.

CONICET comprises four general areas of knowledge to enable comprehensive development of scientific and technological research: a) Agrarian, Engineering and Material Sciences b) Biological and Health Sciences, c) Exact and Natural Sciences, and d) Social Sciences and Humanities. The first of these areas includes Information & Communications Technologies (ICT) as a sub-area, which in practice concentrates the Argentinian Computer Science community. From now on we will use ICT and CS interchangeably.

To analyze the research outputs of CS people, a list of researchers belonging to the ICT sub-area was crawled from the CONICET Web site with the help of a tool using text mining techniques specially developed for this purpose. Table 1 shows the number of researchers in the database as of April 2013. We considered that this sample of researchers is not only representative of the entire CS community of Argentina (there are few exceptions of non-CONICET researchers doing high-quality research and its scientific production is indirectly considered because of their co-authorships with CONICET members) but they are most likely affected and interested in this discussion.

Table 1: CONICET researchers in the ICT area used for the analysis

Category	Number of researchers
Research Assistant	40
Research Associate	27
Independent Researcher	8
Principal Researcher	5
Superior Researcher	1

CS researchers have been immersed in different areas of CONICET for many years, but it was not until the early 2000s that the own entity of CS was recognized and the ICTs sub-area was created. Since then, the discipline has shown steady growth in the number of members. Considering the 81 members the ICT area has at present, Figure 1 gives an idea of their insertion in CONICET since 1998. Specifically, the figure shows how many of these researchers were active in a certain year, defining "active" as those having at least one paper published during that year (the papers considered were those extracted from Google Scholar and filtered as described in Section 2.2).

Using the CS areas identified in [21], the analyzed CS researchers can be categorized as follows: 21 in Artificial Intelligence, 15 in Software Engineering, 10 in Bioinformatics, 9 in Management Information Systems, 8 in Distributed Computing, 5 in Operational Research and Optimization, 4 in Compilers and Programming Languages, 3 in Theory, 2 in Machine Learning and other 1 in each of the four areas Communications and Networking, Human-Computer Interaction, Image Processing and Computer Vision, and Security. As stated in [21], the distinction in CS areas is very important since they imply also different publication practices. However, the goal of this paper is to assess the overall impact of conferences and journals papers, irrespective of the area.
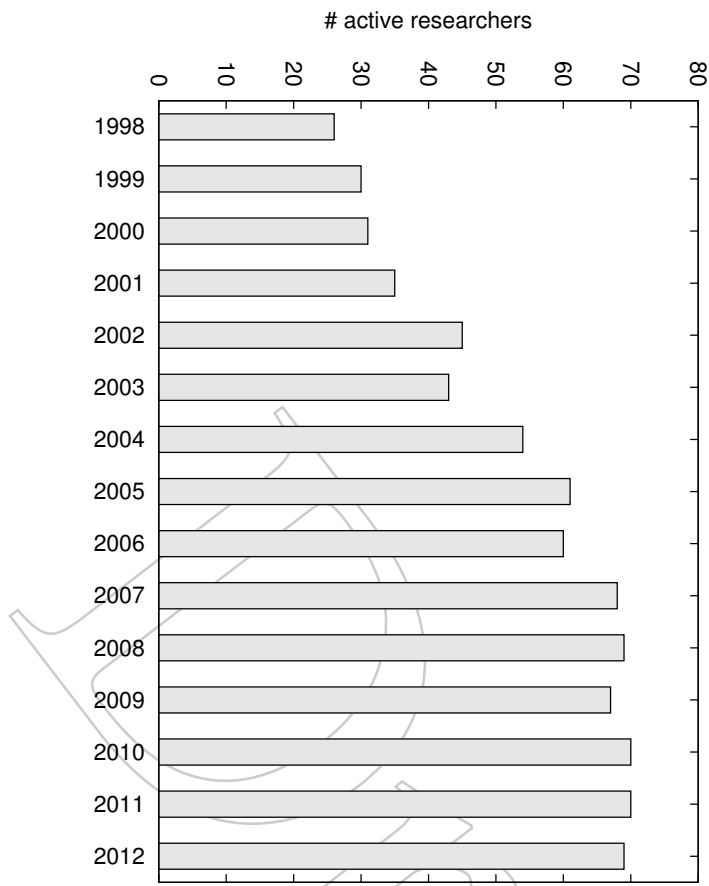
Fig. 1: CONICET active CS researchers (January 1998- December 2012)

2.2 Bibliometric Information Sources

In order to analyze the scientific production of CONICET CS researchers, we extracted information of two different bibliographic databases, namely Scopus[9] and Google Scholar[10], which have distinctive nature and characteristics. Scopus, officially named SciVerse Scopus, is a bibliographic database containing abstracts and citations to academic journal and conference articles. Google Scholar is a search engine that indexes the full text of scholarly literature, and also includes theses, papers stored in Websites from academic institutions, and even patents. In the analysis reported in Section 3.1, we profiled CS CONICET researchers by using both Scopus and Google Scholar data. Conversely, for building the results reported in Section 3.2, we relied on data gathered from Google Scholar only since it has a much more wider indexing scope compared to Scopus.

---

[9] http://www.scopus.com

[10] http://scholar.google.com

2.2.1 Scopus

Scopus allows searching for author profiles, showing their affiliations, number of publications and bibliographic data, along with references and details on the number of citations each published document has received. For each researcher, the data obtained from Scopus in this study included: number of documents, number of journal articles (classified as *Article*, *Review*, *Article in Press* in Scopus), number of conference papers (classified as *Conference Paper* in Scopus), total number of citations and h-index.

In some cases, an individual's information is split into two or more profiles as a consequence of the automatic summarization method used by Scopus. In such situations the number of documents and citations of all profiles corresponding to the same researcher was added, whereas the highest h-index was considered.

2.2.2 Google Scholar

In Google Scholar authors can create their profiles featuring fields of interest, publications and citations, with an e-mail address usually linked to an academic institution. Google Scholar automatically calculates and displays the individual total citation count for a profile, the h-index and the i10-index, among other information.

For each researcher considered in the study, all publications were gathered from Google Scholar by querying the search engine with the complete researcher name. The results were filtered according to the study goals, excluding non-English articles since their audiences –and, consequently, their impact in citations– are limited. In a second step, papers from national conferences were removed, independently of their language, given the limited diffusion of publications in these events. Since the intended goal of this study was to evaluate and compare the impact in terms of citations of conference and journal papers, to include Spanish-written papers would have been unfair to conferences as national papers would contribute with very few citations. Lastly, documents in neither conferences nor journals, such as book chapters, theses and technical reports, were also eliminated since their scientific value was not under examination. For each document, the following information was collected: article title, link to the article, number of citations, publication year, list of authors and publication venue.

Once the information of each paper was extracted from Google Scholar, it was automatically classified into conference or journal paper based on hand crafted rules applied to the article links. Basically, the link allows accessing publisher Websites, which provide article meta-data via meta-tags embedded within the Web page of each article[11]. The information crawled from the Web pages includes title, authors, media name, ISSN/ISBN and DOI, among other data. The ISSN of each paper enabled us to automatically cross-reference it with impact factor lists from the Journal Citation Reports (JCR) of the corresponding publication year. Ultimately, papers with non extractable information starting from their links were manually classified.

As a result of the crawling and filtering processes, we obtained 1872 papers published by the 81 researchers involved in the analysis, where the oldest paper is dated 1979 and the most recent ones are from the beginning of 2013. Out of this total number of articles, 1114 correspond to conferences (59.50%) and 758 to journals (40.50%).

---

[11] http://scholar.google.com/intl/es/scholar/inclusion.html#indexing

3 Findings

In this section, we first present a bibliometric analysis of the Argentinian CS community aiming to have a picture of the scientific production of the overall community. Based on this analysis, described in Section 3.1, we look for answers to the first research question stated: where Argentinian CONICET CS researchers publish, in conferences or journals? and how scientific production of researchers relates to their CONICET categories?. Then, in Section 3.2 we analyze the impact of research produced by this community in terms of citations received according to the publication venues in order to evaluate the effectiveness of the distinctive publication practices found.

3.1 Bibliometric Analysis

Figures 2(a) and (b) illustrate the total number of publications of the CS community separated into papers published in conferences and journals with data obtained from Scopus and Google Scholar, respectively. Both figures show the number of publications per researcher in each CONICET category for the two types of venues. In the case of Google Scholar the number of publications are those considering the filtering explained in subsection 2.2.2, so it is in most cases lower than the total number of documents in the actual Google Scholar researchers' profiles. It is worth noting that, for the sake of significance, we have excluded from the figure the Superior Researcher category since it has only one member.

Figure 2 as well as most of the figures in the rest of the paper are box-plots. A box-plot represents a statistical distribution of values, in which a box is drawn around the region between the first and third quartiles, with a horizontal line at the median value. Whiskers extend from each end of the box for a range equals to 1.5 times the interquartile range. Points lying outside the range of the whiskers are considered outliers and are drawn individually.

Note that the distribution of citations to papers published in CS publication venues is skewed and the tail of the citation distribution for CS papers has a power law behavior [6]. As a consequence, the mean is not an appropriate measure of the central tendency of citations received by articles. Therefore, we use the median, which is more appropriate in case of skewed distributions [6].

The number of journals in both cases is similar, but there is an important difference in the number of conferences reported by both sources, because Scopus does not contain as many conferences as Google Scholar. From these figures, as expected, it can be seen that researchers in the lowest category have fewer papers in journals compared to researchers in other categories. This is due to several reasons, including their lower production per year rate and journal papers processing times. Moreover, this tendency is not clear for Associate and Independent researchers, but Principal researchers have more papers in journals precisely because of their annual production rates and higher seniority. The average number of publications, papers in conferences and journals as well as the average number of produced articles per researcher is summarized in Table 2.

In addition to the overall number of papers in each venue, we were interested in observing how the publication practices in the Argentinian CS community had evolved. With this purpose, we calculated the average number of publications in each venue per researcher in a year using Google Scholar data and based on the number of active researchers per year shown in Figure 1. Figure 3 depicts the evolution of averages over time. Note that standard deviations are wide as the average accounts for researchers in all categories, from Research Assistants having a few papers a year to Principal Researchers with a substantial annual
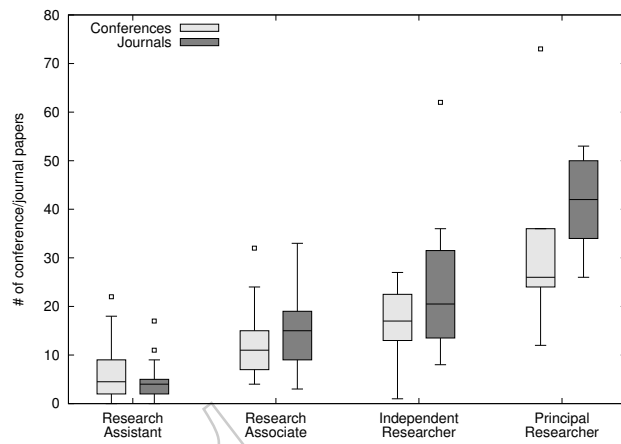
(a) Scopus
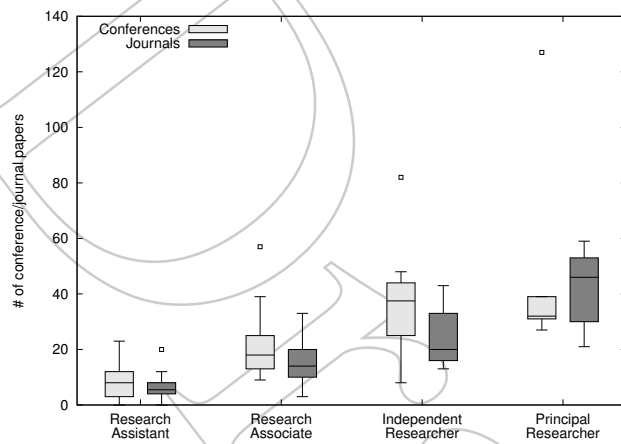
(b) Google Scholar

Fig. 2: Number of articles published per venue by CONICET researcher category

scientific production. It can be deduced from the figure that the gap between conference and journal articles published annually by researchers has been steadily decreasing. In fact, in the last three years considered, both values are quite near, being almost identical in 2012. During that year, researchers published an average of 2.81±1.82 articles in conferences and 2.80±1.50 in journals.

Several CS researchers [19,10,9,8,4] had pointed out that CS should switch as soon as possible to publishing in journals rather than defending a position that is not shared by the vast majority of scientists across all disciplines. Common arguments from their counterparts are that the pressure for moving to JCR indexed journal publications will lead researchers to choose low tier journals. Moreover, some researchers argue that top conferences surpass the impact of journals in the bottom half of the JCR [7].
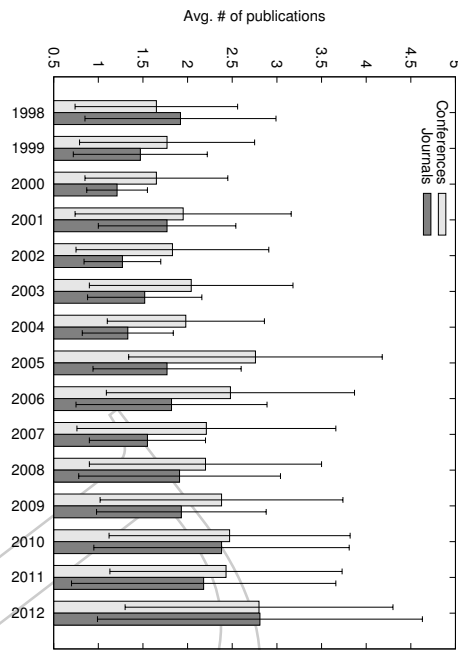
Fig. 3: Average number of publications per researcher (1998-2012)

Table 2: Summary of scientific production of CS researchers

Category	Scopus			Google Scholar		
	publications	conferences	journals	publications	conferences	journals
Research Assistant	10.63±4.94	6.23±4.43	4.33±2.01	14.80±6.23	8.73±5.35	6.08±2.50
Research Associate	28.37±8.26	12.63±4.95	15.37±6.31	35.93±10.13	20.44±7.37	15.48±5.90
Independent Researcher	43.13±16.13	16.63±5.88	25.13±12.41	62.13±16.38	37.88±14.13	24.25±9.06
Principal Researcher	77.80±19.28	34.20±16.24	41.00±8.80	93.00±32.00	51.20±30.32	41.80±13.04

To have some insights about the quality of published articles by CONICET CS researchers, both in journals and conferences, we disaggregated the total number of papers according to impact factors (see explanation below) in one case and conference publisher in the other one. The goal was to determine whether such a shift even started in the Argentinian CS community and, if so, whether it has led to an increase in low impact journal publications. In other words, we tried to find out to what extent researchers published their articles in high impact journals and conference proceedings backed by renowned editorials.

Figure 4(a) shows the number of articles published in journals and their distribution in impact factor tiers. Figure 5(a) shows the articles from events such as conferences, workshops and symposia, available in digital libraries such as the SpringerLink database (only LNCS series), IEEE Xplore, AAAI and the ACM Digital Library (including both ACM and non-ACM events). It is worth noting that there is neither a commonly agreed list nor a bib-

liometric measure for events in CS to separate them in tiers. On the other hand, journals are listed by Thompson Reuters in its yearly JCR and classified by their *impact factor* (IF), a well-known measure reflecting the average number of citations to articles published in the journal during the two preceding years. A first observation that can be made considering both figures together is that journals exhibit an increasing tendency, while conferences remain in the same level since 2005 (in spite of the growing number of members in the community as depicted in Figure 1).

Figure 4(b) shows the proportion of articles in each journal tier out of the total number of journals published. Clearly, the proportion of non-indexed[12] journals has decreased over time on account of the growing importance of publications included in the JCR. Proportionally, there has been a clear increment of publications in higher impact factor journals, IF>=2 as well as 1<=IF<2. In the two inferior tiers, 0.5<=IF<1 and IF<0.5, the percentage of published papers has not significantly changed in the last 10 years. These data suggests that not only the number of articles published in journals has increased during the last years, but also that researchers have been targeting more influential journals as measured by the IF. This is likely a consequence of doing more important, higher impact work during the years.

In regard to conference papers, the largest fraction of articles have been published by Springer-Verlag as part of the Lecture Notes in Computer Science (LNCS) series, including its sub-series Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), as illustrated in Figure 5. LNCS volumes are a special case of publication venue, since they were included in the JCR until 2006. However, as from October 2006, Thomson decided to move LNCS from JCR to the Conference Proceedings Citation Index because the LNCS series does not, strictly speaking, publish journal papers, but conference proceedings. The exclusion of the LNCS series from the JCR implied a devaluation of LNCS publications for most councils. In spite of Thomson's decision, LNCS publications seem to be still highly appreciated when selecting conference venues by the analyzed researchers. These numbers, however, should be analyzed in light of the number of conferences covered by each publisher, shown in Figure 6, as extracted using the advanced search mechanisms of the corresponding Web sites. The number of articles published in IEEE conferences by Argentinian researchers was not in tandem with the important increase in coverage of IEEE proceedings. ACM and AAAI publications have neither important variations over the years nor an important trend difference with the number of proceedings published. Instead, the number of LNCS articles grows in spite of the fact that the number of volumes of this series is quite stable in the last years, confirming a slight preference for this venue.

3.2 Citation Analysis

Traditionally, the impact of scientific publications is measured by the number of citations they receive [12]. Likewise, and despite its known flaws [12], the impact of the research conducted by individual authors, institutions and publication venues is still assessed through citation counts. Then, irrespective of the publication forum, citations are an important tool to assess CONICET researchers performance. The citation analysis of this section was conducted using the whole article database, this is the 1872 papers corresponding to the 81 researchers involved (1114 conferences and 758 journals).

Citations from Google Scholar were considered in this study since Scopus covers only a fraction of scientific publications in computer science, containing few conference articles

---

[12] By "non-indexed" we refer to journals not included in the Thomson's JCR

compared with journals. Furthermore, Scopus misses many citations originated in not covered articles, including numerous conferences and other type of documents, such as theses and technical reports.

In the previous section, we focused on where CONICET CS researchers publish, finding that journals have gradually tend to prevail over conferences in the last years. Another view of the data can be seen in Figure 7(a), which depicts the proportion of papers published in conference proceedings and journals every year out of the total scientific production of CS researchers. Figure 7(b) shows the percentage of citations, out of the total citations received by papers published in a year, grouped by conferences and journals.

From Figures 7(a) and (b) we can infer that in most years, even those in which conference papers outnumber journal ones (all years except the two last ones), the proportion of citations to papers appearing in journals is superior to the percentage of citations to papers in conference proceedings (for example, in 2006 only 33% of the papers were published in journals, but almost half of the citations correspond to those papers). Indeed, considering the complete period from 1998 to 2012, an average of 38.40%±5.32 of the papers were published in journals, but an average of 50.32%±9.08 of the citations of the corresponding years were received by these papers. Since 2010, papers in journals have received more than 60% of the citations.

This result is rather surprising, because conferences in CS are often considered as the best venues for gaining visibility and citations. It also contradicts the popular idea that conference publications in CS ensure high impact [13,7], in terms of citations, compared to journal papers.

Going a step further into this analysis, Figure 8 presents the distribution of citations received by papers published in journals and conferences over the years. Only in 3 years, out of the 15 years analyzed, the median number of citations received by conference papers surpassed that of journal papers. Outliers are not depicted in the figure for the sake of clarity, but the most cited journal paper has 513 citations and the most cited conference paper has been cited 489 times. In the top-10 cited papers, 7 correspond to journals and 3 to conferences. This view of citation counts confirms the previous results: conference papers, at least in the analyzed data, do not have higher impact than journal papers.

The statistical significance of per-year differences was assessed with the two-sample Kolmogorov-Smirnov test (KS-test) [17], a nonparametric test that compares the cumulative distribution function of two independent samples. KS-test makes no assumption about the distribution of data, which is particularly suited for skewed distributions as the produced by citation counts [6]. The test protocol is based on the principle that if there is a significant difference at any point along the two functions, it can be concluded that the samples are likely to be derived from different populations [16]. The test statistic, denoted D, is defined by the point that represents the greatest vertical distance at any point between the two cumulative distribution functions.

The KS-test was performed over each year conferences and journals distribution of citations, obtaining the D statistics shown in Figure 8(b) with the p-values indicated at the top of the bars. Basically, the p-value indicates what is the probability that the two cumulative distribution functions would be as far apart as observed if the two samples were randomly sampled from identical populations. For 2006 on, the small values of p indicate that the citation numbers of journals is above the citations of conferences with a statistically significant difference. Moreover, if we consider the citations to the 1019 articles published in conferences and the 660 in journals in the complete period 1998-2012, the maximum discrepancy between the two cumulative distribution functions is $D = 0.1341$ with $p = 0.000$.

The publication practices of researchers varied according to the reasons exposed before, either in favor of conferences or journals, and sometimes influenced by external factors, such as the traditions in the place where they received their academic training (e.g., where they got their PhD degree, country, institution, etc.), the publication practices of the CS sub-area (for example, 70% of the papers in bioinformatics are published in journals, while 83% of the papers in compilers and programming languages are published in conferences [21]), the available economical resources (e.g., for attending conferences), the evaluation mechanisms they have been subject to, among others.

Individualizing the citation analysis, Figure 9 aims to answer the question about the publication practices of CS researchers and whether these practices influence the number of citations each individual researcher receives. The x axis of the figure contains researchers sorted by their productivity in conferences, i.e., the proportion of their academic production published in conferences. Thus, researchers to the left have a more journal-oriented profile, researchers in the middle have a mixture of conferences and journals, and researchers to the right have a conference-oriented profile. We omitted Assistant researchers in the figure because they have a few papers and thus a yet undefined profile.

Consistently with the previous finding, the figure shows that in most cases the proportion of citations to conference papers is inferior to the proportion of papers actually published in such kind of venue. In the figure, the crossed line depicting the percentage of published conference papers surpasses the light gray bars, meaning that in terms of citations it is more rewarding to publish in journals than in conferences. Only for 5 out of 41 researchers being CONICET Research Associates or up (87% of these researchers) the total percentage of citations to conferences papers (light gray bars) overpass the crossed line depicting the proportion of published conference papers. All in all, researchers do not receive citations to conference papers to the same extent they have published on them. The results of a Kolmogorov-Smirnov test to the citation counts per year (absolute number of citations, not proportional to the whole researcher production) to conferences and journals of these 41 researchers gives a maximum distance between the cumulative distribution functions of $D = 0.3414$ with a corresponding $p = 0.012$, which implies a statistically significant difference.

Table 3: CONICET CS researchers and their publication profiles

Category	Journal-oriented	Mixed	Conference-oriented
Research Associate	8	10	9
Independent Researcher	3	1	4
Principal and Superior Researchers	3	2	1

This individualized analysis of citations leads us to another question: how much a conference or journal paper pays in terms of citations to each researcher profile? For example, a conference-oriented researcher invests in travel and attendance time plus money on conference papers, which tend to be short and only contain the most exciting part of the research, with the promise of high impact and rapid diffusion of results [13, 8, 7]. On the other hand, journal papers tend to be longer and may contain much more detailed information allowing replication and full understanding of the results [15], and thus their production consumes more effort and time. How has each venue rewarded researchers in terms of citations in the experience of the CONICET CS community?

Trying to answer this question, Figure 10 shows the proportion of citations received by conference and journal papers over the total number of citations to each researcher, grouped according to the researchers' publishing practices or profiles. For the same reason as in Figure 9, we omitted Assistant researchers. With the remaining 41 researchers, we defined three profiles according to the ratio of papers a researcher published in journals over conferences: journal-oriented (below percentile 0.33), conference-oriented (above percentile 0.66) and mixed (in between percentile 0.33 and 0.66). The dashed line indicates the average percentage of conference publications for each profile. Then, we calculated the proportion of citations received by conference and journal papers in each profile. Table 3 summarizes the number of researchers in each profile according to their categories.

On the left of the figure, researchers who have journal-oriented profiles received 78% of their citations from journal papers and 22% from conference papers (i.e., 56% less). On the other extreme, conference-oriented researchers receive 31% of their citations from journal papers and 69% from conference papers (i.e., 38% less). As expected, researchers in the mixed profile receive 63% of their citations from journal papers and 37% from conference papers. The main conclusion we can draw from this figure is that despite the visibility conferences in computer science deem to provide, the proportion of citations received by conference papers in the conference-oriented profile is less than those of journal papers in the journal-oriented profile. This observation is in line with the results shown in Figure 9. In addition, the percentage of citations to conference articles is consistently below the actual proportion of articles published in that venue by researchers in all profiles. These results are consistent with the study presented in [5], where the author selected papers indexed in the Digital Bibliography and Library Project (DBLP), whose titles included popular keywords such as "genetic algorithms" or "Internet". He discovered that 78% of these papers were published in conferences, but the journal papers in DBLP have on average 4.51 citations, as opposed to 0.71 citations per publication for conference papers.

4 Conclusions

The analysis carried out throughout this paper led to several interesting conclusions. Specifically, with respect to the first question stated at the beginning of the paper, we could observe that in the big picture of the analyzed articles there is a uniform tendency: across the years, the gap between the number of journal and conference articles published has decreased considerably. This suggests that the CONICET CS community is already shifting to journal publications. In fact, in the last years, the percentage of high-impact journal papers has increased as well, and thus it can be seen that there is an incipient trend in the analyzed community towards publishing results in journals with high impact factor.

Despite the fact that CONICET evaluation policies favor indexed journals over non-indexed journals and conferences, there are many cases of CS researchers for which the strategy of concentrating their scientific production in prestigious conferences has not prevented them from reaching the upper categories in the CONICET scientific career. Nevertheless, many researchers clearly prefer publishing their research in indexed journals. In the future, this self-inflicted "publish indexed papers or perish" effect, together with the monetary costs associated to conference publications, might further lower the interest of CONICET researchers in conferences. It is worth noting that this would create the opportunity to partially redirect monetary resources for other purposes, such as scholarships, equipments, etc.

Another factor that enforces this situation is the impact in terms of citations of published papers by the analyzed researchers. We found that, for 87% of the researchers, it is more rewarding in terms of citations to publish in journals than in conferences. Broadly, these results suggest that, if CS researchers seek to increase citation count, they should concentrate their efforts on publishing in journals. This has the added benefit of being aligned with the unavoidable fact that the use of bibliometrics is widespread to assess the impact or quality of research, particularly in multi-disciplinary organizations such as CONICET. More importantly, the shift of CS towards journal driven publication practices, the prevailing academic standard, will make computer scientists more competitive with other well established disciplines.

References

1. Anderson, T.: Conference reviewing considered harmful. ACM SIGOPS Operating Systems Review **43**(2), 108–116 (2009)
2. Birman, K., Schneider, F.B.: Viewpoint: Program committee overload in systems. Communications of the ACM **52**(5), 34–37 (2009)
3. Eckmann, M., Rocha, A., Wainer, J.: Relationship between high-quality journals and conferences in computer vision. Scientometrics **90**(2), 617–630 (2012)
4. Fortnow, L.: Viewpoint: Time for computer science to grow up. Communications of the ACM **52**(8), 33–35 (2009)
5. Franceschet, M.: The role of conference publications in CS. Communications of the ACM **53**(12), 129–132 (2010)
6. Franceschet, M.: The skewness of computer science. Information Processing & Management **47**(1), 117–124 (2011)
7. Freyne, J., Coyle, L., Smyth, B., Cunningham, P.: Relative status of journal and conference publications in computer science. Communications of the ACM **53**(11), 124–132 (2010)
8. Halpern, J.Y., Parkes, D.C.: Journals for certification, conferences for rapid dissemination. Communications of the ACM **54**(8), 36–38 (2011)
9. Hermenegildo, M.: Conferences vs. journals in CS, what to do? Evolutionary ways forward and the ICLP/TPLP model. In: Dagstuhl Seminar 12452 - Perspectives Workshop: Publication Culture in Computing Research (2012)
10. Jagadish, H.V.: The conference reviewing crisis and a proposed solution. ACM SIGMOD Record **37**(3), 40–45 (2008)
11. van Leeuwen, J.: Where to send your paper? In: Dagstuhl Seminar 12452 - Perspectives Workshop: Publication Culture in Computing Research (2012)
12. Lindsey, D.: Using citation counts as a measure of quality in science measuring what's measurable rather than what's valid. Scientometrics **15**(3-4), 189–203 (1989)
13. Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J.: Viewpoint: Research evaluation for computer science. Communications of the ACM **52**(4), 31–34 (2009)
14. Moed, H.F.: Citation Analysis in Research Evaluation, *Information Science and Knowledge Management*, vol. 9. Springer (2005)
15. Montesi, M., Owen, J.M.: From conference to journal publication: How conference papers in software engineering are extended for publication in journals. Journal of the American Society for Information Science and Technology **59**(5), 816–829 (2008)
16. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press (2011)
17. Smirnov, N.V.: Estimate of deviation between empirical distribution functions in two independent samples. Bulletin Moscow University **2**, 3–16 (1939)
18. Vanclay, J.K.: An evaluation of the Australian Research Council's journal ranking. Journal of Informetrics **5**(2), 265–274 (2011)
19. Vardi, M.Y.: Conferences vs. journals in computing research. Communications of the ACM **52**(5), 5–5 (2009)
20. Vasilescu, B., Serebrenik, A., Mens, T., van den Brand, M.G., Pek, E.: How healthy are software engineering conferences? Science of Computer Programming **89**(Part C), 251–272 (2014)
21. Wainer, J., Eckmann, M., Goldenstein, S., Rocha, A.: How productivity and impact differ across computer science subareas. Communications of the ACM **56**(8), 67–73 (2013)

(a) Total number of published articles
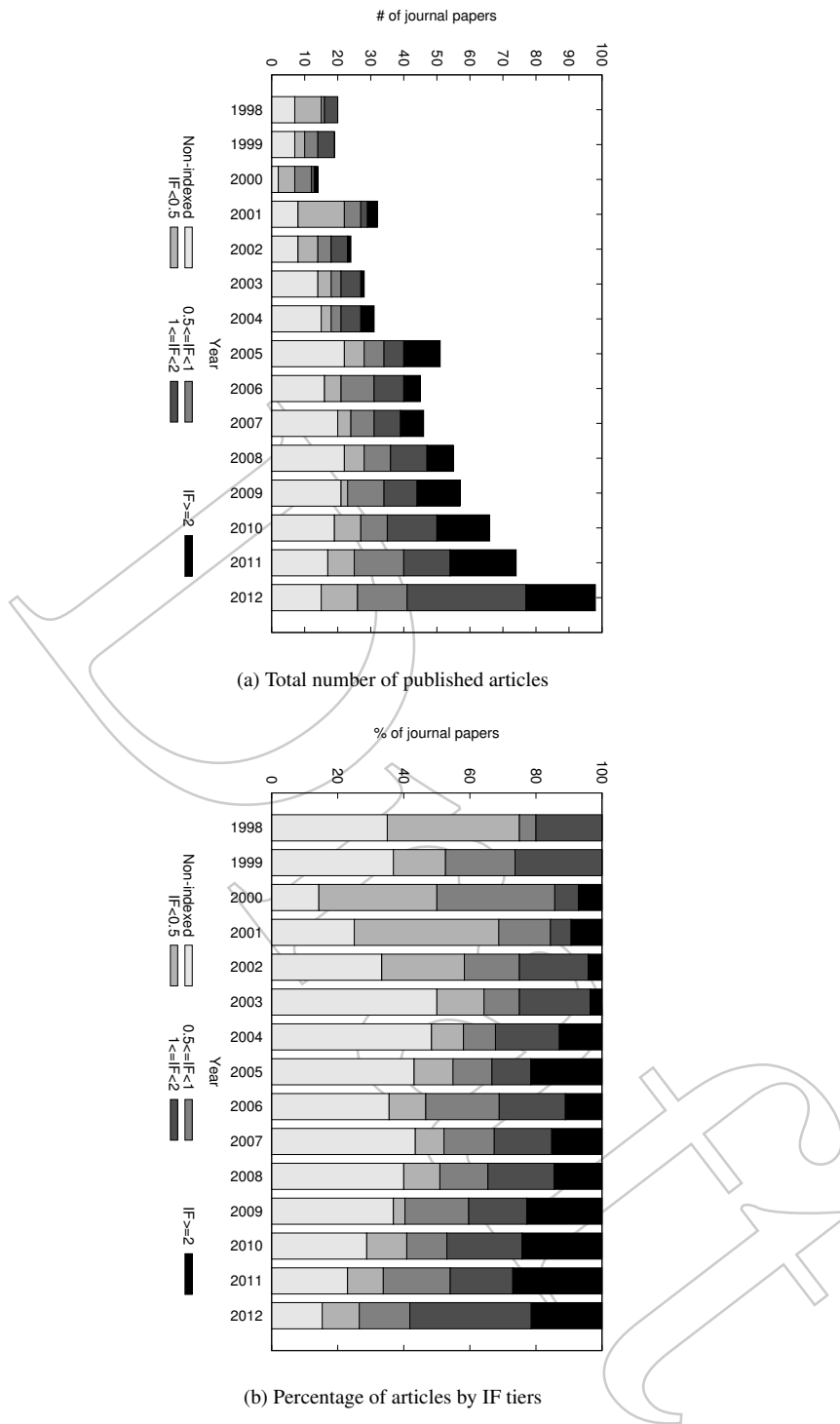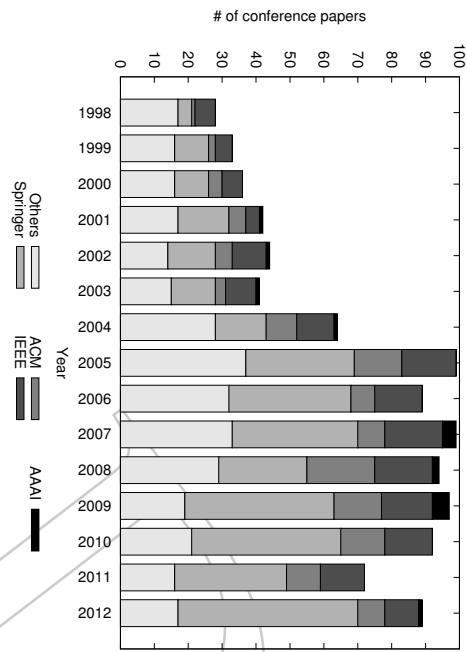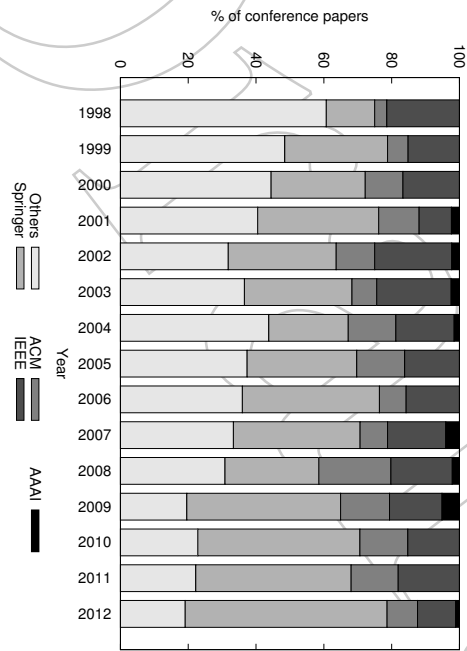

(b) Percentage of articles by IF tiers

Fig. 4: Journal publications by impact factor tiers

(a) Total number of published articles

(b) Percentage of articles by publisher
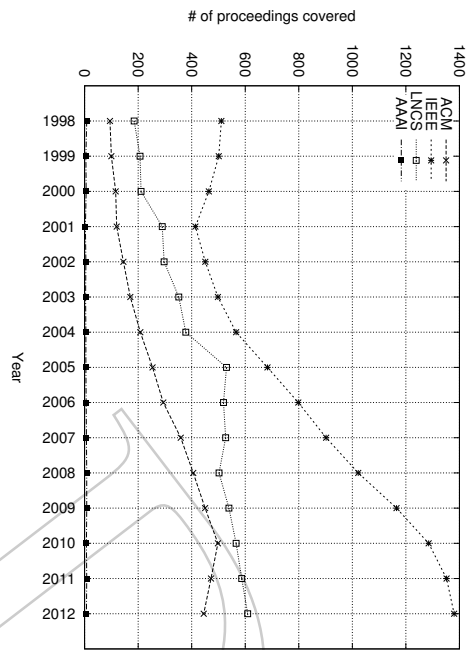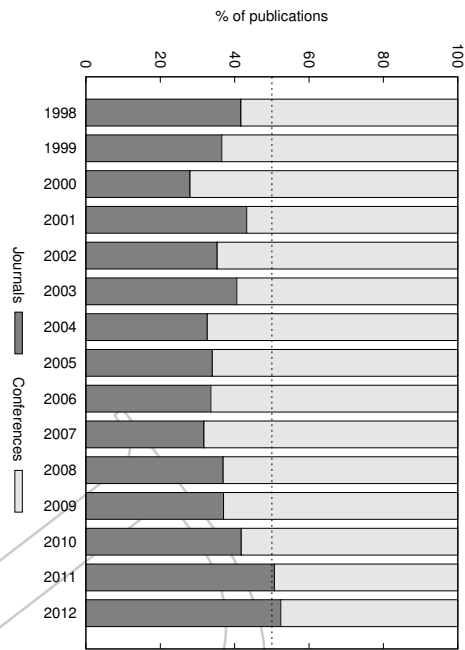
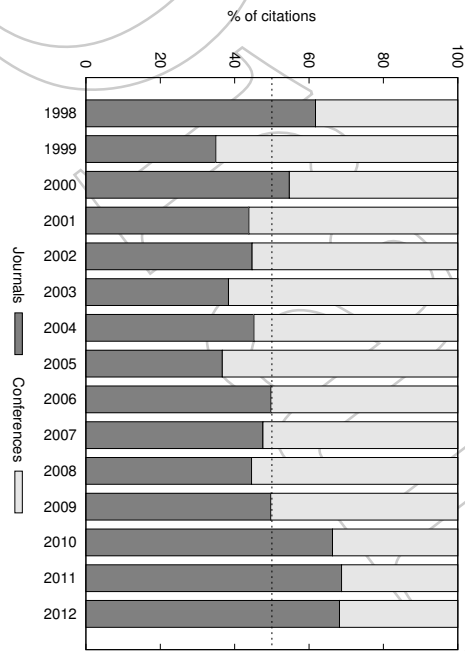Fig. 5: Conference publications by publishers

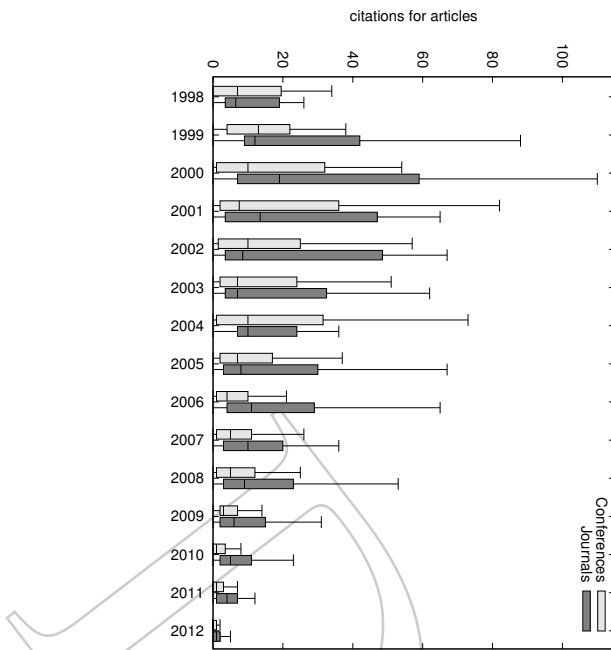Fig. 6: Number of proceedings included in different digital libraries
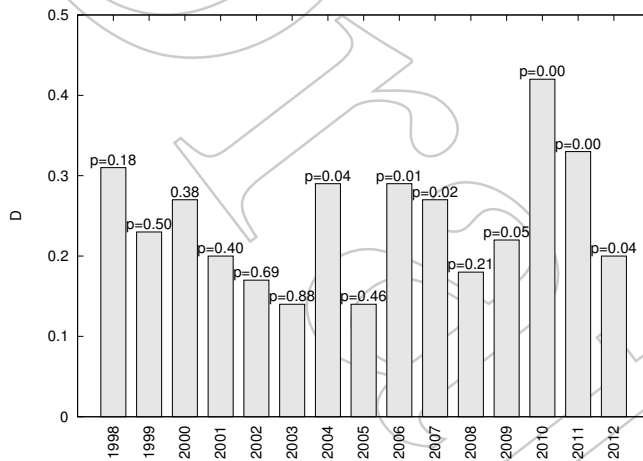
(a) Percentage of publications in conferences and journals

(b) Percentage of citations received by publications in conferences and journals

Fig. 7: Percentage of published papers and citations received per venue

(a) Number of citations received by conference and journal papers

(b) Kolmogorov-Smirnov test for different year citation distributions

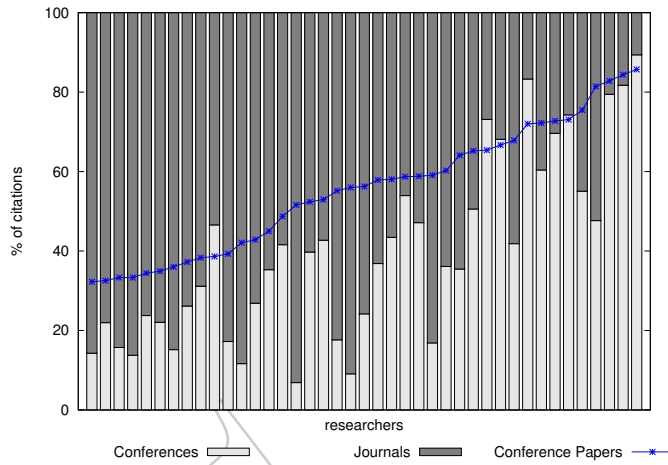Fig. 8: Differences in citations between conference and journal papers

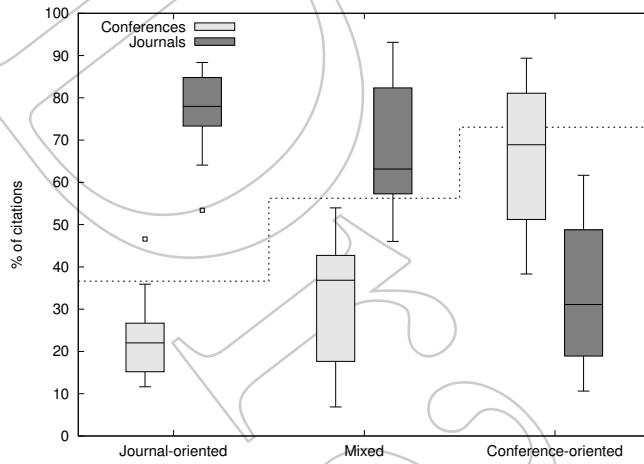Fig. 9: Individual publication profiles and citation sources

Fig. 10: Publication profiles and proportion of citations to journal and conference papers