

Conferences vs. journals in computer science

George Vrettas and Mark Sanderson; RMIT University

1 Abstract

The question of which type of Computer Science (CS) publication— conference or journal – is likely to result in more citations for a published paper is addressed. A series of datasets are examined and joined in order to analyse the citations of over 195,000 conference papers and 108,000 journal papers. Two means of evaluating the citations of journals and conferences are explored, h5 and average citations per paper; it is found that h5 has certain biases that make it a difficult measure to use (despite it being the main measure used by Google Scholar). Results from the analysis show that CS, as a discipline, values conferences as a publication venue more highly than any other academic field of study. The analysis also shows that a small number of elite CS conferences have the highest average paper citation rate of any publication type, although overall, citation rates in conferences are no higher than in journals. It is also shown that the length of a paper is correlated with citation rate.

2 Introduction

In many fields of academic study, a journal paper is the established destination for reporting research. In Computer Science (CS), however, conferences are also valued as a dissemination route. Unlike most other fields of academia, CS academics often view a conference publication as the final venue to describe a piece of work. Within CS, there are around fifty conferences, where having a paper accepted is a notable mark of academic success. The papers are typically between 4,000-7,000 words in length, which enables sufficient detail of the work to be reported. Although CS journals are often willing to publish extensions of conference papers, often CS academics choose not to do this. As a consequence, a number of the CS conferences have become the sole record of research advances.

Compared to other academic fields, this is an unusual publication practise, which has for many years prompted a debate on which is the better publication forum: conferences or journals? In past work, Vardi argued that despite the popularity of conferences for publication, the reviewing quality was not good enough (Vardi, 2009). Fortnow agreed stating that the discipline of CS should “*grow up*” and start publishing in journals where submission deadlines and page limits would not impact on the quality of what was published (Fortnow, 2009). Van Hemert claimed that the reviewing time of CS Journals was far slower than conferences, which resulted in conference being preferred by academics (van Hemert, 2009). Vardi, in reply, called for changes to the journal reviewing process (Vardi, 2010).

One reason why the debate has not been resolved has been a lack of data to measure and compare the relative worth of each publication type. There are well established citation based measures for journals such as two and five year impact factors, as well as the more recent Eigenfactor (Bergstrom, 2007). The custodians of these measures have not focussed attention on applying these measures to conferences. The reasons for this appear to be that in many scientific disciplines, conferences are not valued, and the quality of conference bibliographic metadata is often poorer than it is for journals, making it harder to measure conference impact.

Relatively recently, services, such as Google Scholar¹ (GS) and Microsoft Academic Search² (MAS) have overcome some of these metadata problems enabling a more detailed comparison of the citation rates from journals and conferences. While, there have been past studies examining such questions (Rahm & Thor, 2005), there have been shortcomings of that work, which make comparisons less valid. Google Scholar recently introduced its Metrics (GSM) service, which enables comparison of journal and conference citations. However, the service uses a measure that while well-established, has not been used to assess the citation impact of conferences or journals. It was decided therefore to undertake a study to compare the citation rates of journals and conferences, exploring the following research questions:

- Are the publication practises of CS academics different from other academics?
- Do CS conferences have a stronger citation impact than journals?
- How does such a comparison vary between the two across different quality journals and conferences?
- Do different lengths of paper get different levels of citation?

The paper starts with a summary of past work in this area, followed a description of the data sets used in the study as well as the methodology and measures of our work. The paper next details the results of our study before discussing the results and finally concluding.

3 Past work

Measuring citation has long been used as a means to evaluate publications. Garfield's *impact factor* is probably the best known (Garfield, 1972). Citations to the papers in a publication are measured over a particular period of time (typically two or five years), the higher the aggregate citation count, the higher the publication's impact. There is an assumption that a paper appearing in a high impact publication will over time receive more citations, therefore, such publications tend to be the most desirable venues for academic work. Larivière and Gingras, using a novel methodology, confirmed this assumption showing that the impact factor of a journal significantly affects the number of citations a paper receives (Larivière & Gingras, 2010). See (Leydesdorff, Radicchi, Bornmann, Castellano, & Nooy, 2013) for a review, test, and comparison of proposed alternatives to impact factor.

Impact is commonly calculated for journals only. Though there is no methodological reason for the measure not to be computed for conferences, a lack of coverage of these publications in the traditional citation databases, such as Web of Science³ and Scopus⁴, prevents their impact from being calculated. This lack of coverage has been noted several times (Meho & Yang, 2007; Sanderson, 2008); it presents CS academics with a challenge. Said academics assert that CS conferences have a strong impact – indeed there is empirical evidence to support this (Bar-Ilan, 2010) – however, a lack of metrics supported by the citation databases makes it hard to persuade non-CS academics that conference papers have a value. Measures such as the acceptance rate of a conference are sometimes used as a proxy of quality, but cannot be compared with impact.

Researchers have tried to compare the two types of publications. Rahm and Thor examined two highly regarded database conferences and three equally important journals (Rahm & Thor, 2005). The researchers found that the conference papers had, on average, higher citations than journal papers. In contrast, Franceschet conducted an analysis showing that while most academics publish the majority of their work in

¹ <http://scholar.google.com/>

² <http://academic.research.microsoft.com/> (Note at the time of publishing this paper, it was reported that the Microsoft service was no longer updated, the data gathered for the work in this paper pre-dates this more recent change.)

³ <http://thomsonreuters.com/web-of-science/>

⁴ <http://www.scopus.com/>

conferences, the majority of citations come from their journal publications (Franceschet, 2010). The reason for the opposing conclusions between these two works appears to be the data source used for citations: Rahm and Thor used Google Scholar, Franceschet used Web of Science, which has strong journal citation coverage.

Another citation and publication source is the ACM Digital Library⁵ holding both journal and conference papers with broad citation information. The library details (self-reported) acceptance rates for many CS conferences. Chen and Konstan (Chen & Konstan, 2010) examined 24,294 papers appearing in 2,108 conferences and journals. They showed that full papers⁶ published in low acceptance rate conferences (1 in 3 or lower) were cited more than papers published in a basket of CS journals. Comparing a small sub-set of good quality conferences with all the journals in the sample is only valid if the journals in the sample are of a similar high quality. Such an assumption was not tested.

Taking both journal and conference publication quality into account was the focus of an examination of citation rates of 8,000 papers published in 32 CS conferences and journals covering artificial intelligence, machine learning, information retrieval, and computational linguistics (Freyne, Coyle, Smyth, & Cunningham, 2010). Unlike Chen et al, Freyne et al stratified the publications by quality based on data from the Australian government's research assessment exercise *Excellence in Research for Australia* (ERA). Through consultation with academic associations and rounds of wider feedback, ERA employed subject specialists to label thousands of journals and conferences with one of four quality grades (A*, A, B, or C). The researchers found that the most cited conferences were equal in citation rates to journals rated as A in ERA. The journals rated A* were found to have higher average citation rates than any conference in the group studied.

This brief survey of past work shows contrasting conclusions on the question of which produces more citations: journals or conferences. The reasons for the difference appear to be due to limited data sources for each study. It was concluded that a larger study would provide more reliable results.

4 Is CS publication practise different to other fields?

To answer the first research question, the ERA (2010) list of 20,712 journal and 1,952 conference titles – created to assess the research output of Australian academics – was examined. The list could be reasonably inferred to be a good representation of publication venues of each field of academic research. Every publication in the list was assigned at least one Field of Research (FoR) codes: six digit numbers that detail a three level hierarchy of research areas. As described above, the organisers of ERA, after consulting with a wide range of Australian academic groups, manually graded the journals and conferences for quality. Four grades were used for journals (A*, A, B, C) and three for conferences (A, B, C)⁷.

A comparison was made between the practises of CS (publications coded '08') and engineering (coded '09'). The percentage of all journals and conferences that were assigned a primary CS or engineering FoR code was measured and tabulated (Table 1 and Table 2). The CS journals represented 4% of all the journals ERA considered. The number of engineering journals was larger (7%), but this was to be expected as the field is broader than computer science. For both CS and engineering, there were fewer high quality journals than low. This trend was largely in line with the distribution of grades in the rest of the ERA list.

⁵ <http://dl.acm.org/>

⁶ Determined by page length.

⁷ The current ERA process no longer uses such publication rankings. An archive copy of the list can be found http://web.archive.org/web/20110217195308/http://www.arc.gov.au/era/key_docs10.htm

Journals	All	A*	A	B	C
All academic fields	20,712	1,030	3,054	5,667	10,682
Computer Science	744 4%	54 5%	118 4%	204 4%	346 3%
Engineering	1,462 7%	95 9%	245 8%	426 8%	673 6%

Table 1: Fraction of CS and engineering journals broken down by ERA 2010 grade

Table 2 shows that a substantial percentage of conferences (76%) were from CS, and there were more CS conferences listed than journals (1,489 vs. 744). Engineering was the next largest research field with 12% (236 conferences vs. 1,462 journals). Examining the percentage of conferences across the three grades, it is notable that a large percentage of lower grade conferences (B or C) were CS. In engineering, by contrast, the majority of conferences were graded A. This is not to suggest that engineering conferences were high quality in general, instead, it would appear that those who compiled the engineering list focussed on the best conferences. Those who compiled the CS list appear to have included a wider selection.

Such a strong focus on conferences as noteworthy publication venues place CS in strong contrast to all other fields of academic study.

Conferences	All	A	B	C
All academic fields	1,952	414	527	1,006
Computer Science	1,489 76%	232 56%	382 72%	875 87%
Engineering	236 12%	120 29%	63 12%	52 5%

Table 2: Fraction of CS and engineering conferences broken down by ERA 2010 grade. Please note, one of the Engineering conferences was not graded and is not listed in the table.

5 Data, methodology and measures

In this section, the data sets and the methodologies used in the paper are described.

5.1 Data

The aim of this study was to examine the widest possible list of CS journals and conferences, publication quality information, lists of papers published in the journals and conferences, paper citation counts, and page length information. The data sets chosen were as follows:

- **ERA 2010 conference and journal list.** As described in Section 3, ERA compiled a list of 1,952 conferences and 20,712 journals, each of which was manually assigned a primary FoR code. The 1,489 CS conferences and 744 journals from the ERA 2010 list were extracted.
- **Microsoft Academic Search.** Lists of papers and citation data were taken from Microsoft Academic Search (MAS). According to (Roy et al., 2013), in 2013, MAS held 50 million papers, written by 19 million authors. Rahm showed that an early version of MAS (called Microsoft Libra) had good coverage of citations to CS conferences compared to CS journals (Rahm, 2008). Jacsó (Jacsó, 2011) also states that computer science coverage (including conferences) is the most developed in MAS. An API is provided by MAS to its data and search functions. There are some challenges in joining the ERA list with the data in MAS due to variations in the names of the conferences and journals. Means of dealing with this are described in Section 5.2.1. Several hundreds of thousands of papers were collected from MAS covering the years 2007-2011.
- **The DBLP Computer Science Bibliography.** Unfortunately, page length information was not easily available from MAS. The data set from the DBLP Computer Science Bibliography (Ley, 2009) is accurate, structured, and free to download. Therefore, the title of each listed paper was looked up in DBLP to obtain page length information.

- **ERA 2010/CORE 2008 conference and journal ratings.** Each publication in the ERA 2010 list was manually assigned a rating. The journals were given one of four categories: A*, A, B, and C⁸. Conferences in the 2010 list were placed into three categories: A, B, and C. Unfortunately, these ratings fail to capture information on a small group of elite conferences that are often viewed by the CS community as exceptionally strong. In 2008, a list of elite conferences was compiled by CORE (an Australian academic CS association) using similar criteria to ERA and using all four of the journal grades. Around 5% of conferences were assigned an A*. Therefore, for those conferences rated A* in the CORE ratings, the CORE 2008 rating subsumed the ERA 2010 rating.

5.2 Methodologies to extract paper, citation, and page length information

To extract all of the required data, it was necessary to join the different data sets, which is described here.

5.2.1 Joining the ERA and MAS data

An algorithm was developed to match the conference and journal names from the ERA list to those within MAS. Broadly, the algorithm worked by submitting a query, via the MAS API, based on the name of the conference or journal, and examining the returned results. The name of each returned conference or journal was compared to the original query using a string matching metric. The closest match was considered the MAS conference or journal entry that matched to the original ERA listed name. In a pilot study, it was found that matching on journal names was 42% more accurate than matching on conferences. Therefore, before running the matching algorithm against our entire data set, a random set of 250 conference names were chosen to manually test and tune the matching process.

The first version of our algorithm involved submitting, as a query, the full ERA name of the 250 test conferences to MAS. If more than a single result was returned, the one with the lowest Edit Distance score (compared to the query) was considered to be the matching conference. Testing this version of the algorithm 127 (50.8%) of the conference names matched correctly; 106 (42.4%) did not match to anything; and 17 (6.8%) matched the wrong conference.

The ERA list also provided a short name, typically a unique acronym, for each conference, which was tested as a query to MAS, again comparing the short names of the results using Edit Distance. Now, 134 (53.6%) names were found to be correct matches; 87 (34.8%) no matches; and 29 (11.6%) mismatches. Combining the full and short name search results for a conference name and selecting the match with the shortest Edit Distance gave 152 (60.8%) correct matches; 73 (29.2%) no matches; and 25 (10%) mismatches.

An examination of search results revealed two problems with the algorithm. First, providing the full name of a conference proved restrictive in that commonly occurring words within the name of a conference sometimes resulted in the MAS not finding a match. For example, the conference listed in ERA as “*International Conference on Intelligent Tutoring Systems*” existed within the MAS database only as “*Intelligent Tutoring Systems*”; providing the full ERA name resulted in the MAS not finding a match. Second, Edit Distance takes into account the order of the words and letters in a name, but word order is perhaps less important in conference or journal name matching.

To solve the first problem, the most frequent terms from a conference name were removed before searching MAS. The full list of words that were removed from the conference and journal names was: a, ACM, an, and, annual, conference, for, IEEE, IFIP, international, journal, of, on, SPIE, symposium, the, to, with⁹.

⁸ A small number of journals in ERA list were classed as “not ranked”, these were discarded from our data.

⁹ There was one problem with the stopping procedure; the “Journal of the ACM” is composed of only stopped words. For this particular journal, an exception was written into the algorithm and the full name was submitted as the query to the MAS.

The string comparison metric was changed to the Jaccard Measure, (Jaccard, 1912) which computes a simple word overlap, ignoring word order. Based on an analysis of the sample, a Jaccard score threshold of ≥ 0.2 was required before a match was judged to have occurred. The results of the keyword and short name search algorithm including this threshold were: 167 (66.8%) correct matches; 76 (30.4%) no matches; and 7 (2.8%) mismatches.

The results produced by this final combination of keyword and short name search were considered acceptable and this version of the algorithm was used to search for all the CS conferences and journals from the ERA list. The number of identified conferences was 1,021 and the number of journals was 568, which was manually checked for further errors. In addition, any conferences or journals for which MAS returned fewer than ten papers were discarded to try to avoid citation measures being overly influenced by outliers. This left 630 conferences and 290 journals, within which, there were 195,513 and 108,600 papers respectively.

While undoubtedly, some mismatches exist in the final data set, no evidence was found during the checking of mismatches that would affect the comparison of journals against conferences, the core research question of this paper. Note, this data will be made available for download for other to work with.

5.2.2 Determining paper length

Publications that accept short papers typically reduce reviewing criteria for such papers. As conferences, in particular, can accept a large number of such papers, it was decided to separate such short papers out. An algorithm was developed to look up the papers in DBLP based on their titles and year of publication. First a case-normalisation on the titles in both the candidate and DBLP papers was performed. The DBLP paper with the shortest Edit Distance, published in the same year as a query paper, was selected. Edit Distance was used over Jaccard as here word order was important. Matches with Edit Distance ≤ 5 were selected. The algorithm was applied to a set of 200 randomly chosen papers and the matches were manually inspected. The algorithm was found to correctly match 183 (91.5%) of the papers. The algorithm also 'correctly missed' 12 (6%) of the papers. For 5 (2.5%) of the papers, the algorithm matched the wrong paper. Such results were considered acceptable and no further refinement was made.

With the lengths of each paper established, the papers were assigned to one of three groups: short papers (with between one and four pages); long papers (page length greater than or equal to five); and unknown (papers for which page length could not be determined). A lack of page length information was either due to DBLP not having that information, or the paper title matching algorithm failing to find a match. The number of papers in each group is shown in Table 3. The number of short papers for conferences was notably higher than the number of short papers for journals.

Length	Long (≥ 5 pages)		Short (≤ 4 pages)		Unknown		All
Journal	87,082	80%	5,742	5%	15,776	14%	108,600
Conference	134,417	69%	32,987	17%	28,129	14%	195,513

Table 3: The number of journal and conference papers in the sample broken down by length

5.2.3 Final data set

The final step in processing the data was to stratify the conferences and journals into the ERA/CORE grades. As can be seen in Table 4 and Table 5, the stratification was necessary, as the balance of both publications and papers in the samples was different. The majority of conferences papers were B and C grade; whereas for journals, the grades were A and B.

Grade	A*		A		B		C		Total
Journals	29	10%	65	22%	85	29%	111	38%	290
Conferences	45	7%	119	19%	221	35%	245	39%	630

Table 4: The number of journals and conferences in the sample, broken down by ERA/CORE grade

One can compare the number of journals and conferences that were left after joining all the data sets (in Table 4) with the number of CS journals (Table 1) and conferences (Table 2) in the original ERA lists. From the comparison it can be seen that 39% of the journals and 43% of the conferences remained after the joining process. The '08' FoR code is divided into nine sub-categories, such as Artificial Intelligence, Information Systems, etc. It was decided to not examine citations at this finer grained subject level as the number of publications in each of the grades, particularly for A and A*, was small and there was a strong risk of averages being adversely affected by outliers.

Grade	A*		A		B		C		Total
Journal papers	20,666	19%	29,867	28%	34,991	32%	23,076	21%	108,600
Conference papers	22,099	11%	43,084	22%	63,444	32%	66,906	34%	195,533

Table 5: The number of journal and conference papers in the sample, broken down by ERA/CORE grade.

The average number of papers per journal and conference was examined; the results are shown in Table 6. Examining the overall average first, it can be seen that the average number of papers per conference (measured over the 5 years, 2007-2011) is lower than the average number of papers per journal. The reasons for this difference were not studied, although, in general, journals produce multiple issues per year and conferences typically only produce one proceedings per year, so it is perhaps not surprising that such a difference exists.

The decline in numbers of papers per publication as the grade of the publications reduces was less expected. We speculate that the quality of metadata in MAS that records the papers for each conference or journal is not as good as it is for the better quality publications.

Grade	A*	A	B	C	All grades
Papers per journal	713	459	412	208	374
Papers per conference	491	362	287	273	310

Table 6: Analysis of papers per publication type stratified by ERA/CORE grade measured over 5 years of output

With this data set in place, the final decision was which evaluation measure to use.

5.3 Choosing an evaluation measure

Different evaluation measures are used to assess the publications. We examined two measures: average citations per paper measured over a five year period and the h (Hirsch, 2005) of a publication also measured over five years (h_5), which is used by Google Scholar with its Metrics service (GSM). The creators of GSM took a liberal view of what counts as a publication including journals, conferences, workshops, and refereed online archives of papers.

On the GSM web site h_5 is defined as “...the largest number h such that h articles published in 2008-2012 have at least h citations each”. When measuring the impact of an individual academic, the h family of measures give credit for papers that are cited highly, but are unaffected by papers that have little or no citation. Hirsch appears to have created the measure to reward academics who produce a “broad and sustained impact” (p. 16569). In theory, an academic could obtain a high h by publishing a vast number of papers, hoping that by chance enough will be highly cited. However, in practice no individual would be able to produce the volume of papers required for such a strategy to work.

When measuring h on a publication, such volumes of papers are possible, as can be seen in the 2013 GSM list of the top 100 publications in the world. There, seven of the ‘publications’ are sections of the refereed pre-print web site arXiv¹⁰. The “High Energy Physics - Phenomenology” section has an h_5 of 141, placing it 25th in the top 100. However, over the five years citations were counted (2008-2012) 21,817 papers were deposited in this section of the archive. While an h_5 of 141 suggests that the archive produced a broad and sustained

¹⁰ <http://arxiv.org/>

impact, in the context of nearly 22 thousand uploads, that measure looks less impressive. This bias in h has been observed before (Harzing & Van Der Wal, 2009), it is examined in more detail here.

The relationship between the volume of papers deposited and h_5 was measured over sixteen sections of arXiv, see Figure 1. It can be seen that there is a relationship between the size of an archive and h_5 ; a Spearman's rank correlation coefficient was 0.93.

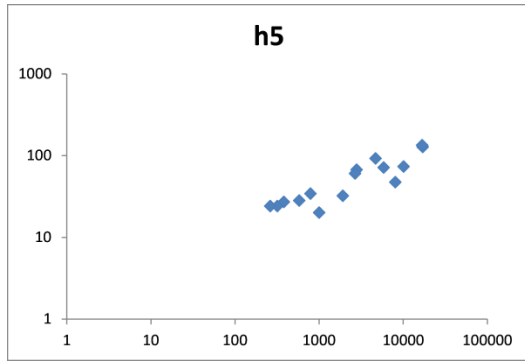


Figure 1: Plot showing the relationship between publication size and h_5 . Axes on log scale.

Note also, in his paper, Hirsch appears to suggest that h was devised to ensure that academics with “low productivity” should not be rewarded by the measure. While it is arguable that such an approach is reasonable for academics, it is not so clear that publications with low productivity should be penalized. Just to match the 141 h_5 of “High Energy Physics - Phenomenology”, a refereed journal or conference has to produce 141 papers in the five year measurement period (28 and 29 papers annually) each cited at least 141 times. In practise, a publication needs to publish many more in order to obtain an h_5 as high as 141. This is a difficult requirement for many low volume conferences and journals.

From such examples, it is less clear if h_5 is an ideal measure for ranking publications. A simple approach to fixing these problems is to calculate mean citations per paper in a publication. Therefore a study was conducted comparing h_5 with average citation rate.

5.3.1 Comparing h_5 with average citation rate

Across all the conferences and journals described in Section 5.1, the average citation rate and the h_5 of each publication was calculated, see Figure 2, Figure 3, Figure 4, and Figure 5.

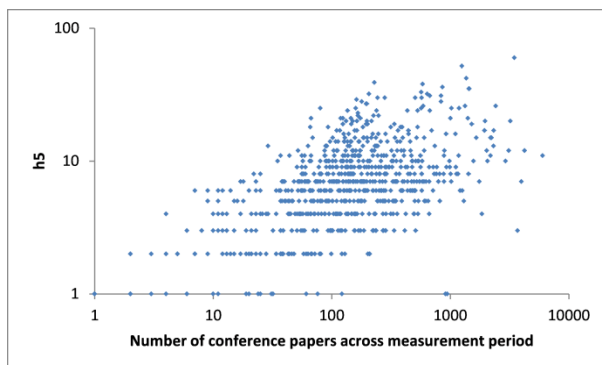


Figure 2: Scatter plot showing the relationship between conference size and its h_5 score. Axes on log scale.

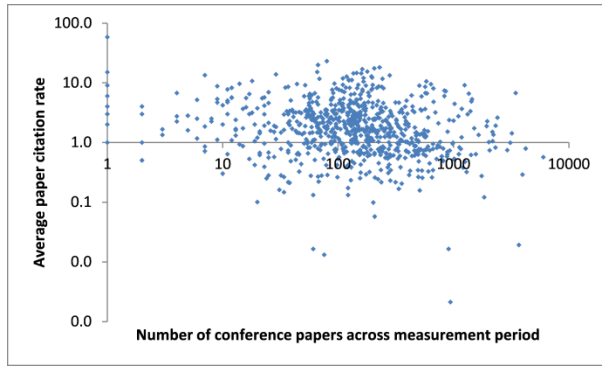


Figure 3: Scatter plot showing the relationship between conference size and its average citation rate. Axes on log scale.

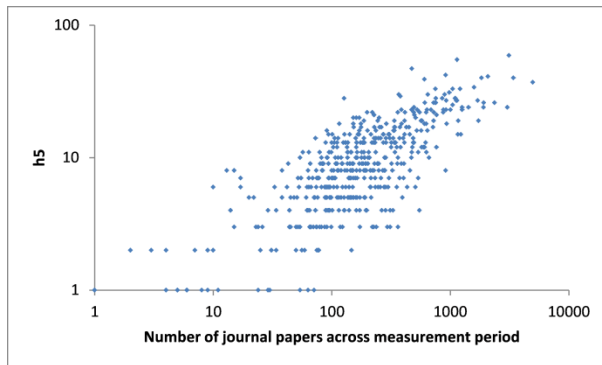


Figure 4: Scatter plot showing the relationship between journal size and its h_5 score. Axes on log scale.

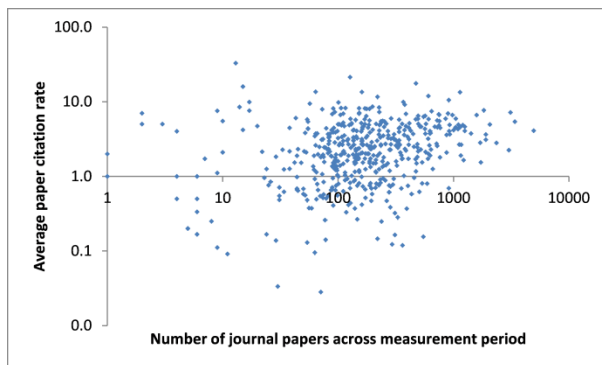


Figure 5: Scatter plot showing the relationship between journal size and its average citation rate. Axes on log scale.

A Spearman's rank correlation coefficient was measured on the four graphs. The correlation between the number of papers and h_5 on conferences and journals was respectively 0.39 and 0.70; between average paper citation rates, the correlation was -0.03 and 0.16.

As with the analysis of the arXiv data, there is a correlation between publication size and h_5 . The correlation for conferences is not as strong as it is for journals and both correlations are less than that measured on the arXiv data. The reason for the lower correlation compared to arXiv is likely to be the influence of reviewing quality for the different venues. However, all correlations for h_5 are much higher than average paper citation rate, for which no discernible correlation exists.

It would appear that if one measures a publication venue using h_5 , there is a risk the score will be notably influenced by the size of the venue; therefore, it was decided to measure publications using average citation per paper.

6 Results

The next research question to examine was do CS conferences have a stronger citation impact than journals? The citation rate for each long paper published in the last five years was looked up in MAS. Then for each ERA/CORE grade, the average citation rate was measured, see Table 7.

Grade	A*	A	B	C	Average
Journal	7.0	4.8	3.5**	2.6**	4.5
Conference	9.5*	4.8	2.7	1.8	4.7

Table 7: The (micro) average citation rates of journal and conference papers by ERA/CORE grade.

The overall average for the conferences and journals are similar: 4.5 vs. 4.7. Two-tailed two-sample unequal variance t-tests were applied to the differences: starred values in the table indicate that the number is significantly higher than the corresponding number in the same column (* $p \leq 0.05$, ** $p \leq 0.01$). It can be seen that A* conferences obtained a significantly higher average citation rate than A* journals. Conferences and journals graded A showed no statistical difference. Journals graded B and C had a significantly higher citation impact over conferences with the same rating.

The distribution of journal and conference citations was plotted, see Figure 6, Figure 7, Figure 8, and Figure 9. The distributions across the two publication types are similar, even though for three of the grades (A*, B, C) the differences in average citation was statistically significant. Generally, as the grade of the publication reduces, the distributions become increasingly skewed to low average citation rates. For both A* and A, there appears to be a greater number of journal papers with low citation rates and a relatively small number of conferences with particularly high citation rates.

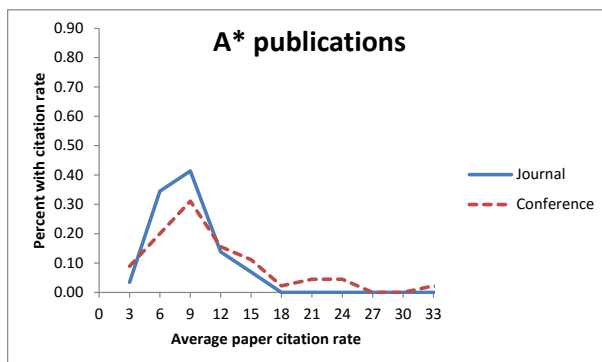


Figure 6: Plot showing the distribution of average citation rates for A* journals and conferences.

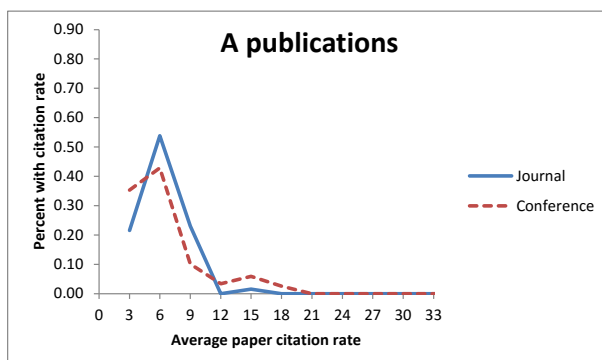


Figure 7: Plot showing the distribution of average citation rates for A journals and conferences.

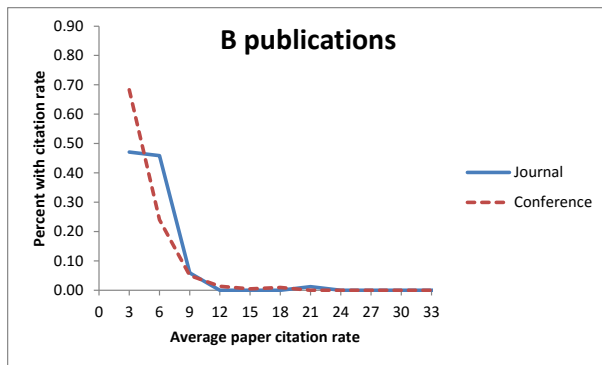


Figure 8: Plot showing the distribution of average citation rates for B journals and conferences.

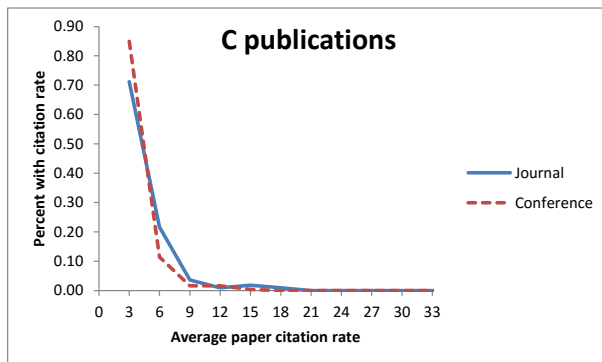


Figure 9: Plot showing the distribution of average citation rates for C journals and conferences.

The standard deviation of the averages in each grade was measured, see Table 8. For journals, the deviation is relatively constant; for conferences particularly A* and A, the deviation was higher, suggesting that there is greater variability in the citation rates for conferences in those two grades.

Grade	A*	A	B	C
Journal	2.8	2.2	2.4	2.5
Conference	6.0	3.6	2.4	1.9

Table 8: The standard deviation of average citation rates, broken down by ERA/CORE grade.

6.1 Examining years

The citation data for each of the five years was analysed (see Table 9) as it was speculated that the growth in citations over the years might be different for journals and conferences. In general, the broad trends seen in Table 7 were reflected here: papers appearing in A* conferences have higher citation rates than those appearing in A* journals; the citation rates of papers in A conferences and journals are similar; and citation rates in B and C journals are higher than rates in the equivalent conferences. There are some differences between the micro and macro averages between the two tables, which were explained by a slightly uneven distribution of paper numbers across the five years. There was no discernable difference between conferences and journals in regard to the growth of citations over the five years studied.

6.2 Examining short papers

The impact of paper length on citation rate was explored: the average citation rate of short and long papers was compared. For journals, the citation rates were 1.4 and 4.2 respectively; for conferences, they were 1.0 and 2.6. Using a same t-test as above, both differences were found to be significant ($p < 0.01$).

	Year	2007	2008	2009	2010	2011	Average (macro)
A*	Journal	14.0	9.5	5.6	2.3	0.6	6.4
	Conference	16.9	12.0	6.9	2.5	0.3	7.72
A	Journal	10.1	7.4	3.8	1.8	0.5	4.72
	Conference	9.8	6.5	3.7	1.4	0.4	4.36
B	Journal	7.4	5.2	3.1	1.4	0.4	3.5
	Conference	4.4	3.8	2.2	0.9	0.2	2.3
C	Journal	5.2	4.7	2.5	1.1	0.3	2.76
	Conference	3.2	2.4	1.7	0.6	0.1	1.6

Table 9: Average citation per rate measured across the five years data was gathered for

Grade	A*	A	B	C
Journal	1.4	0.9	1.0	1.1
Conference	2.4	1.6	0.9	0.9

Table 10: The average citation rates of short journal and conference papers in the sample, broken down by ERA/CORE grade.

The short paper citation rates were calculated for each of the four ERA/CORE grades. The results shown in Table 10 reflect the long paper citation rates in Table 7: the pattern of citations is similar. A short paper in an A* conference on average obtained a similar number of citations (2.4) to a full paper in a C journal (2.6). None of the differences between journal and conference were statistically significant.

The relationship between paper length and citation count was examined in more detail. Conference and journal papers were sorted by their length and the average citation count of all papers for each integer page length was calculated. For any length for which there were fewer than 100 papers, the calculation was ignored as it was assumed that the average could be skewed by outliers. The graphs in Figure 10 and Figure 11 show the relationship between page length and average citation rate for journal and conference papers.

For the conference papers, there is a linear trend, for journal papers, the trend appears to be a logarithmic. The commonest length of conference papers was eight pages, for journal papers, twelve. The distribution of lengths is shown in Figure 12 and Figure 13. The number of citations per page for each paper was also calculated for both conferences and journals, see Figure 14 and Figure 15. Here a difference emerged: for conferences, the rate of citations per page appeared to be relatively constant; for journals, however, the rate peaked at around 5 pages, before dropping off steadily.

It should be noted that the standard deviation of each citation average was substantial: 13 for journals, 8 for conferences. While page length influences citation rate, there is very large variation behind the averages calculated.

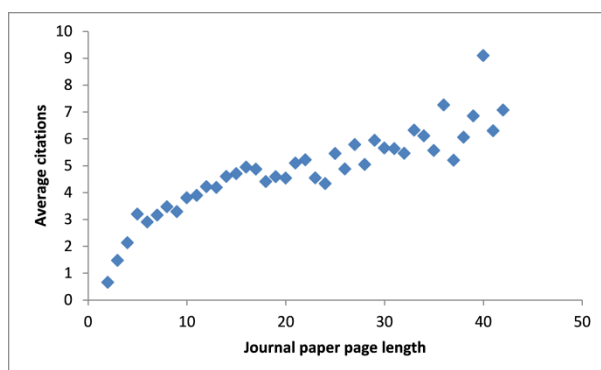


Figure 10: Scatter plot showing the relationship between journal paper page length and average citation rate.

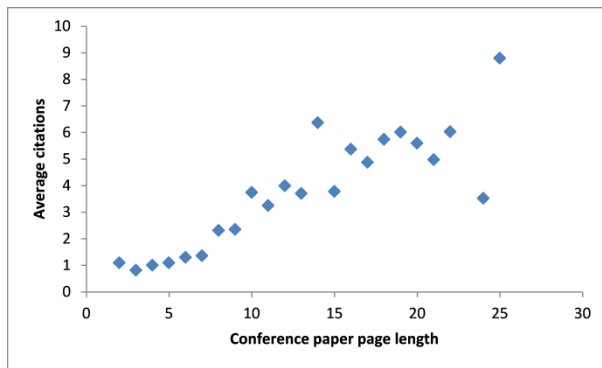


Figure 11: Scatter plot showing the relationship between conference paper page length and average citation rate.

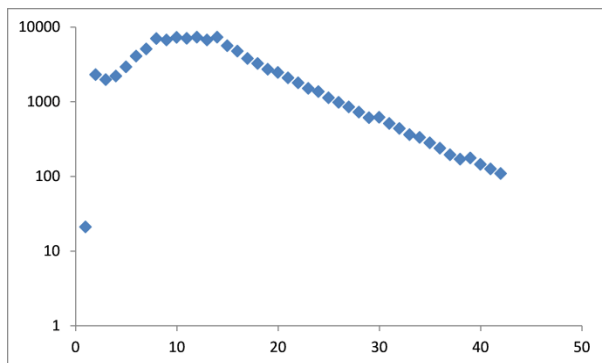


Figure 12: Scatter plot showing the distribution of journal paper page lengths as plotted on a log scale.

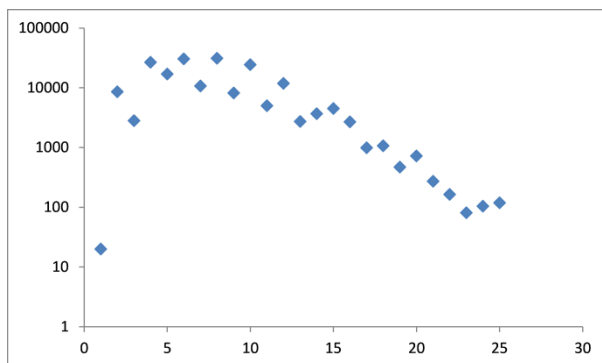


Figure 13: Scatter plot showing the distribution of conference paper page lengths as plotted on a log scale.

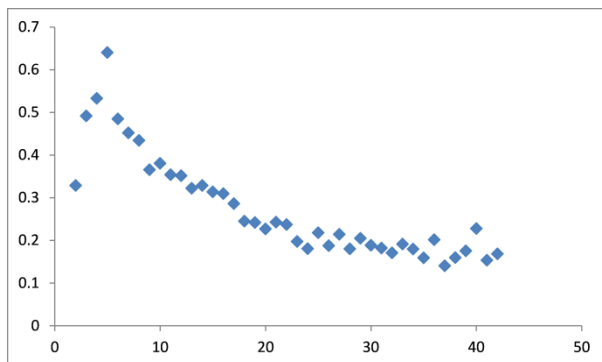


Figure 14: Scatter plot showing a calculation of citations per page for different lengths of journal papers.

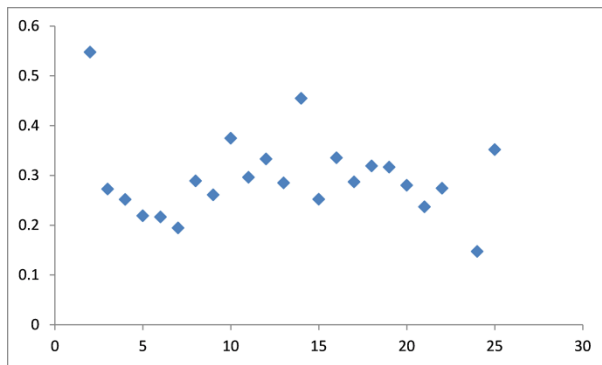


Figure 15: Scatter plot showing a calculation of citations per page for different lengths of conference papers.

7 Discussion

This work studied the citation impact of CS conferences vs. journals. Such research questions were examined before, however, past work did not explore this topic on the scale or breadth conducted here. Rahm and Thor (Rahm & Thor, 2005) and Freyne et al (Freyne et al., 2010) each used fewer than fifty publications from particular sub-areas of CS. The wide standard deviations measured over the MAS dataset in this paper, indicate that using a large data set was necessary to obtain reliable statistics. Chen and Konstan (Chen & Konstan, 2010) showed that conferences with low acceptance rates had higher average citation rates than a wide basket of journals, but the lack of stratification of the journals made comparison between their work and that presented here hard.

Past work does not appear to have studied page length in the way that was conducted here. As can be seen, the graph (see Figure 10 and Figure 11) showed a link between page length and citation. The curves in the two graphs also appeared to be somewhat different with short conference papers receiving consistently low citations and only longer papers received more notable levels of citation. Causes for this apparent pattern are likely to be multiple. Longer papers on average will report more results than shorter papers; longer papers tend to be reviewed more carefully and so should represent more substantial (more citable) work; shorter papers might cover published materials that are less likely to be cited, such as abstract of keynotes; or letters to the editor.

The examination of the balance between conferences and journals in the CS ERA lists (described in Section 4) showed that the venues that a group of CS academics wished to be considered in a research evaluation exercise was dominated by conferences. However, a great many of the venues were weak. It can be seen in Table 2 that there are five times more B and C conferences than there are A. If one were to measure citations per paper across the ERA listed CS conferences regardless of grade and compared that to citations per paper across all ERA journals, journals have a higher average citation rate.

It was only when the quality of publication types was considered, that some citation rates in conferences were measured to be higher than journals. Here, A* conferences had a higher average citation rate (36% higher) compared to A* journals. The difference was statistically significant, though due to high variability of citation rates for conferences, the level of significance was not high. For A journals and conferences there appeared to be little difference and for B and C grade publications, journals had a higher citation rate.

8 Conclusions

Research was described in this paper that addressed the following research questions:

- Are the publication practises of CS academics different from other academics?

- Do computer science conferences have a stronger citation impact than journals?
- How does such a comparison vary between the two across different quality journals and conferences?
- Do different lengths of paper get different levels of citation?

Based on the data analysed in this study, the overall answer to the first question was yes, CS publication practises a very different from other academics. The answer to the second research question is that there is little difference between journals and conferences in terms of citation. However, once journals and conferences were stratified based on a manual assessment of the quality of each publication (the third research question), it was found that for a small number of elite conferences, average citation rates were notably higher than for elite journals. Papers in lower quality journals have more citations than papers in lower quality conferences. The length of a paper was found to have an impact on citation rate. The relationship between length and citation exists across all ranges of page length.

The fact that there continues to be a debate in CS on publication in conference vs. journals perhaps indicates that the rest of academia (as represented on tenure, promotion, or grant award committees) is not yet convinced of the value of conference publication in CS. The uniqueness of the publication profile of CS academics (see Section 4) is perhaps not fully understood by the rest of academia. However, the apparent value that CS academics place on publication in a large number of low grade conferences perhaps weakens the arguments of CS academics on the clear importance of publication in the elite CS conferences.

Perhaps a way to resolve the debate with others on whether journals or conferences are better is for CS academics to not try to suggest that all CS conferences are good publication venues, and instead the field should focus on promoting the virtues of publication in only the very best such forums. The remaining lower grade venues should be viewed like conferences in any other academic field of study: valuable meeting places for interaction and exchange of ideas, but as a suitable place for publishing the final word on a piece of work, their value is questionable.

8.1 A note of caution

While the length of the paper, the type of publication (conference or journal) or the quality of the publication it is published in does appear to influence the number of citations, behind the averages computed in this study, was substantial variation. This has been pointed out in the past, Seglen stated that many papers published in high impact journals are poorly cited (Seglen, 1997). Lozano et al presented results indicating that journal impact factors have a weaker influence on citation than they did in the past (Lozano, Larivière, & Gingras, 2012). While the nature of this variation was not examined in this paper, we speculate that it is simply due to the influence of the quality of the work reported in each individual paper. While attempting to get a paper accepted in the very best type of publication has been shown to influence subsequent citation rates, the principle means of influencing citations is the quality and clarity of the paper's content.

9 Acknowledgements

Many thanks to Dr. Tim Cahill and the paper's reviewers for reading and providing valuable comments on early drafts of the paper.

10 References

Bar-Ilan, J. (2010). Web of Science with the Conference Proceedings Citation Indexes: the case of computer science. *Scientometrics*, 83(3), 809–824.

- Bergstrom, C. (2007). Eigenfactor: measuring the value and prestige of scholarly journals. *College & Research Libraries News*, 68(5), 314–316.
- Chen, J., & Konstan, J. A. (2010). Conference paper selectivity and impact. *Communications of the ACM*, 53(6), 79–83. doi:10.1145/1743546.1743569
- Fortnow, L. (2009). Viewpoint Time for computer science to grow up. *Communications of the ACM*, 52(8), 33–35.
- Franceschet, M. (2010). The Role of Conference Publications in CS. *Commun. ACM*, 53(12), 129–132. doi:10.1145/1859204.1859234
- Freyne, J., Coyle, L., Smyth, B., & Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*, 53(11), 124–132.
- Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation; Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479.
- Harzing, A.-W., & Van Der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, 60(1), 41–46.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11(2), 37–50.
- Jacsó, P. (2011). The pros and cons of Microsoft Academic Search from a bibliometric perspective. *Online Information Review*, 35(6), 983–997.
- Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, 61(2), 424–427.
- Ley, M. (2009). DBLP: some lessons learned. *Proceedings of the VLDB Endowment*, 2(2), 1493–1500.
- Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., & Nooy, W. (2013). Field-normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs. *Journal of the American Society for Information Science and Technology*, 64(11), 2299–2309.
- Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, 63(11), 2140–2145.

- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.
- Rahm, E. (2008). Comparing the scientific impact of conference and journal publications in computer science. *Information Services and Use*, 28(2), 127–128.
- Rahm, E., & Thor, A. (2005). Citation analysis of database publications. *ACM Sigmod Record*, 34(4), 48–53.
- Roy, S. B., De Cock, M., Mandava, V., Savanna, S., Dalessandro, B., Perlich, C., ... Hamner, B. (2013). The Microsoft academic search dataset and KDD Cup 2013. In *Proceedings of the 2013 KDD Cup 2013 Workshop* (p. 1).
- Sanderson, M. (2008). Revisiting h measured on UK LIS and IR academics. *J. Am. Soc. Inf. Sci. Technol.*, 59(7), 1184–1190. doi:<http://dx.doi.org/10.1002/asi.v59:7>
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314(7079), 498.
- Van Hemert, J. (2009). Letters to the editor. *Communications of the ACM*, 52(11), 6–7.
- Vardi, M. (2009). Conferences vs. Journals in computing Research. *Communications of the ACM*, 52(5), 5.
- Vardi, M. (2010). Revisiting the Publication Culture in Computing Research. *Communications of the ACM*, 53(3), 5.