<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A quantitative evaluation of the relative status of journal and conference publications in computer science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>Freyne, Jill; Coyle, Lorcan; Smyth, Barry; Cunningham, Pádraig</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2010-11</td>
</tr>
<tr>
<td><strong>Publication information</strong></td>
<td>Communications of the ACM, 53 (11): 124-132</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>ACM</td>
</tr>
<tr>
<td><strong>Link to online version</strong></td>
<td><a href="http://dx.doi.org/10.1145/1839676.1839701">http://dx.doi.org/10.1145/1839676.1839701</a></td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/2525">http://hdl.handle.net/10197/2525</a></td>
</tr>
<tr>
<td><strong>Publisher's statement</strong></td>
<td>This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Communications of the ACM Volume 53, Number 11 (2010), Pages 9-9 <a href="http://doi.acm.org/10.1145/10.1145/1839676.1839701">http://doi.acm.org/10.1145/10.1145/1839676.1839701</a></td>
</tr>
<tr>
<td><strong>Publisher's version (DOI)</strong></td>
<td>10.1145/1839676.1839701</td>
</tr>
</tbody>
</table>
A Quantitative Evaluation of the Relative Status of Journal and Conference Publications in Computer Science

Jill Freyne *
CLARITY: Centre for Sensor Web Technologies,
University College Dublin, Belfield, Dublin 4, Ireland
jill.freyne@csiro.au

Lorcan Coyle
Lero, University of Limerick, Limerick, Ireland
lorcan.coyle@lero.ie

Barry Smyth
CLARITY: Centre for Sensor Web Technologies,
University College Dublin, Belfield, Dublin 4, Ireland
barry.smyth@ucd.ie

Pádraig Cunningham
Complex Adaptive Systems Laboratory
University College Dublin, Belfield, Dublin 4, Ireland
padraig.cunningham@ucd.ie

May 21, 2009

Abstract

While it is universally held by computer scientists that conference publications have a higher status in computer science than in other disciplines there is little quantitative evidence in support of this position. The importance of journal publications in academic promotion makes this a big issue since an exclusive focus on journal papers will miss many significant papers published at conferences in computer science. In this paper we set out to quantify the relative importance of journal and conference papers in computer science. We show that computer science papers in leading conferences match the impact of papers in mid-ranking journals and surpass the impact of papers in journals in the bottom half of the ISI rankings – when impact is measured by citations in Google Scholar. We also show that there is a poor correlation between this measure of impact and conference acceptance rates. This indicates that conference publication is an inefficient market where venues that are equally challenging in terms of rejection rates offer quite different returns in terms of citations.

*This author has moved to the Tasmanian ICT Center, CSIRO, Australia
1 Introduction

How to measure the quality of academic research, and how to evaluate the performance of particular researchers, has always led to a significant degree of debate. Many feel that performance assessment is an exercise in futility, in part because they believe that academic research cannot be boiled down to a set of simple performance metrics, and in part because they argue that any attempt to introduce such metrics will expose the entire research enterprise to manipulation and gaming. On the other hand, many do recognise the need for some reasonable way to evaluate academic performance, and argue that even imperfect systems help to shed sufficient light on the quality issue, enough to guide funding agencies and tenure committees to make more informed decisions, for example.

One long-standing way to evaluate academic performance is by an assessment of publication output. It is best practice for academics to write-up key research contributions as a scholarly article for submission to a relevant journal or conference, and the peer-review model has stood the test of time as a way to govern the quality of accepted articles. However, the modern world of academic publication now accommodates a broad range of publication opportunities which have led to a continuum of quality, where there is a very significant gap between the lower and upper reaches of this continuum; for example, journal papers are routinely considered to be superior than conference papers, which in turn are generally considered to be superior to papers presented at smaller workshops or local symposia. A number of techniques are routinely used for the purpose of evaluating publications and publication outlets, mostly targeted at journal publications. For example, organisations such as the Institute for Scientific Information (ISI) record and assess the number of citations that are attracted by leading journals (and some high-ranking conferences) in order to compute metrics such as the impact factor of a journal as a basic measure of its ability to attract citations to its contents. Less reliable indicators of publication quality are available to judge conference quality; for example the rejection rate of a conference is often cited as a quality indicator, on the grounds that high rejection rates suggest a more selective review process that is capable of turning out higher quality papers. However, as always, the devil is in the details, and the details in this case can vary greatly between academic disciplines and sub-disciplines.

In this paper we examine the issue of publication quality from a computer science/engineering perspective. We will describe how common publication practices differ from those of other disciplines, in that computer science/engineering research is predominantly published in conferences rather than in journals. This presents an important challenge when it comes to evaluating computer science research output because traditional impact metrics are better suited to evaluating journal rather than conference publications. In an effort to legitimize the role of conferences papers to the wider community we present an impact measure, based on an analysis of Google Scholar citation data, that is well suited to computer science conferences. We validate this new measure with a large-scale experiment, covering 8,764 conference and journal papers, to demonstrate a strong correlation between traditional journal impact and our new citation score. The results of this study speak to that quality of conferences within computer science. They
highlight how leading conferences compare favorably to mid-ranking journals, surpassing the impact of journals in the bottom half of the traditional ISI ranking. In turn, we discuss a number of interesting anomalies within the computer science conference circuit, highlighting how conferences with similar rejection rates (the traditional way to evaluate conferences) can attract quite different citation levels. We also note some interesting geographical distinctions in this regard, particularly with respect to European and U.S. conferences.

2 Publication Practice in Computer Science

Computer science, is a relatively new field of study — the first schools of computer science emerged as recently as the 1980’s — and it can be differentiated from many disciplines in some important respects. Certainly, in recent times the pace of innovation in the field has proven to be unusually rapid and this has, in part, led to some unusual publication practices, at least by the standard of other more traditional disciplines. As already stated, in computer science, conference papers are considered to be a more important form of publication than is generally the case in many other scientific disciplines.

When computer scientists have some interesting or significant research results to report they prepare a conference paper to present at the international conference for their research community, for example at IJCAI or AAAI, for researchers working in Artificial Intelligence or at COLING for researchers in computational linguistics. If the research is accepted for publication at the conference this will normally count as the ‘archival’ publication for that research.

In some circumstances the research might also be published in an extended form in a journal. For example the Artificial Intelligence in Medicine journal publishes extended conference papers from various related medical conferences. In the computer science field conference papers are usually submitted as full papers and they undergo a comprehensive peer-review evaluation, usually involving 3-5 external reviewers. As a result, the leading computer science conferences can have very high rejection rates, resulting in high quality papers that attract considerable attention (and citations) from other researchers. This is in stark contrast to the role of conferences in other disciplines, where conference papers usually take the form of extended abstracts that are not subject to the full rigor of peer-review, and that, as a consequence, rarely attract the same critical attention of other researchers. In these other disciplines the only archival publications are journal papers.

This matters because of the important role that publications play in academic promotions and other forms of research assessment. When such assessments typically span many disciplines, it is not unusual to find conference papers excluded, with only journal papers (and perhaps books and book chapters) considered as eligible. Given the preponderance of conference papers in computer science, this places computer scientists at a distinct disadvantage relative to their peers in other disciplines. The situation is exacerbated by the fact that there is often a complete lack of understanding of the other point of view on this issue. Many computer scientists feel that conferences are a timely and appropriate means of disseminating research results and some view publishing in
journals as somewhat superfluous – the relatively low ISI impact factors for computer science journals are evidence of this. The tradition of publishing work in a conference venue rather than a journal is advantageous to researchers in the computer science field for many reasons. Firstly the field of computer science research tends to be fast paced and the “publish or perish” dictate holds strong. The time frame from submission to publication release for a conference is often less than half that of a similar journal allowing the latest finding to be public knowledge rapidly. A further traditional research trend in Computer Science is the sharing of findings with other researchers in the area? This is facilitated firstly by the peer review model where valued feedback is provided, through the oral presentations at conferences and the physical co-location of experts in the field during the event. By contrast, researchers from other fields feel that if the research had any merit it would have been published in a journal and argue that the solution for computer scientists has to be to focus on journal publications as a matter of urgency.

The computer science viewpoint is greatly weakened by the obvious variability that exists among computer science conference and by the lack of any real objective measure, comparable to journal impact, as a way to evaluate these conferences. There is certainly a lot of variability between computer science conferences. Some of the leading conferences have rejection rates as high as 70–90%, while others reject significantly less papers than they accept. However, the rejection rate of a conference, the argument usually goes, is not an adequate measure of quality. A more objective measure of conference quality is needed, one that can be readily computed, and that approximates a measure of conference impact.

3 Methodology

Until recently, the ability to offer large-scale bibliographic database services, such as the type of citation analysis at the foundation of journal impact assessment, was limited to organisations such as ISI/Thompson Reuters, who maintained citation databases covering thousands of academic journals. This service is now available online via ISI’s Web of Knowledge service, which provides researchers and other interested parties with access to a wide range of bibliographic and citation analysis services. However, the coverage of the ISI’s Web of Knowledge is limited. It focuses mainly on journals and, as such, excludes many of the common publications targeted by computer science researchers.

Recently the emergence of services such as CiteSeer and Google Scholar have helped to address this gap by maintaining online citation databases that provide better coverage of conference proceedings as well as journals?. For example, Google Scholar automatically extracts citation information from a wide range of online publications sources, from large-scale professional services such as Springer-Link to the publication lists that are

---

1 ISI’s Web of Knowledge website is here: [www.isiknowledge.com](http://www.isiknowledge.com)
2 CiteSeer’s webpage is here: [http://citeseer.ist.psu.edu](http://citeseer.ist.psu.edu)
3 Google Scholar’s webpage is here: [http://scholar.google.com/](http://scholar.google.com/)
maintained by individual researchers on their home pages. As a result it is now feasible
to perform large-scale citation analyses by mining the Google Scholar database. In this
section we describe one such study, which covers more than 8,000 articles from a wide
range of conferences and journals, in an attempt to develop a measure of impact, based
on Google Scholar data.

3.1 ISI Impact

By far the most popular single score used to assess the impact of research journals is the
ISI journal impact factor. This is based on the seminal work on citation analysis by 

In a given year the ISI impact of a journal will be the average number of citations in the
previous year to articles in the two preceding years. In 2008 the ISI impact factor for
a journal is the average number of citations found in 2007 publications to papers from
that journal published in 2005 and 2006. In our analysis we use the average of the three
most recent published ISI scores, i.e. the ISI impact factors from 2005, 2006 and 2007.

Because of what the ISI Impact factor sets out to do – assigning a numeric score on
the quality of scientific publications – it has inevitably drawn a lot of criticism over the
years (???). Some of the key criticisms are (?):

- The impact factor correlates poorly with the citation counts for individual articles.
- There is no correction for self-citation, either at an author or journal level.
- The coverage of the database is not complete.
- The score is based on a short time window.
- Long papers with long lists of references (e.g. review papers) distort the score.

Some of the criticisms of the ISI Impact factor are addressed by Garfield in a paper
published in the British Medical Journal (??). In particular he addresses the issue that
this score is based on publications in a two year time window and thus may not be
appropriate for disciplines with longer research cycles and a longer citation ‘half-life’.
Garfield demonstrates that the ranking of medical journals based on a cumulated impact
factor (cumulated over a 14 year period) correlates well with the ISI Impact as actually
calculated.

In this paper we are not concerned with criticising the ISI Impact factor, instead we
wish to demonstrate that our impact factor based on Google Scholar (GS) correlates well
with it. A sample of the ISI impact factors for a subset of computer science journals are
presented in Table ??; this table also includes the GS impact factors for these journals,
which we will return to in a later section. For now it is sufficient to note that these
journals span a range of ISI impact factors with the top ranking International Journal of
Computer Vision enjoying an impact of 4.37 while Artificial Intelligence for Engineering
Design, Analysis and Manufacturing presents with a more modest impact of 0.27. In the
period under consideration (2005 to 2007) the median ISI impact factor for computer
science journals varied between 0.80 and 0.84. Given this median we divide these journals
into three tiers according to their impact factors: high-ranking (A*) journals have impact factors $\geq 2$; medium-ranking (A) journals have impact factors $\geq 0.9$ but $< 2$; low-ranking (B) journals have impact factors $< 0.9$. So our A and A* categories represent journals in the top half of the ISI ranking. Computer science journals not considered in the ISI ranking (of which there are many) are not considered in the study.

<table>
<thead>
<tr>
<th>Category</th>
<th>Journal Name</th>
<th>ISI IF</th>
<th>GS IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*</td>
<td>International Journal of Computer Vision (IJCV)</td>
<td>4.37</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)</td>
<td>3.9</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Journal of the ACM (JACM)</td>
<td>2.73</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Artificial Intelligence (AI)</td>
<td>2.63</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Machine Learning (ML)</td>
<td>2.47</td>
<td>32.5</td>
</tr>
<tr>
<td>A</td>
<td>IEEE Trans. on Knowledge and Data Engineering (TKDE)</td>
<td>1.93</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>AI in Medicine (ARTMED)</td>
<td>1.77</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Information Retrieval (IR)</td>
<td>1.5</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>Computational Intelligence (CI)</td>
<td>1.3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Data and Knowledge Engineering (DKE)</td>
<td>1.2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Decision Support Systems (DSS)</td>
<td>1.07</td>
<td>17.5</td>
</tr>
<tr>
<td>B</td>
<td>AI Communications (AICOM)</td>
<td>0.57</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>International Journal of Pattern Recognition (IJPRAI)</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Pattern Analysis Applications (PAA)</td>
<td>0.5</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)</td>
<td>0.27</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: This table summarises the impact factors of the journals included in the evaluation. The column ISI IF gives the mean 2005-2007 ISI Impact Factor and the column GS IF gives the impact factor we have calculated based on Google Scholar.

### 3.2 Google Scholar & DBLP

Google Scholar is an attempt to bring Google-like search to bibliographic data. As such it provides a familiar Google-style search interface, allowing searchers to locate research articles based on a wide range of features, including article features (author, title etc.), publication features (type, year, name etc.), and a range of subject filters. Google Scholar responds to a query with a result-list in which each result corresponds to a particular article. For example, Figure ?? presents a portion of the results for a search for articles by **Marvin Minsky**. Each result includes information about the article type, the article title, the authors and publication data. Importantly, each result also contains citation data. For example, the **first** result, which corresponds to Minsky’s famous “A Framework for Representing Knowledge” article (?), indicates that Google Scholar has found 3350 citations to this book in its database. Moreover, following this citation link
will lead directly to a list of these 3350 citations.

Figure 1: The results of a Google Scholar query for documents authored by Mavin Minsky.

Thus Google Scholar provides access to the necessary bibliographic data to perform a large scale citation analysis of a wide range of computer science conferences and journals. – given an article Google Scholar will provide its citation count and full details on the other articles that cite it – all that is required is a suitable list of articles to seed the analysis. This is where a service like the Digital Bibliography & Library Project (DBLP) comes in to play (??). DBLP is a publication database of a different type. It is a large archive of publication records, documenting the papers published in a wide range of conferences, journals, as well as some workshops and symposia. DBLP does not provide any citation data but it can be used to provide a suitable list of seed articles for analysis.

3.3 A Measure of Impact Based on Google Scholar

For the purpose of our study we chose to focus on articles from conferences and journals in the areas of Artificial Intelligence and Machine Learning. The reasons are two-fold. First, both areas are very mature and active areas of research within computer science.
7. ICCBR 2007: Belfast, Northern Ireland, UK


Invited Papers

- Hans-Dieter Burkhard, Ralf Berger:
  Cases in Robotic Soccer. 1-15
  Electronic Edition (link) Șănceş
- Larry Kerschberg, Hanjo Jeong, Yong Uk Song, Wooyu Kim:
  A Case-Based Framework for Collaborative Semantic Search in Knowledge Sifter. 16-30
  Electronic Edition (link) Șănceş
- Eva Armengol:
  Usages of Generalization in Case-Based Reasoning. 31-45
  Electronic Edition (link) Șănceş

Research Papers

- Raquel Ros, Ramon López de Mántaras, Josep Luis Arcos, Manuela M. Veloso:
  Team Playing Behavior in Robot Soccer: A Case-Based Reasoning Approach. 46-60
  Electronic Edition (link) Șănceş
- Sutanu Chakraborti, Nirmana Wiratunga, Robert Leblanc, Stuart Watt:
  Acquiring Word Similarities with Higher Order Association Mining. 61-76
  Electronic Edition (link) Șănceş
- Klaus Brinker, Eyke Hüllermeier:
  Label Ranking in Case-Based Reasoning. 77-91
  Electronic Edition (link) Șănceş

Figure 2: A screenshot showing DBLP’s page for ICCBR 2003. The conference’s details are shown as well as a list of all of its papers. Here is listed three invited talks and the first three of 51 full papers (listed as Scientific Papers here).
Second, they are areas that are very familiar to the authors of this paper, which proved useful when it came to selecting appropriate seed conferences and journals and also when it came to independently verifying the results of the study.

For this study then, we chose to focus on 15 conferences and 15 journals between the years of 2000 and 2003, inclusive; see Tables ?? and ?? . These conferences and journals included first, second, and third tier venues, roughly in line with ISI’s rankings. All of the listed conferences and journals were covered by DBLP and the 15 conferences provided access to 3,258 articles, while the 15 journals provided access to 5,506 articles.

Each article was extracted from the DEP XML records and submitted in turn to Google Scholar. Using Google Scholar’s advanced search options queries were constructed to return the Google scholar entry for each seed article. Each query was made up of the title, author names and year of publication. Some initial processing was carried out on the article details to construct a query. All punctuation and non standard characters were removed from titles and the year of publication was converted into a 2 year window rather than a fixed year. Google Scholar returned a list of 50 potential matches and an article was identified when the title, year and at least one author name are matched. Once identified, the necessary citation data was extracted for each article by iterating through the list of articles reported to cite the article to produce our Google Scholar impact factor. The “Cited By x” figure reported by Google for each article was often inaccurate so each citing article was retrieved, checked and counted in turn. In some cases Google listed articles more than once in a list, in others the number of articles in the list simply did not match the figure shown. It was usually the case that the number reported was 1-2 less than the number presented in the list. We found that 10.5% of the articles in our seed sets were not indexed by Google Scholar, 5.4% of conference papers could not be retrieved versus 6.2% of journal articles.

During this process care was taken not to overload the system with requests and the data was gathered over a time frame of a few weeks in early 2008. As mentioned Google Scholar compiles its repository from professional services and researcher maintained pages. In terms of coverage we found that Google Scholar successfully returned citation lists for 89.5% of our seed set with the missing articles being evenly distributed between the journal and conference seed paper sets. The end result is that each article successfully located using Google Scholar is associated with a total number of citations, and these basic citation counts can then be aggregated at the level of individual conferences (or conference series) and journals.

Table ?? shows both the ISI impact factors and our Google Scholar impact factors for the 15 test journals for which we had reliable impact factors available from ISI. A key question concerns the correlation between the ISI impact and Google Scholar impact. If there is not a strong correlation between these impact factors then we cannot be confident that the Google Scholar impact factor is capturing the essence of impact. These ISI and Google Scholar impact factors are plotted in Figure ?? and it should be clear from

\(^{4}\)The DBLP XML records are available here: [http://dblp.uni-trier.de/xml/](http://dblp.uni-trier.de/xml/)
Table 2: The overall Google Scholar Impact Factor and years covered (this refers to the years each conference was held) for each conference proceedings published during the years 2000-2003

<table>
<thead>
<tr>
<th>Conference Name</th>
<th>GS IF</th>
<th>Years Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAI</td>
<td>20</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>IJCAI</td>
<td>17</td>
<td>2001, 2003</td>
</tr>
<tr>
<td>AH</td>
<td>17</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>ECCV</td>
<td>13.5</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>COLING</td>
<td>9</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>ECCBR</td>
<td>8</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>ECAI</td>
<td>7</td>
<td>2000, 2002</td>
</tr>
<tr>
<td>ICEBER</td>
<td>6</td>
<td>2001, 2003</td>
</tr>
</tbody>
</table>

4 Comparison of Journal and Conference Citations

In the previous section we proposed a straightforward metric for assessing the impact of scientific publications based on an aggregate citation count using Google Scholar citation data. We validated this GS impact factor by demonstrating a strong correlation between this metric and the more standard ISI impact factor, across a validation-set of journals. In this section we describe the GS results over the larger set of conferences and journals,
Figure 3: A scatter plot showing the correlation of the Google Scholar Impact Factor and the ISI Impact Factor.
which form the basis of our more comprehensive analysis.

<table>
<thead>
<tr>
<th>Journals</th>
<th>GS</th>
<th>Conferences</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>A</em> Journals</em>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JACM</td>
<td>25</td>
<td>AH, IJCAI</td>
</tr>
<tr>
<td>PAMI</td>
<td>20</td>
<td>ICML, ECCV</td>
</tr>
<tr>
<td>ML</td>
<td>15</td>
<td>UAI, NIPS</td>
</tr>
<tr>
<td><strong>A Journals</strong></td>
<td></td>
<td>COLING, ECCB, TREC</td>
</tr>
<tr>
<td>DSS</td>
<td>10</td>
<td>IJCAI</td>
</tr>
<tr>
<td>TKDE</td>
<td></td>
<td>ECCB, TREC</td>
</tr>
<tr>
<td>ARTMED</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI, INFFUS, PE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DKE, PAA</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>B Journals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIEDAM</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>AICOM, IJPRAI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICANN</td>
</tr>
</tbody>
</table>

Figure 4: This diagram presents a unified picture of CS journals and conferences ranked according to our Google Scholar impact factor.

As mentioned previously we have divided our validation-set of journals into three basic classes, A*, A, and B, according to their ISI impact factor. Having computed the GS impact factors for these journals we can now place them on a continuum, as shown in Figure ???. Figure ?? also shows the positions of the 15 test conference series on this same continuum as the basis of a more direct comparison between journals and conferences.

Overall the results clearly show that many of the conferences under evaluation perform well compared to the benchmark journals. Clearly the A* journals stand out in their ability to attract citations and we note how the leading journals such as the Journal of the ACM, Pattern Analysis and Machine Intelligence, and Machine Learning succeed
in obtaining GS impact factors above 30. That being said, it is interesting to note how
well many of the conferences perform, particularly in relation to the category A jour-
nals. Top of the conference ranking is AAAI, with a GS impact factor of 20 placing it
in the lower reaches of the A* journals so that it compares favourably with the Journal
of Artificial Intelligence. The category A journals correspond to GS impact factors from
8 to 19, which accommodate a wide range of computer science conferences, from the
International Joint Conference on Artificial Intelligence (IJCAI) and the International
Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH), with GS
impact factors of 17, down to the European Conference on Case-Based Reasoning (EC-
CBR) and the Text Retrieval Conference (TREC), which achieve GS impact factors of
8.

5 Conference Impact and Rejection Rates

One commonly held view is that conference rejection rates can be seen to serve as a
useful proxy for future impact 5. Indeed rejection rates are sometimes accepted for this
purpose in various research assessment exercises, including academic promotions. It is
useful then to consider the relationship between conference rejection rates and expected
citation count, based on our GS impact factor, to see whether this view holds up. Figure
?? presents these data as a scatter plot of GS impact factor against conference rejection
rates for 23 individual conferences across the 15 conference series under evaluation. Note
that the data points in Figure ?? reflect a subset of the full set of conferences, namely
those conferences for which reliable rejection rates could be obtained.

While the results indicate that there is some correlation between GS impact factor
and rejection rates the Pearson correlation score of 0.54 is not very convincing and
reflects the considerable variation that exists when it comes to the relationship between
conference rejection rates and the ability of conferences to attract citations. There
are conferences with similar rejection rates but which achieve very different GS impact
factors. Equally, there are conferences with very different rejection rates which achieve
very similar GS impact factors. For instance, AAAI achieves a GS impact factor of 20
with a rejection rate of about 65-75%. By comparison, its European sister conference,
ECAI, is just as selective but achieves a median GS impact factor of only 7 (as shown in
Table ??). As another example, the 2002 European Conference on Case-Based Reasoning
(ECCBR) had a rejection rate of approximately 0.33 and a GS impact factor of 7,
achieving a citation rate that is better that conferences such as ECML 2000 and ECAI
2002, which have twice its rejection rate.

It is interesting to note the apparent bias that exists between well cited U.S. confer-
ences and their similarly selective, though less well cited European counterparts. The
AAAI-ECAI example above is a case in point. Both conferences target the same re-
search area and attract their submissions from a similar community of researchers in a
way that is equally selective. And yet the US-centric AAAI enjoys an expected citation

5Further information on the sources of these rejection rates are available on the living dataset website
for this research at [http://lorcancoyle.org/research/citationanalysis](http://lorcancoyle.org/research/citationanalysis)
Figure 5: Scatterplot of GS impact factors versus rejection rates for a subset of conferences where reliable rejection rates were available.

<table>
<thead>
<tr>
<th>Year</th>
<th>AAAI GS</th>
<th>AAAI RR</th>
<th>ECAI GS</th>
<th>ECAI RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>24.5</td>
<td>67%</td>
<td>9</td>
<td>69%</td>
</tr>
<tr>
<td>2002</td>
<td>14</td>
<td>75%</td>
<td>6</td>
<td>72%</td>
</tr>
<tr>
<td>Overall</td>
<td>20</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: GS impact factors of AAAI versus ECAI in 2000 and 2002 with corresponding rejection rates (RR). (AAAI and ECAI are both held every two years.)
count (computed from the product of the median citation count and the rejection rate of the conference) that is more than twice that of ECAI; see also Table ??.

This regional bias is also evident in another pair of related conferences, namely ICML and ECML; see Table ??°. Once again, the more US-centric ICML conference series is capable of attracting far more citations that a similarly selective Euro-centric ECML conference series. ICML has an expected citation count that is twice that of ECML. There are a few possible explanations for this disparity. One explanation could be that non-European researchers are likely to miss publications at European venues (e.g. ECAI or ECML) and consequently papers at these conferences pick up references from European researchers only. An alternative explanation is that there is a pre-selection process where researchers keep their best work for the international conference.

<table>
<thead>
<tr>
<th>Year</th>
<th>ICML GS</th>
<th>ICML RR</th>
<th>ECML GS</th>
<th>ECML RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>15</td>
<td>56%</td>
<td>6</td>
<td>57%</td>
</tr>
<tr>
<td>2001</td>
<td>16</td>
<td>68%</td>
<td>10.5</td>
<td>62%</td>
</tr>
<tr>
<td>2002</td>
<td>11</td>
<td>67%</td>
<td>6</td>
<td>63%</td>
</tr>
<tr>
<td>2003</td>
<td>14</td>
<td>68%</td>
<td>9</td>
<td>78%</td>
</tr>
<tr>
<td>Overall</td>
<td>14</td>
<td></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: GS impact factors of ICML versus ECML for the years 2000 and 2003 inclusively with corresponding rejection rates (RR).

To test the correlation between GS impact factor and rejection rate further we examined the data available for three conference series which took place in each of the four years covered by our study; UAI, ICML, and ECML are the only conferences that occurred and had published rejection rates available in every year of our study. Table ??° shows the GS impact factors and rejection rates for each year in the study. What is interesting here is that there is no significant correlation between rejection rate and GS impact factor. The Pearson score for UAI is only 0.27, for ICML it is -0.24 and for ECML it is 0.42. This suggests that, at least for these conferences, the yearly changes in rejection rates have little bearing on the expected citation count.

In summary then, these results highlight that any assumed relationship between conference rejection rates and the ability of conferences to attract citations is at best weak and in reality other factors must play a more critical role when it comes to influencing future citations.

6 Conclusions

Evaluating research output is a complex, challenging and contentious issue. Funding agencies argue the need for comprehensive research evaluation metrics so that they may
Table 5: Median GS impact factors for UAI, ICML and ECML from 2000 and 2003 with corresponding rejection rates (RR).

<table>
<thead>
<tr>
<th>Year</th>
<th>UAI GS</th>
<th>UAI RR</th>
<th>ICML GS</th>
<th>ICML RR</th>
<th>ECML GS</th>
<th>ECML RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>17</td>
<td>64%</td>
<td>15</td>
<td>56%</td>
<td>6</td>
<td>57%</td>
</tr>
<tr>
<td>2001</td>
<td>11</td>
<td>60%</td>
<td>16</td>
<td>68%</td>
<td>10.5</td>
<td>62%</td>
</tr>
<tr>
<td>2002</td>
<td>16</td>
<td>66%</td>
<td>11</td>
<td>67%</td>
<td>6</td>
<td>63%</td>
</tr>
<tr>
<td>2003</td>
<td>10</td>
<td>66%</td>
<td>14</td>
<td>68%</td>
<td>9</td>
<td>78%</td>
</tr>
</tbody>
</table>

objectively assess the return on their research investment, and increasingly, academic institutions are relying on similar metrics when it comes to guiding academic promotions. When it comes to publication output there is general consensus about the importance of citations as an important factor when it comes to evaluating research papers and a number of well-defined objective metrics are commonly used. For example, the ISI maintains comprehensive records of the citations attracted by leading academic journals, providing the raw materials for tried and tested metrics such as impact factor. Unfortunately not all disciplines are equally well served by such approaches, especially given their bias towards journal papers. In this paper, for example, we have highlighted how computer science research has traditionally placed a greater emphasis on conference publications and how, as a result, computer science researchers can suffer when it comes to ISI-based research assessment. Simply put, conference papers are generally excluded from such evaluation exercises.

In this paper we have examined the issue of publication quality from a computer science/engineering perspective. Our main focus has been to justify the common publication practices in these disciplines by demonstrating how leading computer science/engineering conferences can attract significant citations scores that are in-line with leading journals. We do this by performing a citation analysis on almost 9,000 conference and journal papers, drawing on citation data from Google Scholar, and aligning these citations with ISI journal rankings. The results highlight a number of important points:

1. There is a strong correlation between the citation scores computed from the Google Scholar data and comparable data from the ISI index, thus validating the use of our new GS score as an alternative citation-based evaluation metric that is applicable to both journals and conferences.

2. The conferences surveyed in this analysis perform well in comparison to the journals. A significant number of conferences achieve median citation rates that are comparable with category A (ISI) journals for example.

3. The long-held view that conference rejection rates are a good proxy for conference quality did not hold up to scrutiny as we found a low coefficient of correlation between the rejection rate of a conference and its GS score.
4. There is evidence of a strong regional bias between similar conferences, with US-centric conferences attracting much greater citation scores than their non-US counterparts.

7 Acknowledgments

This work has been supported by Science Foundation Ireland through grants 07/CE/I1147, 04/RPI/1544, 03/CE2/I303.1 and 05/IN.1/I24.
Bibliography


