Again, the Role of Conference Papers in Computer Science and Informatics

Luís Caires
NOVA Laboratory for Computer Science and Informatics
Departamento de Informática — Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

December 5, 2015

Abstract

It is often claimed that in the Computer Science and Informatics many top tier conferences are high profile venues, with journal-equivalent status or even higher. The goal of this note is to highlight further evidence, based on recent publicly available indicators from the UK REF 2014 evaluation exercise and Google Scholar Metrics, of the special nature and purpose of conference venues in CS and related fields, not shared by most conference venues in other fields of science and engineering. We believe that our analysis is particularly interesting for anyone involved in research evaluation at large, but not necessarily familiar with the CS field. In particular, we focus on the comparison between the status of conferences and journals in different science and engineering fields, and discuss the status and role of conference and journal papers within the CS field itself.

1 Introduction

Scientific publication practices vary between different research fields and communities, and even between different subfields and communities of a single major field, with differences of various accounts being frequently reported. Whereas in the Natural Sciences journal are the key publication channel, other research fields credit and assign significant levels of prestige to a wider variety of publication channels, such as, for example, books and chapters in the Social Sciences.

In Computer Science and Informatics (CS), many top tier conferences are highly respected publication venues, arguably with journal-equivalent status [1].Core conferences in the CS field, only publish original research, as journals do, forbid simultaneous submission to other conferences and journals, and their review process is through and detailed, with two phase review processes, mas journals do. The situation is often misunderstood by other science and engineering fields, which, because the term "conference" is so much overloaded, may develop a biased appreciation of CS publication profiles [7].

The main goal of this note is to record compelling confirmation, based on publicly available data extracted from Google Scholar Metrics and the UK REF 2014 evaluation reports, of the important role of conference publication venues in CS, not comparable with the generally quite different, and sometimes much less

important role of conference venues as publication outlets for original research, in other fields of science and engineering. The evidence reported here may be of particular interest to researchers or research managers concerned about research evaluation at large, but not necessarily familiar with the CS field.

2 Status of conferences: informal evidence

Because conferences are a dissemination outlet present in almost all field of Science, even if more frequent in engineering disciplines, it is easy, even for an informed observer, to vaguely equate the nature of conference venues among fields. This is a serious, even if understandable, mistake. For example, conferences in most engineering fields, including the most prestigious ones, usually play the role of dissemination and networking events, focusing on the presentation of research abstracts, invited talks and keynotes.

On the other hand, certain top tier CS conferences, notwithstanding also combining a component of dissemination and networking, also are bona-fide periodic channels for the publication of original world-class research, and implement detailed evaluation processes, in parallel with the traditional peer-reviewed journal system. In CS, top conference venues are highly prestigious and selective [1] and play, in the perception of the community and individual researchers, a role equivalent to prestigious and selective journals in other science and engineering fields in which conferences, if available, always play a secondary role.

Top tier CS conferences always have a clear international scope, their scientific program committees include leading and rising star researchers from all over the world, with membership overlapping with editorial boards of top journals in the field. They are organised periodically (annually or biannually) at a well-defined season, for many decades now, with the oldest ones founded in the late 60s of the XX century, and endorsed by established international scientific associations, such as ACM, IEEE, AAAI, EATCS, EAPLS, ALP, ETAPS. They tend to be very selective. They provide to researchers, including to those playing scientific leadership roles, the most visible competition arena for getting their results revealed, accepted and validated by the community. Indeed, conferences are also venues of choice for publication of seminal contributions, being clear that many of the most relevant research accomplishments, including some that have shaped the field, have been and continue to preferably be first published at conferences. Computer scientists often struggle to keep their most exciting results under reserve, only to publish them at the next year's edition of some targeted top conference. This situation sharply contrasts with the motivation that drives, e.g., natural scientists such as chemists to feature their work at networking conferences in their field: here, contributions have usually been already published in relevant journals, the goal being essentially to promote their dissemination and impact, with researchers often refraining from presenting new results to avoid loosing their lead.

In general engineering and natural sciences conferences the peer-review evaluation process is mostly based on abstracts (some times 1 page or a few paragraphs long). In the case of CS conferences, no abstracts are accepted as regular submissions, only full papers, reporting on original research, are considered (around 15-25 pages, taking ACM or Springer-Verlag LNCS formats as reference; while some top conferences also accept short papers and posters, these

pieces appear in separated tracks and sections in the proceedings, where they are clearly distinguished from the main contributions). Evaluation and selection is performed by means of thorough peer-reviewing processes, which normally involves many external reviewers and lengthy discussions and consensus in the program committee final meeting. This means that conference program committees work hard in every such venue to evaluate and select submissions. Review processes, usually takes around 3 months from submission to notification, roughly the same amount of time mediating between submission and publication in prestigious journals in high impact natural sciences and engineering fields. The final revised version of the accepted (full) papers are finally collected in proceedings, either as volume or special journal issue edited by an well-known international scientific publisher. For some other cultural reasons, review processes for journal publications in CS often exceed 1 year, which, as frequently recognised, is a bit inconvenient for a fast evolving field such as CS.

Acceptance rates for top CS conferences are publicly known by the community, and usually clearly stated in the preface of proceedings, or found elsewhere, e.g., the ACM Digital library also list historical acceptance rates for some relevant conferences. The number of available publication slots varies between conference venues, but it is often in the range 25-50 papers, which means that the competition overall is really fierce. Some top quality venues may accept a larger quantity of papers, but not at the cost of raising acceptance rates.

For the most selective top tier conferences one may find acceptance rates between 10% and 25%. Excellent papers may get rejected at a conference just because they do not fit in the previously defined acceptance rate, this may be frustrating to their authors, who certainly do not face the same challenge in their journal submissions (acceptance rates for journals, in any field, are not easy to find). Publishing in certain CS conferences may be more challenging than in some top journals. We may cite a few examples: ICSE 2015 (18.5% acceptance rate), POPL 2014 (23% acceptance rate), ESOP 2014 (25% acceptance rate), EuroSys 2013 (15% acceptance rate), SOSP 2011 (17% acceptance rate), IJCAI 2011 (17% acceptance rate), CHI 2014 (23% acceptance rate). There are many other highly relevant conferences.

This state of affairs is also reflected on impact indicators based on citations. Remarkably, it seems that neither acceptance rates for conferences nor trustable citation data for conference papers are easily available on disciplines other than CS. While significative levels of prestige seem to be assigned by the respective communities at least to some conferences in several engineering fields, acceptance rates for those are rarely published if ever, reflecting a lesser (if any) concern with the role of conferences as a first-class venue for dissemination of original and competitive research results and breakthroughs. Most citations in the classical sciences and general engineering fields refer to journal articles, which is not surprising given that those are most relevant publication venues for new results. On the other hand, in Computer Science, top tier conference papers actually receive a substantial amount (if not sometimes the lion's share) of citations, and are highly visible in top positions in the h-index of leading researchers.

3 Status of conferences: statistical evidence

It is not easy to find many sources of information or reliable studies comparing the relevance of different publication venues, namely conferences versus journals across fields. In this note, we collect some observations on this matter, building on two recently made available open very large data sets. One is Google Scholar (GS) Metrics, which provides a per scientific field, 5-year window, impact analysis of various kinds of scientific venues, based on the underlying GS database. In this analysis, the top 20 venues with higher h5-index (last 5 years h-index) are presented in a sorted list. It is also possible to browse each venue's publications that contribute to each record. Although GS is sometimes accused from suffering from suboptimal data quality (e.g., occasional mistakes in citations), its coverage is extremely high, with around 99.3 million documents, 87% of the total number of scholar documents found on the web [5]. GS increase of data quality and stability over time [6], and having all raw data publicly available for scrutiny makes GS attractive for supporting broad observations as done here.

Our second source of information is the REF 2014 data set. The Research Excellence Framework is the system for assessing the quality of research in UK higher education and research institutions; REF 2014 refers to the last conducted evaluation exercise [10]. REF 2014 evaluated 154 universities and a more than 190 000 research outputs, following a careful preparation process (started 2008). REF 2014 is extremely well documented, both in the website and in the many preliminary, interim, and final reports. Remarkably, not just overall evaluation results and reports have been openly published, but also all individual submitted research outputs, in particular publications classified by type, with full citation data.

Submissions were collected by units of assessment, which are mostly disciplinary fields, thus making the collected data also useful for comparative purposes. It is also interesting to notice that REF 2014 follows a sampling methodology based on short-lists of 5 research outputs per researcher. Each researcher contribution for an institutional submission is restricted to just five research outputs – e.g., books, papers or patents - from the previous five years, thus allowing visibility only to the most important and relevant items, in the perspective of each researcher and or team. The particular choices made by researchers while selecting such top five outputs have left a trace in the REF 2014 submission data and allow us to observe the type and identity of publication venues that have been valued the most in terms of relevance, prestige, and impact by the various different fields, for the purposes of the REF 2014 evaluation exercise.

3.1 Google Scholar Metrics Venue Rankings

In this section, we observe the type – journal or conference – of the top 20 publication venues in the fields of science and engineering listed in the disciplinary ranked lists provided by Google Scholar Metrics (GSM). If one picks a sample subfield of Computer Science, say Software Systems, we find in the top 20 entries 11 conferences and 9 journals. In some subfields of Computer Science, the bias towards conferences is even higher, e.g., 13 conferences / 7 journals in Computing Systems and 14 conferences / 6 journals in Computer Security and Cryptography; in all other subfields the balance between conferences and top journals is always relevant. In subfields such as Computing Systems, Computa-

Google Scholar Metrics field designation	**C**	**J**	**%C**
Artificial Intelligence	4	15	21
Automation & Control Theory	2	18	10
Aviation & Aerospace Engineering	3	17	15
Bioinformatics & Computational Biology	1	19	5
Biomedical Technology	0	20	0
Ceramic Engineering	0	20	0
Civil Engineering	0	20	0
Combustion & Propulsion	0	20	0
Computational Linguistics	15	5	75
Computer Graphics	7	13	35
Computer Hardware Design	12	8	60
Computer Networks & Wireless Communication	5	15	25
Computer Security & Cryptography	14	6	70
Computer Vision & Pattern Recognition	9	11	45
Computing Systems	13	7	65
Data Mining & Analysis	9	11	45
Databases & Information Systems	13	7	65
Environmental & Geological Engineering	0	20	0
Evolutionary Computation	7	13	35
Human Computer Interaction	10	10	50
Information Theory	13	7	65
Manufacturing & Machinery	1	19	5
Materials Engineering	0	20	0
Mechanical Engineering	0	20	0
Metallurgy	0	20	0
Microelectronics & Electronic Packaging	7	13	35
Mining & Mineral Resources	0	20	0
Multimedia	10	10	50
Nanotechnology	0	20	0
Ocean & Marine Engineering	2	18	10
Power Engineering	6	14	30
Quality & Reliability	3	17	15
Remote Sensing	3	17	15
Robotics	4	16	20
Signal Processing	3	17	15
Software Systems	11	7	61
Structural Engineering	0	20	0
Sustainable Energy	0	20	0
Theoretical Computer Science	7	13	35
Transportation	2	18	10
Water Supply & Treatment	0	20	0

Table 1: GSM split of top 20 h5-index venues in Engineering and CS fields.

tional Linguistics, Computer Security & Cryptography, Data Mining & Analysis, Databases & Information Systems, Human Computer Interaction, Software Systems, and Theoretical Computer Science, conference venues appear among the top 2 leading publication venue, and in the very top publication venue in all of these fields except one. Considering all subfields in Computer Science we find in average a 50/50 split between conference and journal venues.

The balance between journal and conference venues is quite different in other fields. For example, in Chemical & Materials Science, no conference is listed among the top 20 venues for any of the 19 subfields considered (thus, among approximately 380 venues). The same can be observed for Life Sciences & Earth Sciences, for any of the considered 39 subfields (approximately 780 venues), except for the presence of 3 conferences in the Bioinformatics & Computational Biology subfield. In Physics & Mathematics (covering 10 Subfields in Math and 14 subfields in Physics, for a total of approximately 480 venues), no conference venue is listed other than 1 conference in Computational Geometry (Math) and 9 conferences in Quantum Mechanics, Optics & Photonics, Acoustics & Sound (Physics), covering topics of Signal Processing, Speech Recognition, Quantum Information, Fibre Optics and Communications, and Quantum Electronics.

In some fields of Engineering other than CS we also find evidence of relevance for at least some conference venues. Examples include Computer Hardware Design with 12 conferences and 8 journals in the top 20 positions. Microelectronics & Electronic Packaging, with 7 conferences / 13 journals, Robotics, with 4 conferences and 16 journals, and Bioinformatics & Computational Biology with 3 conferences and 17 journals. In no field except Computer Science and some subfields of Electrical and Computer Engineering conferences seem to significantly appear in the top 20 positions. And when they do, the presence is residual, never in a leading position, or refer to CS biased venues. An example is Quality & Reliability, with 3 conferences and 17 journals, where the mentioned conferences are actually CS conferences on software testing and reliability topics.

Noticeably, the top venue lists for Civil Engineering, Structural Engineering, Environmental & Geological Engineering, and many others do not refer to any conference. A more detailed overview of the collected statistics for subfields of Engineering & Computer Science is shown in Table 1.

3.2 REF 2014 Submission Data

Data collected from the UK REF 2014 [10, 11] also supports an assessment of the breakdown between journal and conference submissions from a rather different, yet quite informative, as it turns out, perspective. The input dataset consists of only 5 outputs per FTE researcher, classified by types. In our analysis, we report on journal (type D) and conference (type E) publications, these types of output group the vast majority of submissions in science and and engineering. Overall, the collected statistics are summarised in Table 2.

The data displayed is consistent with the conclusion that the relative relevance of journal and conference publications differs substantially among scientific fields and subfields. Moreover, the data also substantiates the conclusion that the relevance of conference publications for the Computer Science and Informatics field is much higher when compared to other fields. The REF 2014 data gives particularly strong evidence for this fact, given that researchers were required to pick just their top 5 results. Clearly, only in the Computer Sci-

REF 2014 Unit of Assessment (UOA)	**C**	**J**	% **C**
Aero., Mech., Chem., and Manuf. Engineering	24	4110	1
Biological Sciences	4	8577	0
Chemistry	2	4688	0
Civil and Construction Engineering	16	1348	1
Computer Science and Informatics	1902	5555	26
Earth Systems and Environmental Sciences	4	5195	0
Electric & Electronic Eng., Metallurgy, Materials	28	3984	1
General Engineering	89	8553	1
Mathematical Sciences	17	6731	0
Physics	18	6376	0

Table 2: Journal / Conference papers in REF 2014 UOA

ence and Informatics field the presence of conferences seems to be statistically meaningful, with more than one out of four publications being a conference submission, while in other areas the proportion seem to be only incidental or very residual, in the range of one in one hundred.

That theses differences show up so clearly in the top 5 selections by researchers and institutions is a clear indicator of the high relevance conference publications actually have in the perception of the CS community, given that so many good journal publications were clearly superseded by those in conferences in the competition for the top 5 items. Moreover, we can notice that, e.g., 96% of all UCL submissions (including 97 conference and 164 journal articles) where graded at 4* (61%) level and 3* (35%) level [13]. It is also interesting to observe the distribution of conference / journal submissions to the REF 2014 unit of assessment Computer Science and Informatics for the departments classified in the top 15 positions, which we present in the Table 3. Notice that the average proportion of conference papers in these 15 (top) departments is a bit higher (33%) than the UK average (25%), even if there is a considerable variety in distribution of conference versus journal papers. Recall that this data refers to the supposedly best five contributions of each FTE researcher integrated in these UK Computer Science and Informatics research institutions.

We conclude that the presence of conference publications in the REF 2014 Computer Science and Informatics UOA is substantial (26%), unlike in all other science and engineering research fields, in which they are clearly residual ($\leq 1\%$).

4 Concluding Remarks

Ultimately, what really matters for each scientific community is the actual scientific content, the formal structure, and the quality of the review process of any published material reporting on research results, not so much the adopted publication channel. For some fields, it is just wrong to think that only journal publications are considered the most relevant, or that journal publications are always ranked higher than conference publications, as that may depend a lot on the particular journals and conferences considered. In other fields, it would be just non-sensical to assign journal status to conference publications. All this crucially depends on the particular nature of conferences in each field.

University	**C**	**J**	% **C**	REF 2014 Rank
University College London	97	164	37	1
Warwick	15	84	15	2
Imperial	63	137	32	3
Manchester	31	148	17	4
Sheffield	20	89	18	5
Cambridge	107	67	61	6
Oxford	73	186	28	7
York	33	104	24	8
Newcastle upon Tyne	34	64	35	9
Liverpool	20	73	22	10
Queen Mary	21	62	24	11
Lancaster	60	60	50	12
Nottingham	44	96	31	13
Kings College London	21	134	14	14
Edinburgh	184	197	48	15
Total	823	1665	33	

Table 3: Journal / Conference papers in REF 2014 top 15 CS Departments

Based on some publicly available data from Google Scholar Metrics and the UK REF 2014 submission statistics, our overview corroborates the particular high status of certain conferences in the field of Computer Science and Informatics, which are considered equivalent to journals in terms of publication prestige.

The observations summarised in this note also confirm that Computer Science and Informatics, some subfields of Electrical and Computer Engineering, and several computing-oriented subfields of other disciplines share a specific publication culture that highly values top conferences as publication venues. Likewise, we have noticed that in traditional engineering disciplines where conferences are also organised, for example Chemical or Civil Engineering, the culture is quite different in that conferences not really count as fully fledged publication venues for original research, or may be do so only very marginally.

Of course, research in Computer Science and Informatics is also routinely published in journal venues, even if the ratio of articles in journals versus conferences varies a bit among subfields. Some top CS conferences already publish their complete proceedings as journal issues (e.g., ICLP/TPLP [12], VLDB).

In fact, a growing number of proposals within the CS community have suggested mechanisms to formally conform papers accepted at selected CS conferences to the status of journal papers. A recent one, launched within ACM, suggest the creation of a new journal, tentatively named "Proceedings of the ACM", "a journal series specifically created to publish the proceedings of ACM's highest quality conferences" [4]. Such proposals are subject to ongoing debate [8, 3, 2], but evidence such as the REF 2014 submission profiles seems to clearly show how many CS conference proceedings are "de facto" being competitively shortlisted alongside journal venues, a very distinctive profile. Independently of what the future might bring in this particular case, it is clear that different publication cultures do exist, and, while they do, they should be seriously considered and better understood by the various scientific communities.

Acknowledgments. For useful suggestions or criticism, we would like to thank José Alferes, Rocco De Nicola, Alan Dix, João Leite, Nuno Preguiça, Marc Shapiro, Pedro Tavares, Jan Vitek, and Vasco T. Vasconcelos [...].

References

[1] Bertrand Meyer, Christine Choppy, Jrgen Staunstrup, Jan van Leeuwen. Research Evaluation for Computer Science. *Communications of the ACM, Vol. 52 No. 4, Pages 31-34*, ACM, 2009.

[2] David S. Rosenblum. The Pros and Cons of the 'PACM' Proposal: Counterpoint. *Communications of the ACM, Vol. 58 No. 9, Pages 44-45*, ACM, 2015.

[3] Kathryn S. McKinley. The Pros and Cons of the 'PACM' Proposal: Point. *Communications of the ACM, Vol. 58 No. 9, Pages 43-44*, ACM, 2015.

[4] Joseph A. Konstan, Jack W. Davidson. Should Conferences Meet Journals and Where?: A Proposal for 'PACM'. *Communications of the ACM, Vol. 58 No. 9, Page 5*, ACM, 2015.

[5] Madian Khabsa, C. Lee Giles. The Number of Scholarly Documents on the Public Web. *PLoS ONE 9(5)*, 2014.

[6] Anne-Wil Harzing. A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics, Volume 98, Issue 1, pp 565-575*, 2014.

[7] Moshe Y. Vardi. Conferences vs. Journals in Computing Research. *Communications of the ACM, Vol. 52 No. 5, Page 5*, 2009.

[8] J. Vitek, J. Gibbons. Who owns your research? Results of SIGPLAN Open Access survey SIGPLAN Executive Committe, 2015.

[9] Dan S. Wallach. Rebooting the CS Publication Process. *Communications of the ACM, Vol. 54 No. 10, Pages 32-35*, 2011.

[10] REF 2014 Manager's Report, 2015.

[10] REF2014. Research Excellence Framework 2014: Overview report by Main Panel A and Sub-panels 1 to 6, 2015.

[11] REF2014. Research Excellence Framework 2014: Overview report by Main Panel B and Sub-panels 7 to 15, 2015.

[12] Manuel V. Hermenegildo. Conferences vs. Journals in CS, what to do? Position paper for Dagstuhl meeting 12452: Publication Culture in Computing Research, 2012.

[13] Morris Sloman. Sub-panel 11 Computer-Science and Informatics REF Analysis, 2015.