DOI:10.1145/1859204.1859234

Massimo Franceschet

Virtual Extension
The Role of Conference Publications in CS

A bibliometric view of the publishing frequency and impact of conference proceedings compared to archival journal publication.

THE ROLE OF conference publications in computer science is controversial. Conferences have the undeniable advantages of providing fast and regular publication of papers and of bringing researchers together by offering the opportunity to present and discuss the paper with peers. These peculiar features of conferences are particularly important because computer science is a relatively young and fast-evolving discipline. The fundamental role of conferences in computer science is underlined with strength in the best-practices memo for evaluating computer scientists and engineers for promotion and tenure published in 1999 by the U.S. Computing Research Association[a] (CRA) and, more recently, in a study of the Informatics Europe, whose preliminary results are summarized in Choppy et al.[2]

Recently, *Communications* published a series of thought-provoking Viewpoint columns and letters that swim against the tide.[1,3,7,9] These contributions highlight many flaws of the conference system, in particular when compared to archival journals, and also suggest a game-based solution to scale the academic publication process to Internet scale.[7] Some of the mentioned flaws are: short time for referees to review the papers, limited number of pages for publication, limited time for authors to polish the paper after receiving comments from reviewers, and overload of the best researchers as reviewers in conference program committees. The result is a deadline-driven publication system, in which "*we submit a paper when we reach an appropriate conference deadline instead of when the research has been properly fleshed out*,"[3] that "*encourages and rewards production of publishing quarks—units of intellectual endeavor that can be generated, summarized, and reviewed in a calendar year*"[8] (interestingly, the author of the latter claim is CRA Board Chair Dan Reed). Furthermore, the current conference system "*leads to an emphasis on safe papers (incremental and technical) versus those that explore new models and research directions outside the established core areas of the conferences.*"[3] "*And arguably it is the more innovative papers that suffer, because they are time consuming to read and understand, so they are the most likely to be either completely misunderstood or underappreciated by an increasingly error-prone process.*"[1] Are we driving on the wrong side of the publication road? The question is raised by Moshe Vardi in a May 2009 *Communications* editor's letter.[9]

This article gives an alternative view on this hot issue: the bibliometric perspective. Bibliometrics has become a standard tool of science policy and re-

Table 1. Most prolific authors according to DBLP.

Author	Pubs	Journal	Conference	Venue
Philip S. Yu	547	177 (32%)	362 (66%)	ICDE (C, 49)
Chin-Chen Chang	509	318 (62%)	188 (37%)	Fund. Inf. (J, 34)
Elisa Bertino	494	180 (36%)	294 (60%)	TKDE (J, 31)
Thomas S. Huang	481	126 (26%)	346 (72%)	ICIP (C, 69)
Edwin R. Hancock	449	105 (23%)	340 (76%)	ICPR (C, 52)
Sudhakar M. Reddy	447	144 (32%)	303 (68%)	TCAD (J, 60)
Wen Gao	442	81 (18%)	360 (81%)	ICIP (C, 40)
Grzegorz Rozenberg	438	263 (60%)	109 (25%)	TCS (J, 73)
Alberto Sangiovanni-Vincentelli	426	122 (29%)	301 (71%)	DAC (C, 76)
Mahmut T. Kandemir	412	84 (20%)	326 (79%)	DATE (C, 33)
Mean	**464**	**160 (34%)**	**293 (63%)**	

Pubs: number of publications in DBLP; **Journal**: number of journal publications; **Conference**: number of conference publications; **Venue**: the venue at which the author published most of the papers. The additional information in this column is whether the venue is a journal (J) or a conference (C), and the number of papers published by the author at the venue. Source: DBLP. Date: August 1, 2009.

a http://www.cra.org/reports/tenure_review.html search management in the last decades. In particular, academic institutions increasingly rely on bibliometric analysis for making decisions regarding hiring, promotion, tenure, and funding of scholars. I investigate the frequency and impact of conference publications in computer science, comparing with journal articles. I stratify the set of computer science publications by author, topic, and nation; in particular, I analyze publications of the most prolific, most popular, and most prestigious scholars in computer science.

Table 2. Most popular authors according to the author h index.

Author	H	Pubs	Journal	Conference	Venue
Terrence J. Sejnowski	92	112	49 (44%)	63 (56%)	NIPS (C, 52)
Hector Garcia-Molina	89	370	112 (30%)	294 (69%)	SIGMOD (C, 29)
Tomaso Poggio	89	89	37 (42%)	50 (56%)	IJCV (J, 9)
Jeffrey D. Ullman	87	241	108 (45%)	123 (51%)	SIAM J. Comp. (J, 18)
Robert Tarjan	82	242	151 (62%)	91 (38%)	SIAM J. Comp. (J, 44)
Deborah Estrin	80	145	44 (30%)	100 (69%)	SenSys (C, 20)
Christos H. Papadimitriou	79	322	148 (46%)	170 (53%)	FOCS (C, 29)
Don Towsley	77	339	134 (40%)	205 (60%)	INFOCOM (C, 74)
Ian Foster	73	271	101 (37%)	168 (62%)	HPDC (C, 29)
Scott Shenker	71	170	41 (24%)	128 (75%)	SIGCOMM (C, 41)
Mean	**82**	**230**	**92 (40%)**	**135 (59%)**	

H: the h index computed on Google Scholar; **Pubs**: number of publications; **Journal**: number of journal publications; **Conference**: number of conference publications; **Venue**: the venue at which the authors published most of the papers. The additional information in this column is whether the venue is a journal (J) or a conference (C), and the number of papers published by the author at the venue. Sources: Google Scholar for the h index; DBLP for publication data. Date: August 1, 2009.

Table 3. Prestigious authors (winners of the ACM A.M. Turing Award; only winners during the past 10 years are shown).

Author	Year	Pubs	Journal	Conference	Venue
Barbara Liskov	2008	109	27 (25%)	80 (73%)	SOSP (C, 10)
Edmund M. Clarke	2007	221	67 (30%)	148 (67%)	CAV (C, 21)
E. Allen Emerson	2007	102	28 (27%)	71 (70%)	CAV (C, 10)
Joseph Sifakis	2007	114	25 (22%)	86 (75%)	CAV (C, 9)
Frances E. Allen	2006	13	6 (46%)	7 (54%)	IBM Sys. J. (J, 2)
Peter Naur	2005	32	25 (78%)	7 (22%)	CACM (J, 20)
Vinton G. Cerf	2004	23	11 (48%)	12 (52%)	CACM (J, 4)
Robert E. Kahn	2004	5	4 (80%)	1 (20%)	CACM (J, 2)
Alan C. Kay	2003	18	2 (11%)	16 (89%)	C5 (C, 5)
Ronald L. Rivest	2002	144	48 (33%)	89 (62%)	CRYPTO (C, 10)
Adi Shamir	2002	146	40 (27%)	105 (72%)	CRYPTO (C, 27)
Leonard M. Adleman	2002	49	14 (29%)	33 (67%)	FOCS (C, 11)
Ole-Johan Dahl	2001	11	4 (36%)	5 (45%)	FOCS (C, 11)
Kristen Nygaard	2001	9	2 (22%)	5 (56%)	ECOOP (C, 3)
Andrew Chi-Chih Yao	2000	128	64 (50%)	64 (50%)	FOCS (C, 23)
Frederick P. Brooks, Jr.	1999	43	14 (33%)	29 (67%)	SIGGRAPH (C, 9)
Mean		**73**	**24 (33%)**	**47 (65%)**	

Year: the award assignment year; **Pubs**: number of publications; **Journal**: number of journal publications; **Conference**: number of conference publications; **Venue**: the venue at which the authors published most of the papers. The additional information in this column is whether the venue is a journal (J) or a conference (C), and the number of papers published by the author at the venue. Sources: ACM Web site for ACM A.M. Turing Award winners; DBLP for publication data. Date: August 1, 2009.

The Frequency of Conference Publications in Computer Science

I use the DBLP computer science bibliography (faceted search) to retrieve journal and conference publication counts for scholars. Unfortunately, DBLP does not record citations. For the analysis of impact based on citations and on the h index,[5] I take advantage of Google Scholar and of Thomson-Reuters Web of Science enhanced with the brand-new Conference Proceedings index.

I start with the analysis of the publications of the most prolific computer science authors according to DBLP (see Table 1). Only two scholars, Chin-Chen Chang and Grzegorz Rozenberg, published more journal papers than conference papers. On average, two-thirds of the author publications are conference papers and one-third are journal articles.

Moreover, I analyzed the publications of the most popular computer science authors according to the author h index computed on Google Scholar. The h index of a scholar is the highest number h of papers published by the scholar that have each received at least h citations.[5] I took advantage of the h index compilation maintained by Jens Palsberg[b]; Table 2 illustrates the outcomes of the analysis. The frequency of conference papers is higher than the frequency of journal papers: on average, 59% of the author publications are in conference proceedings, and 40% are in journals. Only one author, Robert Tarjan, published more journal articles than conference papers. Notice, however, that the average share of journal articles is higher for popular scholars (40%) than for prolific scholars (34%). Furthermore, the average number of publications of prolific scholars (464.5) is two times higher than the mean number of publications of popular authors (230.1). Hence, high-impact scholars publish significantly less than prolific ones, and more frequently in journals.

Finally, I analyzed the publications of prestigious computer science scholars. I identified prestigious scholars as the winners of the ACM A.M. Turing Award

b The ranking is available at http://www.cs.ucla.edu/~palsberg/h-number.html (see Table 3). Once again, the share of conference publications (on average 65%) dominates that of journals articles (on average 33%). Only two authors, Peter Naur and Robert E. Kahn, published more in journals than in conference proceedings (notice, however, that I found only five publications for Robert E. Kahn in DBLP). The average share of journal publications for prestigious scholars (33%) is close to that of prolific authors (34%), but lower than the one for popular authors (40%). Furthermore, prestigious scholars published on average approximately one-third of the papers published by popular authors, and one-sixth of the articles published by prolific authors.

I offer two additional observations to conclude this section. The first one is well known to the computer science community: books do not represent a frequent publication in computer science (from 1% to 3% in the analyzed samples). This is a difference with respect to the humanities and to (some of) the social sciences. The second observation is more intriguing: the concepts of productivity, popularity, and prestige are well separate in computer science: the pairwise intersections of the corresponding top-10 compilations are always empty. The divergence between the bibliometric concepts of popularity and prestige is a known phenomenon for the hard sciences, but it is less evident for biology-medicine disciplines and for the social sciences.[4]

The Impact of Conference Publications in Computer Science

I analyzed the frequency and impact of conference and journal publications that contain the 10 most popular computer science topics (see Table 4). As found earlier, conferences are more popular than journals: for each topic in the list, the number of conference papers containing the topic phrase in title, keywords, or abstract is significantly greater than the number of journal papers containing the same topic. On average, 78% of the publications containing some of the hot topics are conference papers, and 22% of them are journal articles. Nevertheless, journal papers collect more citations (67%) than conference papers (33%). This means that journal papers have a much higher impact (on average, 5.41 citations per paper) than conference papers (on average, only 0.71 citations per paper). The higher impact of journals with respect to conferences is confirmed when inspecting the h index column: on average, at most 16 conference papers are cited at least 16 times, while at most 27 journals papers are cited at least 27 times. The topic with the highest conference impact is security, and that with the highest journal impact is data mining. Notice that both the number of citations and the h index for topic security are higher for conference papers than for journal papers, and this is the only exception among all topics. On the other hand, the topic with the lowest conference impact and that with the lowest journal impact is QoS.

Academic institutions increasingly rely on bibliometric analysis for making decisions regarding hiring, promotion, tenure, and funding of scholars.

Table 4. Most popular topics according to DBLP.

Topic	Pubs	Cites	Impact	H	J-Citer	C-Citer
genetic algorithms (C)	1598	1102	0.69	11	394 (39%)	621 (61%)
genetic algorithms (J)	653	4112	6.30	25	1964 (56%)	1563 (44%)
security (C)	6877	8244	1.20	28	1177 (21%)	4404 (79%)
security (J)	1221	5483	4.49	26	2950 (62%)	1809 (38%)
data mining (C)	2548	1752	0.69	13	564 (37%)	964 (63%)
data mining (J)	752	5513	7.33	27	2311 (51%)	2235 (49%)
simulation (C)	18280	10727	0.59	25	3749 (38%)	5998 (62%)
simulation (J)	4304	19198	4.46	36	9720 (59%)	6795 (42%)
clustering (C)	4070	3346	0.82	18	1086 (36%)	1894 (64%)
clustering (J)	1270	9064	7.14	32	3963 (53%)	3468 (47%)
scheduling (C)	3641	2721	0.75	15	780 (33%)	1599 (67%)
scheduling (J)	1147	5028	4.38	23	2036 (47%)	2315 (53%)
QoS (C)	2190	876	0.40	9	229 (28%)	581 (72%)
QoS (J)	594	2226	3.75	18	794 (41%)	1159 (59%)
Java (C)	1776	1594	0.90	13	309 (22%)	1074 (78%)
Java (J)	444	2275	5.12	24	1298 (61%)	817 (39%)
Internet (C)	5785	2464	0.43	12	704 (31%)	1561 (69%)
Internet (J)	1689	7396	4.38	28	3330 (54%)	2803 (46%)
neural networks (C)	4131	2760	0.67	14	1034 (42%)	1409 (58%)
neural networks (J)	1673	11335	6.78	36	5244 (61%)	3296 (39%)
Mean (C)	**5090**	**3559**	**0.71**	**16**	**1003 (33%)**	**2010 (67%)**
Mean (J)	**1375**	**7163**	**5.41**	**27**	**3361 (56%)**	**2626 (44%)**

Pubs: number of publications containing the topic in title, keywords, or abstract; **Cites**: total number of citations received by the publications; **Impact**: average number of citations per publication; **H**: value of the h index on the publication set; **J-Citer**: number of citing publications that are journal papers; **C-Citer**: number of citing publications that are conference papers. Each information is shown for both conference publications (C) and journal publications (J). Sources: DBLP for the most popular topics; Web of Science for publication and citation data. Target period: 2005–2006. Census date: August 1, 2009.

I also analyzed the citation patterns of conference and journal publications. Conference papers are strongly cited by conference papers (67%), while journal papers are cited more by journal papers (56%). However, journal papers are also significantly cited by conference papers (44%).

Finally, I analyzed the frequency and impact of conference and journal publications per country. I assigned a publication to a country if at least one author is affiliated with the country. I restricted the investigation to the 10 nations with the highest scientific impact according to the share of top 1% cited papers.[6] Table 5 lists the statistics I computed. Once again, the percentage of conference papers (on average 76%) dominates that of journal papers (on average 24%), but journal articles harvest more citations (57%) than conference articles (43%): the average impact of journal articles (7.73) is more than four times higher than the impact of conference articles (1.77). Moreover, the average journal h index is 27 and dominates the average conference h index, which is 16. Conference papers are mostly cited by conference papers (67%), while journals are cited by both publication sources, but more by journals (56%). Notice that Japan represents an exception: conference papers collect more citations than journal papers and have the same h index. Moreover, journal papers are cited more by conference papers.

The nation with the highest productivity (number of papers) and also that with the highest scientific impact (number of citations or h index) is, not surprisingly, the U.S. The top-three in number of journal papers are the U.S., England, and Canada; the top-three in number of conference papers is the U.S., Germany, and Japan. The U.S., England, and Germany are the countries that receive most of the journal citations and that have the highest journal h number, while the U.S., Germany, and France are the nations with the highest number of conference citations and with the largest conference h index. Interestingly, Switzerland is the nation with the highest journal impact, followed by The Netherlands and England (the U.S. in only 7th here), while the top-three with respect to conference impact is Switzerland, The Netherlands, and France (the U.S. is 4th).

Although it is more difficult to get published in journals, the effort is ultimately rewarded with a higher impact.

Table 5. Nations with the highest scientific impact.

Topic	Pubs	Cites	Impact	H	J-Citer	C-Citer
U.S. (C)	9959	19244	1.93	41	5398 (33%)	10910 (67%)
U.S. (J)	2760	19446	7.05	43	9239 (54%)	7757 (46%)
England (C)	2370	4019	1.70	19	1368 (37%)	2305 (63%)
England (J)	1021	8808	8.63	36	4793 (62%)	2937 (38%)
Germany (C)	2956	5617	1.90	23	1687 (33%)	3361 (67%)
Germany (J)	838	7069	8.44	32	4281 (69%)	1932 (31%)
Japan (C)	2642	2117	0.80	15	650 (33%)	1309 (67%)
Japan (J)	507	1435	2.83	15	633 (46%)	749 (54%)
France (C)	2189	4542	2.07	24	1266 (31%)	2781 (69%)
France (J)	672	5505	8.19	31	2785 (56%)	2169 (44%)
Canada (C)	1876	2562	1.37	17	870 (37%)	1508 (63%)
Canada (J)	907	6246	6.89	29	3226 (57%)	2477 (43%)
Italy (C)	1860	2966	1.59	18	930 (35%)	1728 (65%)
Italy (J)	718	4398	6.13	25	2147 (53%)	1895 (47%)
Switzerland (C)	635	1974	3.11	18	551 (30%)	1276 (70%)
Switzerland (J)	230	2830	12.30	26	1470 (57%)	1103 (43%)
The Netherlands (C)	895	2012	2.25	17	672 (36%)	1170 (64%)
The Netherlands (J)	280	2644	9.44	24	1449 (61%)	912 (39%)
Australia (C)	1346	1333	0.99	14	386 (31%)	841 (69%)
Australia (J)	399	2974	7.45	25	1517 (56%)	1187 (44%)
Mean (C)	**2673**	**4639**	**1.77**	**21**	**1378 (34%)**	**2719 (66%)**
Mean (J)	**833**	**6135**	**7.73**	**29**	**3154 (58%)**	**2311 (42%)**

Pubs: number of publications having at least one author affiliated in the nation; **Cites**: total number of citations received by the publications; **Impact**: average number of citations per publication; **H**: value of the h index on the publication set; **J-Citer**: number of citing publications that are journal papers; **C-Citer**: number of citing publications that are conference papers. Each information is shown for both conference publications (C) and journal publications (J). Source: King[6] for the list of nations; Web of Science for publication and citation data. Target period: 2005. Census date: August 1, 2009. Note: the data for U.S. conference papers is an approximation since the size of the query result is beyond the maximum limit of Web of Science.

Conclusion

The main conclusions are:

- computer scientists publish more in conference proceedings than in archival journals;
- the impact of journal publications is significantly higher than the impact of conference papers.

The fundamental message for the computer science community is: although it is more difficult to get published in journals, the effort is ultimately rewarded with a higher impact. From a bibliometric perspective, the best strategy to gain impact seems to be that of publishing few, final, and well-polished contributions in archival journals instead of many premature "publishing quarks" in conference proceedings. ■

References

1. Birman, K. and Schneider, F.B. Program committee overload in systems. *Commun. ACM 52*, 5 (May 2009), 34–37.
2. Choppy, C. et al. Research evaluation for computer science. *Commun. ACM 52*, 4 (Apr. 2009), 31–34.
3. Fortnow, L. Time for computer science to grow up. *Commun. ACM 52*, 8 (Aug. 2009), 33–35.
4. Franceschet, M. The difference between popularity and prestige in the sciences and in the social sciences: A bibliometric analysis. *Journal of Informetrics 4*, 1 (2010), 55–63. DOI: 10.1016/j.joi.2009.08.001.
5. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of USA 102*, 46 (2005), 16569–16572.
6. King, D.A. The scientific impact of nations. *Nature*, 430 (2004), 311–316.
7. McKeown, N., Crowcroft, J., and Keshav, S. Scaling the academic publication process to Internet scale. *Commun. ACM 52*, 1 (Jan. 2009), 27–30.
8. Reed, R. Publishing quarks: Considering our culture. *Computing Research News 21*, 2 (Feb. 2009).
9. Vardi, M. Conferences vs. journals in computing research. *Commun. ACM 52*, 5 (May 2009), 5.

Massimo Franceschet (massimo.franceschet@dimi.uniud.it) is a researcher in the Department of Mathematics and Computer Science at the University of Udine, Italy.