Computer Science Journals—An Iterated Citation Analysis

GRAEME HIRST AND NADIA TALENT

Abstract—Citation data from the 1975 and 1976 *Journal Citation Reports* were used to develop a "computer science impact factor" for ranking "core" journals in computer science. The starting set of source journals included *Communications of the ACM*, *Computer*, *IEEE Transactions on Computers*, and *Journal of the Association for Computing Machinery*, to which *Computer Journal* was added in iteration. The problems and limitations of citation analysis are discussed and results are compared with a previous analysis [K. Subramanyam, *IEEE Trans. Prof. Commun.*, vol. PC-19, no. 2, pp. 22–25, Dec. 1976]. Librarians' use of the results should be made in the light of their own experience of user requirements.

Choice of Journals

If one is to conduct a study of important journals in a particular field, one is immediately faced with this "chicken-and-egg" problem: To find citations to identify important journals, it is necessary to look in those important journals.

Manuscript received August 10, 1977; revised September 16, 1977.

The authors are with the Department of Engineering Physics, Research School of Physical Sciences, Australian National University, P.O. Box 4, Canberra, A.C.T., Australia 2600; phone (062) 49 3064.

One cannot identify core computer science journals by analyzing *The Journal of Experimental Psychology* any more than the *IEEE Transactions on Computers* can be used to identify core psychology journals.

It is necessary, therefore, to take an iterative approach, using as a starting point a journal or set of journals which one reasonably believes to be important. This requires either one's own expert knowledge of the field or the opinions of others known to be expert. The results of a citation analysis on this first set must be carefully examined in order to detect any bias in the data.

Two types of bias may be present. First, most journals tend to cite themselves fairly frequently. The *Journal Citation Reports*® (*JCR*™) [2, 3] shows that most journals cite themselves more than they do almost any other journal—in general about 20 percent of a journal's references are to itself [2, p. 7; 3, p. 7; 4; 5] and that most journals are cited by themselves more than they are by almost any other. It is not surprising, therefore, if the journals used in a core-list study end up close to the top of the list. If analysis ceased at this point, one could demonstrate, using almost any journal *A*, that *A* is a core journal in its field.

Second, if the journals chosen represent not the core of the field itself, but rather a subfield (such as numerical methods or programming languages, in the case of computer science), then other journals in the same subfield will appear to be more important than those outside it, for they will be cited more frequently.

For these reasons then, one must regard the results of the first analysis as an approximation only—as an *initial* iteration. To obtain a better approximation, more iterations must be performed. Journals which appear to be important in the results of the initial iteration, but which were *not* used to generate it, are taken and citation data derived from them. The data from this iteration are then merged with those from the initial one. It may be decided that further iterations are necessary, or that the data from one or more of the journals are so biased that they should be eliminated.

This iterative process should soon converge; after a few iterations, a new one should not substantially change the overall results. However, even if convergence is achieved, unbiased results cannot be guaranteed. To see this, we need only consider the pathological case of a fringe journal that cites only itself. If it is chosen as a starting point, the process will converge in one iteration with the erroneous result that this journal is the only one that matters. Similarly, it is possible for the process to converge on what is essentially a mutually citing set of journals which represents the core of a subfield but not the field as a whole.

It is therefore necessary for the whole task to be carried out by, or in collaboration with, someone (personally unbiased, of course) with enough familiarity with the literature of the field to detect whether or not such bias is creeping into the data.

One problem lies in the choice of *any* journal as the *sole* primary journal. In reference [1] the choice was the *IEEE Transactions on Computers* (*IEEE TC*). That journal tends to emphasize computer hardware, number theory applications in hardware, and computer applications, at the expense of other areas of computer science—which is hardly surprising or inappropriate for a journal published by the Institute of Electrical and Electronics Engineers. Being the only primary journal in the analysis [1], *IEEE TC*'s self-citing rate was high enough to give it 3.9 times as many citations as the second most-cited journal, *Communications of the ACM* (*CACM*).

If we were to perform an analogous citation analysis, i.e., one based only on absolute citation counts, we would choose *CACM* as our starting primary journal. It publishes high quality research papers which range over a wider area of computer science. However, it is not fully representative of the field either; it tends to be biased toward software and theory. To try to counter-balance this, we would also use *IEEE TC*, but not assign to it a dominant position.

A second concern is that the source journals be compatible and adequately representative of the "core" of computer science. For example, if there is no overlap among the sets of journals cited by the source journals, then it would seem that one or more of the latter group is not pointed toward the core.

We do not disagree with the use [1] of *Computing Reviews* (*CR*) as a representative secondary journal, although we would expect it to be mildly biased in favor of obscure fringe journals rather than core publications. (One of its functions is to draw attention to items that are worthy of note but which might otherwise be overlooked by most people.) However, we would also consider the *Journal of the Association for Computing Machinery* (*JACM*), a high quality, wide ranging, theoretical journal; *Computing Surveys*, which publishes review articles concentrating on software and theory; and *Computer*, a review-type publication with a bias toward hardware.

Also to be considered in developing a core list of journals is currency. Computer science is a rapidly evolving field and some of its journals are too young to have a substantial history of citation (e.g., *Artificial Intelligence* and *Acta Informatica*). Moreover, in any field it is to be expected that some journals will decline in importance over time, just as others will ascend. Therefore, a study which draws from recent sources is more timely and useful to librarians.

Citation Analysis Is Not Enough

"No one criterion used in isolation can give a realistic indication of the relative importance of journals. Development of 'core lists' of journals based on one measure, however sophisticated, is of little use" [6, p. 370]. We agree completely with this statement, made in 1975 by Subramanyam. The problems of using the single measure of citation analysis are well known [4, 7, 8].

We briefly mention three inherent limitations of citation analysis. The first is well illustrated by the observation that the *Proceedings of the Cambridge Philosophical Society* received 11 citations in Subramanyam's data. We located ten of those 11 citations, and all were to papers published in the period 1947 to 1958. Whereas this journal was once in the core of computer science, before the advent of many specialized journals in the field, it has long since been displaced.

The second general limitation with citation analysis is that there are many important journals that are not frequently cited. These include such journals as *New Scientist* and *Scientific American* [7] and others read for news or current awareness [6] but rarely cited. Below we identify some of these in computer science.

The third problem is that citation analysis does not take into account the different sizes of journals: A journal which publishes more papers will, other things being equal, be cited more often [6, 7]. The use of a journal's *impact factor* [2, 3, 6, 7, 9], which is essentially the average number of times each paper in it is cited, eliminates journal size as a confounding factor because it is a measure of the journal's citedness-to-size ratio.

Subramanyam [6] provides a good review of various methods for determining core lists of journals and concludes that none is fully adequate. The methods fall into two classes: (1) those based on observation of actual use of, and requests for, the journals in a library; and (2) those, such as citation analysis, based on data derived from the journals themselves. The importance of the first class of methods should not be downplayed, even though it serves mainly as a rejector mechanism for unnecessary journals already in the library, rather than as an identifier of needed ones that are not there. What constitutes a core list depends ultimately not on the discipline but on the library—i.e., on the particular teaching or research interests of its users. Therefore, librarians should temper any results derived from methods of the second class with their own experience and the needs and requests of their patrons. The data we present below must be construed not as some absolute core list which defines the heart of computer science literature, but merely as a starting point from which each library can develop its own core list.

Impact Factor, 1974–1975

As noted above, absolute citation counts do not take into account different journal sizes, thus favoring larger journals unfairly. We decided, in constructing an approximation to a core list based on recent data, to compute a *computer science impact factor* (CSIF) [10]. This is like the *JCR*'s impact factor [2, 3, 7, 9] except that it measures the relative number of times that a source article is cited in the computer science literature rather than in the whole scientific literature.

A CSIF is computed thus: A starting set of computer science journals is taken and a citation count made. For each cited journal, the number of citations it received is divided by the number of citable items it published over a certain period. The result is the CSIF for that period. Since the data were readily available in the *JCR*, we looked at citations made in 1974–1975 (the most recent years available) and used as the divisor the total number of citable items published in 1972–1975. The choice of this divisor was somewhat arbitrary, and determined by the available data, but since the CSIF is relative, this shouldn't matter as long as the divisor is held constant. We expect the method to be mildly biased in favor of longer established journals because they have more citable items before 1972. This bias could have been eliminated by looking only at the citations to the years 1972–1975, but only at the expense of adding another bias, namely, to those journals with a higher *immediacy index* [2, 3, 7], which are generally those journals with a higher publication frequency.

This method depends, like absolute citation counting, on the starting set of journals and, similarly, it can and should be iterated to compensate for this. It also inherits from citation analysis the problem that journals which are no longer core journals will have deceptively high CSIFs due to the continuing citation of their old papers. This is often apparent from the *JCR* data; otherwise a knowledge of the literature of the discipline can be used to get around this problem.

We took our citation data from the *JCR*, which does not usually report journals that a citing journal cites less than six times in one year. This should be kept in mind when looking at our results. Because of this, some data which should have been included have not been. However, this does not affect *major* trends in the results.

Communications of the ACM (*CACM*)
Computer
IEEE Transactions on Computers (*IEEE TC*)
Journal of the Association for Computing Machinery (*JACM*)

Unfortunately, *Computer* was not covered in the 1976 *JCR*. *Computing Surveys*, which we would have liked to include, was not covered in either year. In some cases where the *JCR* did not give the number of source items for a journal, we were able to obtain this datum by checking the journal itself. Where this wasn't possible, a CSIF couldn't be calculated.

When we looked at the results of using this journal set, we found there were many journals that were cited by *IEEE TC* and not by the others. This could be explained by the fact that it is a larger journal, and therefore it is more likely that some fringe journals—but not the smaller journals—will be cited in *IEEE TC* six or more times. To be sure, we performed another iteration with *Computer Journal* (*Comp J*), which was the highest ranking *general* journal in the results of the previous iteration. This added a few new journals to the bottom of the list, including the *Computer Journal*'s sister publication, *Computer Bulletin*. However, we appeared to have reached a reasonable convergence. Since our initial set of journals was large, it is not surprising this happened so soon. The results can be seen in Table I.

Discussion

In Table I the CSIFs fall rapidly. Only 21 journals had a CSIF greater than 0.05, below which level a journal cannot be said to have much impact. (This point is indicated by a horizontal line in the table.) Journals that fall below this level tend to be journals that are important in computer science *application* fields, rather than in computer science *per se*. Unsurprisingly, these tended to be cited only by *IEEE TC*.

We submit the journals listed above the line in Table I as an approximation to a computer science journal core list, subject to consideration of the additions we make in the following section. This list should be interpreted by users in the light of their own library experience and journals below the line in Table I should be considered by libraries where there is interest in the applications they represent.

Surveys and Opinions

We have seen that there are many reasons why an important journal may not appear to be important in the results of some forms of citation analysis. In this section we identify some of these journals in computer science; we rely on our own experience and the opinions of others.

To obtain these opinions, we compiled an informal survey that was sent to 26 faculty and graduate students in the Department of Computer Science and the academic section of the Computer Centre at the Australian National University. The survey took the form of a tentative core list to which respondents were asked to add (or delete) journals. To test response reliability, some important journals were omitted from the list and some obscure or irrelevant ones were included.

Only 17 responses were received and they were somewhat disappointing in that many people were loathe to delete any journal from the list. However, there were some consistent trends and we took account of these. In Table II we present a supplementary list of journals which we feel should be held, or at least seriously considered, by any library serving computer science. For each journal, we have given a brief statement on why we are including it.

TABLE I
NUMBER OF CITATIONS AND CALCULATED COMPUTER SCIENCE IMPACT FACTOR (CSIF) FOR TECHNICAL JOURNALS AND PUBLICATIONS

Rank	Cited Journal	Citing Journal								
		IEEE TC	*CACM*	*JACM*	*Computer*	*Comp J*	Total	1972–1975 Source Items	CSIF	Rank in Ref. [1]
1	*IEEE Transactions on Computers*	951	7	18	55	0	1031	650	1.59	1
2	*Journal of the Association for Computing Machinery*	85	63	119	8	38	313	216	1.45	3
3	*Communications of the ACM*	78	324	54	27	128	611	526	1.16	2
4	*Machine Intelligence*	–(a)	7	21	–	–	28	25	1.12(b)	60
5	*Computing Surveys*	7	18	–	–	11	36	44	0.82	
6	*IBM Systems Journal*	15	17	10	10	9	61	76	0.80	24
7	*Computer Journal*	22	31	7	–	153	213	302	0.71	6
8	*Information and Control*	36	6	33	–	11	86	281	0.31	9
9	*IBM Journal of Research and Development*	53	–	–	–	11	64	231	0.28	7.5
10	*Artificial Intelligence*	–	6	6	–	–	12	54	0.222	
11	*Pattern Recognition*	20	–	–	–	–	20	96	0.208	33.5
12	*IEEE Transactions on Information Theory*	70	–	–	–	–	70	341	0.205	7.5
13	*Numerische Mathematik*	–	12	19	–	17	48	258	0.186	4
14	*SIAM Journal on Computing*	–	–	18	–	–	18	122	0.148	
15	*Bell System Technical Journal*	30	8	–	9	–	47	401	0.117	17
16	*Operations Research*	6	15	31	–	–	52	447	0.116	
17	*Journal of Computer and System Sciences*	9	–	9	–	–	18	158	0.114	19
18	*BIT* (c)	–	11	6	–	–	17	216	0.079	13
19	*Proceedings of the IEEE*	53	–	–	–	–	53	716	0.074	10
20	*IEEE Transactions on Acoustics, Speech, and Signal Processing*	23	–	–	–	–	23	314	0.073	25
21	*Philips Research Reports*	13	–	–	–	–	13	182	0.071	
22	*IEEE Spectrum*	19	–	–	–	–	19	386	0.049	54
23	*Management Science*	–	–	27	–	–	27	563	0.048	60
24	*SIAM Journal on Numerical Analysis*	–	7	6	–	–	13	309	0.042	5
25	*Psychometrika*	6	–	–	–	–	6	148	0.041	
26	*IEEE Transactions on Systems, Man, and Cybernetics*	9	–	–	–	–	9	262	0.034	36.5
27	*Journal of Symbolic Logic*	–	–	7	–	–	7	228	0.031	
27	*Mathematics of Computation*	–	7	–	–	6	13	414	0.031	49.5
28	*American Journal of Mathematics*	6	–	–	–	–	6	208	0.029	65
29	*Computer* (h)	–	–	–	6	–	6	217	0.028	54
30	*Journal of the Franklin Institute*	7	–	–	–	–	7	296	0.024	22.5
31	*Journal of the Optical Society of America*	15	–	–	6	–	21	905	0.023	
32	*American Mathematical Monthly*	6	–	6	–	–	12	585	0.021	33.5
33	*Datamation*	–	7	–	–	–	7	348	0.020	11.5
34	*Electronics Letters*	30	–	–	–	–	30	1712	0.018	38

TABLE I (Continued)

Rank	Cited Journal	Citing Journal: *IEEE TC*	*CACM*	*JACM*	*Computer*	*Comp J*	Total	1972–1975 Source Items	CSIF	Rank in Ref. [1]
34	*Applied Optics*	31	–	–	–	–	31	1730	0.018	
35	*IEEE Transactions on Circuits and Systems*	–	–	7	–	–	7	453	0.015	20.5
35	*IEEE Transactions on Communications*	13	–	–	–	–	13	848	0.015	39
36	*SIAM Journal on Applied Mathematics*	6	–	–	–	–	6	506	0.012	65
37	*Annals of Mathematics*	6	–	–	–	–	6	617	0.010	
38	*Electronics*	–	–	–	12	–	12	1305	0.009	65
38	*Bulletin de l'Academie Polonaise des Sciences. Series des Sciences, Mathematiques, Astronomiques et Physiques*	–	–	–	7	–	7	807	0.009	
38	*Journal of Theoretical Biology*	7	–	–	–	–	7	762	0.009	
39	*Proceedings of the Institution of Electrical Engineers [London]*	8	–	–	–	–	8	960	0.008	60
40	*Transactions of the American Mathematical Society*	9	–	–	–	–	9	1432	0.006	
41	*Optics Communications*	6	–	–	–	–	6	1235	0.005	
	Computer Bulletin (d)	–	–	–	–	33	33	(e)		15
	Sigmicro Newsletter	18	–	–	–	–	18	(e)		
	Programming Languages	–	10	–	–	6	16	(e)		
	EE Systems Engineering (f)	–	–	–	9	–	9	(e)		
	Electronic Design	–	–	–	9	–	9	(e)		
	Information Processing (g)	–	8	–	–	–	8	(e)		
	Computer Design	–	–	–	7	–	7	(e)		
	Acta Informatica	–	6	–	–	–	6	(e)		
	Sigplan Notices	–	6	–	–	–	6	(e)		
Subtotal		1673	576	404	165	423	3241	22892	0.142	
All other references		1775	1014	622	393	818	4622			
Total		3448	1590	1026	558	1245	7867			

(a) – indicates that the datum did not appear in the *JCR*, and that therefore the number of citations is less than six.

(b) This figure is deceptively high because *Machine Intelligence* did not appear in 1973–1975. If it had been published in those years with the same number of source items each year as in 1972, namely 25, it would still have a relatively high CSIF of 0.280, even if these hypothetical issues were never cited. This publication sits on the intersection of journals, conference proceedings, and monograph series.

(c) Also known as *Nordisk Tidskrift for Informations-Behandling.*

(d) This journal changed in September 1974 from monthly to quarterly publication, with an accompanying change in format; these data refer to the old series.

(e) The number of source items was not available for this journal; hence it was not possible to compute a CSIF for it.

(f) We have not been able to identify this journal.

(g) This probably refers to *Information Processing and Management.*

(h) Previously *IEEE Computer Group News.*

TABLE II

SUGGESTED ADDITIONS TO THE CORE LIST

Acta Informatica A young, quality, theoretical journal. "A must," commented one survey respondent.

ACM Transactions on Database Systems A new journal published by ACM, likely to be important in its field.

ACM Transactions on Mathematical Software A new journal published by ACM, likely to be important in its field.

Computer A journal of review and tutorial papers, mostly in hardware and applications.

Computing Reviews Publishes abstracts and short reviews of current computer science literature.

Datamation Provides current awareness on commercial applications and products. "To remind academics of how bad the real world is," said one survey respondent.

Electronics Provides current awareness on computer hardware developments.

IEEE Transactions on Software Engineering A new journal, frequently mentioned by survey respondents.

Information Processing Letters Frequently mentioned by survey respondents.

Journal of the Institute of Mathematics and its Application Many respondents mentioned this journal, although that may reflect the bias of local interests.

Journal of Optimization and Application Many respondents mentioned this, although that may reflect the bias of local interests.

Management Science This had a relatively high CSIF, and was allowed by most survey respondents.

Mathematical Programming Many respondents mentioned this, although that may reflect the bias of local interests.

Mathematics of Computation This had a relatively high CSIF, and was allowed by most survey respondents.

Software–Practice and Experience Mentioned by several survey respondents.

[ACM SIG and SIC publications] These were mentioned by several survey respondents. Most provide good current awareness for news and events in their particular computer science subfield.

[Local computer society journal] In many places the local chapter of the national computer society publishes a newsletter, which would be useful for awareness of current events, meetings, etc.

[Local trade journal] In most countries, local newspapers or journals are published for the computer industry. These are useful for current awareness of local events.

SUMMARY

We have shown the problems inherent in core list construction and how some may be solved, and have provided a new approximation to a core list for computer science. We stress the word "approximation" throughout this paper; the most important point we make is that it is the librarians who must continue to make the final judgments for their own libraries. Computational methods can only produce data to assist librarians in decision making; they cannot make the decisions for them.

REFERENCES

[1] K. Subramanyam, "Core Journals in Computer Science," *IEEE Trans. Prof. Commun.*, vol. PC-19, no. 2, pp. 22–25, Dec. 1976.

[2] Eugene Garfield (compiler and editor), *Journal Citation Reports*, vol. 9 of *Science Citation Index 1975*, Institute for Scientific Information, Philadelphia, PA, 1976.

[3] Eugene Garfield (compiler and editor), *Journal Citation Reports*, vol. 9 of *Science Citation Index 1976*, Institute for Scientific Information, Philadelphia, PA, 1977.

[4] Eugene Garfield, "Is Citation Frequency a Valid Criterion for Selecting Journals?" *Current Contents*, no. 14, pp. 5–6, Apr. 5, 1972.

[5] Eugene Garfield, "Journal Citation Studies. 18. Journal [Self-] Citation Rates–There's a Difference," *Current Contents*, no. 52, pp. 5–7, Dec. 25, 1974.

[6] K. Subramanyam, "Criteria for Journal Selection," *Special Libraries*, vol. 66, pp. 367–371, 1975.

[7] Eugene Garfield, "Citation Analysis as a Tool in Journal Evaluation," *Science*, vol. 178, pp. 471–479, Nov. 3, 1972; reprinted in [2], pp. 21–28, and in [3], pp. 27–34.

[8] Eugene Garfield, "Citation Analysis and the Anti-Vivisection Controversy," *Current Contents*, no. 17, pp. 5–10, Apr. 25, 1977.

[9] Eugene Garfield, "What Scientific Journals Can Tell Us About Scientific Journals," *IEEE Trans. Prof. Commun.*, vol. PC-16, no. 4, pp. 200–203, Dec. 1973.

[10] Graeme Hirst, "Discipline Impact Factors: A New Method for Developing Core Journal Lists," in preparation.