Patterns of bibliographic references in the ACM published papers

Jacques Wainer[a,*], Henrique Przibisczki de Oliveira[a], Ricardo Anido[a]

[a] *Computing Institute*
University of Campinas
Av. Albert Einstein 1251
13083-852 Campinas - SP - Brazil

Abstract

This paper analyzes the bibliographic references made by all papers published by ACM in 2006. Both an automatic classification of all references and a human classification of a random sample of them resulted that around 40% of the references are to conference proceedings papers, around 30% are to journal papers, and around 8% are to books. Among the other types of documents, standards and RFC correspond to 3% of the references, technical and other reports correspond to 4%, and other Web references to 3%. Among the documents cited at least 10 times by the 2006 ACM papers, 41% are conferences papers, 37 % are books, and 16% are journal papers.

Keywords: Bibliometrics; citations; computer science; conference papers

1. Citation analysis in Computer Science

A scientific document, which we will call a *source*, makes *references* to other scientific documents (which we will call *destinations*) in order to acknowledge the relation of its content with the wider scientific field. In Computer Science (CS) tradition, references are listed at the end of the paper, in a references or bibliography section. If a source document A makes a reference to a destination document B, we will say that B was *cited* (by A). Thus, in this paper we will use *cited* to refer to destinations. Scientific documents may be published in different venues - such as journals, conferences proceedings, technical reports, books and book chapters, and so on, which we call *document types*.

Citation analysis is the study of the references mainly from the destination's point of view. The intuition is that important results are cited many times, and thus the count of how many times a paper, an author, or a journal have been cited is a proxy for its "importance." Citation analysis has been used

*Corresponding author. Phone: 55(19)3521-5871; fax 55(19)3521-5847
Email addresses: `wainer@ic.unicamp.br` (Jacques Wainer), `henrique.prz@gmail.com` (Henrique Przibisczki de Oliveira), `ranido@ic.unicamp.br` (Ricardo Anido)

more commonly as a tool to evaluate journals (Garfield, 1972), universities, departments, and researchers (Moed, 2005; Ren and Taylor, 2007). On the other hand, CS research is not well represented by the standard research evaluation metrics; Parnas (2007) and Meyer et al. (2009), among others, have pointed out that CS research has a high respect for conference publications, which are not included in most citation count metrics and thus, places the field in disadvantage when compared with other Sciences.

CS researchers agree that conferences are important; some CS subareas attribute more prestige to having papers accepted to particular conferences than to journals. Many arguments have been put forth for the importance of conferences as a publication venue for CS, including the speed of publications of results and the dynamic character of CS research areas which may not fit well with the standard view of a journal with a well defined, fixed scope of interest.

One of the goals of this research is to evaluate how important are conference papers as references of CS papers. Do CS papers, in general, make references to conference papers? Are references to conference papers the majority of the references in CS papers? Are conference papers among the most cited destinations of CS papers? The answers to these questions provide us with some evidence of the importance of conference papers in Computer Science research.

1.1. ACM Digital Library

The Association for Computing Machinery (ACM) publishes in the Web a collection of scientific publication data called The Guide. The Guide is a collection of *information pages*[1] on articles, papers, and books; each information page lists the title, author, publication date, publisher, and when available, abstract, references (made by the document), and citings (references made by other document in the Guide to this destination). The information pages are publicly available. According to ACM, the Guide includes over 1.2 million information pages for documents published by over 3000 publishers.

The subset of the Guide of journal and conference papers published by the ACM is called the Digital Library (DL) and their information pages contains a link to the full text of the documents, only available to subscribers.

1.2. This paper

This paper studies the reference patterns of CS papers published in 2006 by the ACM. We call this set of publications the *(2006) source set*. The source set includes conference papers, journal papers, and articles published in the ACM magazines and newsletters. We believe that this source set can be seen as a (unfortunately biased) sample of the whole CS literature, and thus, this research attempts to measure some characteristics of the bibliographic references in the CS literature as a whole.

We computed some descriptive statistics such as the distribution of number of references per paper and the distribution of the ages of the cited documents

[1] ACM calls these pages *citations*.

in the source data set. These statistics may be of interest when comparing the CS reference patterns with that of other scientific areas, but we do not pursue these comparisons in this paper.

The second set of results is the analysis of the document types of the documents cited in the source set. We developed a program that classifies the destination of a bibliographic reference as a conference paper, a journal paper, a book (or book chapter), or other document types. The program uses approximate pattern matching associated with a large base of journal and conference names, plus a set of heuristics. The program was applied to all the references in the source set, and we determined the proportions of different types of destinations.

A second analysis was a manual evaluation of a randomly selected subset of the references in the source set. The manual evaluation was used not only to evaluate the accuracy of the automatic classifier, but also to understand the importance of a broader set of document types such as technical reports, white papers, software manuals, specifications, Web pages, patents, and so on, in the CS literature.

Finally this paper also analyzed the most cited documents (of any year and publisher) referred to by the papers in the source set. We selected all destinations that were cited 10 or more times, and compared the proportion of conference papers, journal papers, and books.

The set of most cited destinations may capture two distinct groups. Some of these papers may be fundamental results or fundamental concepts in CS, whereas others may be papers that introduced short term "hot" topics. To evaluate these two groups, we also determined the set of most cited documents for all ACM papers published in 2003. We believe that the highly cited destinations that are common to both 2003 and 2006 represent fundamental papers in CS, whereas destinations in only one the these sets are more likely to represent the "hot" topics papers. The standard understanding of the role of conferences in the CS scientific process, as a venue to quickly publish worthy ideas, would indicate that conference papers should be more frequent in the latter group, and books and journal papers should be more frequent in the fundamental publication groups. We were be able to evaluate this claim.

1.3. Related literature

There has been some publications on the relative importance of conference papers in scientific research, for example Glänzel et al. (2006) and Lisee et al. (2008). Glänzel et al. (2006) justified the study of proceedings publications based on the importance of conferences to Engineering and Applied Sciences, which includes CS. The analysis of the Web of Science (WoS) proceedings database, from 1994 to 2002, showed that almost half of the total number of publications indexed by WoS in the area of Engineering are proceedings.

Lisee et al. (2008) collected all references from papers in the WoS indexed journals, from 1980 to 2005, and counted how many of those references were to papers published in proceedings. Overall, only 1.7% of the Natural Sciences and

Engineering references were to proceedings, and 2.5% for the Social Sciences and Humanities, and these proportions were decreasing in time. But within Natural Sciences and Engineering, Computer Science papers had on average 19.8% of their references to proceedings. CS is the subfield with the highest of such proportion, followed by Electrical Engineering and Electronics (13.1%). Lisee et al. (2008) also analyzed the age of the destinations, that is, the time interval between the year the source was published and year the document cited in the reference was published. For the whole set of destinations, the median age (or half-life) is 4.0 years for proceedings references, and 6.1 for all references. For Engineering (which includes CS), the average age of all references was 10.9 years, but the average age cited proceedings papers was only 7.1 years. Thus, the age data corroborates the idea that conference papers are more representative of forefront scientific results, which will obsolesce faster.

The most complete reference analysis of the CS literature was reported in Goodrum et al. (2001). They compared references from journals indexed by WoS with sources available in the CiteSeer repository of CS literature. The comparison is interesting because the WoS sources and the CiteSeer sources are in some way at extremes of the spectrum of CS literature: WoS indexes almost exclusively journals, where as CiteSeer, at that time, indexed only freely available documents, mainly technical reports and other documents made public by their authors. Of interest to this research is the comparative analysis of the destination types of the most cited documents available in CiteSeer and the ones indexed by WoS, displayed in Table 1. Unfortunately, from a methodological point of view, the data in Table 1 cannot be generalized to the whole CS literature because it was collected from the most cited sources, and thus cannot be seen as "general cases." But it is the only published data on the proportion of references to books, journals, and conference proceedings to compare to our results.

2. Method

2.1. Data collection

The set of source documents used in the analysis were all information pages in the ACM DL, for papers published in 2006 and in 2003. As mentioned above, most of the analysis was performed in 2006 data set, and the 2003 data was used only for comparisons regarding the most cited papers.

In September of 2007, when we collected the data, the URL to the ACM paper information pages in the portal had a single identity number. A typical URL would be `http://portal.acm.org/citation.cfm?id=301974`. Unfortunately, by the ID only there was no way of deciding if the paper belonged to the ACM DL, and thus, if it was published in one of the ACM sponsored conferences and journals, or whether it belonged to the ACM Guide.

We sequentially generated and requested all pages with id numbers from 858404 to 1248459 (we did not find any paper from 2006 with id lower than 858404 and higher than 1248459). In each page, there is a string "Year of

Publication:" and we used that information as the publication year. We selected only the information pages that contained a link to a full text, which indicated that the document was published by ACM.

2.2. Manual analysis

We randomly selected 665 references from all the documents in the 2006 source set. Two of us with more experience, JW and RA, independently classified the random references into: a) *books* (reference to whole books), b) *book chapter*, c) *conference paper* (including workshops), d) *journal papers*, e) *magazine papers* (non-peer reviewed articles in magazines such as ACM newsletters, ComputerWorld, and so on.), f) *standards* (including RFCs), g) *manuals* (manuals for software or devices), h) *news* (newspapers or news services news), i) *patents*, j) *thesis* (no distinction between master or PhD thesis), k) *reports* (technical reports from universities, white papers from other research organizations and consortia), l) *data sets* (URL address of data sets), m) *software* (URL address of software), n) *course materials*, o) *Web address* (to companies and consortia such as OMG and W3C, but not pointing to a report, standard, manual, or other of the classes before), and p) *others* (all other types of references). When there was conflicting opinions regarding the classification, the case was discussed, and in some cases we retrieved the document, in order to agree on its classification.

2.3. Most cited destinations

As we mentioned, the ACM paper information page (both for the DL and The Guide) has a section on the references of the paper. ACM uses some algorithm of *entity resolution* to discover which destination corresponds to each reference, and if the destination is one of the documents in the Guide, ACM will create a link from the reference to the destination's information page. In this research we relied on the ACM entity resolution algorithm and we assumed that ACM would include in the Guide the information pages for the most common destinations. Thus, we only counted how many times each URL for the destinations appeared in the source set.

2.4. Confidence intervals

There are two applications of confidence intervals in this paper: the confidence intervals on the proportions of documents types in the manual evaluation and in the automatic evaluation.

The manual evaluation was performed on a *sample* of the reference, and thus the confidence intervals for proportions on the different document types are the standard sample intervals for very large populations (the set of all references in the source set). Assuming a 95% confidence level, the formula for the confidence intervals is:

$$CI_{95\%} = p \pm 1.96\sqrt{\frac{p(1-p)}{n}} \quad (1)$$

where p is the proportion of a particular document type in the sample, and n the sample size, in this case 665.

The proportions obtained using the automatic procedure refer to the whole population of 2006 references and thus there is no sampling error to be accounted for by the confidence interval. But there are errors due to the automatic procedure itself. The manual classification is able to evaluate the accuracy of the automatic classification and one can use the error rate to calculate the confidence interval of the proportions.

Let p_t be the proportion of documents of type t discovered by the automatic procedure, and let e_t be the error rate for the procedure for that type. The probability distribution for the number of errors is a binomial distribution which can be approximated by a normal distribution with mean Ne_t and with standard deviation $\sqrt{Ne_t(1-e_t)}$, where N is the total number of references in the source set. The 95% confidence for the number of errors can be approximated by $e_{max} = Ne_t + 1.96 * \sqrt{Ne_t(1-e_t)}$, that is, the $P(X < e_{max}) = 0.95$. We split these errors equally around the number of documents of type t encountered. Thus the 95% interval of confidence for proportion p_t is

$$p_t \pm \frac{e_{max}}{2N} \tag{2}$$

3. Results

The 2006 source set contained 10,704 documents from ACM, mainly conference and journal papers, but also editorials, book reviews, magazine articles, and so on. The 2003 source set contained 6,557 documents. There were 200,018 references in the 2006 source set.

Figure 1 is the histogram of the number of references per document for the documents with less than 100 references. There are 308 documents with more than 100 references. The average number of references per document is 21.26 (median 18). This can be compared with some data on references per articles for other scientific areas. Biglu (2007) reports that the average number of references per document for all papers indexed by Thomson Reuters Science Citation Index (SCI) for 2005 was 34.63. Vieira and Gomes (2010) also uses SCI data for 2004, and reports on average 18.25 and 24.11 reference per article for Mathematics and Physics respectively. The relatively low figures of references per article for Computer Science may have some consequences on impact factors of CS publications venues (Biglu, 2007).

We performed a separate analysis of the distribution of the ages of the references for journal and conference source documents. The results are in Table 2. The first three columns show the number of references to journals, conferences, and others document types, for journal and conference source documents (the two lines). The last two columns show the mean age of the references. The data show that CS researchers have different reference practices for journal and conference papers. The proportion of references in journal papers are 32% to journals and 38% to conferences; whereas in a conference paper, only 28% of the references are to journals and 43% to conferences. The differences are significant (X-squared=392.4977, df = 1, p-value ¡ 2.2e-16).

The mean age of the references are also different for journal and conference publications. References made by journal papers are older than the references made by conference papers, and the difference is significant (t = 7.9629, df = 31037.99, p-value = 1.738e-15 for references to journals and t = 23.9982, df = 32334.66, p-value ¡ 2.2e-16 for references to conference papers). Thus, it seems that for a journal paper, a CS researcher will look for more established or "older" research, and will prefer more references to journal papers.

3.1. Document types

The main results regarding both the manual and automatic classifications of the references are displayed in Table 3.

The first column identifies the document type, ordered by decreasing of frequency. The next three columns show the results for the manual analysis, the following three to the automatic analysis: the first figure is the number of document types found, the second the proportion to the total, and the third, the 95% confidence interval as discussed in section 2.4.

Notice that the classification of books for the automatic analysis includes references to book chapters, which is treated separately in the manual analysis. The same is true for conference papers, which in the automatic analysis includes magazine papers. Finally, the class of others for the automatic classification corresponds to any reference type that is not book (including book chapters), conference, and journal papers, whereas Others for the manual analysis refer to any reference type that is not among one of the 15 types discussed in section 2.2.

Regarding the manual classification, the document types of papers in magazines, manual, news, thesis, data sets, patents, and course materials, each correspond to 1% or less of the references. These document types were not included in the table, and thus the sum of references in Table 3 does not add to the 665 references that were analyzed.

To build the confusion matrix for the automatic classifier we grouped book and book chapters into books, and journal and magazine papers into journal, and all other classifications as "Other". Table 4 displays the confusion matrix.

The error rate for the automatic detection of conference papers is the number of incorrectly classified references (as a conference or not conference) divided by the total number of references, that is, (1+5+3+7+18)/665 or 0.0511. Similarly, the error rate for book classification is 0.0466, for journal classification is 0.0541, and for others, 0.1083.

3.2. Most cited documents

We manually classified all 580 destinations (in the Guide) that received 10 or more references in the 2006 source set. Table 5 displays the proportion of books (or book chapters), journal, and conference papers of the 10 top most cited documents (actually 11, since both the 10th and the 11th most cited had the same number of references), the top 20 (actually 21), top 50, top 100 (actually

103), top 200 (actually 213), and top 500 (actually 580). Among the 580 most cited destinations there is one technical report and 4 RFC.

But, as we discussed in the introduction, Table 5 may include two different components: papers that are in some way "fundamental" to CS, and papers that present "hot topics" in 2006. In an attempt to separate these two components, we computed among these 580 references, which ones were also present among the most cited documents in 2003, and the ones that were not. Only 145 of the 580 most cited destinations were also among the 2003 most cited references. Table 6 presents the proportions of document types of the most cited destinations common for 2006 and 2003. They likely represent the papers that presented a "fundamental" result in CS. Table 7 presents the proportions for the documents that were only present in 2006 as a most cited. We believe that they are an approximation of papers that presented topics that became important in 2006, and thus we call them "hot topic" papers. The first two columns are blank because there was no hot topic papers among the top 10 and top 20 most cited references.

4. Discussion and Conclusions

As expected, conferences are important to Computer Science, and that is represented in the number of references to conferences in the 2006 source set. They represent 39% of all bibliographic references in 2006, but somewhat more surprising, they represent 50% of the 500 most cited destinations. And if we consider only the subset of "fundamental results" within the 500 most cited (as described in section 2.3), conferences still amount to almost 40% of such documents. That is, conferences are venues where final, finished results are also published. The importance of conferences was severely underestimated by the analysis using the WoS data and even the data using the CiteSeer documents performed by Goodrum et al. (2001). But, as we discussed, the numbers obtained by Goodrum et al. (2001) cannot be generalized to the whole CS literature, even for the years the data refer to, since they represent a non-random sample.

Journals are also important as CS references: 27% of all references are to journals, and they correspond to 25% of the 2006 most cited destinations.

A somewhat untold story is the importance of books in the CS literature. Although the numbers were available in Goodrum et al. (2001) analysis, that paper did not elaborated this fact. Books amount to 6% of all references in 2006, but they have larger proportion among the most cited destinations types. A possible explanation to the importance of books, especially textbooks is a process of transferring important ideas in CS to the public domain and thus loosing the original authorship of the idea. If one wants to provide a reference to the result that the vertex cover problem is NP-complete or a reference to the Floyd-Warshall algorithm, one adds a bibliographic reference to a complexity textbook (the second most cited destination in the 2006 source set) or an algorithms textbook (the third most cited destination) instead of the respective original papers.

The 2006 most cited destination, the book *Design Patterns: Elements of Reusable Object-Oriented Software* by Gamma, Helm, Johnson, and Vlissides seems to point to a different pattern. A complex concept was "first" widely presented as a whole book, instead of as conference or journal papers[2], which follows the style of publications in the social sciences.

The importance of books and in particular textbooks as bibliographic references in the CS literature has parallels in other scientific areas. Some researchers have pointed out that review papers (papers that summarize different empirical research) are more cited than others, and, in particular, they are more cited than the primary research they summarize (Peters and van Raan, 1994; Aksnes, 2006). Clearly, textbooks are the "review papers" of Computer Science.

Another interesting conclusion of this research is that computer scientists behave somewhat different when writing conference or journal papers in terms of reference choices. Journal papers make more reference to other journal papers and such references point deeper into the past, than conference papers. Some differences in the proportion of journal and conference papers for the most cited WoS and CiteSeer documents is also reported in Goodrum et al. (2001) (see Table 1), but the magnitude of changes we encountered are higher.

Regarding the age of references, we can only compare our data with the one presented in Glanzel and Schoepflin (1999), who collected the mean age of references for papers published in 1994 indexed in SCI. According to Glanzel and Schoepflin (1999), Mathematics and Electronic Engineering had mean reference age of 11.3 and 8.6 years respectively. That should be compared with 7.8 which is the mean age for references in journal papers. Unfortunately, it is yet unclear if the younger mean reference age is a intrinsic characteristic of Computer Science, or if the difference is due to the 12 years difference between the data collected.

The analysis of the most cited documents, specially the comparison to 2003, seems to indicate that CS research's waves of interests last for a short time. Only 25% of the documents that were cited 10 times or more in 2006 also appeared among the documents cited 10 times or more in 2003. This fast lived waves of interests must be further explored, since they are not fully reflected in the mean life of references: on average papers refer to documents from 5 to 6 year old, but in three years 3/4 of the most cited papers changed.

Books are the most frequent document types among the documents that remain important between 2003 and 2006, followed, surprisingly, by conference papers. Among the representatives of the new wave of interest, conferences, as expected, play the most important role.

The manual analysis of the random sample of the references also reveals another interesting phenomena. Beyond conference and journal papers, and books, the next most cited document types are: reports, other Web addresses, standards, book chapters, software, and manuals.

We were somewhat surprised with the low number of references to PhD

[2]The Wikipedia text on design patters points out that there were previous conference papers on patterns by other authors than the book authors.

thesis. This maybe due to the fact that PhD candidates are encouraged/required to publish partial results of their work and thus, references to thesis may have migrated to the corresponding conference and journal papers. We were also surprised with the relatively high number of references to newspaper or news service news - this seems to indicate that some CS research is either inspired by or try to argue its relevance based on current events.

Aksnes, D. W., 2006. Citation rates and perceptions of scientific contribution. Journal of the American Society for Information Science and Technology 57, 169–185.

Biglu, M. H., 2007. The influence of references per paper in the sci to impact factors and the matthew effect. Scientometrics 74 (3), 453–470.

Garfield, E., 1972. Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science 178 (4060), 471 – 479, dOI: 10.1126/science.178.4060.471.

Glänzel, W., Schlemmer, B., Schubert, A., Thijs, B., 2006. Proceedings literature as additional data source for bibliometric analysis. Scientometrics 68 (3), 457–473.

Glanzel, W., Schoepflin, U., 1999. A bibliometric study of reference literature in the sciences and social sciences. Information Porcessing and Management 35 (1), 31–44.

Goodrum, A., McCain, K., Lawrence, S., Lee Giles, C., 2001. Scholarly publishing in the Internet age: a citation analysis of computer science literature. Information Processing and Management 37 (5), 661–675.

Lisee, C., Lariviere, V., Archambault, E., 2008. Conference proceedings as a source of scientific information: A bibliometric analysis. Journal of the American Society for Information Science and Technology 59 (11).

Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J., 2009. Viewpoint research evaluation for computer science. Commun. ACM 52 (4), 31–34.

Moed, H. F., 2005. Citation analysis in research evaluation. Springer.

Parnas, D., 2007. Stop the numbers game. Commun. ACM 50 (11), 19–21.

Peters, H. P. F., van Raan, A. F. J., 1994. On determinants of citation scores: A case study in chemical engineering. Journal of the American Society for Information Science 45 (1), 39 – 49.

Ren, J., Taylor, R., 2007. Automatic and versatile publications ranking for research institutions and scholars. Commun. ACM 50 (6), 81–85.

Vieira, E., Gomes, J., 2010. Citations to scientific articles: Its distribution and dependence on the article features. Journal of Informetrics 4, 1–13.

5. Figure headings

- Figure 1: Histogram of the number of reference per paper.

6. Tables

	CiteSeer 488 most cited documents	WoS 515 most cited documents
Journal articles	182 (37%)	205 (39%)
Book/book chapters	207 (42%)	290 (56%)
Conference proceedings	77 (15%)	18 (3%)
Technical reports	8	0
Computer documentation	2	1
Miscellaneous	12	1

Table 1: Comparison of the most cited CiteSeer and WoS documents (from Goodrum et al. (2001))

Source	Destination: Number of references			Destination: Mean age of references		
	Journal	Conference	Others	Journal	Conference	All
Journal	16,818	20,070	15,532	9.8	6.1	7.8
Conference	38,614	58,828	39,178	9.1	5.0	6.6
Total				9.3	5.3	7.0

Table 2: Number of references and mean ages of references for different source types.

Document type	Manual Analysis			Automatic Analysis		
	number	percentage	conf. int.	number	percentage	conf. int.
Conference	271	41%	37% - 44%	79,088	39%	37% - 42%
Journal	198	30%	26% - 33%	54,763	27%	25% - 31%
Book	67	10%	8% - 12%	12,208	6%	4% - 8%
Reports	25	4%	2% - 5%			
Web address	24	4%	2% - 4%			
Standard	18	3%	2% - 4%			
Book chapter	13	2%	1% - 3%			
Software	11	2%	0% - 2%			
Others	10			53,959	27%	21% - 32%
Total	665			200,018		

Table 3: Number of references for each destination type for both the manual and automatic analysis. Automatic analysis does not include the types "Report" and the others left blank. See text for the difference in meaning for "Book", "Journal", and "Other" for both analysis. Destinations types that received 1% or less of the references are not included, therefore the discrepancy between total and the sum of references for the manual analysis.

Automatic classification	Manual classification			
	Conference	Book	Journal	Other
Conference	246	1	5	3
Book	0	38	0	0
Journal	7	2	184	4
Other	18	29	18	110

Table 4: Confusion matrix for the automatic classifier.

	top 10	top 20	top 50	top 100	top 200	top 500
Books	72.7%	57.1%	46.0%	41.7%	34.3%	24.6%
Conferences	18.1%	38.1%	38.0%	40.7%	43.6%	49.1%
Journals	9.1%	4.8%	16.0%	16.5%	21.1%	25.3%
Number of documents	11	21	50	103	213	580

Table 5: Proportion of document types in the most cited destinations in the 2006 source set

	top 10	top 20	top 50	top 100	top 200	top 500
Books	72.7%	57.1%	51.1%	50.7%	41.6%	37.9%
Conferences	18.1%	38.1%	34.1%	31.7%	34.7%	37.9%
Journals	9.1%	4.7%	14.6%	17.5%	23.7%	23.5%
Number of documents	11	21	41	63	101	145

Table 6: Proportion of document types in the most cited destinations which are common to the 2006 and 2003 source sets, which may represent more stable, fundamental results in CS.

	top 10	top 20	top 50	top 100	top 200	top 500
Books			22.2%	28.2%	27.7%	20.2%
Conferences			55.5%	53.8%	51.8%	52.9%
Journals			22.2%	15.4%	18.7%	26.0%
Number of documents	0	0	9	39	112	435

Table 7: Proportion of document types in the most cited destinations which are only present in the 2006 source sets, which may represent more “hot topics” results in CS.

7. Figures