Scientometrics (2017) 110:529–539
DOI 10.1007/s11192-016-2214-9

Evolution in the number of authors of computer science publications

João M. Fernandes[1] · Miguel P. Monteiro[2]

Received: 14 December 2015 / Published online: 17 December 2016

Abstract This article analyses the evolution in the number of authors of scientific publications in computer science (CS). This analysis is based on a framework that structures CS into 17 constituent areas, proposed by Wainer et al. (Commun ACM 56(8):67–63, 2013), so that indicators can be calculated for each one in order to make comparisons. We collected and mined over 200,000 article references from 81 conferences and journals in the considered CS areas, spanning a 60-year period (1954–2014). The main insights of this article are that all CS areas witness an increase in the average number of authors, in every decade, with just one slight exception. We ordered the article references by number of authors, in ascending chronological order and grouped them into decades. For each CS area, we provide a perspective of how many groups (1-author papers, 2-author papers and so on) must be considered to reach certain proportions of the total for that CS area, e.g., the 90th and 95th percentiles. Different CS areas require different number of groups to reach those percentiles. For all 17 CS areas, an analysis of the point in time in which publications with $n + 1$ authors overtake the publications with n authors is presented. Finally, we analyse the average number of authors and their rate of increase.

Keywords Scientific authorship · Number of authors · Scientific publication · Computer science

✉ João M. Fernandes
jmf@di.uminho.pt

Miguel P. Monteiro
mtpm@fct.unl.pt

[1] Dep. Informática/Centro ALGORITMI, Universidade do Minho, Braga, Portugal

[2] Dep. Informática/NOVA-LICS, Universidade Nova de Lisboa, Almada, Portugal

Introduction

The pressure faced by researchers in order to publish papers (the well-known "publish or perish") implies the existence of some practical and ethical problems concerning the assignment of authorship in academic and scientific publications (Abt 1981; Bennett and Taylor 2003; Solomon 2009). In fact, the career and prestige of a researcher is strongly and essentially dependent on the number and quality of (co-)authored publications.

Previous recent work by Cavero et al. (2014) and Fernandes (2014) has shown that the number of authors per scientific articles is increasing, whenever data for CS in general or software engineering in particular are analysed. In the present article, we extend previous studies on authorship trends (e.g., Gu 2002; Greene 2007; Cohoon et al. 2011; Wainer et al. 2013; Cavero et al. 2014; Fernandes 2014; Garousi and Fernandes 2016) to a significant number of areas within CS, on the basis of a sample of almost 190,000 curated references. As explained in detail in the next section, the sample used in this work follows the framework described by Wainer et al. (2013). The main rationale of their work is that productivity of researchers in different CS areas may be different, but that there is no clear evidence about that fact.

The main purpose of this article is to analyse if there are some significant differences among the different areas that constitute CS in terms of authorship. This may be relevant for comparing performance of individuals, research groups, or even departments that operate in distinct CS areas. For instance, if two people apply for a professorship position in a CS department, one with a publication track in Artificial Intelligence and another one in Software Engineering, are they expected to have published similarly or very differently? As another example, we may be interested in understanding if a CS department whose members publish in the Bioinformatics and Database areas is comparable to another department whose members typically publish in the Security and Theory areas. Answering questions like those requires analysis of authorship patterns across distinct CS areas, an effort to which this article aims to contribute.

Method

This study includes articles published in conferences and journals, since both types are prestigious in computer science (CS). Patterson (2004) and Meyer et al. (2009) indicate that in computing it is common to prefer conferences over journals. Freyne et al. (2010) show that CS conference papers have a similar impact to mid-ranking journals. This situation is in contrast to the prevailing academic tradition where the primary means of publishing is in archival journals (Vardi 2009).

We use articles listed in the DBLP website (Ley 2009), since it is specifically devoted to bibliographic information on major CS journals and conference proceedings. Other studies (e.g., Elmacioglu and Lee 2005; Laender et al. 2008; Solomon 2009; Biryukov and Dong 2010; Franceschet 2011; Cavero et al. 2014; Fernandes 2014) have also used DBLP to obtain bibliographic data. Additionally, DBLP can be interfaced by automatic mechanisms, making it adequate to support automatic retrieval processes. The bulk of the data for the study was obtained on March 3, 2015, by downloading from the DBLP database all entries for the considered conferences and journals.

This article requires a scheme to structure CS into its constituent areas. Unfortunately, there is no universal agreement in the scientific and professional communities on how to

Table 1 CS areas: names, abbreviations and corresponding seed venues

Area	Abbr.	Seed venues
Artificial Intelligence	AI	AIJ, JAIR, JAR, AAAI, IJCAI
Bioinformatics	BIO	BMC Bioinf, Bioinformatics, JCB, RECOMB, TCBB
Communications and Networking	COMM	TON, TCOM, Mobicom, Sigcomm, Infocom
Compilers and Programming Languages	C+PL	OOPSLA, POPL, PLDI, TOPLAS, CGO
Computer Architecture	ARCH	ISCA, MICRO, DAC, ASPLOS, TCAD, SC
Computer Graphics	GRAPH	TOG, CGA, TVCG, SIGGRAPH
Database	DB	TODS, VLDB, Sigmod
Distributed Computing	DC	TPDS, JPDC, ICDCS, ICPP
Human–Computer Interaction	HCI	TOCHI, IJMMS, UMUAI, CHI, CSCW
Image Processing and Computer Vision	IPCV	IJCV, TIP, CVPR, ICIP
Machine Learning	ML	JMLR, ML, NECO, NIPS, ICML
Management Information Systems	MIS	ISR, MANSCI, JMIS, EJIS, MISQ
Multimedia	MM	MMS, TMM, IEEEMM, MM, ICMCS
Operational Research and Optimization	OR	Math Prog, SIOPT, C&OR, Disc Appl Math
Security	SEC	TISSEC, JCS, IEEESP, SP, USS, CSS
Software Engineering	SE	TSE, TOSEM, ICSE, TACAS, ESE
Theory	TH	JACM, SICOMP, STOC, FOCS, SODA

accomplish that purpose. For example, ACM and IEEE each divide CS into different areas—ACM, through special interest groups, or SIGs, and IEEE, through technical committees, or TCs. Moreover, some of these divisions reflect historical decisions that may be less relevant today. Microsoft Academic Search, Scopus and others classify different CS areas but none describes how they arrived at their classifications.

This work adopts the set of CS areas proposed by Wainer et al. (2013) as listed in Table 1. Their rationale to structure CS into its areas was to obtain a broad coverage of CS that includes the more traditional areas (e.g., communications and networking, programming languages, databases, computer architecture, distributed computing, and software engineering). Additionally, their proposal also aimed to cover both new areas (e.g., bioinformatics and security) and some areas on the "fringe" of CS that are not always present in university CS departments in different countries (e.g., operations research and management information systems). Wainer et al. do not claim those areas are the only, or most important, areas of CS and neither do we. Nevertheless, this set of CS areas serves as a suitable basis for our study. For a more detailed explanation of the methods adopted by their study—including how each venue was associated with one, and only one, CS area—see Wainer et al. (2013).

In a subsequent step, Wainer and his colleagues defined a set of venues clearly representative of each area, which they called "seed venues". The set was validated on the basis of feedback provided by colleagues in each area and based on information regarding citations per published paper, acquired from sources such as Microsoft Academic Search and Thompson Reuters Journal of Citation Reports. We base our study on the selection of conferences and journals used by Wainer et al., because it is compatible with our aims.[1] Table 1 shows the 81 seed venues in its rightmost column (using their usual abbreviations) for all selected CS areas.

There is one difference between Table 1 and that presented by Wainer et al. that warrants an explanation. We noticed that TSE (IEEE Transactions on Software Engineering) was missing in the CACM-paper by Wainer. Being from the SE area, the authors of the present article know that TSE is considered one the top venues for SE and were surprised by this omission. TSE is included in a list of SE venues presented in a technical report referred to in the article by Wainer et al. In that report, all seed venues appear before any other, non seed venues, in all CS areas. TSE is the very first venue to appear in the list for SE, further suggesting its omission was a mistake. The authors of the present article contacted Wainer on this issue. Though Wainer couldn't be sure on account of the study having been several years before, he admitted the omission might indeed be a mistake. On the strength of these facts, we included the data from TSE on the sample analysed in this article.

We developed a software program through which we downloaded all DBLP references from the venues from Table 1, comprising a total of 202,755 references. Subsequently, the program discarded all references that did not relate to scientific articles, which includes entries with no authors (e.g., lists of reviewers and programme committees), editions of proceedings, editorials, prefaces, acknowledgments, messages from the editors, forewords, special issue introductions, introductions to ''in honour/in memoriam'' issues, tributes, obituaries, errata, corrigenda, book reviews, comments to articles, and their replies. We also discarded entries with less than four pages, which typically relate to editorials, invited papers, posters, tool papers, workshop summaries, and similar short contributions that are not generally considered regular scientific articles.

All the above steps were monitored and subject to several manual tests by the authors to make sure that the automatic processing performed correctly and in accordance to the aims of this study. Despite our best efforts to ensure an exhaustive evaluation of the processed information, it is still possible that some bibliographic entries were wrongly filtered, due to a number of reasons (e.g., incorrect spellings, missing data, and wrong or unexpected formatting). Nevertheless we believe the number of such cases is relatively low and we do not expect it to significantly impact overall results.

[1] The current paper utilizes only the seed venues indicated by Wainer et al. (2013). These seed venues are the most favoured by researchers, who recognize the seed venues as "central" and "important" to contributed articles to their respective areas. The interested reader could argue that this study could have included all the venues from the longer list in the appendix of the technical report referred to in that paper. To check if the resulting numbers would be significantly different, we experimented performing the study using the longer list of venues for two areas—C+PL and DB. For instance, the average numbers of authors for C+PL are 1.727, 2.129, 2.626 and 3.306, respectively for decades 75–84, 85–94, 95–04 and 05–14. The equivalent numbers from the longer list are 1.747, 2.042, 2.367 and 3.021. Similarly, for DB the average numbers of authors are 1.948, 2.305, 3.183 and 3.792. The equivalent numbers from the longer list are 1.905, 2.255, 2.973 and 3.686. Though the numbers are not exactly the same—as it would be expected—the conclusions to be derived from the longer list are broadly the same as with the seed venues.

Similar conclusions apply to the numbers related to overtakes (Table 3). For C+PL, the years in Table 3 are 1973, 1979, and 2002, while if the longer list of venues is used the years would be 1973, 1979, and 2009.

Analysis

For the bulk of the analysis, data for the various CS areas were grouped into decades, starting in 1955. Thus, decade groups comprise intervals 1955–1964, 1965–1974, etc., up to 2005–2014. We included data from a CS area for a given decade if there were events spanning at least 3 years in that decade. When years are fewer than three, data are discarded. For instance, our dataset includes papers relating to the MM area from 1993 onwards, so decade 1985–1994 was not considered for this area as it covers just 2 years. This grouping gives rise to the data presented in Table 2. The resulting references comprise a total of 185,402.

Establishing 3 years as minimum to represent a decade may seem a somewhat low threshold. It is important to note that some venues do not take place every year, so in a few cases the norm may not be much higher. It should be noted that the primary focus of this analysis—number of authors—is relatively independent of number of editions. Therefore, admitting a low number of editions is not expected to compromise results.

It is also important to note that whenever the analysis is not based on the grouping into decades (e.g., computing overall averages or stating the first year considered for each CS area), data from all years are considered.

Table 2 shows the 17 areas selected for our study, based on the proposal by Wainer et al. In this table, areas are ordered from oldest to youngest, based on the first year considered for any venue from that area. Figure 1 shows in a graphical form the same values in Table 2, for the nine oldest areas.

A first thing to note from Table 2 (and Fig. 1) is that all CS areas witness an increase in the average number of authors, in every decade, with just the (slight) exception of AI, in

Table 2 CS areas: first year considered, abbreviations and average number of authors in each decade

1st year	Area	1955–1964	1965–1974	1975–1984	1985–1994	1995–1904	2005–2014
1954	TH	1.413	1.478	1.749	2.148	2.432	2.631
1964	ARCH		1.754	2.113	2.603	3.226	3.756
1969	AI		1.70438	1.70356	1.809	2.172	2.908
1973	C+PL			1.727	2.129	2.626	3.306
1974	OR			1.821	1.935	2.147	2.529
1975	DB			1.955	2.322	3.132	3.797
1975	SE			1.900	2.194	2.659	3.174
1980	SEC			1.536	2.076	2.415	2.922
1981	DC			2.128	2.282	2.693	3.438
1982	GRAPH			1.780	2.302	2.895	3.820
1983	COMM				2.170	2.548	3.255
1984	MIS				2.116	2.272	2.576
1986	ML				2.119	2.355	2.886
1986	HCI				2.527	2.810	3.545
1987	IPCV				2.302	2.637	3.179
1993	MM					2.731	3.559
1994	BIO					3.382	4.181

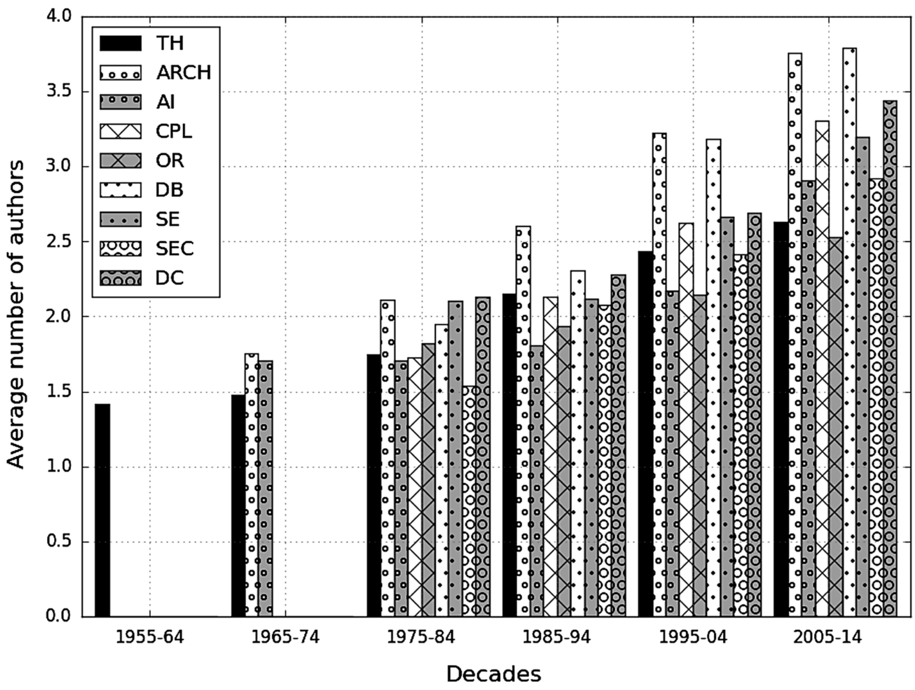

Fig. 1 Evolution in the average number of authors for the oldest CS areas

which the average for decade 1975–1984 (1.70356) signals a slight decrease from the previous decade (1.70438). Even in that case, the decrease requires a minimum of four decimal points to be noticed.

Figure 2 shows the percentages of 1-author papers, 2-author papers and so on, up to and until eight authors, for each of the CS areas. From this figure, we can see how many groups—1-author papers, 2-author papers and so on—we need to keep adding to reach the bulk of all papers from each CS. Looking at the 90th percentile (i.e., 90% of all papers) as well as the 95th, serves to indicate how significant is the tail of the distribution. Figure 1 also indicates the point when 95th percentile is reached. It turns out that in all CS areas, we need to include at least 4-authors to reach the 90th percentile. In some CS areas we need to include five authors—ARCH, C+PL, DB, SEC, DC, HCI, IPCV, MM. For GRAPH, we need to include six and for BIO we must go as far as seven. We also see that the proportion of publications with just one author or with two authors tend to be lower in the younger CS areas. It is the case of the youngest—BIO, MM—but also COMM, HCI and IPCV.

In all CS areas, the upper limit in the number of authors for 90 and 95% only differs by 1. In some cases (TH, DC, OR, MIS, IPCV), it is the same number. For example, in the case of DC, 90% of the papers have at most 5 authors, and the same goes for 95%. This means that the tail of the distribution is still somewhat “tall”, or “fat”, for 5 authors. To reach the 95th percentile, we need to include the first four groups in three CS areas—MIS, OR, and TH—five groups for six CS areas—AI, COMM, DC, IPCV, ML, and SE—and six groups for six CS areas—ARCH, C+PL, DB, HCI, MM, and SEC. For GRAPH we need to include the first seven groups and for BIO we must include eight.

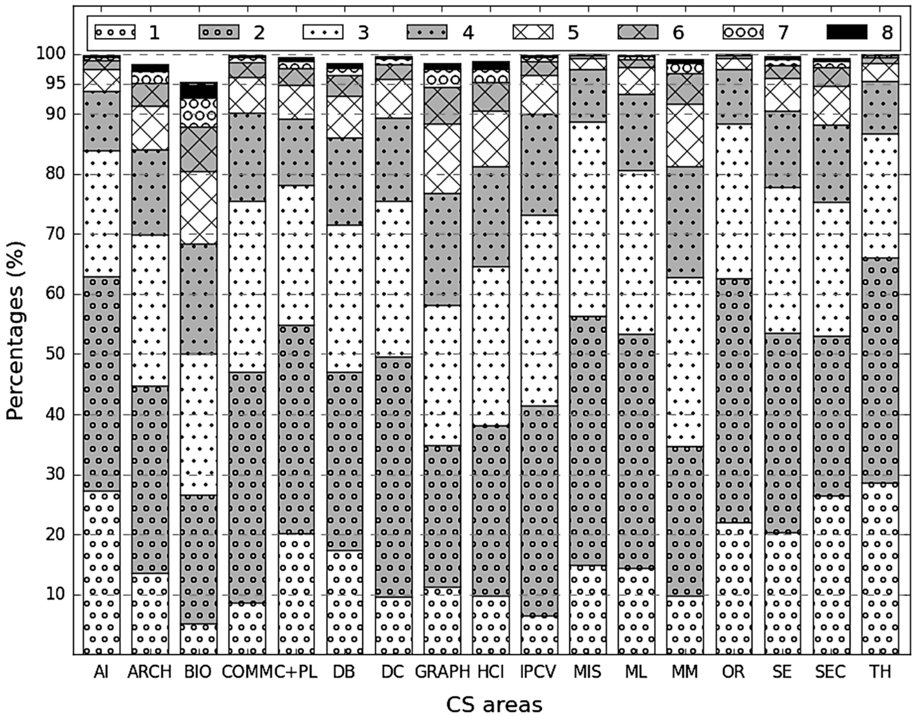

Fig. 2 Percentages of 1-author papers, 2-author papers, and so on, up to and until 8-author papers, for al 17 CS areas. Aggregate values from the full periods considered are used

Table 3 Year in which publications with a given author count first become the majority

CS area	1st year	2 overtakes 1	3 overtakes 2	4 overtakes 3
AI	1969	1994	2012	–
ARCH	1964	1983	1996	–
BIO	1994	1994	1999	–
C+PL	1973	1979	2002	–
COMM	1983	1984	2011	–
DB	1975	1975	1993	2009
DC	1981	1981	2007	–
GRAPH	1982	1982	2005	2008
HCI	1986	1986	2008	–
IPCV	1987	1990	2009	–
MIS	1984	1986	2011	–
ML	1986	1986	2013	–
MM	1993	1993	2001	–
OR	1974	1978	–	–
SE	1975	1981	2001	–
SEC	1980	1986	1996	–
TH	1954	1985	–	–

In a previous study, Fernandes (2014) used around 70,000 DBLP entries from 122 venues from the SE CS area, for the period 1971–2012. From the analysis of that sample, Fernandes concluded that single-author articles comprise over 50% of the total in the first decade and a half starting at 1971, while 3- and 4-author articles presently dominate. His analysis also indicates that until 1980, the majority of the articles have a single author, while presently articles with 3 or 4 authors comprise almost half of the total.

It is interesting to perform a similar analysis on our broader sample. To that effect, Table 3 shows, left to right, the CS area name, the first year with publications for that CS area, the first year in which 2-author publications outnumber 1-author publications, the first year in which 3-author publications outnumber 2-author publications and the first year in which 4-author publications outnumber 3-author publications. In all CS areas, there are temporary reversals in the overtakes. For instance, after AI 2-author publications overtake 1-author publications in 1994, there is still 1 year in which 1-author publications in the area of AI are in the majority (1998). Nevertheless, the tendency for the number of authors to increase is clear in all CS areas.

In all CS areas, 2-author publications end up being in the majority at some point. Five areas already start with 2-author publications in the majority: BIO (1994), DB (1975), DC (1981), GRAPH (1982), HCI (1986), ML (1986) and MM (1993).

At the end of the 1970s, 2-author publications overtake 1-author publications in 2 areas—C+PL (1979) and OR (1978). There are 6 such overtakes in the 1980s—ARCH (1983), COMM (1984), MIS (1986), SE (1981), SEC (1986) and TH (1985)—and 3 overtakes in the 1990s—AI (1994), IPCV (1990) and MM (1993).

Table 4 Average number of authors and rate of increase

CS area	First year	Average nbr. of authors	Average nbr. authors for the 2 latest decades	Derivative nbr. authors w.r.t. time	Derivative for the 2 latest decades	Total of papers
TH	1954	2.253	2.538	0.2436	0.200	13,657
ARCH	1964	3.110	3.528	0.5005	0.530	16,306
AI	1969	2.365	2.687	0.3009	0.736	13,794
C+PL	1973	2.678	3.035	0.5262	0.680	4730
OR	1974	2.306	2.388	0.2361	0.382	13,033
DB	1975	2.956	3.472	0.6139	0.665	5573
SE	1975	2.658	2.958	0.4245	0.515	7336
SEC	1980	2.672	2.782	0.4619	0.507	2880
DC	1981	2.839	3.112	0.4366	0.745	11,456
GRAPH	1982	3.420	3.543	0.6798	0.925	6647
COMM	1983	2.850	2.986	0.5425	0.707	17,704
MIS	1984	2.438	2.483	0.2298	0.304	5064
ML	1986	2.623	2.696	0.3837	0.531	13,233
HCI	1986	3.256	3.320	0.5088	0.734	7025
IPCV	1987	2.948	2.975	0.4387	0.542	29,907
MM	1993	3.265	3.287	0.8285	0.829	6269
BIO	1994	4.053	4.056	0.7983	0.798	15,221

Three-author publications end up overtaking 2-author publications in all CS areas except two—OR and TH. DB and GRAPH also reach a point in which 4-author publications overtake 3-author publications—in 2009 and 2008 respectively. In DB, the group in the majority keeps oscillating between 3-author and 4-author publications since the overtake of 2009. In GRAPH, 4-author publications are steadily in the majority since 2011.

We now turn the focus to the average number of authors throughout the history of each CS area. Table 4 shows the 17 CS areas, again ordered from oldest to youngest (columns 1 and 2 from left). Column 3 presents the average number of authors throughout the life of CS area. Column 4 presents the average number of authors for the two most recent decades, i.e., the period 1995–2014.

The CS area with the lowest average number of authors is also the oldest—TH—with an average of 2.253 authors between 1955 and 2014. Next comes AI, which is also one of the three oldest CS areas. However, ARCH—the second oldest—has a higher average (3.110) and is placed in the upper half of the list, when CS areas are ordered from lowest average to highest. The CS areas with the highest average number of authors are BIO (4.053) and MM (3.265), which are also the two youngest.

Many areas are several decades old and it may be inappropriate to use an average extending so far back into the past. Patterns of research and authorship evolved significantly and may currently be very different from what they were at the start of the considered period. The difference between the average from the two most recent decades and the overall average is positive in all CS areas, which shows very clearly that the rate of increase in the average number of authors per article is accelerating.

Even when we consider just the two most recent decades (column 4), TH is still one of the three areas with the lowest average number of authors, with 2.253 authors. Two of the three other CS areas with lowest averages—OR and MIS—are both “fringe” CS areas. The other is AI, which appears between OR and MIS.

BIO is still the CS area with the highest average number of authors per article, even when considering the two most recent decades only (with 4.056). This was expected, since both averages yield the same value for BIO, which is also the overall maximum. However, ARCH—one of the oldest CS areas—has the second highest average from the two most recent decades (3.110). ARCH seems to be a special case: while the other older CS areas tend to have relatively low average number of authors, the numbers for ARCH keep increasing at one of the highest rates.

Column five of Table 4 shows the derivative of the average number of authors with respect to time. These values are an indicative of the overall rate of increase throughout the decades. We compute it by dividing the difference between the most recent value and the oldest, by the number of decades less one. For instance, for TH $\frac{\mathrm{AVG}[2005-2014]-\mathrm{AVG}[1955-1964]}{6-1}$ yields $\frac{2.631-1.413}{5}$, which results in 0.2436.

Computed values for the above derivative are all positive, which is another indicator that the average number of authors is increasing. Values range from 0.2298 (for MIS) to 0.8285 (for MM). The CS areas with the lowest rate of increase are MIS, OR (0.2361) and TH (0.2436), in that order. The two CS areas with the highest rate of increase are also the two youngest—MM (0.8285) and BIO (0.7983).

If we compare the overall rate of increase in the average number of authors with the rate from just the two most recent decades (column 6), we get a glimpse of how the rate of increase is evolving. Values from range from 0.200 for TH to 0.925 for GRAPH. The landscape does not change significantly with respect to the previously mentioned derivative. The three CS areas with the lowest overall rate of increase—MIS (0.2298), OR

(0.2361) and TH (0.2436)—are also the ones with the lowest rate of from the 2 most recent decades—TH, MIS and OR, by this order. Thus, only the relative order of the laggards is different when we switch from the overall rate of increase to the recent rate of increase. The same is observed for the three CS areas with the highest rate of increase: it comprises BIO, MM and GRAPH in both cases. Only the relative order from the former is the opposite for the latter. So, besides GRAPH, the two areas with the biggest rate of increase are the two youngest—BIO and MM.

When we look at the total number of papers published in the seed venues, throughout the life of each CS area, we notice that the areas with the highest total of papers are not the oldest, with the exception of ARCH. This is not what we expected, since older areas had more time to accumulate a higher paper count. Instead, the highest totals belong to IPCV (29,907), COMM (17,704) and ARCH (16,306—which again follows a different trend from that of the other older areas). It is striking that, next to these, the CS area with most papers is BIO (15,221), one of the youngest.

It is also striking that three relatively old CS areas have the three lowest totals of papers among the corresponding seed venues: SEC (2880), C+PL (4730) and DB (5573). SEC is an extreme case: our sample starts at 1980 and spans 35 years. Yet, we have information of just 2880 papers—less than a hundred per year. Opportunities for future work include finding out why there are so few publications in these CS areas, particularly SEC, when the analysis is restricted to the seed venues. One hypothesis is that those areas yield fewer opportunities for easily contributing with new results that advance the field due to their accumulated body of knowledge. Another hypothesis is that the sample under study—comprising just some seed venues—does not adequately represent the population.

Conclusions

This article presents an analysis of the evolution in the number of authors of scientific publications in computer science (CS), based on a framework that structures CS into 17 constituent areas (Wainer et al. 2013). The sample following that framework comprises almost 190,000 curated references from 81 conferences and journals in the considered CS areas and spanning the period 1954–2014.

The most significant findings are as follows.

- All 17 CS areas witness an increase in the average number of authors per paper, in all decades. This is not unexpected, since similar observations exist in other works domains (Abt 1981; Cavero et al. 2014; Fernandes 2014). We detect just one exception—AI—whose rate of increase remained steady in its first two decades. After that, the numbers for AI started increasing and presently shows one of the highest rates of increase.
- The CS areas with the highest rates of increase in the average number of authors per paper tend to be the youngest. Though all CS areas witness increases, the rate of increase for some of the older areas never reached the levels of most other CS areas. It is the case of Theory (TH) and Operational Research and Optimization (OR). Computer Architecture (ARCH) is an exception among older CS areas in that its numbers keep increasing at one of the highest rates.
- Contrary to what would be expected, some of the oldest CS areas have the lowest aggregate paper count, when the analysis is restricted to the seed venues. Among those,

Security (SEC) is an extreme case whose count starts at 1980 and spans 35 years and yet registers just 2880 papers.

- In all CS areas, 2-author publications end up outnumbering 1-author publications from a certain point on. With the exception of OR and TH, 3-author publications also end up outnumbering 2-author publications. Two CS areas—Computer Graphics (GRAPH) and Database (DB)—reach the point in which 4-author publications from a given year outnumber any other group from the same year.

Acknowledgements This work was supported by FCT—Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

Abt, H. A. (1981). Some trends in American astronomical publications. *Publications of the Astronomical Society of the Pacific, 93*(553), 269. doi:10.1086/130820.

Bennett, D. M., & Taylor, D. M. (2003). Unethical practices in authorship of scientific articles. *Emergency Medicine Journal, 15*(1), 263–270. doi:10.1046/j.1442-2026.2003.00432.x.

Biryukov, M., & Dong, C. (2010). Analysis of computer science communities based on DBLP. In M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, & I. Frommholz (Eds.), *Research and advanced technology for digital libraries, volume 6273 of LNCS* (pp. 228–235). Berlin: Springer. doi:10.1007/978-3-642-15464-5_24.

Cavero, J. M., Vela, B., & Cáceres, P. (2014). Computer science research: More production, less productivity. *Scientometrics, 98*(3), 2103–2111. doi:10.1007/s11192-013-1178-2.

Cohoon, J. M., Nigai, S., & Kaye, J. (2011). Gender and computing conference papers. *Communications of the ACM, 54*(8), 72–80. doi:10.1145/1978542.1978561.

Elmacioglu, E., & Lee, D. (2005). On six degrees of separation in DBLP-DB and more. *ACM SIGMOD Record, 34*(2), 33–40.

Fernandes, J. M. (2014). Authorship trends in software engineering. *Scientometrics, 101*(1), 257–271. doi:10.1007/s11192-014-1331-6.

Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology, 62*(10), 1992–2012. doi:10.1002/asi.21614.

Freyne, J., Coyle, L., Smyth, B., & Cummingham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM, 53*(11), 124–132. doi:10.1145/1839676.1839701.

Garousi, V., & Fernandes, J. M. (2016). Highly-cited papers in software engineering: The top 100. *Information and Software Technology, 71*, 108–128. doi:10.1016/j.infsof.2015.11.003.

Greene, M. (2007). The demise of the lone author. *Nature, 450*(7173), 1165. doi:10.1038/4501165a.

Gu, Y. (2002). An exploratory study of Malaysian publication productivity in computer science and information technology. *Journal of the American Society for Information Science and Technology, 53*(12), 974–986. doi:10.1002/asi.10125.

Laender, A., de Lucena, C., Maldonado, J., de Souza e Silva, E., & Zivianim, N. (2008). Assessing the research and education quality of the top Brazilian computer science graduate programs. *ACM SIGCSE Bulletin, 40*(2), 135–145.

Ley, M. (2009). DBLP—Some lessons learned. In *Proceedings of the VLDB endowment (PVLDB 2009)* (Vol. 2, Number 2, pp. 1493–1500).

Meyer, B., Choppy, C., Staunstrup, J., & van Leeuwen, J. (2009). Research evaluation for computer science. *Communications of the ACM, 52*(4), 31–34.

Patterson, D. A. (2004). The health of research conferences and the dearth of big idea articles. *Communications of the ACM, 47*(12), 23–24. doi:10.1145/1035134.1035153.

Solomon, J. (2009). Programmers, professors, and parasites: Credit and co-authorship in computer science. *Science and Engineering Ethics, 15*(4), 467–489. doi:10.1007/s11948-009-9119-4.

Vardi, M. Y. (2009). Conferences vs. journals in computing research. *Communications of the ACM, 52*(5), 5. doi:10.1145/1506409.1506410.

Wainer, J., Eckmann, M., Goldenstein, S., & Rocha, A. (2013). How productivity and impact differ across computer science subareas. *Communications of the ACM, 56*(8), 67–73. doi:10.1145/2492007.2492026.