Scientometrics
https://doi.org/10.1007/s11192-021-04035-5

Author placement in Computer Science: a study based on the careers of ACM Fellows

João M. Fernandes[1] · António Costa[1] · Paulo Cortez[2]

Received: 11 February 2021 / Accepted: 7 May 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract
The order in which authors of a scientific paper place their names on the byline follows in many research fields some implicit rules. In most fields, the first author is considered to be the one who contributed most to the intellectual effort described in the paper. Additionally, the last author is normally the most senior researcher and in many situations the contribution to the paper is more indirect. In this manuscript, we intend to analyse the evolution of the positions of computer science (CS) researchers on the bylines of scientific papers throughout their careers. In particular, this bibliometic study considers the Association for Computing Machinery (ACM) Fellows (the most prestigious members) that present a long and rich publication record. Our hypothesis is that young CS authors tend to have their names placed in the first positions of the bylines, while senior CS researchers are often considered as last authors. Several statistical analyses were conducted by using biblometric data collected from ACM Fellows and other CS researchers. Overall, the obtained results do confirm our initial hypothesis.

Keywords Bibliometrics · Scientific authorship · Authors order · Scholarly publication

Introduction

The strong pressure faced by researchers in order to publish scientific papers (known as the "publish or perish" dilemma) implies some practical and ethical issues related to authorship (Abt, 1981; Bennett and Taylor, 2003; Solomon, 2009). In fact, many dimensions of a researcher career, such as funding, professional promotions, tenure, prestige, and collegial respect, are strongly dependent on the number and quality of the scientific publications.

✉ João M. Fernandes
jmf@di.uminho.pt

António Costa
costa@di.uminho.pt

Paulo Cortez
pcortez@dsi.uminho.pt

[1] ALGORITMI Center, Department of Informatics, University of Minho, 4710-057 Braga, Portugal

[2] ALGORITMI Center, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal

This paper focuses on a particular relevant authorship issue that is related with the positions of the authors names on the bylines, which may influence the assumptions that readers make about the exact author contributions to the research (Maciejovsky et al., 2008; Fox et al., 2018).

Different research fields tend to follow their own implicit rules to establish the order in which authors of a scientific paper place their names on the byline. In a small number of fields (e.g., Economics, Mathematics, and Business, Management and Accounting), the order is highly alphabetical (Henriksen, 2019; Fernandes and Cortez, 2020). In the scientific areas where authors are not listed alphabetically, which corresponds to most cases (Fernandes and Cortez, 2020), common field practices are often applied. For instance, in the biomedical sciences, the order of the authors reflects the role they play in the process of writing the articles (Marschke et al., 2018).

Several authors argue that only the first and last authors have a more general accepted meaning across different research fields (Reisenberg and Lundberg, 1990; Rennie et al., 1997; Kennedy, 2003; Wren et al., 2007). The advantage of being the first author of a scientific paper is a well-studied question (Engers et al., 1999; Krasnova et al., 2012; Ackerman and Brânzei, 2017). The first author (of a non-alphabetically ordered paper) is usually considered to be the one who has taken the initiative and responsibility for the research and has developed most of the work. Often, the first author is considered to be the one who contributed most to the intellectual effort described in the paper. While it is generally agreed that the first author is the primary contributor for the work, the last one is usually the principal investigator, who supported the work. If the first author is a student or a subordinate scientist, then the last author is traditionally her/his supervisor or mentor. Indeed, the last author is often the most senior researcher and in many situations the direct contribution to the paper is not minimal but of a different type (Buehring et al., 2007; Kosmulski, 2012). There are other implicit authorship rules in some scientific areas. For instance, Zuckerman (1968) uses the term “noblesse oblige” to indicate that Nobel laureates allow their co-authors to be the first ones, even when their own contributions is higher than the other co-authors’.

This manuscript aims to analyse the evolution of the positions of researchers on the bylines of scientific papers throughout their careers. We consider researchers from the specific field of computer science (CS) and that present a long and rich publication record. Our hypothesis, as an implicit rule in this field, is that more junior authors tend to have their names placed firstly (thus in the “left” of the bylines), while senior researchers are often placed lastly (in the “right” side of the bylines). This hypothesis is rooted in two assumptions: (1) CS researchers tend to reduce their time to truly perform scientific work as they advance in their careers, devoting more of their attention to managerial duties and roles; (2) first positions on the byline of a CS paper are typically occupied by the researchers who made more intellectual contribution, while the last ones are occupied either by those who made the smallest contributions or by the most senior researchers. It should be noted that Association for Computing Machinery (ACM) is the world’s largest scientific CS society, supporting directly several top tier journals and conferences and known prizes (e.g., ACM Turing Award). Moreover, ACM Fellows are considered as career awards, given that they are related to a small and special selection of top 1% ACM members that produced outstanding contributions to the CS field. Thus, most of these researchers should have a strong scientific merit. Using the collect ACM Fellows data, we then performed several statistical analyses to check if our hypothesis is valid, which include the calculation of an author position index (API) and diverse data distribution graphs. The obtained results were complemented by considering similar analyses over another dataset with a random and larger selection of CS researchers. Overall, both ACM Fellows and generic CS researchers exhibit similar author placement patterns.

The paper is organized as follows. "Related work" section describes the state of the art. Then, the adopted research methodology is presented ("Methodology" Section). Next, the obtained results are presented and analysed ("Results" section). Finally, limitations are discussed in "Limitations" section and conclusions and future work are presented in "Conclusions" section.

Related work

The decision on the order of the authors of a scholarly publication can follow several approaches, as indicated by Peidu (2019):

1. Amount of contribution;
2. Alphabetical order;
3. Multiple first author or multiple last author;
4. Seniority or reverse seniority;
5. Raffling or lottery system; and
6. Negotiation or mutual understanding.

The choice of the approach is strongly dependent on the common practices of the research field . A small number of scientific fields, such as Economics and Mathematics, adopt mostly the alphabetical order (Fernandes and Cortez, 2020). However, other scientific communities follow distinct implicit rules. In the particular analysed CS field, the first author of a multi-authored paper is often considered to be the most important contributor. Similarly, the last position tends to be used for the most senior, older or most prestigious author. Despite this perception, comprehensive studies on the authors order in CS are not abundant. We next describe some related works that address the authors order in different fields, aiming to provide an overall perspective on how distinct bibliometric researchers address this topic.

Liang et al. (2004) analysed three Chinese universities, aiming to study the fraction of co-authored publications where the graduate student's name precedes that of the supervisor's (the *g*-ratio concept). They found that the doctoral *g*-ratios of all three universities are as high as 80%, which reflects a regular structure of the scientific collaboration between doctoral candidates and their supervisors. They have also shown that in general master students *g*-ratios are smaller than doctoral level *g*-ratios.

In another study, Moore and Griffin (2006) analysed the factors that affect the placement of names in co-authored publications in education-related journals. The obtained results indicate that both contribution amount and idea origination were typically used to determine name placement, but authorship credit was also assigned based upon criteria like seniority and assistance to colleagues. More recently, Costas and Bordons (2011) presented a study that analysed the order of authorship for more than 1000 permanent Spanish scientists from three scientific fields (Biology and Biomedicine, Materials Science, and Natural Resources). They have shown that there is a trend for younger researchers to appear in the first position, while more senior ones are more likely to sign in the last position. Although these two articles have objectives that are similar to our work, they are related with different scientific fields and thus cannot be transposed to CS.

In another bibliometric research, Liu and Fang (2014) examined the authorship preferences of scientific group leaders for seven research fields and eleven geographic locations. In Mathematics and "Physics, Particles & Fields", the typical rule is for authors to be listed alphabetically. However, scientific group leaders from Egypt and Shanghai usually list their names either first or last in the byline, the same as group leaders in other research fields. Senior authors from Egypt often appear as the first authors, a pattern that is not observed to the group leaders from other locations.

Cabanac et al. (2015) presented an analysis of the publication records of 3860 CS researchers with the objective of studying the evolution patterns of their co-authorships. Their contribution is however different from ours, since they are focused in the authorship collaborations (i.e., their co-authors) that those researchers have established.

Abramo et al. (2016) explored the relationships among research performance, age, and seniority of full professors in Italy. They analysed a 5-year period (2006–2010), using performance indicators that take into account the positions of the names of those professors on the papers bylines, for eleven fields (they disregarded the ones where the practice is to place the authors in alphabetical order).

Another interesting study, also based on data obtained from the DBLP service, but limited to USA and Canada, is presented by Way et al. (2017). They show evidences of a gradual shift toward last-authorship position, with the relative first/last proportion reaching stability around the 8th year. Faculty members at the top institutions have their average proportion of last-author papers significantly higher than those of other faculty, which confirms the idea that professors at elite institutions tend to begin working with students earlier and have larger or more productive research groups.

The related works are summarized in Table 1, which assumes a chronological order and the following characterizing columns: Geogr—the researchers geographic location region; Fields—the analysed scientific fields; Source—the bibliographic data source; Period—the time period of analysis of the publications; Authors—the number of considered researchers; and Pubs—the number (in thousands) of analysed publications. The last row of Table 1 positions our research, showing that we consider a wider research geographic region (the

Table 1 Summary of the related work

Study	Geogr[a]	Fields[b]	Source[c]	Period	Authors	Pubs (10^3)
Liang et al. (2004)	China	32	CDDB+CSCD	1989–1998	14,953[d]	13
Moore and Griffin (2006)	USA	Edu.	n.a.	n.d.	60	n.d
Costas and Bordons (2011)	Spain	3	WoS	1994–2004	1064	25
Liu and Fang (2014)	11 c.	7	WoS	2002–2011	n.d.	275
Abramo et al. (2016)	Italy	11	WoS	2006–2010	11,989	n.d.
Way et al. (2017)	USA+Canada	CS	DBLP	1970–2011	2453	200
This work	World	CS	DBLP	1953–2021	636	161
				1956–2021	18,649	2432

[a]Geography locations of the analysed researchers (11 c.—11 countries).

[b]Analysed research fields, in terms of number of fields or specific field (Edu.—Education; *CS* Computer Science).

[c]Bibliographic source (*CDDB* Chinese Dissertation Document Bibliography Database; *CSCD* Chinese Science Citation Database; *WoS* Web of Science)

[d]Pairs of authors (student/supervisor) *n.d* non disclosed

entire world), a larger data collection period and also a larger number of authors and publications (in particular for the CS Others dataset, "Bibliometric data" Section).

Moreover, there is only one work that targets the specific CS domain (Way et al., 2017), although with a narrower geography coverage and time span. Furthermore, it should be noted that previous works only consider first and last author positions (e.g., Costas and Bordons, 2011; Way et al., 2017). Thus, our work is the only one that uses a numeric indicator (the Author Position Index) that measures how close an author name is to the beginning or to the end of a list of authors. This is an important aspect, since, contrarily to the widely adopted binary indicator (first or last), it provides a better measurement of the ordering level that was adopted by the authors.

Methodology

Research goal

The research approach we have used in our study is the Goal, Question, Metric (GQM) methodology (Basili, 1992). Following the GQM goal template, the goal of this research work is to study the authorship positions of CS researchers throughout their careers. To tackle this goal, the main research question (RQ) is the following:

> Is there a tendency for CS researchers, as their careers advance, to move their names on the paper bylines from the first (leftmost) positions to the last (rightmost) ones?

Bibliometric data

This bibliometric study aims to perform a comprehensive analysis of the evolution of the positions occupied by senior CS researchers throughout their careers. In this manuscript, the initial list of potential senior authors includes those that are ACM Fellows. The Association for Computing Machinery (ACM), founded in 1947, is the world's largest scientific and educational computing society (https://www.acm.org). The ACM Fellow title, established in 1994, is the most prestigious member grade and recognizes the top 1% of its members for their outstanding accomplishments in the computing field or outstanding service to ACM and the computing community. A candidate for Fellow must have at least five years of professional membership within the last 10 years. As of September 2020, the list of ACM Fellows (http://awards.acm.org/fellows/award-winners) contains a total of 1221 members that received the award from 1994 until 2020. We adopt this list in this paper, since it provides a valuable set of researchers that obtained a world-class scientific level in the CS field.

Based on the list of all ACM Fellows, the next step was to obtain their list of scientific publications. For this purposes, we adopted the DBLP website (Ley, 2009), since it is specifically devoted to bibliographic information on a vast list of CS journals and conference proceedings. Other studies (e.g., Elmacioglu and Lee, 2005; Fernandes, 2014; Fernandes and Monteiro, 2017; Kim, 2018) have also used DBLP to obtain bibliographic data. DBLP can be interfaced by automatic mechanisms, which eases the retrieval process. The bulk of the data for the study was obtained on November 21, 2020, by downloading from the DBLP database all (160,955) publication entries associated with 929 ACM Fellow profiles.

For each exact name in the ACM Fellows list, a search query with that name was sent to the DBLP server through its public application programming (API) interface (https://dblp.org/search/author/api). Each query string was pre-processed in order to comply with the DBLP formats. For example, the publications of ACM Fellow Albert R. Meyer, which is listed in the ACM list as "Meyer, Albert R", can be retrieved if the query uses the substring "Meyer$+Albert+R" for the identification of the author. The complete URL to query the DBLP API server for this particular example is: https://dblp.org/search/author/api?q="Meyer$+Albert+R"&format=xml. The server responds with an XML file with a list of the matching DBLP authors. When the number of returned authors is either 0 or greater than 1, the author is ignored. During our data collection procedure, 292 ACM fellows were in this situation. For the remaining 929 cases, exactly one author was returned. In those cases, the retrieved XML content was further processed to obtain the DBLP id and the corresponding URL for that author. For Albert R. Meyer, the URL is the following: https://dblp.org/pid/m/ARMeyer. To download the Bibtex file with all the DBLP publications of Albert R. Meyer the following URL was used: https://dblp.org/pid/m/ARMeyer.bib. This process was automated with an R script, allowing to fetch all ACM Fellows Bibtex/XML files, which were saved into a local server.

The collected Bibtex files were then manually edited to allow the parsing mechanism to use one unique name. We identified a few cases where an author has used throughout her/his career two or more different name versions in the DBLP papers (e.g., John A. Stankovic and Jack A. Stankovic; Paramvir Bahl and Victor Bahl; Rodney G. Downey and Rod Downey; Marilyn Wolf, Marilyn Claire Wolf, Wayne H. Wolf, and Wayne Hendrix Wolf). Thus, we reviewed manually these cases in order to produce a unique name identifier for all papers of the same author.

In order to include a minimum research career time span and amount of publications, the collected DBLP data was further filtered by considering all papers for each ACM Fellow that: (1) has at least 20 years of publication (when comparing his first and last DBLP publications); and (2) has an average of two or more papers per year when considering only papers that have at least two authors. The resulting dataset includes a total of 131,041 publications (authored by two or more researchers), for a total of 636 ACM Fellows. Table 2 presents the bibliometric data attributes that were considered and grouped in terms of two main items: publications and authors.

Our approach to identify senior researchers has some similar aspects to the one followed by Cabanac et al. (2015). They also obtained their data from the DBLP service, which contains 3860 researchers who published at least 15 papers in CS conferences and journals,

Table 2 Adopted bibliometric data attributes for the ACM Fellows dataset

Context	Attribute	Description
Publication	A	ACM Fellow unique identification (e.g., "A0001", "A0009")
	Y	Publication year (1953 to 2021)
	R	Research experience of A (in years, from 1 to 60)
	N	The total number of publication authors ($N \in \{2, 3, ..., 124\}$)
	P	Position of A in the publication: from 1 (first author) to N
Author	F	Year in which the ACM Fellow title was granted (1994 to 2019)
	G	Generation of the ACM Fellow ({1950,1960,...,2000})

provided that they started publishing in the period 1980–1985, and were still active from 2005 onward. So, they also considered CS authors with at least 20 years of publication.

We highlight that the ACM Fellows dataset bibliometric dataset includes only publications that have two or more authors, since these are the ones where the analysis of authorship position makes sense. Nevertheless, the research experience attribute is computed as $R = Y - Y_f + 1$, where Y_f denotes the year in which the first DBLP paper was published (regardless of how many authors appear in such paper). We further note that in this paper, the research experience (R, in years) is used as a reasonable measure of seniority (the higher the value, the longer is the research career). Regarding the authors (total of 636 unique A values), the year in which the ACM Fellow title was granted was retrieved from the ACM Fellows webpage. As for the generation attribute (G), it corresponds to the decade in which the first DBLP paper is inserted (Y_f). The dataset includes the following generation distribution of the 636 ACM Fellows: 1950—1; 1960—5; 1970—29; 1980—120; 1990—179; 2000—302.

For comparison purposes, a second bibliometric dataset was retrieved from DBLP, termed here CS Others. It includes a random selection of 18,649 CS researchers that match the selection criteria (minimum of 20 years of publication record and average of two or more papers per year) and that are not ACM Fellows. The obtained dataset includes a total of 2,432,307 papers that were published from 1956 to 2021. For each publication, we stored the same publication context attributes shown in Table 2 (e.g., the research experience R varies from 1 to 62). The contents of both datasets are available at http://www.gcom.di.uminho.pt/pubs/scim21.

Bibliometric statistics

In this study, we define the author position index (API), within the range $[-1.0.. + 1.0]$, and that expresses the position of an author's name on the byline:

$$\mathrm{API}_p = \frac{2(P-1)}{N-1} - 1 \tag{1}$$

P denotes the position of author A in a publication p that has a total of N authors. The extreme values -1.0 and $+1.0$ indicate that the author name is in the first and last positions of the byline, respectively. If the author is in a position closer to the first position than to the last one (in the leftmost part of the list), the resulting API_p value is negative (e.g., -0.5). Similarly, positions closer to the last position than to the first one (in the rightmost side of the list) are assigned with positive API_p values (e.g., +0,25). It should be noted that the API_p index is only computed for papers with two or more authors, which corresponds to our ACM Fellow DBLP dataset. For exemplification purposes, Table 3 presents all possible API_p values for papers with $N = 2$ to $N = 8$ authors.

Several of the API_p analyses of "Results" section assume their global evolution in terms of the research experience attribute (R) from Table 2. To simplify these analyses, this attribute assumes a total of 21 bins, corresponding to the first 20 years plus all other years (summed into the "> 20" bin). To define a global API_p value (for a particular R bin), we adopt the common average and median statistics. In some graphical plots, the average point is complemented by its Student's t-distribution 95% confidence interval. We also consider two main API_p aggregation methods: by publication (BP) and by author (BA). The BP method assumes all publications associated with a particular R bin. As for BA, it assumes a first computation of the average (or median) API_p value for all papers of author A for the

Table 3 Example of the Author Position Index (API_p) values

N	$P=1$	$P=2$	$P=3$	$P=4$	$P=5$	$P=6$	$P=7$	$P=8$
2	−1	+1	–	–	–	–	–	–
3	−1	0	+1	–	–	–	–	–
4	−1	$-\frac{1}{3}$	$+\frac{1}{3}$	+1	–	–	–	–
5	−1	$-\frac{1}{2}$	0	$+\frac{1}{2}$	+1	–	–	–
6	−1	$-\frac{3}{5}$	$-\frac{1}{5}$	$+\frac{1}{5}$	$+\frac{3}{5}$	+1	–	–
7	−1	$-\frac{2}{3}$	$-\frac{1}{3}$	0	$+\frac{1}{3}$	$+\frac{2}{3}$	+1	–
8	−1	$-\frac{5}{7}$	$-\frac{3}{7}$	$-\frac{1}{7}$	$+\frac{1}{7}$	$+\frac{3}{7}$	$+\frac{5}{7}$	+1

particular *R* bin. Then, the computed values are aggregated by computing their average (or median) values for all *A* authors associated with the *R* bin.

Results

Using the R tool, we first computed the API_p scores for all 131,041 papers available in the ACM Fellows dataset. Then, we performed several aggregated analyses, which are shown here in terms of graphs.

Figure 1 shows the main statistical analysis, plotting in the *y*-axis the aggregated API_p scores (average and respective 95% confidence intervals; median) versus the years of research experience (in *x*-axis, *R* bin). It displays the result of two aggregation methods: BP (publication based, top graph) and BA (author based, bottom graph). The most important result is that both graphs and aggregation statistics (average and median) show the same growing trend evolution of the API_p scores as the research career proceeds. In effect, the more junior researchers ($R \leq 5$) present negative aggregated API_p scores, thus confirming that they tend to be positioned on the leftmost side of the author paper bylines. Then, there is mid-career phase ($5 < R < 10$), where authors obtain mostly zero API_p scores, denoting either a mid-paper byline position or an equal mixture of first and last paper authorships. Finally, there is a senior-career phase ($R \geq 10$) where both average and median curves move into more positive API_p values (the rightmost side of the bylines). In particular, the 95% confidence interval whiskers confirms that the differences are significant (visible when two confidence intervals do not overlap) for the evolution of the average API_p values.

To compare how representative are ACM Fellows for the CS field, we have also computed the API_p scores for all 2,432,307 papers from the CS Others dataset. Figure 2 presents the respective aggregated results, when considering the BP (top graph) and BA (bottom plot) methods. It should be noted that the CS Others dataset includes a larger number of researchers and papers when compared with the ACM Fellow data, thus the confidence intervals in Fig. 2 are naturally smaller when compared with Fig. 1. More importantly, the CS Others results exhibit the same generic pattern previously identified for the ACM Fellows data. Indeed, for both BP and BA graphs, there is a consistent increase of the API_p scores when the years of research experience evolve. For instance, when using the BP aggregation method, the average API_p when $R = 1$ is − 0.21 for the ACM Fellows data and − 0.27 for the CS Others records. The same average increases to 0.46 when $R > 21$ for the ACM Fellows and to 0.38 for the CS Others researchers. The dataset differences are rather small when considering the extreme *R* points: 0.06 points when $R = 1$ and 0.08 points when $R > 21$. Thus, these results confirm that both ACM Fellows and generic CS authors tend to move their names in the paper bylines through their careers in a similar way.

In Fig. 3, we detail the evolution of the average API_p scores for different generations of ACM Fellows (G). The decades of 1950 and 1960 were omitted in this analysis, since they correspond to a very small number of ACM Fellows ($N = 6$), thus with lack of statistical robustness. Figure 3 shows that the previously identified global API_p score career pattern was followed by different generations of researchers, with all generation curves being aligned with a negative (e.g., $R \leq 5$) to positive (e.g., $R \geq 10$) average API_p career shift. An interesting pattern is related to the most recent generation (G =2000), which contains more researchers (N =302). For these researchers, while the general average API_p career increase is still visible, there is a higher variability for small R differences (e.g., there is a strong decrease from $R = 1$ to $R = 2$ and strong increase from $R = 7$ to $R = 8$), when compared with the previous generations. The identification of the correct explanation for this higher variability for the 2000 generation researchers authorship pattern would require data that we do not consider (e.g., analysis of the full researcher CVs) and thus is left for future work. Nevertheless, we hypothesise that it might be due to two factors. Firstly, several recent CS researchers work in companies, thus their contributions might be focused towards being more directly involved with research (when compared with academic scholars). Secondly, there is currently more collaboration among researchers, which might translate into a wider range of author paper roles. For instance, one researcher could act as the principal investigator of a given R&D project, while collaborating in another project managed by a colleague.

The aggregated ACM Fellows BP method results from Fig. 1 (top graph) are further inspected by performing a more fine-grained Beeswarm plot analysis (Fig. 4). When compared with other summarisation graphs, Beeswarm plots (also termed as violin scatter plots) have the advantage of providing an easily visualisation of the density of the data distributions in a single axis value (Kabacoff, 2020). In Fig. 4, a gray and black colouring is used to denote papers published before (gray, Fellow = "no") and after (black, Fellow = "yes") receiving the ACM Fellow title. The graph shows a clear point density and darker color increase when moving from the bottom left (young researchers, more negative API_p scores) to the top right (senior researchers, more positive API_p scores). Thus, the fine-grained results confirm our CS career authorship placement hypothesis.

The last two graphs in Fig. 5 complement Fig. 4 by better characterizing the first dataset in terms of the year when the ACM Fellow awards were granted. The left of Fig. 5 presents the temporal (in years) of the time required for a CS researcher to become an ACM Fellow (measured from the year when $R = 1$). The right of Fig. 5 plots the total number of papers published within our dataset range (x-axis, from 1960 to 2020). This graph includes the evolution of two curves, papers published before (Fellow = "no") and after (Fellow = "yes") becoming ACM Fellow. The left plot shows that there is a general decrease, with more recent CS researchers requiring less time to obtain the ACM Fellowship. As for the right graph, it is a natural consequence of our data collection method, all researchers are ACM Fellows and thus the number of papers published after receiving the award title becomes the majority at the end of our collection period (specifically, after the year of 2006). Finally, to better characterize the ACM Fellows dataset, Fig. 6 shows the evolution (the publication year, x-axis) of the distribution of the papers in terms of the total number of authors (N, y-axis). Similarly to what has been shown in Fernandes and Monteiro (2017), the plot confirms that there is a growth in the total number of CS paper authors through time.

Fig. 1 Evolution of the author position API_p index (y-axis) according to the number of years of research ▶ experience (R, in the x-axis) for the *ACM Fellows* dataset. Top graph shows aggregated values for all publications (BP method), while bottom graph plots aggregated values for each researcher (BA method). The black line and circle points denote the average value and respective 95% confidence intervals. The gray line and diamond points represent the median values.)

Limitations

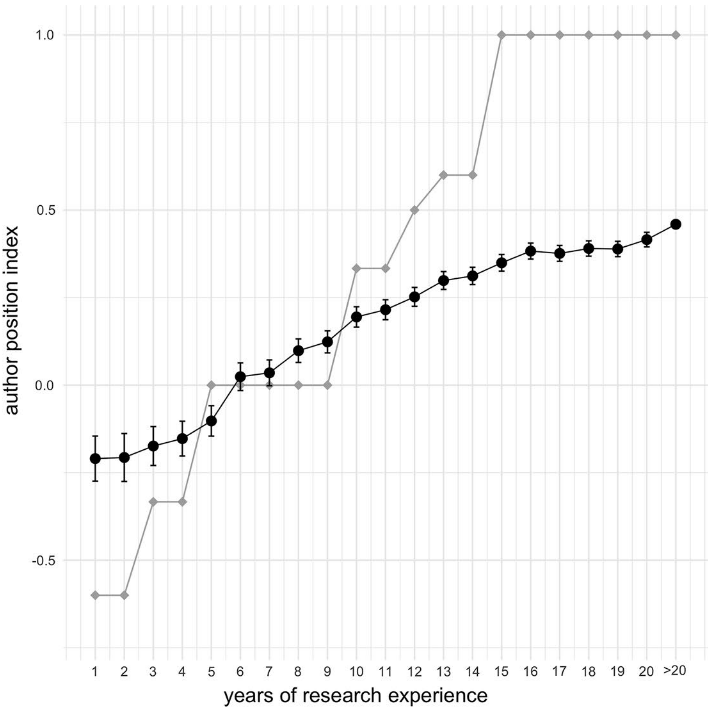
While interesting results were achieved, some study limitations are here discussed. We only considered bibliometric data related with the DBLP publication database. While the DBLP database provides a strong coverage of CS publications, there might be interdisciplinary papers, co-authored by CS authors, that are not listed there. Moreover, we only consider author placement data. As already discussed, a richer analysis could be achieved if other data sources were included, such as public researcher curriculum profiles. Nevertheless, retrieving such data gives rise to additional challenges that are out of the scope of this research (e.g., handling unstructured texts or lack of standardisation of the research curricula).

Conclusions

This paper provides a systematic bibliometric study on the evolution of the positions of the names of computer science (CS) researchers throughout their publication careers. The researchers considered in this manuscript are 636 ACM Fellows, with publication records that span for 20 years and with an average of two or more papers per year. The publication details were collected from the DBLP database, producing a total of 131,041 author paper placement records. Using these records, we have held several statistical analyses, based on a proposed author position index (API) and distinct aggregation measures. For comparison purposes, we retrieved another dataset (CS Others) from the DBLP database and that includes 2,432,307 publications from 18,649 randomly selected researchers. The results confirm our hypothesis that CS authors tend to have their names more often placed at the leftmost positions on the bylines during the beginning of their careers and at the rightmost part at the end. Indeed, similar author placement results were obtained for both ACM Fellows and CS Others datasets.

The CS field as many specific bibliometric patterns, like for example, conferences being often the preferred venues for publication (Kim, 2019; Franceschet, 2011). Thus, it worth to analyse authorship placement specifically for this field, as conducted in this research. The particular proposed metric (API) allowed to confirm the initial hypothesis that junior CS authors tend to place their names in the first positions of the bylines, while senior ones normally assume the last positions. Thus, the API metric and associated aggregation measures can be used as a reliable proxy for determining the level of seniority of CS researchers based on their publication record. These metrics could be potentially used to define specific research policies for junior or senior scholars. For instance, funding agencies or governmental bodies could discriminate positively junior researchers, when analysing research proposals.

In future work, we intend to approach bibliometric data related with researchers from other scientific fields, to check if similar findings can be found. It would also be


author position index
1.0
0.5
0.0
-0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >20
years of research experience

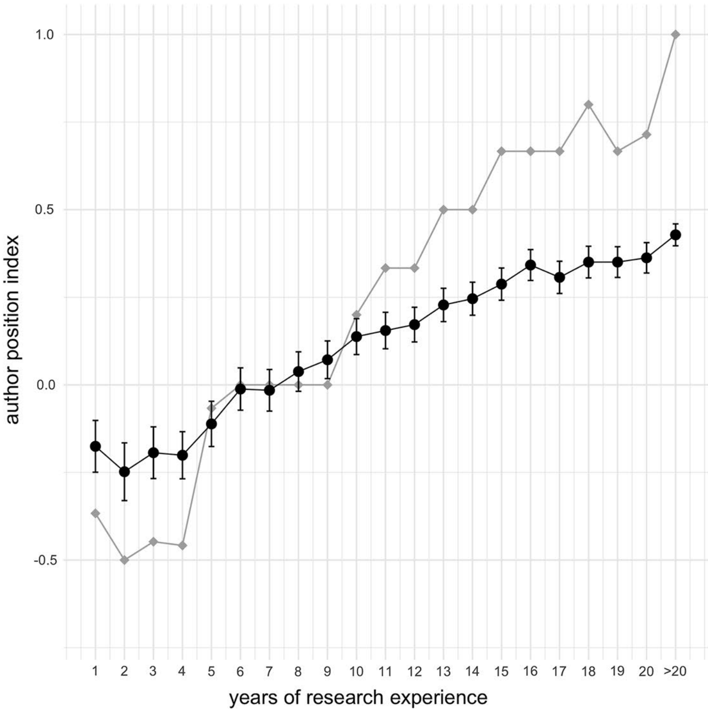
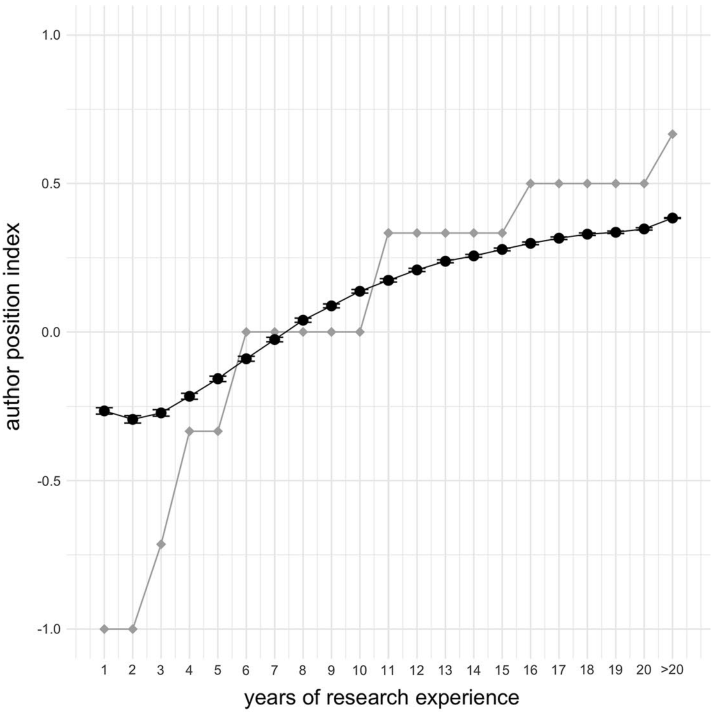

author position index
1.0
0.5
0.0
-0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >20
years of research experience

Fig. 2 Evolution of the author position API_p index (y-axis) according to the number of years of research experience (R, in the x-axis) for the *CS Others* dataset. Top graph shows aggregated values for all publications (BP method), while bottom graph plots aggregated values for each researcher (BA method). The black line and circle points denote the average value and respective 95% confidence intervals. The gray line and diamond points represent the median values.) ▶

interesting to analyse more career details, such as curricula or citations, to further clarify and explain some API evolution patterns, such as the generation differences.


1.0
0.5
0.0
-0.5
-1.0
author position index
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >20
years of research experience

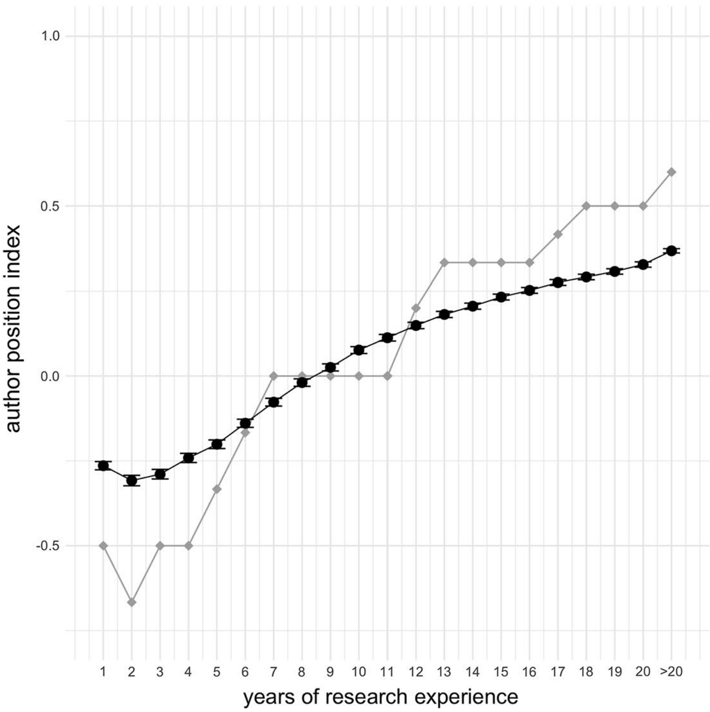

1.0
0.5
0.0
-0.5
author position index
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >20
years of research experience

Fig. 3 Evolution of the ACM Fellows average author position API_p index (y-axis) for all papers (BP method) and different generation researchers (G) according to the number of years of research experience (R, in the x-axis)

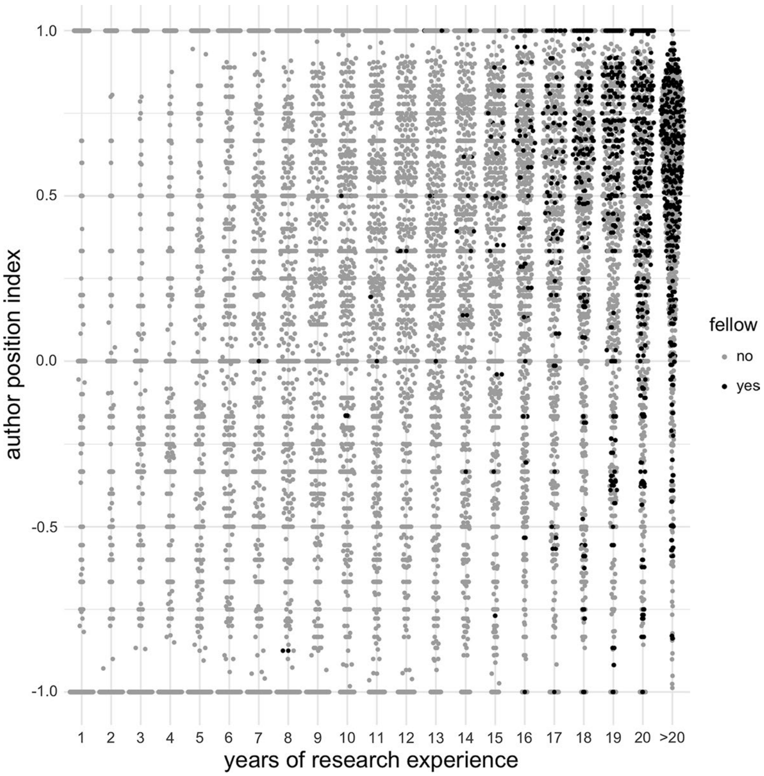


Fig. 4 Evolution of the ACM Fellows beeswarm plots of the author position API_p index (*y*-axis) for all researchers according to the number of years of research experience (*x*-axis)

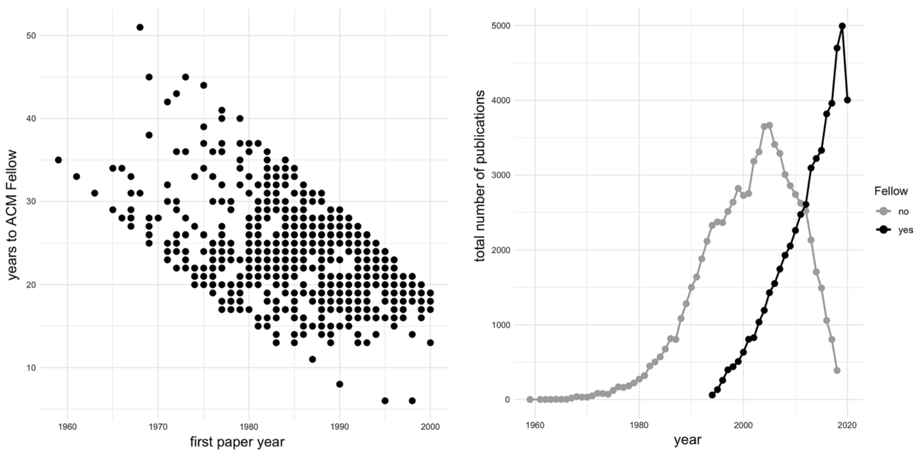


Fig. 5 Temporal evolution (*x*-axis, in years) of the number of years to Fellow (*y*-axis, left plot) and total number of ACM Fellows publications (*y*-axis, right plot)

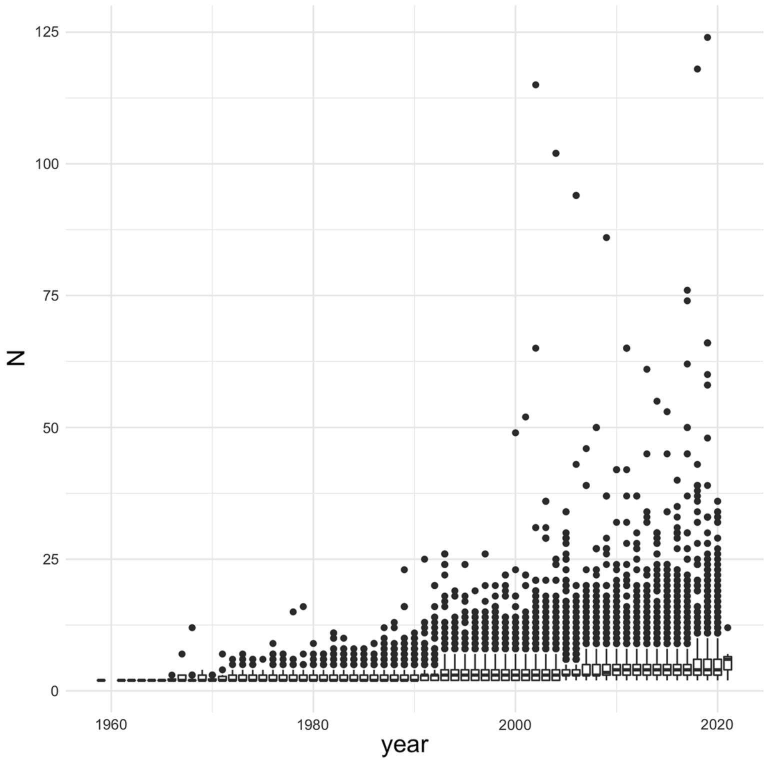


Fig. 6 Temporal evolution (x-axis, in years) of the boxplots for the total number of paper authors ($N \geq 2$, in y-axis) and ACM Fellows data

Acknowledgements This work has been supported by FCT—Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. We would also like to thank the anonymous reviewers for the insightful comments that helped improving the contents of the manuscript.

Author Contributions J. M. Fernandes performed the conceptualisation, methodology, software, investigation, resources, formal analysis, writing—original draft, writing—review and editing. A. Costa was responsible for software, writing—review and editing, and data curation. P. Cortez contributed with methodology, software, investigation, validation, formal analysis, writing—review and editing, and visualisation.

References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2016). The combined effects of age and seniority on research performance of full professors. *Science and Public Policy*, *43*(3), 301–319. https://doi.org/10.1093/scipol/scv037.

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

Ackerman, M., & Brânzei, S. (2017). The authorship dilemma: Alphabetical or contribution? *Autonomous Agents and Multi-Agent Systems*, *31*(5), 1077–1093. https://doi.org/10.1007/s10458-016-9351-7.

Basili, V. R. (1992). Software modeling and measurement: The goal/question/metric paradigm. Technical report, University of Maryland.

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

Buehring, G. C., Buehring, J. E., & Gerard, P. D. (2007). Lost in citation: Vanishing visibility of senior authors. *Scientometrics*, *72*(3), 459–468. https://doi.org/10.1007/s11192-007-1762-4.

Cabanac, G., Hubert, G., & Milard, B. (2015). Academic careers in computer science: Continuance and transience of lifetime co-authorships. *Scientometrics*, *102*(1), 135–150. https://doi.org/10.1007/s11192-014-1426-0.

Costas, R., & Bordons, M. (2011). Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective. *Scientometrics*, *88*(1), 145–161. https://doi.org/10.1007/s11192-011-0368-z.

Elmacioglu, E., & Lee, D. (2005). On six degrees of separation in DBLP-DB and more. *ACM SIGMOD Record*, *34*(2), 33–40. https://doi.org/10.1145/1083784.1083791.

Engers, M., Gans, J., Grant, S., & King, S. (1999). First-author conditions. *Journal of Political Economy*, *107*(4), 859–883.

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

Fernandes, J. M., & Cortez, P. (2020). Alphabetic order of authors in scholarly publications: A bibliometric study for 27 scientific fields. *Scientometrics*, *125*(3), 2773–2792. https://doi.org/10.1007/s11192-020-03686-0.

Fernandes, J. M., & Monteiro, M. P. (2017). Evolution in the number of authors of computer science publications. *Scientometrics*, *110*(2), 529–539. https://doi.org/10.1007/s11192-016-2214-9.

Fox, C. W., Ritchey, J. P., & Paine, C. E. T. (2018). Patterns of authorship in ecology and evolution: First, last, and corresponding authorship vary with gender and geography. *Ecology and Evolution*, *8*(23), 11492–11507. https://doi.org/10.1002/ece3.4584.

Franceschet, M. (2011). Collaboration in computer science: A network science approach. *J. Assoc. Inf. Sci. Technol.*, *62*(10), 1992–2012. https://doi.org/10.1002/asi.21614.

Henriksen, D. (2019). Alphabetic or contributor author order: What is the norm in Danish economics and political science and why? *Journal of the Association for Information Science and Technology*, *70*(6), 607–618. https://doi.org/10.1002/asi.24151.

Kabacoff, R. (2020). *Data Visualization with R*. Quantitative Analysis Center: Wesleyan University.

Kennedy, D. (2003). Multiple authors, multiple problems. *Science*, *301*(5634), 733. https://doi.org/10.1126/science.301.5634.733.

Kim, J. (2018). Evaluating author name disambiguation for digital libraries: a case of DBLP. *Scientometrics*, *116*(3), 1867–1886. https://doi.org/10.1007/s11192-018-2824-5.

Kim, J. (2019). Author-based analysis of conference versus journal publication in computer science. *Journal of the Association for Information Science and Technology*, *70*(1), 71–82. https://doi.org/10.1002/asi.24079.

Kosmulski, M. (2012). The order in the lists of authors in multi-author papers revisited. *Journal of Informetrics*, *6*(4), 639–644. https://doi.org/10.1016/j.joi.2012.06.006.
Krasnova, H., Kummer, T., Schäfer, K., and Veltri, N. F. (2012). Let's collaborate, but I will be the first author! Exploring the importance of the first authorship for IS researchers. In *20th European Conference on Information Systems (ECIS 2012)*, page 144.
Ley, M. (2009). DBLP—Some lessons learned. *Proceedings of the VLDB Endowment*, *2*(2), 1493–1500. https://doi.org/10.14778/1687553.1687577.
Liang, L., Liu, J., & Rousseau, R. (2004). Name order patterns of graduate candidates and supervisors in Chinese publications: A case study of three major Chinese universities. *Scientometrics*, *61*(1), 3–18. https://doi.org/10.1023/B:SCIE.0000037359.22151.6f.
Liu, X. Z., & Fang, H. (2014). Scientific group leaders' authorship preferences: An empirical investigation. *Scientometrics*, *98*(2), 909–925. https://doi.org/10.1007/s11192-013-1083-8.
Maciejovsky, B., Budescu, D. V., & Ariely, D. (2008). The researcher as a consumer of scientific publications: How do name-ordering conventions affect inferences about contribution credits? *Marketing Science*, *28*(3), 222–227. https://doi.org/10.1287/mksc.1080.0406.
Marschke, G., Nunez, A., Weinberg, B. A., & Yu, H. (2018). Last place? The intersection of ethnicity, gender, and race in biomedical authorship. *AEA Papers and Proceedings*, *108*, 222–227. https://doi.org/10.1257/pandp.20181111.
Moore, M., & Griffin, B. W. (2006). Identification of factors that influence authorship name placement and decisions to collaborate in peer-reviewed, education-related publications. *Studies in Educational Evaluation*, *32*(3), 125–135. https://doi.org/10.1016/j.stueduc.2006.04.004.
Peidu, C. (2019). Can authors' position in the ascription be a measure of dominance? *Scientometrics*, *121*(3), 1527–1547. https://doi.org/10.1007/s11192-019-03254-1.
Reisenberg, D., & Lundberg, G. D. (1990). The order of authorship: Who's on first? *Journal of the American Medical Association*, *264*(14), 1857.
Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable. *Journal of the American Medical Association*, *278*(7), 579–585.
Solomon, J. (2009). Programmers, professors, and parasites: Credit and co-authorship in computer science. *Science and Engineering Ethics*, *15*(4), 467–489. https://doi.org/10.1007/s11948-009-9119-4.
Way, S. F., Morgan, A. C., Clauset, A., & Larremore, D. B. (2017). The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences*, *114*(44), E9216–E9223. https://doi.org/10.1073/pnas.1702121114.
Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position: A survey of perceived contributions to papers based on byline position and number of authors. *EMBO Reports*, *8*(11), 988–991. https://doi.org/10.1038/sj.embor.7401095.
Zuckerman, H. A. (1968). Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology*, *74*(3), 276–291.