Authorship trends in software engineering

João M. Fernandes

Abstract This paper aims to examine authorship trends in software engineering, especially those related to the number of authors, of scientific publications. We collected and mined around 70.000 entries from DBLP for 122 conferences and journals, for the period 1971–2012, in order to process several bibliometric indicators. We provide evidence that the number of authors of articles in software engineering is increasing on average around +0.40 authors/decade. The results also indicate that until 1980, the majority of the articles have a sole author, while nowadays articles with 3 or 4 authors represent almost half of the total.

Keywords Scientific authorship · Number of authors · Scientific publication · Software engineering

Mathematics Subject Classfication 68N01

Introduction

Authorship can be claimed by those contributing intellectually to the completion of the described research. In science, a published article is the primary means where new work is publicly made available. Career progression and academic promotion are greatly influenced by authorship. According to Bennett and Taylor (2003), the benefits of authorship are numerous and include (1) Contribution to the progress of science; (2) Personal sense of achievement; (3) Evidence of an individual's professional reputation; and (4) Creation of currency for academic appointment, promotion, research funding and entry to professional bodies.

J. M. Fernandes (✉)
Departamento de Informática/Centro Algoritmi, Universidade do Minho, Campus de Gualtar,
4710-057 Braga, Portugal
e-mail: jmf@di.uminho.pt

🌀 Springer

According to Greene (2007), until about 1920, sole authorship was the predominant tradition in science. As this manuscript shows, sole authorship was also the dominant situation in software engineering until 1990. However, there has been a trend in multiple authorship (Kennedy 2003) that is also taking place in the software engineering domain. Modern research is becoming very complex and specialised in many disciplines. In fact, many research topics require different skills and competencies in methodologies, analysis capabilities, computer tools, statistics and mathematics, as well as intense knowledge of a field. It becomes almost impossible for a single researcher to master all these specialised issues alone and consequently to conduct the work individually. Thus, most research work is nowadays developed collaboratively. Articles in some fields already exhibit a huge number of authors. For example, (The Atlas Collaboration 2008), an article in high-energy physics, is co-authored by 2,926 scientists from 169 research institutions. Another example is (The Gusto Investigators 1993), an article in medicine co-authored by 972 authors.

This study examines trends in authorship, especially those related to the number of authors of scientific articles in software engineering. We collected and processed almost 70.000 DBLP entries for 122 conferences and journals, in the period 1971–2012.

Method

This study includes articles published in conferences and journals, since both types are prestigious in software engineering. Nuseibeh (2011) considers that it is difficult to find significant distinctions between conference articles and journal articles. Patterson (2004) indicates that in computing it is common to prefer conferences instead of journals. Freyne et al. (2010) present quantitative evidence that articles in leading computing conferences match the impact of articles in mid-ranking journals and surpass the impact of articles in journals in the bottom half of the Thompson Reuters rankings. This contrasts with the prevailing academic tradition where the primary means of publishing is in journals (Vardi 2009).

We have decided to use articles listed in the DBLP website (Ley 2009), since it is specifically devoted to bibliographic information on major computer science journals and proceedings. DBLP includes the major software engineering conferences and journals. Additionally, DBLP can be accessed through automatic means, easing the download process of the information. We obtained the data for the study on May 17, 2013, by automatically downloading from the DBLP database the entries for the conferences and journals that we have chosen and that we believe to be closely related to the software engineering community. We hereafter designate this collection of bibliographic entries as DBLP-SWEng.

With a web crawler, 69,325 bibliographic entries, published from 1971 until 2012, in the software engineering domain were downloaded. All the entries were automatically processed and analysed with a Ruby program and the ones that were not related to scientific articles were withdrawn. This set 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 decided also to ignore the entries that have less than 4 pages. In this group, we typically can find editorials, invited papers, posters, tool papers, workshop summaries, and similar short contributions that are not regular scientific articles. Despite our efforts to make a comprehensive evaluation of the

processed information, it is possible that some entries were not correctly filtered, due to different designations or spellings; we believe that these situations are relatively small and should not affect the results in a significant measure. A total of 5,635 entries have been rejected, representing around 8.1 % of the original sample. So, DBLP-SWEng contains 63,690 entries related to software engineering articles.

In total, 21,881 journal articles (34.4 %) and 41,809 conference articles (65.6 %) were processed, whose annual distribution is depicted in Fig. 1. The number of articles until 1983 seems relatively small (less than 400 each year). There is a steady increase in the number of published articles, in accordance with the conclusions by Larsen and von Ins (2010). In total, the study includes 122 venues: 31 journals and 91 conferences. The complete list of journals and conferences is listed in the appendix. The 32 journals are from eight publishers: ACM, Cambridge University Press, Elsevier, IEEE, IET (formerly, IEE), Springer, Wiley, and World Scientific. Table 1 shows the number of considered journals from each of these publishers and the respective number of articles (and its percentage).

We present in this section various statistical analysis related to authors of the considered articles. We use the notions of new and active authors presented in Fig. 2 (top) shows the number of new authors (i.e., those who co-authored an article in DBLP-SWEng for the first time) every year. The software engineering community progressively grows every year. On average in the period 1982–2012, there are 150 more new authors in a given year than in the previous year. Since 2005, more than 3,000 new authors joined the community each year. Figure 2 (bottom) shows the cumulative new authors in each year. In total, 60,443 authors joined the software engineering community in the period 1971–2012.

We now examine the activity of the authors. An author is considered active in a given year if he/she publishes at least one article in that year. Figure 2 (top) also shows the number of active authors every year. It can be seen that since 1996, more than 3,000 authors were active each year. In particular, more than 9,000 authors were active in 2012.

Figure 3 shows the ratios between new authors and active authors in each considered year in the period 1971–2012. There is an evident tendency for these ratios to decrease along the time and since 1994, new authors represent around 50 % of the active authors in every year. Finally, Fig. 4 shows the average number of articles per active author in each year. They indicate the average productivity of the software engineer researcher. These values are obtained by dividing the number of articles published in a given year by the number of active authors in that same year. The values for this indicator in the last 3 decades are decreasing gradually and are located in the 0.4–0.6 range.

Results based on the number of authors

The total results (i.e., for both journal and conference articles) are shown in Table 2 and Fig. 5. More than half of the processed articles have either 1 or 2 authors. In particular, 1/3 of the articles have exactly 2 authors. If we consider articles up to 5 authors, the percentage is bigger than 96 %. Articles with 8 or more authors represent less than 1 % and articles with 20 or more authors represent a residual percentage of 0.0126 % (8 articles out of 63,690).

For science and engineering, the number of authors per article by field (with an author from the USA) are reported by NSF (2012). The average number of authors per article in all fields grew from 3.2 (1990) to 5.6 (2010). By 2010, among the 13 considered fields, computer science (in which software engineering is included) presents the third smallest
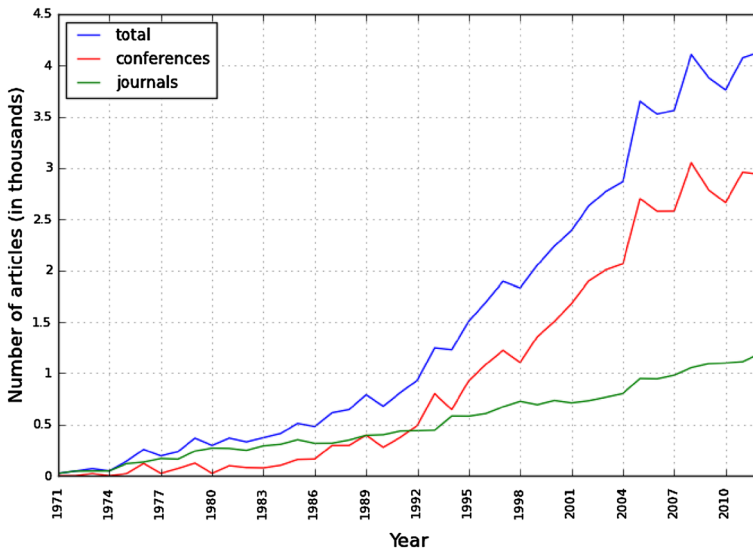
Fig. 1 Number of published articles, per year, in software engineering (1971–2012), for conferences and journals

Table 1 Journals considered in this study

| Publisher	#Journals	#Articles	%Articles
Elsevier	6	7,130	32.6
Springer	12	4,008	18.3
Wiley	4	3,198	14.6
IEEE	1	2,875	13.1
ACM	4	2,631	12.0
Cambridge	2	846	3.9
World scientific	1	787	3.6
IET	1	406	1.9
Totals	31	21,881	100.0

The columns are, from left to right, name of the publisher; number of journals for each publisher; number of published articles by each publisher; and percentage of the published articles by each publisher with respect to the total number of journal articles

average (3.3), just above mathematics (2.2) and social sciences (2.1). The ranking is leaded by astronomy (13.8) and physics (10.1).

Figure 6 depicts, for each year in the period 1971–2012, the distribution (in percentages) of the articles among the different number of authors. The synthesised results, based on the values for all years in the considered period, demonstrate that, until 1980, 1-author articles represent more (or slightly less) than 50 % of the total. By 1981, 1-author articles represent 41.3 %. By 1988, the percentage for 1-author articles (38.9 %) is for the first time less than 40 %. The year 1990 is the last one, where there are more 1-author articles than 2-author ones. The decline of the percentage for 1-author articles is continuous and by
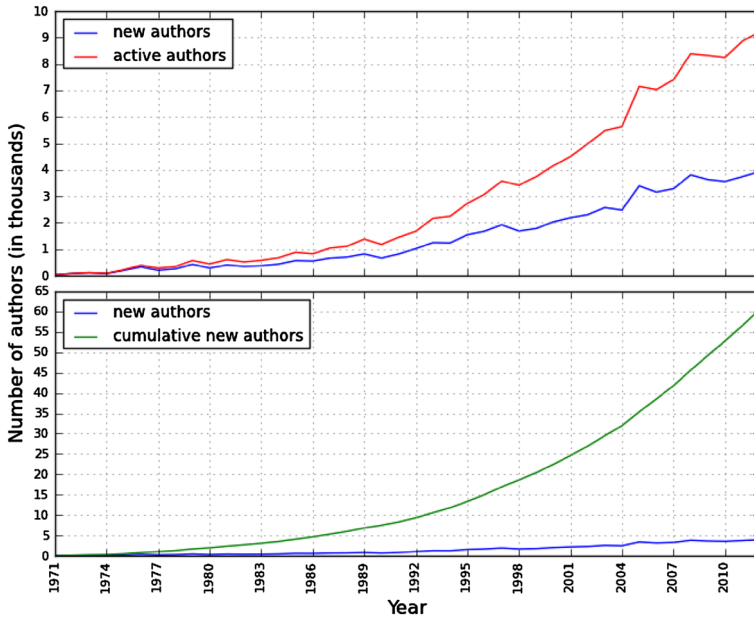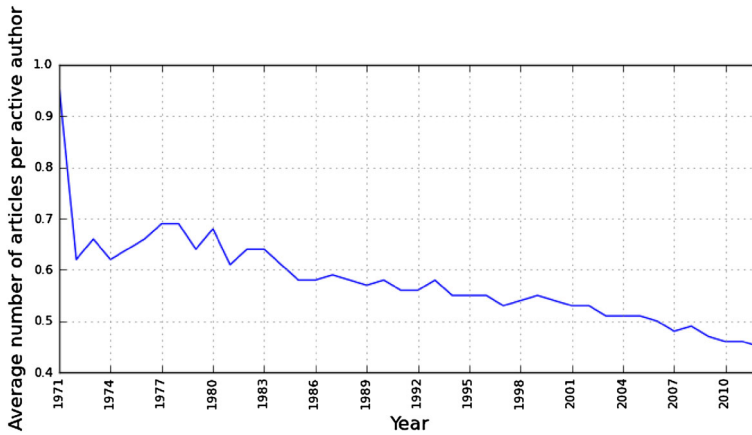
Fig. 2 New and active authors in software engineering (1971–2012), for conferences and journals; (top) new and active authors; (bottom) new authors and cumulative new authors
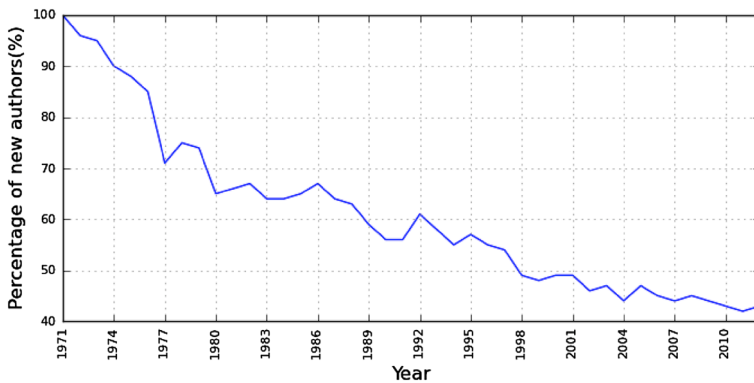
Fig. 3 Percentages of new authors with respect to the active authors in each year (1971–2012)

2001, 3-author articles supplanted them in number for the first time. Since 2009, 1-author articles represent less than 1/8 of the total for each year.

From 1993 until 2008, 2-author articles dominate the scene, with percentages between 30 and 40 %. The percentage of 2-author articles grows until 1994 (reaching a value of 40.1 %). Since then, that value shows a general tendency to decrease, with a few exceptions. The first year where the dominance belongs to 3-author articles (with 29.9 %) is 2009. However, the percentage of 3-author articles is decreasing since then, so it is expected that the dominance will belong to 4-author articles in a few years. In 2012,

**Fig. 4** Average number of articles per active author in each year (1971–2012)

Table 2 Percentages of articles in software engineering conferences and journals for different number of authors (1971–2012)

| #Authors	#Articles	%	Cumul. %
1	13,519	21.2	21.2
2	20,936	32.9	54.1
3	15,768	24.8	78.9
4	8,019	12.6	91.5
5	3,271	5.1	96.6
6	1,227	1.9	98.5
7+	950	1.5	100.0
Total	63,690	100	

The columns are, from left to right, number of authors; number of articles considered in this study for each number of authors; percentage of the articles for each number of authors with respect to the total number articles; and the cumulative percentage of the articles (i.e., the percentage of the articles with at most the number of authors in that row)

4-author articles represent already 20.5 % of the total. This analysis shows that there is a clear tendency for a growth in the number of authors of articles in software engineering.

We have also calculated the average number of authors for articles in software engineering, in a yearly basis. Figure 7(top) shows the results for this metric, which grows very smoothly along the timeline. For instance in 1971, the 92.3 % of the 26 considered articles have just 1 author. The average number of authors is 1.08. In 2012, the value is 3.23. Based on the Ordinary Least Squares method, we can say that from 1971 until 2012, there is an average growth of +0.40 authors in every decade, as given by the linear approximation depicted in Fig. 4a. Stated in a different way, we can say that on average, an additional author is added to all articles every 25 years. However, if the analysis is focused on the last two decades, the growth is around +0.60 authors/decade: +0.524 from 1992 to 2002, and +0.664 from 2002 to 2012). This rate implies that only 17 years are needed to have an additional author is all articles.
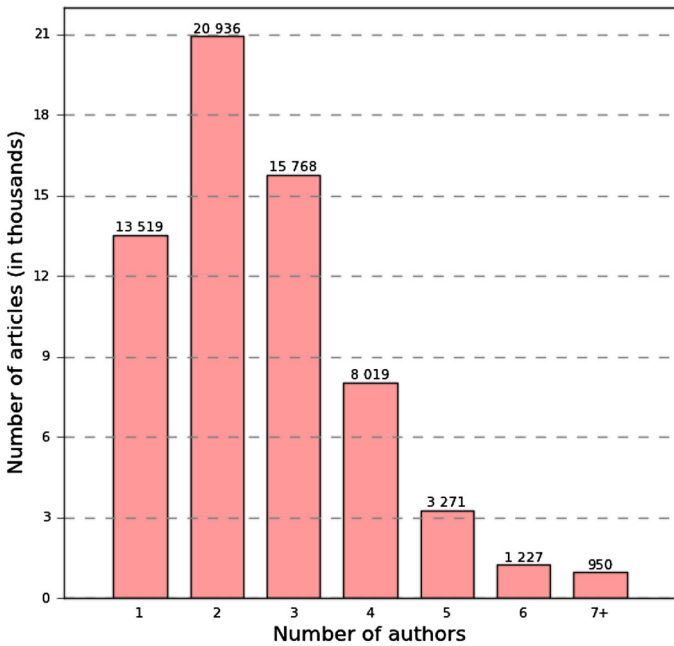
Fig. 5 Number of articles (in thousands) for different number (1, 2, 3, 4, 5, 6 and 7 +) of authors, in software engineering (1971–2012)
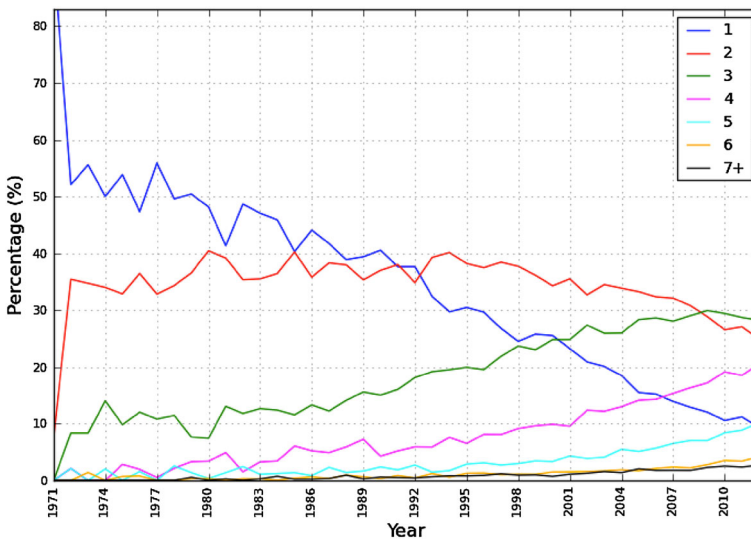
Fig. 6 Percentages of articles with different number of authors, per year, in software engineering (1971–2012)

We also observe, from Fig. 7b, that the maximum number of authors, per year, tends to grow, although the tendency is a bit erratic. From 1992, this value is constantly above 10, which shows a recent tendency for articles to have tens of authors. The article written by Moreira et al. (2007) lists 46 co-authors, which is the maximum for our sample.
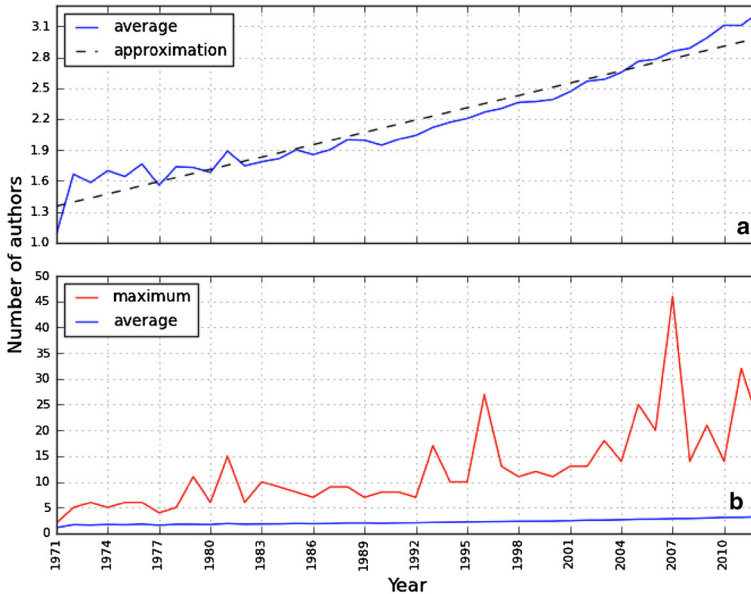
Fig. 7 **a** Average number of authors for articles, per year, in software engineering (1971–2012). The *blue line* shows the actual values calculated for each year, while the *dashed black line* shows the linear approximation for the trend, which provides a growth of +0.40 authors/decade. **b** Average and maximum number of authors for articles, per year, in software engineering (1971–2012)

The results, shown in Table 3, for conference articles and journal articles, considered separately, are similar to the overall results but with some relevant differences. There is a significant difference in terms of the percentages for 1- and 2-author articles. Articles with 1 or 2 authors represent a little more than 50 % for conferences, but are 59 % for journals. On average, conference articles have more authors than journal articles. For the full period (1971–2012), the average number of authors for conference articles is 2.68, while for journal articles it is 2.48. So, the difference is as short as 0.20.

Results based on the names of authors

We have computed some figures based on individual authors. Please note that DBLP does not have the notion of unique key for differentiating the authors. This means that authors are distinguished by their names. Mistakes are thus possible, because some authors do not always use the same spellings and because different authors may have the same name. No special measure, like the one proposed by Newman (2004), was taken to overcome this problem, since the differences in the results are not significant. Anyway, the results shown in this section must be taken with caution, because they may not be totally correct.

The average number of articles per author is 2.73. This is a low number that is mainly due to the fact that 37,351 authors (out of 61,032, i.e., 61 %) have (co-)authored just one article. Figure 8 depicts a histogram that shows the number of authors that have precisely

Table 3 Percentages of articles in software engineering for different number of authors, for conferences and journals, considered separately (1971–2012)

#Authors	Conferences			Journals		
	#Articles	%	Cumul. %	#Articles	%	Cumul. %
1	8,253	19.8	19.8	5,266	24.1	24.1
2	13,263	31.7	51.5	7,673	35.0	59.1
3	10,809	25.8	77.3	4,959	22.7	81.8
4	5,616	13.4	90.7	2,403	11.0	92.8
5	2,316	5.6	96.3	955	4.3	97.1
6	893	2.1	98.4	334	1.6	98.7
7+	659	1.6	100.0	291	1.3	100.0
Total	41,809	100	–	23,461	100	–

The columns are, from left to right, number of authors; number of considered conference articles for each number of authors; percentage of the conference articles for each number of authors with respect to the total number conference articles; and the accumulated percentage of the conference articles (i.e. the percentage of the articles with at most the specific number of authors). Similarly, the next three columns are for journal articles

co-authored a specific number of articles. It shows how many authors have authored just one article, how many have authored two articles, and so on (until 150 articles).

In total, 90 % of the authors have co-authored less than 6 articles. Our results are in accordance with Lotka's law that states that there are many authors who only publish a small number of articles and a small number of authors who publish a large number of articles. According to this law, the number of authors writing n articles is about n^{-a} of those writing one article, where a nearly always equals two. For DBLP-SWEng (1971–2012), the a value equals 2.045.

The average number of co-authors considering all authors in the full period of this study is 4.58. If we restrict our analysis to each year individually, the figures are obviously smaller, as is depicted in Fig. 9. However, we can see that the number of collaborators for all active authors in a given year is gradually growing from 0.148 in 1971 up to 3.751 in 2012. Based on the Ordinary Least Squares method, we can say that from 1971 until 2012, there is an average growth of +0.69 co-authors in every decade, as given by the linear approximation depicted in Fig. 9. Stated differently, on average, an additional co-author is added to all articles every 14.5 years. This clearly shows that in general researchers of the software engineering field are collaborating with more co-authors.

In Tables 4 and 5, we show four different tops for DBLP-SWEng authors, based on different metrics that are calculated by different methods. The list on the left hand-side in Table 4 shows the authors with the most number of co-authors. In this list we can find some authors with a relatively low number of co-authored articles (e.g., Julio Cesar Sampaio do Prado Leite and Uirá Kulesza), because many of their articles include lists of 8, 10, 11 and even 32 co-authors (Chavez et al. 2011). The right hand-side list in Table 4 shows the DBLP-SWEng authors that have co-authored more articles. It ranks the authors according to the inflated co-author credit, assigning 1 unit of authorship credit to every co-author.

The two lists in Table 5 show the authors that have contributed the most to the literature of software engineering, by using the fractional and harmonic authorship credit. In these two procedures, the "value" of each article is always 1, which means that is
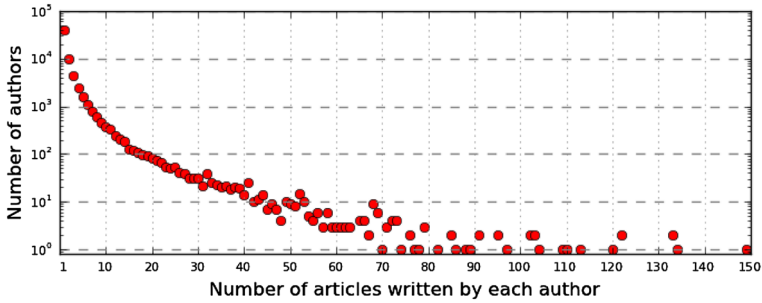
Fig. 8 Histogram of the number of authors (software engineering researchers) with a given number of authored articles (1971–2012). The y-axis uses a logarithmic scale
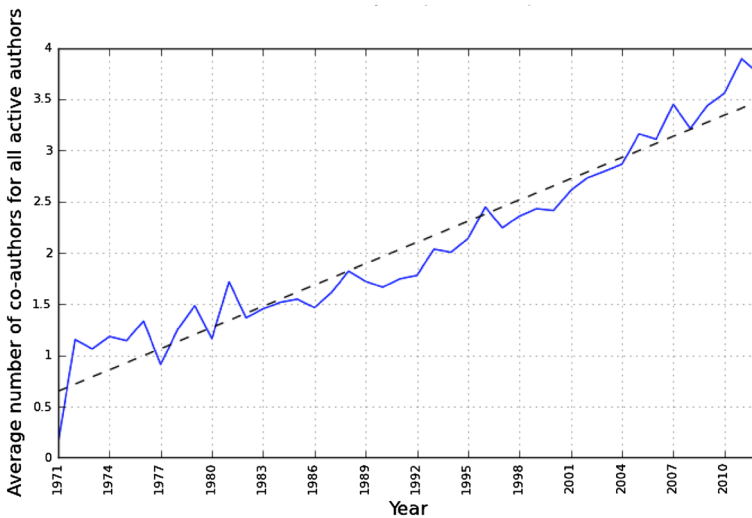
Fig. 9 Average number of co-authors for all active authors in each year (1971–2012)

independent of the number of co-authors. To quantify the credits attributed to each co-author, a formula is needed. In the fractional credit (Hagen 2014), multiple co-author articles are counted fractionally according to (the inverse of) the number of authors. So, if an article has N co-authors, each one is credited 1/N. This is obviously only a rough approximation: in reality co-authors of an article do not contribute equally. However, in the absence of other data, it is the obvious first approximation to make. This approach is in accordance with the method proposed by Ren and Taylor (2007). We can see in the left hand side list of Table 5, for example, that Bertrand Meyer is third in this top, because he has written many articles alone and when he collaborates the number of co-authors is relatively low (typically 3 or 4). Norman F. Schneidewind is also well positioned (11th) despite the fact that he has "just" 45 articles. However, he is the only author in 32 of those articles.

In the harmonic authorship credit, the credit for the ith author of a publication with N co-authors is calculated according to the following formula (Hagen 2014):

Table 4 The authors with the highest number of co-authors and the highest number of co-authored articles

	Author	**CoA**	Infl	Frac	Harm	Author	**Infl**	Frac	Harm	CoA
1	Victor R. Basili	**146**	133	52.97	51.46	Lionel C. Briand	**149**	52.28	61.42	111
2	Barry W. Boehm	**127**	110	47.53	48.88	Mark Harman	**134**	43.65	43.45	105
3	Hong Mei	**125**	122	31.79	19.87	Victor R. Basili	**133**	52.97	51.46	146
4	John Mylopoulos	**121**	102	35.07	25.66	Giuliano Antoniol	**133**	36.67	37.97	106
5	Lionel C. Briand	**111**	149	52.28	61.42	Mario Piattini	**122**	34.83	21.59	108
6	Carlos José Pereira de Lucena	**109**	86	23.08	15.87	Hong Mei	**122**	31.79	19.87	125
7	Tao Xie	**109**	89	29.03	24.52	Andrea De Lucia	**120**	31.93	38.48	73
8	Mario Piattini	**108**	122	34.83	21.59	Paolo Tonella	**113**	40.13	35.18	70
9	Giuliano Antoniol	**106**	133	36.67	37.97	Barry W. Boehm	**110**	47.53	48.86	127
10	Mark Harman	**105**	134	43.65	43.45	Mary Jean Harrold	**109**	41.61	33.78	80
11	Gregg Rothermel	**99**	102	34.18	28.74	Tsong Yueh Chen	**104**	35.87	40.39	62
12	Michael D. Ernst	**94**	69	22.87	17.48	Claes Wohlin	**103**	47.84	41.64	62
13	Tim Menzies	**93**	66	24.01	26.36	Ahmed E. Hassan	**103**	32.90	26.52	82
14	Don S. Batory	**93**	70	26.93	26.60	John Mylopoulos	**102**	35.07	25.66	121
15	José Carlos Maldonado	**92**	58	14.85	10.05	Gregg Rothermel	**102**	34.18	28.74	99
16	Reidar Conradi	**90**	85	33.06	31.52	Thomas W. Reps	**97**	39.37	34.17	68
17	Julio Cesar Sampaio do Prado Leite	**87**	30	9.81	8.60	Gerardo Canfora	**95**	29.52	41.57	60
18	Robert B. France	**87**	62	19.35	19.35	Massimiliano Di Penta	**95**	28.01	25.37	83
19	Richard F. Paige	**87**	66	23.86	21.83	Bertrand Meyer	**91**	50.07	44.31	78
20	Hongji Yang	**87**	76	23.00	17.79	Stéphane Ducasse	**91**	29.34	27.89	77
21	Katsuro Inoue	**86**	73	18.78	12.81	Tao Xie	**89**	29.03	24.52	109
22	W. Eric Wong	**85**	77	25.89	27.27	Taghi M. Khoshgoftaar	**88**	33.55	41.60	51
23	Uirá Kulesza	**84**	35	6.51	5.97	Carlos José Pereira de Lucena	**86**	23.08	15.87	109
24	Qing Wang	**83**	69	16.85	13.92	Reidar Conradi	**85**	33.06	31.52	90
25	Jean-Marc Jézéquel	**83**	79	25.75	19.17	Arie van Deursen	**85**	29.91	27.63	72
26	Massimiliano Di Penta	**83**	95	28.01	25.37	Bernhard Steffen	**82**	30.92	25.09	66
27	William C. Chu	**82**	60	16.93	15.92	Brian Henderson-Sellers	**79**	42.20	41.34	67
28	Ahmed E. Hassan	**82**	103	32.90	26.52	Rajiv Gupta	**79**	32.70	28.85	39
29	Laurie J. Hendren	**81**	67	23.48	16.99	Jean-Marc Jézéquel	**79**	25.75	19.17	83
30	Mary Jean Harrold	**80**	109	41.61	33.78	Jin Song Dong	**78**	24.72	25.02	64

The CoA columns denote the number of co-authors for the corresponding author. The Infl columns indicate the inflated credit, which is the same as the number of articles co-authored by the respective author. The Frac and Harm columns show respectively the fractional and the harmonic credit of each author

Table 5 The authors with the highest contributions to the software engineering literature based on the fractional and harmonic authorship credits

	Author	**Frac**	Harm	Infl	CoA	Author	**Harm**	Frac	Infl	CoA
1	Victor R. Basili	**52.97**	51.46	133	146	Lionel C. Briand	**61.42**	52.28	149	111
2	Lionel C. Briand	**52.28**	61.42	149	111	Victor R. Basili	**51.46**	52.97	133	146
3	Bertrand Meyer	**50.07**	44.31	91	78	Barry W. Boehm	**48.86**	47.53	110	127
4	Claes Wohlin	**47.84**	41.64	103	62	Harry M. Sneed	**47.64**	47.00	59	18
5	Barry W. Boehm	**47.53**	48.86	110	127	Bertrand Meyer	**44.31**	50.07	91	78
6	Harry M. Sneed	**47.00**	47.64	59	18	Mark Harman	**43.45**	43.65	134	105
7	Mark Harman	**43.65**	43.45	134	105	Claes Wohlin	**41.64**	47.84	103	62
8	Brian Henderson-Sellers	**42.20**	41.34	79	67	Taghi M. Khoshgoftaar	**41.60**	33.55	88	51
9	Mary Jean Harrold	**41.61**	33.78	109	80	Gerardo Canfora	**41.57**	29.52	95	60
10	Paolo Tonella	**40.13**	35.18	113	70	Brian Henderson-Sellers	**41.34**	42.20	79	67
11	Norman F. Schneidewind	**40.12**	40.64	45	26	Norman F. Schneidewind	**40.64**	40.12	45	26
12	Jan Bosch	**39.90**	34.60	73	52	Tsong Yueh Chen	**40.39**	35.87	104	62
13	Thomas W. Reps	**39.37**	34.17	97	68	Andrea De Lucia	**38.48**	31.93	120	73
14	Simon L. Peyton Jones	**38.64**	37.32	61	60	Robert L. Glass	**38.08**	37.58	46	11
15	Robert L. Glass	**37.58**	38.08	46	11	Giuliano Antoniol	**37.97**	36.67	133	106
16	Giuliano Antoniol	**36.67**	37.97	133	106	Simon L. Peyton Jones	**37.32**	38.64	61	60
17	Tsong Yueh Chen	**35.87**	40.39	104	62	Paolo Tonella	**35.18**	40.13	113	70
18	John Mylopoulos	**35.07**	25.66	102	121	Jan Bosch	**34.60**	39.90	73	52
19	Mario Piattini	**34.83**	21.59	122	108	Thomas W. Reps	**34.17**	39.37	97	68
20	Steven P. Reiss	**34.21**	33.74	46	25	Mary Jean Harrold	**33.78**	41.61	109	80
21	Gregg Rothermel	**34.18**	28.74	102	99	Steven P. Reiss	**33.74**	34.21	46	25
22	Jens Palsberg	**33.67**	31.44	69	54	Mira Kajko-Mattsson	**31.96**	29.82	52	60
23	Taghi M. Khoshgoftaar	**33.55**	41.60	88	51	Reidar Conradi	**31.52**	33.06	85	90
24	Reidar Conradi	**33.06**	31.52	85	90	Jens Palsberg	**31.44**	33.67	69	54
25	Ahmed E. Hassan	**32.90**	26.52	103	82	Robert M. Hierons	**30.21**	30.29	71	59
26	Rajiv Gupta	**32.70**	28.85	79	39	Hassan Gomaa	**30.03**	30.12	53	41
27	Elaine J. Weyuker	**32.33**	28.28	58	22	Martín Abadi	**29.32**	22.87	52	42
28	Andrea De Lucia	**31.93**	38.48	120	73	Alexander Egyed	**29.23**	31.65	68	67
29	Hong Mei	**31.79**	19.87	122	125	David Binkley	**29.13**	26.05	72	58
30	Alexander Egyed	**31.65**	29.23	68	67	Rajiv Gupta	**28.85**	32.70	79	39

The Frac and Harm columns show respectively the fractional and the harmonic credit of each author. The Infl columns indicate the inflated credit, which is the same as the number of articles co-authored by the respective author. The CoA columns denote the number of co-authors for the corresponding author

$$\text{Harmonic } i\text{th author credit} = \frac{\frac{1}{i}}{1 + \frac{1}{2} + \cdots + \frac{1}{N}}$$

The result for harmonic authorship credit is shown in the right hand side list of Table 5. Although it differs from the result for the fractional formula, 24 authors (80 %) are listed in both top-30 lists. It is not clear which formula is more adequate for software engineering, since different practices are used to indicate the co-authors in the byline. We hope to be able to investigate this issue in the near future.

Solo authors (i.e., those that have exclusively published articles alone) are 3,320 in the full period. This figure represents 5.1 % of the total number of authors in DBLP-SWEng. A vast majority (86 %) of the solo authors have published just one article. In total, solo authors have published 3,979 articles, a figure that represents 6.2 % of the total.

Conclusions

This manuscript shows that the number of authors of scientific articles in the software engineering domain is increasing. This situation is not unexpected, since similar observations exist in other domains (Abt 1981; NSF 2012; Zetterström 2004). From a situation in software engineering, where 1-author articles were more than 50 % of the total during 15 years, 3- and 4-author articles dominate nowadays.

Since the average number of authors of scientific articles is increasing, the system of authorship is consequently becoming inappropriate, in the sense that it becomes more difficult to credit all the authors for the specific contributions they made to each article. Therefore, the community must establish an agreed publishing standard to define how to assign the academic contribution to all collaborators of a research project, as proposed by Solomon (2009). How one can measure the effective contribution of a given co-author in a 20-author article? To answer questions like this one, proposals similar to the one by Marcos et al. (2012), which suggests articles to be explicit about the role/contribution of each involved person, need to be debated.

The notion of author is becoming useless, since many unacceptable and inappropriate forms of authorship exist: honorary, gift, guest, ghost, and coercive (Gasparyan et al. 2013). An author is someone who has made substantive intellectual contributions to a study and is responsible for a component of the work (Greenland and Fontanarosa 2012). The scientific community needs the more realistic notion of contributor (Rennie et al. 1997). The observed growth in the number of authors will imply in the near future new metrics for the productivity of both researchers and institutions, as they tend to work more in cooperation. For instance, some recent proposals suggest new forms of assigning citations to co-authors (Ausloos 2013; Liu and Fang 2012; Zhang 2009).

In our opinion, the research community needs to be aware of the trends and changes in academic authorship. Articles about authorship in software engineering are not common, with some exceptions like Ren and Taylor (2007) or the annual publication-based assessment of scholars and institutions that is published in the Journal of Systems and Software (JSS) since 1994 (Wong et al. 2011). Some of the results of this study differ with respect to the ones reported, for example, by Cohoon et al. (2011) that indicate a small increase in the number of articles published by lone authors at ACM-sponsored conferences for the computing domain. They also report that by 2008, 97 % of all conference articles have 2 or more authors, which substantially differs from our results (87.1 %). More studies about academic authorship are definitely needed to provide a clear view of the reality. In particular, studies that provide information for the different subdomains of computing are of paramount importance, to allow a fair comparison of researchers in the computing field.

The bibliographic data from DBLP used in this article and the full rankings of software engineering authors can be downloaded from www.di.uminho.pt/∼jmf/DBLP-SWEng.

Appendix

List of journals

ACM: ACM J Exp Algorith, ACM T Math Software, ACM T Progr Lang Sys, ACM T Softw Eng Meth; *Cambridge*: J Funct Program, Theor Pract Log Prog; *Elsevier*: Adv Eng Softw, Comp Lang Syst Struct, Inform Soft Tech, J Syst Soft, Sci Comput Program, J Vis Lang Comput; *IEEE*: IEEE Trans Soft Eng; *IET*: IET Softw; *Springer*: Ann Soft Eng, Autom Soft Eng, Constraints, Empir Soft Eng, Formal Asp Comput, Int J Par Progr, Innov Syst Soft Eng, Req Eng J, Soft Syst Model, Software Qual J, Soft Concepts Tools/Struct Program, Soft Tools Tech Transf, TAOSD; *Wiley*: J Soft Maint, Soft Focus, Soft Pract Exper, Softw Test Verif Rel; *World Scientific*: Int J Soft Eng Knowl Eng.

List of conferences

AGILE, AOSD, APSEC, ASE, ASEC, CBSE, CC, COMPSAC, COORDINATION, CP, CSEET, CSMR, DBPL, DSVIS, ECMFA, ECOOP, EDOC, EMSOFT, Ershov, ESEC, ESEM, ESOP, ESSoS, EUROMICRO-SEAA, EWSPT, FASE, FATES, FM, FMCO, FORMATS, FoSSaCS, FSE, GPCE, ICCBSS, ICECCS, ICFEM, ICGSE, ICMT, ICPC, ICPE, ICSE, ICSM, ICSP, ICSR, ICST, IFM, ISESE, ISSRE, ISSTA, IWPSE, IWSSD, METRICS, MoDELS, MSR, OOIS, OOPSLA, PASTE, PLDI, POPL, PPOPP, PROFES, QEST, QSIC, RE, REFSQ, RTA, SAS, SBES, SC, SCAM, SCM, SEFM, SEKE, SEL-MAS, SEW, SLE, SOFTVIS, SPIN, SPLC, SSR, TACAS, TAPSOFT, TOOLS, VEE, VMCAI, WCRE, WICSA, WIKIS, XP.

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.

Ausloos, M. (2013). A scientometrics law about co-authors and their ranking: The co-author core. *Scientometrics, 95*(3), 895–909. doi:10.1007/s11192-012-0936-x.

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

Chavez, C., Kulesza, U., Soares, S., Borba, P., Lucena, C. J. P., Masiero, P. C., et al. (2011). The AOSD research community in Brazil and its crosscutting impact. *SBES, 2011*, 72–81.

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.

Freyne, J., Coyle, L., Smyth, B., & Cunningham, 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.

Gasparyan, A. Y., Ayvazyan, L., & Kitas, G. D. (2013). Authorship problems in scholarly journals: Considerations for authors, peer reviewers and editors. *Rheumatology International, 33*(2), 277–284. doi:10.1007/s00296-012-2582-2.

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

Greenland, P., & Fontanarosa, P. B. (2012). Ending honorary authorship. *Science, 337*(6098), 1019. doi:10.1126/science.1224988.

Hagen, N. H. (2014). Counting and comparing publication output with and without equalizing and inflationary bias. *Scientometrics, 8*(2), 310–317. doi:10.1016/j.joi.2014.01.003.

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

Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*(3), 575–603. doi:10.1007/s11192-010-0202-z.

Ley, M. 2009. DBLP: Some lessons learned. In *Proceedings of the VLDB endowment (PVLDB 2009)*, (Vol. 2, No. 2, pp. 1493–1500).

Liu, X. Z., & Fang, H. (2012). Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index. *Scientometrics, 91*(1), 37–49. doi:10.1007/s11192-011-0571-y.

Marcos, E., Vara, J. M., & Castro, V. (2012). Author order: What science can learn from the arts. *Communications of the ACM, 55*(9), 39–41. doi:10.1145/2330667.2330680.

Moreira, J. E., Salapura, V., Almási, G., Archer, C., Bellofatto, R., Bergner, P., et al. (2007). The Blue Gene/L supercomputer: A hardware and software story. *International Journal of Parallel Programming, 35*(3), 181–206. doi:10.1007/s10766-007-0037-2.

National Science Foundation, 2012. Science and Engineering Indicators 2012, chap. 5.

Newman, M. E. J. 2004. Who is the best connected scientist? A study of scientific coauthorship networks. In Ben-Naim, E., Frauenfelder, H., Toroczkai, Z. (Eds.), Complex networks, Lecture notes in physics. Springer, (Vol. 650, pp. 337–370). doi:10.1007/978-3-540-44485-5_16.

Nuseibeh, B. (2011). Editorial: What makes a publication archival? *IEEE Transactions on Software Engineering, 37*(2), 145. doi:10.1109/TSE.2011.34.

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.

Ren, J., & Taylor, R. N. (2007). Automatic and versatile publications ranking for research institutions and scholars. *Communications of the ACM, 50*(6), 81–85. doi:10.1145/1247001.1247010.

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable. *JAMA, 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. doi:10.1007/s11948-009-9119-4.

The Atlas Collaboration. (2008). The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation, 3*(8), S08003. doi:10.1088/1748-0221/3/08/S08003.

The Gusto Investigators. (1993). An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *The New England Journal of Medicine, 329*(10), 673–682. doi:10.1056/NEJM199309023291001.

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

Wong, W. E., Tse, T. H., Glass, R. L., Basili, V. R., & Chen, T. Y. (2011). An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008). *Journal of Systems and Software, 84*(1), 162–168. doi:10.1016/j.jss.2010.09.036.

Zetterström, R. (2004). The number of authors of scientific publications. *Acta Paediatrica, 93*(5), 581–582. doi:10.1111/j.1651-2227.2004.tb02980.x.

Zhang, C. T. (2009). A proposal for calculating weighted citations based on author rank. *EMBO Reports, 10*(5), 416–417. doi:10.1038/embor.2009.74.