A Closer Look into Collaborative Publishing at Software-Engineering Conferences

Rand Alchokr[1][0000−0003−0112−5430], Jacob Krüger[2][0000−0002−0283−248X], Yusra Shakeel[1,3][0000−0001−5135−4325], Gunter Saake[1][0000−0001−9576−8474], and Thomas Leich[4][0000−0001−9580−7728]

[1] Otto-von-Guericke University, Magdeburg, Germany
{rand.alchokr,saake}@ovgu.de
[2] Ruhr-University, Bochum, Germany
jacob.krueger@rub.de
[3] Karlsruhe Institute of Technology, Karlsruhe, Germany
yusra.shakeel@kit.edu
[4] Harz University & METOP GmbH, Wernigerode, Germany
tleich@hs-harz.de

Abstract. Computer science and particularly software engineering is a rapidly evolving research discipline increasingly conducted by large, collaborative teams. Unfortunatly, there is little research on the underlying publication activity and collaboration patterns in software engineering. To address this gap, we study two properties of research collaborations in software engineering: the number of collaborators (i.e., authors of a paper) and their academic age (i.e., their experience of working in research). More precisely, we investigate collaborations for papers published at all main tracks of three top-level software-engineering conferences (i.e., ASE, ESEC/FSE, ICSE) and one top-level reference conference (i.e., JCDL), including a total of 5,188 papers and the corresponding 8,730 unique authors. Our results indicate that collaboration is more prevalent now than ever before, with a decline in the proportion of researchers who contribute single-author papers. Moreover, our analysis revealed that the ideal team size seems to range from two to four researchers, and that junior researchers seem to need the support of more experienced co-authors to get published at such top-level conferences. Ultimately, our goal is to understand how collaborations in software engineering have evolved and impact different researchers (e.g., newcomers, juniors), helping to highlight potential impediments and consequent improvements regarding the quality of research, collaborations, and mentoring.

Keywords: Software engineering · Publications · Scientific collaboration · Junior researchers

1 Introduction

Collaboration is key in research to cope with the complex nature and rapidly evolving corpus of scientific work—specifically in computer-science-related disciplines with their high pace of advancements. Consequently, there has been an

increased emphasis on collaboration as a tool of science [20]. Scientific collaboration refers to a number of individuals (e.g., researchers, students, practitioners) working together on a research problem that leads to a co-authored research paper. Collaboration is a complex task that depends on the involved researchers' attitude towards it and involves numerous social factors that may impede or facilitate its cooperative aspects. Most studies build on the underlying assumption that collaborative activity increases research productivity [11, 25]. Still, other studies revealed contribution challenges that certain groups of researchers face, for instance, new researchers (e.g., juniors, newcomers) [1, 3, 9, 19]. As a consequence, it is particularly important to understand how different groups of researchers are involved in and impacted by scientific collaborations.

For our work, we consider the activity of writing and publishing papers to reflect on the interaction between researchers. In this paper, we take a step towards understanding how junior researchers are involved in scientific collaborations by quantitatively analyzing their publishing activity, co-authors, and evolutionary patterns. For this purpose, we elicited data from digital libraries with *the goal of understanding scientific collaborations' impact on publication productivity inside the software-engineering (SE) community over time with a focus on the involvement of junior researchers.* To address this goal, we defined two research questions (RQs):

RQ$_1$ *Are collaboration patterns in software-engineering stable over time?*
RQ$_2$ *What are frequent collaboration patterns for software-engineering juniors?*

Precisely, we tracked the collaborations of authors at the following three top-level software-engineering conferences (1, 2, 3) and one reference conference (4), each of which involves more junior researchers as active participants over the past years: 1) IEEE/ACM International Conference on Automated Software Engineering (ASE), 2) ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), 3) IEEE/ACM International Conference on Software Engineering (ICSE), and 4) ACM/IEEE Joint Conference on Digital Libraries (JCDL). These conferences have a high reputation, which is why most researchers of any academic age and reputation aim to publish at them.

We report the results of our quantitative analysis involving 5,188 main-track papers written by 8,730 authors over a time period of 43 years, from 1975 to 2020. Our complete dataset is available as an open-access repository.[5] Unfortunately, it is not possible to study the papers that were rejected at each conference, which is why we have to be careful with interpreting our results since they build only on accepted and published papers. Still, the results of our analysis reveal important insights concerning SE research, collaboration patterns, and the involvement of junior researchers. Our findings uncover a changing trend in each of the selected conferences from papers with few or single authors towards multi-authored papers, indicating an ideal team size of two to four researchers. The analysis also indicates juniors' needs for collaboration to be successful in getting their papers accepted at such conferences. Ultimately, we hope that the

[5] https://doi.org/10.5281/zenodo.6824306

results of our study provide a better understanding of scientific collaboration and serve as a foundation for further research to understand whether certain groups of researchers are over- or underrepresented.

2 Background and Related Work

Collaboration is key for advancing science and creating new knowledge; generally defined by two elements: working together for a common goal and sharing knowledge [12]. Researchers tend to collaborate due to different factors, such as their different specializations and the growth of interdisciplinary fields [4]. Their different skills and expertise benefit the development of research projects and corresponding papers, which is why multi-authored papers are common in science [4,12]. Eventually, research on scientific collaborations has aimed to understand collaboration patterns to better comprehend the scientific process [17,18]. For instance, Costas and Bordons [8] present results regarding collaborations between members from more than one group. The results indicate a constantly increasing number and frequency of collaborations over time. Related bibliometrics study [10,13] focus on authorship trends in software engineering, and found that the number of authors is increasing on average with around 0.40 authors/decade until 1980. Other studies concentrate on more specific collaboration aspect, such as co-authorships (e.g., author order in multi-authored papers), that are critical components for successful collaborations.

Since most research teams comprise both early and later career scientists, studies concerning these two groups have been conducted. For example, Zhou et al. [24] found that newcomers tend to collaborate more with existing group members than with other newcomers to gain more experience and reputation. Similarly, juniors (i.e., researchers that have worked up to three years in academia [1, 14]) were the focal point of several studies due to their essential role in providing innovative ideas, and broadening the scope of collaborations [16] as well as their high motivation that can inspire others and improve the work atmosphere [1,2,3]. While most of such studies report interesting findings, the datasets used are old and they do not investigate collaboration over time, specifically between authors with different levels of expertise. Out study in this paper fills this gap by extracting recent papers and analyzing evolution trends.

3 Methodology

We extracted data for four top-level conferences, because computer-science (and particularly software-engineering) research is generally more focused on those instead of journals [7, 15, 22]. Namely, we studied the main research tracks of these conferences from their first edition (in parentheses) until 2020, with three software-engineering ones (ASE 1991; ESEC/FSE 1987; ICSE 1976) and one partially software-engineering related (JCDL 2001). We chose JCDL because it is more general than the other conferences, and thus serves as a reference to comparing software engineering to other computer-science fields.

Table 1: Overview of our dataset.

conference	period	# papers	# authors	# unique authors
ASE	1991–2020	1,069	3,740	2,482
ESEC/FSE	1987–2020	1,239	4,264	2,614
ICSE	1976–2020	2,300	7,434	4,380
JCDL	2001–2020	580	2,087	1,393
total		5,188	17,525	8,730

For our analysis, we collected data by automatically crawling *dblp*,[6] which provides structured bibliographic data and distinguishes authors with identical names. To improve the quality of our data, we studied only main research-track papers and and manually compared the session information in dblp to official information in the ACM Digital Library[7] to identify mislabeled papers. However, some older conferences did not clearly label their papers, which us why we excluded a paper if it comprised fewer than seven pages. Overall, our analysis resulted in a total of 5,188 main-track papers, which we summarize in Table 1. Note that the total number of unique authors (8,730) is not the sum of the last column, since we counted each author only once across all conferences. For each extracted paper, we additionally crawled and extracted the corresponding authors' bibliographic data from dblp.

4 RQ_1: Collaboration Patterns

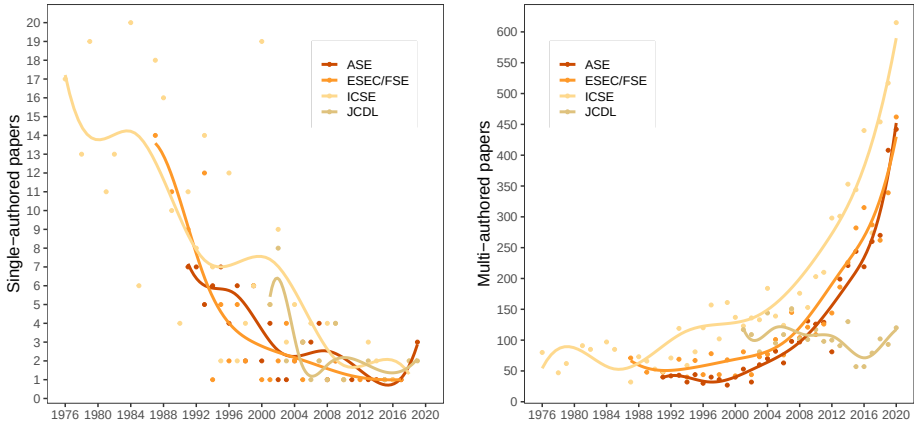
Measurements. To understand how collaboration patterns between researchers changed over time ($\mathbf{RQ_1}$), we measured for each conference individually the total number of single-authored papers, the single-junior-authored papers, the number of multi-authored papers, and the number of papers written by multiple authors where a junior author is the first author. We distinguish junior authors from other authors based on the academic age of each author using the authors' first publication year ($Year_{firstPaper}$) and a paper's publishing year ($Year_{paper}$), extracted from dblp individually for each (author, paper) pair. The academic age is the time span for which a researcher has actively published papers, which we computed as follows:

$$Age_{academic} = Year_{paper} - Year_{firstPaper} + 1 \tag{1}$$

Junior researchers have an academic age that ranges from one up to three years.
Results. The average number of authors for the papers during the observation period increases on average (year/average):

- ASE: 1991/2,0 – 2000/2,8 – 2010/3,7 – 2015/4,1 – 2020/4,8
- ESEC/FSE: 1987/2,2 – 2000/2,5 – 2010/3,3 – 2015/3,8 – 2020/4,6

[6] https://dblp.uni-trier.de/
[7] https://dl.acm.org/proceedings



(1) Single-authored papers. (2) Multi-authored papers.

Fig. 1: Authorship trends at the four conferences.

– ICSE: 1978/1,8 – 2000/2,5 – 2010/3,7 – 2015/4,1 – 2020/4,8
– JCDL: 2001/3,0 – 2010/3,7 – 2015/3,2 – 2020/3,6

The average increases and seeing that the changing number of authors does not fully align with the number of accepted papers (e.g., ICSE'12: 87; '13: 85; '14: 99; '15: 82; '16: 101) [5], it seems that the continuous increase is caused by more collaboration [6]. This is a widely acknowledged pattern caused by the number of authors on papers increasing over time [23].

We display the number of papers written by single and multiple authors in Figure 1. Here, we can see that collaboration has been trending upwards. The high variance particularly in the earlier years may have various reasons, such as the number of active researchers at that time or the popularity of the computer-science domain and software engineering. Moving to juniors' contributions and the level of collaboration they need to reach top-level conferences, we display in Table 2 the number of single-authored papers and single-junior-authored papers—in addition to the papers where a junior is the first/lead author. It is clear that juniors tend to participate at a higher rate in multi-authored papers in subsequent positions. These findings are not surprising, because juniors are less experienced and knowledge is a cumulative process that needs time. So, a senior is more likely to have better abilities to write papers [21]. Our data may also signalizes that the reputation and quality of the conferences could have a negative impact on juniors, since they may be discouraged to submit alone.

Table 2: Overview of author collaborations and juniors' involvement.

conference	single-authored papers / juniors	%	multi-authored papers / juniors as 1st author	%
ASE	81 / 17	21	3,659 / 397	10.8
ESEC/FSE	85 / 24	28.2	4,179 / 395	9.4
ICSE	281 / 96	34.1	7,153 / 665	9.2
JCDL	34 / 3	8.8	2,053 / 189	9.2
all	481 / 140	29.1 %	17,044 / 1,646	9.6 %

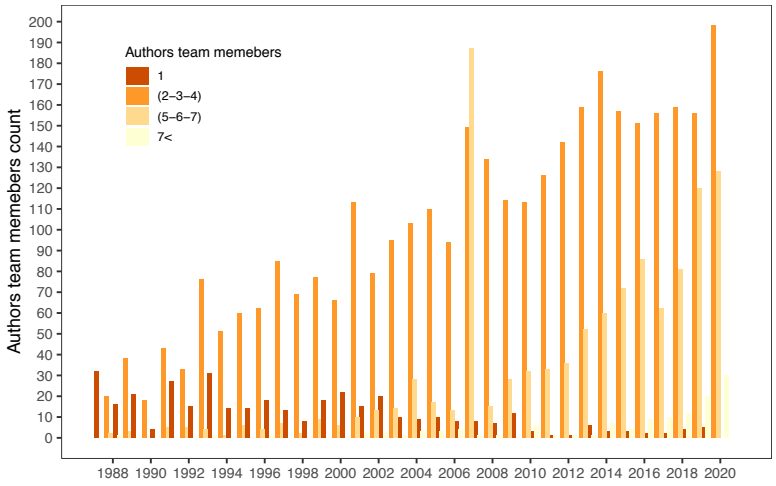


Fig. 2: Collaborating author teams' members count of the four conferences.

5 RQ_2: Frequent Collaboration Patterns

Measurements. To discover the most frequent number of authors collaborating together to publish a paper, we analyzed all papers' authors for each year. Specifically, we measured for each conference individually and combined the number of authors participating in a paper (multi-authored paper). We categorized the results into four groups: single-authored, (2-3-4) authors, (5-6-7) authors, more than 7 authors.

Results. We can see in Figure 2 that most authors' team sizes is within the (2-3-4) category, whereas papers written by more than seven authors represent the lowest over all years. So, the most frequent combination are two to four authors, and as we are focusing on accepted papers at top-level conferences, we can consider this combination as the most successful team size for collaborations. The exact numbers for ASE and ICSE are two authors, while it is three authors for ESEC/FSE and JCDL.

6 Conclusion

In this paper, we studied researchers' collaborations at four top-level conferences. By analyzing 5,188 main-track papers and their 8,730 authors, we observed a rise in the average number of co-authors that does not fully align with the number of published papers. More precisely, the results indicate an increasing level of collaboration and decreasing number of single-author papers. This illustrates potential challenges certain groups of researchers may encounter, since juniors comprise approximately 29 % of the authors of single-authors papers, whereas their chances as first authors decline to 9.6 % in multi-authors papers. The results reveal that the most frequent collaboration pattern ranges from two to four authors. In the future, we plan to expand on these findings to improve our understanding of juniors' impediments.

References

1. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: Understanding the contributions of junior researchers at software-engineering conferences. In: Joint Conference on Digital Libraries (JCDL). IEEE (2021)
2. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: On academic age aspect and discovering the golden age in software engineering. In: International Conference on Cooperative and Human Aspects of Software Engineering (CHASE). ACM (2022)
3. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: Peer-reviewing and submission dynamics around top software-engineering venues: A juniors' perspective. In: International Conference on Evaluation and Assessment in Software Engineering (EASE). ACM (2022)
4. Bordons, M., Gomez, I.: Collaboration networks in science. In: The Web of Knowledge: A Festschrift in Honor of Eugene Garfield (2000)
5. Bultan, T., Whittle, J.: ICSE 2019 PC Chairs Report (2019)
6. Casadevall, A., Semenza, G., Jackson, S., Tomaselli, G., Ahima, R.: Reducing bias: Accounting for the order of co–first authors. Journal of Clinical Investigation **129**(6), 2167–2168 (2019)
7. Chen, J., Konstan, J.A.: Conference paper selectivity and impact. Communications of the ACM (2010)
8. Costas, R., Bordons, M.: Do age and professional rank influence the order of authorship in scientific publications? some evidence from a micro-level perspective. Scientometrics **88** (2011)
9. Diem, A., Wolter, S.: The use of bibliometrics to measure research performance in education sciences. Research in Higher Education **54**(1) (2013)
10. Fernandes, J.: Authorship trends in software engineering. Scientometrics **101**, 257–271 (2014)
11. Godin, B., Gingras, Y.: Impact of collaborative research on academic science. Science and Public Policy **27**(1), 65–73 (02 2000)
12. Hara, N., Solomon, P., Kim, S.L., Sonnenwald, D.H.: An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the American Society for Information Science and Technology **54**(10), 952–965 (2003)
13. Laurance, W.: Second thoughts on who goes where in author lists. Nature **442**(26) (2006)
14. Li, W., Aste, T., Caccioli, F., Livan, G.: Early coauthorship with top scientists predicts success in academic careers. Nature Communications (2019)
15. Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J.: Viewpoint research evaluation for computer science. ACM **52**(4) (2009)
16. Perretti, F., Negro, G.: Mixing genres and matching people: A study in innovation and team composition in hollywood. Journal of Organizational Behavior **28**, 563 – 586 (2007)
17. Qi, M., Zeng, A., Li, M., Fan, Y., Di, Z.: Standing on the shoulders of giants: the effect of outstanding scientists on young collaborators' careers. Scientometrics **111** (2017)
18. Regalado, A.: Multiauthor papers on the rise. Science **268**(25) (1995)
19. Rørstad, K., Aksnes, D.: Publication rate expressed by age, gender and academic position – a large-scale analysis of norwegian academic staff. Journal of Informetrics **9**(2) (2015)

20. Simonin, B.L.: The importance of collaborative know-how: An empirical test of the learning organization. Academy of Management Journal **40**(5), 1150–1174 (1997)
21. Tien, F., Blackburn, R.: Faculty rank system, research motivation, and faculty research productivity: Measure refinement and theory testing. Journal of Higher Education **67**(1), 2–22 (1996), cited By 108
22. Tran, H., Cabanac, G., Hubert, G.: Expert suggestion for conference program committees. pp. 221–232 (05 2017)
23. Wren, J.D., Kozak, K.Z., Johnson, K.R., Deakyne, S.J., Schilling, L.M., Dellavalle, R.P.: The write position. EMBO reports **8**(11), 988–991 (2007)
24. Zhou, M., Mockus, A.: What make long term contributors: Willingness and opportunity in oss community. In: International Conference on Software Engineering (ICSE). pp. 518–528 (2012)
25. Zuckerman, H.: Nobel laureates in science: patterns of productivity, collaboration, and authorship. American sociological review **32 3**, 391–403 (1967)