Understanding the Contributions of Junior Researchers at Software-Engineering Conferences

Rand Alchokr*, Jacob Krüger*†, Yusra Shakeel*, Gunter Saake*, Thomas Leich‡
*Otto-von-Guericke University Magdeburg, †Ruhr-University Bochum, ‡Harz University Wernigerode; Germany
Email: {rand.alchokr | shakeel | saake}@ovgu.de; jacob.krueger@rub.de; tleich@hs-harz.de

***Abstract*—Junior researchers play a key role in advancing research by providing diverse and novel points of view. However, their participation in the scientific community and especially computer science is not well-understood. In this paper, we describe our first steps towards understanding the contribution (i.e., in terms of publications) of junior researchers to computer science. More precisely, we investigated to what extent junior researchers contribute publications to four highly reputable software-engineering conferences. We collected data on 5,188 main-track research papers and the corresponding 8,730 authors. The incipient results indicate a decline in the proportion of junior researchers contributing to the main-tracks of these conferences. Moreover, their ratio of contribution is highly related to collaborations with more experienced researchers. With this pilot study, we aim to show that the analysis method we employed can foster a more detailed understanding of the status and development of junior researchers' contributions.**

***Index Terms*—academic age, junior researchers, collaboration**

I. Introduction

Junior (or early-career) researchers are defined as those researchers who have only recently started to work in research; typically, they have up to three years of research experience [5]. Such junior researchers can be key drivers of novel research by providing innovative ideas, different points of view, and a high motivation. They can broaden the knowledge and perspectives of a research team or whole community, potentially disrupting and opening up long-standing groups that may have unintentionally reduced their external collaborations. Consequently, junior researchers represent a novelty-enhancing part of their teams [4]. For this reason, it is essential to understand their involvement in the research community and analyze their collaboration with more experienced researchers, since exploring their contributions can help reveal hidden obstacles regarding their involvement. For example, research suggests that the actual as well as academic age of researchers impact their publication rate, indicating that junior researchers could face challenges of contributing to research [3], [6].

In this paper, we study a factor that has received little attention when analyzing the contributions of junior researcher in software engineering: their *academic age*, which we define as the time span a researcher has actively published (i.e., junior researchers have an academic age of up to three years). For this purpose, we present an initial overview of junior researchers' contributions as well as their collaborations at major scientific conferences. Precisely, we extracted all papers and the corresponding author data for the main tracks of four well-established (mainly software-engineering) conferences: 1) International Conference on Automated Software Engineering Conference (ASE); 2) Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE); 3) International Conference on Software Engineering (ICSE); and 4) Joint Conference on Digital Libraries (JCDL). We chose these conferences due to their high reputation and because they are at least as important to their research communities (i.e., SE) as journals [2]. Consequently, researchers of any academic age aim to publish at these conferences, making them ideal subjects for studying the contributions of junior researchers as well as potential obstacles they could face. We collected data on 5,188 main-track papers and their 8,730 authors for all instances of the conferences until 2020. This data serves as a basis for our initial analysis, and thus as a starting point for future research on the involvement of junior researchers.

TABLE I: Overview of our dataset.

conference	period	# papers	# authors	# unique authors
ASE	1991–2020	1,069	3,737	2,482
ESEC/FSE	1987–2020	1,239	4,312	2,614
ICSE	1976–2020	2,300	7,434	4,380
JCDL	2001–2020	580	2,087	1,393
total		5,188	17,570	8,730

II. Methodology

To collect our data, we crawled *dblp*.[1] We chose dblp, because its data is structured by conferences, covers our four subjects completely, is open-access, and has a high data quality (e.g., distinguishing authors with the same names by author IDs). To further improve the comparability and quality of our data, we extracted only main-track papers of each conference. Consequently, we had to analyze the session information (provided as labels) in dblp to identify main-track papers. Moreover, we manually compared these labels to official information in the ACM Digital Library[2] to identify mislabeled papers. We experienced that, particularly for older editions of the conferences that involve only a single volume of proceedings for all tracks, main-track papers are not clearly labeled across these sources. For this reason, we decided to enforce one more proxy criterion if the above validation was not conclusive: If we could not clearly label a paper as belonging to

[1] https://dblp.uni-trier.de/
[2] https://dl.acm.org/proceedings

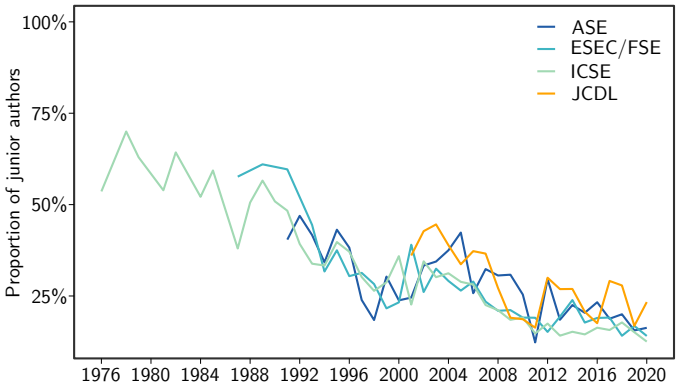

Fig. 1: Proportion of junior researchers over years.

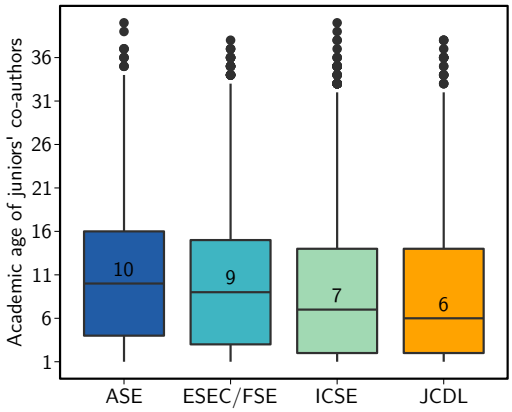

Fig. 2: Distribution of the academic age of juniors' co-authors.

the main track, we excluded it if it comprised fewer than seven pages. Thus, we excluded, for instance, vision and short papers, journal-first and keynote abstracts, as well as tool demos that were part of the same proceedings as the main-track papers.

In Tbl. I, we summarize our dataset. Note that the total of unique authors (8,730) is not the sum of the last column, since we counted each author only once in and across all conferences. We calculated the academic age of an author as follows:

$$Age_{academic} = Year_{paper} - Year_{firstPaper} + 1 \quad (1)$$

We calculated the academic age individually for each author and published paper ($Year_{paper}$) based on the authors' first publication ($Year_{firstPaper}$). We extracted the data from dblp and considered the actual first publication of an author, not the first paper at one of the four conferences. Based on Eq. 1, junior researchers have an ($Age_{academic}$) of one to three years.

III. Results and Discussion

Contributions by junior researchers. We display the ratios at which junior researchers contributed to each conference over the years in Fig. 1. As we would expect, the ratios start high and decline over time. Still, we can see three separate phases: First, while establishing conferences (until 1996), the ratio of junior researchers was comparatively high. Second, the ratios varied heavily between 1996 to 2012. Interestingly, at JCDL the first phase was essentially skipped, indicating that conferences as publication venues got established—not the individual conferences only. Third (since 2012), the ratios stabilized at an even lower level across all conferences, with few outliers (e.g., JCDL 2016 had almost 30 % junior researchers). This evolution is highly interesting, for instance, the third phase may be caused by more collaborations between researchers, a higher reputation of the conferences, or gate-keeping by reviewers. Potentially, this had a negative impact on junior researchers, either directly (e.g., gate-keeping) or indirectly (e.g., discouraging submissions).

Researchers' collaborations. Collaboration is an essential activity in a researchers' work to obtain new experiences, and it is rapidly increasing in almost all research communities [1]. Interestingly, 9.3 % (481) of all 5,188 papers have been written by a single author, but only 140 by junior researchers—which are also very old papers (i.e., 80 % before 1995). 4,707 papers have been written by multiple authors and 2,083 involve at least one junior researcher. We studied such collaborations in more detailed, which is why we computed the box plots we display in Fig. 2 that show the distribution of the academic age of junior researchers' co-authors. As we can see, the median differs between the conferences, but they still indicate a median academic age of six to ten years. Consequently, it may be challenging for junior researchers to contribute to one of the conferences without having co-authors that have extensive knowledge on how to write and polish papers accordingly.

IV. Conclusion

In this paper, we presented a first overview of the degree of contribution of junior researchers to four major (mainly software-engineering) conferences. Our findings indicate a steady decline in the ratio of junior authors contributing to the conferences that stabilized in recent years. Moreover, our results show that junior researchers at these conference extensively collaborate with more experienced researchers. Note that our findings are preliminary and we require more research to understand the actual reasons for our observations.

In that direction, we plan to investigate whether the decreasing ratios of junior researchers are a naturally occurring phenomenon (e.g., obtaining the expertise to conduct high-quality research) or caused by artificial barriers for junior researchers (e.g., gate-keeping, paper engineering). Currently, we focus on potential biases (e.g., academic age, reputation, institution) that may be caused or mitigated by different reviewing models. Based on our work, we hope to help the research community improve and foster the involvement of junior researchers.

References

[1] J. Adams, "Collaborations: The rise of research networks," *Nature*, 2012.
[2] J. Chen and J. A. Konstan, "Conference paper selectivity and impact," *Communications of the ACM*, 2010.
[3] A. Diem and S. C. Wolter, "The use of bibliometrics to measure research performance in education sciences," *Research in Higher Education*, 2013.
[4] R. Katz, "The effects of group longevity on project communication and performance," *Administrative Science Quarterly*, 1982.
[5] W. Li, T. Aste, F. Caccioli, and G. Livan, "Early coauthorship with top scientists predicts success in academic careers," *Nature Communications*, 2019.
[6] K. Rørstad and D. W. Aksnes, "Publication rate expressed by age, gender and academic position - a large-scale analysis of Norwegian academic staff," *Journal of Informetrics*, 2015.