On Academic Age Aspects and Discovering the Golden Age in Software Engineering

Rand Alchokr
Otto-von-Guericke University
Magdeburg, Germany
rand.alchokr@ovgu.de

Jacob Krüger
Ruhr-University Bochum
Bochum, Germany
jacob.krueger@rub.de

Yusra Shakeel
Otto-von-Guericke University
Magdeburg, Germany
shakeel@ovgu.de

Gunter Saake
Otto-von-Guericke University
Magdeburg, Germany
saake@ovgu.de

Thomas Leich
Harz University & METOP GmbH
Weringerode & Magdeburg, Germany
tleich@hs-harz.de

ABSTRACT

Background: Physical aspects are essential human factors that play a key role in a researcher's career and development. Aging is one of the most important physical aspects that can impact the productivity of a researcher (e.g., in terms of publications). In parallel, aging adds experience and proficiency on the scientific research work, such as assuring the quality and reliability of research.
Objective: We aim to understand the impact of aging and the academic age on research publications productivity of research software engineers - the people actively developing software or conducting research in an academic research environment - and explore their *Golden Age* aspect.
Method: We performed a first study on the age distribution of researchers who have published at three famous and prestigious software-engineering conferences: ASE, ESEC/FSE, and ICSE, including 4,620 research-track papers and their 7,337 authors.
Results: The results suggest that the academic productivity is maximized at year 15 (*Golden Age*) and it is held roughly constant for further 15 years before it declines. The results also find, that half authors disappear after their first publication year, reflecting dropout rates that academia suffers from.
Conclusion: Through this pilot study, we share insights on the age distribution, and thus representation, of software-engineering researchers at major conferences and try to understand whether certain groups of researchers are over- or underrepresented.

CCS CONCEPTS

• **General and reference**; • **Software and its engineering**;

KEYWORDS

Software engineering, publications, academic age

CHASE 2022, May 21–22, 2022, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Rand Alchokr, Jacob Krüger, Yusra Shakeel, Gunter Saake, and Thomas Leich. 2021. On Academic Age Aspects and Discovering the Golden Age in Software Engineering . In *Proceedings of the 15th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE), May 21–22, 2022, Pittsburgh, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Exploring scientific communities in their societal context (i.e., aging, productivity, characteristics) is an important research direction to understand their composition, inclusiveness, and impact. In this paper, we report an initial bibliometric study on one of the most essential factors impacting a researcher's career in Software Engineering (SE): their age. Biologists define *aging* as an age-dependent decline of intrinsic physiological function that impacts everyone's abilities [3, 5, 20, 22]. Researchers have conducted numerous studies across various research communities on aging in the context of diseases, functional decline, mortality, "successful" aging, and productivity limits. Similarly, researchers explored how aging impacts scientific work and attempted to define the most productive ages in their fields. One of the most distinguished works in this regard has been conducted by Lehman [18], who attempted to determine the age at which individuals are most likely to make notable contributions or achievements in their professional field. Already in 1966, Lehman [15] defined the concept of the so-called *Golden Age*, referring to the age at which researchers are most productive. Such studies have been replicated in several other communities (e.g., physics, mathematics, chemistry, engineering). However, many of these studies exhibit contradicting results or involved only a small sample of researchers.

Building on such research, but concentrating on the SE community, *our goal is to study the relation between the academic age (i.e., the period in which a researcher published actively) of SE researchers and their scientific output, namely, when does the highest output of published papers in top venues start, and when does it end.*

For this purpose, we extracted data on papers and authors of the: 1) International Conference on Automated Software Engineering (ASE); 2) Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE); and 3) International Conference on Software Engineering (ICSE). We chose conferences, since computer-science (and particularly software-engineering) research compared to other communities is generally more focused on those instead of journals [4, 19, 23]. These conferences have a high reputation, which is why most researchers of any academic age and reputation aim for them. Therefore, by concentrating on these conferences, we ensure that our analysis should cover fundamental contributions to the community. We elicited and analyzed data of the 4,620 papers (and their 7,337 authors) published at the main tracks of all instances of these conferences until 2020. Our dataset is available as an open-access repository.[1]

In this paper, we report and discuss our analysis of how the academic age impacts SE researchers' productivity; and define their *Golden Age*. According to CRA guidelines: "success as a researcher is then not primarily a matter of numbers[2]". Thus, this research does not aim to measure a researcher's overall scientific impact or success, as measuring success is relatively a more complex task that includes further significant actors such as citation counts.

Since the existing literature provides only few insights on the impact of aging in the SE community, our results shed a light into important issues, such as the involvement of new researchers and help to understand whether certain groups of researchers are over- or underrepresented. Most importantly, we identify important future research directions to help reveal potential biases or fundamental barriers for researchers. Ideally, our results can help to tackle misrepresentation of different academic ages, for instance, in hiring committees or faculty positions, and facilitate their involvement in research.

2 BACKGROUND AND RELATED WORK

Aging is a biological fact that occurs beyond human control. Researchers have analyzed aging as a natural progression from a biological point of view and also measured the productivity connected to it [3, 5, 7, 10, 20]. However, perspectives on aging differ between societies and shift during human evolution. For instance, Gorman [10] exemplifies that most developed countries define a retirement age between 60 to 70 years. In contrast, in many developing countries the definition of old age depends on various factors that may or may not indicate an individual's contribution to society. Such factors can involve losing the ability to handle responsibilities or to participate actively in social activities.

Aligning to the second perspective, researchers studied at what age professionals are likely to make notable contributions to their field. Particularly, Lehman [15] found that researchers seem to be most productive during their thirties and early forties. Yet, in multiple research communities, for instance, physics and mathematics, this peak could be reached earlier [15, 17]. In chemistry, Lehman [16] analyzed the profiles of 2,500 researchers and found that chemists seem particularly productive at the age of 30 to 34. Likewise, achievement peaks occur around the thirties in several engineering and other technology communities, according to an analysis of 119 contributors from the field of electrical engineering [18]. However, exceptions to this age trend are considerable, and outstanding contributions may occur at any age, from late teens to the 80s. Overall, the existing studies exhibit somewhat contradicting results.

Both the physical and academic age can impact the publication rate [6]. As explained by Rørstad and Aksnes [21], there is an age-related rise and decline pattern and the relationship between age and publication rate can be curvilinear, most prominently in engineering and technology. Age is an influential factor in publications production where it increases with age and then declines [1, 2]. Győrffy et al. [11] used another definition of *Golden Age*, which comprises the years between the maximum growth and maximum number of citations in a year. They analyzed the number of citations of high-quality papers of Hungarian researchers from 2014 to 2019 to evaluate the performance evolution. The findings show that the average age with the highest citation growth for the researchers was 41.53 years, but it varied between different communities. In summary, the study indicates that scientific careers require decades to reach their peak number of citations, with the Golden Age starting only in the second half of a researcher's 40s.

Gingras et al. [9] investigated the research careers of 6,388 university professors in Quebec to identify "turning points." They found that the first turning point occurs at the age of 40, when the productivity begins to slow down after a sharp increase at the beginning of the careers. A second turning point is evident at around 50 years, when there is a peak of productivity but the lowest average scientific impact. At this point, most researchers tend to publish less as first authors and move more into a supervisor role. For active researchers, the maximum productivity tends to remain stable until they retire. According to Hunt [12], Laurance [14], Tscharntke et al. [24], typically the first author receives most of the credit for being the main contributor, whereas the position of subsequent authors is usually decided by contribution, alphabetical order, or reverse seniority. However, there might be different methods and criteria that committees adopt to quantify author's contributions due to different traditions across countries and research fields [14]. While such studies concentrate mainly on other communities, we focus on SE specifically and take a broader perspective.

3 METHODOLOGY

Dataset. We studied all main-track papers of three major SE conferences starting from the first edition of each conference (ASE 1991; ESEC/FSE 1987; ICSE 1976) until 2020. Additionally, we collected information on all authors. We selected these conferences because they have a high reputation, indicating that researchers from different academic ages submit high-quality papers, and they cover a wide range of topics in SE.

To collect our data, we automatically crawled *DBLP*,[3] which provides bibliographic data structured (among others) by publication venues. We chose DBLP because it covers all three conferences completely, is open-access, and has a high data quality (e.g., distinguishing authors with the same names by providing a web page for each author). As suggested in a study by Kim [13], scholars can consider DBLP data as highly accurate in disambiguating author names, but in some homonym cases may not be properly distinguished. Note that while DBLP has a single website for each ASE and ICSE, it has two websites for ESEC/FSE: one covering ESEC

[1] https://www.dropbox.com/s/pwow4agtjxgx3nr/Goldenagepaper.zip?dl=0
[2] https://cra.org/resources/best-practice-memos/incentivizing-quality-and-impact-evaluating-scholarship-in-hiring-tenure-and-promotion/
[3] https://dblp.uni-trier.de/

Table 1: 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
total		4,620	15,483	7,337

and one covering FSE until both merged completely in 2017. For this reason, we extracted all data from both websites and deleted duplicate entries afterwards.

To improve the comparability and quality of our data, we chose to only study the main-track papers of all three conferences. Thus, we removed all non-main-track papers from our dataset. As a quality check, we manually compared the final numbers of papers against the official statistics in the ACM Digital Library,[4] and compared the session information on research tracks (provided as labels) in DBLP with the information provided in the ACM Digital Library. We noticed that, particularly for older editions of the conferences, main-track papers are not clearly labeled. Therefore, we decided to enforce a proxy criterion if the label was not clear. Namely, we excluded any paper that comprises fewer than seven pages. Considering the expected paper lengths of the three conferences over time, we argue that this is a reasonable proxy criterion to validate the remaining papers. This validation helped us to select 4,620 main-track papers out of 11,106 papers in our initial dataset. The extracted data from DBLP comprises mainly standard bibliographic data for each paper and each author, such as, DOI, conference, year, page count, as well as the years an author started and stopped publishing.

In Table 1, we summarize the properties of our dataset. We can see that we extracted 4,620 main-track papers that have been written by 15,663 authors. Note that the column “# authors” includes authors multiple times when they wrote more than one paper.

Academic Age. The academic age is the time span during which a researcher has actively published. We calculate it as follows:

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

We computed the academic age individually for each author and published paper ($Year_{paper}$) based on the authors’ first publication ($Year_{firstPaper}$). The academic age of a researcher is the age of the eldest paper published by the researcher [8]. To eliminate bias, it was important to be precise and differentiate between the current academic age of an author and the author’s age when a paper was published. So, a similar equation applies for computing the current academic age for an author using ($Year_{lastPaper}$) instead of ($Year_{paper}$). We extracted the data from DBLP and remark that we considered the actual first publication of an author, not the first paper at one of the three conferences.

Productivity. We computed percentiles of productivity at conference level and overall of the authors using the following equation:

$$AvgProd_{(age,conf)} = \sum MT / \sum Authors \quad (2)$$

Precisely, for each academic age and conference, we computed the ratio between the main-track papers (MT) and the total number of authors. This provides an understanding to what extent researchers with a specific academic age contributed to a conference.

[4] https://dl.acm.org/proceedings

4 RESULTS AND DISCUSSION

Measurements. To understand how the publishing activity of the authors in our dataset evolved, we measured for each academic age

- the number of authors at that academic age and their position on their papers; as well as
- the average number of papers that have been published by authors at that academic age.

This provides an intuition regarding the involvement (or productivity) of SE researchers in general.

We started by automatically computing the authors’ academic age based on Equation 1, and identifying their positions on papers using a script. That script scans the papers and authors in our dataset and assigns the corresponding position in the author list to each author. Afterwards, we used a simple grouping function to identify which papers have been published by researchers at what academic age. For the second measure, we used Equation 2 to compute the average productivity for each academic age.

Results. In Figure 1, we display an overview of the authors’ academic age and position in the author list when they published a paper at any of the three conferences. The data shows that most authors are actually in their first publishing year when getting their papers accepted at those top conferences, and that this ratio declines over time. Additionally, in Figure 2 we display how many papers the authors of a certain academic age published on average at each conference separately and combined. As we can see, for the entire dataset the average number of papers increases with the academic age until a certain point, at which it remains rather stable before dropping. Finally, we present details on the Golden Age of SE researchers in Table 2, summarizing information on researchers with the most papers in one year.

Observations. From our results, we derive four key observations:

O_1 Most authors accepted at the conferences have been in their first publication year.

O_2 The number of first authorships declines with a higher age.

O_3 The number of authors with an academic age above one drops drastically (by more than 50 %).

O_4 The average number of papers published at a certain academic age peaks at around 15 years and remains stable before it declines in the 30s.

Discussion. O_1 and O_2 indicate age-related trends. Both observations imply that the SE community involves junior researchers as students and practitioners into research. In detail, authors in their first year of their publication activity are by far the largest subgroup (for the three conferences over years combined). Many authors have only started publishing in that year, and seeing the large drop afterwards, may have published only this paper. This is a typical situation for students working on a thesis that they publish with their advisors, but who decide against a scientific career. Similarly, collaborators from industry are usually not publishing much on their own, which could explain particularly the higher number of non-first authors with an academic age of one. These observations are underpinned by O_3, which is also reasonable, indicating that more experienced researchers switch into advisory roles. In contrast, junior researchers have to build their reputation and careers, for which they need to publish their own research and focus

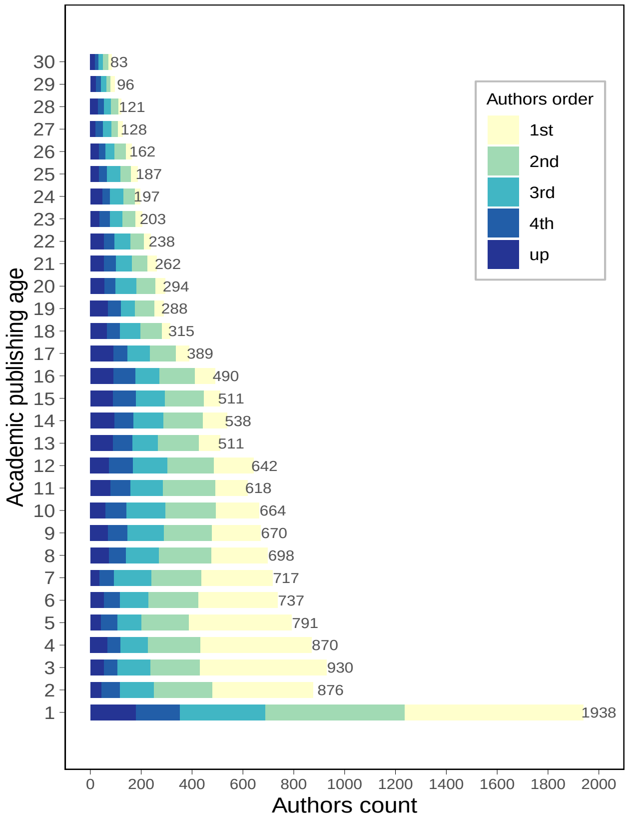

Figure 1: Total number of authors at all three conferences of a specific academic age and their position on a paper.

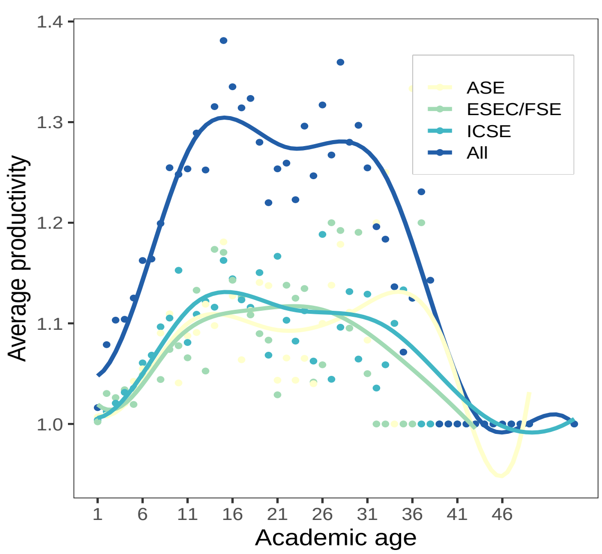

Figure 2: Average number of papers published per author of a certain academic age.

Table 2: Overview of the Golden Age.

# papers	# authors	avg. age (years)	age details (years)
9	1	14	14
8	5	12.6	(7, 9, 14, 15, 18)
7	6	15.6	(12, 2*13, 15, 19, 22)
6	12	16.5	(3*9 , 11, 14, 15, 17, 18, 20, 24, 26, 27)
5	33	14.2	(3, 3*8, 9, 21*(10-19), 2*23, 27, 28)
Golden Age:		72.9/5 = 14.58	

more on individual papers. Overall, these observations confirm that our dataset is sensible and align to what we would expect.

O_4 indicates the productivity peak for SE researchers who have published at these three conferences. As we can see in Figure 2 and Table 2, these peaks somewhat harmonize among the conferences, with an overall Golden Age of approximately 14.58 years. This peak remains relatively stable until it starts to decrease in the early 30s. Moreover, there are a few researchers with a surprisingly high academic age, for instance over 50 years of working and publishing research (which is longer than ICSE exists). While our dataset comprises only three major conferences, we still believe that it yields quite robust conclusions in spite of its seemingly limited size. In particular, it seems that SE researchers require quite some experience to publish at these conferences, but their productivity remains rather stable for quite long time. So, they seem to be just as creative and productive at an older age (i.e., an academic age of 30 is around an actual age in the 50s, depending on the country).

5 THREATS TO VALIDITY

A possible threat to the validity is the metrics we used to compute the academic age and the productivity of authors. These may be computed differently, depending mainly on the perspective of the researcher conducting the study. Accordingly, these metrics could get more complex to also consider, for example, other bibliometrics such as citations, or social constructs that characterize the research communities and the perception of "productivity" or "age" differently. Our focus on three SE conferences with high reputation may bias our findings, since there are numerous other conferences and journals that most researchers publish in. Therefore, as we do not know to which extent the results are generalizable to the whole SE society, we consider this study as a first step towards an overly generic one. Additionally, when analysing authors position in the authors list, we considered the most common aspect in that regard according to literature mentioned in Section 2 where the first author has the highest effort in the paper. Thus, other institutes, groups, and faculties, follow different conventions in author list including: alphabetical order, shuffling of positions, order of seniority and more. The correctness of our bibliographic data is also a threat to validity. Despite using a reliable data source and validations, some parts of our data may be incorrect—especially for older conference editions. Moreover, data interpretation may vary between researchers. For this reason, we publish our dataset and analysis scripts for others to verify.

6 CONCLUSION AND FUTURE WORK

In this paper, we reported an empirical analysis on the relationship between one's tenure in the academic SE research community and the number of publications at three top SE conferences to discover the impact of the academic age on the publication activity. For this purpose, we mine data from DBLP and study 4,620 main-track papers and their 7,337 authors of ASE, ESEC/FSE, and ICSE conferences. Overall, we found that the different roles of researchers (e.g., students, advisors, industrial collaborators) can be somewhat observed in the publication data. Moreover, there seems to be a Golden Age for SE researchers starting at an academic age of around 15 years, which remains rather stable for around 15 years. Interestingly, while this Golden Age starts only at 15 years, first-year researchers are actually the largest group among the authors with a drastic decline afterwards (over 50 %). As a result, we argue that publishing at these conferences requires researchers to have the motivation and the desire to pursue. Still, the decline of newer researchers is concerning.

Consequently, improving our understanding of researchers' involvement at top conferences, barriers they face, and their productivity are interesting directions for future research. As a first future step, we plan to facilitate further discussion on some relevant aspects as including the number of cites (using other well-known, available sources others than DBLP). In addition, we plan for turning the observations from this small empirical study into a longer-term and more impactful research program. To this end, more in-depth studies involving a larger dataset considering other conferences and journals, interviews and surveys are needed to articulate the vision of this study to eventually achieve the desired goals in facilitate inclusion in the scientific community.

REFERENCES

[1] Dag W. Aksnes, Kristoffer Rorstad, Fredrik Piro, and Gunnar Sivertsen. 2011. Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology* 62, 4 (2011), 628–636.
[2] F. Barjak. 2006. Research productivity in the internet era. *Scientometrics* 68 (2006), 343–360.
[3] Anne Bronikowski and Thomas Flatt. 2010. Aging and its demographic measurement. *Nat Educ Knowl* 1 (01 2010).
[4] Jilin Chen and Joseph A. Konstan. 2010. Conference Paper Selectivity and Impact. *Commun. ACM* 53, 6 (2010), 79–83.
[5] Ward Dean. 1992. *Evolutionary Biology of Aging.*
[6] Andrea Diem and Stefan C. Wolter. 2013. The Use of Bibliometrics to Measure Research Performance in Education Sciences. *Research in Higher Education* 54, 1 (2013), 86–114.
[7] Charles Fox, Derek Roff, and Daphne Fairbairn. 2001. *Evolutionary Ecology: Concepts and Case Studies.*
[8] Massimo Franceschet. [n.d.]. A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics : an international journal for all quantitative aspects of the science of science, communication in science and science policy.* 83, 1 ([n. d.]).
[9] Yves Gingras, Vincent Larivière, Benoît Macaluso, and Jean-Pierre Robitaille. 2008. The Effects of Aging on Researchers' Publication and Citation Patterns. *PLOS ONE* 3, 12 (2008).
[10] M. Gorman. 1999. Development and the rights of older people. *In: Randel J, et al., Eds. The ageing and development report: poverty, independence and the world's older people. London, Earthscan Publications Ltd.* (1999).
[11] Balázs Győrffy, Gyöngyi Csuka, Péter Herman, and Ádám Török. 2020. Is There a Golden Age in Publication Activity?—An Analysis of Age-Related Scholarly Performance Across All Scientific Disciplines. *Scientometrics* 123 (2020).
[12] R. Hunt. 1991. Trying an authorship index. *Nature* 352, 187 (1991).
[13] Jinseok Kim. 2018. Evaluating author name disambiguation for digital libraries: a case of DBLP. *Scientometrics* 116 (2018), 1867–1886.
[14] W. Laurance. 2006. Second thoughts on who goes where in author lists. *Nature* 442, 26 (2006).
[15] Harvey Lehman. 1953. *Age and Achievement.* Princeton University.
[16] Harvey Lehman. 1958. The Chemist's Most Creative Years. *Science* 127, 3308 (1958).
[17] Harvey Lehman. 1960. The Age Decrement in Outstanding Scientific Creativity. *American Psychologist* 15, 2 (1960).
[18] Harvey Lehman. 1966. The Most Creative Years of Engineers and Other Technologists. *The Journal of Genetic Psychology* 108, 2 (1966).
[19] Bertrand Meyer, Christine Choppy, Jørgen Staunstrup, and Jan van Leeuwen. 2009. Viewpoint Research Evaluation for Computer Science. *ACM* 52, 4 (2009).
[20] Linda Partridge and Nicholas Barton. 1996. On Measuring the Rate of Ageing. *Proceedings of the Royal Society of London* 263, 1375 (1996).
[21] K. Rørstad and Dag W. Aksnes. 2015. Publication rate expressed by age, gender and academic position - A large-scale analysis of Norwegian academic staff. *J. Informetrics* 9 (2015), 317–333.
[22] M Tatar. 2001. "Senescence," in: Evolutionary Ecology – Concepts and Case Studies, eds C. *W. Fox, D. A. Roff, and D. J. Fairbairn (New York: Oxford University Press)* (2001), 128–141.
[23] Hong Tran, Guillaume Cabanac, and Gilles Hubert. 2017. Expert suggestion for conference program committees. 221–232.
[24] Teja Tscharntke, Michael E Hochberg, Tatyana A Rand, Vincent H Resh, and Jochen Krauss. 2007. Author Sequence and Credit for Contributions in Multiauthored Publications. *PLOS* (2007).