CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Comeback or dropout: study of discontinued researchers at early career stage
Authors	Gayen, Avijit;Chakraborty, Somyajit;Mitra, Saikat;Jana, Angshuman
Publication date	2025-02-05
Original Citation	Gayen, A., Chakraborty, S., Mitra, S. and Jana, A. (2025) 'Comeback or dropout: study of discontinued researchers at early career stage', Scientometrics, pp.1-36. https://doi.org/10.1007/s11192-025-05243-z
Type of publication	Article (peer-reviewed);journal-article
Link to publisher's version	10.1007/s11192-025-05243-z
Rights	© 2025, Akadémiai Kiadó Zrt. Published by Springer Nature Limited. This is a post-peer-review, pre-copyedit version of an article published in Nature Sustainability. The final authenticated version is available online at: https://doi.org/10.1007/s11192-025-05243-z
Download date	2026-11-02 00:53:35
Item downloaded from	https://hdl.handle.net/10468/17134



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Comeback or Dropout: Study of Discontinued Researchers at Early Career Stage

Avijit Gayen[1], Somyajit Chakraborty[2], Saikat Mitra[3] and Angshuman Jana[1]

[1] {avijit.gayen, angshuman}@iiitg.ac.in
Indian Institute of Information Technology, Guwahati, India.
[2] somyajitChakraborty@ucc.ie
University College Cork, Cork, Ireland.
[3] Saikat.Mitra2021@iem.edu.in
Institute of Engineering and Management, Kolkata, India.

Abstract

In scientific research, understanding the dynamics of early-career researchers who discontinue their academic pursuits is paramount. However, this study aims to investigate the phenomenon of research discontinuation among early-career researchers, shedding light on the factors influencing their decisions and the potential for their return to the research arena. This work uses a mixed-methods approach to analyze the large-scale bibliography dataset. We extensively investigate the phenomenon of research discontinuation among early-career scholars. As the discontinuation of research activity at the early career stage could be an indication of drop-out in further activities, we have focused on the first 10 years of research tenure of the researchers. In our work, we specifically consider the three consecutive "non-publishing" years as the discontinuation of the research scholar. We conduct an experimental assessment of our proposed method using the AMiner Author dataset and note that among 113,637 researchers, – 89% of researchers have discontinued three years in their initial 10 years of research tenure. We further observe that out of the total researchers, 90% discontinued researchers drop out of their careers in the first ten years of their tenure of work.

Keywords: Discontinued, dropout, Collaboration network, Early career stage, Co-authorship network, citation.

1 Introduction

Scientific research and development plays a key role in driving the progress of any nation. In the current era of science & technology, the economic and social development of any nation is majorly dependent on research and development activities[1]. Collaborative research activity not only at the national level but also across the country becomes popular and effective in various domains of research. The joint participation of researchers in solving a complex problem helps to improve the quality of work and also fosters knowledge among the collaborative researchers [2]. It promotes new skills and techniques among researchers [3]. In the realm of current research activities, the demand for academic and scientific collaborations has steadily increased over the past few decades [4]. The expansion of research and development activities has resulted in the generation of vast and rapidly growing bibliographic metadata. This electronic bibliographic data provides researchers with valuable opportunities to analyze and comprehend the patterns of scientific progress and its dynamics, which play a crucial role in shaping research policies.

The role of early career-stage researchers in collaborative research work has been identified as a significant contribution to the current research trend [5]. In literature, [6–8] early career researchers are usually characterized as young researchers of less than 35 years. This set of researchers has been considered as either they are pursuing a doctorate or holding a doctorate and continuing as post-doctorate researchers [9]. Despite having relatively less experience and limited knowledge, early career researchers (ECRs) demonstrate a high level of enthusiasm and curiosity for new ideas, often driving research to greater heights. Their ability to tackle research challenges, combined with their dedication and perseverance, enables them to solve complex research problems effectively. The collaborative research activity amongst experienced researchers with ECRs develops the research skills among the ECRs. On the other hand, the persistent efforts of ECRs to address complex problems are highly commendable [10]. As a result, ECRs play a significant role in driving transformations within academia, which in turn fosters the nation's research and development.

In spite of the significant contributions of ECRs, identifying effective collaborations to address potential research problems remains a considerable challenge for them [11]. While this issue is common to researchers at any career stage, ECRs may face greater difficulty in selecting the right academic collaborator due to limited domain knowledge and a lack of access to domain experts. This can lead to unsuitable collaborations [12, 13] or missed opportunities for valuable partnerships, ultimately resulting in fewer high-quality collaborations during the early stages of their research careers. This situation is over-complicated due to the constantly increasing bibliographic data of the numerous research articles. Thus, this situation drives away ECR from the research activity which further makes them disassociate and drop out of the research community. Besides these major threats, ECRs also encounter several critical challenges that hinder their career progression. It has been

noted that, due to insufficient focus on ECRs, they face intense competition with well-established researchers or research groups to secure funding for their projects [14]. Additionally, their work on identified problems often gets overshadowed by the contributions of more recognized researchers [15]. In this context, it has been observed that a large number of researchers discontinue their research activities which is reflected in their non-publication tenure. Although a few researchers do a comeback after a few years of non-publication and continue their research work which has been observed from their bibliographic records, amongst those discontinued researchers a major chunk of researchers drop out from the research activity further. Though the discontinuation in research work can happen at any stage of the research career, it has been observed that a large set of researchers discontinue their research activity at the early career stage [9, 10]. Although discontinuation in any kind of socio-technical activity in any social system is common and obvious, the discontinuation of a large set of researchers at their early career stage has a very strong impact. The discontinuation turned into dropout could be a great loss for any nation in the course of the advancement of technology, science, and academics. Moreover, research discontinuation may impact the scholarly community [16].

The nature of discontinued research activity could be caused due to many issues i.e., change in the portfolio (moving to the administration, moving to industry, etc.), lack of time, less enthusiasm at a higher age, etc. amongst the researcher community. Though these socio-economic events in the research community contribute a significant amount of discontinuation in research activity, there exist various causes that trigger some serious situations for ECR which drives ECR to disassociate from research activity. In this work, we are addressing the issue of discontinuation of ECR marked by their non-publication in their research tenure. Non-publication of the researchers have been addressed as discontinuation in many earlier works [16–18].

In a work [19], The author examined research dropout through a gender-specific lens, analyzing data over a span of approximately five decades. The study revealed that gender inequality in research advancement—specifically in hiring practices and project funding—adversely impacts female researchers. Additionally, the authors noted that women are less involved in collaborative research, which contributes to a higher rate of dropout among female researchers. In another work [20], authors cited the major causes of dropout from Ph.D. work. The authors identified the factors contributing to Ph.D. program dropouts, highlighting mental health challenges, financial difficulties, and issues related to the relationship with Ph.D. supervisors as key concerns. In the context of dropout of post doctoral researcher, the authors mentioned imbalance nature of work life, less rewarding, stress of work pressure as well of job security [21]. They also pointed out the gender specific biases of work stress towards female researchers. However, in the literature, we find very few works apart from the above mentioned works addressing the discontinued nature of research activities. It is observable that there is a substantial research gap

which could figure out the nature of the discontinuation of research activities in the research community. By addressing these research gaps and continuing to investigate the challenges faced by early-career researchers, we can work toward fostering a more sustainable, inclusive, and resilient academic community. This, in turn, will contribute to the longevity and diversity of talent in research and innovation, benefiting society as a whole. Promoting collaborative research environments and addressing the challenges faced by discontinued researchers within their co-authorship networks can contribute to a more sustainable and inclusive research ecosystem. The study of the discontinued nature of research activities would help us to understand the root cause of their discontinuation and further help in research policy making. Hence, in this work, we address the following research questions to understand the major issues related to the discontinuation of research activity.

- **RQ1**: *What are the key indicators in publication and collaboration data that can reliably identify early-career researchers experiencing a three-year publication discontinuation within their first ten years of research?*
- **RQ2**: *What are the temporal and collaboration factors that influence the likelihood and timing of early-career researchers resuming academic publishing after a period of publishing discontinuation?*
- **RQ3**: *What underlying publication trends and collaboration patterns differentiate early-career researchers who permanently drop out from those who return to academia after a discontinuation period?*

To address the above research questions, we use large bibliographic metadata of Aminer [22]. First, we removed duplicates and updated any missing parameters in the dataset for our ease of use. In this work, we have classified the researchers as *'discontinued'* researchers if he/she has consecutive three years tenure of non-publication. In some relevant work, non-publication period of $24 - 36$ months has been considered as discontinuation of research and the authors showed the discontinuation has been considered as 'academic loss' [16]. On the other hand, the discontinuation may be attributed to either the completion of their graduation from the institution where they conducted their research or a change in career over time. Thus, the rationale behind the three-year gap in research publication to be assumed as *'discontinuation'* of the research work is to embark on this period which could be a sufficient time window to identify the discontinued researchers as the researchers to can move on to other domain of work i.e., transitioning to administration or entering the industry after graduating from the institution is common. In our dataset, we found that approximately 90% of the researchers in the complete dataset, after preprocessing, have not published any work in the three consecutive years of their research careers. To analyze and understand the nature of discontinuation, we further categorized the *'discontinued'* researchers into ***'come-back'*** and ***'drop-out'***. Our analysis reveals the collaboration and publication patterns of those ***'come-back'*** and ***'drop-out'*** researchers in their careers. The cues might be useful for understanding the root cause of a large

set of dropouts from the research community. Throughout the paper, we have used "Researchers" and "Authors" term synonymously. Although the "Collaborative Work" have a broader sense in the context of research activities, in our paper, we have meant it for "Co-authorship" activity among the researcher which had been identified from their bibliographic records.

2 Related Works

In this section, we distinguish the contributions associated with early career researchers (ECRs) and the significant career challenges they face. First, we emphasize the contributions to scientific collaborative research, highlighting recent trends in collaboration developments. Next, we outline the roles of ECRs in research and identify the problems and challenges they encounter. We also focus on the contributions related to "discontinued" researchers at the early stages of their careers in various research domains. Finally, we discuss the scope of our work, addressing the research gaps identified in the literature.

Trends in Collaborative Research

The need for collaborative research work has been identified long ago. The complex research problems are being addressed by multiple collaborators using their own skills and expertise. A significant increase in collaborative research activities has been noted across various fields [23]. Over the past few decades, the demand for academic and scientific collaborations has experienced substantial growth [4]. The advancement of communication technology has given the immense opportunity to develop collaboration among distant researchers [24]. Thus not only local collaborations but also online collaborations have become popular in the last two decades [2]. In [25], authors noted a continually rising demand for international scientific collaboration within the research community. In a recent study [26], they highlighted the rapidly growing need for new researchers. Additionally, Melin et al. found in their research [3] that 38% of academic respondents identified "Increased Knowledge" as the primary benefit gained from collaborations. The growth of knowledge among researchers advances in the development of science and technology.

Early career stage Researcher

The research efforts of early career researchers (ECRs) in collaborative projects are vital to addressing highly complex scientific challenges within the research community. Several previous studies [6, 9, 27–32] have provided qualitative definitions of ECRs, generally considering individuals at the Ph.D. or post-Ph.D. level. In earlier work [27], it was observed that initiatives by ECRs promote interdisciplinary research, thereby advancing developments in science and technology. Another study [33] highlighted that international exposure to ECRs accelerates knowledge acquisition and education consistently and steadily.

Challenges and Problems of ECR

The importance of ECRs' involvement in addressing critical problems is undeniable across all research fields. However, they encounter various challenges and issues not only in cross-domain research activities but also within their specific areas of interest [34]. One of the major issues of ECR in their early stage is "time management" as they perform multiple activities e.g., teaching, guiding projects of under-graduates, etc. apart from their own research activities [35, 36]. In several significant studies [37, 38], the authors noted that funding and recognition for ECRs significantly influence their performance. A recent study [39] identified major obstacles faced by ECRs in multidisciplinary areas of marine scientist and further proposed strategies to address these challenges. They pointed out the following five major barriers that limit ECRs' capacity for interdisciplinary research: (a) demanding workloads (b) stress linked to funding, publishing, and employment uncertainty (c) limited support for balancing personal and professional commitments (d) ineffective supervisory support and (v) the steep learning curve associated with interdisciplinary research.

In [40, 41], the authors noted that current higher education policies foster a 'risk-career' environment for early career researchers (ECRs), making their careers highly unpredictable and unstable. This uncertainty often drives promising researchers to seek alternative career paths. McDowell et al. examined the review processes of standard peer-reviewed journals and identified the early-stage 'ghost' review writing done by ECRs without any acknowledgment[42]. This situation not only reveals the compromising nature of the review process, which can undermine research quality but also highlights the exploitation of ECRs. Additionally, in another study [43], Drosou et al. emphasized the challenges ECRs face when trying to publish their research reports in prestigious journals.

Discontinuation in research work:

Though the discontinuation from any kind of socio-economic activity at any time stage in the system is common and obvious, the discontinuation in research activity at the early career stage has a strong impact in the growth of science & technology. There exist very few works in the literature where authors address this discontinuation in research work issues, rather it has been assumed as the natural continuous process in the research domain. In some work [19], the author showed the nature of drop-out in research activity from a gender-specific perspective. Authors discussed in their study of a longer period ($\approx$ five decades) that the strong presence of gender inequality in research promotion in terms of the hiring process as well as the funding for the projects affects female researchers. On the other hand female researchers are less adopted in the development of collaboration which in terms leads to more dropouts from research. In another work [20], authors observed the causes of dropout from Ph.D. work. They specifically mentioned the issues related to mental health issues, economic crises as well as problems regarding the Ph.D.

supervisor. In a work [44], authors studied the nature of discontinued journals in several domains. Though the discontinuation further turned into ***'drop-out'*** can not be identified as the age-specific activity, the several issues in the early career stage might be the driving force of the discontinuation of ECR in their career.

The lack of earlier works related to ***'discontinued'***-nature of researchers, motivates the strong need to study the nature of those discontinued researchers in the large research community. While the literature acknowledges the challenges faced by ECRs, it is surprising that none of the studies have attempted to provide substantial solutions to these issues as a primary focus of their research. On the other hand, the analysis of discontinued researchers has not been given any such focus in the literature though it might be a major flaw of the research policy making. Thus there is a strong need of a formal methodology to study and analyze the discontinued researcher in the various domains of the research community to concentrate on the issues related to them. In this paper, we present a method utilizing an electronically updated bibliographic dataset to identify discontinued researchers and to further investigate their various research activities.

3 Methodology

In this section, we first present a comprehensive description of the dataset utilized in our study. Specifically, we summarize the dataset, including a description of each field. We initially describe the method of analysis of academic demographics of the researchers whom we have considered in our experiment to understand the context of this study. Additionally, we outline the methodology employed to study the discontinued researchers and their characteristics to get insight into the possible causes of discontinuation. As the discontinuation in research activity at the early career stage could be an indication of drop-out in further activities, we have focused on the first 10 years of research tenure, i.e., the period of 10 years counting from the first publication of the researcher in the dataset, of the researchers. In our work, we specifically consider the three consecutive "non-publishing" years as the "publishing discontinuation" of the research scholar.

In addition to describing the proposed analysis method, we include a schematic diagram in Figure 1 to illustrate the workflow for implementing the method. We utilized a structured methodology to analyze the dynamics of publication discontinuation among researchers and their career trajectories. Initially, a raw bibliographic dataset containing publication and author records was preprocessed to remove duplicates and fill in missing values. The dataset was segregated into preprocessed paper and author datasets, which were further filtered to identify researchers with publication discontinuation by excluding those with consistent publication activity. Researchers were classified as having publication discontinuation if they exhibited a consecutive three-year gap in their publication history. These researchers were then categorized

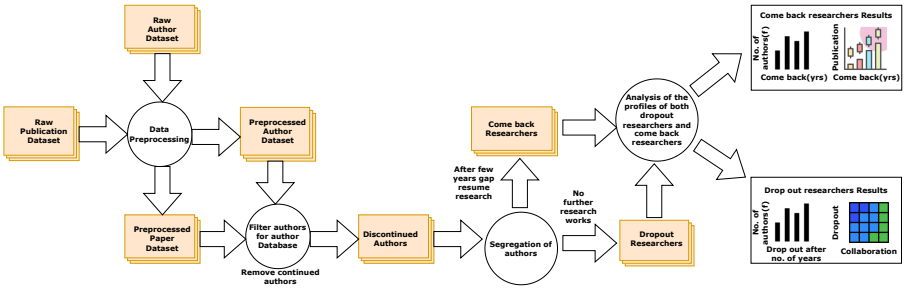


Fig. 1: Figure represents the schematic diagram to describe workflow adopted in the proposed method.

into two groups: come-back researchers, who resumed publishing after the gap, and drop-out researchers, who ceased academic publishing entirely. Profiles of both categories were analyzed to study their publication and collaboration patterns. For come-back researchers, we examined the timeline and circumstances of their return, while for drop-out researchers, we assessed their active research period before cessation. The outcomes for both groups were evaluated, focusing on factors such as collaboration networks, duration of inactivity, and their impact on the number of authors over time. This comprehensive analysis provided insights into the behaviors and trends associated with publication discontinuation, informing strategies to address and mitigate such gaps in academic publishing.

3.1 Dataset Summary

In our work, the AMiner dataset [22] has been used for the analysis. The dataset is open-source and available under a Creative Commons license (CC BY-NC-SA 4.0), maintained by K. Scott Mader. The last update was on December 12, 2018, and it includes bibliographic information for the publications along with the author's detailed information and their collaboration in the duration of 1936—2014. Though the dataset is large and contains extensively detailed information, in our analysis, we have been able to parse a part of it. We utilized the publication information and detailed data about the authors and their co-authorship relationships. The dataset is divided into three separate files, organized according to the attributes listed in the table 1.

3.2 Proposed Method

Initially, we eliminated duplicate data points and filled in the missing values in the dataset for our ease of use. We have considered the researchers as *'discontinued'* researchers if he/she does not publish any work in three consecutive years. The event of publication discontinuation might occur in the early career stage either as a result of graduating from the institution or due to a career change over time. Even if the discontinuation might have been considered natural at any stage of the career, there is a strong urge to find out the causes of

Paper		Author	
Attributes	Attribute Definition	Attributes	Attribute Definition
index	paper ID	index	author ID
*	paper title	n	author name
@	authors	a	affiliations
o	affiliations	pc	publication count
t	published year	cn	number of citations
c	publication venue	hi	h-index
%	the ID of references	pi	p-index(equal A-index)
!	abstract	upi	p-index(unequal A-index)
-	-	t	area of interest

Table 1: Dataset attributes' description

Arnet Miner Author dataset	Raw	Filtered
Number of authors	-	17,12,433
Number of valid authors	-	9,76,883
Number of author parsed	4,75,445	4,75,441
Number of parsed discontinued authors	-	417719($\approx$ 88%)

Table 2: Basic information of raw and filtered dataset(Publication Period: 1936—2014).

Statistics of the Processed Data	Value
No. of Author parsed	1,13,637
No. of Discontinued author	1,00,000
No. of come-back author	4,291
No. of drop-out authors	90,963
Avg. no. of authors per paper (drop-out)	1.89
Avg. no. of authors per paper (come-back)	7.94
Avg. no. of collaborator (drop-out)	2.85
Avg. no. of collaborator (come-back)	3.16

Table 3: Basic statistics of the processed dataset in our work (Publication tenure 1980—2010).

disassociation from regular publishing activity at the early career stage of the researcher. This insight might be useful for making better research policy for the ECR.

In the initial study, it has been observed that there exists $\approx$ 88% of researchers in the complete dataset, after preprocessing, experienced discontinuation, meaning they had no publications in three consecutive years during their entire research tenure. Though we have scanned the entire research career of every *'discontinued'* researcher, We primarily concentrated on the first 10 years of a researcher's career, as this period is critical and often presents significant challenges for them. Our observation reveals that a large set of researchers discontinued their research activities in this period. Although there exists a large set of discontinued authors, we have taken an ample amount of data for analysis further and primarily concentrated on the early stage of their career, specifically the first 10 years.

To analyze and understand the nature of discontinuation, we further categorized the *'discontinued'* researchers into ***'come-back*** and ***'drop-out'***. The ***come-back*** researchers are those who have three years of discontinuation of publication in their research career but break the discontinuation later on and have some publication thereafter. The study of ***come-back*** researcher might be useful to understand the nature of career gap which might reveal the underlying issues behind it at their early career stage. In our study, we identify the researchers as ***'drop-out'*** amongst the *'discontinued'* researchers who do not have any publications further in their careers. Our analysis reveals the collaboration and publication patterns of those ***'come-back*** and ***'drop-out'*** researchers in their careers. The cues might be useful for understanding the root cause of a large set of drop-outs from the research community.

3.2.1 Study of Academic Demographics

In this section, we examine the academic demographics of early-career researchers (ECRs) to understand the broader contexts influencing publication discontinuation and re-engagement. By analyzing the geographic distribution of researchers, institutional affiliations, and publication outputs, we aim to uncover patterns that shape the trajectories of *"come-back"* and *"drop-out"* researchers. This analysis includes the distribution of affiliations among these two groups, categorizing publications by type (e.g., journal articles, conference papers, workshop paper, etc.), and evaluating contributions across regions and institutions. By comparing the prevalence and characteristics of *"come-back"* and *"drop-out"* researchers across geographic and institutional landscapes, we aim to identify disparities and trends within the academic ecosystem that influence research continuity. These findings will provide critical insights into systemic barriers and opportunities, guiding the development of interventions to support sustainable academic careers for ECRs.

3.2.2 Analytical Methods

To study the phenomenon of research discontinuation among early-career researchers, we employed a structured analytical framework using bibliographic data from the AMiner Author dataset. The methods focused on identifying discontinued researchers, analyzing their publication and collaboration patterns, and categorizing them into come-back and drop-out groups.

Temporal Modeling:

We develop a probability distribution model that has been used to understand the temporal patterns of comeback and dropout events. For come-back researchers, we propose a model $p(CB, n)$ to describe the likelihood of returning n years after a publishing discontinuation period of three years. For drop-out researchers, a similar probability model $p(DO, n)$ was used to capture the likelihood of leaving academic publishing after n years of active research.

The parameters of these models were optimized using non-linear regression techniques.

Publication Analysis:

For come-back researchers, the analysis focused on the distribution of publication counts during the comeback years and subsequent research periods. This involved calculating descriptive statistics, such as the mean, median, and quartiles of publication counts, to examine patterns across different comeback durations. We further propose a model $p(CB, n, k)$ to describe the likelihood of publication of k number of papers returning n years after a publishing discontinuation period of three years. For drop-out researchers, publication records prior to their discontinuation were analyzed to assess productivity levels during active years. We perform the statistical analysis similar to the come-back researcher. We also propose a model $p(DO, n, k)$ to describe the likelihood of publication of k number of papers on n^{th} year(considering the year of first publication of the researcher as year 1) before their dropout from the research activity.

Collaboration Network Analysis:

We analyze the collaboration patterns based on the popularity index (pi) of collaborators. The popularity index (pi) quantifies a researcher's influence within collaboration networks, based on bibliometric data like publications, citations, and co-authorships. It reflects their relative academic standing. To understand the attachment probabilities with collaborators of varying popularity levels for come-back researchers, we perform a heatmap analysis to understand their collaboration preferences. Similarly, we develop a heatmap analysis to evaluate the relationships between the popularity indices of drop-out researchers and their collaborators, identifying patterns that might influence research continuation or discontinuation.

4 Results & Discussion

In this section, we discuss the results obtained from the various observations to understand the nature of discontinuation. Initially, we study the academic demographics of early-career researchers to understand the broader contexts influencing publication discontinuation and re-engagement of publication activities. We further specifically studied the publication and collaboration patterns of the ***'come-back'*** and ***'drop-out'*** category of the *'discontinued'* researchers. We have studied 113637 researchers out of which $\approx$ 89% researchers have discontinuation of three years in their initial 10 years of research tenure. We further observe that out of total researchers, 90% discontinued researchers drop out of their careers in the first ten years of their tenure of work. We study the come-back researchers to understand the underlying factors that drive them to return to their research work later on.

4.1 Academic Demographics

In this section, we outline the results of our analysis of academic demographics to understand the factors influencing research discontinuation among early-career researchers. Initially, we examine the overall geographic and institutional distribution of researchers and publications, highlighting disparities in global participation and research output. Further we focus on the *'come-back"* researchers, analyzing their geographic and institutional affiliations as well as their publication contributions to identify potential factors aiding their re-engagement in research. Later, we delve into the academic demographics of *'drop-out"* researchers, exploring their distribution and publication patterns to uncover systemic barriers that may contribute to their departure from the research ecosystem. Together, these sections provide a comprehensive view of the academic landscape and its influence on the trajectories of early-career researchers.

4.1.1 General Statistics

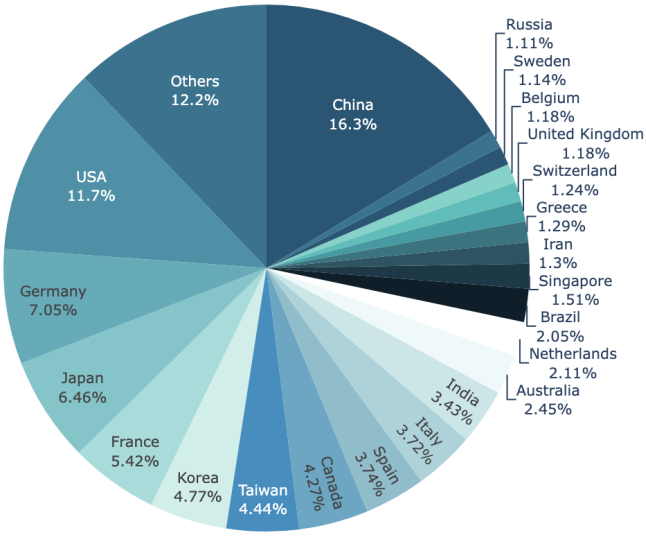


Fig. 2: Figure represents a pie chart to reveal the percentage of authors from each countries extracted from our database. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

In this section, we observe the three major statistics —a) distribution of researchers across the countries as well the institute b) distribution of publications across the countries as well the institute c) nature of publication venue

(i.e., Conference, Journal, Workshop etc.), of our dataset to reveal the demographics of researchers whose bibliographic data has been used for the entire study in this paper.

In figure 2, we illustrate the geographic distribution of researchers, highlighting the proportion of authors from different countries. A notable observation is the dominance of a few countries, such as the United States, China, Germany, Japan, France, Korea, Tiwan, Canada, Spain, Italy and India, which collectively contribute a significant share of researchers. In contrast, several regions, particularly those with developing economies, are underrepresented, with minimal contributions to the global research community. The observation reveals the systemic challenges faced by researchers in underrepresented regions i.e., limited access to research funding, mentorship, and collaboration opportunities which may led to the early discontinuation of academic careers in these areas. Such geographic imbalances highlight the importance of addressing structural inequalities within the global research ecosystem, particularly for supporting early-career researchers and mitigating factors that drive research discontinuation.

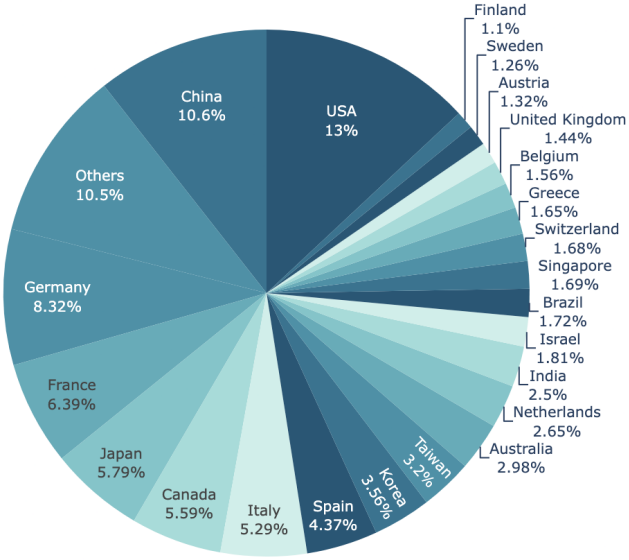


Fig. 3: Figure represents a pie chart to reveal the percentage of publication count from each countries extracted from our database. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

Figure 3 highlights the distribution of publications across countries, offering insights into the global research output. The major observations are as

follows—a) a few countries, notably the United States, China, and a handful of European nations, dominate the total publication counts, b) countries with fewer authors, as identified in Figure 2, also tend to have lower publication counts, c) high-author regions do not always correlate directly with high publication outputs. For instance, some countries with a significant number of researchers exhibit relatively modest publication counts. This trend suggests a concentration of research activity in regions with strong academic infrastructure and funding reinforcing the link between regional participation in research and productivity. Interestingly, some high-author region potentially pointing to inefficiencies or systemic barriers such as limited access to publishing resources, high rejection rates, or lack of collaborative networks. This observation highlights disparities in research productivity that could influence early-career researchers' likelihood of continuation or discontinuation. Addressing these gaps by fostering equitable access to resources and publication opportunities is critical to supporting a more balanced global research ecosystem.

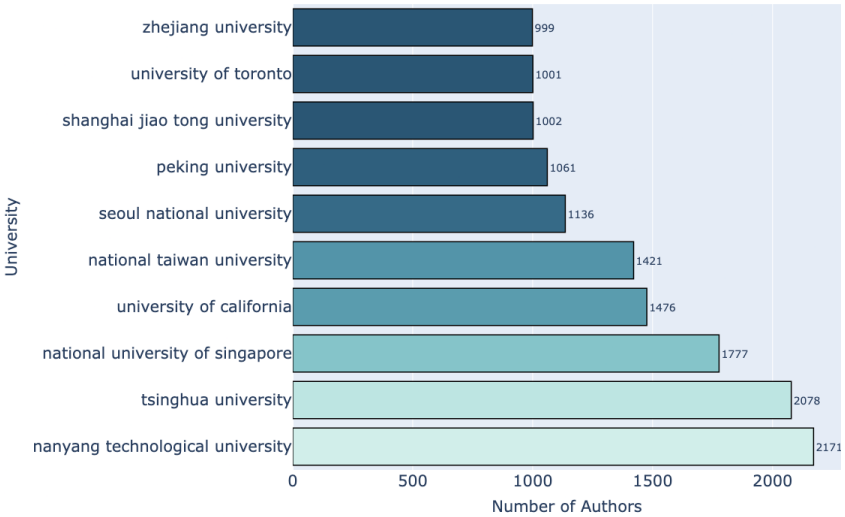


Fig. 4: Figure represents a horizontal bar chart showing the top 10 institutes based on author count as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of authors.

In figure 4, we highlight the top 10 institutes contributing to the dataset in terms of author count, which showcase a concentration of researchers in prominent academic institutions. Similarly, we observe in figure 5 which illustrates

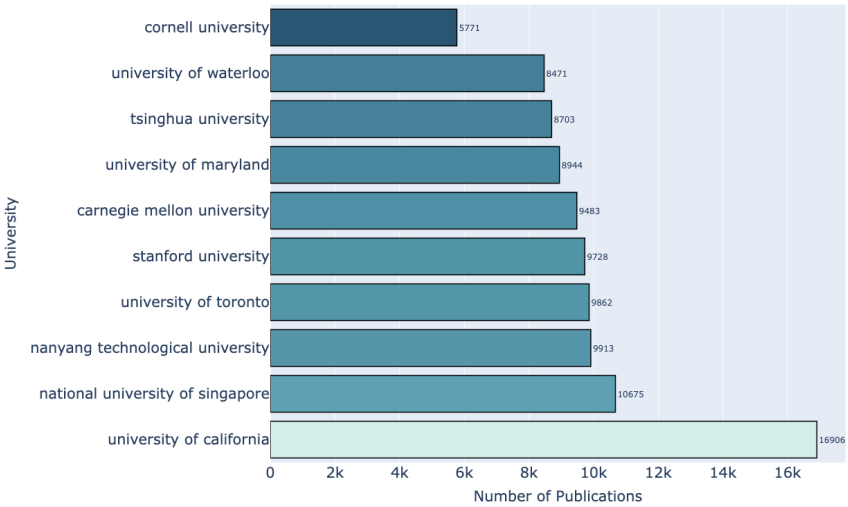


Fig. 5: Figure represents a horizontal bar chart showing the top 10 institutes based on publication count contributed by the authors as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of publications contributed by the authors.

the top 10 institutes but ranked by publication count. The positive correlation between author count and publication output at these institutes, emphasizing the role of institutional support and resources in driving both researcher affiliation and productivity. However, disparities among institutions suggest varying levels of contribution and support for academic engagement and publishing. This insight underscores the importance of institutional policy in fostering research output and mitigating potential early career discontinuation trends.

Figure 6 categorizes publications by type, providing an overview of the preferred channels for disseminating research within our dataset. The key observation is the dominance of conference proceedings as the primary publication type, followed by journal articles, while workshops, books, and other formats contribute relatively smaller shares. This trend is particularly prominent in domains where conferences play a crucial role in sharing preliminary findings and fostering collaborations. Based on this observation, we also suggest that the reliance on conference publications may reflect the nature of research fields covered in the dataset, where fast-paced dissemination of ideas is valued. However, it may also indicate challenges in accessing journal publication opportunities, which often require more time, resources, and mentorship, particularly for early-career researchers.

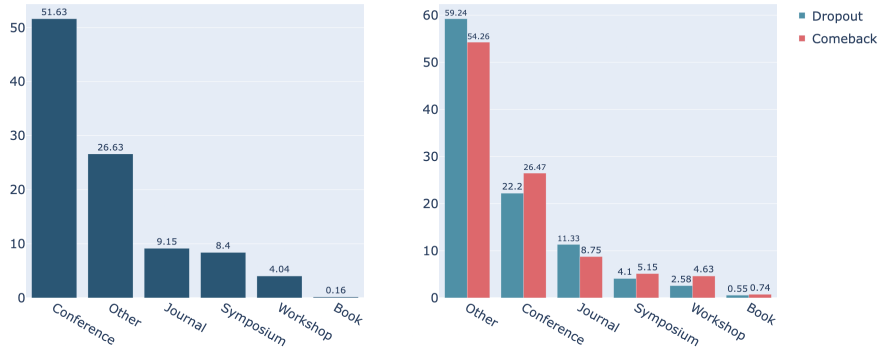


Fig. 6: Figure represents two bar charts showing publication type, i.e., "Conference", "Journal","Symposium", "Workshop", "Book" and "Others". In our dataset, if we do not find any publication marked of any above mentioned type, in the bar chart, we have marked them as "others". The left bar chart shows the distribution of publication types on above mentioned categories whereas the right bar chart compares publication types between *"come-back"* and *"drop-out"* authors. In both of the graph, along x-axis, we plot the categories of publication and along y-axis, it represents the percentage of that category publication.

4.1.2 Come-back Researchers' Statistics

In this section, we observe the three major statistics —a) distribution of researchers across the countries as well the institute b) distribution of publications across the countries as well the institute c) nature of publication venue (i.e., Conference, Journal, Workshop etc.), of our dataset to reveal the demographics of *come-back* researchers whose bibliographic data has been used in our study in this paper.

In figure 7, we observe the geographic distribution of "*come-back*" researchers. It is represented as a pie chart showing the percentage of such researchers by country. We observe that a significant dominance of a few countries, such as the United States, China, and some selected European nations, which contribute the majority of come-back researchers, while developing nations show minimal representation. This highlights systemic disparities in research support systems across regions, where developed nations offer better opportunities for re-engagement in academic publishing after atleast of three year publication discontinuation. This scenario reflects the availability of institutional resources, funding, and collaborative networks in certain countries which plays a critical role in enabling researchers to return after an academic publishing gap. It also underscores the need for global efforts to support researchers in underrepresented regions to bridge this gap.

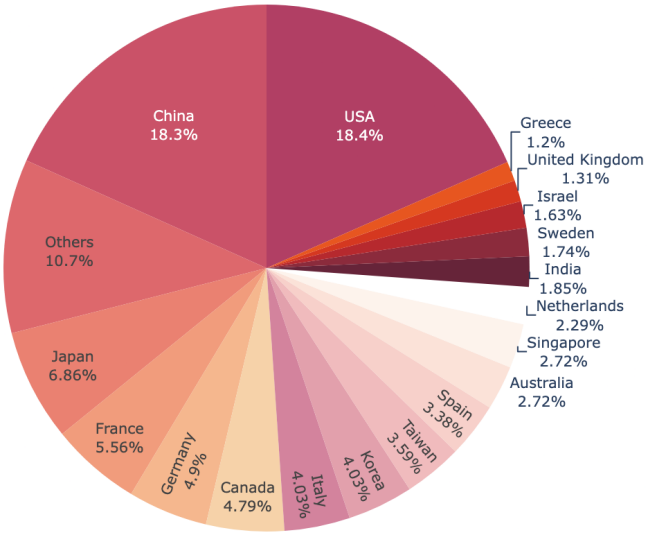


Fig. 7: Figure represents a pie chart to reveal the percentage of *'come-back'* authors from each countries segregated from the all authors in the dataset. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

In figure 8, we show the publication distribution of "*come-back*" researchers by country. It reveals similar trends to figure 7 but with notable differences in proportional contributions. For example, while countries like the United States and China dominate in both figures, their share of publications in figure 8 is higher than their share of researchers in figure 7, that indicates greater productivity among em come-back researchers in these regions. Conversely, some countries with a noticeable share of researchers in figure 7 contribute proportionally fewer publications in figure 8. This highlights that while institutional support may facilitate re-entry into academia, variations in productivity likely stem from differing levels of access to resources, collaboration networks, and publication opportunities across regions.

Figures 9 and 10 examine institutional contributions of *"come-back"* researchers, where figure 9 shows the ranking the top 10 institutions by the number of such researchers and in figure 10 shows the ranking the top 10 institutions by their publication counts. A key observation is that while certain institutions excel in both metrics, others show discrepancies—some with a large number of *come-back* researchers but lower publication outputs, and others with fewer researchers but higher productivity. This suggests that institutions with a higher number of researchers may focus on broad reintegration efforts, while those with fewer researchers but higher outputs emphasize

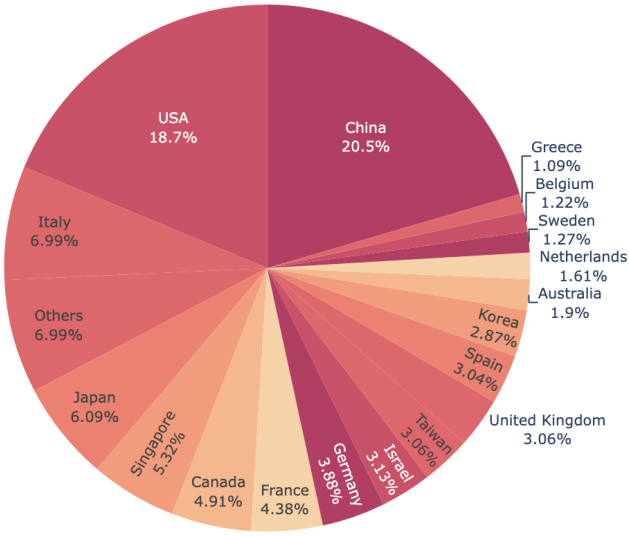


Fig. 8: Figure represents a pie chart to reveal the percentage of publication count by *'come-back'* authors from each countries extracted from the dataset. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

research quality and productivity. The nuanced difference between these observations highlights varying institutional strategies: institutions excelling in re-engagement in academic publication activity may face challenges like resource allocation or mentoring, limiting productivity, whereas those driving higher outputs may provide concentrated support to fewer researchers. This underscores the need for institutions to strike a balance between enabling researcher re-engagement and fostering sustained academic productivity.

Figure 6 showcases the types of publications produced by *"come-back"* researchers, represented as a pie chart categorizing outputs into conferences, journals, symposiums, workshops, books, and others. We observe a significant dominance of conference publications, followed by journals, while other types like workshops and books contribute minimally. This may indicates that *"come-back"* researchers predominantly engage in conference venues, likely due to their faster turnaround times and opportunities for networking, which are critical for re-establishing their academic presence. Their higher reliance on conference publications reflects both the time-sensitive nature of re-entry and the necessity to rebuild collaborative networks, while the relatively lower journal output highlights potential challenges, such as stringent review processes and resource constraints, which may hinder publishing in more traditional, high-impact venues.

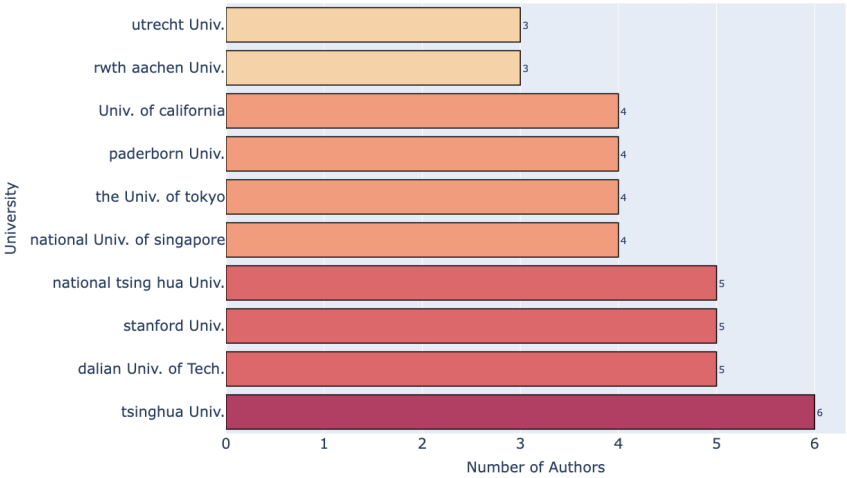


Fig. 9: Figure represents a horizontal bar chart showing the top 10 institutes based on *'come-back'* author count as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of authors.

4.1.3 Drop-out Researchers' Statistics

This section examines three key statistics: (a) the distribution of researchers across various countries and institutions, (b) the distribution of publications across these countries and institutions, and (c) the types of publication venues (e.g., conferences, journals, workshops, etc.) within the dataset, to analyze the demographics of *drop-out* researchers whose bibliographic data has been utilized in this study.

In figure 11, we present a pie chart analyzing the geographic distribution of *"drop-out"* researchers. It reveals significant disparities among countries. The chart indicates that developed nations such as the United States and China dominate in the number of drop-outs, reflecting their larger research communities. However, the proportional representation of drop-outs is also notable in developing regions, which underscores systemic challenges like improper allocation of funding, precise mentorship, and institutional support in those areas. It suggests that while a higher absolute number of drop-outs in developed countries aligns with their expansive research systems, the relatively high proportions in underrepresented regions highlight a critical need for targeted interventions.

We observe the geographic distribution of publications by *"drop-out"* researchers in figure 12, which complements the findings from figure 11. While figure 11 focuses on the distribution of *"drop-out"* researchers themselves across countries, this observation delves into their publication outputs. Both figures

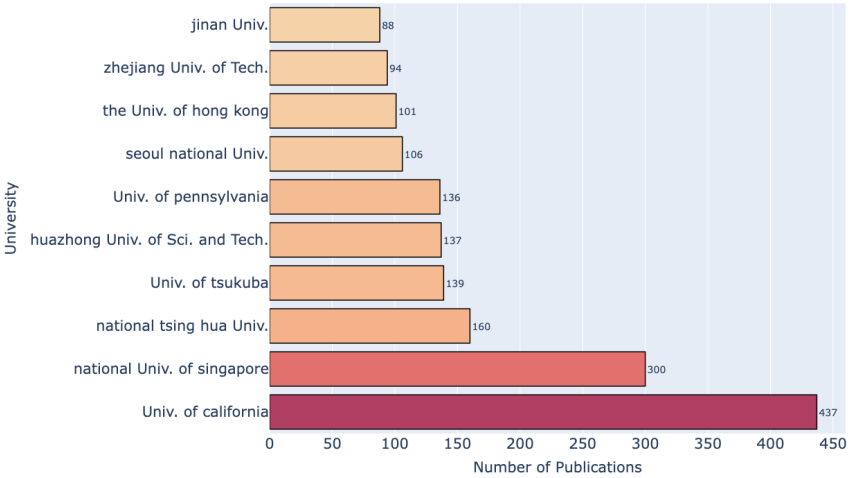


Fig. 10: Figure represents a horizontal bar chart showing the top 10 institutes based on publication count contributed by the *'come-back'* authors as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of publications contributed by the *'come-back'* authors.

highlight that developed nations, such as the United States and China, dominate in terms of both drop-out numbers and publication outputs. However, we insight a key difference that although these countries have higher absolute numbers of drop-outs, they also maintain substantial publication outputs, suggesting that even researchers who eventually leave academia in these regions contribute significantly to scholarly work. Conversely, in developing regions, lower shares in both figures underscore systemic challenges such as inadequate funding and institutional support. Together, these insights suggest that while drop-outs from developed countries tend to leave behind a more substantial academic footprint, those from underrepresented regions face barriers that limit their productivity and contributions before exiting. This comparison highlights the critical need for global policies addressing disparities in research ecosystems to foster both retention and impactful participation.

In figures 13 and 14, we observe the institutional distribution of *"drop-out"* researchers and their publication contributions, respectively. Figure 13 highlights the top 10 institutions with the highest number of drop-out researchers, showcasing that prominent institutions, particularly from the United States, China, and a few European countries, dominate. Figure 14, on the other hand, reveals that while these institutions also lead in publication outputs by drop-outs, although their productivity levels vary significantly. Their is a very crucial

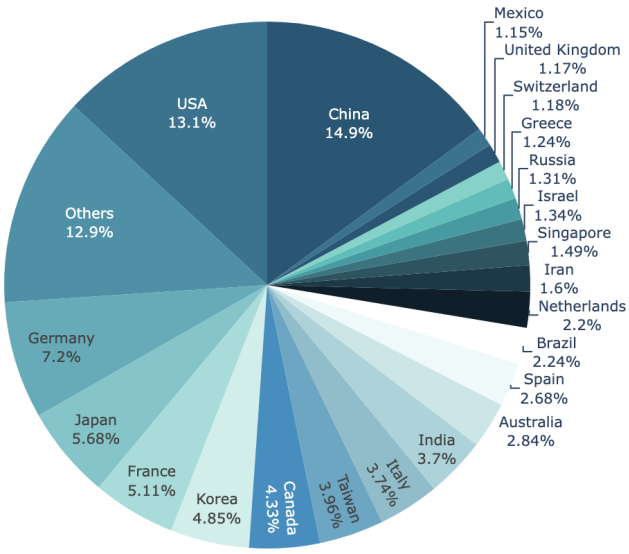


Fig. 11: Figure represents a pie chart to reveal the percentage of *'drop-out'* authors from each countries segregated from the all authors in the dataset. In the left side of the pie chart shows the list of top countries indicating the exact number of *'drop-out'* authors from each country within the chart. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

observation is that certain some institutions with high drop-out numbers contribute disproportionately low publication counts, suggesting that structural factors such as institutional policies, mentorship quality, and research support may play a role in influencing researcher retention and productivity. In contrast, institutions with fewer drop-outs but higher publication outputs may focus on fostering high-impact research and offering better support systems. Together, these figures underscore the critical importance of targeted interventions, such as enhanced mentorship and resource allocation, to address the challenges faced by early-career researchers, particularly in institutions with high drop-out rates, to maximize both retention and scholarly contributions.

Figure 6 shows the categorizes of the publication produced by *"drop-out"* researchers. In the pie chart, it illustrates a significant preference for conference proceedings, followed by journal articles, with other types such as workshops, books, and symposium contributing minimally. This trend suggests that drop-out researchers tend to focus on venues with quicker dissemination timelines, such as conferences, likely due to the time-sensitive nature of their academic engagement or resource constraints. The relatively lower representation in journals and other high-impact formats highlights potential challenges such

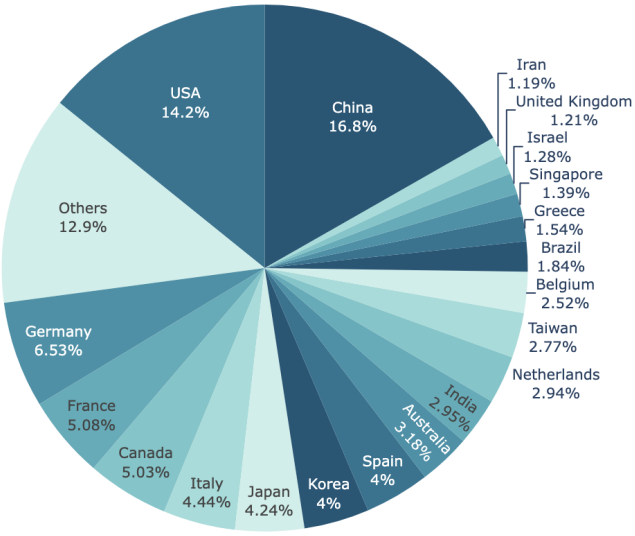


Fig. 12: Figure represents a pie chart to reveal the percentage of publication count by *'drop-out'* authors from each countries extracted from the dataset. In the left side of the pie chart shows the list of top countries indicating the exact number of publication count contributed by *'drop-out'* authors from each country within the chart. The countries contributing $\leq 1\%$ of share in the pie chart, we have marked them as "others".

as limited access to mentorship, longer review processes, or resource-intensive requirements for journal submissions. The findings imply that while drop-out researchers contribute to the academic ecosystem, their outputs often lean toward shorter-term or preliminary findings rather than sustained, high-impact work. Addressing these disparities through support for journal submissions, enhanced collaboration opportunities, and improved access to resources may help improve the research footprint of at-risk early-career researchers before disengagement.

Insight of Academic Demographics

The pair-plots observed in figure 15 and 16 reveal the relationships between the number of unique affiliations, publications, citation counts, and h-index values for *"come-back"* and *"drop-out* researchers. It provide valuable insights into their academic productivity and affiliations. Both groups predominantly show only one unique affiliation, indicating institutional stability and suggesting that authors tend to remain affiliated with a single institution during their active periods. However, there are authors with multiple affiliations, hinting at potential collaborations or shifts in institutional support. The findings suggest that a single affiliation may not sufficiently foster diverse academic exposure

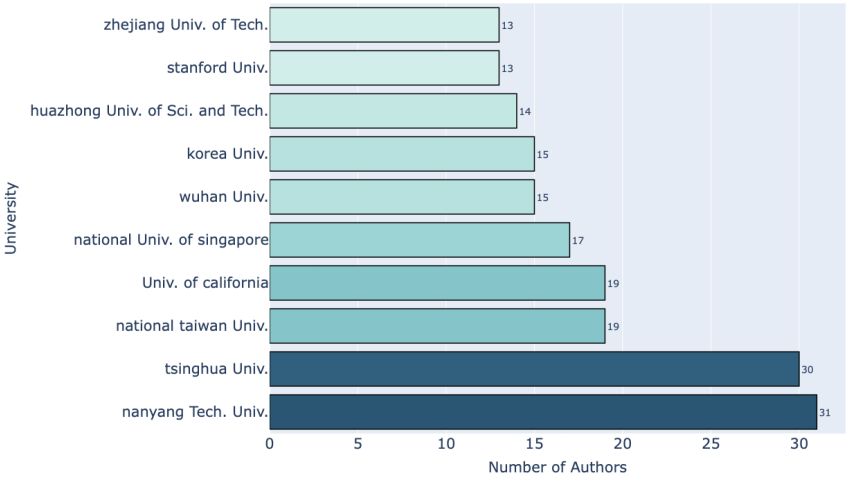


Fig. 13: Figure represents a horizontal bar chart showing the top 10 institutes based on *'drop-out'* author count as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of authors.

for *drop-out* researchers, potentially limiting their career growth. In the context of *come-back* researchers, multiple affiliations may serve as a valuable resource for re-entry, enhancing access to collaborative networks and institutional resources. A clear positive correlation is observed between the number of papers published, citations received, and h-index in both groups, as expected, since these metrics are interdependent. However, there are notable differences in the patterns between the two groups. Comeback authors exhibit a more concentrated cluster in the mid-to-high range for productivity metrics (publication count, citation count, and h-index), indicating a consistent level of impactful contributions. This suggests that when authors return to academic work after a period of inactivity, they often contribute meaningfully to their fields, reinforcing the importance of supporting researchers in re-establishing their careers. In contrast, *drop-out* authors demonstrate a broader distribution, with many concentrated in the lower productivity range. This may reflect challenges faced by *drop-out* researchers, such as reduced resources, diminished academic engagement, or other external factors contributing to their discontinuation. However, some outliers among *drop-out* authors indicate exceptional contributions, showing that even those who leave academia can sometimes leave a significant mark. These patterns underline the importance of fostering an environment that retains productive researchers and provides pathways for re-entry for those who take a hiatus. Supporting *come-back* authors, given their propensity for meaningful contributions, and addressing the challenges faced

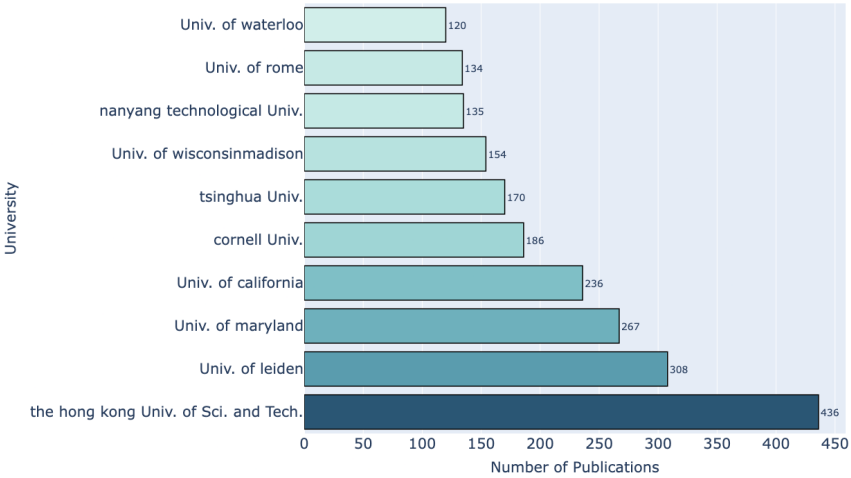


Fig. 14: Figure represents a horizontal bar chart showing the top 10 institutes based on publication count contributed by the *'drop-out'* authors as extracted from our dataset. In figure, along y-axis, we represent the institute name, and along the x-axis, it shows the number of publications contributed by the *'drop-out'* authors.

by *drop-out* researchers can strengthen the overall academic ecosystem. Finally, we outline some interesting takeaway from this academic demographics:

- There is a distinct disparity between top universities by count of author and number of publication contributed by those authors. We observe that comparatively more concentration of researchers in Asian and south Asian universities on the other hand in terms of number of publication, USA, Canada and UK are predominant.
- It is prominently observable that the *drop-out* researchers from top universities produce an ample amount of publication contribution in their respective domain which and might be the indication of their quality of research. We intuit that these researchers either moving to their industry portfolio after finishing their graduation from the institute or they are dropping out from due to the competitive environment of top universities.
- In the context of Asian and south Asian countries *come-back* and *drop-out* statistics, in India the *come-back* researchers are quite low(1.74%) in terms of number of authors and the number of publication contributed by them is merge. On the other hand, the *drop-out* researchers are almost 3.7% in terms of number of researchers count and $\approx 2.95\%$ of the total publications by the *drop-out* researchers are contributed by Indian *drop-out* researchers. However neighborhood country like China, there is a parity with similar statistics in

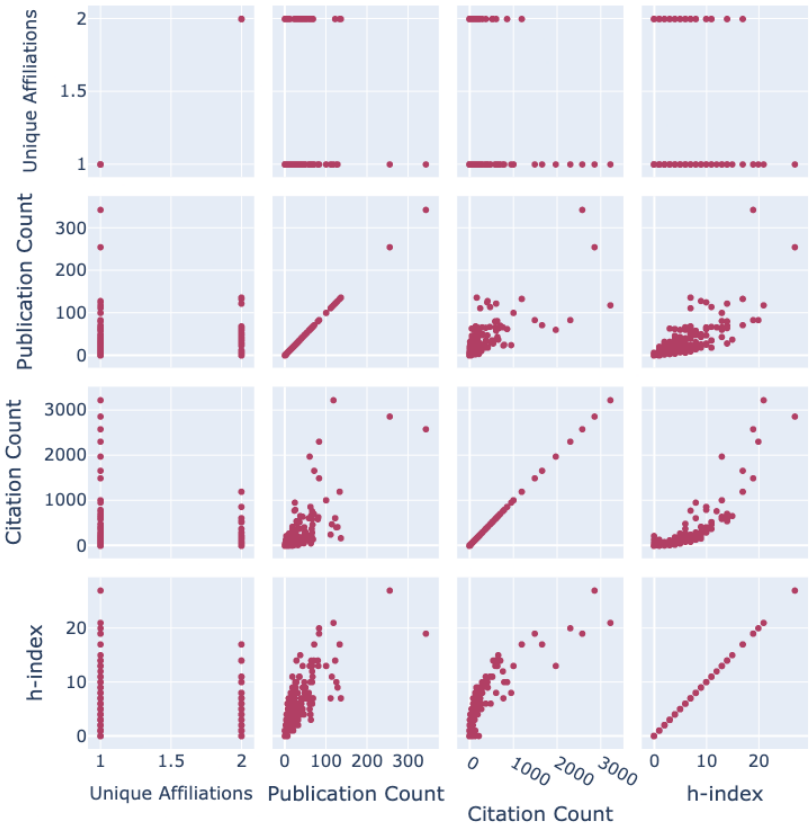


Fig. 15: Figure represents a pair-plot (correlation) among the No. of unique affiliation, No. of publication, citation count, and h-index of the *come-back* authors.

come-back and *drop-out*. Interestingly, Japan has reverse scenario where the *come-back* researcher count is higher than its *drop-out*.

4.2 Analysis of *C*ome-back Researcher

In this section, we study the researcher's nature of come-back after the publishing discontinuation period. We specifically observe the distribution of come-back researchers after the discontinuation. We also study their publication distribution after their come-back. We restrict the observation in the initial 10 years of their research tenure. Further, we also study the collaboration pattern of the come-back researchers. To do it, we observe the association of the come-back researchers with their collaborators based on the popularity index.

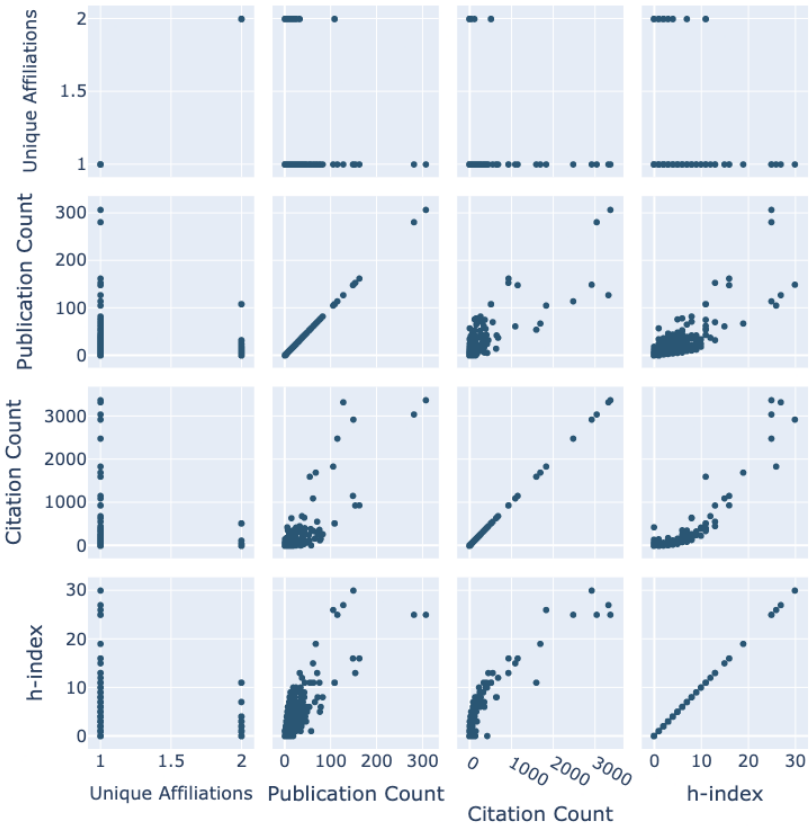


Fig. 16: Figure represents a pair-plot (correlation) among the No. of unique affiliation, No. of publication, citation count, and *h*-index of the *drop-out* authors.

Distribution of come-back researcher:

In this section, we study the distribution of come-back researchers. We further propose a model as given in eq. 1, which is able to mimic the distribution nature. We observe in figure 17 the temporal probability distribution $p(CB, n)$ of researchers resuming academic publishing after a publication discontinuation period. Here, n represents the number of years after a three-year discontinuation. It shows the comparison of the empirical result of the come-back statistics with the proposed model to leverage the nature of come-back. In the figure 17, we represent the year 1 as the year of publication of the come-back researcher with a new publication after the 3 years of publishing discontinuation. In the proposed model, $p(CB, n)$ represents the probability distribution of the researcher's come-back probability where we consider n(year) as the

model parameter and $\alpha = 0.61$ & $\beta = 0.48$ as the two model constants.

$$p(CB, n) = \alpha * e^{-\beta n} \tag{1}$$

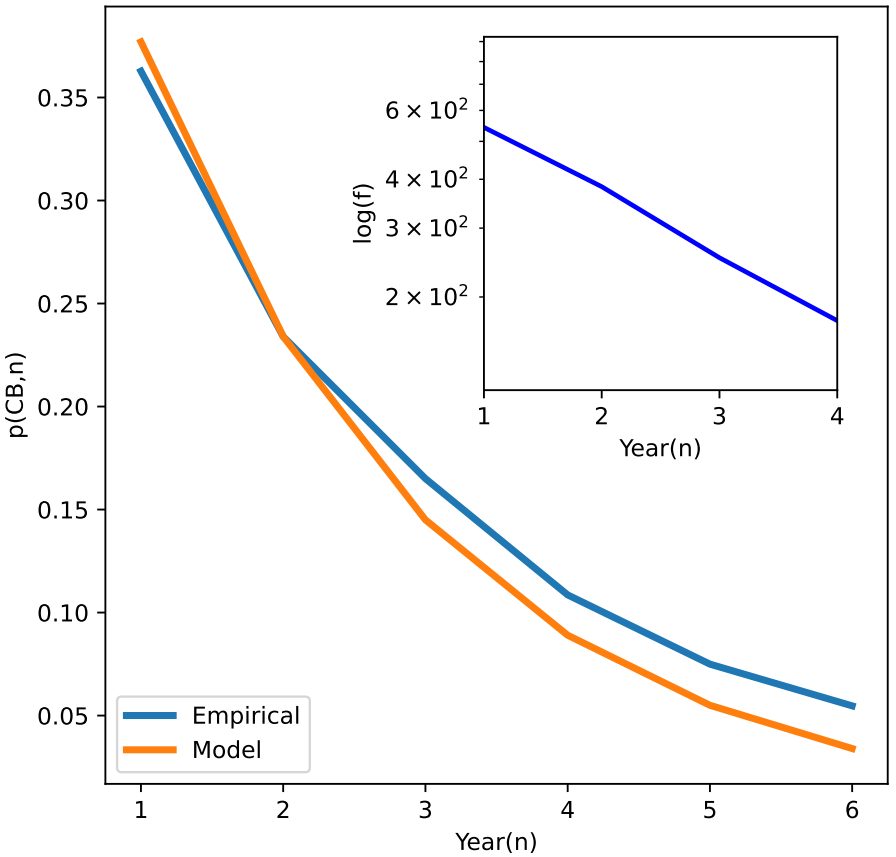


Fig. 17: Figure represents the comparison of the probability distribution of empirical result and model ($p(CB, n)$, given in eq. 1, RMSE=0.017) of come-back researchers. The y-axis represents the probability of coming back after the discontinuation of research work, while the x-axis shows the year they come back after the discontinuation. We also plot the same in the log-normal scale in the inset of the figure.

The proposed model closely aligns with the empirical data (RMSE = 0.017), illustrating that the likelihood of returning decreases exponentially as the gap in publishing lengthens. The inset log-normal plot highlights the steady decline in probability, emphasizing the diminishing chances of reentry with prolonged absences. Though the come-back process steadily slows down across

the year, the come-back happens over the year in the early career stages. This nature could be explained as the challenges of publication in the early career stage of the researchers, where it takes a longer period to get a publication at a reputed venue.

Publication Distribution of come-back Researcher:

In this section, we analyze the publication patterns of researchers who returned to their work after a period of publication discontinuation. We specifically observe the number of publications during the tenure of the first year (after the publication discontinuation of three years) to the end of the first 10 years of her career. In figure 18, we consider year 1 for those come-back researchers who have published at least one research work in the beginning year of their publication activity and afterward have 3 consecutive nonpublishing years. Figure 18 presents a box plot illustrating the publication distribution of researchers who returned after a break. The plot displays the minimum, first quartile, median, third quartile, and maximum number of publications for each year after their return. The box plot captures that the publication output of come-back researchers is skewed towards lower counts, with the median number of publications remaining below 10 for most years. This decline highlights the challenges faced by researchers in re-engaging with the academic ecosystem after a pause in publishing activity. In figure 18, we also observe that the inter-quartile range narrows significantly after the third year of return, reflecting stabilization in publication trends, though at lower levels. It reflects that researchers returning sooner after discontinuation ($n \leq 2$) exhibit marginally higher productivity than those returning after longer gaps. We observe that the come-back researchers tend to produce a low number of publications when the discontinuation tenure increases. This nature reflects the disassociation from the research activity at the early career stage and a longer period of discontinuation could be identified as a signal of some major issues in his/ her research activity such as limited access to resources or weaker institutional support. It can be inferred that the researchers with shorter discontinuation periods tend to reintegrate more effectively into academic publishing, possibly due to maintained expertise and professional connections.

Figure 19 shows the observed probability distribution of publications $p(CB, n, k)$ from the empirical result, where, n represents the number of year and k represents the number of publication. In the figure 19, we observe that though the probability of low publication could be comparatively high, it tends to increase with discontinuation tenure. On the other hand, the probability of a high number of publications decreases with increased discontinuation tenure. Figure 19 shows that the probability of low publication counts ($k \leq 5$) remains significantly high, particularly for researchers returning after longer gaps ($n > 3$). Conversely, the likelihood of high productivity ($k > 15$) diminishes sharply as n increases, indicating that extended absences negatively impact research output. For $n \leq 2$, the probability of $k > 10$ publications is noticeably higher, reflecting the relative success of early-return researchers.

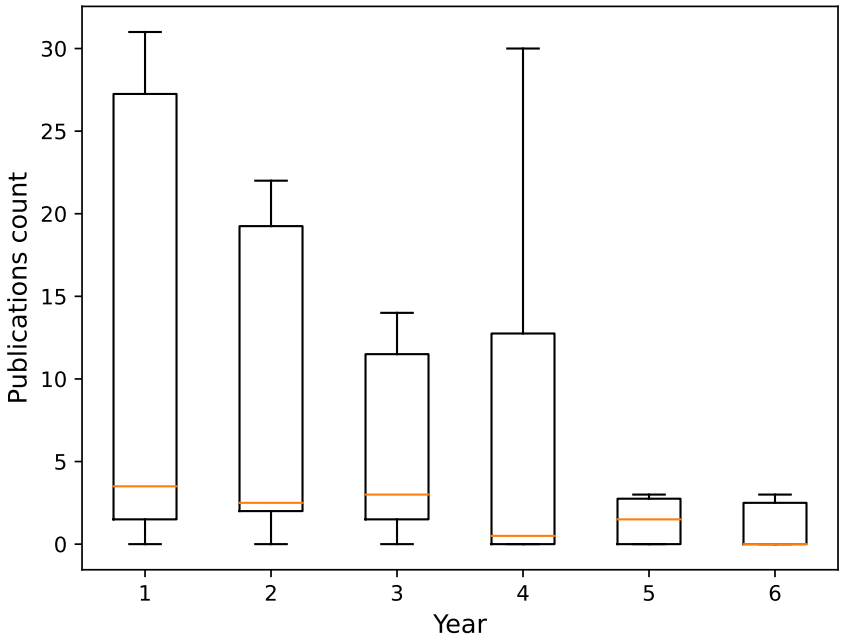


Fig. 18: Figure represents the publication distribution of the come-back researchers across the year after their discontinuation. Along the y-axis, we represent the number of publications of come-back researchers after the discontinuation tenure. The x-axis represents the come-back year after the discontinuation.

The publication distribution exhibits a long tail for higher k values, indicating a small subset of highly productive researchers who resist the overall trend. This overall situation signifies that though the come-back researchers break the discontinuation with their "come-back" publication in the early career stage, there exist few come-back researchers who are productive. For the majority, the extended absence negatively impacts their ability to resume high levels of research output.

Popularity of the collaborators of the Come-back Researchers:

In this section, we study the collaboration relationship of the come-back researchers with their collaborators. In this context, we study the attachment probability of come-back researchers with their collaborators concerning their popularity index. This study reveals the nature of the collaboration of the researchers who come back after some certain discontinuation period. We specifically study the same concerning *pi* of the researchers. To understand the attachment probability of the researchers at a specific popularity index, we represent the attachment probability in a heatmap.

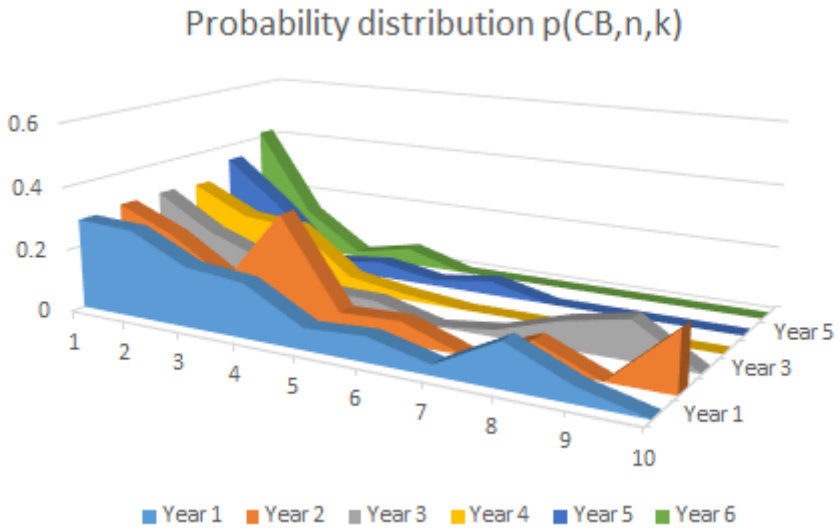


Fig. 19: Figure represents the probability distribution ($p(CB, n, k)$) of publications of the come-back researchers across the year after their discontinuation. The y-axis represents the probability value. On the other hand, the z-axis, represents the come-back year after the discontinuation. Along the x-axis, we represent the number of publications after the come-back.

In figure 20, we observe the attachment probability of the drop-out researchers with their collaborators concerning pi. The range of pi value for the come-back researchers and their collaborators are $(0-385)$ and $(0-3825)$ respectively. We have categorized the come-back researchers and their collaborators based on the distribution of the pi values. The collaboration patterns of come-back researchers, as shown in Figure 20, highlight distinct tendencies in their association with collaborators of varying popularity levels. Researchers with lower popularity indices predominantly collaborate with peers of similar stature. In contrast, those with higher popularity indices exhibit a more selective pattern, showing a reduced likelihood of partnering with less prominent collaborators. This disparity suggests that reputation and influence in academia play a role in shaping collaborative networks, even after a researcher's return from a break.

4.3 Analysis of *d*rop-out Researcher

In this section, we study the drop-out researchers. We find in our study that $\approx 90\%$ of the researchers dropped their research activity at their early career stage. In this section, we initially observe the drop-out researchers' distribution concerning time. We further analyze their publication distribution which they have done before dropping their research career. We also study the nature of their collaboration in terms of the collaborator's popularity index.

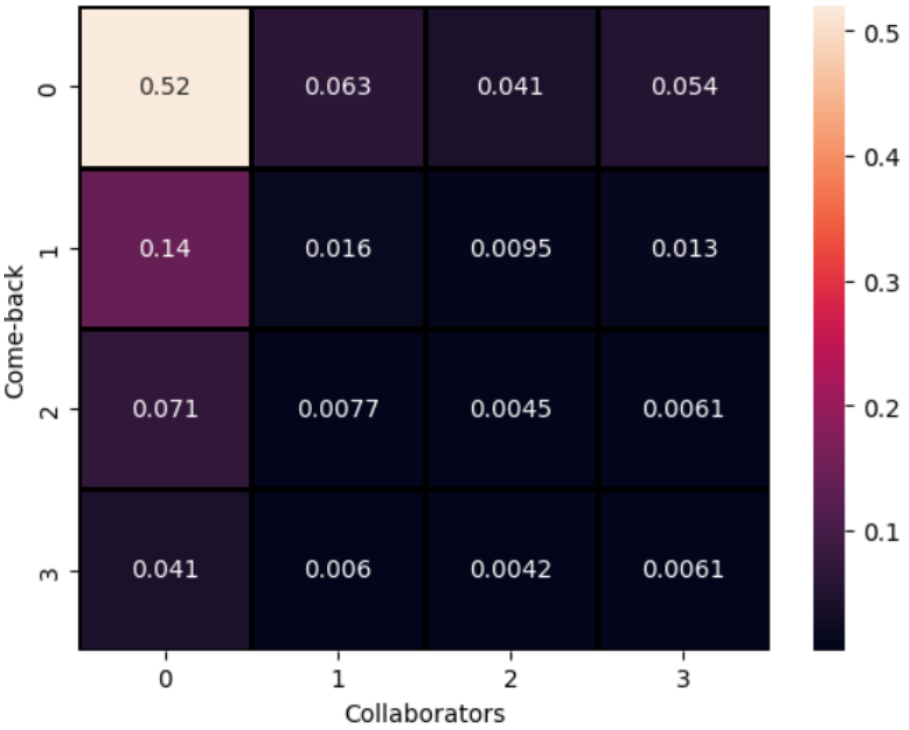


Fig. 20: Figure represents the heatmap based on the popularity index (pi) of the researchers. Along the row, we plot the come-back researchers and along the column, we plot their collaborators. Each cell represents the attachment probability of the come-back researchers with their collaborators for the concerned pi value.

Distribution of Drop-out researcher:

In this section, we observe the distribution of drop-out researchers. Since our focus is on the early career stages, we limit our analysis to the author dataset, identifying dropouts within the first 10 years of a research career. Figure 21 presents the empirical results, displaying the distribution of dropout researchers based on the length of time they were active in research before leaving the community. We further propose a model as given in eq. 2 to understand the dynamics and show the comparison in figure 21.

$$p(DO, n) = \gamma * e^{-\delta n} \tag{2}$$

In the proposed model, $p(DO, n)$ represents the probability distribution of the researcher's drop-out probability where we consider n(year) as the model parameter and $\gamma = 0.99$ & $\delta = 0.69$ are the two model constants. Here, n represents the number of active research years before publication discontinuation.

The model achieves a root mean square error (RMSE) of 0.123, confirming its robustness for longer tenures ($n \geq 4$).

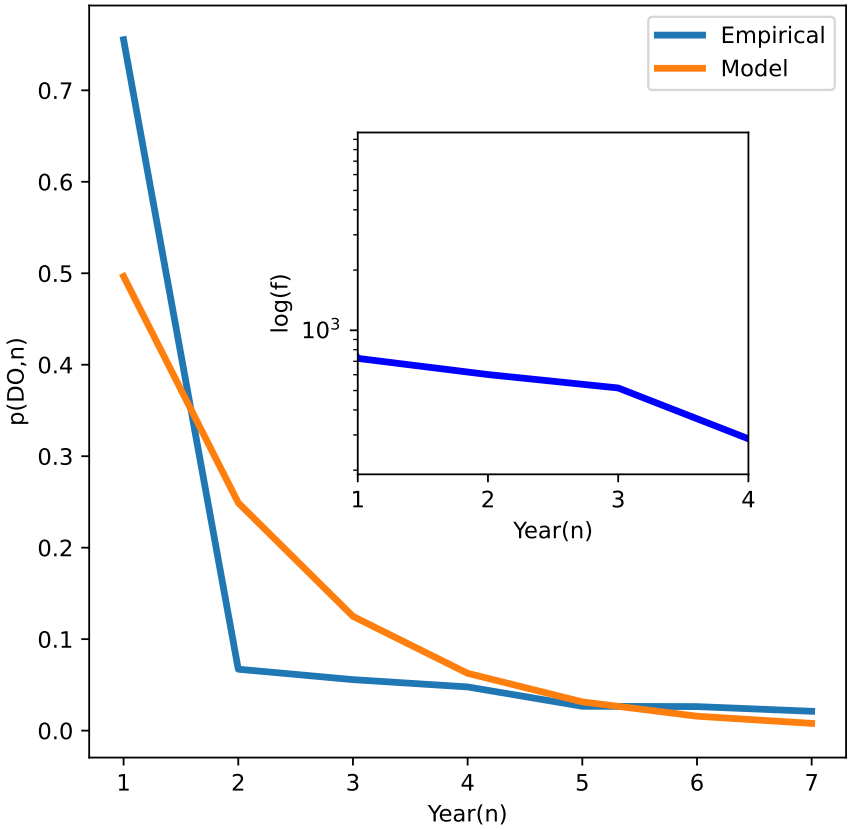


Fig. 21: Figure represents the distribution of drop-out researchers across the year. Along the y-axis, we represent the number of authors who drop out of the research work. The x-axis represents the year in which they dropped out of the research community after doing active research activity. We also plot the same in the log-normal scale in the inset of the figure. We also show the proposed model in the graph and the RMSE value of the Model with the empirical result is 0.123.

In the figure 21, we observe that though the proposed model is not closely comparable for very early drop-out researchers, it can perfectly model the researchers who drop out of their career after 4 or more than 4 years of work. The empirical result reflects that the probability of drop-out decreases sharply after 2 years. This could be an indication of the disassociation of researchers from research activity after the completion of their graduation. On the other hand, there is a steady decrement in drop-out probability after 2 years. This

nature shows that a longer period of work could be an indication of settlement in their research career. These results emphasize the critical challenges faced by early-career researchers, particularly in the initial phase of their academic journey. Support mechanisms during this period may help reduce drop-out rates. The above observation reveals that early-career dropout is strongly associated with the initial adjustment period, suggesting that external support mechanisms (e.g., mentorship programs, and funding opportunities) during the first three years could significantly reduce attrition rates. on the other hand, the gradual decline in dropout probability for researchers with longer tenures implies that sustained engagement in research correlates with career stabilization.

Publication Distribution of Drop-out Researcher:

In this section, we analyze the publication distribution of the drop-out researcher. We specifically study the publication distribution of them before they drop out at their early career stage. In figure 22, the publication patterns of researchers prior to their dropout. The publication counts are analyzed over their active years (n). We observe that across all n, the publication count exhibits a broad range, with a consistent trend of low productivity as n increases and median publication counts remain low (< 10), and the interquartile range shows minimal variation, particularly after the fourth year. The figure also shows that though the ranges of publication count vary across the year, the average number of publications of the drop-out researchers becomes significantly low and steady across the years of active research period before their drop-out. This nature deduced that the drop-out researchers might face a lot of challenges at the early career stage which is reflected in their low count of publications before they drop out of their career in research. Despite some variability in publication counts, the overall trend suggests that many dropout researchers struggle to achieve significant research output before leaving academia. The low publication counts may indicate challenges such as limited access to resources, ineffective collaborations, or difficulties in navigating the academic publishing landscape. These findings suggest that low research productivity could serve as an early warning indicator for potential drop-outs. It is quite alarming that researchers dropping out after longer active tenures (e.g.,$n > 5$) often show consistently low productivity, underscoring the need for ongoing interventions to address systemic barriers to research performance.

Figure 23 illustrates the observed probability distribution of publications $p(DO, n, k)$ or researchers who dropped out during their early career stages. Here, n represents the number of active research years before dropout, and k denotes the number of publications in those years. In figure 23, we observe that The probability of low publication counts (k) is significantly higher for shorter active tenures (n), indicating that early dropouts tend to publish fewer papers before leaving the research field. Conversely, the probability of high publication counts diminishes sharply with increasing n, reflecting the challenges

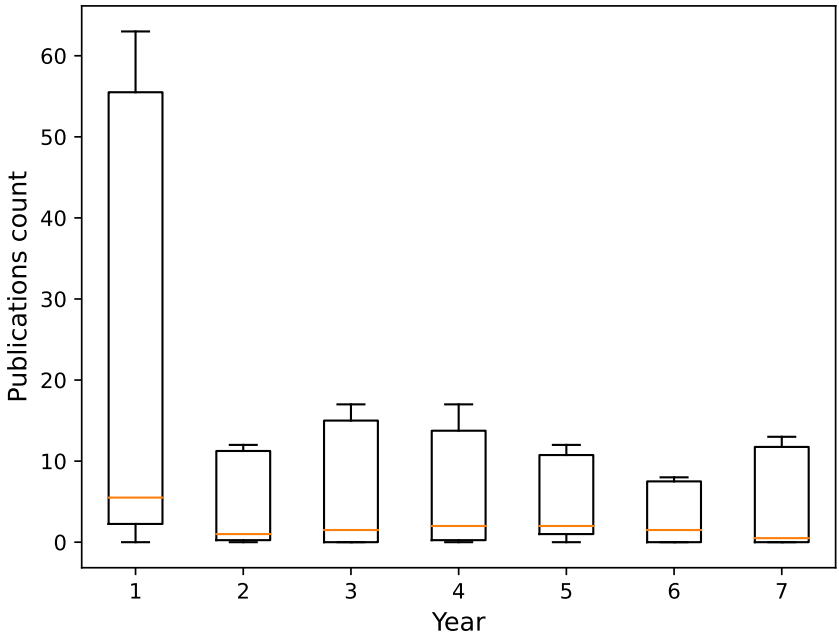


Fig. 22: Figure represents the probability distribution of publication of the drop-out researchers across the year before their discontinuation from the research activity. Along the y-axis, we represent the number of publications of drop-out researchers. Along the x-axis, we represent the number of years of research activity before drop-out.

faced by researchers who struggle to maintain consistent academic output. We further observe that the distribution highlights a steady decline in publication productivity as the researchers approach their dropout year. The most interesting part of the result reflected in figure 23 is that for $n > 5$, a noticeable plateau suggests that even among longer-tenured researchers, a lack of significant publications contributes to eventual disengagement. This overall situation significantly differentiates the drop-out researchers from come-back researchers which could be used as a cue to further predict the possible drop-out researchers from the community. This publication pattern underscores the potential role of publication-related pressures in driving dropout decisions. To mitigate the early-stage dropouts from the research community, it could be suggested that proper interventions, such as mentorship or targeted funding for researchers with low publication rates could be a planned approach.

Popularity of the Collaborators of the Drop-out Researchers:

In this section, we study the attachment of the drop-out researchers with their collaborators concerning their popularity index. This study reveals the nature of the collaboration of the researchers who dropped out from the research

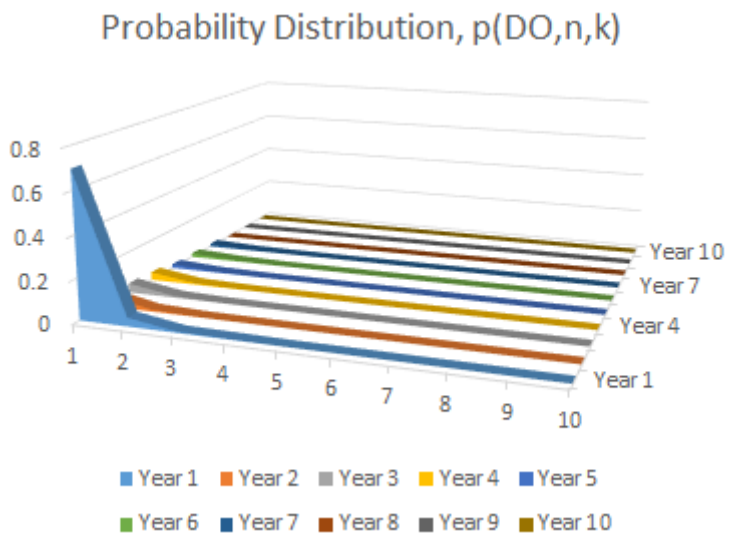


Fig. 23: Figure represents the probability distribution of publication of the drop-out researchers across the year before their dropout from the research activity. The y-axis represents the probability value while, the z-axis, represents the number of years of research activity before drop-out and along the x-axis, we represent the number of publications

activities after certain years. We specifically study the same concerning the *pi* of the researchers. The range of *pi* value for the drop-out researchers and their collaborators are $(0 - 99.49)$ and $(0 - 99.9)$ respectively. In figure 24, presents a heatmap of the attachment probabilities between drop-out researchers and their collaborators, categorized by their respective popularity indices *pi*.

Drop-out researchers exhibit a higher probability of collaborating with highly popular researchers ($pi > 75$). We also observe that despite this collaboration trend, their overall network density remains low, suggesting limited collaborative diversity. The result also shows that drop-out researchers collaborating with high-*pi* individuals often exhibit lower personal popularity, implying asymmetric relationships that may lack mutual reinforcement. On the other hand, the attachment probability sharply declines for collaborations between two low-*pi* researchers, indicating a possible systemic barrier to sustaining such partnerships. This could be inferred from the above observation that reliance on a few high-profile collaborators might hinder skill diversification and network growth for drop-out researchers. Thus to foster more inclusive and resilient academic networks, balanced collaborations, especially among less popular researchers is encouraged for the overall development of the research community.

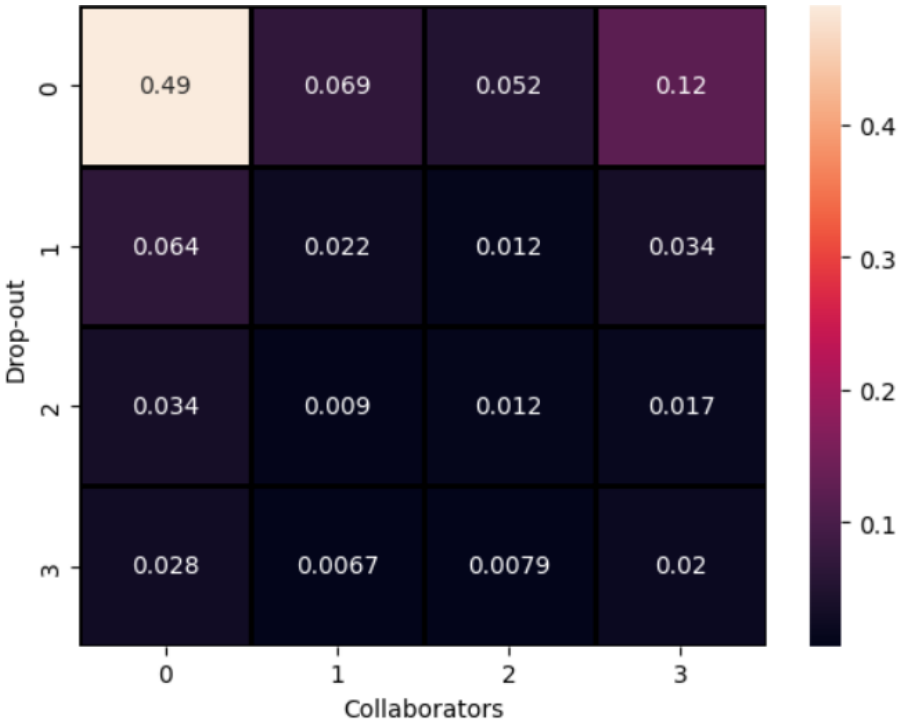


Fig. 24: Figure represents the heatmap based on the popularity index(pi) of the drop-out researchers. Along the row, we plot the drop-out researchers and along the column, we plot the collaborators of the drop-out researchers. Each cell represents the probability of attachment to the drop-out researchers and their collaborators for the concern pi value.

Academic performance difference of 'come-back' & 'drop-out': Although the nature attachment probability of the drop-out researchers with their collaborators is more or less the same as that of the come-back researchers, we find a comparatively higher probability of attachment of the drop-out researchers with high popularity valued collaborators. This nature could be explained as the dissimilar collaboration among the researchers at their early career stage which tends to discontinue later on. The impact of dissimilar collaboration can lead to dropout activity in the early career stage. The distinction between come-back and drop-out researchers during their early career stages is statistically evident in their productivity, collaboration networks, and temporal behavior. Drop-out researchers, comprising approximately 90% of discontinued researchers in the dataset, exhibit significantly lower publication outputs before cessation, with a median publication count consistently below 10 across active years. Their collaborations are marked by a reliance on highly popular individuals, yet their overall network diversity remains limited, suggesting asymmetric and unsustainable partnerships. Conversely, come-back

researchers demonstrate a higher average collaboration count (3.16 vs. 2.85) and more balanced networks, leveraging both low- and high-profile connections for reentry. Temporally, the likelihood of dropout peaks sharply within the first two years, while come-back probabilities are highest in the first year post-discontinuation, declining exponentially thereafter. This contrast highlights that early re-engagement and robust collaborative ecosystems are critical for sustained academic activity, whereas the lack of support and low publication rates are strong predictors of permanent dropout. These findings emphasize the need for interventions, such as targeted mentorship and funding, to foster inclusion and mitigate attrition during this vulnerable career phase.

5 Discussion and Conclusion

In the pursuit of understanding the complex phenomenon of research discontinuation among early-career researchers, this study has harnessed the power of a large-scale bibliography dataset to shed light on the intricacies of their academic journeys. This work presents crucial patterns and trends, providing a deeper understanding of the factors influencing early-career researchers' decisions to either return to research (come-back) or permanently withdraw (drop out) from the academic arena. The results highlight clear disparities in publication productivity, collaboration patterns, and re-entry probabilities between these groups, emphasizing systemic and individual factors influencing career trajectories. Drop-out researchers consistently exhibit low publication outputs and rely on asymmetric collaborations with high-profile individuals, reflecting limited productivity and restricted access to diverse networks, which are major predictors of academic attrition. In contrast, come-back researchers show slightly higher productivity and benefit from balanced collaboration networks, particularly when their discontinuation periods are short. The likelihood of re-entry is highest within the first year post-discontinuation and decreases exponentially, while dropout rates peak within the first two years, underscoring the critical challenges of the early career stage. These findings corroborate previous research, which highlights the significance of early interventions, supportive networks, and sustained research momentum in shaping ECR trajectories, emphasizing the need for targeted strategies to foster retention and re-engagement. The study's limitations include reliance on the AMiner dataset, which restricts generalizability across disciplines, and the use of simplified metrics like publication count and popularity index, which overlook qualitative aspects such as research impact and interdisciplinarity. The dataset does not provide information on the researcher's status (e.g., Ph.D. candidate or Ph.D. holder) as well as the status of their publication(e.g., academic research work or industry research work), limiting insights into career stages. The focus on the first 10 years of academic careers may miss long-term trends, and while correlations are identified, causality is not established. Additionally, the dataset lacks contextual factors such as funding, mentorship quality, and geographic influences, which could significantly impact research discontinuation. Based on

the findings, we recommend establishing structured mentorship programs to guide researchers through their initial years and introducing targeted funding for those with low productivity or narrow collaboration networks. Institutional platforms should foster equitable and diverse collaborations, while reintegration programs, such as grants or workshops, can support researchers returning after a gap. Predictive models can identify at-risk individuals early, enabling tailored interventions. Additionally, inclusive policies addressing structural barriers and resource inequities can promote a diverse and resilient academic community, enhancing the contributions of early-career researchers to science and innovation. Through the analysis of large-scale bibliographies dataset, we have gained valuable insights into the factors contributing to early-career researchers' decisions to discontinue their research pursuits and the potential pathways for their re-engagement with the academic community. The low publication rates and asymmetric collaboration patterns emerge as significant predictors of discontinuation among ECRs. Drop-out researchers often lack diverse networks and struggle to maintain consistent productivity, making them more susceptible to disengagement. On the other hand, the probability of re-entry is highest within the first year post-discontinuation, emphasizing the importance of timely interventions. Successful reintegration is strongly linked to shorter discontinuation periods, balanced networks, and higher pre-discontinuation productivity. We observe that drop-out researchers exhibit lower productivity, narrower networks, and heavier reliance on high-profile collaborators, while come-back researchers show more balanced collaborations and slightly higher productivity, particularly after short gaps.

This work makes it evident that research discontinuation is a multifaceted issue that warrants continued attention and action. As research discontinuation remains a pertinent challenge in the scholarly landscape, this research contributes to the collective effort to retain and rejuvenate early-career research talent, with the ultimate aim of advancing knowledge and societal progress. This work can be further extended to propose a closed-form model including the impact of other factors e.g. venue of publication, number of citations, etc. which might be useful to predict the possibility of drop-out of the researcher at their early career stage. This prediction can be useful for tailoring research policy and funding promising researchers at early career stages and it may be applicable for the termination of 'mass' drop-outs from the research community by identifying their proper measures. As a future extension of this work, we can expand the dataset to include data from diverse disciplines and regions, capturing a wider range of academic trajectories.

Author contributions. All the authors contributed equally to this work.

Data availability. Used Dataset is publicly available.

Declaration

Conflict of interest. The authors declare that they have no conflict of interest.

References

[1] MS Windows 10 Kernel Description. https://www.imf.org/en/Blogs/Articles/2021/10/06/blog-ch3-weo-why-basic-science-matters-for-economic-growth. Accessed: 2023-05-23

[2] Abbasi, A., Altmann, J., Hwang, J.: Evaluating scholars based on their academic collaboration activities: two indices, the rc-index and the cc-index, for quantifying collaboration activities of researchers and scientific communities. Scientometrics **83**(1), 1–13 (2010)

[3] Melin, G.: Pragmatism and self-organization: Research collaboration on the individual level. Research policy **29**(1), 31–40 (2000)

[4] Wagner, C.S., Park, H.W., Leydesdorff, L.: The continuing growth of global cooperation networks in research: A conundrum for national governments. PloS one **10**(7), 0131816 (2015)

[5] Nicholas, D., Watkinson, A., Boukacem-Zeghmouri, C., Rodríguez-Bravo, B., Xu, J., Abrizah, A., Świgoń, M., Herman, E.: Early career researchers: Scholarly behaviour and the prospect of change. Learned Publishing **30**(2), 157–166 (2017)

[6] Gayen, A., Mukherjea, T., Jana, A.: Characterization of 'early career stage'researchers from their co-authorship network. In: 2023 IEEE Guwahati Subsection Conference (GCON), pp. 1–6 (2023). IEEE

[7] Gayen, A., Chakraborty, S., Chakraborty, N., Jana, A.: How can i grow!!: Analysis of co-authorship network of early career stage researchers. In: International Conference on Smart Computing and Communication, pp. 341–351 (2024). Springer

[8] Gayen, A., Mukherjea, T., Mukherjee, S., Jana, A.: Identification of 'early career stage' researchers from large scale bibliographic dataset. In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7 (2023). https://doi.org/10.1109/ICCCNT56998.2023.10307427

[9] Nicholas, D., Watkinson, A., Boukacem-Zeghmouri, C., Rodríguez-Bravo, B., Xu, J., Abrizah, A., Świgoń, M., Clark, D., Herman, E.: So, are early

career researchers the harbingers of change? Learned Publishing **32**(3), 237–247 (2019)

[10] Evans, M.C., Cvitanovic, C.: An introduction to achieving policy impact for early career researchers. Palgrave Communications **4**(1), 1–12 (2018)

[11] Lopes, G.R., Moro, M.M., Wives, L.K., De Oliveira, J.P.M.: Collaboration recommendation on academic social networks. In: International Conference on Conceptual Modeling, pp. 190–199 (2010). Springer

[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] Sampson, V., Clark, D.: The impact of collaboration on the outcomes of scientific argumentation. Science education **93**(3), 448–484 (2009)

[14] Murray, D.L., Morris, D., Lavoie, C., Leavitt, P.R., MacIsaac, H., Masson, M.E., Villard, M.-A.: Bias in research grant evaluation has dire consequences for small universities. PloS one **11**(6), 0155876 (2016)

[15] Browning, L., Thompson, K., Dawson, D.: From early career researcher to research leader: Survival of the fittest? Journal of Higher Education Policy and Management **39**(4), 361–377 (2017)

[16] Pica, N., Bourgeois, F.: Discontinuation and nonpublication of randomized clinical trials conducted in children. Pediatrics **138**(3) (2016)

[17] Chapman, S.J., Shelton, B., Mahmood, H., Fitzgerald, J.E., Harrison, E.M., Bhangu, A.: Discontinuation and non-publication of surgical randomised controlled trials: observational study. Bmj **349** (2014)

[18] Johnson, A.L., Fladie, I., Anderson, J.M., Lewis, D.M., Mons, B.R., Vassar, M.: Rates of discontinuation and nonpublication of head and neck cancer randomized clinical trials. JAMA Otolaryngology–Head & Neck Surgery **146**(2), 176–182 (2020)

[19] Jadidi, M., Karimi, F., Lietz, H., Wagner, C.: Gender disparities in science? dropout, productivity, collaborations and success of male and female computer scientists. Advances in Complex Systems **21**(03n04), 1750011 (2018)

[20] Larcombe, W., Ryan, T., Baik, C.: What makes phd researchers think seriously about discontinuing? an exploration of risk factors and risk profiles. Higher Education Research & Development **41**(7), 2215–2230 (2022)

[21] Dorenkamp, I., Weiß, E.-E.: What makes them leave? a path model of postdocs' intentions to leave academia. Higher Education **75**, 747–767 (2018)

[22] Arnetminer. https://www.kaggle.com/kmader/aminer-academic-citation-dataset (accessed 20th February, 2021)

[23] Borgman, C.L., Furner, J.: Scholarly communication and bibliometrics. Annual review of information science and technology **36**(1), 2–72 (2002)

[24] Kong, X., Jiang, H., Yang, Z., Xu, Z., Xia, F., Tolba, A.: Exploiting publication contents and collaboration networks for collaborator recommendation. PloS one **11**(2), 0148492 (2016)

[25] Leclerc, M., Gagné, J.: International scientific cooperation: The continentalization of science. Scientometrics **31**(3), 261–292 (1994)

[26] Montoya, F.G., Alcayde, A., Baños, R., Manzano-Agugliaro, F.: A fast method for identifying worldwide scientific collaborations using the scopus database. Telematics and Informatics **35**(1), 168–185 (2018)

[27] Bridle, H., Vrieling, A., Cardillo, M., Araya, Y., Hinojosa, L.: Preparing for an interdisciplinary future: A perspective from early-career researchers. Futures **53**, 22–32 (2013)

[28] Laudel, G., Bielick, J.: How do field-specific research practices affect mobility decisions of early career researchers? Research Policy **48**(9), 103800 (2019)

[29] Wang, Y., Jones, B.F., Wang, D.: Early-career setback and future career impact. Nature communications **10**(1), 4331 (2019)

[30] Allen, C., Mehler, D.M.: Open science challenges, benefits and tips in early career and beyond. PLoS biology **17**(5), 3000246 (2019)

[31] Morriss, L.: Enhancing the visibility of early career researchers in social work. SAGE Publications Sage UK: London, England (2019)

[32] Thomsen, P., Lopesi, L., Leenen-Young, M.: Introduction: Charting provocations and exploring new directions for pacific research in Aotearoa-New Zealand from pacific early career academic (PECA) perspectives. Victoria University of Wellington, Stout Research Centre Wellington, NZ (2021)

[33] Djerasimovic, S., Villani, M.: Constructing academic identity in the european higher education space: Experiences of early career educational researchers. European Educational Research Journal **19**(3), 247–268

(2020)

[34] Ortlieb, R., Weiss, S.: What makes academic careers less insecure? the role of individual-level antecedents. Higher Education **76**(4), 571–587 (2018)

[35] Maer-Matei, M.M., Mocanu, C., Zamfir, A.-M., Georgescu, T.M.: Skill needs for early career researchers—a text mining approach. Sustainability **11**(10), 2789 (2019)

[36] Barkhuizen, G.: Identity dilemmas of a teacher (educator) researcher: Teacher research versus academic institutional research. Educational Action Research **29**(3), 358–377 (2021)

[37] Horta, H., Cattaneo, M., Meoli, M.: Phd funding as a determinant of phd and career research performance. Studies in Higher Education **43**(3), 542–570 (2018)

[38] Castelló, M., McAlpine, L., Sala-Bubaré, A., Inouye, K., Skakni, I.: What perspectives underlie 'researcher identity'? a review of two decades of empirical studies. Higher Education **81**, 567–590 (2021)

[39] Andrews, E.J., Harper, S., Cashion, T., Palacios-Abrantes, J., Blythe, J., Daly, J., Eger, S., Hoover, C., Talloni-Alvarez, N., Teh, L., *et al.*: Supporting early career researchers: insights from interdisciplinary marine scientists. ICES Journal of Marine Science **77**(2), 476–485 (2020)

[40] Skakni, I., Calatrava Moreno, M.d.C., Seuba, M.C., McAlpine, L.: Hanging tough: post-phd researchers dealing with career uncertainty. Higher Education Research & Development **38**(7), 1489–1503 (2019)

[41] Castelló, M., Kobayashi, S., McGinn, M.K., Pechar, H., Vekkaila, J., Wisker, G.: Researcher identity in transition: Signals to identify and manage spheres of activity in a risk-career. Frontline Learning Research **3**(3), 39–54 (2015)

[42] McDowell, G.S., Knutsen, J.D., Graham, J.M., Oelker, S.K., Lijek, R.S.: Co-reviewing and ghostwriting by early-career researchers in the peer review of manuscripts. Elife **8**, 48425 (2019)

[43] Drosou, N., Del Pinto, M., Al-Shuwaili, M.A., Goodall, S., Marlow, E.: Overcoming fears: a pathway to publishing for early career researchers. Disaster Prevention and Management: An International Journal **29**(3), 340–351 (2020)

[44] Krauskopf, E.: An analysis of discontinued journals by scopus. Scientometrics **116**(3), 1805–1815 (2018)