Gender inequality in academia: Opportunities and challenges in the era of globalization and digitalization

Xinyi Zhao

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

Acknowledgements

As I reach the culmination of my Ph.D. journey, I would like to express my heartfelt gratitude to everyone who has made this achievement possible. First and foremost, I am profoundly grateful to my supervisors, Ridhi Kashyap and Emilio Zagheni, whose exceptional guidance and unwavering support have been instrumental in the successful completion of my Ph.D. study and the completion of this thesis. Their invaluable suggestions at the outset provided me with a clear direction and purpose for my Ph.D. journey. Both of them are excellent researchers in Demography and Computational Social Science, and every discussion with them was very inspiring and helpful for framing my research questions on gender disparities and exploring the role of big data in my study. During the collaboration with Ridhi Kashyap in Chapters 3 and 4 and the collaboration with Emilio Zagheni in Chapters 2, 3, and 4, their meticulous feedback on my research and writing significantly improved the quality of my research. Their encouragement was pivotal in helping me publish my findings. I deeply appreciate their dedication and the immense amount of time they invested in my development as a researcher.

I would like to extend my sincere thanks to my coauthors besides my supervisors, Samin Aref (Chapter 2), Guy Stecklov (Chapter 2), Aliakbar Akbaritabar (Chapters 3, 4), Gemma E. Derrick (Chapter 5), Cassidy R. Sugimoto (Chapter 5), and Vincent Larivière (Chapter 5). They are all terrific coauthors and their expertise and dedication have significantly enriched the quality of our work. The collaboration with them has been a truly rewarding experience. I am also thankful to my thesis committee of Charles Rahal and Francesco Rampazzo, who gave me insightful suggestions in my annual assessments and encouraged me to think more deeply about the implications of my work in Sociology.

I had the great fortune to have two superb academic homes - one in Oxford at the Department of Sociology, Leverhulme Centre for Demographic Science (LCDS) and St-Hughs College, and the other in Rostock at the Max Planck Institute for Demographic Research (MPIDR). The facilities, expertise, and supervision provided by the University of Oxford and MPIDR were instrumental in advancing my academic development. In Oxford, I want to acknowledge the rich and supportive academic environment, which created a lively community of scholars that I count myself lucky to be a part of. The diverse and stimulating atmosphere enabled me to explore my research interests and engage with a vibrant academic community of different backgrounds. It helped me find areas of fruitful overlap between demography and sociology. I am particularly grateful for the resources, seminars, and workshops that greatly enriched my understanding and broadened my perspectives. I would like to extend my heartfelt thanks to the excellent researchers at LCDS, who provided invaluable suggestions and insights that greatly enhanced my research. Thank you to friends at the Department and LCDS, Rachel Ganly, Haohao Lei, Naa Adjeley Mensa, Guanghui Pan, Jiaxin Shi, Hanbo Wu, Jiani Yan, Xijia You, and Yan Zhang. Their companionship have made my experience in Oxford immensely rewarding and enjoyable. The talk with Rachel Ganly and Naa Adjeley Mensa helped me spark new ideas in Chapter 5 on Ph.D. mothers and provided me with useful materials for my presentations. A special thank to the Media and Communications Office at LCDS who kindly promoted my research, and the administrative staff at the Department of Sociology, who work diligently to support students and faculty alike.

Being accepted by MPIDR under the team of Emilio Zagheni was a turning point in my life. The institute's commitment to fostering cutting-edge demographic research and its collaborative spirit provided me with unparalleled opportunities for professional growth and development. Thanks to the invaluable support and generous funding from MPIDR, I was able to attend multiple international academic conferences where I could present my work, receive constructive feedback, and engage with leading scholars and peers from around the world. These events greatly broadened my academic horizons and exposed me to the latest and most innovative research in my field. I am grateful to IMPRS-PHDS at MPIDR, particularly Heiner

Maier and Christina Westphal for their support during my Ph.D. journey. Beyond that, I would like to express my profound gratitude to my mentors at MPIDR, Aliakbar Akbaritabar, and Samin Aref for their help during the different Ph.D. stages. Samin Aref introduced me and guided me to the field of Scientiometrics and Aliakbar Akbaritabar further provided me with the tools, insights, and inspiration needed to bridge my study between Scientiometrics and gender disparities. In addition to academic support, their encouragement and advice on my academic plans and personal matters have been equally significant and cannot be overstated. I would like to extend my gratitude to all my colleagues at MPIDR. Their impressive research shaped my understanding of demography and enriched my knowledge and perspective on key demographic issues, enabling me to approach my own research from a deeper and more nuanced perspective. I want to thank my lovely cohort-mates, Carolina Coimbra Vieira, Esther Denecke and Chia-Jung Tsai, and the friends I met through IMPRS-PHDS, Mary Joul Abed Al Ahad and Shuye Yu. Their companionship since joining MPIDR made my Ph.D. life more memorable. I also want thank The friendships and professional relationships I have developed here have enriched my academic journey and personal growth in countless ways. I am also thankful to all the administrative staff at MPIDR, from the receptionist to the IT staff, who ensured that everything ran smoothly, allowing me to focus on my research and studies.

Last but not least, I want to thank my family and friends. They respect and support all the decisions I made, and encourage me to pursue my goals. Their support and companionship have been a constant source of strength and motivation. I am deeply appreciative of your love, and friendship, and for being a steadfast part of my life, no matter where I am.

Gender inequality in academia: Opportunities and challenges in the era of globalization and digitalization

Xinyi Zhao

St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2024

Abstract

This thesis explores the dynamics of gender inequality within scientific production across different dimensions in relation to globalization and digitalization in scientific systems. International mobility of researchers has become an important dimension in the globalized era. **Essay 1** discusses the heterogeneity in the levels of gender disparity in immigration and emigration of German-affiliated researchers and particularly among return migrants. It finds that gender imbalance in the German science system has been intensified as the subgroups who returned to Germany are more male-dominated than the subgroups who left. **Essay 2** extends the discussion to the global scale and provides the first global and dynamic view of gendered patterns of transnational scholarly mobility. While female researchers continued to be underrepresented among internationally mobile researchers, this gender gap has been narrowing at a faster rate than the gender gap in the population of all researchers. Digitalization has changed the way science is practised today. **Essay 3** studies gender differences in online visibility and self-promotion behaviors of early career scholars. Women's publications showed higher online visibility on Twitter, while men were more likely to self-tweet their research and accumulated more additional citations from self-promotion persisting the gender gaps. Finally, establishing a career among highly qualified people has certain implications for demographic life events such as childbirth and parenting. **Essay 4** advances knowledge on the conflict between starting an academic career in terms of Ph.D. attainment and parenting role. Women have been more likely to dedicate themselves to academic pursuits at first and then become parents, and they suffered more from the dual-role conflict in their scientific impacts. This thesis contributes to highlighting how gender gaps in academia persist and shift, which is vital for shaping policies that enable the sustainability of female scientists'

careers and help the global science system benefit from the contributions that highly qualified female researchers could make.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Gender disparity in academia

Women and girls represent half of the world's population and also half of its potential (Madgavkar et al., 2019; Mosse, 1993; Zippel, 2017). However, gender inequality has been a longstanding issue and concern across various domains, including education, employment, and health (Buchmann et al., 2008; Goldin, 2021; Madsen, 2017; Madsen et al., 2015; Seguino, 2000; World Economic Forum, 2024). These systematic gaps stagnate social progress and economic development worldwide. In 1946, the United Nations' Commission on the Status of Women (CSW) was established to promote the advancement of women throughout the world (The Intergovernmental Support Division of UN Women, 2019). "Achieving gender equality and empowering all women and girls" as one of the Sustainable Development Goals (SDGs) reflects the increasing global effort promoting gender equality (Leal Filho et al., 2022). Despite increasing attention and significant progress over the past decades, gender parity remains a distant goal, and inequality against women persists everywhere.

Gender equality in education is one of the most important goals within the framework of SDGs, which has witnessed significant progress in recent years. Girls are getting educated at higher levels today than ever before, and the share of women with tertiary education has risen consistently, reversing the historical gender gap in favour of men (Encinas-Martín and Cherian, 2023). Women are close to reaching gender parity among doctoral graduates, especially in Europe and North America (Korhonen, 2023; Directorate-General for Research and Innovation (European Commission), 2021b). However, when it comes to women's participation in research overall, we continue to see a "leaky pipeline" where only 28% of researchers around the globe were women, based on estimates as of 2015 (Huyer, 2015). Even in European countries, women are still a minority among senior academics including in the Netherlands (18.7%), Germany (19.4%), France (21.9%), Switzerland (23.3%), and the United Kingdom (26.4%) (Directorate-General for Research and Innovation (European Commission), 2019). This indicates that the increasing proportion of women in tertiary education is not necessarily translating into a greater presence in science, and gender gaps widen at the researcher level including full-time faculty and professorship (West et al., 2013).

Highly qualified female academics face a number of barriers to fulfilling their potential. They are hindered from participating in the scientific arena to getting recognition and progression in the research system. Such systematic bias against women in science has been found in salary (Directorate-General for Research and Innovation (European Commission), 2019), training (Venkatraman, 2014), career promotions (Falco et al., 2023), research funding (Van der Lee and Ellemers, 2015), and opportunities to go to academic meetings or sit on committees and panels (Christian, 2018). In terms of credit for research, women have been historically underrepresented in the roles of the first author and last author, who traditionally receive the most credit for the publication (Broderick and Casadevall, 2019; Rexrode, 2016; West et al., 2013). Even when contributing equally to the publication, women are always more likely to rank behind men in authorship (Broderick and Casadevall, 2019). Considering issues around work-family conflict, female researchers are more likely to serve in lead parenting roles and be more engaged in childcare tasks, which inevitably reduces their working hours (Derrick et al., 2022). These conscious or unconscious biases and traditional social norms ultimately lead to male-dominated science lacking gender equality in various facets. Overall, female researchers tend to have shorter, less well-paid careers and receive less recognition in their work than their male counterparts.

Gender equality in science is not a problem that affects just women. It is also a problem that impedes the long-term development of societies, holding back critical advances across multiple aspects of employment, innovation, and the green economy (Bahous, 2022; Nair-Bedouelle, 2023). The imbalanced participation of women and men in science restricts the available pool of expertise capable of addressing intricate natural and social problems, as well as advancing technology and innovation. Specifically, and as one example, inadequate representation of women in the field of Medicine may result in gender-specific health needs being overlooked, leading to subpar healthcare outcomes for women. From an economic perspective, the underrepresentation of women, along with an inadequate distribution of talent, would impede economic growth by causing a decline in productivity and reduced activity in the labour market (Bahous, 2022). The gender imbalance in the pool of available collaborators inhibits the cross-gender collaborations, hence disadvantaging the innovation of scientific output and future improvements (Yang et al., 2022; Zhang et al., 2024). Moreover, a dearth of prominent female role models might restrict the ambitions and possibilities of women and girls who aspire to pursue professions in science, hence perpetuating the cycle of underrepresentation of women in science (Breda et al., 2023; Young et al., 2013).

From “leaky pipeline” and “sticky floor” to “broken ladder” and “glass ceiling”, the obstacles and barriers encountered by female researchers are drawing increased attention from the public. Over the past decades, several entities such as international organizations, regional governments, academic institutions, and feminist organizations, have stepped up efforts to close gender gaps in science through policy and other public action (Muñoz Rojas, 2023). For example, the European Commission is committed to promoting gender equality in research and innovation and has introduced the "Gender Equality Plan" (*GEP*) in Horizon Europe as an eligibility criterion and the mandatory integration of the gender dimension into research and innovation content (Directorate-General for Research and Innovation (European Commission), 2021a). Following this plan, the project "Redesigning Equality and Scientific Excellence Together" (*RESET*) has been launched and led by a consortium of European universities, which aims to place gender equality and diversity at the core of scientific and academic policymaking. The United Nations Educational, Scientific and Cultural Organization (*UNESCO*) is mainstreaming gender equality in its international scientific programmes, such as implementing the project "For Women in Science Programme" with the Fondation L'Oréal to increase the participation of women in solving the great challenges facing humanity and empower them to achieve scientific excellence (Nair-Bedouelle, 2023). Despite the progress towards gender equality in science, there is still a long way to go before achieving a gender-balanced and sustainable science system.

This introductory chapter lays out the main themes that motivate this thesis. Faced with the widely recognized issue of gender disparity in academia, Section 1.2 further discusses this issue in the new era by looking at the opportunities created by globalisation and digitalization for the underrepresented group of female researchers, while work-family continues to be a barrier to their academic careers. Section 1.3 then proceeds to discuss the innovative methods of bibliometic analysis and its combination with other data sources in studying gender disparity in academia across different dimensions. Section 1.4 describes the research questions I aim to address in this article-based thesis, followed by an ethical consideration on the data collection and research design in Section 1.5.

1.2 Rethinking academia in new era

Academia is changing, with opportunities for broader professional participation and wider knowledge dissemination provided by the globalized and digitalized era. International mobility of researchers facilitates the exchange of knowledge, ideas, and skills, and thus contributes to the dynamics and diversity of the global knowledge production system (Bauder, 2015; Netz and Jaksztat, 2017). In light of skill shortages and market needs in the current knowledge-based society, attracting highly skilled people has become increasingly important for countries worldwide (Czaika and Parsons, 2017). Highly skilled individuals, including researchers, appear to be more mobile than the general population (Czaika and Orazbayev, 2018; Docquier and Marfouk, 2004; Drivas et al., 2018). Meanwhile, the advent of digitalization has revolutionized the way we communicate, share information, and connect with others. The emerging digital tools and platforms become invaluable for researchers to achieve academic promotion and increase visibility in cost-effective ways that are unbounded by space and time. These advances in the current era have the potential to create a more inclusive and equal academic environment, but whether this is realised is an empirical question that deserves deeper investigation.

The new opportunities are not always accessible and affordable to everyone, which would interact with the historical roots of gender disparity in science and exacerbate gendered patterns. Furthermore, in addition to technological advancements in science, the new era is also characterized by shifts in the perspectives of women regarding their careers and families. Women with higher education are more likely to have higher levels of occupational aspiration and in particular, women in science expect to advance on the academic career ladder (Drake and Svenkerud, 2023). The road ahead towards gender equality is uphill, and persistent work-family conflict is one of the most difficult barriers to the scientific career trajectories of female researchers, but it may not be the same for male researchers. Women report more time spent on housework and childcare than men (Mason et al., 2013b; Morgan et al., 2021a) and women are also disproportionately affected by other family demands, including elder-care responsibilities and need to accommodate a partner's career (Mody et al., 2021). There is a need to revisit and assess the gender gaps in academia in light of the contemporary context.

1.2.1 Academic migration

In the current era, globalization has facilitated the exchange of knowledge, ideas, and skills and thus contributed to the dynamic development of the knowledge production system through international mobility and global collaboration of researchers (Bauder, 2015; Boekhout et al., 2021; Netz and Jaksztat, 2014; Zhao et al., 2021). International mobility is also increasingly recognized as a key strategy for scientists seeking to participate in global scientific networks and to advance their careers (European Commission and Directorate-General for Research and Innovation, 2018; Weert, 2013). Highly skilled people have become the most mobile population group worldwide (Schiller and Cordes, 2016). In Organisation for Economic Co-operation and Development (OECD) countries, highly qualified individuals including researchers, make up one-third of the immigrant population (Schiller and Cordes, 2016).

On the other hand, international migration can be an important factor that intensifies and amplifies some of the gendered inequalities in contemporary science, and gender inequalities can continue to shape the globalization of science (Zippel, 2017). Previous studies have provided a starting point to examine to what extent the participation of researchers in transnational academic mobility varies by gender. Female researchers were found to be less international than their male colleagues among visiting researchers in Germany, particularly in the natural sciences (Jöns, 2011b). Similar results of gender differences in international scholarly migration were also found in Russia (Malakha, 2002). Despite the increasing attention given to the gender disparity among international mobile researchers, the current scientific knowledge of how these inequalities vary by the origin and destination countries remains limited. Equally importantly, how international migration benefits female researchers and what drives the underrepresentation of female mobile researchers deserve much more attention, in close relation to the development of policy support for women in science.

1.2.2 Online visibility

Digitalization has changed the way people interact with each other. Online social platforms such as Facebook and Twitter (now X), offer a promise of broader professional participation and wider knowledge dissemination, especially for underrepresented groups (Yammine et al., 2018). Getting published is the first step in developing an academic career, and sharing and distributing findings to generate visibility can help maximize scientific impact. Higher online visibility in turn increases the chances of research being noticed, used, and cited, while also helping grow researchers' reputations and enabling future career opportunities (Collins et al., 2016; Thelwall et al., 2013). Researchers have increasingly used social media to share their publications on the internet and reach out to different audiences. From 2012 to 2016, the number of scientific tweeters has significantly increased from 0.21 million to 2.06 million (Yu et al., 2019).

Online visibility provides a new viewpoint to investigate gender inequality in the contemporary science world, with critical implications for creating a more inclusive and equal environment for researchers. The use of social media has the potential to help improve equity, diversity, and inclusion within academia as it provides a widely available and readily accessible platform to scholars, including underrepresented groups such as women (Yammine et al., 2018). However, some evidence shows that the online impact of scientists is unlikely to be gender neutral, much like the scientific success offline (Vásárhelyi et al., 2021) and the scientific communication on social media is still disproportionately male-dominated (Sugimoto et al., 2017; Yammine et al., 2018). Women have been found to gain less attention than men in disseminating their research online, according to tracking their online mentions of research articles (Vásárhelyi et al., 2021). Given such observation, a more comprehensive understanding of the role of social media in gender disparity in online visibility and its linkage with scientific impact is needed.

1.2.3 Work-family conflict

Even though the past several decades have seen great efforts made to increase the participation of women in science, women's representation has not increased apace with the growth of women in higher education (Wolfinger et al., 2008). There is a historically unchanged dilemma for women: work-family conflict, which to some extent contributes to the underrepresentation of women in science (Mason et al., 2013a). Abundant literature has shown female researchers tend to spend more time on domestic chores than male researchers. Particularly, the domestic responsibilities associated with parenting require considerable and fixed time and energy, which unavoidably compete with the demands of academic work (Derrick et al., 2022; Mason and Goulden, 2004; Mason et al., 2013a; Mason et al., 2023; Wolfinger et al., 2008).

It has been shown that parenting engagement is related to decreased research productivity and impact, and women are more likely to serve in lead parenting roles, making them suffer more from motherhood penalties in their academic performance (Derrick et al., 2022). Such academic productivity and impact disadvantages put female researchers in inferior positions in competitions for tenure-track positions and leak out of the pipeline at far greater rates than men. The underrepresentation of women in academia already occurs at the early-career stage, between the pool of Ph.D. candidates and the pool for academic jobs (Anders, 2004; Solga et al., 2023). Female doctorate recipients are more likely to seek careers outside academia, considering the intense conflict between work, family and parenthood. The roles of researcher and mother are both highly demanding, in terms of time, financial, physical, and emotional resources (Mason et al., 2023). The "conventional wisdom" advises women against having children during doctoral studies as the role of mothers is always seen as incompatible and conflicting with doctoral research in a way that fatherhood is not (Kulp, 2019; Mason et al., 2013a; Mason et al., 2023; Morrison et al., 2011; Wolfinger et al., 2008). The unequal impact of parenthood on researchers expands the gender disparity in academia from their early-career stage. The challenges of coordinating parenthood and academic careers among early-career researchers, especially Ph.D. mothers, deserve more attention.

1.3 Bibliometric analysis

1.3.1 Bibliometric data

One of the biggest challenges of studying gender disparity in academia is the lack of unified and comprehensive statistics suitable for making cross-national comparisons of gender disparities in science. Much of the existing research on gender differences is based on case studies limited to subsets of active scientists in specific countries, disciplines, or institutions, making it difficult to compare and generalize the findings to all of science (Huang, Gates, et al., 2020). The fragmented evidence does not allow for drawing a holistic picture of female participation in science, and this evidence is dependent on geographical and political contexts (Halevi, 2019).

Bibliometric data, which refers to the quantitative data derived from bibliographic records, including publications, authors, journals, and citations, is increasingly used to analyze scholarly communication patterns, track research trends, and evaluate the impact of scholarly output. In other words, bibliometric data is the collection of scientific publications' metadata, usually accumulated by companies such as Clarivate (Web of Science), Elsevier (Scopus) and similar that collect publicly available information. This type of data captures the publishing histories of researchers across disciplines, time, and geographical locations, providing the possibility of creating a more comprehensive picture of the gender disparity in science.

Several Application Programming Interfaces (APIs) and portals are available for gathering bibliometric data. Web of Science (WoS) and Scopus are the most frequently used sources of bibliometric data, providing metadata on scientific documents and citation links between these documents (Baas et al., 2020; Birkle et al., 2020; Visser et al., 2021). These publications are reviewed against quality standards before inclusion, so they are considered standard tools for bibliometric analysis and research evaluation. In addition to the main bibliographic databases, there are other newer databases, such as Dimensions (by Digital Science) and OpenAlex (by OurResearch) which are both open-access and make the metadata about articles available for free (BibliometriX, 2023; Herzog et al., 2020; Priem et al., 2022). Several studies have compared the coverage of different bibliometric data sources (Huang, Neylon, et al., 2020; Mongeon and Paul-Hus, 2016; Singh et al., 2021; Visser et al., 2021). Fig. 1.1 by Visser et al. (2021), reused here from the Open Acess publication (licensed under CC BY 4.0), shows that out of the five data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic Graph (MAG), MAG offers by far the most comprehensive coverage of the scientific literature. WoS (curated and complemented by the Centre for Science and Technology Studies at the University of Leiden, CWTS), by comparison, includes the least scientific literature and has the most selective indexing philosophy. Despite the large coverage in Dimensions and MAG, the completeness of the metadata of the articles and patents they include might not be as extensive as the more established databases, and the quality of the data may be a concern for researchers.

In addition to the difference in the coverage, Table. 1.1 lists the strengths and weaknesses of different bibliometric data sources in multiple aspects. Generally speaking, the more established data sources, that is Scopus and WoS, outperform Dimensions and OpenAlex in terms of the quality and accuracy of the data (Baas

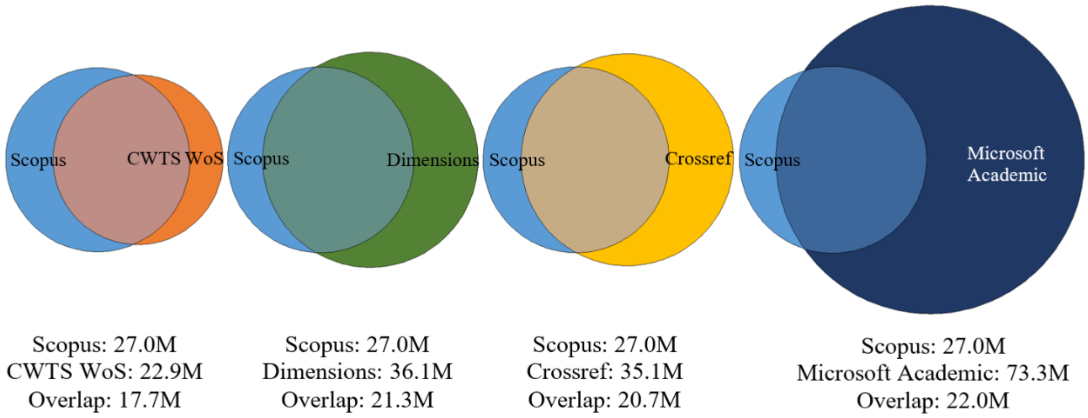


Figure 1.1: Overlap of documents between Scopus and the other data sources. Reused from Open Access publication (licensed under CC BY 4.0) entitled "Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic" by Visser, M., van Eck, N. J., and Waltman, L., 2021, *Quantitative Science Studies*, 2 (1): 20–41.

et al., 2020; Birkle et al., 2020; Visser et al., 2021). A reliable data source of publications with comprehensive information, such as academic affiliation address, is especially important for bibliometric analyses at the country and institutional level. This metadata is crucial for studies on the migration of scholars that rely on changes in these affiliation addresses to identify a migration event. However, a large fraction of Dimensions and OpenAlex documents lack such information (Guerrero-Bote et al., 2021). Despite English publications dominating the datasets, Scopus offers greater coverage of non-English documents (of all types) than WoS (Vera-Baceta et al., 2019). In Scopus, titles from all geographical regions are included and approximately 22% of titles in Scopus are published in languages other than English, adding up to 40 local languages (Elsevier, 2020c). Additionally, Scopus has the advantage of the self-developed author identifier, which allows the disambiguation of authors, with an average precision of 98.1% and a recall of 94.4% measured on a gold set (Precision measures the percentage of author IDs that are associated with the publications of a single individual only; Recall measures the percentage of author IDs that are associated with all of the Scopus publications of an individual) (Baas et al., 2020).

Table 1.1: Comparision of main bibliographic data sources: Scopus, Web of Science, Dimensions, and OpenAlex

Data source	Record Coverage	Strengths	Weaknesses
Scopus	94+ million (by Jan 2024)	• High-level of data completeness and accuracy • Reliable author name disambiguation	• Subscription-based access, not free
WoS	91+ million (by Feb 2024)	• High-level of data completeness and accuracy • Extensive longitudinal coverage going back to 1900	• Subscription-based access, not free • Lower coverage of non-English documents • No reliable author name disambiguation
Dimensions	140+ million (by March 2024)	• High-level of coverage • Free access to basic search functionalities • API access to data (limited to 2k records per export)	• Subscription-based access for advanced features • A large number of records lacking country and affiliation information • Low accuracy of field classification scheme • Unknown quality of author name disambiguation
OpenAlex	250+ million (by March 2024)	• Largest dataset of publications • Free access without subscription barriers • API access to data	• Unknown quality of author name disambiguation

Due to the aim of providing a cross-national measurement of gender inequality across different dimensions, including global scholarly migration, this thesis mainly opted for bibliometric data from Scopus and combined it with other data sources for specific research questions. This thesis got access to the full snapshot of Scopus bibliometric databases provided by the German Competence Network for Bibliometrics (Kompetenznetzwerk Bibliometrie, 2021) and used this dataset to construct researchers' publication histories, migration trajectories, and other events.

Compared to using web interfaces or APIs to acquire data, utilizing the full in-house snapshot of Scopus data, which has undergone quality controls[1], ensures higher data quality and saves time in data collection. In addition, one chapter utilized a global survey of parenting researchers, which has been linked to their publications in WoS by the research team who conducted the survey and is also possible to be linked to Scopus. In this chapter, I relied on the bibliometric data from WoS connected to the survey respondents, because WoS includes publications from more extended time periods and a proportion of respondents are from senior cohorts.

1.3.2 Bibliometric analysis on gender inequality in academia

Bibliometric analysis has become an essential tool for tracking not only research participation and scholarly activity, but also measuring scientific impact and influence. It can be carried out as a systematic analysis of a large body of publications, by using standard bibliometric indicators of quantity (the measurement of the productivity of particular researchers), quality (the measurement of the quality or performance of researchers' output), and structural indicators (the connection between publications, authors, institutes, and countries) (Durieux and Gevenois, 2010).

The studies examining gender disparities in science while using bibliometric methodologies can be grouped into four main areas which include productivity, performance, impact and visibility, and academic standing (Donner, 2021; Halevi, 2019; Larivière et al., 2013; Maliniak et al., 2013; Surawicz, 2016). The concept of "productivity puzzle" has been introduced as women have been found to publish less than men over the first 12 years of careers, after examining the productivity and impact of 263 matched pairs of men and women scientists who received doctorates in 1969–1970 (Cole and Zuckerman, 1984). More recently, Ruggieri et al., 2021

[1] 1) https://bibliometrie.info/en/research/working-reports/; 2) https://www.dzhw.eu/en/forschung/projekt?pr_id=484

carried out a bibliometric analysis on articles published in the period 2016–2018 by the researchers of the Italian National Research Council (CNR), indicating that gender disparities in scientific production consistently persist particularly in the disciplines of Science, Technology, Engineering and Mathematics (STEM), while the gender gaps are the closest to parity in medical and agricultural sciences.

Huang et al. (2020) offered a comprehensive picture of longitudinal gender differences in performance through a bibliometric analysis on 1.5 million gender-identified authors whose publishing careers ended between 1955 and 2010. They found that men and women publish at a comparable annual rate and have an equivalent career-wise impact for the same size body of work. However, this finding is in contrast with the results from Boekhout et al. (2021). Their research found that men produce on average between 15% and 20% more publications than women, and are more likely to be the last author of publications in some disciplines. Evidence is accumulating that female researchers are underrepresented in most scientific disciplines, publish fewer articles throughout their careers, and their work receives fewer citations than male researchers.

1.3.3 Gender detection in bibliometric analysis

Gender detection is one of the main challenges for gender-based analysis using bibliometric data since the only indicator for gender is the author's name (Halevi, 2019). While many names can be easily determined as female or male, there are some names including gender-neutral names and names with regional traits such as Asian names that make gender detection difficult. Currently, the main methods for gender inference include (Halevi, 2019): 1) Manual assignment; 2) Questionnaires 3) National and institutional databases; and 4) Software tools and APIs.

Among them, name-to-gender inference services including software and APIs tend to be increasingly used in identifying the gender of the authors from bibliometric databases (Halevi, 2019; Larivière et al., 2013). Generally, they rely on name repositories to further refine the results by combining additional information including the family name or country of origin. With the help of matching algorithms, these inference services are usually fast, cost-effective, and can be applied retroactively to large datasets. However, the accuracy of these services varies. Comparatively, Gender API ("GenderAPI.io") and NamSor ("Namesor") are normally recognized as the most accurate tools but not free (Karimi et al., 2016; Menéndez, 2020; Sebo, 2021a).

1.3.4 Combination of bibliometric data and other data sources

A variety of bibliometric methods have been developed to investigate the participation and impact of researchers. Bibliometric data retrieved from more established databases such as Scopus and WoS can be used to assess productivity and scientific impact by quantifying the number of publications and the received citations, respectively. Beyond these traditional metrics, bibliometric analysis has the potential to provide richer insights into gender disparities in academia when integrated with other data sources.

With the advent of Web 2.0, Altmetric, standing for "Alternative metrics" (a platform with a similar name: https://www.altmetric.com/), measures and monitors the reach and impact of scholarship and research through online interactions as a complement to traditional citation-based measurements (Huang et al., 2018; Vásárhelyi et al., 2021). As the way scientists exchange ideas and receive attention has become more diverse and open nowadays, combining bibliometric data and Altmetric data can offer a more comprehensive evaluation of the impact of scholarly research on both science and society (Sugimoto et al., 2017). Paul-Hus et al. (2015) compared the event counts from Twitter, blogs, and news collected by Altmetric with citations from WoS and found the scientific community constituting the citing audience is more male-dominated than the social media environment. Vásárhelyi et al. (2021) further integrated the coauthorship network based on the publication history from the Open Academic Graph (OAG) to explore the characteristics associated with online success by gender. The results indicate that the impact of prior work, social capital, and gendered tie formation in coauthorship networks are linked with online success for men, but not for women.

Data from Curriculum Vitae (CV) can also complement bibliometric data for the analysis of gender differences in academic careers. Combining bibliometric data and analysis of the authors' CVs can help establish whether there is a link between academic performance and academic retention, promotion, and recruitment (Halevi, 2019).

For topics such as work-family conflict, where information beyond scientific metrics may be required, surveys and interviews can provide relevant information and insights into the specific research questions. For example, a study by Morgan et al.(2021a) linked survey responses to the number of papers each researcher published in a given year to analyze how productivity changes in response to the presence or absence of the initial parenthood event. Derrick et al.(2022) surveyed researchers who have published in WoS as the first or last authors to explore the associations between parenting models learned from the survey and research outcomes based on the bibliometric analysis. Retrieving the researcher samples from large-scale bibliometric data helps expand the geographic coverage of samples to a global scale and also makes it easier to connect the survey information with the corresponding bibliometrics.

1.4 Overview of the Studies

In light of the presented gaps in the literature, this thesis aims to contribute to the literature on gender inequality in academia against the background of the wider forces of the globalization and digitalization of scientific systems. It presents a more comprehensive assessment of gender differences in science by leveraging both traditional and non-traditional sources of social data, especially "digital trace" data. The thesis draws on a growing body of work studying scientific production through the use of bibliometric data, but further complements this with other available data such as social media data and traditional survey data on science and scientists' behavior to tackle the following questions in a series of empirical studies:

1. Among the researchers who have ever affiliated with Germany, do trajectories and migration trends of internationally mobile researchers differ by gender in terms of out-migration and return migration? (Essay 1, Chapter 2)

2. On a global scale, how does gender inequality in the mobility of academic scientists vary across countries and over time, and how does it affect the demographic composition of the scientific workforce across the origin and the destination countries? (Essay 2, Chapter 3)

3. Are there gender differences in online attention to the publications of early-career researchers and these researchers' self-promotion behaviors? What are the linkages between online visibility and scientific impact among early-career female and male researchers? (Essay 3, Chapter 4)

4. Are there gender differences in the timing and sequencing of earning a Ph.D. and having children? Are these timing and sequencing related to the scientific performance outcomes of researchers? (Essay 4,Chapter 5)

Chapter 6 concludes the thesis, by summarizing the key findings on gender disparity in academia across different dimensions and highlighting the practical and methodological contributions to the literature. Moreover, this thesis calls for equal opportunities for women and men in science and a more inclusive and supportive academic environment, especially for female researchers.

1.5 Ethics statement

This thesis involves the analysis of bibliometric data which were collected from publicly available metadata on scientific publications on publishers' websites by Elsevier in the Scopus database (Essay 1-3), the tweeters of scientific tweets from Altmetric Details Page API and Twitter's public API which were collected before the recent changes in Twitter (now X) in July 2023 (Essay 3), and a global survey on parenting engagement in academia conducted by Derrick et al. (2022) and used as secondary data here (Essay 4).

The data of Scopus was obtained through the German Competence Network for Bibliometrics (Kompetenznetwerk Bibliometrie, grant number 16WIK2101A, https://bibliometrie.info/) and the access they granted to the Max Planck Digital Library. One of the most important ethical considerations in bibliometric analysis is the privacy and confidentiality of researchers and their work (Alfasoft, 2024). All chapters only contain aggregate or summary statistics without individual information, with no reidentification risk for individual researchers. Determining the gender of individuals based on their first name also raises important ethical issues by oversimplifying the concept of gender, even though this method has been widely used in scientific research (Peters and Norton, 2018; Sebo, 2021b). The bias and inaccuracy of gender detection results may perpetuate discriminatory practices, and the gender detection process based on first names undoubtedly resulted in misclassifications of some researchers. The sequential, three-step process for gender detection in this thesis has been validated with an overall high degree of accuracy, reducing the bias as much as possible. I also excluded the researchers whose gender cannot be identified from their first names from the dataset after controlling if they will affect the robustness of results. Also, the dichotomization of gender into female or male risks marginalizing some individuals who do not recognize themselves in this binary differentiation. I admit that this thesis was not able to consider gender minority groups or those identifying themselves with nonbinary genders. Determining gender through self-identification would be preferable and would also increase the accuracy of the result. However, self-identification is resource-intensive and often not feasible for large-scale studies (Sebo, 2021b). For this population-level study on relative female-to-male gender gaps, the results are able to highlight the overall, aggregate gender differences among researchers. There are no individual-level results of gender analysis presented in the study.

The Altmetric Details Page API was subscribed to Research Services by the University of Oxford and the usage of Twitter's public API was granted through an application to Twitter's Academic API for academic purposes, which was approved in 2022 . All data collection was conducted complying with the Terms of Use of publicly available data on both platforms and all results were shown in the form of aggregated numbers without revealing personal and sensitive information. There is no reidentification risk for individual users from the results that are presented in the thesis. I acknowledge that for this population-level study with a large dataset, it may not be possible to reach all Twitter users to obtain their informed consent (Ahmed et al., 2017). Twitter made it clear to its Tweeters what will happen to their data (including that it may be used for research) at the point of signing up to use its service (Nicolas, 2020), so it is a reasonable expectation of Twitter users that their public data exposed on the platform may be viewed and used for research. The linkage between scientific publications and related tweets was done specifically for tweets with the digital object identifier (DOI) of the publication. In other words, this is about scientific dissemination and outreach of individuals who were trying to reach a wider audience and publicize their work. Thus, although personal information was used for the linkage, it is in relation to public tweets about scientific content meant for public dissemination. In the process of identifying the researchers who self-promoted their publications on Twitter by comparing the author names with the user names and handle names of tweeters, this study took the ethical standpoint of not quoting tweets or disclosing the names in the results.

The survey data in essay 4 was collected by researchers at Indiana University and was used by me as a secondary survey data source. The main investigators who collected this survey provided me the data, as collaborators in this essay. The original, primary data collection for the survey was approved by the Indiana University Institutional Review Board on September 25, 2017. The protocol number is 1,706,832,059. Informed consent was obtained from all participants prior to data collection. The bibliometric information on scholars' publications from WoS was linked to the survey responses by the research team, which was also included in the dataset for this study. All methods were carried out in accordance with relevant guidelines and regulations, and all results were shown in aggregated form, not disclosing the individual-level original raw data to comply with the data usage agreement.

The research adheres to the ethical guidelines and codes of conduct outlined by the University of Oxford. The presented research adheres to the ethical guidelines and codes of conduct outlined by the University of Oxford. In conducting this research, I have been committed to practicing academic integrity, transparency and adhering to responsible research practices.

2

Departure and return migration of German-affiliated researchers: Are women less likely to return to Germany?

This chapter is co-authored with Samin Aref, Emilio Zagheni, and Guy Stecklov. Materials from this chapter have been published as Zhao, X., Aref, S., Zagheni, E., and Stecklov, G. (2022). "Return migration of German-affiliated researchers: analyzing departure and return by gender, cohort, and discipline using Scopus bibliometric data 1996–2020." *Scientiometrics*, 127, 7707–7729.

Abstract

The international migration of researchers is an important dimension of scientific mobility, and has been the subject of considerable policy debate. However, tracking the migration life courses of researchers is challenging due to data limitations. In this chapter, we used Scopus bibliometric data on eight million publications from 1.1 million researchers who have published at least once with an affiliation address from Germany in 1996–2020. We constructed the partial life histories of published researchers in this period and explored both their out-migration and the subsequent return of a subset of this group: the returnees. Our analyses shed light on the career stages and gender disparities between researchers who remained in Germany, those who emigrated, and those who eventually returned. We find that the return migration streams were even more gender imbalanced, which points to the need for additional efforts to encourage female researchers to come back to Germany. We document a slightly declining trend in return migration among more recent cohorts of researchers who left Germany, which, for most disciplines, was associated with a decrease in the German collaborative ties of these researchers. Moreover, we find that the gender disparities for the most gender imbalanced disciplines were unlikely to be mitigated by return migration given the gender compositions of the cohorts of researchers who have left Germany and of those who have returned. This chapter uncovers new dimensions of migration among scholars by investigating the return migration of published researchers, which is critical for the development of science policy.

2.1 Introduction

To ensure the dynamic flow of scientific ideas and expertise, and to promote and facilitate knowledge production, national science systems rely on the international exchange of scholars (Conchi and Michels, 2014; Moed et al., 2013; Robinson-Garcia et al., 2019). While the globalization of research has numerous benefits that have been widely acknowledged in the literature (Appelt et al., 2015; Bauder, 2015; Franzoni et al., 2015; Netz and Jaksztat, 2017), the undeniable downside of academic mobility is the potential loss of talent for countries that train and export more researchers than they receive from other countries. Global competition for talent has led to the introduction of a range of policies and economic incentives aimed at encouraging balanced flows of researchers. However, little attention has been paid to returnees who stay in other countries temporarily, and then return to their country of academic origin. These returnees usually bring with them additional skills, newly established connections and collaborative ties, and complementary expertise acquired abroad (Appelt et al., 2015; Organisation for Economic Co-operation and Development (OECD), 2008). Equally importantly, there is evidence that these returnees tend to receive far more citations than their stationary counterparts (Guthrie et al., 2017b), or their internationally mobile counterparts who do not return (Zhao et al., 2021). Thus, for countries that are facing the challenge of the out-migration of researchers exceeding the in-migration of researchers (Zhao et al., 2021), facilitating the return migration of scholars and taking steps to rebalance these trends to their benefit are extremely critical.

Despite being recognized as a science powerhouse, Germany has been sending more highly qualified individuals (including researchers) abroad than it has been receiving according to some reports (Organisation for Economic Co-operation and Development (OECD), 2008; Schiller and Cordes, 2016; Zhao et al., 2021). In recent years, Germany has implemented a range of policies and programs designed to attract students and researchers from other countries (Bardin, 2016; Düvell, 2019; Eule, 2016). There are several return migration programs aimed at maintaining and strengthening ties with previously German-affiliated researchers to facilitate their re-integration into the German science system (Conchi and Michels, 2014). For example, the German Academic International Network (GAIN)[1] supports current and prospective returnees, and facilitates cooperation between researchers in Germany and North America. The German Scholars Organization (GSO)[2] is another initiative aimed at reversing Germany's "brain drain" and turning it into a "brain gain" by offering several services to academic professionals in Germany. Given the practical relevance of this issue to policy development and strategic decisions at a national level, a better understanding of the trajectories and migration trends of researchers formerly affiliated with German institutions and academic returnees to Germany is urgently needed.

Return migration also has potentially large implications for the persistence of gender inequalities in academia (Zippel, 2017). The issue of gender disparities in academia has been extensively documented across many disciplines and in most countries (Huang, Gates, et al., 2020; Larivière et al., 2013; Zhao et al., 2021). Although it has been suggested that breaking the glass ceiling that hinders women's advancement is especially challenging for internationally mobile academics (Zhao et al., 2021; Zippel, 2017), the heterogeneity in the levels of gender disparity among former and current German-affiliated researchers, and particularly among returnees, is not clear. Exploring this topic is the first step towards achieving more balanced gender representation in academia, and ensuring the sustainability of academic careers for women researchers (Weert, 2013; Zhao et al., 2021). Due to the implementation of a wide range of measures and policies aimed at promoting

[1] www.gain-network.org
[2] www.gsonet.org

gender equality in academia, female representation in various disciplines has been increasing (Huang, Gates, et al., 2020; Macaluso et al., 2016; Zippel, 2017). Given these developments, studying the temporal trends at the intersection of gender and international mobility can have important administrative and policy implications. Moreover, the question of how the representation of female returnees in more recent cohorts has changed in response to these policies and developments deserves more attention. An extensive analysis of return migration among German-affiliated researchers disaggregated by discipline, gender, and cohort is critical for making progress towards transforming the German science system into an inclusive and diverse system with balanced migration flows of scholars.

Previous studies have suggested that academic returnees tend to maintain collaborative ties with their previous host countries (Ackers and Gill, 2005; Conchi and Michels, 2014; Franzoni et al., 2014; Guthrie et al., 2017b). As these ties are critical components of knowledge transfers, it is clear that scholarly migration is no longer a zero-sum game. Returnee researchers may face challenges when attempting to incorporate the knowledge they have acquired abroad into another context, and in re-establishing their career in their academic home country (Fernández-Zubieta et al., 2015; Melin and Janson, 2006; Weert, 2013). This may be partly because researchers are disconnected from their home country's academic networks while abroad, which may limit their access to the information and support they would need to find a job in their academic home country. This erosion of connections may, in turn, reduce the willingness of researchers to return (Ackers and Gill, 2005; Baruffaldi and Landoni, 2012). Previous work on this topic has shown that there is a gap in macro-level quantitative research on collaboration and migration among scholars. In particular, the interactions between scholars' collaborative ties with their academic home countries and return migration have not been previously investigated. Therefore, a comprehensive analysis that examines these relationships can provide useful insights into return migration among academics. This can facilitate the development of policies that create additional paths for returnees to re-integrate professionally into their academic home countries.

Motivated by the observations above, this chapter relies on large-scale digitized bibliometric data from Scopus (Burnham, 2006; Mongeon and Paul-Hus, 2016) to investigate the trajectories and migration trends of internationally mobile researchers in Germany as well as their German academic links during the period of being outside of Germany. We analyze German-affiliated published researchers during the 1996-2020 period while taking each researcher's years of experience, gender, discipline, and cohort into account. Specifically, this paper aims to address the following research questions (which are both methodological and empirical), with a focus on the return migration of scholars to Germany:

1. What is the composition of returning researchers based on their gender, years of experience, and previous host countries? (Subsection 2.3.1 and 2.3.2)

2. How does the gender ratio vary by discipline and cohort among researchers who leave Germany, and among researchers who return? (Subsection 2.3.3 and 2.3.4)

3. How does the association between return migration to Germany and collaboration with German institutions vary across disciplines and cohorts? (Subsection 2.3.5)

2.2 Materials and methods

2.2.1 Authorship records of German-affiliated researchers

Scopus is an abstract and citation database of scientific literature (Burnham, 2006) that covers over 77 million publications, according to its 2020 coverage guide (Elsevier, 2020b). From the Scopus database, we obtained the authorship records (linkages between an author's affiliation and a publication) of more than 1.1 million researchers for our analysis, which involves over eight million publications. All of these researchers had used a German affiliation address in at least one publication at some point during the 1996-2020 period (ending in April 2020).

2.2.2 Pre-processing of raw bibliometric data

To ensure the reliability of our results, we pre-processed the raw bibliometric data from Scopus using three sequential steps. These three steps discussed below are: handling the missing countries in the dataset, disambiguating the author profiles, and inferring gender from the authors' first names.

Handling missing countries in the dataset

First, there are 74,430 (5%) authorship records for which the country variable is missing. To handle the missing values systematically, we have developed a neural network algorithm inspired by (Miranda-González et al., 2020) that predicts the missing countries with a high degree of accuracy. This supervised learning algorithm takes an affiliation address as the input and predicts the country as the output. The data used to make the prediction are city, institution, and address. These strings were combined using a bag-of-words method with frequencies (of a given word in a sample) that are normalized (relative to the frequency in the whole dataset) using a *term frequency inverse document frequency* approach (Tokunaga and Makoto, 1994). A simple and standard architecture was used to develop the neural network (*deep feed forward neural network*) with non-linear activation functions. We used a random set of one million authorship records drawn from our dataset that contain country information and split it into training data (80%) and testing data (20%). Other technical details of the development of the neural network have been explained in (Miranda-González et al., 2020). The predictions made based on the test dataset showed that the neural network can correctly predict the country for 98.4% of records, which is a level of accuracy we consider acceptable for handling missing country data.

Author name disambiguation

The second step of our data pre-processing helped us overcome the problems associated with using Scopus author IDs to identify unique authors. It has been shown that Scopus author IDs have high levels of precision and completeness (Kawashima and Tomizawa, 2015; Paturi and Loktev, 2020). Precision measures the percentage of author IDs that are associated with the publications of a single individual only. Completeness measures the percentage of author IDs that are associated with all of the Scopus publications of an individual. The results of an evaluation of the accuracy of Scopus author IDs conducted in August 2020 showed that the precision and the completeness of Scopus author profiles are 98.3% and 90.6%, respectively (Paturi and Loktev, 2020). However, while it appears that the quality of individual-level Scopus data is sufficiently high to enable us to study the migration of researchers (Aman, 2018a; Kawashima and Tomizawa, 2015), there are several notable limitations to keep in mind when using Scopus data for migration research. The precision limits in Scopus author IDs implied that 1.7% of Scopus author IDs may be associated with the publications of more than one person, which could affect the accuracy of the migration events detected by looking at changes in affiliation countries per author ID. Accordingly, as the second step in our data pre-processing, a subset of authorship records that are more likely to have suffered from the precision flaws of Scopus author IDs were analyzed using our conservative author name disambiguation algorithm.

Our author disambiguation algorithm was inspired by the state-of-the-art methods in author name disambiguation (D'Angelo and van Eck, 2020). It assumes that every two authorship records are from distinct individuals (despite sharing a Scopus author ID), unless sufficient evidence is found to the contrary using a rule-based scoring approach and a clustering method. We first calculated the similarity score of each pair of authorship records belonging to the same author ID. The similarity was measured based on author names, coauthor names, subjects, funding information, and grant numbers. The author disambiguation algorithm made all pairwise comparisons between authorship records with the same author ID, and created a distance matrix based on similarities and dissimilarities in the aforementioned features for each pair of records. A clustering algorithm was then used to process the distance matrices and cluster similar authorship records. We then issue revised author IDs based on the resulting clusters. We used the agglomerative clustering algorithm from the scikit-learn Python library (Pedregosa et al., 2011) to cluster authorship records. This algorithm belongs to the family of hierarchical clustering methods. Supporting our conservative approach, it first places each record in its own cluster and then merges pairs of clusters successively if doing so minimally increases a given linkage distance (Pedregosa et al., 2011). As well as being compatible with our conservative approach, agglomerative clustering has the advantage of offering us the flexibility to process any pairwise distance matrix.

We examined the author profiles that are outliers in terms of the number of affiliation countries or the number of publications. In particular, there are 30,715 (2%) author profiles that are associated with more than six countries of affiliation, or more than 292 publications[3] (an average of more than one publication per month across a period of 24 years and four months). These author profiles are more likely than others to be affected by the precision flaws of Scopus author IDs. For example, each Scopus author profile could contain records from more than one individual researcher. Based on these criteria, 25,000 author IDs are classified as suspicious. These author IDs are associated with 2,242,797 publications. After disambiguation, revised author IDs are issued for these records according to their clusters and then merged with the rest of the data in preparation for the third pre-processing step.

[3]These two thresholds are chosen empirically so that a subset of outliers of a size compatible with the results on Scopus author ID precision flaws (Paturi and Loktev, 2020) can be extracted.

Inferring gender from first names

In this chapter, We inferred gender from the first names of researchers by looking up their first names in a large database of names and genders called *Genderize* (genderize.io). After performing basic text operations (like removing middle initials from the first name field), we obtained the gender for 1,117,813 author profiles in our dataset. For the remaining profiles, we manually searched for public author information to determine the gender by checking the individuals' personal homepages, curricula vitae, online profiles, and biographies in publications, as well as other online sources. Using this manual approach, we were able to determine the genders for 3,139 additional author profiles. Finally, the most likely gender for 77% of the author profiles in our dataset was determined through either algorithmic or manual gender detection. For our analyses that involve gender (e.g., measuring gender ratios), we set aside the 23% of author profiles whose gender could not be determined either algorithmically or manually.

2.2.3 Migration event, mobility types, and career stages

The international mobility of researchers was determined by identifying the changes in the affiliation addresses of authors across different publications over time. To more reliably detect migration events, the most frequent (mode) country(ies) of affiliation was extracted for each researcher in each year. In this chapter, a migration event is considered to have happened only if the mode country of affiliation changes for the researcher across different years (Subbotin and Aref, 2021). Accordingly, the *country of academic origin* (*country of academic destination*) is defined as the mode country during the first (last) year of publishing. Based on the individual's migration events or the lack thereof, each researcher can be assigned to one of the following four categories:

1. Non-mover (with Germany being the researcher's mode country in all years);

2. Immigrants and transients (origin: not Germany; but with Germany being the researcher's mode country at some point in time);

3. Outward (origin: Germany; current country: not Germany); and

4. Returnee (origin and current country: Germany; but with another country being the researcher's mode country at some point in time).

Except for non-movers, researchers may move between the categories over time, as an individual's status depends on the period being examined. For example, an individual identified as an outward researcher will become a returnee at the next point in time if a move to Germany is detected.

We defined the *academic age* (*age*) of a researcher as the number of years since his/her first publication. Furthermore, we classified researchers as *early-career* (*senior*) if their academic age is seven years or less (14 years or more). Researchers with an academic age between seven and 14 years are categorized as *mid-career* (Aref et al., 2019). As our dataset covers only the 1996-2020 period, our analysis of some temporal dimensions of the data or cohorts of researchers may suffer from left truncation and/or right censoring. We explain in Section 2.3 some of the resulting limitations of our dataset were explained in Section 2.3 .

2.2.4 Inferring disciplines using a topic model

The Science Journal Classification (ASJC) codes in our bibliometric dataset indicate the fields and disciplines (subfields) of publication venues, which could be used as proxies for determining the disciplines of researchers (Zhao et al., 2021). However, because the links between the disciplines associated with journals and the disciplines of authors are indirect, we used a data-driven method to infer the disciplines of individual researchers. Topic modeling, which is a common unsupervised learning approach for natural language processing, can be used to determine the disciplines of researchers by inferring the latent topical structure of textual bibliometric data (Blei, 2012; Gerlach et al., 2018).

As a flexible topic model, Latent Dirichlet Allocation (LDA) is in essence a generative probabilistic model with three layers: document, topic, and word (Blei et al., 2003; Pritchard et al., 2000). It assumes that each topic is a mixture of an underlying set of words and that each document is a mixture of a set of topic probabilities (Blei et al., 2003; Gerlach et al., 2018). LDA has been shown to perform reliably in automatically identifying semantic topic information from large-scale textual data (Dahal et al., 2019).

From the publications authored by each researcher in our dataset, we extracted publication titles, journal titles, and keywords to generate the individual's text corpus (document in LDA terminology). We then tokenized the text by removing all punctuation and making all of the words lower-case to improve the cohesion of the documents. The remaining words in each text corpus were then lemmatized and stemmed to their root form. This includes being transferred to the first person and the present tense if needed. For some common phases (e.g., machine learning) that are related to discipline topics, we used multi-word expressions, two-gram collocations, and three-gram collocations from all of the text documents according to their frequency of occurrence. The tokenized and lemmatized texts were then abstracted to a *bag of words*, which records the indices of words and the number of times each word appears in an author's LDA document. In LDA, each document can be considered as a mixture of latent topics, each of which is characterized by a distribution of words (Blei et al., 2003). The topic coherence score is a measurement of the semantic similarity between the high scoring words in each topic and represents the interpretability of the topics.

After the implementation of the steps above, the average topic coherence score of all topics was maximized at 0.67 (through trial and error) when we allowed 30 topics for the whole text corpus. Each topic is composed of a set of vocabularies and the corresponding weights that indicate their contributions to the topic. The topic with the highest probability among all of the initial topics was assigned to each author's LDA document as the intermediate result. We manually interpreted the topics based on their most frequent terms and assign titles to them accordingly. When different topics include similar or highly relevant keywords, we combined them into a single discipline. For example, a topic involving "space" and "earth" and a topic involving "galaxy" and "star" were combined into the same discipline: "Earth and Planetary Sciences." Using this approach, we produced 17 distinguishable disciplines[4] to represent the main discipline of each researcher according to their publications. Detailed results on the 30 topics and their mapping to 17 disciplines were provided in the Table. A.1 in Appendix A. The author's LDA document was considered as "Multidisciplinary" when no single topic's contribution percentage exceeded 0.3.

2.2.5 Cohorts leaving Germany and returning to Germany

A cohort is a group of people who have experienced a common event in a selected period, such as birth (Reilly et al., 2005; Rothman, 2012). In our analysis, we used the time of first publication as the common event for defining cohorts of researchers. To reduce the impact of left-truncated data on cohorts, which is more likely for the first few years of our dataset, we used the following three cohorts: 1998-2001, 2002-2005, and 2006-2009.

Person-time rate is an index commonly used in epidemiology and demography to express an incidence rate: i.e., the number of incidents (migration events) per person-time in a population during a period (Rothman, 2012). The denominator of a person-time rate is the total amount of time that the study members are at risk

[4] "Agricultural, Biological and Environmental Sciences," "Biochemistry, Genetics, and Molecular Biology," "Chemistry and Chemical Engineering," "Computer Science," "Earth and Planetary Sciences," "Economics and Social Science," "Engineering," "Energy," "Health Professions," "Immunology and Microbiology," "Materials Science," "Mathematics," "Medicine," "Neuroscience," "Pharmacology, Toxicology and Pharmaceutics," "Physics and Astronomy," and "Psychology."

of a certain incident during a period. One key advantage of using person-time rate for migration is that it enabled us to consider the case that different individuals are exposed to migration events for varying amounts of time. The incidents we were interested in are: (1) leaving Germany for the group of all researchers in Germany, and (2) returning to Germany for the group of all outward researchers. Given a specific period of time t, the *departure rate* per 1,000 person-years for a cohort c is defined in Eq. (2.1).

$$R_{\text{departure}(c,t)} = N_{\text{departure}(c,t)} / PT_{\text{in Germany}(c,t)} \times 1000 \tag{2.1}$$

In Eq. (2.1), $N_{\text{departure}(c,t)}$ represents the number of researchers from cohort c leaving Germany during time period t, and $PT_{\text{in Germany}(c,t)}$ represents the sum of the number of years each researcher from cohort c stays in Germany (and is exposed to leaving Germany) during period t. The denominator of the departure rate takes all of the researchers who are in Germany into consideration as the population exposed to leaving Germany. Similarly, given a specific period of time t, the *return rate* per 1,000 person-years for a cohort c is defined in Eq. (2.2).

$$R_{\text{return}(c,t)} = N_{\text{return}(c,t)} / PT_{\text{outside Germany}(c,t)} \times 1000 \tag{2.2}$$

In Eq. (2.2), $N_{\text{return}(c,t)}$ is the number of returnees from cohort c during period t, and $PT_{\text{outside Germany}(c,t)}$ is the sum of the number of years each outward researcher stays outside of Germany (and is exposed to returning to Germany) during period t. The denominator of the return rate only involves researchers who have left Germany as the population exposed to returning to Germany. We computed the departure rates and the return rates separately for male and female researchers in the three cohorts (1998-2001, 2002-2005, and 2006-2009). Specifically, we considered the departure rates of researchers of different cohorts who leave Germany at the academic ages of one to five, and the corresponding return rates during the first five years after their departure from Germany. Taking the 1998-2001 cohort as an example, the departure rate at academic age one refers to the outward researchers who were “academically born" in 1998 (1999, 2000, 2001) and left Germany in 1999 (correspondingly, 2000, 2001, 2002). For the same cohort, the return rate refers to the researchers who returned during the 2000-2004 (2001-2005, 2002-2006, 2003-2007) period among the outward researchers who left Germany in 1999 (correspondingly, 2000, 2001, 2002).

2.2.6 Collaborations with Germany while away

We defined and used the variable *collaborative ratio* to distinguish the strength of the academic linkages with Germany for outward researchers during the period when they were away from Germany. For the outward researcher, i, during the period when s/he was away from Germany (denoted by t), we calculate his/her collaborative ratio using a simple fraction: $CR_{(i)} = D_{(i,t)}/N_{(i,t)}$. The numerator, $D_{(i,t)}$, counts the publications of outward researcher i in period t that list a German affiliation for i or his/her co-authors. The denominator, $N_{(i,t)}$, is the number of all publications of outward researcher i during period t. If a publication authored by i during period t has at least one author with a German affiliation, it contributes to the collaborative ratio $CR_{(i)}$. Furthermore, the average collaborative ratio for all researchers in each discipline (cohort) was calculated to measure the average strength of the academic collaboration with Germany maintained by the outward researchers in that discipline (cohort).

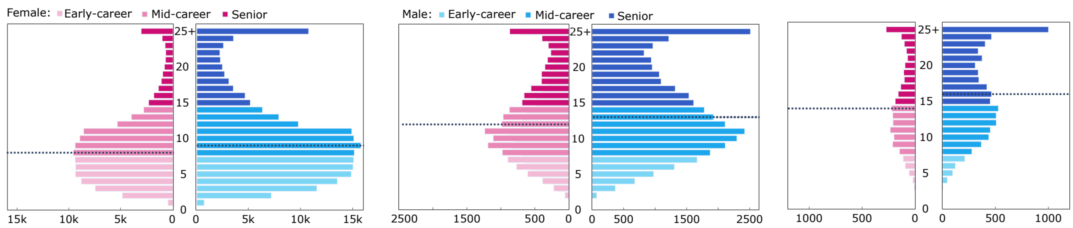
2.3 Results

Using cleaned and processed bibliometric data associated with over eight million Scopus-indexed publications over 1996-2020 from more than one million German-affiliated researchers, we analyzed data on 375,288 female researchers associated with 2,665,139 publications and 745,664 male researchers associated with 6,516,016 publications. Among these researchers, there are 50,803 female mobile researchers (associated with 1,007,606 publications) and 119,298 male mobile researchers (associated with 2,760,282 publications) who have ever migrated between Germany and 194 other countries in our dataset. There are $103,573 \pm 48610$ researchers in each discipline, with medicine having the largest number of researchers (199,658) and health professions having the smallest number of researchers (26,398).

Based on pre-processed data, we provided five analyses to describe different aspects of the emigration and the return migration of these researchers. In Subsections 2.3.1 and 2.3.2, we tracked their career life courses from a temporal perspective, and their geographic trajectories from a spatial perspective. We then compared the departure rates and the return rates of female and male scholars to explore the gender differences across cohorts and disciplines (Subsections 2.3.3 and 2.3.4). Finally, we looked at the association between the return rates of outward researchers and the strength of their collaborative ties to Germany in Subsection 2.3.5. Our inferred migration events dataset is publicly available in a FigShare data repository (Zhao et al., 2022b).

2.3.1 Age and gender composition of researchers

We compared the age and gender compositions of three groups of researchers: non-movers, outward researchers, and returnees. Fig. 2.1 compared the age and gender distribution of these three groups using population pyramids, which include individuals who survived as an active researcher up to a certain date (researchers whose latest publication was in 2010 or later). Ignoring the truncated top age, we can see that both female and male non-movers had a notable and pronounced bulge at the transition from early-career ages to middle-career ages, which was presented as an expansive pattern. However, the non-mover age pyramid showed considerably small proportions at other ages, with a pattern characterized by a sharp decline to age 21, followed by a stable increase until age 25+. The median ages for female and male non-movers are nine and 10 years, respectively. In the categories of outward researchers and returnees, the overall length of academic trajectories has been considerably prolonged for both female and male researchers. Specifically, the median ages of female outward researchers and returnees are 12 and 14, respectively; and the median ages of male outward researchers and returnees are 13 and 16, respectively.



(a) Non-movers (b) Outward researchers (c) Returnee researchers

Figure 2.1: Composition of academic age and gender for non-movers, outward researchers, and returnees. The colored versions of all figures are available online in high resolution.

Overall, these findings suggested that both male and female returnees stayed in academia longer than both outward researchers and non-movers. As more evidence on return migrants emerged, the strengths of returnees were increasingly being seen as valuable. It has, for example, shown that from a historical perspective, returnees tended to make important contributions to local economies and to be relatively successful, both in comparison to people who never migrated and to people who emigrated but did not return (Abramitzky et al., 2019). The findings on the positive impact of return migration were encouraging and suggested that Germany, as well as other sending countries, should embrace international mobility and the return migration of scholars. While we found that both male and female scholars benefited from returning, we also observed that returning to Germany had a more positive impact on the careers of male than of female researchers. For example, the results showed that 64.57% of male returnees, but only 51.13% of female returnees, had become senior professionals (see detailed age composition in 2.1(c)). The smaller benefits found for women are not surprising, and point to the ongoing challenges women in academia face.

2.3.2 Out-migration and return migration by geography

Fig. 2.2 illustrated from a geographic perspective the interplay between outflows of researchers from Germany and the corresponding rates of return to Germany, through a *density equalizing cartogram* (Dorling et al., 2006; Houle et al., 2009). Here, the shape of the map polygons was transformed proportionally to the outflows to different countries. The colors represent the differences in the countries' return rates, as further explained in the legend. The most common host country for researchers from Germany was the United States (US), which received around 24% of the outward researchers from Germany. Next came Switzerland and the United Kingdom (UK), which together attracted 22% of the outward researchers from Germany. In total, these three countries received nearly half of the outward researchers from Germany and had thus become the most appealing options for German-affiliated researchers interested in pursuing an international academic career. These estimates were also consistent with previous findings that the US, the UK, and Switzerland are the most common origin and destination countries for scholarly migration to and from Germany (Organisation for Economic Co-operation and Development (OECD), 2015; Zhao et al., 2021). The observed pattern for the European countries that received researchers from Germany indicated that the countries neighboring Germany and German-speaking countries were among the most popular host countries for scholars who began their publishing activity in Germany.

As the colors on the map showed, the rates of return from the most common receiving countries that were larger in size were all below 36%; meaning that about one-third of German-affiliated researchers moved back to Germany, while nearly two-thirds continued their research abroad. While the US hosted the largest share of researchers from Germany, the rate of return to Germany from the US was also relatively high, at 34%. Similarly, while the UK and France were among the top host countries for researchers from Germany, the rates of return to Germany from these countries were also high, at 30% and 29%, respectively. By contrast, the rates

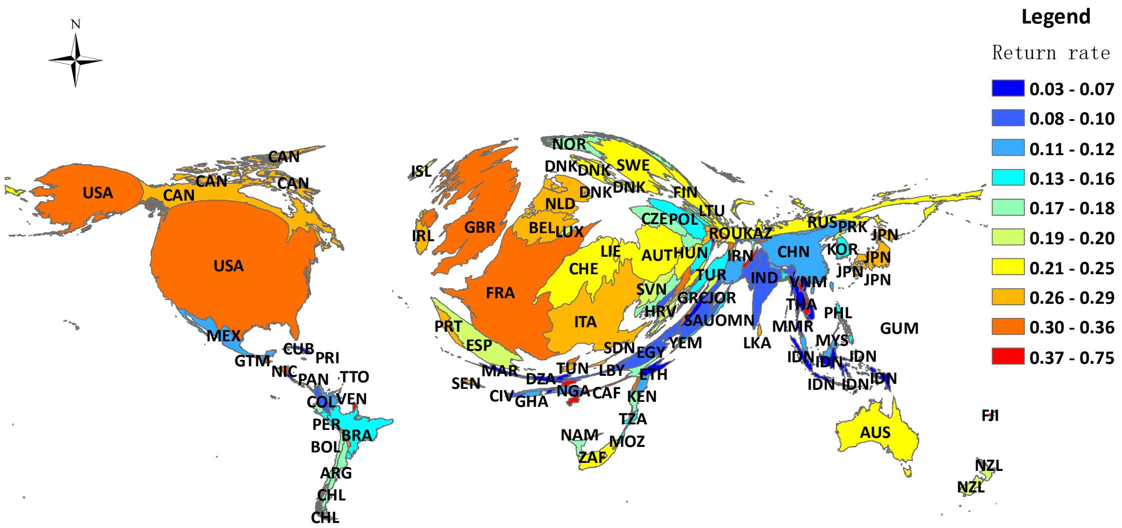


Figure 2.2: Outward flows (from Germany) and respective return rates across countries. The sizes of the countries are proportional to the flows of outward researchers from Germany. The colors indicate the differences in the return rates of the German-affiliated researchers returning to Germany from each country.

of return to Germany were below one-quarter for German-affiliated researchers in Switzerland, Sweden, Austria, and Australia; and the return rate was especially low for previously German-affiliated researchers in Switzerland, at only 20%. It thus appears that researchers who moved from Germany to these four countries were comparatively less likely to return. The lower propensity to return may be partly explained by the higher spending on Research & Development (R&D) in these countries. In 2017, Switzerland, Sweden, Austria, and Australia spent about 3.18%, 3.36%, 3.00%, and 3.08% of their GDP, respectively, on R&D – far above the OECD average of 2.67%, ahead of the US (2.85%), the UK (1.68%), and on levels competing with Germany (3.07%) (Organisation for Economic Co-operation and Development (OECD), 2021). In addition, the lower return rate of previously German-affiliated researchers in Switzerland was broadly consistent with our expectations, given that approximately 1.2% of all scientific papers worldwide were produced by Swiss-affiliated researchers, which was remarkable given the country's small population (Turney, 2019).

2.3.3 Rates of departure and return across cohorts

Fig. 2.3 illustrated the departure rates (left) and the return rates (right) per 1,000 person-years, disaggregated by cohort and gender. The academic age at departure is on the y-axis for both outward researchers and returnees. For returnees, the length of time away from Germany was also reported by the use of ombre colors. Taking the cohort 1998-2001 as an example, out of 1,000 researchers, around eight women and nine men with a German academic origin in this cohort moved abroad at academic age one. For every 1,000 outward researchers who left Germany at academic age one, around 215 women and 278 men had returned to Germany within five years. Among them, 74 women and 88 men had returned to Germany after one year, making the first year the most likely year of return for that cohort. In general, there was a slight but stable decline in the departure rates with academic age for all three cohorts. However, the most striking pattern was observed for the 2006-2009 cohort: the departure rates of female researchers exceeded those of male researchers for most ages, especially at academic ages one, two, and three. Specifically, we found that 11 out of 1,000 female researchers in this cohort left Germany at academic age one, while only nine out of 1,000 male researchers left Germany at that age. This result indicated that in this cohort, more female than male researchers chose to migrate early in their careers. Meanwhile, the return rates of the female researchers of all three cohorts were much lower than those of their male counterparts. This difference may be partly related to the longer average length of academic life for male returnees, as Fig. 2.1 showed. Taken together, these results indicated that female outward researchers had a greater tendency than their male counterparts to remain abroad for longer periods or possibly to settle down in other countries, which may have exacerbated the gender disparities in the German science system. Thus, the findings suggested that out-migration trends may increase gender disparities within the German academic system unless further action is taken.

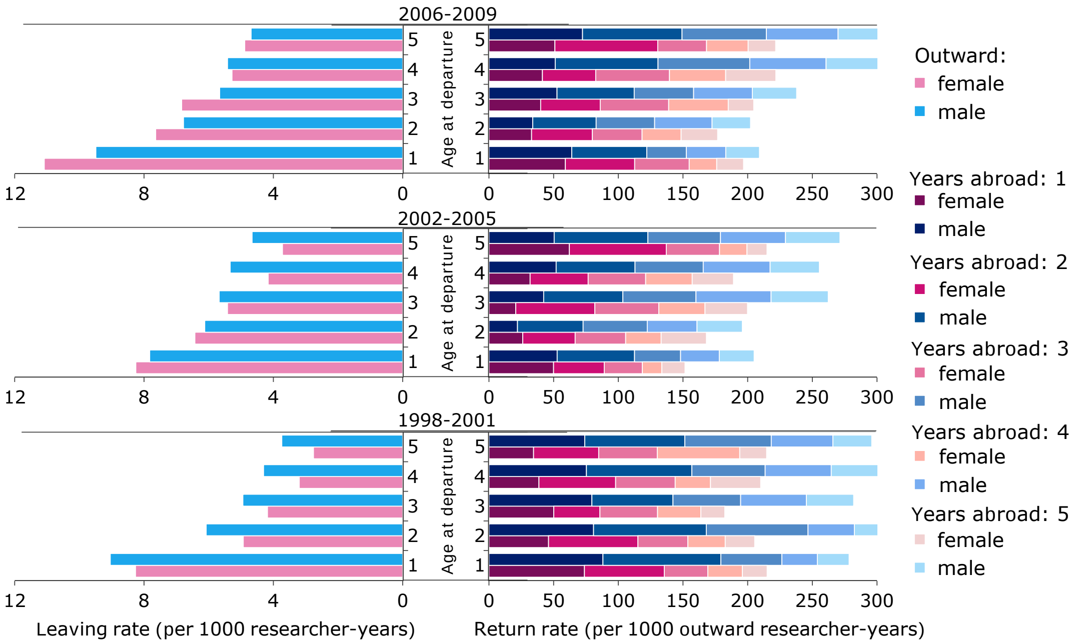


Figure 2.3: The rates of leaving Germany within the first 5 years since first publication per 1000 person-years (left), and the rates of return to Germany within the first 5 years after departure per 1000 person-years (right).

We also observed that the return rates were generally higher for researchers who moved out of Germany in their later years and tended to increase with academic age. This trend was more noticeable among male researchers and in the two latest cohorts. The more pronounced increase in return migration at later academic ages for men than for women suggested that there were structural processes that operated at specific moments of the academic life course and that these processes could further extend the gender differences in German academia.

2.3.4 Gender composition of outward and return streams by discipline and cohort

Considering that the male-to-female ratios of researchers vary across disciplines (Zhao et al., 2021), we took a further look at the gender disparities disaggregated by discipline for the three cohorts, as shown in Fig. 2.4. The colors in the heat map showed that the representation of female researchers varied by discipline in the horizontal dimension, and by cohort in the vertical dimension. The bottom row of the map represents the overall proportion of female researchers in each discipline in Germany over the 1996-2020 period, as a baseline for comparing the variability in the representation of females among the researchers who left and returned over time.

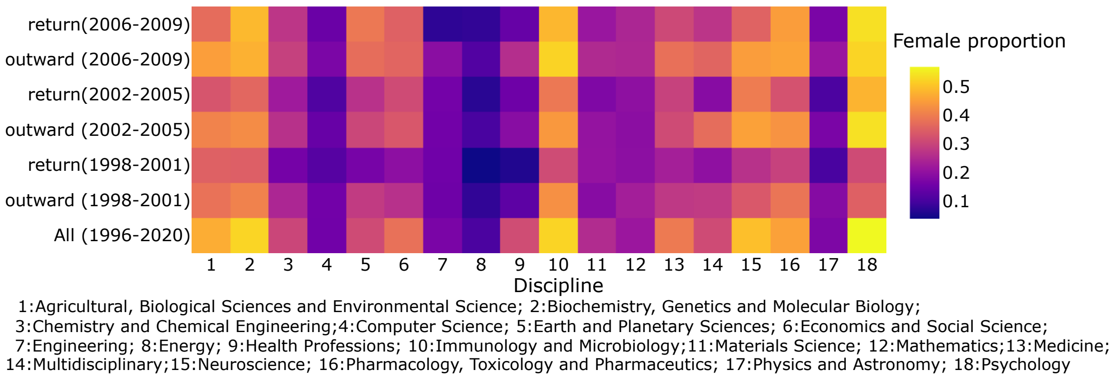


Figure 2.4: Proportion of female researchers in different groups by discipline and cohort.

Compared to the baseline, almost all disciplines appeared to be more male-dominated over time among both outward and returnee researchers, albeit to varying degrees. One exception was the field of mathematics (12), in which female researchers accounted for a higher proportion of each of these two migrant categories in the latest cohort (2006-2009), relative to the long-term pattern. Despite the lower representation of female researchers in both the outward and returnee groups, for the majority of the disciplines, we saw an increasing trend in the proportion of female researchers with each successive cohort, in line with our earlier discussion in Subsection 2.3.3.

When comparing the categories of outward researchers and returnees in the same discipline and cohort, we observed that the proportion of female returnees was generally smaller than the proportion of female outward researchers. For example, when we looked at the latest cohort of researchers in the field of energy, we found that the female proportion among returnees was much smaller than the female proportion among outward researchers. The overall impression provided by these data was that most disciplines were experiencing rising gender disparities, in part because female scientists who left Germany were less likely than their male counterparts to return. Despite substantial efforts to increase gender equality in academia, gender disparities seem to remain substantial across disciplines.

2.3.5 Collaborative ties with Germany and rates of return

In this section, we examined the association between the levels of academic collaboration with German researchers maintained while abroad and the corresponding rates of return to Germany. Fig. 2.5 showed a scatter plot of the return rates (y-axis) and the average collaborative ratios (x-axis) for each discipline. Note that the collaborative ratio is the fraction of publications of an outward researcher (during the period outside of Germany) with a German affiliation. The horizontal (vertical) line indicated the overall average return rates (average collaborative ratio) for outward researchers across all disciplines. The number of returnees for each discipline was represented by the size of the circles. Overall, the Pearson correlation coefficient between the collaborative ratio and the return rates was 0.45, indicating a moderate positive association. Researchers in most health science and life science disciplines, including medicine, health professions, and psychology, were more likely to return to Germany than researchers in other disciplines, as indicated by the higher return rate over the average rate. When we looked at the returnees' levels of academic collaboration with Germany while abroad, we found that health science returnees, as well as researchers in some physical science disciplines, like earth and planetary science, were more likely to maintain academic ties with Germany, as shown by collaborative ratios that exceed the mean values of 33%. Specifically, we observed that health science researchers maintained stronger collaborative ties with Germany, and were more likely to return; whereas researchers in STEM fields, who tended to leave Germany without maintaining as many collaborative ties, were less likely to return.

Next, we looked at the association between collaborative ties and return rates among outward researchers by cohort. The results disaggregated by cohort were

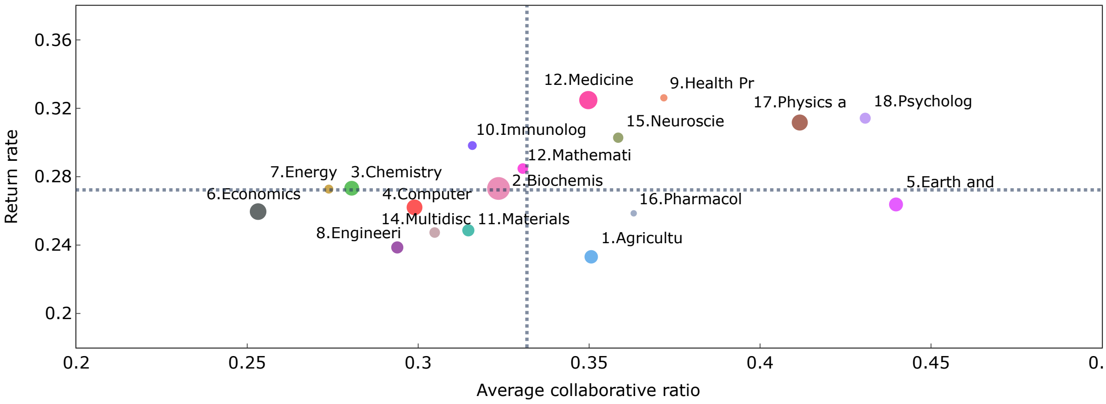


Figure 2.5: Return rates and collaborative ratios across disciplines.

shown in Fig. 2.6, with the average return rates and collaborative ratios in each cohort represented by the horizontal lines and the vertical lines, respectively. Our results showed an overall decreasing trend in rates of return by cohort, but the left-truncation of the data complicated the reliable investigation of trends involving the first cohort. Despite the general trend, researchers in health professions and medicine were more likely to return than researchers in other disciplines. The collaborative ratios grew slowly but steadily with each cohort; thus, researchers in the latest cohort maintained relatively strong collaborative ties to Germany. Similar patterns can be observed separately for most disciplines.

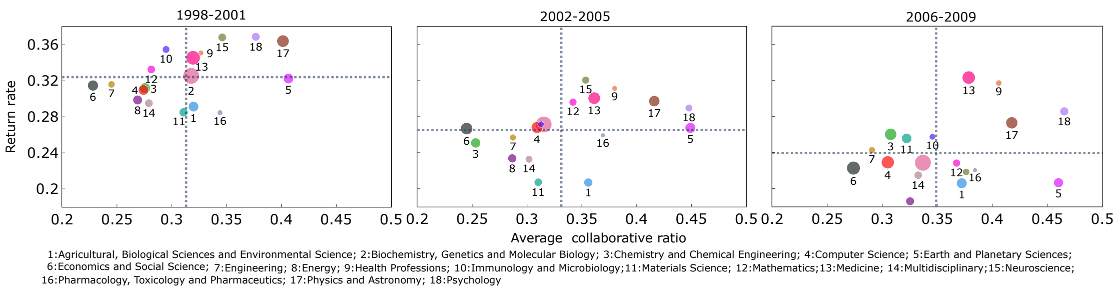


Figure 2.6: Return rates and collaborative ratios by discipline and cohort.

The correlations found between the collaborative ratios and return rates in the first two cohorts were in line with the overall pattern shown in Fig. 2.5, with Pearson correlation coefficients of 0.43 and 0.41, respectively. This association became much weaker (the correlation coefficient was 0.29) in the latest cohort, whose discipline averages appeared to be scattered widely across the four quadrants. Between cohorts 2 and 3, neuroscience dropped from quadrant 1 to quadrant 4, indicating a sharp decrease in return rates, despite an increase in academic links with Germany. Between cohorts 2 and 3, we saw an increase in collaborative ratios among the outward researchers in the fields of chemistry and chemical engineering, accompanied by stable return rates. For most other disciplines, however, the return rates tended to decrease, as shown in Fig. 2.6.

2.4 Discussion and future directions

As “science brokers," researchers develop innovative ideas and make scientific contributions by combining information and resources in various domains using specialized skills and knowledge, which they acquire at different institutions and geographical locations (Williams, 2007). International experience can play a substantial role in helping researchers accumulate knowledge, information, and capital, and can thus contribute to their scientific research and academic careers (Teichler, 2015; Wang, 2020). The previous study found that internationally mobile researchers accounted for over 16% of the population of Scopus-published researchers who had affiliation ties to Germany over the 1996-2020 period (Zhao et al., 2021). We also observed that despite representing a minority in the German science system, mobile researchers make substantial contributions, as evidenced by the finding that compared to non-mover researchers in Germany, they had higher annual citation rates (Zhao et al., 2021). Because of their more nuanced trajectories and international experience, returnees can make important contributions to the German science system. Here, we analyzed the return migration of researchers to Germany from several perspectives; i.e., by taking into account their disciplines, cohorts, genders, and levels of collaboration with Germany while abroad.

Our quantitative results for Scopus-published researchers with ties to Germany provide further evidence to support previous findings. The results of our comprehensive analysis of emigration and return migration as two outcomes for researchers who left the German science system indicate that the age and gender compositions of outward researchers and returnee researchers differed from those of non-movers. The median age for returnee researchers was up to six years higher than that of non-movers, which suggests that there were substantial differences in their levels of experience. All three groups of researchers differentiated by their levels of experience, from early-career to senior, were heavily dominated by men. The ongoing gender disparities we found throughout the academic life cycle were in line with the findings of previous studies (Vásárhelyi et al., 2021). In particular, we observed that the publishing careers of male returnees were, on average, longer than those of other groups, with more than half of them being in their senior career stage.

The countries receiving the largest flows of researchers from Germany were shown to have some of the highest return rates as well. However, we also found that of the large numbers of German-affiliated researchers who moved to Switzerland, Sweden, Austria, and Australia, relatively small proportions returned to Germany. Three of these countries have linguistic, cultural, and geographic proximity to Germany. Moreover, they all have higher R&D spending per GDP (Organisation for Economic Co-operation and Development (OECD), 2021) than the UK and the US (and three have higher R&D spending than Germany), which has enabled them to succeed in attracting and retaining published researchers from Germany.

Supporting the representation of female researchers in academia through equitable policies is imperative for Germany (Lutter and Schröder, 2019), and for other countries (Morgan et al., 2021a). The trajectories of internationally mobile female researchers are a particularly important dimension in evaluating a national science system. We analyzed the gender differences among outward researchers and returnees. Our results indicated that the gender disparities in the German science system tend to be intensified over cohorts. Consistent with evidence showing that the representation of female researchers in academia has risen over time (Huang, Gates, et al., 2020), we found that the proportion of female researchers has increased among both outward researchers and returnees across cohorts, taking into account the number of years between first publication, departure from Germany, and return to Germany. However, the proportion of female researchers among those who returned to Germany was lower than it was among those who left, indicating that female outward researchers have a greater tendency than their male counterparts to live abroad for longer periods, or possibly to settle down in other countries. When we looked at the proportions of female researchers in the two subpopulations of interest disaggregated by cohort and discipline, we found that both the outward and the returnee subpopulations in most disciplines were more male-dominated than the overall population of researchers in that discipline, in line with the greater gender disparities observed among all German-affiliated migrant researchers in most disciplines (Zhao et al., 2021). These findings suggested that the gender imbalance in the German science system (with respect to scholars who started publishing in Germany) may be intensified by the subgroups who were returning to Germany being more male-dominated than the subgroups who were leaving Germany.

Finally, we looked at the interplay between the degree to which researchers continued to collaborate with German institutions while abroad, and their corresponding return rates. The results showed a positive, moderate association between collaboration and return rates across disciplines. After cohorts were introduced into the analysis, the return rates decreased with successive cohorts, while the collaborative ratios increased on average. In the fields of medicine, health professions, physics, and psychology, the likelihood of collaborating with Germany and that of returning to Germany were both higher than the total averages. In contrast, researchers in the fields of engineering, computer science, and economics had both lower collaboration and lower return rates than the total average. To tackle the challenge of talent loss in STEM fields, and to attract and retain STEM researchers from abroad, Germany, –which already has a large number of initiatives for international researchers, like GAIN and GSO– would likely benefit from developing additional programs focused on STEM fields (Organisation for Economic Co-operation and Development (OECD), 2015).

This chapter has several limitations, which can be addressed only through ongoing work and additional efforts. Our bibliometric analysis was based on the higher quality signals for researchers who have higher publication rates. Therefore, the reliability of our findings may not be the same for all fields, given that their average publication rates vary (e.g., physics vs. history). Another limitation is that we could not analyze migration events that were not captured in the publication data. In addition, because of the possible differences between publication years and migration years, the temporal patterns of the data should be interpreted with caution.

We recognize that bibliometric data, like other sources of big data, are not produced for use as research data, and are therefore susceptible to potential biases or errors. In our materials and methods, we outlined a series of pre-processing steps for systematically dealing with some of the data quality issues in our application context. Additional scientometrics research is needed to better identify the potential quality problems with bibliometric data, and to find systematic and effective remedies for addressing them.

As well as contributing to the literature on the migration of researchers (Aman, 2018a, 2018b; Andrey and Elena, 2019; Aref et al., 2019; El-Ouahi et al., 2021; Miranda-González et al., 2020; Moed and Halevi, 2014; Robinson-Garcia et al., 2019; Subbotin and Aref, 2021) in the context of Germany (Netz and Jaksztat, 2014; Parey et al., 2017; Zhao et al., 2021), more importantly, our research fills a critical gap in the research on the return migration of scholars, which is a novel subject in the bibliometric analysis of academic migration. This work, which represents a continuation of (Zhao et al., 2021), was aimed to provide a policy-relevant descriptive analysis of return migration among researchers by taking their levels of experience, gender, disciplines, and cohorts into account. Obtaining insights into researchers who have left Germany, including their age, gender, and characteristics that could influence their potential return to Germany, is a key step towards understanding migration among scholars as a concept that is more nuanced than a one-off relocation event.

Several interesting questions remain to be investigated, including the question of what personal and professional factors drive the international migration of researchers. Differences in levels of support for parenthood between Germany (Gangl and Ziefle, 2009; Lutter and Schröder, 2019) and other countries (Morgan et al., 2021a) may have a bearing on some of the observed gender disparities. Combining different data sources could allow us to expand the analysis and examine other critical topics, like parenthood policies. Investigating the citation performance of outward and returnee researchers could provide us with additional insights into the individual-level consequences of scholars' migration decisions. In addition, the observed association between return migration and personal and professional factors, including disciplines and collaborative ties, can be further investigated to find the mechanisms involved, such as the emergence of discipline-specific centers that are particularly attractive for migrant researchers.

2.5 Data and Materials Availability

The inferred migration events dataset is publicly available in a FigShare data repository:https://doi.org/10.6084/m9.figshare.18433139. Scripts and data which allow for replication of our analysis are publicly accessible on GitHub under https://github.com/zxy919781142/Return-migration-of-German-affiliated-researchers.

3

A gender perspective on the global migration of scholars: Do more female researchers than ever migrate internationally?

This chapter is co-authored with Aliakbar Akbaritabar, Ridhi Kashyap, and Emilio Zagheni. Materials from this chapter have been published as Zhao, X., Akbaritabar, A., Kashyap, R., and Zagheni, E. (2023). "A gender perspective on the global migration of scholars." *Proceedings of the National Academy of Sciences*, 120(10).

Abstract

Although considerable progress toward gender equality in science has been made in recent decades, female researchers continue to face significant barriers in the academic labor market. International mobility has been increasingly recognized as a strategy for scientists to expand their professional networks, and that could help narrow the gender gap in academic careers. Using bibliometric data on over 33 million Scopus publications, this chapter provides a global and dynamic view of gendered patterns of transnational scholarly mobility, as measured by volume, distance, diversity, and distribution, from 1998 to 2017. We find that, while female researchers continued to be underrepresented among internationally mobile researchers and migrated over shorter distances, this gender gap was narrowing at a faster rate than the gender gap in the population of general active researchers. Globally, the origin and destination countries of both female and male mobile researchers have become increasingly diversified, which suggests that scholarly migration has become less skewed and more globalized. However, the range of origin and destination countries continued to be narrower for women than for men. While the United States remained the leading academic destination worldwide, the shares of both female and male scholarly inflows to that country declined from around 25% to 20% over the study period, partially due to the growing relevance of China. This chapter offers a cross-national measurement of gender inequality in global scholarly migration that is essential for promoting gender-equitable science policies and for monitoring the impact of such interventions.

3.1 Introduction

Over the past 50 years, women have made enormous strides in scientific research, including in the fields of science, technology, engineering, and mathematics (STEM) (Elsevier, 2020a; Directorate-General for Research and Innovation (European Commission), 2021b). Nonetheless, women continue to face a number of barriers to participation, recognition, and progression in the scientific arena (Christian, 2018; Larivière et al., 2013; Macaluso et al., 2016). In the current era of globalization, international mobility is increasingly recognized as a key strategy for scientists seeking to participate in global scientific networks and collaborations, and to advance their careers (European Commission and Directorate-General for Research and Innovation, 2018; Weert, 2013). However, less attention has been paid to gender differences in international scholarly migration, especially on a global basis (Christian, 2018; Huang, Gates, et al., 2020; Macaluso et al., 2016; Zhao et al., 2021). This chapter considers the interplay between the globalization of scientific knowledge, the internationalization of academia, and gender inequalities in the academic labor market (Jonkers and Cruz-Castro, 2013; Jöns, 2011a; Shauman, 1996), with the aim of providing substantive support for policies that advance gender equality in academia.

While the population of female scientists and scholars has more than doubled since 1993, and a wide array of programs promoting gender equality in academia have been launched, gender disparities persist in nearly all facets of academia and sciences (Huang, Gates, et al., 2020; Directorate-General for Research and Innovation (European Commission), 2019). In 2016, women researchers held 41% of academic positions across the 28 countries of the European Union (EU-28). However, in many European countries, including in the Netherlands and Germany, women held fewer than one in five senior academic positions (Directorate-General for Research and Innovation (European Commission), 2019). Women are also underrepresented as researchers in Asian countries such as Japan, where they account only for approximately one in four full-time faculty members (Government of Japan, 2019). Female researchers in the Global South are relatively "invisible" compared to those in the Global North (Fresnoza-Flot, 2022), and their representations among researchers in Guinea (6%), Ethiopia (7.6%), and Mali (10.6%) show more alarming gender disparities (Sougou et al., 2022). While it is clear that the sciences and academia continue to be dominated by males at the global scale, there is also substantial variation in levels of gender inequality across countries. Unfortunately, unified and comprehensive statistics suitable for making cross-national comparisons of gender disparities in the sciences do not exist (Akbaritabar and Squazzoni, 2021; Auschra et al., 2022; Teele and Thelen, 2017), let alone statistics on gender disparities in global brain circulation. A first goal of this chapter is to document cross-national trends in a systematic way.

The existing research that has considered the gender dimension in international scholarly migration has mainly focused on either emigrants from an origin country perspective (Docquier et al., 2009; Dumont et al., 2007), or on immigrants from a destination country perspective (Jöns, 2011a; Muchomba and Kaushal, 2022; Nascia et al., 2021). Although some of these studies have discussed scholarly mobility involving several countries, most have paid little attention to their separate roles as receiving and sending countries, mainly owing to a lack of relevant data (Dumont et al., 2007). A clear-cut division between immigration and emigration countries can deepen the understanding of the extent to which female and male researchers respond to the push and pull forces in any given country, which in turn shape the gender composition of national science systems. Within the global migration system more broadly, the increasing interconnectedness and integration of countries around the world have led to a more distinct pattern of international migration, where migrant populations have been coming from an increasingly diverse range of countries, but have been moving to a shrinking number of prime destination countries (Czaika and de Haas, 2014). A similar pattern has also been observed among highly-skilled migrants (Kerr et al., 2016). Looking specifically at the population of academic scientists, movements from the Global South to the Global North, and from Asia to English-speaking countries like the US, the UK, and Canada, have been the long-established paths (Kirloskar and Inamdar, 2021; Song and Liang, 2019; Wiesel, 2014). In recent decades, also as a result of numerous programs aimed at attracting overseas researchers, and of changes in socioeconomic and geopolitical conditions, non-English-speaking Western countries, like Switzerland, Germany and France (Franzoni et al., 2015; Organisation for Economic Co-operation and Development (OECD), 2013), as well as some Asian countries, such as China and Singapore (Cao et al., 2020; Chou, 2020; Science|Business, 2022), have become increasingly attractive for international researchers. However, the extent to which the diverging destinations are shifting, or the established asymmetric and skewed migration patterns among researchers are changing, has not been assessed in depth.

Emerging evidence has shown that men and women do not respond to the regional pull and push factors with the same intensity when making migration decisions (Docquier et al., 2009). Despite the increasing average migration distances for mobile researchers (Czaika, 2018; Czaika and Orazbayev, 2018), women are more likely to concentrate in the largest urban centers. Also, wherever they reside, women are less likely than men to relocate (Kulis and Sicotte, 2002; Yoon and Kim, 2018). Generally, female researchers are less geographically mobile than their male counterparts (Rosenfeld and Jones, 1987). From the perspective of reconciliation of career and family, married academic men are more likely than academic women to relocate to small communities where there are fewer academic positions available and are less likely than women to choose positions in large metropolitan areas (Kulis and Sicotte, 2002). By contrast, women academics are more likely than their male counterparts to make job shifts within the same locality instead of pursuing cross-border mobility. Given these considerations in the literature, we hypothesize that female researchers generally migrate shorter distances and have a lower diversification level in both scholarly immigration and emigration, compared to their male counterparts.

Existing research highlights geographic constraints on women's careers in academia (Marwell et al., 1979), and evidence is accumulating that large differences in labor market conditions and in women's rights between the origin and the destination countries have led to larger migration flows of highly-skilled females to specific destinations (Adsera et al., 2022; Kerr et al., 2016; Nejad and Young, 2014). Thus, gender disparities in specific bilateral migration corridors may vary substantially, contributing to distinct migration trajectories and distributions of female and male researchers. More broadly, it has been argued that social inequalities are produced and reproduced within the globalization of the science system (Bilecen and Van Mol, 2013; Findlay, 2010; Torche, 2014). In particular, gender plays a significant role in shaping international academic mobility patterns (Bauder, 2015; Bilecen and Van Mol, 2013; Holloway et al., 2012; Shinozaki, 2017; Sondhi and King, 2017), and underrepresentation of female migrant researchers has been observed, in various degrees, in country-specific analyses (Bilecen and Van Mol, 2013; Guthrie et al., 2017a; Leemann, 2010; Rosenfeld and Jones, 1987). Due to the lack of time series of comparative and gender-disaggregated data on migration of scholars, we do not have a clear picture of gender inequalities in global patterns of migration of scholars. However, based on the literature, we hypothesize that, while women are generally under-represented in academia and among internationally mobile researchers, there is also substantial heterogeneity in levels of gender inequality across countries.

The migration literature, more broadly, has highlighted a growing feminization of migration, indicating an increasing share of women among all migrants and the tendency of women migrating more independently of men (Donato and Gabaccia, 2015, 2016; Gouws, 2010). The idea of feminization of migration has grown in importance in the new age of international migration and globalization, with the doubling of female migrants during the period 1960–2015, and with relatively equal shares of women and men in the migrant population (Donato and Gabaccia, 2015, 2016; Gabaccia, 2016). The feminization of international migration not only relates to the increasing figure of female migrants, but also to the fact that women increasingly migrate independently, in search for jobs, instead of depending on marriage and families (Andall, 2013; Caritas Internationalis, 2012). Research has also shown increasing proportions of well-educated women from the Global South (or low- and middle-income countries) concentrating in more economically developed countries in the Global North (Docquier et al., 2009; International Organization for Migration and Organisation for Economic Co-operation and Development (OECD), 2014; Nejad and Young, 2014). For example, the comparison between the education- and gender-specific worldwide migration in 1990 and that in 2000 indicated an increasing participation of high-educated women in South-to-North emigration (Docquier et al., 2009). However, the question of whether the feminization process has also emerged for global transnational scholarly mobility, and exhibits similar patterns, has received limited attention. In light of these trends suggestive of a feminization of international migration, and also because of the emergence of an increasing number of programmes for supporting female researchers in academic mobility to Global North destinations (Leggon et al., 2015), we hypothesize that the under-representation of women among internationally mobile scholars has been decreasing over time, and that an increasing number of female scholars from the Global South (or low- and middle-income countries) have been concentrating in more economically developed countries in the Global North.

To fill these and related gaps in our understanding of gender inequalities in scholarly migration, and to test our hypotheses, we use data from Scopus on over 33 million publications. Our data and methods enable us to estimate international mobility of researchers from around the world, and by gender, during the 1998–2017 period. We aim to assess how gender inequality among mobile academic scientists varies across countries and over time on a global scale, and how it affected the demographic composition of the scientific workforce across the origin and the destination countries, by answering the following questions:

1. To what extent does gender inequality among transnational mobile researchers vary by country and time? How does it compare to the gender composition of overall researcher population worldwide?

2. At the global level and at the level of individual countries, how do the scholarly migration spread by gender, among both emigrants and immigrants? How do these gendered spread patterns evolve with time?

3. What are the dominant destinations for globally mobile researchers at different time periods? More specifically, how does the gender difference in destinations and its distribution vary by origin countries?

3.2 Data, Methods, and Measurements

3.2.1 Bibliometric data on global publications from Scopus

This chapter relied on a large-scale bibliometric dataset consisting of more than 33 million Scopus article and review publications, and on an exhaustive dataset covering 10 million published researchers worldwide between 1996 and 2020, which have been disambiguated using Scopus author identification numbers (Baas et al., 2020). The researchers in this dataset not only include the academics affiliated to universities and institutions, but also involve those working in industry with research publications in Scopus. While Scopus covers a longer time period, we used the period window starting in 1996 due to license limitations and the quality of the metadata. We cleaned and pre-processed bibliometric information according to the steps in Section 2.2.2, Chapter 2, and further mapped the geographic locations of the published researchers by linking each author's affiliation and publication (Zhao et al., 2021, 2022a). These processing steps laid the foundation for identifying mobile researchers and their transnational trajectories, and for analyzing patterns of global scholarly mobility over time.

3.2.2 Active academic publishing period and researcher population

In this chapter, we first constructed the active academic publishing period for each author. For any author A, we can obtain a sequential collection of publications, composed of the corresponding publishing years and affiliation countries. The publication collection is defined as:

$$\text{Publications}(A) = \{(T_1, C_1), (T_2, C_2), ..., (T_i, C_i), ..., (T_n, C_n)\} \tag{3.1}$$

Here, T_i and C_i represent the publishing year and mode residence country(ies) of the ith publication and n is the number of all publications of author A.

Given that authors may not publish every year but are still active in research work, some host countries could be missed during the gaps between the publication years. These authors did not contribute to the researcher population in those non-publication years, which could result in a discontinuous academic period and lead to an underestimation of the overall researcher population. To prevent this, we implemented a two-year padding (vicinity) to fill the gaps in the years between publications and to estimate the annual researcher population. Given the two sequential publication instances (T_i, C_i) and (T_{i+1}, C_{i+1}), when T_i and T_{i+1} are not consecutive years, we fill the gap to estimate the researcher population by considering the two-year padding of (T_i, C_i) and (T_{i+1}, C_{i+1}). Table 3.1 lists all possible situations that can be used to fill the gap between T_i and T_{i+1}.

Table 3.1: Two-year padding for the residence countries between publishing years

Case	Year gap	Filling instance
1	$(T_{i+1} - T_i) = 2$	$(T_{i+1} - 1, C_{i+1})$
2	$(T_{i+1} - T_i) = 3$	$(T_i + 1, C_i), (T_{i+1} - 1, C_{i+1})$
3	$(T_{i+1} - T_i) = 4$	$(T_i + 1, C_i), (T_{i+1} - 2, C_{i+1}),$ $(T_{i+1} - 1, C_{i+1})$
4	$(T_{i+1} - T_i) \leqslant 5$	$(T_i + 1, C_i), (T_i + 2, C_i),$ $(T_{i+1} - 2, C_{i+1}), (T_{i+1} - 1, C_{i+1})$

Similarly, the filling instances before T_i and those after T_i can be determined by considering the extracted information from adjacent publications. The filling instances were then used to estimate the researcher population for countries in each year. For example, in case 2, author A can be considered an active author in C_i during the years T_i and $T_i + 1$. In addition, the author is assumed to be involved in the researcher population of C_{i+1} during the years $T_{i+1} - 1$ and T_{i+1}.

After constructing the active academic period of each researcher, we can obtain the research population in year i and country c from the number of active researchers in year i and country c.

3.2.3 Migrant researchers

Transnational academic migrants were identified based on whether the authors had ever been affiliated with universities or institutes in a country other than their country of origin through their publications (Elsevier, 2020a; Miranda-González et al., 2020; Zhao et al., 2021). Similar to the method in Chapter 2, the migration of researchers was identified by the changes in the affiliation addresses across different publications over time. The main difference here is that the identification process is based on the constructed active academic period of each researcher after filling some years without publications. The residence country(ies) in each year was assigned to each author according to his or her most frequent (mode) country(ies) during the publishing year and that of the most recent years of publishing, if needed. We assumed that an author migrated when his or her country of affiliation changed (Moed and Halevi, 2014; Zhao et al., 2021).

More specifically, to determine the moving year between T_i and T_{i+1} in the case of migration when C_i is different from C_{i+1}, we first assumed the middle year between T_i and T_{i+1} is T_{mid}, which was calculated by the ceiling method (ceiling is a function that maps any real number to the least integer greater than or equal to this number). For example, T_{mid} is 2003 when T_i and T_{i+1} are 2001 and 2004. We then considered the following circumstances to determine the migration events of author A:

(a) If T_{mid} is within the two-year padding threshold of T_{i+1}, that is, $(T_{i+1} - T_{mid}) \leqslant 2$, we take T_{mid} as the moving year T_{move} in which the author moves from C_i to C_{i+1}. For example, when T_i and T_{i+1} are 2001 and 2004, we consider the movement that happened in 2003.

(b) If T_{mid} is outside the two-year padding threshold of T_{i+1}, that is, $(T_{i+1} - T_{mid}) > 2$. In this case, there is a large period gap between the two publishing years. Thus, we apply the two-year padding (vicinity) method by taking $(T_{i+1} - 2)$ as the moving year T_{move}. For example, when T_i and T_{i+1} are 1999 and 2007, the year 2005 is determined to be the moving year. This is based on the assumption that each publication signals the output of work that has happened in the past two years (to account for the publication delays in different disciplines) (Björk and Solomon, 2013).

In this case, we can obtain this movement event for author A:

$$\text{Migration} = \{T_{move}, C_{origin}(\text{i.e.,}C_i), C_{destination}(\text{i.e.,}C_{i+1})\} \tag{3.2}$$

Similarly, the author who has at least one movement during her or his publishing career can be identified as a migrant researcher, and the respective movements can be extracted from the publication information.

To reduce the effects of yearly fluctuations and to compare the migration trends over time, we considered four time periods (1998–2002, 2003–2007, 2008–2012, and 2013–2017) in this study, and grouped the migration events according to the migration year. Notably, we excluded the years of 1996, 1997, 2018, and 2019 in the time periods due to the application of the two-year padding method in the population estimation.

3.2.4 Gender detection from first names

Gender detection is a key step in measuring the relative participation of female versus male authors in bibliometric analysis (Halevi, 2019). The only form of Scopus metadata that could be used to infer an author's gender was the author's first name (Halevi, 2019). Based on the gender detection method mentioned above (Section 2.2.2, Chapter 2), this chapter further elaborated the steps of inferring gender from first names in the following steps:

Problematic first names and implemented solutions

Before making name-to-gender inferences, we identified the issues related to problematic first names in the Scopus data, including *inconsistent names for a single author*, *combined names*, and *unavailable names*. The following strategies were used to deal with the different cases of problematic names:

(a) *Inconsistent names*: One author ID (i.e., the unique identification number assigned by Scopus to differentiate and disambiguate authors (Baas et al., 2020)) may have multiple name variations (e.g., the records of the same author ID may have the names of Christine and Christina, as illustrative examples). **Solution**: The most frequent (mode) name was extracted to assign to this author in order to ensure that each author ID has only one single first name.

(b) *Combined names*: Some first names include special characters or are combined with a middle name (e.g. David A., H. Okti, and Yan Ying).

Solution: If the first name string contains the initials and names (e.g. David A., H. Okti), the non-initials were kept as their first names (e.g., David, Okti); if the string contains multiple non-initials (e.g., Yan Ying), the first non-initial was taken as the first name (Yan).

(c) *Unavailable names*: Some first names may only have initials (e.g. A., H.)

Solution: The genders of these authors were categorized as "unknown".

Detection process

After handling the problematic names, We elaborated a sequential, three-step process to infer the gender identities of researchers from their first names by using 1) a large dictionary of names and genders, i.e., a worldwide gender-name dictionary (WGND) that includes 6.2 million names from 182 different countries (Raffo, 2016); 2) a gender detection tool *Demographicx* based on a deep learning *Bidirectional Encoder Representations from Transformers* (BERT) embedding model with sub-word tokenization (Devlin et al., 2019; Liang and Acuna, 2021); and 3) an application programming interface called *genderize.io*. The genders of the majority of names were identified through WGND, and the remaining first names without genders were further processed, first using by *Demographicx*, and then using *genderize.io*.

To ensure the reliability of our gender detection process, we validated the developed method against two established databases of first names and genders (Akbaritabar and Squazzoni, 2021; Winkelmann, 2022), and we also compared the detection performance of our method and another gender detection method (Kozlowski et al., 2022) (see Section B.1 in Appendix B). Overall, we considered this level of accuracy acceptable to ensure a high level of gender inference results (the validation and comparison results can be found in Supplementary Information). We found that 31.55% of the authors were female and 57.72% of the authors were male, while the genders of the remaining 10.73% of authors were labelled as "unknown". To test the robustness of our results given the missing gender, we assigned the gender of female or male to the researchers missing specific gender information using *Multiple Imputation by Chained Equations* (MICE) (more detail in Section 3.4.1, Robustness Analysis). The results after gender imputation displayed a consistent pattern with the original results before gender imputation.

3.2.5 Measures of migration spread by gender

To explore the diversification of the origin and the destination countries over our study period, we used a migration spread measure (Czaika and de Haas, 2014) that showed how dispersed or concentrated migration trajectories were across all observed bilateral migration corridors in terms of both the origin and the destination countries. More specifically, emigration spread measures the extent to which the migrant population is spread across destination countries, while immigration spread indicates the extent to which migrants come from a diverse set of origin countries. As a key concept of the globalization of science, the migration spread of researchers is closely related to how knowledge and information are transferred and interconnected through the diversity and scope of migratory pathways. We disaggregated the migration spreads by gender to further examine whether the female and the male researchers came from and moved to an equally diverse range of countries.

Country-level migration spread

At the country level, the emigration spread (ES_t^i), defined in (3.3), was used to measure the extent to which bilateral migration flows from any given country i (M_t^{ij}) are diverse in the destination countries during a specific period of time t

$$ES_t^i = 1 - \sum_{j=1}^{n_t} (\frac{M_t^{ij}}{EM_t^i})^2 \tag{3.3}$$

n_t is the number of countries involved in the global scholarly migration during the period t. M_t^{ij} indicates the size of the emigration flows from country i to country

j, and EM_t^i measures the overall volume of the emigration flows from country i by adding up all bilateral emigration flows from country i.

Similarly, we calculated the immigration spreads (IS_t^i) in (3.4), to measure the diversity level of the inflows M_t^{ji} to country i, relative to all the immigration flows to country i (IM_t^i).

$$IS_t^i = 1 - \sum_{j=1}^{n_t} (\frac{M_t^{ji}}{IM_t^i})^2 \tag{3.4}$$

Global-level migration spread

Similar to the country-level migration spread, the global emigration spread (ES_t^{Global}, (3.5)) and the global immigration spread (IS_t^{Global}, (3.6)) measure the extent to which the global migration flows were dispersed across all destination and origin countries during the period t.

$$ES_t^{Global} = 1 - \sum_{i=1}^{n_t} (\frac{EM_t^i}{M_t})^2 \tag{3.5}$$

$$IS_t^{Global} = 1 - \sum_{i=1}^{n_t} (\frac{IM_t^i}{M_t})^2 \tag{3.6}$$

Additionally, given the hypothesis that the migration spreads varied across countries over time, different countries have played different roles in the global migration spread. We also measured the weighted average of the country-specific emigration spread ($ES_t^{Weighted}$, defined in (3.7)) to represent the overall level of emigration spread during the period t. Accordingly, the global weighted average of the country-specific immigration spread ($IS_t^{Weighted}$, defined in (3.8)) can be calculated.

$$ES_t^{Weighted} = \sum_{i=1}^{n_t} ES_t^i \cdot \frac{EM_t^i}{M_t} \tag{3.7}$$

$$IS_t^{Weighted} = \sum_{i=1}^{n_t} IS_t^i \cdot \frac{IM_t^i}{M_t} \tag{3.8}$$

3.3 Results

3.3.1 Gendered patterns and global trends of researcher mobility

Worldwide, the number of researchers who have published in Scopus-indexed outlets has increased considerably during the past decades. Our analyses showed that the number of female published researchers in the 2013–2017 period was nearly three times as large as in the early 1998–2002 period, rising from approximately 0.7 to 1.7 millions. There was also a substantial increase in the number of male published researchers; over the same period, the number roughly doubled from 1.5 to 3 million. We also found a considerably increasing migration intensity of female researchers, in terms of both absolute numbers and the proportion of all female published researchers. The number of female mobile researchers nearly tripled over this period, rising from 29,000 (4.3% of all published female researchers) in 1998-2002 to 79,000 (4.6%) in 2013–2017. By contrast, over the same period, the population of male mobile researchers grew more slowly, roughly doubling in absolute count from 92,000 (6% of all published male researchers) to 167,000 (5.6%). This trend suggests a growing feminization of international scholarly migration, in line with our hypothesis. Table. B.2 in Appendix B shows more details on the migration intensity of specific countries.

To better understand the gendered patterns of global scholarly migration over time, we computed the female-to-male gender ratios in the groups of published researchers and migrant researchers by individually dividing the number of female researchers by the number of male researchers in each group. This measurement can help us assess the gender gaps among all published researchers (vertical axis) and migrant researchers (horizontal axis), as shown in Fig. 3.1. The migrant researchers depicted here include both emigrants from, and immigrants to, each country (The separated result by incoming and outgoing researchers was shown in Fig. 3.7, Robustness Analysis). Fig. 3.1 showed that the overall median gender ratio in both groups increased over the study period, from 0.47 to 0.64 among all published researchers and from 0.32 to 0.5 among the subgroup of mobile researchers. The gap between the fitted regression line and the 45° line shank gradually, declining from 0.42 to 0.24. This pattern indicated that the increase in the share of females among migrant researchers outpaced the increase of females among all researchers. In other words, the gender gap in scholarly mobility was decreasing at a faster pace than the overall gender gap among researchers.

Gender parity and even higher representation of female researchers (female-to-male ratio of one and above) in the number of researchers was achieved in only a small fraction of countries, most noticeably in Portugal and Serbia, where the gender ratios for mobile researchers also approached one. While the values of the two gender ratios were correlated (i.e., in most countries, if female researchers were less represented among all published researchers, they were also less represented among migrant researchers), the values of the gender ratios among migrant researchers were typically smaller than the values of the gender ratios among all researchers. This indicates that female researchers were less internationally mobile than male researchers in almost all of the shown countries. A similar pattern can also be observed in both migration inflows and outflows (see Fig. 3.7). Japan was an example country with a significant gender disparity with a female-to-male ratio below 0.3 throughout the four periods. Yet interestingly, Japan also stood out as the only anomaly country with a relatively higher proportion of females among mobile researchers than all published researchers (see the data points close to or over the 45° line in all four periods for Japan).

In addition, Fig. 3.1 also showed that three clusters of countries have emerged over time. In the first cluster of countries, located mainly in the bottom-left quadrant, there were marked gender disparities over time (e.g., Saudi Arabia, Pakistan, and South Korea). Female scholars in these countries were underrepresented in the population of scholars and in the subgroup of mobile researchers, and their situation

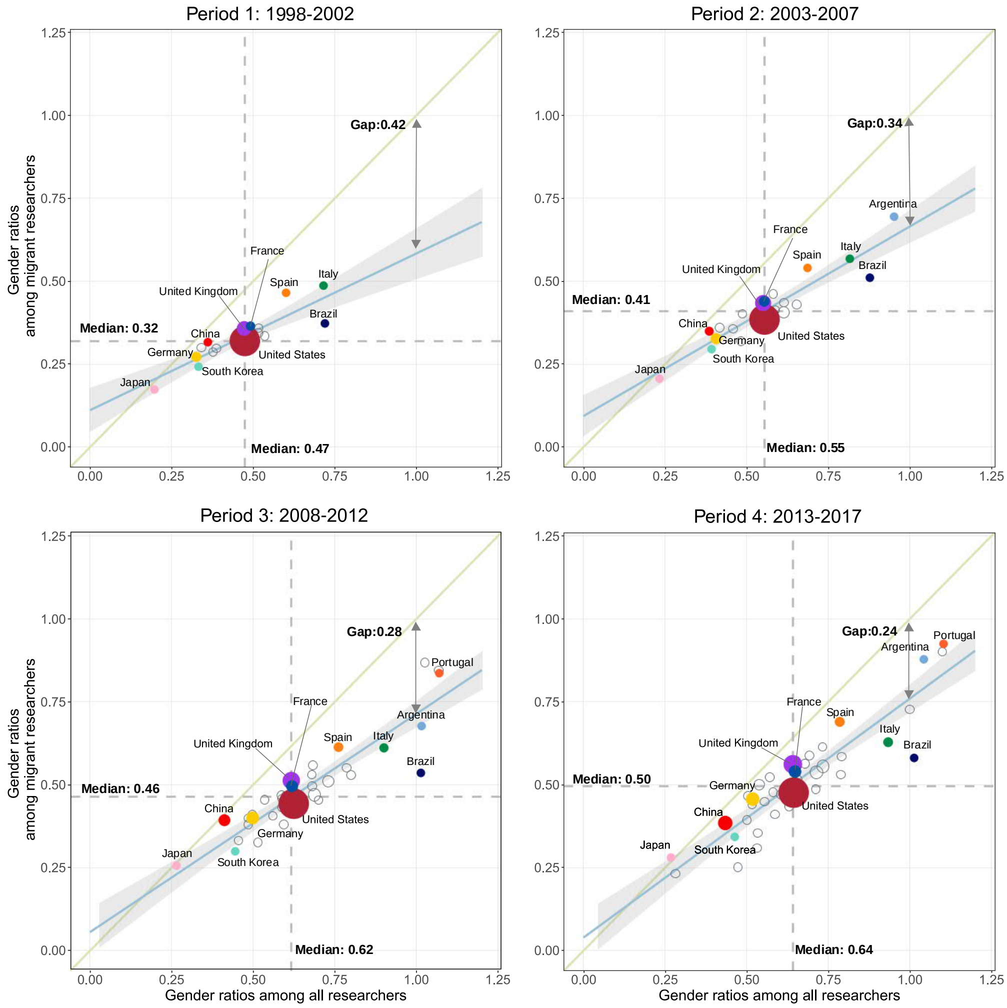


Figure 3.1: Gender ratios among all published researchers (X-axis) and migrant researchers (Y-axis). In the sub-figure for each period, only the countries with over 500 female mobile researchers are shown, as including the countries with small populations of mobile researchers may give rise to bias in the ratio measurements. The size of each country's circle is proportional to the number of female researchers who migrated from and to this country. Notably, to increase the identifiability, the countries with no more than 2,000 female migrant researchers are set to the minimum size. The vertical and the horizontal dashed lines indicate the median gender ratios of all published researchers and of mobile researchers in each period. The 45° line in each sub-figure is used to help compare the gender ratios of these two categories, with another double-arrowed line underlining the distance between it and the fitted regression line at the X value of one. This helps us track the convergence tendency of female representation in the group of mobile researchers versus that in the total researcher population, and how it changed over the four time periods.

did not improve substantially over the four periods. In the second cluster comprising a large share of countries, the proportions of female researchers among mobile scholars and among all active scholars have remained close to the global median values over time. This group included the largest and more established science systems (e.g., USA, UK, Canada, and Germany). The third cluster comprised the group of countries that achieved gender equality at the level of the population of researchers, and that also had a relatively high female-to-male ratio among mobile researchers (e.g., the countries in the top right parts of the figure panels, such as Serbia, Argentina, and Portugal).

The stable and slow growth in the female-to-male ratios with time shown in Fig. 3.1 is consistent with the disaggregated results after controlling the migration direction (inflows and out flows; see Section 3.4.2, Robustness analysis), discipline (Section 3.4.3, Robustness analysis) and income groups of countries (Section 3.4.4, Robustness analysis).

3.3.2 Diversification and diffusion of scholarly migration by gender

Previous literature has shown an increase in the distances of general migration (Czaika and Orazbayev, 2018). This pattern was also seen in the international scholarly migration, where both female and male mobile researchers migrated increasingly longer distances, despite a slight drop in the migration distance of female researchers over the period 2013-2017 (see Fig. B.1, Appendix B). Comparably, male researchers tended to migrate longer distances than female researchers, except for a few countries such as China and South Korea (see Fig. B.2, Appendix B).

To deepen our understanding of the characteristics of global scholarly migration by gender, we further investigated the spreads of migration outflows and inflows. These measurements can help to quantify the extent to which mobile researchers were dispersed across destination countries (emigration spread), and the extent to which they came from a diverse range of origin countries (immigration spread) (Czaika and de Haas, 2014; Czaika and Orazbayev, 2018) (see the definition of migration spread measures in Section 3.2.5). The aim of gender-disaggregated analyses is to assess whether female and male outflows were spread equally across the destination and the origin countries.

When we looked at emigration, we saw that across the four periods, there were, respectively, 103, 119, 137, and 148 distinct origin countries for female researchers, and 141, 149, 160, and 164 distinct origin countries for male emigrant researchers. Conversely, when we looked at immigration, we observed that across the four periods, there were 105, 126, 140, and 147 countries that received female immigrant researchers, and 144, 154, 168, and 173 countries that received male immigrant researchers. Overall, these numbers indicated that more countries were engaged in international academic circulation over time, but also that male researchers originated from and moved to more countries than female researchers in each period, albeit with a narrowing of the gender gap across periods. While the numbers of distinct origin and destination countries were relatively balanced for female mobile researchers, the volume of academic destinations for male researchers increased faster than the range of origin countries.

Fig. 3.2 showed the emigration (left) and immigration (right) spreads of a selection of countries, with the global (country-weighted) levels of spread indicated by the solid line (further information on methods and analyses with different global measures is provided in Section B.4, Appendix B). The countries were selected as the representatives of the three clusters in Fig. 3.1, where gender disparities among mobile researchers were relatively large in South Korea and China, but small in Brazil and Italy. Germany and the US were located in the largest cluster of countries with moderate values of gender ratios, close to the global median. These two countries are also among those that have well-developed science systems and are established academic magnets for global researchers. Additionally, the six countries are located across Asia, Europe, North America and South America.

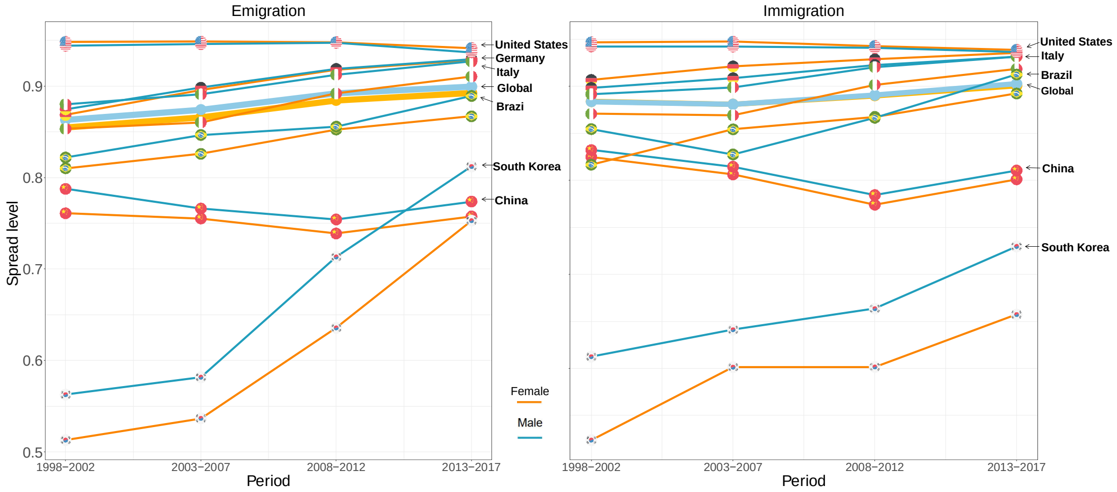


Figure 3.2: Scholarly emigration (left) and immigration (right) spreads, by gender and across four periods, for selected countries (labeled lines) and at the global level (solid, thicker line without a flagged circle).

The solid lines without flags in Fig. 3.2 indicated that global migration spreads among both female and male researchers underwent a stable increasing trend along the two dimensions of emigration and immigration. By the fourth period, the values of the emigration and the immigration spread had converged to a similar level. However, the gender gap in global emigration spreads (blue solid line versus orange solid line) suggested that female mobile researchers were, overall, more likely to move to a narrower range of destination countries across the four periods. Especially, the emigration spreads which started off from a relatively lower level of diversity has nearly kept up with the same level of immigration diversity in the most recent period. National context played an important role in shaping the diversification of scholarly migration in both directions.

Increases in the dispersion of outflows and inflows were broadly observed in most countries, but to varying degrees. The most noticeable increase that can be seen in Fig. 3.2 occurred in South Korea, a country that lagged behind other countries in earlier periods, and then experienced a rapid growth in both immigration and emigration spreads. For China, by contrast, we observed a declining trend in the diversification of both outflows and inflows relative to the earlier periods. These trends indicated that, over time, the outflows of academic researchers of both genders from China have tended to concentrate in a narrower pool of destination countries, and the inflows have been more unevenly dispersed in a shrinking number of origin countries over time. The increasing intensity in the volume of scholarly migration to and from China, but with decreasing migration spreads, indicates strengthening scientific relationships between China and some specific countries. For example, the proportion of high-tech research in China that was conducted in collaboration with the US increased continuously, from 4.6% in 2009 to 16.9% in 2018 (Zhu et al., 2020). While the US remained the country with the largest emigration and immigration spreads, throughout the four periods, it also saw a slight decrease in the migration spreads in the most recent period.

In most countries included in Fig. 3.2, the spread values were lower for female mobile researchers, which indicates that there was less diversification in both origin and destination countries among female mobile researchers, than among male mobile researchers. South Korea stood out as having a clear-cut gender gap in migration spreads: compared to their female counterparts, male researchers leaving South Korea consistently migrated to a broader range of destinations, and male researchers entering South Korea also had higher levels of diversification in their origin countries. The rapid increase in the migration spreads of both female and male researchers did not help to narrow the gender gap in migration diversities. In contrast, in the US and Germany, female mobile researchers had higher levels of immigration diversity than male mobile researchers, which indicated that the female researchers who moved to these countries were more evenly dispersed in their origin countries than their male counterparts. Another striking finding is during the period 2003-2007, Brazil experienced a notable decline in the diversity of origin countries of male inflows, which was temporarily surpassed by the increasing spread of female immigration at the same period. It may be accompanied by a huge influx of male researchers from specific countries and a growing appeal of Brazil for global female researchers. Despite this, the diversification of male inflows increased since then and rebounded to a higher level, even exceeding the global baseline.

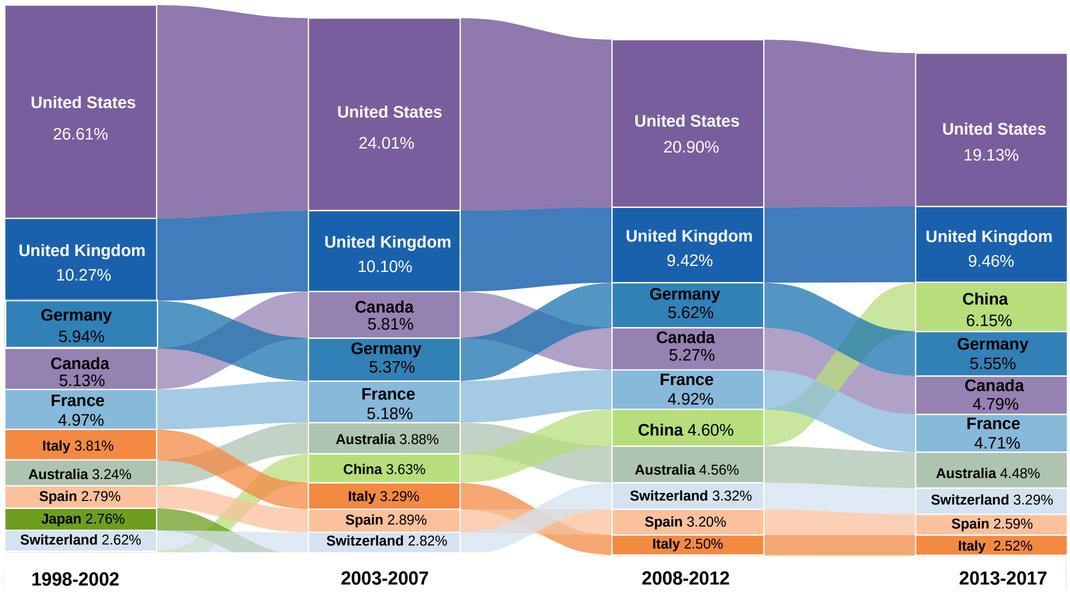
The trends of the scholarly immigration and emigration spreads by gender on a global scale are presented in Section B.4, Appendix B. To investigate the possible divergence of migration paths from and to each country by gender, we compared the average gender difference in the distributions of each country's scholarly inflows and outflows in Section B.5, Appendix B.

3.3.3 Distribution of bilateral scholarly migration flows by gender

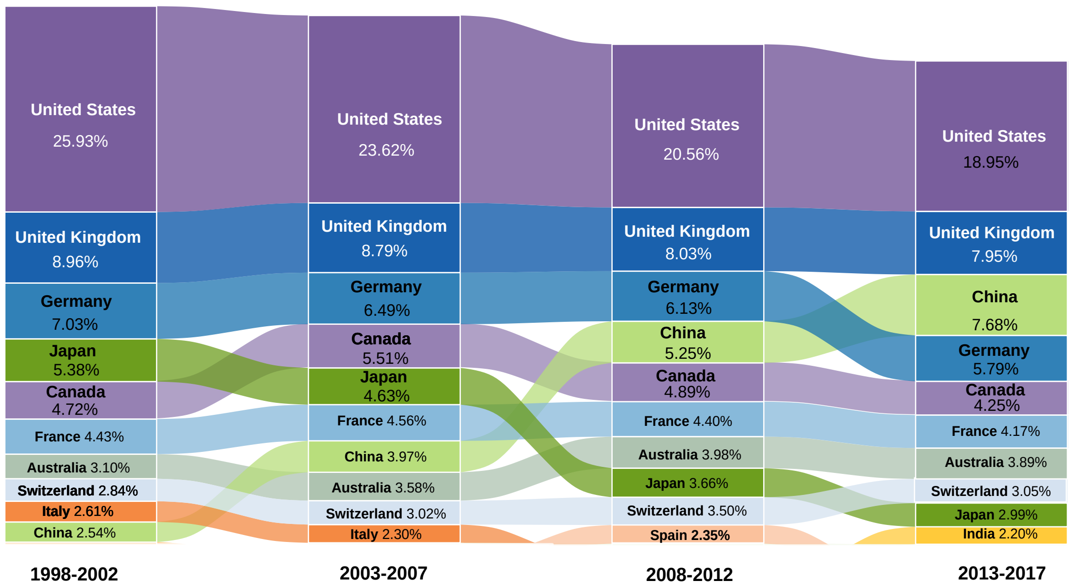
Favored destinations at the global level

The increases in emigration spreads indicated that mobile researchers became increasingly distributed across a more diverse range of destination countries. This trend has shifted the landscape of globally attractive destinations for research talent. Fig. 3.3 shows in more detail the dynamics of the 10 most preferred destination countries, together with the respective gender-disaggregated shares of inflows.

The US, the UK, and Germany were consistently attractive to large shares of female and male mobile scholars. For researchers of both genders, the US had the largest incoming flows, receiving more than one-quarter of global mobile researchers in the first period, and nearly one-fifth of global mobile researchers in the latest period. In the most recent period, China replaced Germany as the third-most popular host country for mobile researchers of both genders. Following closely behind were Canada and France, which received comparable shares of mobile researchers. Japan, by contrast, has been losing its attractiveness for mobile researchers over time. In addition, Japan was an academic destination that was more favored by male researchers than by female researchers. More specifically, the share of female inflows coming to Japan comprised 2.76% of total female migration flows during the first period, which was only half of the proportion of



(a) Female migrant researchers



(b) Male migrant researchers

Figure 3.3: The 10 most preferred destinations for global mobile researchers by gender. The labeled number for each country is the proportion of inflows each country received among all global migration flows by gender. The order of the 10 countries from the top to the bottom reflects the level of their attractiveness from the highest to the relatively lowest.

male inflows to Japan (5.38%). Across periods, Japan has become a less attractive destination for male researchers, falling from fourth to ninth position, as well as female researchers, for whom it has not been a top ten destination since the first period. This finding to some extent explained the aforementioned extreme gender disparity among academic migration across Japan.

More distinct gender disparities can be observed among countries in the lower positions, among the top 10 destinations. The composition and the ranks of the countries ranking seven to 10 for male researchers showed greater variability over time, than the preferred destinations for female migrant scholars. For instance, India emerged recently as the 10th most preferred destination, accounting for more than 2% of male inflows, overtaking Spain in the third period, and Italy in the second period. By the fourth period, both Italy and Spain were no longer among the 10 most preferred destinations for male researchers. In comparison, Italy and Spain continued to receive larger shares of female inflows, in line with the observation in Fig. 3.1 that the female-to-male ratios of migrant researchers in the two countries were relatively high. Over time, the migration flows for researchers of both genders became more evenly distributed across the top 10 destinations, and became less concentrated among the top three destinations. This finding is consistent with the long-term increases in emigration spreads, as shown in Fig. 3.2.

Favored destinations at the country level

Fig. 3.4 presented the three most preferred destinations for female and male mobile researchers from the six countries with different levels of diversification (shown in Fig. 3.2). The differences in the distribution of outflows provided a more detailed explanation of the varying patterns of the country-level migration spreads and migration paths. The outflows to the top three destinations across these countries ranged from a minimum level of concentration of 25% (see the case of the US) to a high level of concentration of 75% (see the cases of South Korea in the two earliest time periods). In the case of the US, the more balanced distribution of outflows to host countries corresponded to a much higher level of emigration spreads from the US, as shown in Fig. 3.2. Meanwhile, scholarly emigration from South Korea was considerably skewed towards the US, indicating the lower diversity of scholarly outflows.

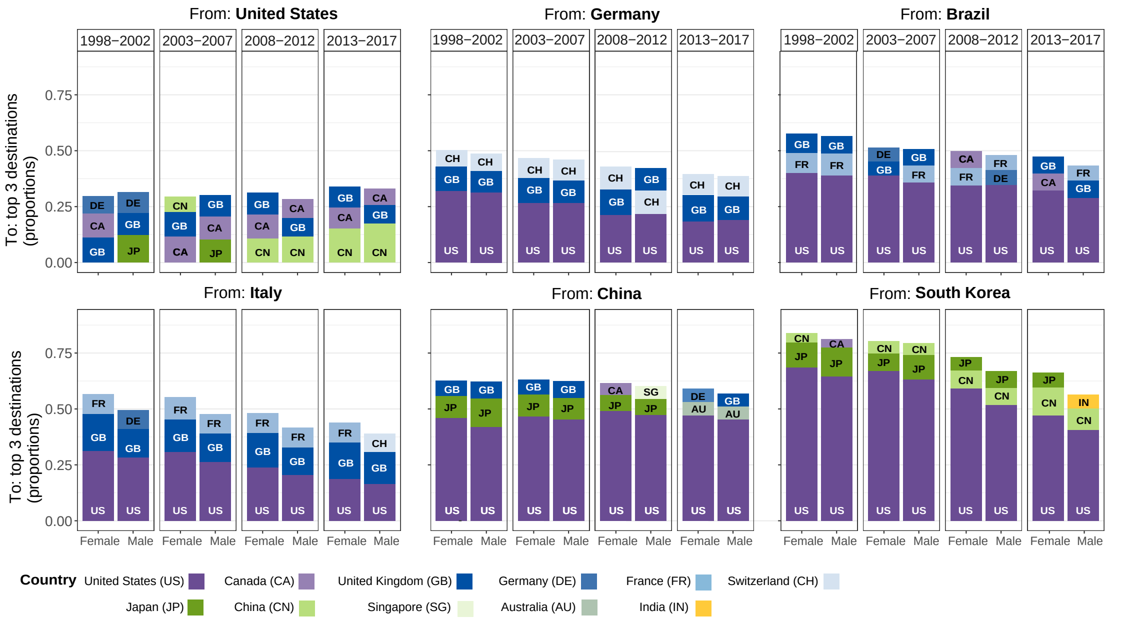


Figure 3.4: Top three destinations for mobile researchers by gender at the country level. Stacked bars are ordered based on the share of mobile researchers who migrated to each of the top three destination countries (indicated by different colors and labels).

Looking specifically at these countries, the US attracted the largest share of mobile researchers from other countries and was especially favored by female researchers. This observation can help to account for the lower emigration spreads observed in these countries. In contrast, the spreads of the outflows of female and male researchers from the US were more balanced across the top three emigration destinations. A notable change for the US in the two most recent periods was the growing share of its outflows to China, which have helped China become one of the most popular destination countries from a global perspective (shown in Fig. 3.3).

The destinations of mobile researchers from each country differed by gender to varying degrees. For instance, while Japan is among the top three destinations for outgoing males from the US in the first two periods, it was not so for females. Female mobile scholars from all selected origin countries, with the exception of the US in the two earliest periods, were always more likely to concentrate in the top three destination countries with slightly larger proportions. This greater concentration of female scholars among the top three destinations was consistent with the generally lower emigration spreads among female scholars than among male scholars. While the largest shares of migration inflows were to the US, the migrations of researchers of both genders tended to link the countries that had stronger cultural ties, and close geographic proximity. For instance, Switzerland (as a German-speaking country) was among the top destinations for Germany's outgoing scholars. Despite the US leading the other countries with a large share of immigrant researchers from South Korea by a wide margin, China was also consistently preferred as a destination by migrants from South Korea and other Asian countries. This preference for destinations with cultural similarities and geographic proximity was consistent across the four periods.

3.4 Robustness analysis

3.4.1 Imputation for missing gender

In our gender identification for authors using their first names (as described above), we had 10.73% of authors whose gender was labelled with "unknown". To assess the robustness of our results, we further examined the impact of these researchers without given genders and to assess the sensitivity of our results to missing names. We imputed the missing genders for the researchers with "unknown" gender using *Multiple Imputation by Chained Equations* (MICE), which is a widely-used method for dealing with missing data (Azur et al., 2011). More specifically, instead of randomly assigning genders to these researchers, we selected the researchers' fields of research (e.g., Agricultural Sciences, Engineering and Technology, Humanities, Medical and Health Sciences, Natural Sciences or Social Sciences, the assignment method is discussed above) and time periods of their first publications (1, 2, 3, or 4 for 1998–2002, 2003–2007, 2008–2012, and 2013–2017, respectively) as features to predict the gender (female or male) of researchers with "unknown" gender. We aimed to see how gender ratios change given that we have relatively reasonable gender information imputed for these researchers. The original global-level gender ratios without "unknown" gender researchers and the result involving these researchers after assigning imputed genders were compared in Fig. 3.5.

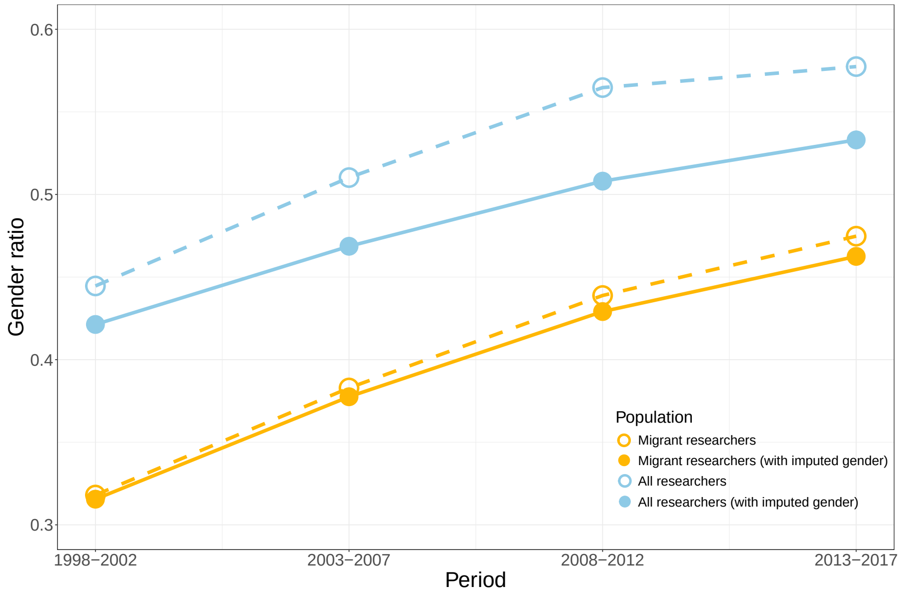


Figure 3.5: The comparison between the gender ratios among all published researchers and migrant researchers before and after imputing missing genders.

For migrant researchers (orange lines), the results showed little difference before and after imputing missing genders. While for the population of all researchers, the gaps became wider at some time periods, the overall trends kept consistent over time.

In Fig. 3.1, the gaps between the 45° lines and the fitted regression lines modelling the relationship between the gender ratios of migrant researchers and those of all researchers are 0.42, 0.34, 0.28, and 0.24 at the X value of one in each time period without gender imputation. After incorporating the researchers with imputed gender, the corresponding gaps change to the values of 0.45, 0.34, 0.30, and 0.25, shown in Fig. 3.6 with the country-level gender ratios of both researcher populations. The gaps throughout the four periods remained consistent, and it is reasonable to believe our findings are not skewed by the composition of those researchers for whom we cannot infer gender by names.

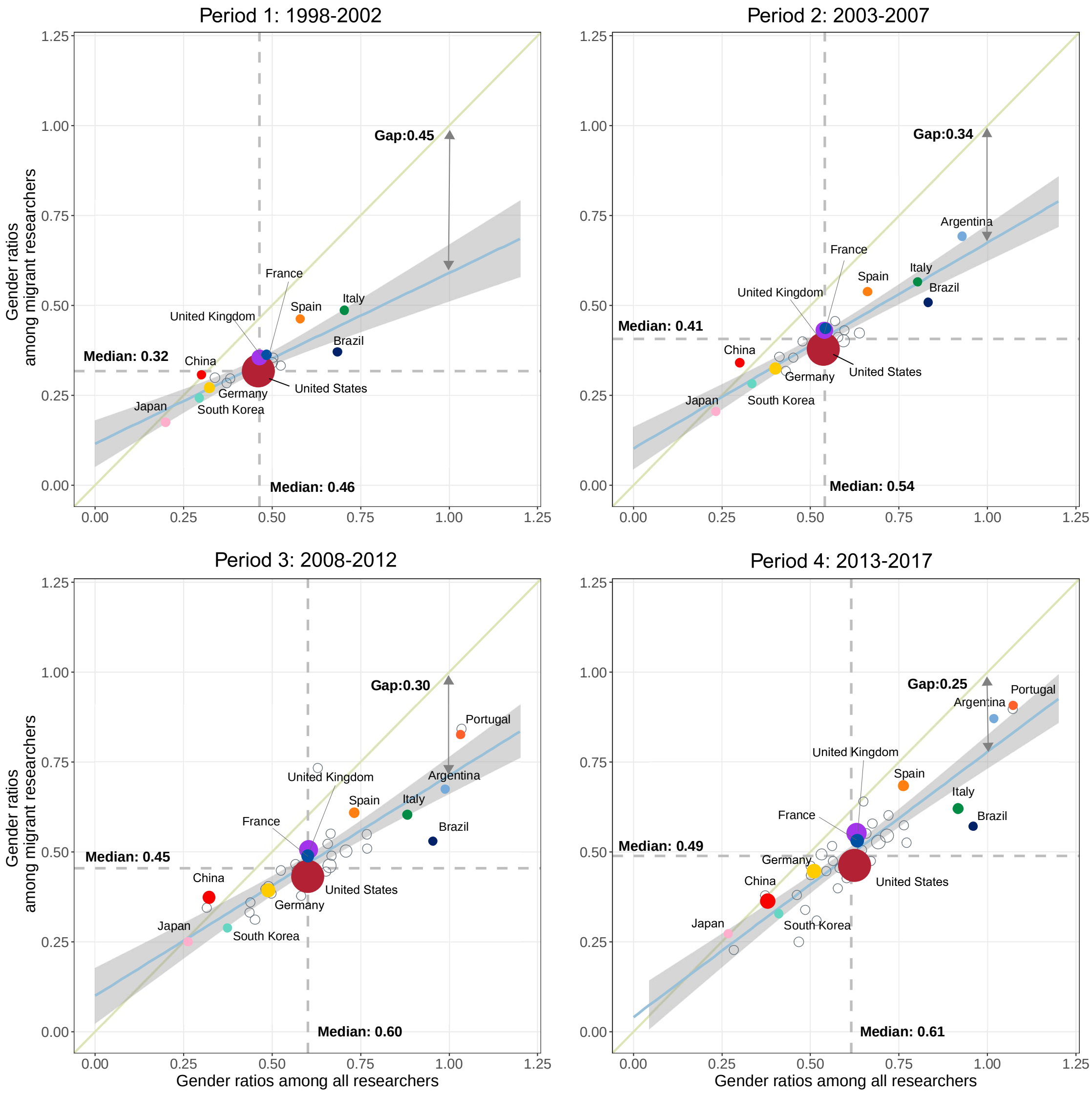


Figure 3.6: Gender ratios among all published researchers (X-axis) and migrant researchers (Y-axis) on country level (after imputing missing genders).

3.4.2 Gender disparities by outflow and inflow

The gender ratios varying in the two directions can contribute differently to the country-level gender disparity. Also, the overall gender ratio among migrants may obscure divergent outcomes for transnational scholarly migration in both directions. To deepen the understanding of how gendered patterns of scholarly mobility have varied globally, we separately look at the gender ratios among outflows and inflows (Fig. 3.7).

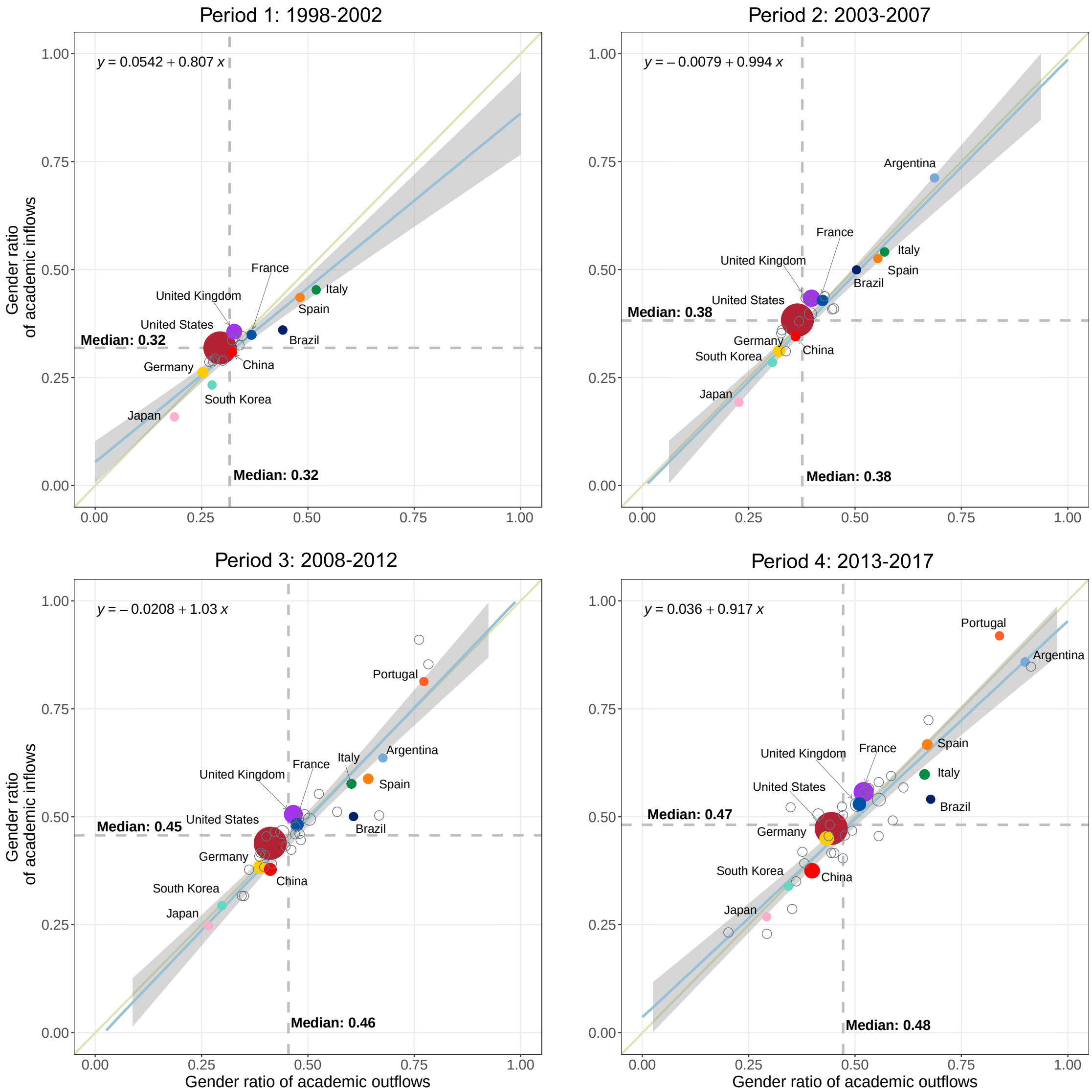


Figure 3.7: The comparison of gender ratios among scholarly outflows (X-axis) and inflows (Y-axis) during the four periods

Generally, the countries appearing in all four figure panels experienced stable and slow growth in the female-to-male ratios in both directions. In addition, the countries emerging in the sub-plots of recent periods (especially period three and period four) that had sufficient numbers of female migrant researchers were more likely to display identifiable patterns of gender ratios far from the median values. These countries either had significant gender disparities (such as Pakistan and Saudi Arabia) or relatively balanced gender relationships (such as Portugal and Serbia), presenting discernible gendered patterns in terms of parity and inequality.

Taken as a whole, the gender patterns of scholarly inflows and outflows across most major countries were consistent over time. A few countries stood out as having a skewed gender ratio in one direction. One example is Brazil. In the most recent period of 2013–2017, around seven female researchers were leaving Brazil for every 10 male researchers who were emigrating. However, the number of female researchers moving to Brazil was nearly half that of male researchers. This indicates that the level of gender balance was higher among the researchers leaving Brazil than among the researchers moving to Brazil.

3.4.3 Gender disparities by discipline

We further examined the gender ratios across different fields of specialty, that is, Agricultural Sciences, Engineering and Technology, Humanities, Medical and Health Sciences, Natural Sciences, and Social Sciences. The category is a mapping of OECD categories (Organisation for Economic Co-operation and Development (OECD), 2007) to the All Science Journal Classification Codes (ASJC) from Scopus. The mapping is done by the German Competence Network for Bibliometrics (Kompetenznetzwerk Bibliometrie, 2021). Only researchers with over 70% of their publications in a given field were assigned to that field of specialty to favor the field with the most proportion in the case of scholars with multidisciplinary publications.

Gender ratios among migrant researchers versus all researchers

We first looked at the female-to-male gender ratios among both the whole population of researchers and migrant researchers in the six fields of specialty at different time periods, shown in Fig. 3.8.

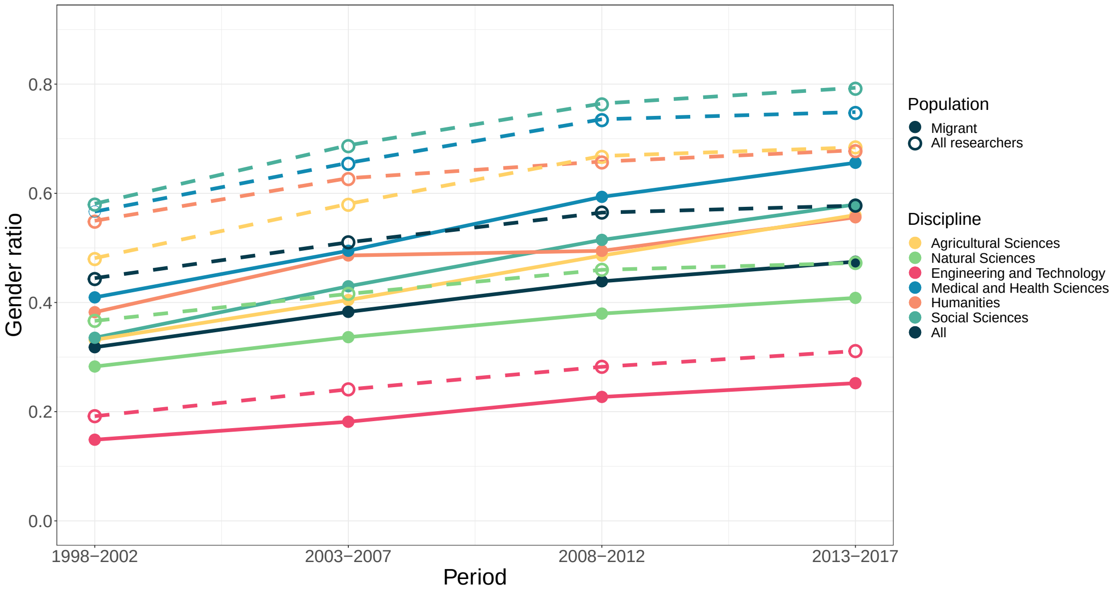


Figure 3.8: Gender ratios among all published researchers and migrant researchers by discipline.

Comparatively, Social Sciences featured more balanced gender ratios among all researchers, followed by the Medical and Health Sciences. However, among migrant researchers, the gender ratios in Medical and Health Sciences surpassed those in the Social Sciences, which indicated women studying Medical and Health Sciences were more likely to engage in global academic migration. By contrast, across the whole population of researchers or migrant researchers, female researchers were most significantly underrepresented in the field of Engineering and Technology. Despite the difference in gender ratios by discipline, all fields showed a tendency towards a closing gender gap among both general researchers and migrant researchers over time, generally in line with the overall trends across all disciplines combined (black lines).

Field-standardized gender ratios

Given the differences in gender ratios across fields of study, and to adjust for any compositional differences in the size of fields over time, we standardized the gender ratios by using a fixed (standardized) distribution of researcher populations by field as the weight. More specifically, we applied the field distribution of global researchers in the first period to the field-specific gender ratios of researchers at each period to obtain the (field-) standardized gender ratios. The results before and after standardization based on the fields of speciality shown Fig. 3.9 displayed the consistent trends in gender ratios. It indicated that the changes in the composition of disciplines over time did not drive the trend, i.e., the observed changes in the gender gap in scholarly mobility over time that we presented above.

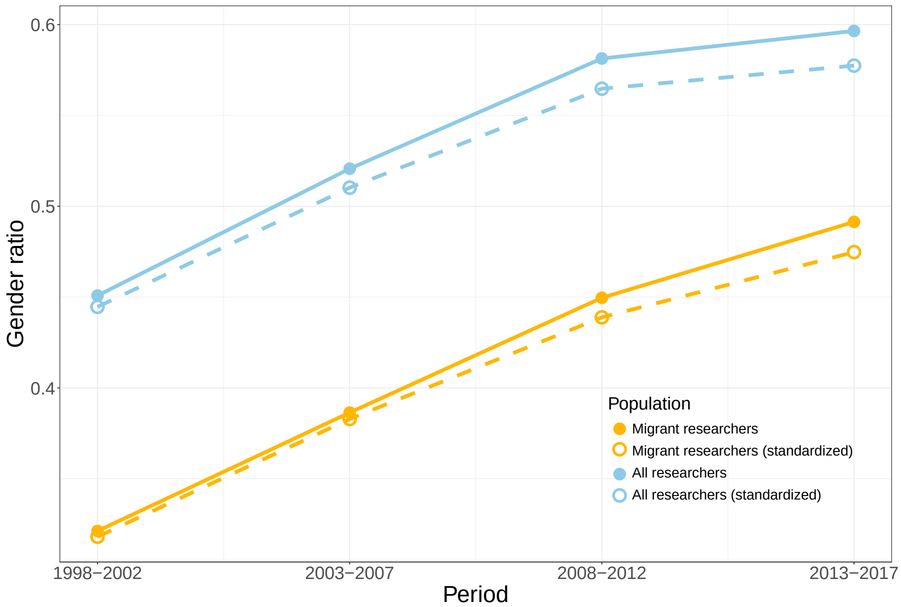


Figure 3.9: The comparison between the gender ratios among all published researchers and migrant researchers before and after field-standardization.

3.4.4 Gender disparities by income level

To examine if the closing gender gaps are consistent in both the Global North and Global South, we classified the countries into four groups (i.e., low-income economies, lower-middle-income economies, upper-middle-income economies, and high-income economies) based on the World Bank classification of income groups (Bank, 2021). In our dataset, there were 28, 49, 51, and 68 countries in the groups of low-income economies, lower-middle-income economies, upper-middle-income economies, and high-income economies, respectively. We considered those in the group of low- and lower-middle-income economies as being in the Global South.

Gender ratios among migrant researchers versus all researchers

Overall, the Global North countries, which include high-income countries and upper-middle-income countries achieved higher level of gender parity in global scholarly migration, shown in Fig. 3.10. By contrast, low-income countries saw the largest gender disparities among mobile researchers while lower-middle income countries continued to show the most under-representation of female researchers in the group of all published researchers.

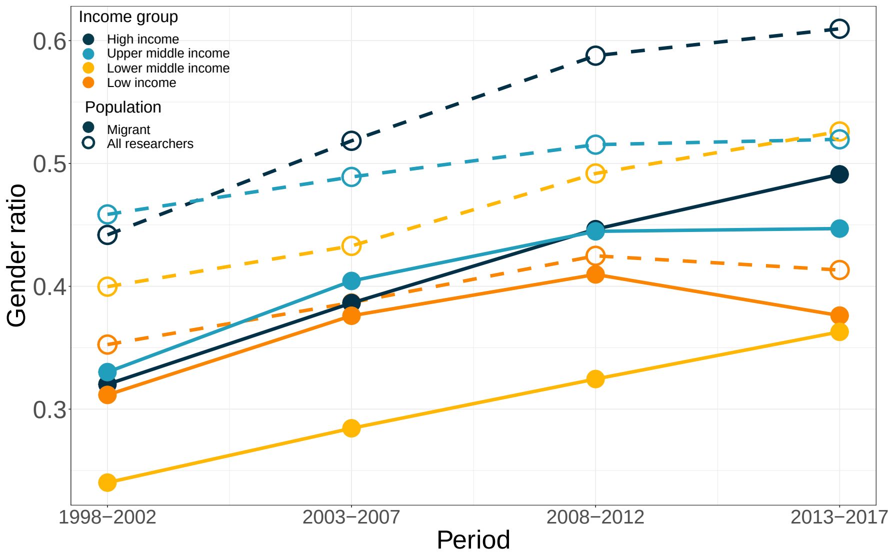


Figure 3.10: Gender ratios among all published researchers and migrant researchers across income groups (The classification of income groups are assigned by World Bank).

Generally, the trend of gender ratios over the period 1998-2017 indicated an overall increasing representation of female researchers in both populations of researchers, except for the group of low-income countries with a declining trend of gender ratios in the most recent period (2013-2017). Despite this, low-income countries had the smallest difference between the gender ratios among migrant researchers and the gender ratios among all researchers. This suggested that in low-income countries, female researchers may be a highly selected group, given the overall larger gender inequalities in academia in these settings. It also may reflect the feminization of migration in the Global South where gender norms and poor job opportunities may act as push factors for more skilled women to migrate (Bailey and Mulder, 2017; Donato and Gabaccia, 2015). For other groups of countries, especially the high income countries that have remained the dominant academic hubs accommodating a large part of researchers all over the world, an increase in the female representation of migrant researchers at a faster pace echoed the findings in 3.3.1 that globally, female researchers have been more internationally mobile than ever before.

Emigration and immigration spreads

The increasing diversification of origin and destination countries among both male and female mobile researchers on the global level can also be observed in different income groups, shown in Fig. 3.11, especially the emigration spreads which indicates migrant researchers of both genders have been more widely dispersed in the destination countries. However, both emigration and immigration spread among researchers across the Global South were at a higher level than those in the upper middle- and high-income countries. It echoed the effect of "rich-get-richer" where researchers from high-income countries tended to concentrate in other high-income countries, in contrast to those from low- and lower-middle-income countries, who tended to migrate to a broader range of destinations.

Despite the gender trend of higher diversification across destinations for researchers from the Global South, male mobile researchers from the Global South were still more diverse in a relatively broad range of destination countries, in line with our previous findings. Low-income countries exhibited larger gender gaps in the migration spreads in both directions, which tended to be wider in the most recent period.

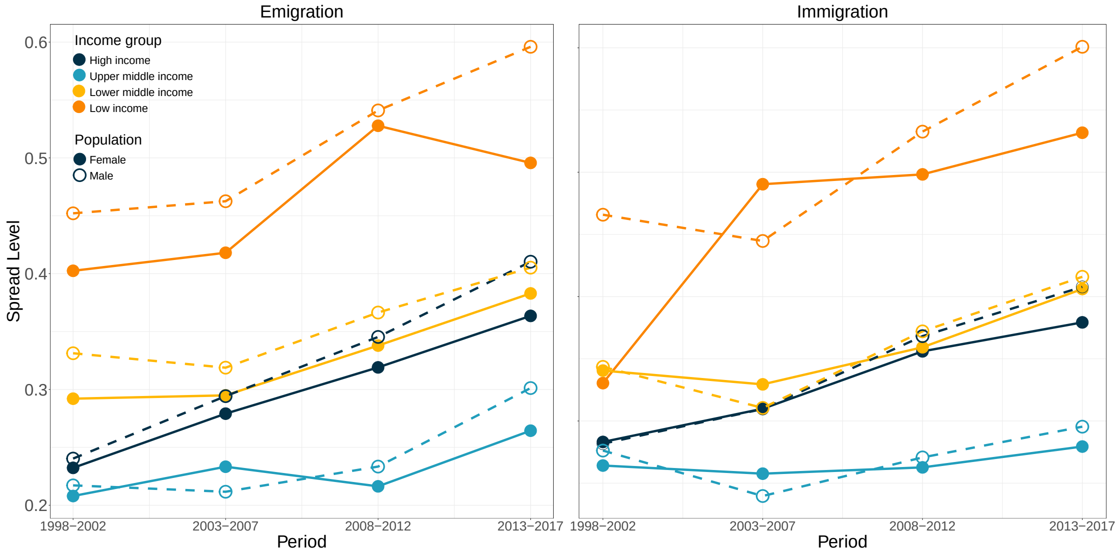


Figure 3.11: Scholarly emigration (left) and immigration (right) spreads, by gender and across income groups.

3.5 Discussion

Using bibliometric data on over 33 million publications, this chapter provided a global view of transnational scholarly migration by gender over the 1998–2017 period. Our analysis revealed a gender gap in favor of males among all published researchers, and an even larger gender gap among internationally mobile researchers. However, the rate of increase in female representation among global mobile researchers outpaced the rate of increase among all scholars, which suggests that female researchers have become more internationally mobile over the study period, in both absolute and relative terms. These trends indicate that broader patterns of increasing feminization of international migration have also occurred for global scholarly migration, a specific type of highly-skilled mobility. Despite these increases, significant cross-national heterogeneity persisted, with some countries, such as Serbia, Argentina, and Portugal, having near gender parity (female-to-male ratio of one) among migrant researchers, while some countries, such as Japan and South Korea, having significant gender gaps in favor of men (around 0.25). The global talent hubs like the US and Germany saw gender gaps around the global median levels (around 0.6).

In contrast to the skewed patterns of globalization in international migration, which indicate that migrants tended to move from an increasingly diverse range of origin countries to a shrinking pool of prime destination countries (Czaika and de Haas, 2014), our analysis showed that, for scholarly migration, there were simultaneous trends toward increasing migration distance and diversification of origin and destination countries among both male and female mobile researchers. The declines in transportation and communication costs, together with the strengthening of collaboration between national universities, likely contributed to greater levels of knowledge diffusion within a more balanced, globalizing science system. Despite this increasing diversification in the global scholarly migration system, women continued to migrate shorter distances, on average, and had lower emigration spreads, as they moved to a narrower range of destination countries than men. This gender gap in emigration spreads was more pronounced in countries that had large gender gaps in transnational scholarly mobility, such as South Korea. China, as an exception, showed a downward in the diversity level of both academic outflows and inflows. The Increasing number of mobile researchers concentrating in specific origin and destination countries hinted at a consolidated linkage between China and some countries.

The existing literature has pointed out that the global scientific system is largely shaped by highly-resourced nations (Lee and Haupt, 2021), with the US being the primary destination of choice for researchers (Franzoni et al., 2015; Guthrie et al., 2017a; Maier et al., 2007; Zhao et al., 2021, 2022a). Our results revealed a more nuanced picture, showing that the share of researchers moving to the US steadily declined over time, by around 7% for both male and female inflows over 20 years, despite its unchanged position as the top destination. Additionally, female researchers were consistently more likely than their male counterparts to choose the US as their academic destination, from both a global and a country-level perspective. Meanwhile, China has emerged as a prime destination with continuously increasing shares of female and male immigrant researchers, primarily from the US and from neighbouring Asian countries. In the meanwhile, a large number of PhD students and early-career researchers funded by the Chinese government to study abroad are required to return after finishing the study, which also makes China an emerging destination for researchers who had originally moved from China (identified as Chinese returnee researchers) (Fedasiuk, 2020). The increasingly diverse destinations for global migrant researchers provide additional empirical evidence for our finding that scholarly emigration has tended to be less skewed than before.

While this chapter has provided a novel and comprehensive picture of the gendered pattern of global scholarly migration, it also has limitations that we would like to acknowledge. We considered researchers who are published scholars whose information was retrieved from the Scopus database. This is a database which is dominated by English language journals mostly situated in western countries (e.g., EU countries, the USA and the UK). Despite this, approximately 22% of titles in Scopus are published in languages other than English (adding up to 40 local languages), and more than half of Scopus content originates from outside North America (Elsevier, 2020c). In other words, Scopus has a relatively large coverage of documents in non-English language. In addition, internationally-mobile scholars are more likely to publish at least part of their scholarship in English. Given the prominent role of English in science, we do not have reasons to believe that this bias would affect the key results of the article in a substantial way. However, it is important to remember that Scopus is not representative of all scholars, as non-English publications are under-represented. Future research could explore avenues for combining different and complementary data sources in order to make these types of data more broadly representative. Another limitation is potential inaccuracies in the gender detection results for Chinese researchers. Compared to the detection results for other countries, our design process, which is consistent with state-of-the art approaches, is less accurate in detecting genders from transliterated

Chinese names. For the purposes of this study, we performed a sensitivity analysis by imputing missing gender (see section *Imputation for missing gender*, Supplementary Information), which confirmed the robustness and reliability of the results that we present in the article. A promising direction for future research on gender detection includes applying our gender detection methods to names written in Mandarin, rather than to names transliterated in Roman alphabet (Elsevier, 2020a; Sebo, 2022). We also note that this chapter identifies scholars' origins in terms of their country of academic origin (see Method section for more details), which may not necessarily be their country of nationality or ethnic origin.

Our findings describe the global gender imbalances in scholarly migration, but do not assess the specific mechanisms or factors underlying the imbalances, or the factors driving changes in the gender gap. The promising trend toward the closing of the gender gap suggests that the targeted scholarship and fellowship programs funded by governments, multilateral organizations and private foundations, have helped women advance their academic careers through relocation, at least to some extent (Fiske, 2012; Huyer, 2015). However, other factors, such as relocation for family reasons, might have also pushed female researchers to migrate more frequently (or be considered non-movable and be excluded from the hiring pool (Rivera, 2017)). Thus, the relocation decisions of female scholars may not be attributable to their individual motivations for career advancement or research opportunities (Adsera et al., 2022). Therefore, the autonomy and the freedom to engage in international migration should also be considered when examining the decision-making of female researchers or the negotiations of academic couples (Schaer et al., 2017). This chapter represents an initial key step towards improving our understanding of global patterns over time, which is essential for promoting gender-equitable science policies and interventions. We encourage future research, both within specific country contexts and from a comparative perspective, to deepen our understanding of the factors that contribute to the aggregate-level patterns that we observed.

3.6 Data and Materials Availability

Scripts and data which allow for replication of our analysis are publicly accessible on GitHub under https://github.com/zxy919781142/A-gender-perspective-on-the-global-migration-of-scholars.

4

Gender differences in online visibility of early career researchers: Do female researchers benefit more from attention on social media?

This chapter is co-authored with Aliakbar Akbaritabar, Ridhi Kashyap, and Emilio Zagheni. It is currently under review for publication in an international journal.

Abstract

Social media in the digital era has become a powerful tool for disseminating research to a broad audience and has been widely used to increase the visibility of researchers, especially for underrepresented groups, including female researchers. Gaining visibility is particularly important for early-career female researchers to accumulate scientific impact, thus helping narrow the gender gaps in academic careers. Combining publication data from Scopus with information about research dissemination from both Altmetric and Twitter over a period spanning 2012–2023 (prior to Twitter's change to X), this chapter examined gender differences in online attention to the publications of early-career researchers and these researchers' self-promotion behaviors. We find that the first publications of female researchers were more likely to be mentioned on social media. However, male researchers were more likely to self-promote their first publications. These gender gaps in the probability of self-promotion were more pronounced in Social Sciences and Humanities, especially when first publications appeared in top-ranked journals. To further assess the impact of online visibility, we matched researchers who were mentioned online with those who were not and found that early-career female researchers gained more citations if their first publications were mentioned compared with their male counterparts. However, the gender differences reversed for self-promotion, whereby early-career male researchers gained more citations from self-promotion than females. Our results highlight how social media can help improve the visibility of female researchers and amplify their scientific impact. Yet, persisting gender gaps in self-promotion may limit this potential and reproduce gender inequalities.

4.1 Introduction

Social media has substantially affected how scholars promote their research, with platforms such as Twitter (named X since July 2023), Facebook, LinkedIn, and others being widely leveraged by researchers to share and access scholarly materials (Collins et al., 2016; Sugimoto et al., 2017). Higher online visibility generated through dissemination on social media increases the chances of research being noticed and used, while also helping grow researchers' reputations and enabling future career opportunities (Collins et al., 2016; Thelwall et al., 2013). The benefits of online visibility are likely to be particularly salient for early-career researchers. Getting published is the first step in developing an academic career, and sharing and distributing findings to generate visibility can help maximize scientific impact. For early-career researchers, however, traditional promotion strategies like attending conferences and workshops are usually costly, time-consuming, and less accessible (Vásárhelyi et al., 2021; Yammine et al., 2018). In an increasingly digital world, social media presence offers an efficient way for early-career researchers to distribute their research to a broader community. Moreover, increased interaction with the scientific community and the public could lead to a growing impact network for early-career researchers, which is essential for a long-term career in academia. Despite the increasing discussion on the role of social media in the visibility of researchers and their scientific work Collins et al., 2016; López-Goñi and Sánchez-Angulo, 2017; Peng et al., 2022; Vásárhelyi et al., 2021; Yammine et al., 2018, only a few studies have paid attention to the online visibility of early-career researchers and the consequences of this visibility for researchers' careers and impact (Haase and Müller, 2020).

Disparities exist in various aspects of science, including a lack of visibility for under-represented groups of researchers (Desai et al., 2021; Leahey, 2007; Liu et al., 2023; Petersen et al., 2019; Vásárhelyi et al., 2021). From a gender perspective, across different fields, women are underrepresented at traditional venues for gaining visibility and promoting research, e.g., academic conferences (Corona-Sobrino et al., 2020; Débarre et al., 2018), and are less likely to receive invitations to speak at colloquia and seminars (Nittrouer et al., 2018). Research also shows that women may benefit less from traditional forms of offline networking in terms of establishing new network contacts (Bapna and Funk, 2021), but once a virtual format is adopted for the conferences, the participation of women is substantially increased (Skiles et al., 2022). The scientific contributions of female researchers have shown to be less cited and undervalued compared to those of their male counterparts (Dion et al., 2018; Teich et al., 2022; Vásárhelyi et al., 2021), which leads to female scientists gaining less visibility in both academia and among the public. Such conscious or unconscious gender biases against women could in turn influence the career promotion and success of female researchers (Fortin et al., 2021).

The use of social media has the potential to serve an important role in the movement towards increased equity, diversity, and inclusion within academia as it provides a widely available and readily accessible platform to scholars, including underrepresented groups such as women (Yammine et al., 2018). Disseminating scientific findings through social media and breaking free from the constraints of space to aggregate mentions from vast audiences may help democratize attention to scholarly output, thus reducing gender bias in academia (Fortin et al., 2021). Despite this potential for equalizing engagement, research has indicated that social inequalities are often reproduced online. Female researchers have been found to receive lower online visibility, and such gender gaps vary by discipline on Twitter, which has been the most popular social media platform used in academia(Paul-Hus et al., 2015). Nevertheless, gender gaps in impact achieved online tend to be smaller than the traditional scientific impact measured by citations (Vásárhelyi et al., 2021). However, it is not clear whether gender differences in online visibility appear to also exist among early-career researchers and how these patterns have changed with the increase in social media adoption among early-career academics. We hypothesize that while there are gender differences in online attention disfavouring women among early-career researchers, they are likely to have become smaller across cohorts. One of the explanations behind the lower visibility of female researchers on social media is that women are significantly less likely than men to self-promote their papers (Peng et al., 2022; Vásárhelyi et al., 2021), partly due to the “feminine modesty effect" (Altenburger et al., 2017) and the fear of backlash (Henley, 2015). Looking specifically at early-career cohorts, we hypothesize that the gender gap in self-promotion persists but should become smaller across successive cohorts.

The correlation between online visibility and scientific impact in terms of citations has been well-studied (Bornmann and Haunschild, 2018; Costas et al., 2014; Eysenbach, 2011), and considerable evidence indicates that online metions on papers increased citations and such citation advantage appeared across disciplines and countries of authors (Eysenbach, 2011; Palamar and Strain, 2021; Shu and Haustein, 2017). However, limited research has sought to disentangle self-promotion from online visibility, which could lead to different impacts varying by the disseminators. Compared with being mentioned by others, publicizing one’s own scientific papers is a more efficient way to draw attention from targeted readers, especially from scholars in related fields (Howoldt et al., 2022; Klar et al., 2020). Given that, it is worthwhile to investigate whether additional advantages of self-promotion exist in the scientific impact of researchers in terms of the citation of their publications. The differences between how mentioning and self-promotion behaviours translate into citations and whether these benefits accrue unevenly by gender have received limited attention. A common limitation in studies that explore the relationship between online visibility and its impacts on citation is that they are often unable to account for the influence arising from potential confounders that drive selection both into social media visibility and citation impacts. Applying a matched-pairs design, as we do here, can help reduce the impact of confounders (e.g., journal ranking and discipline) and provide a better understanding of the association between online visibility as well as self-promotion on social media and cumulative citations (Kotsemir and Nefedova, 2022). Our hypothesis posits that, generally, online mentions can help early-career researchers gain more citations, but we expect these advantages of self-promotion to accrue more strongly for men who may use online platforms to self-promote more. Furthermore, we hypothesize that self-promotion has an independent and amplifying effect on citations.

Twitter (and also X) has been the predominant platform for researchers to share and distribute scientific findings over the past years and has played an important role in reshaping the way of scientific communication (Haase and Müller, 2020; Howoldt et al., 2022). About one in five journal articles published in 2012 was tweeted at least once (Haustein et al., 2015). From 2012 to 2016, the number of scientific tweeters significantly increased from 0.21 million to 2.06 million, and the number of scientific tweets even increased at a higher rate, from 0.98 million to 17.45 million (Yu et al., 2019). Scientific tweets, as one important way of online dissemination for research on social media, led to higher online visibility for both research and researchers (Collins et al., 2016; Thelwall et al., 2013). The literature also points to scholarly tweets having positive impacts on citations, with evidence showing that tweeted articles on average received 20% more citations than those that were not tweeted (Costas et al., 2014; Dehdarirad, 2020; Eysenbach, 2011; Shu et al., 2018; Yu et al., 2019). Despite changing to X, and broader changes in the management of the platform in mid-2023, Twitter has so far remained the most popular platform for research dissemination (Sugimoto et al., 2017; Thelwall et al., 2013) and the leading platform for studying social media attention (Kidambi, 2024). Particularly for the period under analysis in this chapter (prior to April 2023), Twitter remains of significant value as a widely used platform for learning how research generates online visibility on social media and how online visibility, in turn, translates to scientific impact.

To assess gender differences in online visibility and self-promotion, and their cumulative impacts, this chapter identifies 567,162 early-career researchers who started publishing during the period of 2012-2016, combining bibliometric information on their publications from Scopus together with online visibility measured by Twitter mentions through Altmetric and the Twitter API. The data collection process using the Twitter API relies on the period until April 2023 and predates recent changes in access to Twitter data. We examine whether social media platforms contribute to the creation of a more inclusive space for female academics to get attention online, or whether they perpetuate gender inequalities already from an early-career stage. We especially assess gender disparities in online visibility by disentangling self-promotion from general online dissemination and examine the impacts of these different forms of social media visibility on citation impacts. By doing so, we aim to provide a more comprehensive understanding of the linkage between online visibility and scientific impact to examine how processes of gender inequalities unfold at an early stage of academic careers.

4.2 Methods

4.2.1 Early-career researchers and their first publications

This chapter used large-scale bibliometric data from Scopus, to identify all authors who started publishing during the period 2012–2016. The starting year of publication was used to define the academic cohort of researchers. We considered the researchers who are in their first three years since the first publication, i.e., years one, two, and three, to identify the early-career researchers. That is, the researcher from cohort 2012 was considered an early-career researcher during the period 2012–2014. In most fields of science, it is acknowledged that the first author contributes the most to the work, by undertaking most of the research and writing most of the paper (Laudel, 2002; Nicholas et al., 2017; Pain, 2021). Based on this, we further selected those who have published at least one first-authored paper in their early-career stage, that is, the first three years of their career. Then we looked at the online visibility of a researcher's first *first-authored publication* which marks the start of a researcher's academic career in the form of scientific publications. Thereafter, we simplified the label first *first-authored publication* as the first publication to make it more concise.

We used the data from 2012 onward to coincide with 1) Altmetric's founding year and 2) higher usage of Twitter for academic purposes based on the reviewed literature (at least until we finished the data collection in April 2023 and before the changes in Twitter in July). This analysis on the online visibility of early-career researchers looked at online mention and self-promotion of their first publication on Twitter (now X), proceeding in the following steps: (1) we first looked at the gender differences in the probability of one's first publication being mentioned on Twitter and estimated the tweet counts received on the first publications by early-career female and male researchers if they were mentioned. (2) we then turned attention to the gender differences in the probability of self-promotion on Twitter and examine whether there is a gender difference in the process; (3) we examined the subsequent impact of this self-promotion on the citations of their first publication by comparing the impact of overall online mentions on citations.

4.2.2 Scientific mentions on Twitter

We utilized the Altmetric Details Page API, one of the most important altmetric tools, to retrieve the online mentions on their first publication on Twitter. Twitter mentions as tracked by Altmetric include all tweets (original), retweets, and quoted tweets that contain a direct link to a scholarly output (Altmertic, 2023). Here we only considered the original tweets to investigate the online visibility of researchers.

We assumed that the counts of Twitter mentions for each early-career researcher's first publication reflect two different processes: First, either other Twitter users saw the publication and decided to post it on Twitter, or the researchers promoted their papers on Twitter. Both of these processes help the publication become visible online. Second, the publication can accumulate more mentions over time. In our dataset, we found only around 28.5% of the researchers received tweet mentions on their first publications, and among the tweeted publications, around 47.8% received only one tweet mentions. This suggested that the distribution of the academic tweet mentions had the features of zero inflation and over-dispersion of one. Hence, to accommodate the excess zeros and over-dispersion of Twitter mention counts, we used a zero-inflated negative binominal (ZINB) regression to model the online visibility of early-career researchers' first publication and then examine whether it differs by gender. A ZINB regression consists of two parts: a logistic component to predict the probability of *certain zero* mentions, and a negative binomial component to model the mention counts if being mentioned (Eq. (4.1))).

$$P(M_i = n) = \begin{cases} \pi_i + (1-\pi_i) \times g(M_i), & \text{if } n = 0 \\ (1-\pi_i) \times g(M_i), & \text{if } n > 0 \end{cases} \tag{4.1}$$
$$g(M_i) = \Pr(Y = M_i|\mu_i, p)$$

$$logit(\pi_i) = \gamma_0 + \gamma_g \times gender_i + \sum_{k=1}^{n}(\gamma_k \times cov_{k,i} + \gamma_k^{'} \times gender_i \times cov_{k,i}) + \sigma_i \tag{4.2}$$

$$ln(\mu_i) = \beta_0 + \beta_g \times gender_i + \sum_{k=1}^{n}(\beta_k \times cov_{k,i} + \beta_k^{'} \times gender_i \times cov_{k,i}) + \tau_i \tag{4.3}$$

π_i is the logistic link function which indicates the probabilities of excess zero mentions online for researcher i (Eq. (4.2)). In addition to the variable of interest, i.e. gender ($gender_i$), we also considered these control variables: author's cohort (2012 is the reference level), the field of specialty (without a discipline assignment is the reference level), the number of authors in the first publication, the relative publication year since the start of the academic career (1, 2, or 3, and 1 is the reference level), the ranking quantile of the published journal in the subject area (journal rank of Q4 is the reference level), and whether the publication is a product of international collaboration or a collaboration with other universities/institutes (i.e., multiple institutions are involved). We also included interactions between gender and other control variables in the model. For the negative binomial component $g(M_i)$, we applied the same control variables as well as their interactions with gender in μ_i to model the probability of receiving n times Twitter mentions for the researcher i (Eq. (4.3)). σ_i and τ_i are the error terms in the two models, respectively.

Considering that the received Twitter mentions may vary by the researcher i's affiliation country j, we extended the models in Eq. (4.2) and Eq. (4.3) to a mixed-effects version by incorporating the random effects for country variability in the intercept and the slope of gender in Eq. (4.4) and Eq. (4.5), respectively.

$$\begin{aligned} logit(\pi_{i,j}) = \gamma_0 + \delta_{0,j} + (\gamma_g + \delta_{g,j}) \times gender_{i,j} \\ + \sum_{k=1}^{n}((\gamma_k + \delta_{k,j}) \times cov_{k,i,j} + (\gamma_k^{'} + \delta_{k,j}^{'}) \times gender_{i,j} \times cov_{k,i,j}) + \sigma_{i,j} \end{aligned} \tag{4.4}$$

$$\begin{aligned} ln(\mu_{i,j}) = \beta_0 + \alpha_{0,j} + (\beta_g + \alpha_{g,j}) \times gender_{i,j} \\ + \sum_{k=1}^{n}((\beta_k + \alpha_{k,j}) \times cov_{k,i,j} + (\beta_k^{'} + \alpha_{k,j}^{'}) \times gender_{i,j} \times cov_{k,i,j} + \tau_{i,j} \end{aligned} \tag{4.5}$$

Before considering all of the control variables and the country-level random effects in Eq. (4.4) and Eq. (4.5) to generate the full ZINB model, we took a step-wise method by gradually adding control variables and their interactions with gender in the Baseline ZINB model which only includes gender as the variable.

According to the estimated results, we predicted the counts of Twitter mentions on the first publications of early-career female and male researchers, and measured the marginal effects of gender (gender gap, i.e. difference between male and female researchers).

4.2.3 Self-promotion on Twitter

Due to licensing restrictions from Twitter, Altmetric only provides the Tweet ID and Twitter User ID (individual tweeter ID) in their Details Page API, lacking the individual information of the tweeter. That means, we can only see whether the publication has been tweeted or not through the publication's Digital Object Identifier (DOI), but with no clues to identify who tweeted it. We further employed Twitter's public API to track more details about the tweeters including their display names and handle names (i.e., username or @name, which is unique to Twitter), using the User ID we obtained from Altmetric. We assessed whether the early-career researchers had self-promoted their first publication by comparing their names retrieved from Scopus with the names of the tweeters who mentioned the publication on Twitter. Although users were not required to use their real names in their Twitter display names or handle names, we believe that the majority of researchers use their real names on Twitter given the promotion purpose, with a stronger focus on facilitating social networking (Howoldt et al., 2022). Author-name matching is also the dominant method to identify researchers on Twitter (Costas et al., 2017; Howoldt et al., 2022).

The name-matching process we elaborated on to find the self-promoted researchers is shown in Fig. 4.1. For each early-career researcher in our dataset, we first combined their first names and last names that were retrieved from Scopus, with or without space, to form the four combinations as the possible author names. We collected all User IDs of those who mentioned the first publication of each early-career researcher, and we further acquired both display names and handle names of the tweeters by entering Twitter IDs in Twitter API. In addition, we also considered an existing open data set of scholars on Twitter which includes 423,920 unique tweeter IDs forming 498,672 unique paired OpenAlex author IDs and tweeter IDs (Mongeon et al., 2023). We can further obtain the author names for each tweeter ID included in this dataset, as one of the references for name-matching.

We labeled the dataset as *OpenAlex Database* in Fig. 4.1 because the dataset was created using the May 2022 OpenAlex data dump by the researchers who provided it (Mongeon et al., 2023). We then compared each category of Twitter names with the four combinations of author names to calculate the similarity scores and found the best-matched name with the highest similarity score for each category. We decided on the final best-matched Twitter name from the three matched names and determined if the author self-promoted their first publication only if the match score for the best-matched name exceeds 0.6. We emphasize here that for researchers who did not use their published name appearing in Scopus for their Twitter profiles, our name-matching procedure failed to find them.

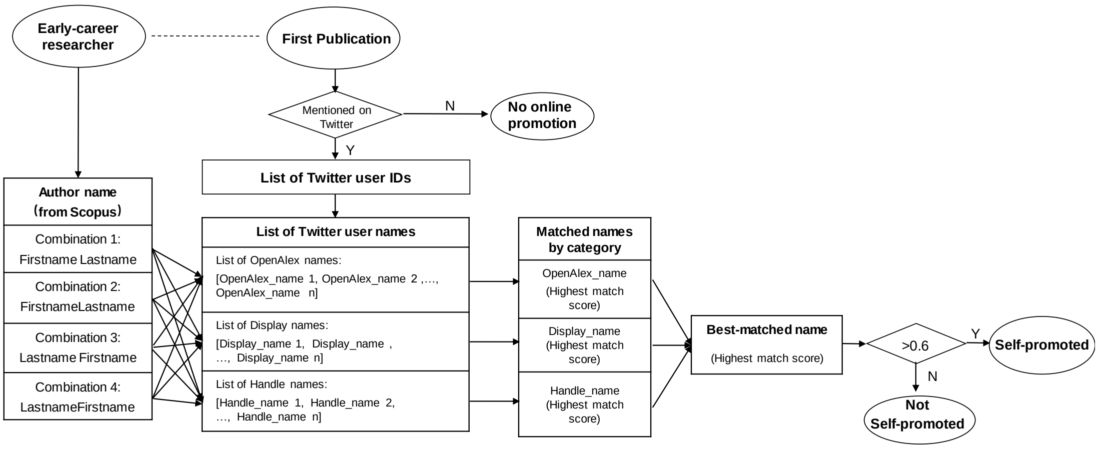


Figure 4.1: The workflow of matching Scopus-published researchers and profiles on Twitter.

After identifying the self-promoted publications, we employed logistic regression to model the probability of self-promotion for the researcher and examine the potential gender differences in this process. We employed logistic regression to model the probability of self-promotion for the researcher i and examine the potential gender differences in this process. In addition to the control variables we mentioned before, we also considered whether their publication has received mentions on Twitter (not being tweeted is the reference level) and the counts of tweets by others.

The mixed-effects version for researcher i in country j is shown as follows:

$$logit(P(self_{i,j} = 1)) = \theta_0 + \eta_{0,j} + (\theta_g + \eta_{g,j}) \times gender_{i,j} + \\ + \sum_{k=1}^{n} (\theta_k + \eta_{k,j}) \times cov_{k,i,j} + (\theta_k^{'} + \eta_{k,j}^{'}) \times gender_{i,j} \times cov_{k,i,j} + \xi_{i,j} \tag{4.6}$$

we also employed the step-wise method by gradually adding control variables in the base logistic model which only includes gender. According to the estimated results, we predicted the probability of self-promotion of early-career female and male researchers and measureed the marginal effects of gender (gender gap, i.e., difference between male and female researchers) in this process.

4.2.4 Matching Researchers with similar characteristics

To explore the impacts arising from general Twitter mentions and early-career researchers' self-promotion on research impact, we first used propensity score matching (*PSM*) to match (1) the researchers who received Twitter mentions on their first publications with those who did not; (2) the researchers who self-promoted their first publication with those who only received tweets by others. The *PSM* begins with a logistic regression to estimate the propensity score of the occurrence of the treatment on observed variables for the researcher i, shown as below:

$$logit(P(Outcome_i = 1)) = \lambda_0 + \sum_{k=1}^{n} (\lambda_k \times cov_{k,i}) + \psi_i \tag{4.7}$$

For the first matching, we defined the treatment group as those who received Twitter mentions on their first publications and the control group as those who did not. We considered gender, academic cohort, discipline, the journal rank of their first publication, affiliated country, the number of authors, and whether the publication is a product of an international collaboration to estimate the propensity score of receiving Twitter mentions. Based on the similar propensity scores, the distribution of observed baseline variables will be similar between treated and control subjects.

That is, each treated subject (i.e., the researchers with Twitter mentions) in the treatment group will be paired with a control subject (i.e., the researchers without Twitter mentions) in the control group. We further confined each pair of researchers who have the same gender, are from the same cohort, do research in the same field of specialty, and published their first article in the journals ranking at the same quantile, to generate the first matching of paired researchers (*Matching 1*).

Similarly, we employed *PSM* to match those who self-promoted their first publication with those who did not (that means, their publication was only being mentioned by others) from the pool of researchers whose first publication has been mentioned on Twitter. In addition to the factors we considered above in the logistic model of *PSM*, we added the tweet counts as the control variable to estimate the propensity score of self-promotion, ensuring a similar level of online exposure on Twitter. We finally generated the second matching of self-promoted researchers paired with those who did not (*Matching 2*).

4.2.5 Impact of online visibility on citations

To analyze whether online mentions, as well as self-promotion, played a different role in the scientific impact of the early-career researchers' first publications, we used the cumulative *discipline-normalized annual citation scores* (DNCS) within five years after publication as the proxy of the impact which allows comparison over cohort and discipline. In other words, the citation counts are normalized by discipline and publishing year to make it more comparable with people in the same research field.

We divided the actual citation of each publication by the average citation of all early-career researchers' first publications in the same discipline and publishing year as the citation scores and accumulated the citation scores for the first 5 years since publication. The measurement of five-year normalized citation impacts for each first publication in the discipline f and the publishing year t is represented as below in Eq. (4.8), where e_{ft} is the average (expected) citation counts for the early-career researchers' first publications:

$$DNCS = \sum_{t}^{t+4} \frac{c}{e_{ft}} \tag{4.8}$$

We employed linear regression to estimate the relationship between online visibility and the five-year cumulative DNCS for the matched researchers by gender. For *Matching 1*, in addition to the treatment —whether a researcher's first publication is mentioned on Twitter— we also included in the model each early-career researcher's gender, cohort, discipline, the journal rank of the first publication, the number of authors, and whether the publication is a product of international collaboration as control variables and their interactions with the treatment in the model. We further measured the marginal effects of Twitter mentions on the citation, that is, the citation difference between the publications with online mentions and those without online mentions for both female and male researchers.

To estimate the marginal effect of self-promotion on citations by gender, we used a similar model with the treatment of self-promotion and the control variables mentioned above plus the number of Twitter mentions for the researchers' population in *Matching 2*. The interactions between treatment and control variables were also included in the model.

4.2.6 Ethical Considerations

The Twitter data collected and used in this chapter consists of all accessible IDs of the tweets that mentioned the DOIs of the publications we studied and the corresponding IDs of the Twitter users who posted the Tweets. The information was accessible from Altmetric and was then used to extract the display names and handle names of these Twitter users. The Altmetric Details Page API was subscribed to Research Services by the University of Oxford, and the usage of Twitter's public API was granted through an application to Twitter's Academic

API for academic purposes, which was approved in 2022. All data collection was conducted complying with the Terms of Use of publicly available data on both platforms. All results were shown in the form of aggregated numbers without any personal information such as author names, names shown on Twitter and publication DOIs that can be used to identify individuals.

I acknowledge that Twitter users might be unaware that their data can be accessed and analyzed in this way, and it may not be possible to reach all Twitter users to obtain their informed consent (Ahmed et al., 2017). Tweeters made it clear to its Tweeters what will happen to their data (including that it may be used for research) at the point of signing up to use its service (Nicolas, 2020), so it is a reasonable expectation of Twitter users that their public data exposed on the platform may be viewed and used for research. Moreover, this research is about scientific dissemination and outreach of individuals who were trying to reach a wider audience and publicize their work. Thus, although personal information was used for the linkage during research but was not shown in results, it is in relation to public tweets about scientific content meant for public dissemination.

In conducting this research, I have been committed to practicing academic integrity, transparency and adhering to responsible research practices.

4.3 Results

4.3.1 Gender differences in online visibility of early-career researchers

We identified 567,162 early-career researchers across five academic cohorts (2012-2016). We defined the cohort of a researcher as the exact year in which the first publication appeared in Scopus, while their academic age was calculated as the number of years since their first publication. We exclusively included the researchers who had at least one publication as the first author at the early-career stage, that is, at their academic ages of 1, 2, and 3 years. Some early-career researchers may have multiple first-authored publications and the first one of these was taken for the subsequent analysis of its social media attention and impacts, which was simplified as *first publications* in the following text.

Among these early-career researchers, 161,884 (28.54%) were detected to receive Twitter mentions on their first publication, and 8,677 (1.53%) researchers self-promoted their first publication. By using a systematic process composed of name-gender detection methods used and validated in previous research (Zhao et al., 2023) and a category of six macro research fields of specialty, we identified the gender and research fields of these researchers. Specifically, 71,660 female researchers (32.75% of all published female authors in our dataset) received Twitter mentions on their first publication, while 90,224 male researchers (25.90% of all published male authors) received mentions. The aggregated results indicated that early-career female researchers were more likely to get mentions online. More details on the descriptive statistics on the sample and self-promotion rates were shown in Table. C.1 in Appendix C.

Online visibility is not only determined by whether the publication has been mentioned online but also by how many mentions it has received, that is, the tweet count. For the researchers whose first publication has been mentioned on Twitter, almost half of them (47.81%) only received one mention. Considering this skewed distribution of mentions, we used zero-inflated negative binomial (ZINB) modeling that simultaneously accounts for the excess "zero" mentions and over-dispersed "one" mentions to predict how many mentions were received by female and male researchers (see Section 4.2.2, Methods). We estimated it from the Baseline ZINB Model (Model 0) and successively added variables to control for compositional confounders that might differentially affect the publications of male and female researchers to understand how these impacted the gender difference in predicted mention counts. These include the author's cohort, the field of specialty, the number of authors of the first publication, academic age (1, 2, or 3 years), the ranking quantile of the published journal in the subject area, whether the publication is a product of international collaboration or collaboration with other universities or institutes in the same country. We also added the interactions of these variables with gender to explore how gender differences in Twitter mentions change across other factors (see the estimate results in Section C.2, Appendix C).

The Baseline ZINB Model indicated that the number of Twitter mentions for a female author's first publication was on average 20 percent higher (Incident Risk Ratio (IRR) = 1.20) than those for a male author's first publication in the left panel of Fig. 4.2. Despite a higher predicted mention count for female researchers, this gap narrowed with the inclusion of controls, suggesting that factors such as journal prestige and institutional collaboration can help account for the higher mention counts that female researchers on average received (see the estimate results of Model 0-6 in Table. C.3, Appendix C). The Full ZINB Model (Model 7) further controled the country-level random effects using a multilevel setup, as the distribution of scientific tweets and tweeters was more concentrated in a few countries, such as the USA, and some European countries. The marginal effects from the Full ZINB Model in Fig. 4.2 showed that the gender difference in the mention count largely disappeared after accounting for factors such as journal prestige and discipline, as well as the country-level random effects. The country-level heterogeneity in the gender difference of mention counts was shown in Fig. C.2 in Appendix C.

As the right panel of Fig. 4.2 showed, gender differences in the numbers of predicted mention based on the Full ZINB Model varied across different factors, although gender differences in most cases are small. For researchers from earlier cohorts (2012, 2013, and 2014), male researchers had more predicted mentions, while the pattern switched by the 2015 cohort, with female researchers receiving more mentions and the gender gap narrowing. While previous research has indicated a higher presence of researchers from Social Sciences and Humanities in Altmetrics and Twitter than from the Natural Sciences (Chen et al., 2014; Costas et al., 2020),

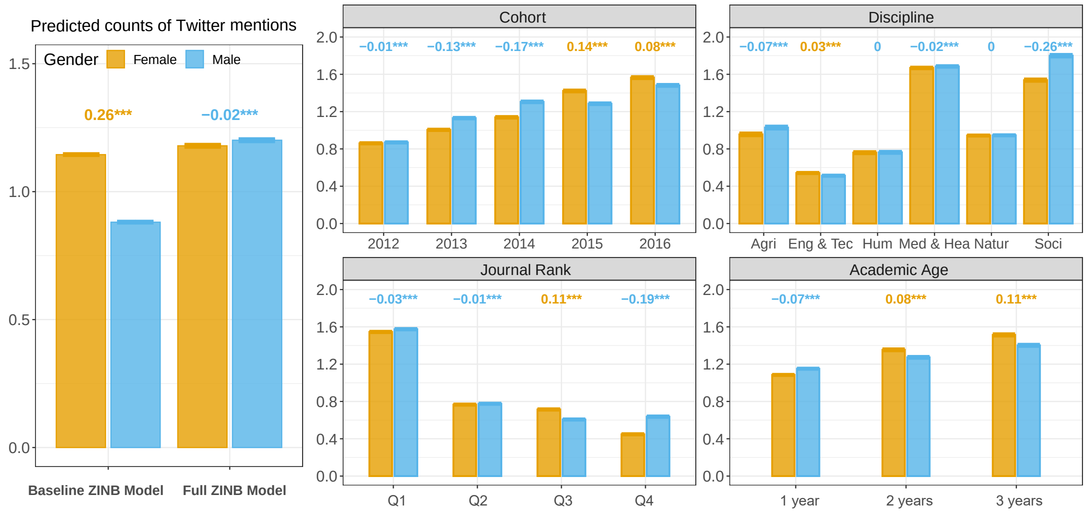


Figure 4.2: Predicted number of Twitter mentions on early-career female and male researchers' first publication, across the full sample (left panel) and disaggregated by cohort, discipline, journal rank and academic age (right panel). The left panel compares the results of the Baseline ZINB Model and the Full ZINB Model, and the disaggregated result in the right panel is based on the Full ZINB Model. The numbers at the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.1 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women and positive values (printed in orange text) shows the opposite (Agri: Agricultural Sciences; Eng & Tec: Engineering and Technology; Hum: Humanities; Med & Hea: Medical and Health Sciences; Natur: Natural Sciences; Soci: Social Sciences).

we further showed that Social Sciences also had the largest female-to-male gender gaps (-0.26) in the counts of online mentions. Early-career male researchers in this area were predicted to receive, on average, the most Twitter mentions (1.81), statistically significantly more than their female counterparts (1.53). However, in Engineering and Technology, which was always dominated by male researchers in terms of population size (Zhao et al., 2023), the first publications by female early-career researchers were predicted to receive slightly more attention on Twitter with a marginal effect of 0.03 in mention counts.

Early-career researchers with the first publications in Q1 (top 25%) journals, based on the journal ranking from SCImago Journal Rank (SCImago), received more Twitter mentions than those who published in journals with lower rankings. Yet the gender gap within these top-ranked journal was negligible, i.e. women and men benefited equally from visibility when they published in these journals.

4.3.2 Gender differences in self-promotions of early-career researchers

When looking at the probability of self-promotion, we also employed the step-wise method by gradually adding control variables in the Baseline Logistic Model (Model 0) which only includes gender. The step-wise modeling of self-promotion was shown in Section C.4, Appendix C. The result in Fig. 4.3 depicted a different pattern relative to the general online mentions shown above, as male researchers always showed a higher predicted probability of self-promoting their first publication. Across the entire sample (left panel), Full Model (Model 8 in Table. C.4, Appendix C), which is the multilevel logistic regression model which further includes whether the publication has been promoted by others and the number of others' promotions as control variables based on the specification used in the Twitter mention analyses above, displayed a larger difference in the probability of self-promotion (-0.43) than the Baseline Model that only includes the factor of gender (-0.04). One of the possible reasons is that in certain countries with larger numbers of early-career researchers, like China and South Korea, the early-career researchers were less inclined to actively promote their first publication, despite smaller gender differences.

Within different categories (right panel), the results based on the Full Logistic Model indicated that, with the cohorts of the researchers, the probability of self-promotion among both female and male early-career researchers had increased, with the smallest gender difference observed in the most recent cohort (2016). Disaggregating by six fields of specialty, this gender difference in self-promotion was larger for those working in Social Sciences (-1.45 percentage points (p.p)) and Humanities (-0.60 p.p.) compared to those in Engineering and Technology (-0.06 p.p.). Also, the probabilities of self-promotion were overall higher in the Social Sciences across both genders compared with Science, Technology, Engineering,

Figure 4.3: Predicted probabilities of early-career female and male researchers self-promoting their first publication across the full sample (left panel) and disaggregated by cohort, discipline, journal rank and academic age (right panel). The left panel compares the results of the Baseline Logistic Model and the Full Logistc Model, and the disaggregated result in the right panel is based on the Full Logistc Model. The numbers on the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.1 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women and positive values (printed in orange text) shows the opposite (Agri: Agricultural Sciences; Eng & Tec: Engineering and Technology; Hum: Humanities; Med & Hea: Medical and Health Sciences; Natur: Natural Sciences; Soci: Social Sciences).

and Mathematics (STEM) fields. The highest rate of self-promotion in the Social Sciences echoed the largest number of online mentions in this field, while in the Medical and Health Sciences where both female and male researchers on average received over 1.5 mentions online, the probability of self-promotion in this field was below 2 p.p., smaller than the researchers in Social Sciences and Humanities. We can also see that male researchers were more likely to promote their first paper when the paper was published in a Q1 (top 25%) journal, as that group has the largest gender gap (-0.59 p.p.). As the probability of self-promotion increased from the first year of their career (1-year of academic age) to the third year (3-years of academic age), the gender difference in self-promotion also gradually increased from -0.41 to -0.5 p.p. Such a widening gap in the willingness to self-promote over time suggested that, in the long run, if the current behavior continues into later career stages, male researchers would be far ahead in self-promoting themselves compared to females. The country-level heterogeneity in the probabilities of self-promoting first publication, which was shown in Fig. C.3 in Appendix C, indicated that in all the countries, early-career male researchers were always more likely to self-promote their first publications.

4.3.3 Gender difference in the scientific impacts of online visibility among early-career researchers

To probe into the longer-term effects of online mentions and self-promotion behavior on early-career researchers' scientific influence in terms of the citation of their first publications, we used propensity score matching to create two matched groups of scholars by considering multiple variables (see the description in Section 4.2.4).

We first modeled the propensity of receiving the treatment, i.e., online exposure on Twitter, from the pool of all early-career researchers across the academic cohorts of 2012–2016 to match researchers with Twitter mentions on their first publication to those without mentions. In the matching process, we considered gender, academic cohort, field of research, journal rank of their first publication, affiliated country, number of authors, and whether the publication was a product of an international collaboration. The result generated the *Matching 1* of 135,562 pairs (271,124 individuals). Fig. C.4 in Appendix C shows the variable distributions are similar between the treatment and control groups in *Matching 1*.

Then, we modeled the propensity of self-promoting the first publications on Twitter from the pool of researchers whose first publications have been mentioned on Twitter. In addition to controlling the factors used in the matching process above, we added the tweet counts as an additional control variable. We produced *Matching 2* of 6,189 pairs (12,378 individuals), after confirming the variable distributions were similar between the treatment and control groups (Fig. C.5, Appendix C). The treatment group here is those who self-promoted their first publications in addition to receiving mentions from others. The control group is those who only received mentions from others. The comparison between the results in *Matching 1* and *Matching 2* helped us investigate gender differences in the impact of online mentions and whether there is an additional advantage in citation performance from self-promotion.

We used the *discipline-normalized annual citation scores* (DNCS) in the first five years since publication to measure the cumulative scientific impact of early-career researchers' first publications. Based on the matched pairs in *Matching 1*, Fig. 4.4 showed the marginal effects of Twitter mentions on the five-year cumulative DNCS for both early-career female and male researchers. The marginal effects can be interpreted as the difference in the predicted DNCS received by the treated researchers versus the control group of researchers. In terms of general online mentions (left panel), both females and males benefited from Twitter exposure for the five-year DNCS of their first publications compared to those publications that receive no mentions, shown from the solid bar contrast to the stripped bar. Female researchers, on average, had a higher increase in DNCS (0.28) as a result of being mentioned online compared to male researchers (0.23). The higher increase in normalized citation scores helped shrink the gender gap in citations but did not change the male-dominated pattern. In other words, the gender difference in citations could be narrowed thanks to online visibility, even though it still existed. To get a better understanding of how strong the impact of online visibility is, we estimated the average increase in DNCS to be 0.27 for early-career researchers of both genders when their first publication was in a Q1 journal compared to being published in a Q2 journal. This effect size is nearly in the same range as the one of being mentioned online. It is a further indication of the importance of online visibility in generating scientific impacts for early-career researchers, especially for females.

The disaggregated marginal effect at different categories (right panel) suggested that there was a slight upward trend in the effects of online mentions on DNCS

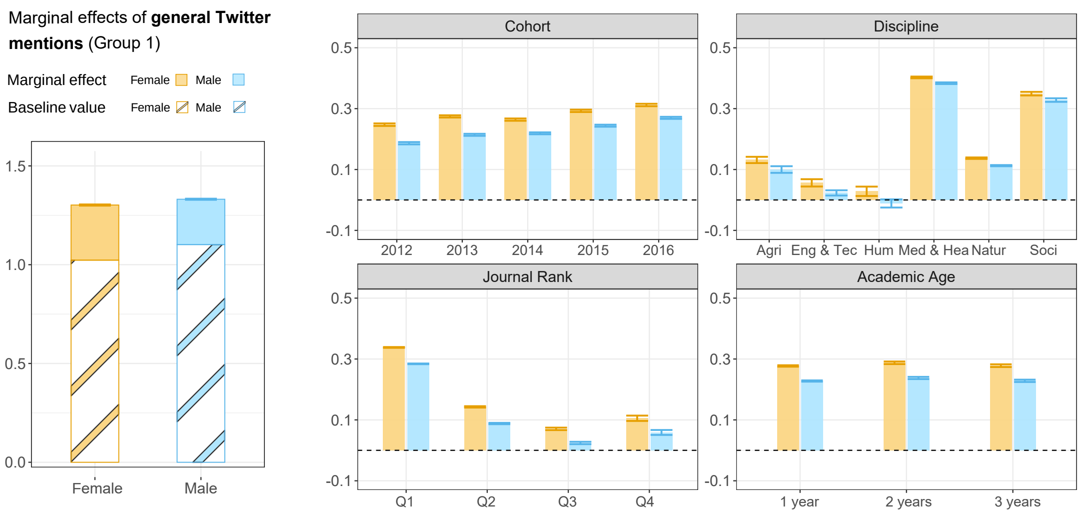


Figure 4.4: Marginal effects of Twitter mentions on the 5-year cumulative *discipline-normalized citation scores* (DNCS) among early-career female and male researchers (*Matching 1*), at the overall level compared to the average DNCS of researchers without Twitter mentions as baseline (left panel) and marginal effects disaggregated by cohort, discipline, journal rank, and academic age (right panel).

over time. This means that researchers from more recent cohorts tended to get more normalized citations on their first publications when those publications were mentioned on Twitter. Meanwhile, the gender gaps in the citation gains tended to be gradually reduced over the cohort, with the smallest gaps for the researchers from cohort 2016. In addition, the interaction between Twitter mentions and the fields of research also determined the extent to which more citations would be accumulated. Researchers from Medical and Health Sciences tended to gain the most additional citations, increasing by around 0.4 units of five-year cumulative DNCS. The field of Social Sciences, which was most likely to have mentions on Twitter (Table. C.1 in Appendix C), saw an increase of 0.35 units of five-year cumulative DNCS in the first publications of both early-career female and male researchers. Other research fields, on the contrary, indicated relatively smaller (e.g. Engineering and Technology, and Agricultural Sciences) or no gains (e.g. Humanities) in the five-year cumulative DNCS for early-career researchers. Besides, the increase in 5-year cumulative DNCS from Twitter mentions for the researchers publishing in Q1 journals was more than twice compared to those publishing in other journals, and female researchers received around 0.05 units of DNCS more than their male counterparts. This suggested that higher quality publications, as proxied by journal ranking, and after controlling for multiple factors such as discipline, gained more from Twitter exposure.

We also measured the marginal effects of Twitter mentions on citations by limiting the treatment group to receiving only one mention, as a robustness analysis. This led to 61,671 pairs (123,342 individuals) of researchers. The results further confirmed that female researchers generally gained higher citation counts when mentioned on Twitter. More details about the results were shown in Fig. C.6, in Appendix C.

For the publications that have already been visible on Twitter, Fig. 4.5 showed that self-promotion was associated with higher five-year cumulative DNCS compared to those only mentioned on Twitter, irrespective of analyzing overall average effects (left panel) or disaggregations across categories (e.g., cohort, field, journal rank, and academic age; right panel). Only the researchers whose first articles were published in Q4 journals are an exception, with an insignificant or even negative effect of self-promotion. Unlike general online mentions that result in greater citation increases for female researchers, early-career male researchers tended to gain more additional citations after self-promotion, which again widened gender gaps.

As online mentions played an increasingly important role in accumulating citations (shown in Fig. 4.4), the relationship between DNCS and self-promotion, on the flip side, had been weakening over the cohort of researchers. The increase in five-year cumulative DNCS related to self-promotion was consistently larger for early-career male researchers compared to female researchers. Across the research areas, early-career researchers from the field of Agricultural Sciences gained the highest marginal increase 5-year cumulative DNCS through self-triggered tweets, with male researchers gaining nearly 0.2 units of additional DNCS relative to their female counterparts. Despite the remarkable increase in self-promotion, the gains from general Twitter mentions relative to those without any mentions in this field were comparatively smaller. A similar case can be seen in the area of Humanities and Social sciences, where self-promotion generates significant increases in 5-year cumulative DNCS while general mentions including those mentions by others did not have the same impact on long-run DNCS. This suggested that beyond the positive effect of Twitter visibility on citations, who is the promoter also played an important role in determining the additional increase in citations.

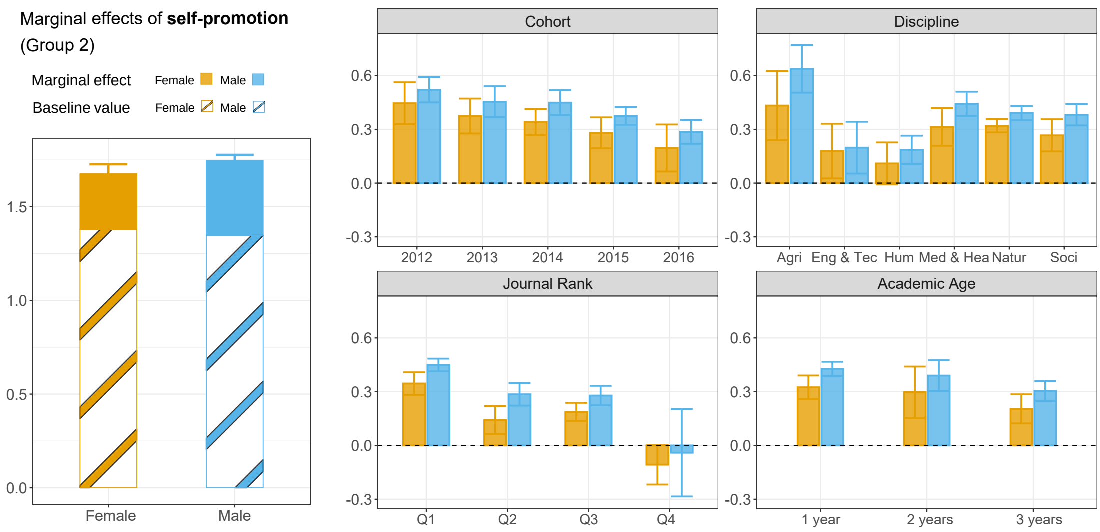


Figure 4.5: Marginal effects of self-promotion on Twitter on the 5-year cumulative *discipline-normalized citation scores* (DNCS) among early-career female and male researchers (*Matching 2*), at the overall level compared to the average DNCS of researchers without self-promoted mentions as baseline (left panel) and marginal effects disaggregated by cohort, discipline, journal rank, and academic age (right panel).

4.4 Discussion

By establishing a connection between the first academic papers published by researchers who started publishing between 2012 and 2016, and the related research dissemination measured from Altmetric and Twitter until April 2023, and before recent changes in Twitter (now X), we examined gender differences in the online visibility of early-career researchers and how the interaction between gender and online visibility impacted on citations. We found that the first publications by female researchers were, on average, more likely to be mentioned online on Twitter. Once we controlled for publication- and researcher-level characteristics (e.g., journal ranking), these gender differences largely disappeared, suggesting that first publications by female researchers were positively selected on attributes that increased promotion by others. However, early-career female researchers were less likely to self-promote their first publications on Twitter, even after controlling for these factors. These gender gaps in self-promotion were consistent with the fact that women were underrepresented among Twitter users, which further reflects broader patterns of gender inequality in the digital sphere (Kashyap et al., 2020). On the other hand, when women try to promote themselves online, they may face similar barriers to those offline, such as gender homophily (McPherson and Smith-Lovin, 1986), the fear of backlash (Henley, 2015; Moss-Racusin et al., 2012) and the “feminine modesty effect" (Altenburger et al., 2017. While our research cannot uncover the exact reasons why women are less likely to promote themselves, we highlight how gender differences in self-promotion are evident at the early stages of the scientific career. Existing evidence suggests most tweeting researchers are from the Social Sciences and Humanities (Costas et al., 2020; Howoldt et al., 2022). We further found that the early-career researchers in these fields are more likely to self-promote their publications, yet also exhibit the largest gender gaps skewing more towards men.

Through matching researchers on an extensive list of characteristics to better control for confounders, we further found that there was a gender difference in the way online visibility translates into scientific impact. Female researchers gained more citations if their first publications were mentioned on Twitter, while males gained more citations from self-promotion than females. Compared to the “Pull Model" in which researchers depend on disseminators to “pull their research in", self-promotion by early-career researchers was found to bring an additional increase in citations compared to those only receiving online mentions by others. Self-promotion helps researchers gain agency by actively promoting their results and by connecting with the community of their research area towards a “Push Model" where researchers as disseminators “push their research out" to public (Howoldt et al., 2022; Klar et al., 2020). Our results indicate that gender differences were produced again in the “Push Model", which benefited males more, with nearly one-third more normalized citations on average than females. Previous work indicates that male researchers are more inclined to communicate intensely through social media channels and gain higher popularity, with more followers and more public lists including them (Klar et al., 2020). Online social networks often mirror real-world networks, and these advantageous networks could help explain the larger citation impacts from self-promotion experienced by men. These patterns indicate that while online visibility may facilitate greater amplification of research by underrepresented groups such as women, persisting gender gaps in online self-promotion behaviors limit the potential for even greater impacts, and reproduce gender disparities. Overall, these findings suggest a process of cumulative disadvantage starting early in scientific careers: men were more likely to self-promote their first publication, with subsequent citation impacts also being larger for men.

This chapter contributes to a comprehensive understanding of gender disparities in online visibility by distinguishing the push and pull mechanisms at play within online academic dissemination. Prior research has examined the broader questions of gender inequalities on social media (Kashyap et al., 2020) and also highlighted the lower online visibility of female researchers (Débarre et al., 2018; Paul-Hus et al., 2015; Vásárhelyi et al., 2021). However, existing research often does not distinguish between the dynamics of other- versus self-promotion in relation to online visibility, and cannot control for confounders that may affect the impacts of online visibility. Our dataset and research design, through a focus on early-career researchers, can help minimize, to some extent, the impact of accumulated resources and networks, and examine the dynamics of online visibility at the start of academic careers. Our findings show the potential positive role of social media for amplifying the research of female scholars at the start of their academic careers – as indicated by the attention received through mentions by others, and their linked citation impacts. However, we also show persisting gender inequalities in self-promotion, which limit the full potential of these impacts and, in turn, threaten to reproduce existing gender inequalities in academia. To further increase the online visibility of female researchers and help them receive more credit for their work, our findings indicate the potential value of integrating training in self-promotion on social media within graduate school curricula and encouraging early-career female researchers to self-promote their work.

This chapter draws on the linkage of bibliometric and social media data, which, while providing a unique data set to observe the impacts of online visibility, also creates some limitations. We acknowledge that Scopus bibliometric data, while including publications from other languages, still is dominated by English-language publications (Baas et al., 2020). Similarly, scientific tweeters are also especially concentrated in the USA and Europe. These data sources potentially underestimate the number of early-career researchers in some non-English-speaking countries. To address these sources of country-level heterogeneity, we employed a multilevel model to control for the country-level variations in the population size of researchers and their tweet activities. Additionally, even though we designed a careful matching procedure to link Scopus authors on Twitter with a relatively high recall rate, in cases where Twitter profiles have names and handles that are not related to the actual authors' names, we are unable to match them. It has been found that 76% of scientific tweeters use their true names for their Twitter accounts (Yu et al., 2019). For those using fake names for their accounts, the effect of self-promotion with mismatched Twitter names and author names on publications can be weakened, which may bias our results.

Lastly, although our data collection and period of analysis predates recent changes in Twitter, we acknowledge that with the transition of Twitter to X in July 2023, the platform's approach to granting access to data for research has changed drastically. Along with this shift, a recent survey indicates that thousands of scientists have left X and joined other decentralized social media platforms like Mastodon, Bluesky, and Threads (Vidal Valero, 2023). Despite these recent trends, Twitter has so far remained the most widely used platform for online dissemination of science by scientists and the public and contained the largest portion of the research dissemination recorded by Altmetric (Hayum, 2023; Kidambi, 2024; Sugimoto et al., 2017; Thelwall et al., 2013). However, the future landscape of whether and how these patterns will change remains uncertain. Future research should investigate how the diversification of platforms will shift the use and impacts of social media, and gender differences in these impacts.

4.5 Data and Materials Availability

Scripts and data which allow for replication of our analysis are publicly accessible on GitHub under https://github.com/zxy919781142/Gender-differences-in-online-visibility-of-early-career-researchers.

5

Gender differences in the conflict between the beginning of an academic career and parenthood: Are female researchers less likely to become Ph.D. parents?

This chapter is co-authored with Gemma E. Derrick, Cassidy R. Sugimoto, and Vincent Larivière. It is currently under review for publication in an international journal.

Abstract

Harmonizing the beginning of an academic career and parenthood is challenging for researchers, particularly for women at the doctoral degree (Ph.D.) stage. Using a global survey on academic parents, this chapter examined the timing and sequencing of obtaining a Ph.D. degree and having children, for men and women. Five career trajectories were identified, and women were more likely to dedicate themselves to academic pursuits of Ph.D. degrees early and become parents later. Men, on the other hand, were more likely to have a child while pursuing a Ph.D. degree, and have three or more children. Despite a growing trend of parenthood among more recent doctoral degree holders, women are still less likely to do so than men. Women with children were also found to have lower probabilities of obtaining a Ph.D. than men. From a career-long perspective, parenthood starting at the early career stages is associated with the largest gender differences in citations to scientific work, with men outperforming women. However, among the researchers who delayed having children after 35, gender differences in citation disappeared. This work provides a strong contribution to identifying the dominant and evolving relationship between parenting and academic work and provides insights into other labor environments.

5.1 Introduction

The past several decades have seen tremendous progress in increasing the participation of women in higher education with the US and many European countries observing gender balance at the doctoral level by 2018 (Directorate-General for Research and Innovation, 2021a, 2021b; Llorens et al., 2021). Gender shifts in the age of doctoral recipients have also been observed. Not only has the median age of earned doctorates decreased over time, but the gender gap in the median ages has also been steadily closing: in 1992, the median age of doctoral completion peaked at 33.3 for men and 36.2 for women; while in 2022, women's median age (31.7 years) was only slightly higher than men's (31.2 years) (National Center for Science and Engineering Statistics, National Science Foundation, 2023). However, this period coincides with the peak reproductive ages for women (Mason, 2009b; Mason et al., 2013a). As parenthood creates a new demand to allocate time to childcare, the majority of women who intend to seek or are currently pursuing a Ph.D. degree face the dilemma of deciding when to have children, so as not to jeopardize future academic career progression (Marcus, 2007; Mason et al., 2013a; Paksi et al., 2016).

Mixing doctoral studies with parenthood is a challenge for Ph.D. students who seek to embrace "dual lives", as the roles of doctoral researcher and parent are both highly demanding and require considerable time and energy, especially for women who give birth and need to recover from childbirth. In the U.S., only 8.3% of female Ph.D. recipients who graduated between 2000 and 2005 had children while pursuing their graduate degrees (Kulp, 2019; Zhou and Mollo, 2023). One of the most important reasons contributing to the minority of Ph.D. mothers is insufficient support. Some higher education institutions (HEIs) and nations have made progress towards creating a more inclusive and supportive environment for academic parents, such as flexible working systems in countries with advanced gender policies (Rafnsdóttir and Heijstra, 2011) and policies towards online working after

COVID-19 (Staniscuaski et al., 2020). However, in most cases, these flexibilities are not available to Ph.D. students who are mostly classified as "students" by their institutions rather than "employees" (Moreau and Kerner, 2013), and hence, they are less likely to have access to the organizational policies that are designed to ease the burden of balancing parenthood with work.

In addition to the inadequate support for Ph.D. parents, the role of mothers is seen as conflicting with doctoral research in a way that fatherhood is not (Mason et al., 2023). Ph.D. mothers report facing marginalisation and discrimination in their doctoral education (Mason et al., 2023), as they do not meet the expectation of a "good student" who should be truly committed to their fields of study (Kulp, 2016). The "conventional wisdom" discourages women from having children in graduate school, recommending they delay family formation even until the point of achieving tenure or professorship (Kulp, 2016; Wolfinger et al., 2009; Wolfinger et al., 2008). However, the additional five to seven years of racing against the tenure clock or towards a professorship after a Ph.D. put women squarely at the end of the normal reproductive cycle (Mason, 2009a). Due to the persistent tensions between work and family, women in science face the dilemma of when to have children relative to starting their academic careers, while men have more flexibility in the timing of having children. We hypothesize that there are gender differences in the timing and sequencing of having children and obtaining a Ph.D. degree; more specifically, women are more likely to delay having children after Ph.D. attainment while men have a higher probability of having children during their Ph.D. at the early-career stages. Also, the evolution of the traditional gender roles of men as breadwinners and women as caregivers has broadly changed women's attitude towards career and family and motivated those from more recent generations to seek to be an "ideal academic" and a "good parent" simultaneously (Kulp, 2016)–a dual role that is not necessarily expected of fathers. We further argue that over time, there has been an increasing trend of delayed parenthood among female researchers.

The impact of parenthood on academic careers, in terms of academic performance, promotion opportunities, and career development, has been well-documented and is shown to be highly gendered (Lutter and Schröder, 2019; Tower and Latimer, 2016; Wolfinger et al., 2008; Wolf-Wendel and Ward, 2014). Pregnancy and childbirth are tremendously taxing on women, both physically and mentally, and require a substantial amount of time for recovery. In addition, academic mothers are reported to allocate more time to childcare and have to reduce their working hours, while fathers have higher research productivity after their first child (Derrick et al., 2022; Lutter and Schröder, 2019; Morgan et al., 2021a). The unequal impacts of parenthood on academic mothers and fathers reflect the existence of the "motherhood penalty" and "fatherhood premium" in academia. Moreover, more parenting engagement in turn lessens various academic engagements such as collaboration, visits, and travel, making researchers less visible in the scientific community and their papers less likely to be cited (Sugimoto and Larivière, 2023). Ph.D. mothers, in particular, are more negatively affected by parenthood (Kulp, 2016; Mason et al., 2013a). They are less likely to be centered on academic events and obtain key career-related resources (Kennelly and Spalter-Roth, 2006; Kulp, 2016). Such drawbacks accumulated from the early career stages may put them at a disadvantage in the long-term scientific impact and career development. Considering that the impacts of parenthood differ by the timing of having children, our hypothesis posits that the gender differences in the effects of parenthood on career outcomes are more pronounced among researchers who become parents during the early stages of their careers compared to those who postpone having children.

To discern the gendered patterns of doctoral education and having children for researchers and their association with research performance and academic career, this chapter makes use of the largest survey of publishing parents (Derrick et al., 2022) to investigate the gender differences in the sequencing of Ph.D. attainment and having children from a longitudinal perspective. Considering that the impacts of parenthood differ by the timing of having children, our hypothesis posits that the gender differences in the effects of parenthood on career outcomes are more pronounced among researchers who become parents during the early stages of their careers compared to those who postpone having children. Furthermore, we assess the relationship between the Ph.D.-parenthood sequencing and the career-long outcomes by gender, in terms of scientific impacts and career development. This research highlights the unique challenges in navigating the decision to have a child faced by early-career researchers and how the demands of parenthood influenced their career outcomes when starting their academic careers and in the long run.

5.2 Materials and methods

5.2.1 Data and sample

This chapter used a global survey on the relationship between parenting engagement and academic performance (Derrick et al., 2022). The survey samples comprises 10,445 parent researchers who were still active in publishing with at least one WoS-indexed paper during the period of 2007–2016 as the first or last author. The respondents account for around 0.40% of the entire population of researchers indexed in WoS and all survey participants are parents. Targeting publishing researchers, not directly from the universities or institutes as other surveys of academics have previously done, allowed for the inclusion of surveyed researchers from both academic and non-academic sectors (e.g., private institutes and governments). Another advantage of this sampling is that the respondents can be matched with their corresponding publication records in the WoS, which allowed us to evaluate their academic performance in their life course. The geographic coverage of respondent researchers spans the globe but is more concentrated in English-speaking countries, such as the United States and the United Kingdom. The respondents work in different research fields, including Social Sciences, Arts & Humanities, Health Sciences, and Natural Sciences.

To track the Ph.D. status and fertility history of researchers, this chapter further pre-processed the survey to exclude the respondents with incomplete or erroneous information. We excluded the respondents who lack the birth year of any of their children (25 individuals). We then excluded the respondents who had children at ages younger than 16 or older than 50 (47 individuals) and those whose ages of earning a Ph.D. younger than 20 (26 individuals) as these data are likely to be erroneous rather than anomalous. Since the survey was implemented in 2018, we also excluded the respondents who indicated they had a child born after 2018 (5 individuals), which we consider a mistake in their answers. This pre-processing reduced the sample size to 10,349 respondents. To investigate the conflict occurring in the early stages of their academic career around Ph.D., we decided to track the life-course trajectories of respondent researchers from age 16 until age 40, as around 90% of the respondents got their Ph.D. degrees before 40. After excluding those born after 1978, that is, who were younger than 40 until the survey year, we arrived at the final sample of 8,097 researchers, which represents 77.5% of the original respondents.

Additionally, we coded the researchers' birth years into five cohorts: before 1960 (birth year earlier than 1960); cohort 1960-1964; cohort 1965-1969; cohort 1970-1974; and cohort 1975-1978. The countries of origin for the researchers cover 119 countries, but the United States, United Kingdom, and Canada are overrepresented with around 60% of the sample. Considering the skewed geographic coverage, we broke down the respondent researchers' home countries into the following regions: the US, the UK, Canada, Western Europe, Northern Europe, Southern Europe, Eastern Europe, Oceania, Asia, Latin America, and Africa. The summary statistics of the respondent researchers by different categories were shown in Table. D.1, Appendix D.

5.2.2 Multichannel sequence analysis and cluster analysis of life-course trajectories of researchers

Sequence analysis is an established method to study life courses in social science by efficiently grouping people with similar life-course trajectories (i.e., sequences) and seeking patterns that show up across a number of trajectories. Since we simultaneously considered two separate life course domains of having a child and earning a doctoral degree, we employed an extended sequence analysis, that is, *multichannel sequence analysis* (MSA) (Gabadinho et al., 2011; Gauthier et al., 2010) followed by cluster analysis to identify common patterns of bidimensional (bichannel) life-course trajectories.

The first step of MSA in this chapter was to create bidimensional sequences of Ph.D. status and fertility history for researchers from age 16 to age 40, resulting in a sequence of 25 age-specific statuses for each channel. The two channels in the analysis were measured in the number of children and the Ph.D. status (No/Yes). The rate of having a child and the rate of earning a doctoral degree vary depending on the age, so we opted for the *dynamic Hamming distance* (DHD) with the advantage of strong timing sensitivity to construct the dissimilarity matrix between the sequences. Based on the dissimilarity matrix, we then applied *Ward's hierarchical clustering* method to group the similar sequences (Ward, 1963) and plotted the results using a dendrogram to decide the number of clusters, that is the life-course trajectories comprising Ph.D. status and fertility histories in our analysis (Macindoe and Abbott, 2004). We used the *average silhouette width* (ASW), *Hubert's C index* (HC), and *point-biserial correlation* (PBC) to measure the cluster quality and evaluate the clustering results. Finally, we assigned the cluster membership for each researcher and plot the clustering results. The sequence analysis and cluster solution were conducted in R using the *TraMineR* package (Gabadinho et al., 2011).

5.2.3 Cox proportional hazard models of gaining a doctoral degree

To investigate whether researchers with children have lower probabilities of Ph.D. attainment, we employed the Cox proportional hazard model to estimate the hazard of obtaining a doctoral degree with different numbers of children by adjusting for other variables. The observation years for an individual are from 16 years old until 40 years old, but either the event occurred (obtaining a doctoral degree) or the observation period expired for those who had not yet obtained a doctoral degree, so-called right-censored data. The Cox model is expressed as follows:

$$H(t|Gender, X) = h_0(t)exp(\alpha_0 \times Gender + \sum_{j=1}^{n}(\alpha_j \times x_j + \alpha_{g_j} \times Gender \times x_j)) \tag{5.1}$$

where the $H(t|Gender, X)$ means the hazard rate of obtaining a doctoral degree at age t by taking the *Gender* (female or male) and other characteristics $X(x_1, ..., x_n)$ into account. The characteristics X include the number of children before obtaining a doctoral degree or until 40 years old without getting a doctoral degree, the researcher's birth cohort, geographical region of birth, research field, and whether the researcher's partner is an academic. We took a step-wise method by gradually adding control variables from X in the baseline model (Cox Model 0) which only includes gender. The full model includes all the control variables and their interactions with gender. The coefficients in the models show the hazard ratios, namely, the multiplicative effects of variables on the hazard of obtaining a Ph.D. (Habicht, 2022).

5.2.4 Logistic model for the likelihood of having a child in the early academic career

To see whether there are gender differences in having a child at the beginning of an academic career, we used a logistic model to estimate the probability of having at least one child during the early career stage, which referred to the period of doctoral study and the years immediately following a Ph.D. in our definition. In our data sample, there are 441 researchers without a Ph.D. degree (5%), and we have no information about when they started their academic careers, so we excluded them from this question and got a sample of 7,656 researchers. The duration of doctoral study varies depending on the discipline and country and also differs by individual. A Ph.D. programme typically takes four to six years; hence, we defined the duration of doctoral education as five years. We also consider the five years following a Ph.D. as another period in the early academic career. We separately estimate the probability of childbirth in these two periods using the logistic model below.

$$
\begin{aligned}
logit(P(child_i = 1)|T) &= \beta_{intercept} + \beta_0 \times Gender_i \\
&+ \sum_{j=1}^{n} (\beta_j \times X_{i,j} + \beta_{g_j} \times Gender_i \times X_{i,j}) + \sigma_i
\end{aligned}
\tag{5.2}
$$

Here, $child_i$ is the logistic link function, which indicates the probabilities of having at least one child for researcher i during the period T ($T = 0$ represents the period during the doctoral study (that is, the five years prior to obtaining a Ph.D. degree) while $T = 1$ represents the first five years after a Ph.D.). In addition to the variable of interest, i.e., gender ($Gender_i$), we also considered these control variables in X: the researcher's birth cohort, geographical region of birth, research field, and whether the researcher's partner is an academic, the number of existing child(ern), and the researcher's age. The interactions between gender and other control variables were added to the model.

5.2.5 Outcome of life-course trajectories of researchers

We took the *mean discipline-normalized citation scores* (MNCS) and the employment type to respectively represent the outcomes of research impact and career situation. The *discipline-normalized citation score* (DNCS) was calculated for each publication by dividing its actual citation by the average citation of all publications in the same discipline and publishing year, to make the citations more comparable across research fields and publishing years. Then we get MNCS for each researcher, by taking the average of the DNCS of all their papers that were published prior to reaching the age of 40. The employment status of the researchers in the original survey consists of five different types: research-based academic positions, teaching-based academic positions, research-teaching-based academic positions, non-academic positions which include positions in governments or private sectors, and others including unemployment. We estimated the MNCS using the linear regression model (Eq. (5.3)) and the probability of employment positions using the multinomial logistic model (Eq. (5.4)).

$$\begin{aligned} MNCS_i = {} & \gamma_{intercept} + \gamma_0 \times Trajectory_i \\ & + \sum_{j=1}^{n} (\gamma_j \times X_{i,j} + \gamma_{t_j} \times Trajectory_i \times X_{i,j}) + \sigma_i \end{aligned} \tag{5.3}$$

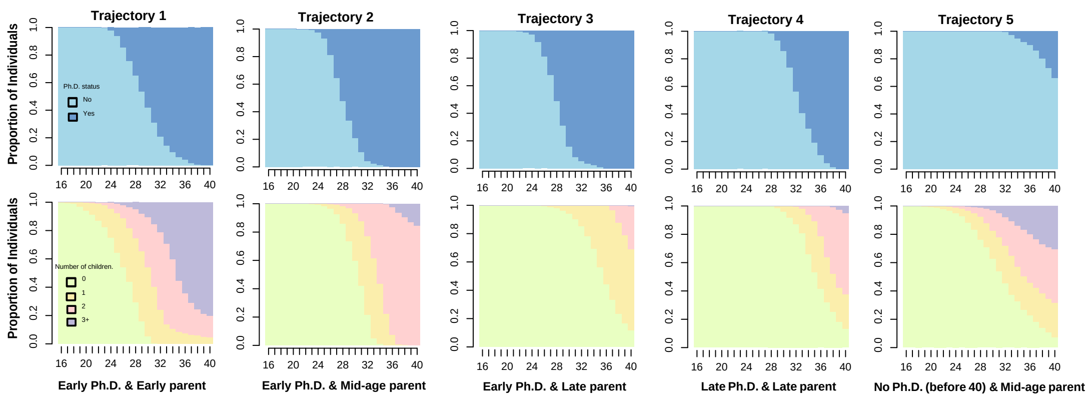
$$\begin{aligned} logit(P(emplo_i = m)) = {} & \lambda_{intercept} + \lambda_0 \times Trajectory_i \\ & + \sum_{j=1}^{n} (\lambda_j \times X_{i,j} + \lambda_{t_j} \times Trajectory_i \times X_{i,j}) + \sigma_i \end{aligned} \tag{5.4}$$

The control variables in X include the researcher's gender, birth cohort, geographical region of birth, research field, and whether the researcher's partner works in academia or not.

5.3 Results

5.3.1 Trajectories of Ph.D. attainment and parenthood

We examined the gender differences in the trajectories of Ph.D. attainment vis-à-vis the decision to have a child, in order to have a comprehensive understanding of the transitions, timing, and sequencing of obtaining a Ph.D. degree and having children among researchers from a gender perspective. We identified five trajectories for publishing researchers in Fig. 5.1: Early Ph.D. and Early parent (T1); Early Ph.D. and Mid-age parent (T2); Early Ph.D. and Late parent (T3); Late Ph.D.and Late parent (T4); No Ph.D. (before 40) and Mid-age parent (T5). Here, we labeled the identified trajectories with an average age of having a first child younger than 30 as "Early parents", while the term "Late parents" was labeled to the trajectories with an average age older than 35. The term "Mid-age parents" described the trajectory with an average age of entering parenthood between 30 and 35. For the channel of earning a Ph.D. degree, the trajectories with an average age younger than 30 were defined as "Early Ph.D.", otherwise, they were defined as "Late Ph.D.". T5 is the only trajectory in which most researchers did not earn a doctoral degree before they were 40 years old, we labeled the researchers in this trajectory as "No Ph.D." since we focused on the trajectories in the early stages of their academic careers before 40. In T2 and T4, researchers tended to have children within the first three years after receiving doctoral degrees. Comparably, the average age of earning a Ph.D. (28.6) is younger in T2 than that in T4 (33.1). Researchers in T3 were likely to prolong having children for over five years after receiving a Ph.D., while T1 is the only group where researchers were likely to become parents first and then earn Ph.D. degrees.



Trajectory	1			2			3			4			5		
	Female	Male	All	Female	Male	All	Female	Male	All	Female	Male	All	Female	Male	All
Num	345	422	**767**	1003	799	**1802**	1212	1005	**2217**	1079	843	**1922**	889	500	**1389**
Average age of Ph.D.	30.72	29.46	**30.03**	28.57	28.59	**28.58**	28.67	28.34	**28.52**	33.07	33.18	**33.12**			
Average age of having first child	25.90	26.78	**26.39**	30.82	30.66	**30.75**	35.63	36.87	**36.19**	35.89	36.89	**36.33**	32.09	33.23	**32.50**
Average Num of children	2.89	3.50	**3.22**	2.19	2.30	**2.24**	1.46	1.64	**1.54**	1.81	1.96	**1.88**	2.13	2.49	**2.26**
Proportion of academic couples (%)	36.16	21.52	**27.92**	33.05	30.91	**32.09**	35.41	34.00	**34.74**	30.48	30.81	**30.63**	20.82	23.01	**21.64**

Figure 5.1: State distribution plots of Ph.D. status (top) and the number of children (bottom) by trajectory. Age is shown on the horizontal axes, and the proportion of researchers belonging to each state at a given time between the ages of 16 and 40 is shown on the vertical axes. The descriptive statistics of the researchers belonging to each trajectory are shown in the table below the plots.

The plurality of researchers can be found in T3: that is, those who had an early Ph.D. and their first child after 35 (2,217, i.e., 27.83%) researchers in the dataset. Researchers usually have only one or two children by the age of 40. That indicated researchers may prioritize obtaining a doctoral degree before having children, resulting in smaller numbers of children overall. The next largest groups are T4 (1,922, 23.74%), T2 (1,802, 22.26%), and T5 (1,389, 17.15%). T1 has the fewest researchers (767, 9.47%) who encompassed the dual lives of a parent and Ph.D. at younger ages (i.e., below 30). The researchers following T1, on average, have around 3 or more children, partly because they started having children at earlier ages.

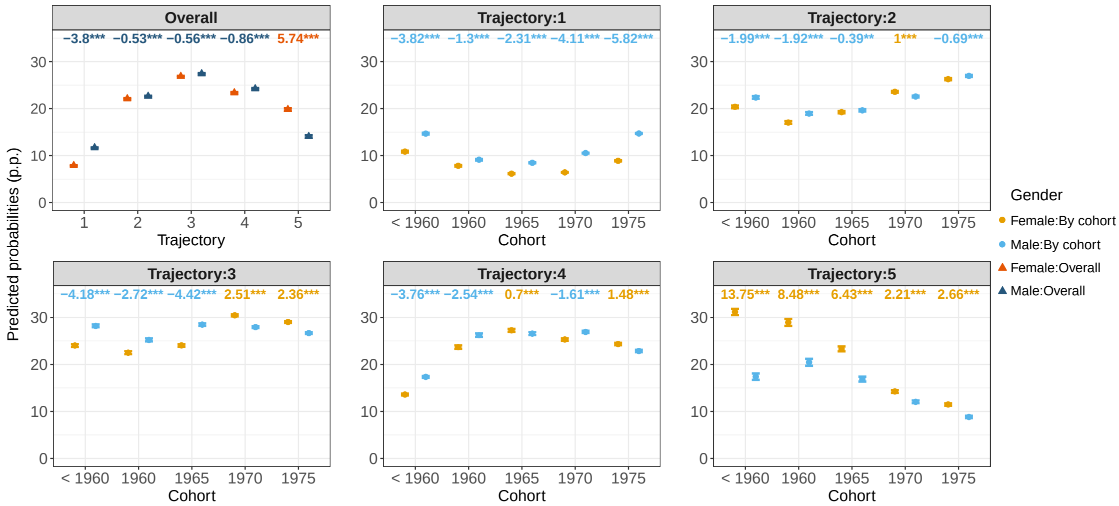


Figure 5.2: Predicted probability of each trajectory by gender across the full sample and disaggregated by cohort. The numbers on the top show the marginal effects of gender on the probabilities in each trajectory with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavoring women.

Fig. 5.2 further explored the gender stratification in the trajectories of Ph.D. attainment and parenthood and how it changed across the cohorts of researchers. From the overall distribution, the largest gender disparities shown in T5 and T1 indicated contrasting patterns of starting an academic career and having children between women and men. Academic mothers without a Ph.D. qualification had a greater likelihood of occurrence. By contrast, men were more likely to pursue a Ph.D. degree and have children at a younger age, before reaching 30 years old. Also, they had a larger likelihood of having more than three children. When looking at the probabilities by cohort, we observed a great shift in the distribution of the trajectories of Ph.D. attainment and parenthood as well as the gender patterns. The occurrence of publishing researchers without a doctoral degree has gradually diminished in academia, as evidenced by the declining probability in T5. This occured along with the decreased gender disparities in this group, dropping from a female-to-male difference of 13.75 percentage points (p.p). in the cohort before 1960 to 2.66 p.p. after 1975. Concurrently, researchers from more recent cohorts were inclined to prioritize completing their Ph.D. studies before having children, typically at the age of around 30 (T2) or even later, after the age of 35 (T3). The pattern of T3 has been increasingly observed among female researchers, especially those from the recent cohorts of 1970 and 1975, reserving the gender difference with a higher probability for female researchers than male researchers. Such shifts in the trajectories indicated a change in gender roles in career and family with time. More well-educated women give priority to their careers before their fertility intentions. Comparably, we also observed a notable increase in the gender gap in the model (T1) where male researchers were more likely to have a child at the early stages of their academic careers.

We further showed how the trajectories of Ph.D. attainment and parenthood differed by interaction between gender and the occupation of the partner in Fig. D.1, Appendix D. It indicated that among the researchers in T1, male researchers were more likely to have a partner not working in academia, while this was not true for female researchers. Academic couples were more likely to be found in the case of T3.

5.3.2 Gender differences in the probabilities of having children during early academic periods

The trajectory of having children during the early stages of one's Ph.D. attainment, known as T1, is characterized by the smallest number of researchers and displays large gender disparities. We further predicted the probabilities of having children during the early academic career, that is, the five years before and five years after obtaining a doctoral degree by cohort (Fig. 5.3 (a)) and country of origin (Fig. 5.3 (b)). By doing so, we can have a closer look at how male and female researchers responded to having children when they started their academic careers around Ph.D. attainment and how the behaviors varied over time and differed by country.

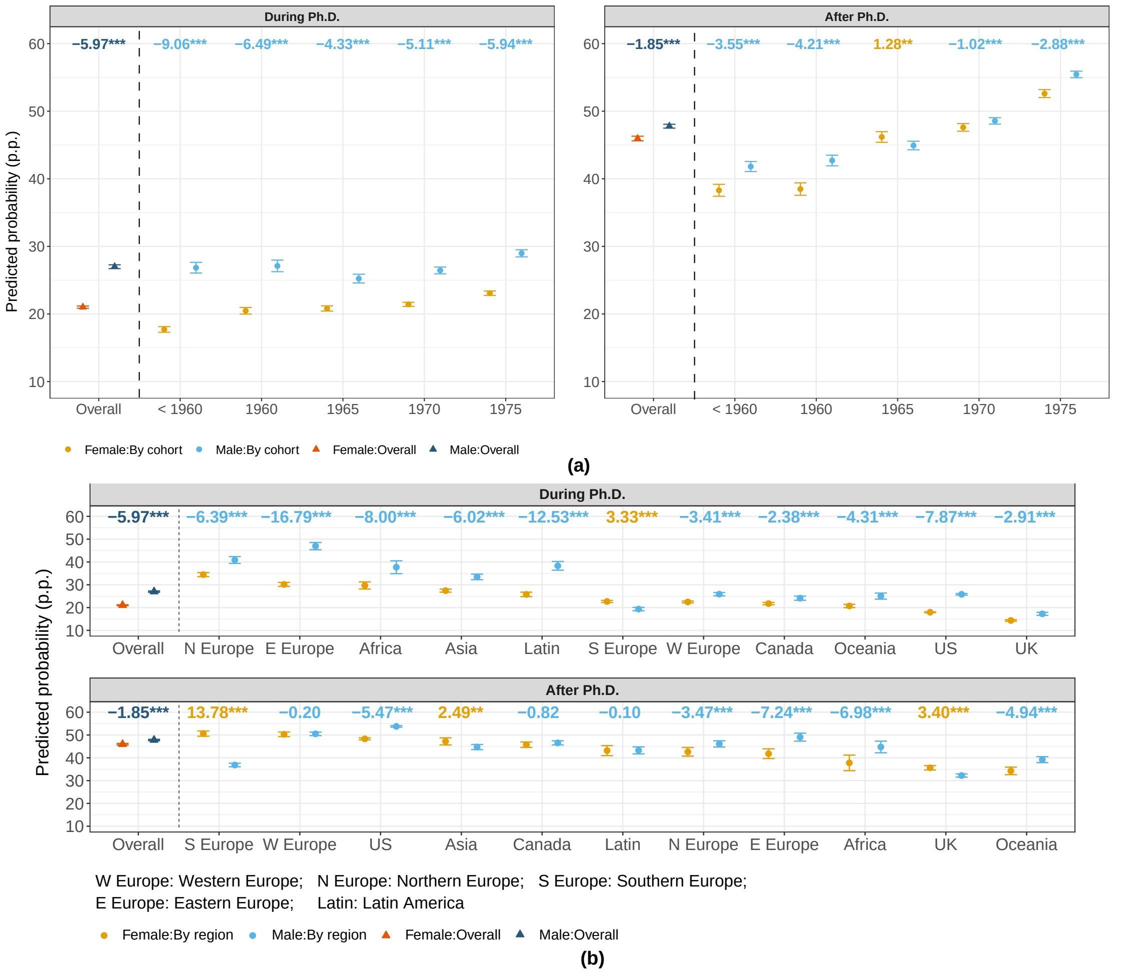


Figure 5.3: Predicted probabilities of having children during the pursuit of a Ph.D. and during the first five years after Ph.D. disaggregated by cohort (a) and geographic region of researchers (b). The geographic regions are ordered by the predicted probabilities of female researchers having children. The first column in each figure shows the overall probabilities. The numbers on the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women.

During the Ph.D stage, Fig. 5.3 (a) showed fewer than one-third of researchers were found to have a child, despite a slight increase over time. Male researchers were comparably more likely to have a child, and the female-to-male differences in the probabilities tend to extend over time, which rebounded from the lowest point of -4.33 p.p. at the cohort of 1965 to -5.94 p.p. among the researchers from the cohort after 1975. This finding was also consistent with the growing occurrence of parenthood among early-career male researchers in the trajectory of T1. When looking at the probabilities across the researcher's origin countries (Fig. 5.3 (b)), Ph.D. parents were more likely to be observed in Northern and Eastern Europe while they were relatively rare in the UK. Eastern Europe also showed the largest gender difference, where nearly half of the male researchers from Eastern Europe were predicted to have at least one child during the Ph.D. stage, 16.79 p.p. higher than their female counterparts. Only in Southern Europe, female Ph.D. students had a higher probability of having children than their male counterparts.

Comparably, both men and women were more likely to have children after rather than during doctoral education. There has been a growing trend among researchers towards this pattern over time, with over 50% of researchers from the cohort after 1975 having at least one child within the first five years after completing their doctoral degree (see Fig. 5.3 (a)). Still, male researchers were more likely to do so than female researchers. The only exception occurred in the cohort of 1965, showing a higher probability among female Ph.D. recipients who were born between 1965 and 1969. When disaggregated by country, female researchers from Southern Europe showed the highest tendency to have children within five years after doctoral study, with a noticeable female-to-male gender difference of 13.78 p.p in probabilities. Although there were fewer Ph.D. parents in the U.S., a comparably higher number of researchers from this country were observed to have children immediately after obtaining their Ph.D. It is important to emphasize that only a small group of researchers form the U.K. were likely to have children during the academic stage around their Ph.D. attainment (both five years before and five years after obtaining a doctoral degree).

5.3.3 Parenthood impact on the rates of Ph.D. attainment

In addition to the individual choice to delay childbearing beyond child-bearing stages, researchers may also choose to postpone parenthood beyond the period of doctoral study because of the negative impacts it may have on completing the Ph.D. We compared the probabilities of earning a Ph.D. degree among researchers with different timings of having children and investigated gender differences in the probabilities using survival analysis.

Fig. 5.4 (a) used Kaplan-Meier (KM) survival curves to show the instantaneous rates of obtaining a Ph.D. between the ages of 16 and 40 for female and male researchers with different numbers of children (0, 1, 2 and 3+; here the number of 0 means no children before obtaining a Ph.D. degree). The visual inspection and the pairwise Log-Rank tests between female and male researchers with the same parenthood status indicated that there was no significant gender difference in the rates of obtaining Ph.D. degrees between researchers without children prior to Ph.D. attainment. However, the rates of Ph.D. completions were negatively associated with the number of children, and the gender differences became visually and statistically distinguishable among the academic parents. It showed that parenthood penalties existed in the early academic career in terms of the rates of obtaining a Ph.D., especially for women.

To further investigate the gender difference in the probabilities of obtaining a Ph.D., we then used a Cox model to estimate the hazard ratios of obtaining a doctoral degree by gender. The baseline Cox model (Cox Model 0, shown in Table. D.2, Appendix D) which only includes gender, indicated that the probability (i.e., hazard ratio in Cox model) of obtaining a doctoral degree for women was on average 17 percentage points (p.p) lower for women than for men. We successively added other variables that might affect the probability of obtaining a Ph.D. degree, including the number of children, the researcher's birth cohort, geographical region of birth, the research field, and whether the researcher's partner is an academic. We also added the interactions between these variables with gender to explore how gender differences in obtaining a Ph.D. degree across other factors (see the estimate results in Table D.2, Appendix D). With the inclusion of the number of

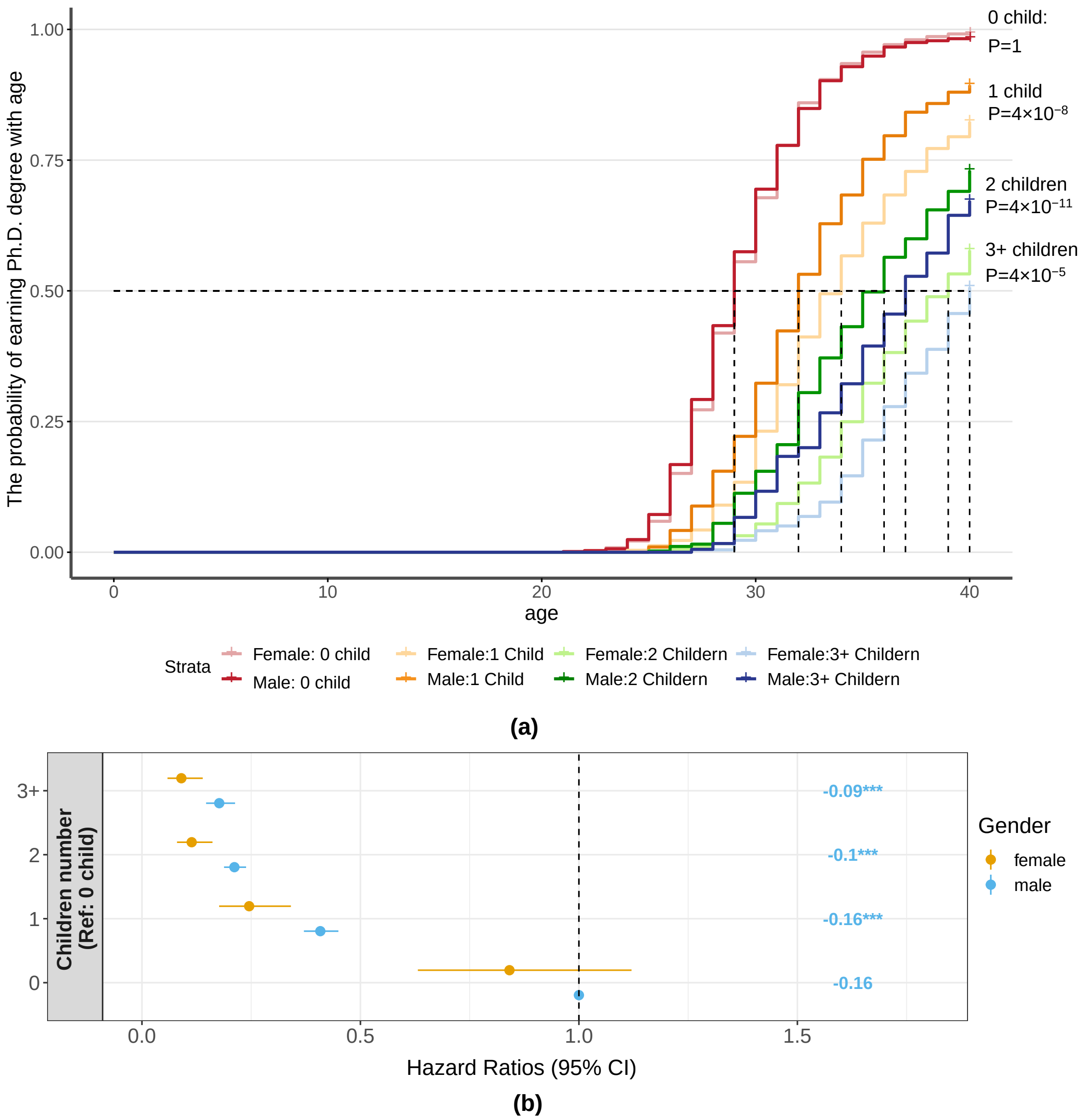


Figure 5.4: (a) Kaplan-Meier (KM) survival curves of time to obtain a Ph.D. for researchers separated by the interaction of gender and the number of children (0,1,2, and 3+). The p values show the significance level from the pairwise Log-Rank test and the dotted lines show the corresponding ages at which the probabilities reach 0.5 (b) Plotted hazard ratios (exponential coefficients) of obtaining a doctoral degree by gender and number of children (1,2, and 3+, reference is childless), with 95% confidence intervals. The values on the right show the gender differences in the hazard ratios with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women

children and other variables, the impact of gender became insignificant; rather, the number of children played a significant role in the probability of obtaining a Ph.D. degree and drove the gender disparity.

Based on the full model (Cox Model 5 in Table D.2), Fig. 5.4 (b) showed the estimated probability of Ph.D. completions conditioned on the gender and the number of children with men without children before Ph.D. as reference (reference value is 1), after considering the interactions between gender and other variables. In line with the previous findings in the KM curves, parenthood was shown to conflict with Ph.D. attainment, as the probability of obtaining a Ph.D. for academic parents falls below half of what it is for men without children. Also, parenthood contributed to the gender difference in Ph.D. attainment, and the largest gender gap disfavoring women was seen in the case of having one child (-0.18 p.p.). The survival analysis on the gender differences in the probabilities of Ph.D. attainment across generations can be found in Fig. D.2, Appendix D.

5.3.4 Career-long outcomes of Ph.D.-parenthood trajectories by gender

Beyond the parenthood penalty at the early career stage in terms of Ph.D. attainment, we further assessed how the timing and sequencing of having a child and obtaining a Ph.D. influenced the career-long outcomes of academic performance (Fig. 5.5) and career paths (Fig. 5.6) among female and male researchers. Fig. 5.5 shows that, on average, male researchers had a higher scientific impact in terms of *mean discipline-normalized citation scores* (MNCS) prior to the age of 40 than their female counterparts. The exception was shown in the trajectory of T5 which consists of researchers without Ph.D. degrees; in this case, female researchers tended to perform better in the receipt of citations. In T3, where researchers who started their academic career before 30 but delayed having children after 35 years old, there were no significant gender differences in the scientific impact. By contrast, the largest gender difference was shown in the group of researchers who became parents and obtained a Ph.D. at young ages and have more than three children by 40 (T1); male researchers received, on average, around 0.33 higher MNCS than female researchers. Shifting from T3 to T2 to T1, the increase in the amount of time spent on parenting responsibilities throughout the academic career was accompanied by a steep decline in citation scores for female researchers. However, male researchers did not experience such a loss in their citations. It implied that female researchers experienced greater negative impacts from parenthood in academic outputs as measured by citation, that is, the intense investments in parenting engagements (Derrick et al., 2022; Morgan et al., 2021a) created accumulative disadvantages for female researchers over time, resulting in the gaps in scientific impact.

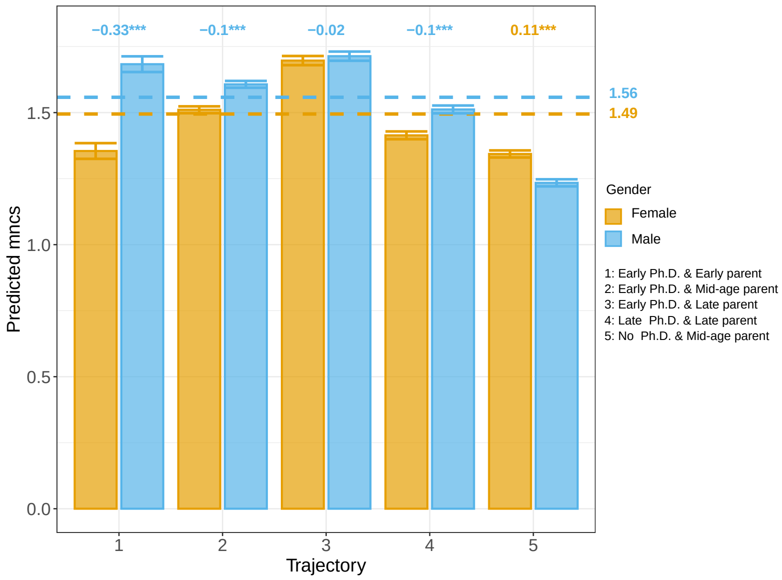


Figure 5.5: Predicted *mean discipline-normalized citation scores* (MNCS) of female and male researchers by trajectory from the start of publishing age until the age of 40. The numbers on the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavoring women. The horizontal lines with annotated values on the right mean the average predicted MNCS across trajectories by gender.

Fig. 5.6 compared the probabilities of working in academia (research-teaching-based academic positions only, left penal) and outside academia (right panel) for female and male researchers aged over 40 across five trajectories. Here, we showed the probabilities of working in the sectors for both research and teaching to represent the likelihood of working in academia, since the majority of publishing researchers

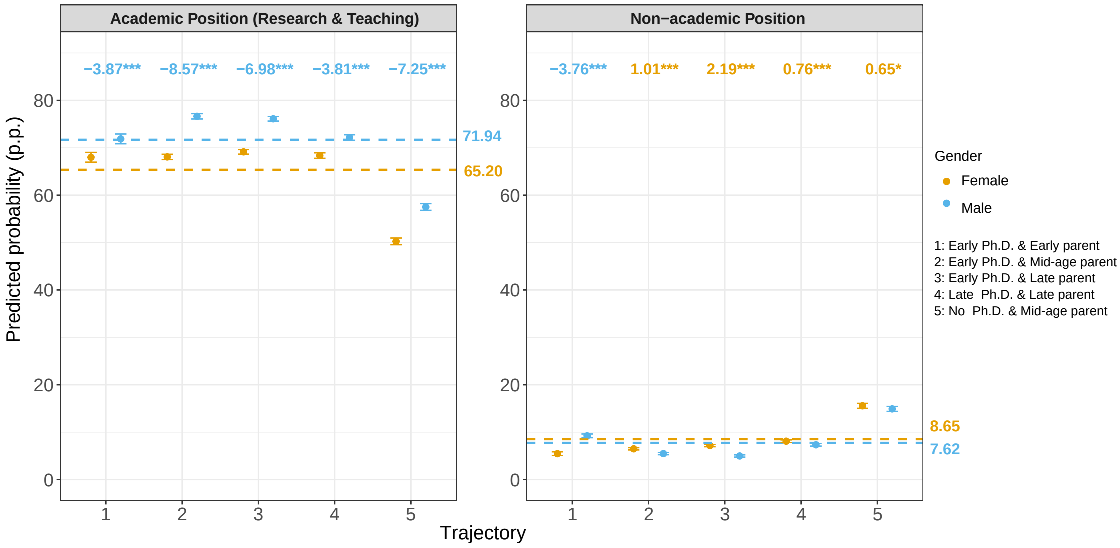


Figure 5.6: Predicted probabilities of working in academic positions (left) and non-academic positions (right) across trajectories by gender. The numbers on the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavoring women. The horizontal lines with annotated values on the right mean the average predicted probabilities across trajectories by gender

who stay in academia are employed in research-teaching-based academic positions (77%). The results for a minority of researchers who were predicted to work in the academic sectors of only research or teaching can be found in more detail in Fig. D.3, Appendix D. The dual-task sectors comprising both research and teaching in academia showed some gender differences, varying by the identified Ph.D.-Parenthood trajectories. Male researchers across all trajectories had higher probabilities of working in academic positions, particularly in the trajectories T2 (-8.57 p.p), T3 (-6.98 p.p), and T5 (-7.25 p.p) with larger gender differences favoring male researchers. Male researchers who obtained Ph.D. early but later entered parenthood after Ph.D. attainment (T2 and T3) were shown to be more likely to work in academic sectors. This can be partly explained by the accumulated advantages for male researchers staying in academia since Ph.D. attainment. Even though generally lower probabilities for both genders in T5, male researchers, mainly those from the previous cohorts when doctoral degrees have not become a critical requirement in the academic position for research and teaching, were still significantly more likely to serve in these positions.

The right panel of Fig. 5.6 shows that a small fraction of researchers who were active in publishing in the past decade work in sectors outside academia, such as government or private companies. Not surprisingly, female researchers with children had a higher probability of working in non-academic positions, which contributes to the "leaky pipeline" in academia. However, male researchers who became fathers at younger ages and have over three children (T1), were also more likely to engage in non-academic employment, with a 3.76 p.p. higher likelihood than their female counterparts.

5.4 Discussion

This chapter analyzed the trajectories of 8,097 academic researchers, in terms of how they sequenced the completion of their Ph.D. relative to having children between the ages of 16 and 40. The life-course analysis allowed the study to capture a historical retrospective view of the career-long cost to female and male researchers of having children, which is otherwise lacking in the literature. The main finding shows that overall, obtaining a Ph.D. degree before 30 and having children after 35, is common for most researchers, especially female researchers born after 1970. Men, comparatively speaking, were more likely to balance early parenthood at the same time as completing their Ph.D. studies. At the beginning of an academic career, women with children are less likely to earn a Ph.D. degree compared to men, even though there was no significant difference in the probability and the timing of obtaining a Ph.D. among women and men without children prior to Ph.D. attainment. From a long-term perspective, academic mothers were disadvantaged in the overall citations of publications compared to their male counterparts, and the gender gaps are most pronounced among those who have children early when starting their academic careers. In addition, academic fathers were more likely to stay in academic positions for longer periods than their female counterparts.

The "leaky pipeline" in academia— where women are more likely to leave academia before reaching the highest eschelons—has been a concern for decades. Parenthood is an important factor of gender disparity in academic advancement, with women leaving academia earlier than men because of greater care responsibilities placed upon them (Mason et al., 2013a). Conventional wisdom has been for women to delay parenthood, until securing a tenured or permanent position. From the five identified trajectories of researchers who were active in publishing until 40 years old, we found that a significant number of academics, particularly women, followed this suggested path and delayed having children until they were over 35 years old, approximately eight years after completing their Ph.D. degrees. This trend echoes the importance of a doctoral degree as a requirement in academia, with female researchers more likely to be dedicated to Ph.D. completion first without having children. On the other hand, it reflects the incompatibility between parenthood and the ambition and motivation of early-career female researchers to advance their academic careers. By contrast, there is no similar assumption of fatherhood in academia. Men are found to have higher probabilities of simultaneously completing a Ph.D. education and having children at young ages below 30. They, accordingly, tend to have more children compared to women overall. The gender inequalities in parenting engagement indicate women are more likely to serve in lead roles (Derrick et al., 2022), allowing men to better balance academic work and family life with more flexibility and fewer professional consequences.

Even though the previous literature attributed the under-representation of women in science to the gender differences in ability and occupational expectations, our results indicate that there is no significant difference in gaining a Ph.D. degree between women and men before having children. From a longitudinal perspective, women and men who pursued advanced academic professions, such as obtaining a

Ph.D., and delayed having children until the age of 35 (i.e., early Ph.D. and late parenthood, T3 in our analysis) do not differ significantly in their scientific impact, as shown from their normalized citations. However, when parenting responsibilities are taken into account, women face additional obstacles, resulting in larger gender inequalities in the rates of earning a Ph.D. that are unfavorable to women. Also, the gender disparity in the impacts of parenthood is accumulated from the early-career stages to the long run. Women who commenced childbearing at a younger age, along with having more children, experienced greater adverse consequences on their scientific impact while no similar impacts were found on men's citations. Such gaps may also be related to differential engagement in parenting: women are reported to be more engaged in parenting with more daily tasks of caring for children but fathers can protect their research time (Derrick et al., 2022). In addition, even though female researchers are reported to have an increasing desire for higher professional ranks (Drake and Svenkerud, 2023; Sharafizad et al., 2020), mothers are found to have a lower probability of staying in academia than fathers, which also contributes to the underrepresentation of women in academia. Some evidence shows that parenthood may also change the work preferences of researchers and women focus more on teaching than research (Morgan et al., 2021a). Overall, in the academic field, parenthood plays a different role in the professional life of women and men while women often suffer more from parenthood.

This chapter contributes to the literature by providing a longitudinal view of the gender differences in the timing and sequencing of starting an academic career by earning a Ph.D. degree and having children. The existing literature mostly focuses on the decision to have children when researchers are already in the stage of habilitation or higher academic positions, missing the important time point of doing a Ph.D. in the pipeline of academic career. The Ph.D. stage normally overlaps with the peak time of childbearing for women, which places unique pressure on the decision of when, and if, to have children. This paper goes earlier to the Ph.D.

period to have a comprehensive discussion on the trajectories of having children before the Ph.D., after the Ph.D., and/or later career stages. The career-long perspective allows us to monitor the career cost of the precarity of decisions about parenthood at early-career stages compared to postponing until later years around the time of achieving tenure. Also, the discussion over the contexts of generations and countries deepens the understanding of how the choices of when to have children and when to start an academic career change with time and differ by country.

Despite providing an analysis of career-long effects of parenthood at various times of the academic career including the period of Ph.D. study, this chapter has some limitations. We acknowledge that WoS is dominated by English-language publications from Western countries, making the researchers in our dataset over-represented in the United States, United Kingdom, Canada, Australia, and New Zealand, accounting for more than 60% (Derrick et al., 2022). The broad geographic coverage of the survey still allows us to observe some geographic heterogeneity in the gendered patterns of the timing and transitions of earning a Ph.D. and having children. Additionally, even though we have considered the number of children in the work-family conflict rather than the simply binary categorization of parenthood (that is, having children or childless) and concluded that the motherhood penalty intensified in the families of more children, we did not have evidence on the extent to which parent researchers differ from childless researchers as the dataset only includes the parenting researchers. Similarly, as we only know the year of obtaining a doctoral degree for each respondent researcher without the time of starting the doctoral study, we assumed the Ph.D period as the five years before obtaining a doctoral degree, which may cause some erroneous categorization of our data. But still, we made the assumption for both women and men and observed some gender differences in the probabilities of having children in the given time. For those who chose to leave academia, there may be other factors that influenced the long-term career outcomes in addition to parenthood. In the next step, we plan to extend the survey to include childless researchers and incorporate more career-wise questions to gain a better understanding of the parenthood penalty in academia.

By giving validation to the testimony of the challenges of combining research careers with parenthood for women at their early career stages, this chapter highlights the bidirectional conflict between Ph.D. research and parenthood while Ph.D. parents, especially Ph.D. mothers, who take on the highly demanding responsibilities of being a doctoral researcher and a mother received limited support. Hence, we call for more support and understanding from the universities and the whole academy to make the role of Ph.D. mothers compatible and acceptable in academia. Formalized parental leave equally for Ph.D. students would give Ph.D. mothers more flexibility to harmonize their studies and family and improve the retention of women in science. Nordic countries are known for their generous parental leave policies which are also available to Ph.D. students (Brandth et al., 2012). Parents in Sweden are allotted 480 days to stay home with their newborn, 360 of which are paid (Cardoso, 2022). However, in the US, UK, and other similar countries, which constitute the majority of our sample, paid maternity or paternity leave is still not widely available for Ph.D. students. Meanwhile, parental rather than just maternal leave policies may incentivize fathers to be more involved in parenting, contributing to the movement for gender equality in academia. In addition, the concept of family-friendly workplaces with accessible and affordable child care, such as breastfeeding rooms and university nurseries, is still new in most universities or institutes (Lantsoght et al., 2021). Moreover, across academia, more actions are needed for implementation to support mothers' academic careers and equalize promotion opportunities. Such examples of positive actions include extending for a few years the window of eligibility for fellowships and grants for academic mothers and creating specific grants or project calls for mothers with long career breaks. Such measures may encourage academic mothers to continue with their scientific careers even after long breaks for parental leave.

5.5 Data and Materials Availability

Scripts and data which allow for replication of our analysis are publicly accessible on GitHub under https://github.com/zxy919781142/Gender-differences-in-the-conflict-between-the-beginning-of-an-academic-career-and-parenthood.

6
Conclusion

In an age when gender diversity and inclusion are of growing importance, this thesis depicts a more comprehensive picture of gender inequalities in science, especially against the background of the wider forces of the globalization and digitalization of scientific systems. Each of the four empirical chapters focuses on a specific perspective of gender disparity in science and research, relying on bibliometric data and other traditional and non-traditional sources of data. The first two studies exploited bibliometric data to establish a comprehensive framework for studying international scholarly migration and assessing the gender disparity in the process using different measures. The second study further extended from the case study of German-affiliated researchers in the first study, to the global pattern of migration of scholars from a gender perspective. The third study shifted attention to gender differences in early-career researchers' online visibility, including their self-promotion behavior, and examined their effect on subsequent scientific impact. The fourth study discussed the gender differences in work-family conflict among researchers by discerning the patterns regarding the timing of doctoral education and having children and further examining whether the patterns were related to the academic outcomes.

Female researchers who leave Germany are less likely than their male counterparts to return. Chapter 2 identified the German-affiliated researchers from Scopus bibliometric data and shed light on the differences in gender composition and career stages between researchers who remained in Germany, those who emigrated, and those who eventually returned. The gender imbalance in the German science system was found to be further exacerbated by migration trends among female and male researchers. The persistent gender disparity can be explained by that female emigrant researchers had a greater tendency than their male counterparts to live abroad for longer periods, or possibly to settle down in other countries. In almost all disciplines, the subgroups who returned to Germany were also more male-dominated than the subgroups who left Germany. These findings are consistent with the previous argument by Zippel (2017) that return migration had potentially large implications for the persistence of gender inequalities in academia. From the career perspective, the returnee researchers of both genders appeared to benefit from their international research experience, as it considerably prolonged their academic careers. However, male returnee researchers were still more likely to stay longer than their female counterparts in academia, intensifying the gender disparities throughout the academic life cycle. The advantages favoring male researchers could partly contribute to more male researchers returning to Germany than female researchers.

This chapter uncovered new dimensions of scholarly migration by investigating the return migration of researchers and how different moving trajectories contribute to gender patterns in the national science system. However, some questions including what personal and professional factors drive the gender differences in the scholarly migration trajectories, could not be answered with bibliometric data alone and need additional measures or policies to foster connections with these previously German-affiliated researchers. Investigating the critical topics, like employability and careers of emigrant and returnee researchers could provide deeper insights into the gender differences in the decision-making of returning. This chapter also emphasizes the need for more support and incentives to attract researchers, especially female researchers, to return and facilitate their reintegration into the German science system. These findings also have significant implications for other countries and also for other occupations that need advanced skills, particularly those experiencing substantial brain drain and gender inequality.

Women in science are less mobile, and they move over shorter distances, but the gender gap has shrunk. Chapter 3 further explored the international mobility of researchers from around the world and provided the first global and dynamic view of gendered patterns of transnational scholarly mobility. While female researchers were less internationally mobile than men, the gender gap has shrunk considerably and decreased at a faster pace than the overall gender gap among researchers. In addition to moving less, female researchers also originated from and moved to fewer countries, as well as migrated shorter distances than their male counterparts. That indicates that female researchers were still restricted in moving globally as freely as their male counterparts. Gender gaps among researchers and international scholars that favor men were smaller in high- and upper-middle-income countries than in low-income countries. The US, the UK, and Germany remained popular for both female and male mobile scholars, but in these global hubs of international science, gender gaps nonetheless persisted.

These results give some positive hints that opportunities to advance their academic careers through international mobility have increased for women. Still, there is a long way to go until female and male researchers in the global science network and transnational scholarly migration have equal opportunities, resources, and outcomes. In addition to limited opportunities, female researchers also experience more challenges in the reconciliation of career and family, which also hinders them from relocating (Kulis and Sicotte, 2002). It is imperative to provide more comprehensive support and incentives to advance the professional development of female researchers, including networking and collaboration opportunities, short-term visiting programs, and funding for academic migration, which can provide female researchers with greater flexibility.

Chapters 2 and 3 both exploited a novel source of digital trace data —bibliometric data— to study the migration of published researchers and the demographic composition of the scientific workforce. It demonstrated that this innovative data can supplement the lack of time series of comparative and gender-disaggregated data on the migration of scholars, to gain a better understanding of the gender disparities in the movements of academic talents. By integrating demographic perspectives in bibliometric research, these studies implied that globalization is an important factor that can intensify some of the gendered inequalities in contemporary science, and meanwhile, gender inequalities continue to shape the globalization of science. Our established methodology and study framework can foster the development of bridges between demographic and scientometric research.

Men are more likely to self-tweet their research and receive additional citations by self-promotion, which again widens the gender gap. Chapter 4 linked the publications of early-career researchers from bibliometric data with social media data to assess the gender differences in the online visibility of early-career researchers and thus examined if social media can help equalize the visibility and impact of researchers and improve gender equality in academia. This study contributed to a comprehensive understanding of gender disparities in online visibility by distinguishing the push (promoted by others) and pull (self-promotion) mechanisms at play within online academic dissemination. In contrast to the previous studies that found lower online visibility of female researchers (Paul-Hus et al., 2015; Vásárhelyi et al., 2021), the focus on early-career researchers, which can to some extent minimize the impact of accumulated resources and networks, showed the potentially positive role of social media for amplifying the research of female scholars at the start of their academic careers. However, the persisting gender inequalities in self-promotion, in turn, limited the full potential of these impacts and reproduced the existing gender inequalities. These findings, on the one hand, revealed the potential of social media for female researchers to increase their visibility and impact, and on the other hand, suggested the value of integrating training in self-promotion on social media within graduate school curricula and encouraging early-career female researchers to self-promote their work. In lower-middle-income countries (LMIC), promoting women's access to technology and building their digital skills can benefit more women in science with an efficient gateway to knowledge exchange and self-promotion. From the perspective of data science, the linkage between bibliometric data and social media data is an innovative and meaningful practice to investigate the interplay between the internationalization of research knowledge, traditional venues of the scientific impact, and the gender disparity in academia. This study gained a deeper insight into the process of gender inequality from online to offline and the interplay between online and offline in the digitalized era.

Having children early in academic careers widens the gender differences in citation scores, with women disadvantaged. Chapter 5 discussed how the timing of parenthood, particularly in relation to Ph.D. attainment, was associated with the scientific outcomes from the gender perspective. Women were more likely to dedicate themselves to academic pursuits of Ph.D. degrees early and delay having children, while men were more likely to balance parenthood and Ph.D. attainment at earlier ages, as well as have more children (>3 children) by the age of 40. The conflict between Ph.D. attainment and parenthood for women was evidenced by the lower possibility of obtaining a Ph.D. degree for mothers than fathers, despite no significant gender difference in the rates of obtaining a Ph.D. degree between those without children. Female researchers also experienced greater negative impacts from parenthood in academic outputs measured by citation, which indicated that investments in parenting engagements created cumulative disadvantages for female researchers over time, resulting in gaps in scientific impact.

This study contributed to providing a longitudinal view of the gender differences in the timing and sequencing of starting an academic career and having children. By giving validation to the testimony of the challenges of combining research careers with parenthood for women at their early career stages, this study highlighted the bidirectional conflict between Ph.D. attainment and parenthood and called for more support and understanding from the universities and the whole academy to make the role of parents compatible and acceptable in academia. It also provided insights into other labor environments to create more supportive workplaces for parents, especially mothers. Additionally, this study was based on a global survey, where the samples were generated from publishing researchers in bibliometric data and further matched to their corresponding publication. The demographic information and life course trajectories obtained from the survey and then complemented by the scientific metrics offered a novel lens to quantify the gender gaps in the cost of having children to academic performance.

This thesis contributed to drawing a more comprehensive picture of the gendered structures in contemporary academic environments. Globalization and digitalization have profoundly shifted the process of knowledge production and transmission, but they did not thoroughly change the gendered pattern in the scientific workforce and academia and even intensified some of the persisting gendered inequalities due to the fact that advanced tools and opportunities are still more likely to be utilized by men. Parenthood is still a significant factor that generates gender disparity in academic careers and trajectories, as evidenced by persisting and longer-term motherhood penalties.

Methodology-wise, the main contribution of this thesis was to integrate bibliometric data with other traditional and non-traditional data to assess gender disparities in academia across different dimensions and in a more cross-disciplinary, scalable, longitudinal, contemporary, and comprehensive way. More broadly, this thesis emphasized the value and potential of digital trace data including bibliometric data and social media data in advancing fundamental population science and sociological studies, through the lens of digital and computational perspectives.

Gender disparity has been deeply rooted in academia, and perpetuated in almost all dimensions until now. Despite this, this thesis uncovered some promising signals for female researchers and aimed to provide advice to mitigate the gender disparity in academia. Female researchers have become more internationally mobile than ever before, and some countries, such as Serbia, Argentina, and Portugal have reached near gender parity among migrant researchers. More targeted scholarship and fellowship programs funded by governments, multilateral organizations, and private foundations are critically needed to further help women advance their academic careers through relocation or international collaboration networks if relocation is not ideal. Moreover, it is necessary to offer some support and help for female migrant researchers to strengthen their integration into the destination community.

The first publications by female researchers were, on average, more likely to be mentioned on Twitter, indicating online platforms provide a channel for female researchers' scientific work to be positively selected on attributes, hence increasing the chance of being seen and shared by a larger audience. Considering the additional benefit of self-promotion on social media compared to being promoted only by others, it is valuable to integrate training in self-promotion within graduate school curricula and encourage early-career female researchers to self-promote their work. In terms of conflict between work and family, women who pursued academic professions first and delayed having children performed as well as their male counterparts. It does not mean women in academia should delay or even avoid having children, but it indicates that parenthood is an important factor that contributes to the gender gaps in academia, which is related to the differential engagement of women and men in parenting. The concept of family-friendly workplaces with accessible and affordable child care, which can help ease the burden on mothers, should be widely adopted in universities and institutes. From the family perspective, encouraging fathers to engage more in parenting responsibilities can help mother researchers protect their research time, which is important for their career advancement (Morgan et al., 2021b). Lastly, we have to admit that there is still a long way off to fully achieve gender equality in academia. More concrete actions are required to address different facets of gender inequality in academia, and these efforts should be made by each entity, from individuals to professional organizations to governments. One day, the glass ceiling in academia can be broken, and the leaky pipeline can be fixed.

Appendices

A

Appendix to Chapter 2

Mapping topics to disciplines

Table A.1 provides the intermediate results from the topic model (30 topics) and how the topics are mapped into 17 disciplines (inspired by the All Science Journal Classification) based on their similarities.

Table A.1: Details of mapping 30 topics to 17 disciplines

Topic	Most frequent keywords	Discipline result
01	space, earth, wave, model, surface, field, datum, rock, seismic, structure	Earth and Planetary Sciences
02	material, property, surface, apply, structure, film, growth, metal, thin film, magnetic	Material Science
03	disease, cardiovascular, patient, clinical, function, heart, medicine, cardiac, lung	Medicine
04	infection, disease, clinical, immunology, virus, patient, vaccine, microbiology, transplantation, blood	Immunology and Microbiology
05	ecology, population, evolution, forest, diversity, specie, genetic, biology, animal, conservation	Agricultural, Biological and Environmental Sciences
06	protein, molecular, gene, genetic, biology, cell, biological, expression, nature, mutation	Biochemistry, Genetics and Molecular Biology
07	model, simulation, engineering, flow, dynamic, design, numerical, structure, modeling, experimental	Engineering
08	network, note computer, subserie note artificial intelligence, note bioinformatic, ieee, information, model, datum, design, engineering	Computer Science
09	patient, surgery, clinical, treatment, cancer, outcome, disease, therapy, pediatric	Medicine

10	plant, cell, metabolism, physiology, metabolic, stress, biology, response, enzyme, arabidopsis	Agricultural, Biological and Environmental Sciences
11	brain, neuroscience, rat, mouse, alzheimer disease, neurology, parkinson disease, disease, model, receptor	Neuroscience
12	laser, optical, engineering, spie society optical, optic, measurement, spectroscopy, fiber, pulse, apply	Physics and Astronomy
13	management, health, economic, education, social, germany, development, policy, review	Economics and Social Science
14	cell, cancer, expression, tumor, gene, molecular, stem, receptor, mouse, apoptosis	Biochemistry, Genetics and Molecular Biology
15	polymer, material, composite, engineering, property, surface, fiber, coating, application, technology	Material Sciences
16	nanoparticle, chemistry, surface, cell, membrane, chemical, physical, spectroscopy, microscopy, electrochemical	Chemistry and Chemical Engineering
17	chemistry, synthesis, chemical, structure, complex, reaction, organic, crystal, molecular,	Chemistry and Chemical Engineering
18	cancer, patient, therapy, treatment, clinical, breast cancer, carcinoma, radiotherapy, tumor, oncology	Medicine
19	galaxy, star, astrophysical, cluster, society, xray, monthly notice royal astronomical, formation, general, evolution	Earth and Planetary Sciences
20	power, ieee, electronic, communication, technology, sensor, ieee transaction, measurement, device, application	Engineering
21	energy, process, technology, production, engineering, gas, atmospheric, environmental, water, development	Energy
22	drug, clinical, skin, exposure, risk, allergy, medicine, test, assessment,	Pharmacology, Toxicology and Pharmaceutics
23	physics, physical review, letter, physical, energy, measurement, letter section, high, nuclear, elementary particle, high energy	Physics and Astronomy
24	water, marine, environmental, new, sediment, climate, carbon, change, soil, ocean	Earth and Planetary Sciences
25	patient, child, disorder, cognitive, treatment, therapy, rehabilitation, physical, depression, pain	Psychology
26	model, theory, datum, mathematical, function, application, problem, dynamic, stochastic, time	Mathematics
27	surgery, bone, injury, treatment, fracture, tissue, trauma, surgical, clinical, implant	Medicine
28	soil, food, plant, apply, microbial, nutrition, activity, production, quality, microbiology	Agricultural, Biological and Environmental Sciences
29	imaging, image, magnetic resonance, mri, medicine, stroke, brain, radiology, ultrasound, compute tomography	Health Professions
30	plasma, physics, solar, nuclear, fusion, beam, ion, electron, radiation, nuclear instrument	Physics and Astronomy

B
Appendix to Chapter 3

B.1 Validation of gender detection

The first established database we used for validating our gender detection process is a name-gender dictionary of 16,921 female first names, 17,740 male first names, and 9,907 unisex first names in 68 countries worldwide (Winkelmann, 2022). The second dataset we used was a manually calibrated dataset of 4,558 authors (1,447 female authors, 2,980 male authors, and 131 authors without identified genders) extracted from top sociology journals (Akbaritabar and Squazzoni, 2021). We employed the indicators of *precision* and *recall* rate to separately judge the detection accuracy for female and male first names in different countries. The *precision rate* of female names quantified the proportion of the correctly identified female first names among all the first names we identified as female, while the *recall rate* quantified the proportion of the correctly identified female first names among all the female first names in the name-gender dictionary. The two indicators were also applied to the male first names.

For the validation of the first dataset, our gender detection process worked well in almost all countries. The evaluation results for a portion of the countries or regions are shown in Table B.1. In general, the *precision rate* and the *recall rate* were over 90%, which can be seen as proof of the overall high degree of accuracy of our gender detection method. China was an exception, mainly because in the name-gender dictionary, a large share of Chinese first names were unisex in the transliterated Roman alphabet.

For the second database, the *precision rate* and the *recall rate* for the female authors' first-

Table B.1: Validation results of the name-based gender inference process

Country/region	Precision rate		Recall rate	
	Female	**Male**	**Female**	**Male**
Arabic/ Persian	94.91%	95.39%	96.96%	92.89%
Chinese	50.55%	87.60%	50.30%	50.05%
German	98.35%	97.54%	98.06%	98.00%
Japanese	90.49%	91.26%	92.34%	98.72%
Indian	90.49%	80.92%	96.25%	96.22%
Russian	97.73%	96.83%	99.23%	97.75%

names were 0.97 and 0.80, respectively; while the *precision rate* and *recall rate* for the male authors' first-names were 0.98 and 0.86, respectively.

To compare our gender detection method with methods used by other research, we applied another name-gender database generated by a gender disambiguation algorithm (Kozlowski et al., 2022) in detecting the genders of the researchers in our dataset. This reference database included the genders of 1,609,107 distinct US first authors from the Web of Science. The comparison of the results of the two methods shows that the genders of 89.27% of global researchers are identified consistently (Specifically, the genders of 94.23% of migrant researchers and 88.69% of non-moving researchers, respectively, were identified consistently).

B.2 Migration intensity of mobile researchers by gender

The dataset we processed for our analysis on gender disparities in global scholarly migration involved 180, 191, 197, and 202 countries during the periods 1998–2002, 2003–2007, 2008–2012, and 2013–2017, respectively. To make the results more comparable, and to capture the main flows and dynamics of global scholarly migration over time, we estimated the migration intensity of mobile researchers by gender which includes both the absolute number of female (male) mobile researchers and their proportions relative to all female (male) researchers (relative number). Table B.2 showed the results of the 18 major countries, which were the origin and also the destination countries with the largest 50 migration flows in any of the four periods. These migration flows across the 18 countries accounted for 94.97%, 93.57%, 88.69%, and 80.88% of all scholarly migration flows over the four periods, respectively. The Table also indicated that global scholarly migration mainly occurred across a small fraction of countries, which led to the skweness of migration flows.

It is evident that the share of scholars and the size of the science system were much larger in some countries than others (e.g., see the case of the USA, which has hosted around one-third of the share of worldwide scholars in all time periods). The countries with the larger and more established science systems greatly influenced the observed migration trends.

Table B.2: The stock and proportion of migrant researchers by gender. In each period, the countries are sorted based on the number of female researchers in a descending order. The last row is the number of all migrant researchers by gender and the respective proportions of researchers across the globe.

	Period 1 (1998-2002)			
	Female migrants		Male migrants	
Country	Stock	Proportion (%)	Stock	Proportion (%)
United States	10,830	4.95	33,808	7.33
United Kingdom	4,464	8.74	12,475	11.53
Germany	2,645	8.80	9,721	10.54
Canada	2,345	8.62	6,815	12.91
France	2,312	8.16	6,334	11.00
Italy	1,719	4.94	3,530	7.26
Australia	1,454	7.71	4,338	12.31
Japan	1,289	3.34	7,414	3.79
Spain	1,231	4.86	2,646	6.27
China	1,223	3.07	3,863	3.50
Switzerland	1,136	16.76	3,785	19.05
Netherlands	994	9.37	3,339	12.19
Sweden	878	8.17	2,442	11.74
Brazil	652	3.33	1,748	6.43
South Korea	522	5.33	2,152	7.30
Israel	502	7.38	1,622	11.71
India	456	3.52	1,503	5.42
Global	29,161	4.28	91,684	5.96

	Period 2 (2003-2007)			
	Female migrants		Male migrants	
Country	Stock	Proportion (%)	Stock	Proportion (%)
United States	13,585	4.52	35,308	6.52
United Kingdom	5,999	8.49	13,809	10.76
Canada	3,428	8.54	8,440	12.90
Germany	3,348	7.45	10,235	9.26
France	3,110	8.29	7,076	10.44
Australia	2,297	7.96	5,350	12.11
China	2,230	2.19	6,379	2.41
Italy	2,009	4.28	3,538	6.15
Spain	1,676	4.33	3,103	5.50
Switzerland	1,614	15.95	4,475	18.45
Japan	1,511	2.93	7,312	3.30
Netherlands	1,237	7.96	3,079	9.66
Sweden	1,018	7.33	2,336	10.32
South Korea	936	4.70	3,168	6.24
Brazil	882	2.11	1,727	3.62
India	726	3.36	2,269	5.07
Israel	457	5.49	1,284	8.42
Global	41,717	3.96	108,958	5.28

	Period 3 (2008-2012)			
	Female migrants		Male migrants	
Country	Stock	Proportion (%)	Stock	Proportion (%)
United States	17,659	4.52	39,872	6.37
United Kingdom	8,295	9.11	16,179	10.96
Germany	5,050	8.11	12,620	10.10
Canada	4,621	8.43	9,833	12.32
France	4,389	9.02	8,857	11.26
China	4,315	2.05	10,974	2.15
Australia	3,827	9.19	7,490	13.12
Spain	2,786	5.12	4,542	6.36
Switzerland	2,675	17.86	6,558	21.78
Italy	2,460	4.23	4,025	6.23
Netherlands	1,961	8.80	4,196	11.28
Japan	1,912	3.07	7,450	3.17
Sweden	1,609	9.50	3,035	12.18
South Korea	1,362	3.79	4,548	5.63
India	1,326	3.38	4,062	5.33
Brazil	1,229	1.65	2,294	3.12
Israel	536	5.71	1,409	8.92
Global	61,699	4.03	140,594	5.18

	Period 4 (2013-2017)			
	Female migrants		Male migrants	
Country	Stock	Proportion (%)	Stock	Proportion (%)
United States	20,600	5.19	43,242	7.01
United Kingdom	10,456	11.13	18,615	12.70
China	6,969	2.24	18,107	2.53
Germany	6,290	10.29	13,764	11.66
Canada	5,437	9.65	10,111	12.80
France	5,351	10.44	9,914	12.54
Australia	4,879	10.58	8,784	13.99
Switzerland	3,454	20.63	6,896	22.17
Spain	3,192	5.58	4,629	6.35
Italy	3,085	5.21	4,910	7.73
Netherlands	2,619	10.85	4,585	12.36
Sweden	2,161	12.70	3,832	15.28
Japan	2,011	3.59	7,162	3.42
India	2,001	3.98	5,639	5.98
Brazil	1,785	2.18	3,073	3.81
South Korea	1,762	4.12	5,137	5.57
Israel	674	7.68	1,642	10.97
Global	79,370	4.65	167,182	5.66

B.3 Migration distance by gender

We measured the migration distances based on the geographic distance between the capital cities of the origin and destination countries. As Fig. B.1 showed, there was a gap between the average distance of male and female researchers that was consistent in all time periods, and males migrated longer distances. The average migration distances for both female and male researchers kept increasing until 2012. During the most recent period 2013-2017, the average migration distance for male researchers still increased but at a slower rate, while female researchers on average tended to migrate shorter distances compared to the previous period. Accordingly, the gender difference regarding the average migration distance became wider.

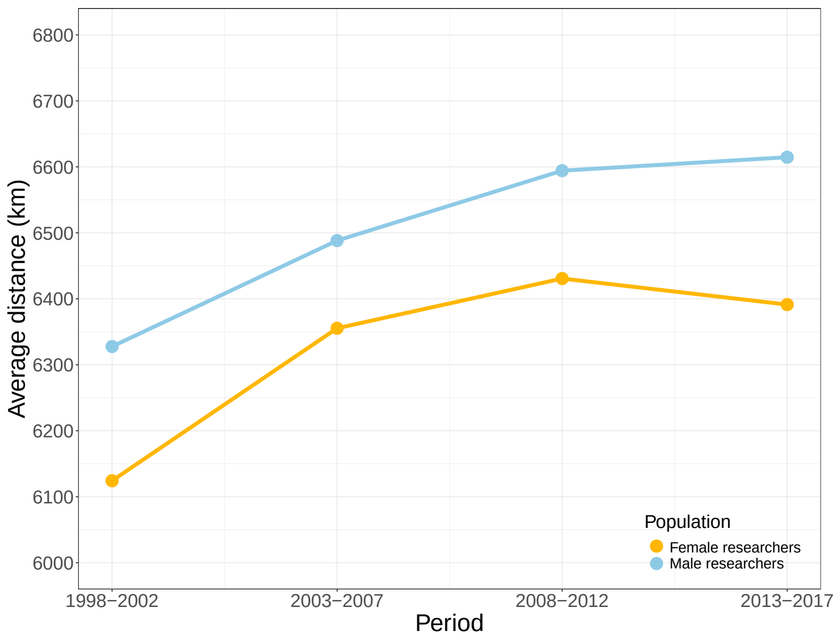


Figure B.1: Average distance between origin and destination of migration among female and male researchers.

We also compared the average migration distances among female and male researchers across six selected countries in Fig. B.2, which are also the countries shown in Fig. 3.2 in our main text. Migrant researchers across Asian countries like China and South Korea usually migrated longer distances compared to researchers in European countries such as Germany. This finding challenged some existing hypotheses that Asian women are more likely to migrate to neighbouring East Asian countries and the Middle East (Caritas Internationalis, 2012). This distinct pattern of migration of female researchers from Asia has been suggested in studies looking at specific groups. For example, South Korean mobile scholars, especially females, have been shown to be geographically concentrated in large metropolitan areas according to in-depth interviews with Korean women who earned doctoral degrees (Yoon and Kim, 2018). The distribution of prime academic destinations played an important role in determining migration distances. That is, the preferred destination countries were clustered in North America and Western Europe, far away from the East Asian countries. Therefore, we attached more attention to the destinations of scholarly migration by gender to dig into the globalization process and compare gender differences. The distribution of top destination countries by migrant researchers was shown in Fig. 3.3 (global level) and Fig. 3.4 (country level) in our main text.

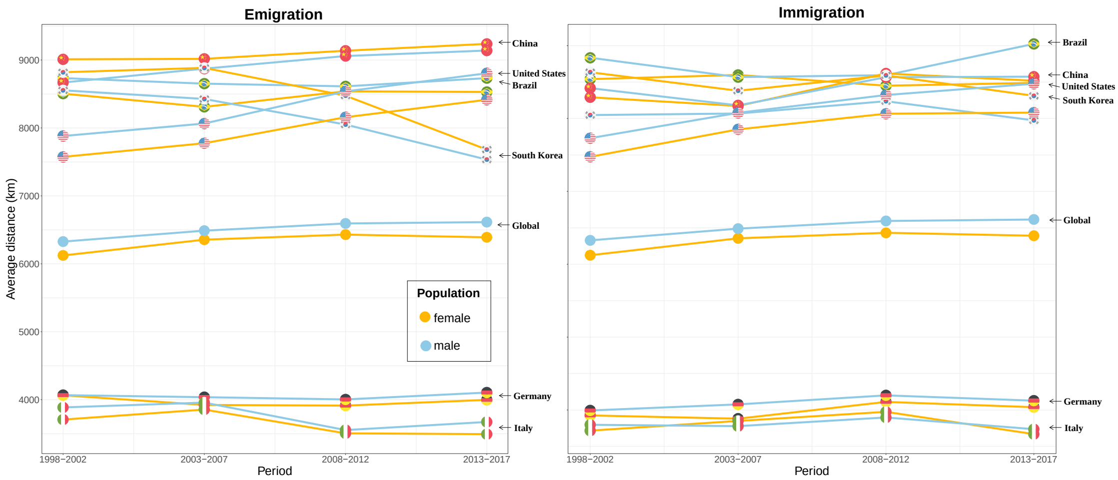


Figure B.2: Average distance among female and male emigrant (left) and immigrant (right) researchers across selected countries (the lines without flags show the global average).

B.4 Global-level migration spread by gender

Using the two measures described in *Methods: Global migration spread*, Fig. B.3 showed that on a global scale, immigration and emigration spreads kept increasing over time, except for a slight decline in the overall (unweighted, lines with squares) emigration spreads and the country-weighted (lines with triangles) immigration spreads during the second period (2003–2007).

The overall (unweighted) migration spreads indicated that mobile researchers of both genders tended to migrate from a relatively small number of countries to a more diverse array of destination countries, especially in the earlier periods. By comparison, in the female migration pattern, there was a larger difference between the emigration and the immigration spreads, which indicated that there was less balance between the outbound and inbound flows among female researchers. That implied that overall, female researchers from a smaller group of countries were able to move internationally, and they were more likely to disperse across a broader range of destination countries.

Compared to the overall global-level migration spreads without weighting, the country-weighted average spread measures indicated that there was less diversification of the average migration spreads with smaller spread values, after incorporating weights that accounted for the size of migration flows across different countries. It showed a more distinct gender pattern, especially among outflows where female mobile researchers were less diverse in their destination countries, which also corresponded to the country-level emigration spreads shown in Fig. 3.2, in the main text. The weighted values, in essence, gave more importance to the major migration countries, which accounted for the largest shares of international migration flows. This indicated that the scholarly inflows and outflows of the major migration countries were less diverse than the overall global scholarly flows, which was in line with the literature showing that a large share of migrant researchers moved from and to only a small number of countries (Macháček et al., 2022; Zhao et al., 2021, 2022a). More interestingly, the figure of country-weighted migration spreads showing more diversification in scholarly inflows relative to scholarly outflows contrasted the figure of the overall migration spreads, especially during the earlier periods. The more diverse range of origin countries from the perspective of specific countries tended to overlap when aggregating, while the pool of destination countries of global outflows was less concentrated. More recently, the emigration and immigration spreads obtained from both methods gradually converged to similar values, which indicated that similar patterns of the diversification of scholarly inflows and outflows were occurring in more countries.

Strikingly, the simultaneously increasing emigration and immigration spreads among mobile researchers can be compared to a migration pattern among the general population by Czaika, and de Haas (2014). Their analysis revealed that the immigration pattern of the general population has become more skewed and concentrated, with average migrants appearing to be concentrated in a smaller pool of fewer selected destination countries (declining emigration spreads), but come from a larger pool of origin countries (increasing immigration spreads). However, our results showed that among scholarly migrants, the pools of both the sending countries and the receiving countries have tended to become more diverse over time. Indeed, the migration patterns of scholars did not have to follow those of the general population. In the past few decades, technological shifts have led to increasing demand and competition for high-skilled workers among a growing number of countries. This trend has facilitated an increase in the diversity of academic destinations for high-skilled migrants, especially for researchers, and has led to more frequent global exchanges.

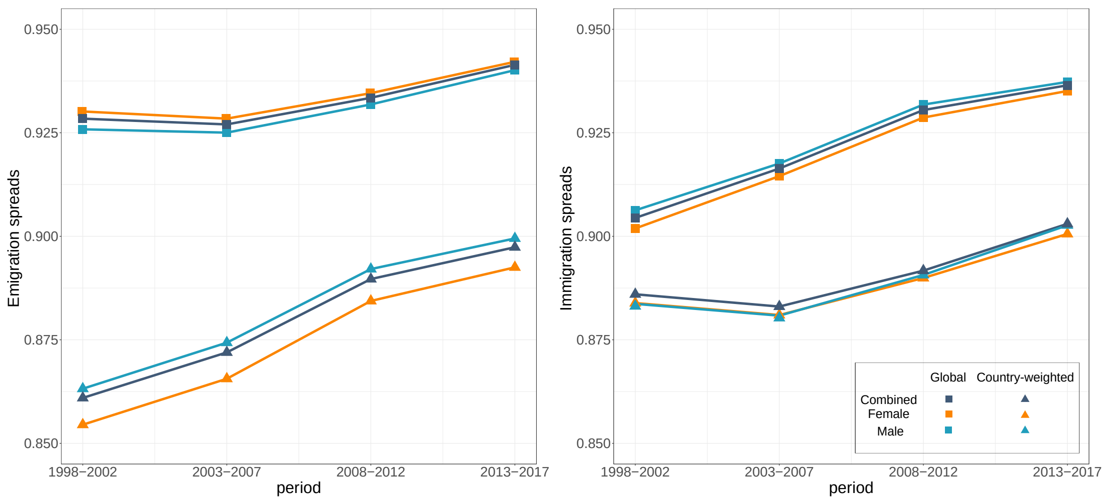


Figure B.3: Global-level emigration (left) and immigration (right) scholarly migration spreads by gender.

B.5 Gender differences in migration distributions

The gender differences in migration distributions point to the possible gender differences in the support provided for or the barriers to leaving a specific country, and also to the benefits or the costs of moving to a particular destination. We calculated the average gender differences among migration inflows and outflows, separately. To a given destination country i, the average gender difference among migration inflows at a specific period of time t is measured using (B.1),

$$Diff\_I_{i,t} = E(|\frac{IM^i_{f,t}}{M_{f,t}} - \frac{IM^i_{m,t}}{M_{m,t}}|/(\frac{IM^i_{f,t}}{M_{f,t}} + \frac{IM^i_{m,t}}{M_{m,t}})) * 100 \quad \text{(B.1)}$$

$IM^i_{f,t}$ and $IM^i_{m,t}$ indicate the immigration flows of female and male researchers to country i from all possible origin countries during the period t. $M_{f,t}$ and $M_{m,t}$ are the number of all migration flows of female and male researchers during the period t, respectively. The ratio of the gender difference $|\frac{IM^i_{f,t}}{M_{f,t}} - \frac{IM^i_{m,t}}{M_{m,t}}|$ to the sum $(\frac{IM^i_{f,t}}{M_{f,t}} + \frac{IM^i_{m,t}}{M_{m,t}})$ reflects the *relative gender difference* along each migration path to country i. Considering the potential disturbance of extreme values measured from the single-gender dominant migration paths, we used the trimmed mean of the middle 70% of the *relative gender differences*. That means that the top 15% and the bottom 15% of the *relative gender differences* will be ignored, and not involved in this measurement. Finally, the trimmed average value $Diff\_I_{i,t}$ indicates the overall gender difference in immigration distribution of country i. If the female and male researchers immigrating to country i share the same distribution of migration paths, $Diff\_I_{i,t}$ will be zero; otherwise, if the two groups follow totally different migration distributions, $Diff\_I_{i,t}$ will reach the maximum value, i.e., one. Similarly, $EM^i_{f,t}$ and $EM^i_{m,t}$ indicate the emigration flows of female and male researchers, respectively, from country i to all possible destination countries during the period t. Accordingly, $Diff\_E_{i,t}$, the overall gender difference in the distribution of migration outflows from country i at time t is calculated as below, ranging from zero to one:

$$Diff\_E_{i,t} = E(|\frac{EM^i_{f,t}}{M_{f,t}} - \frac{EM^i_{m,t}}{M_{m,t}}|/(\frac{EM^i_{f,t}}{M_{f,t}} + \frac{EM^i_{m,t}}{M_{m,t}})) * 100 \quad \text{(B.2)}$$

Fig. B.4 compared the gender differences in the immigration distribution (top half in four panels) and the emigration distribution (bottom half) of mobile researchers across different countries over time. Generally, the gender gaps along the inflow paths of researchers were larger than that along the outflow paths, especially for the first two periods. In particular, the countries such as Russia and Japan consistently had larger average gender differences among immigrants, staying near the top of the ranking. This indicated that for these countries, the trajectories of female and male immigration were less overlapped. By contrast, countries like the US, the UK, and France ranked last in terms of gender gaps in the inflow distributions, with female and male mobile researchers following similar migration paths. Regarding emigration, there were significant differences in the outward trajectories of female and male mobile researchers emigrating from the USA. The gendered distribution of scholarly inflows and outflows, combined with the levels of migration spreads, provided more information about the gender disparities in the national science systems under the current socio-spatial configurations. These insights are critical for understanding the future trends of high-skilled migration by gender.

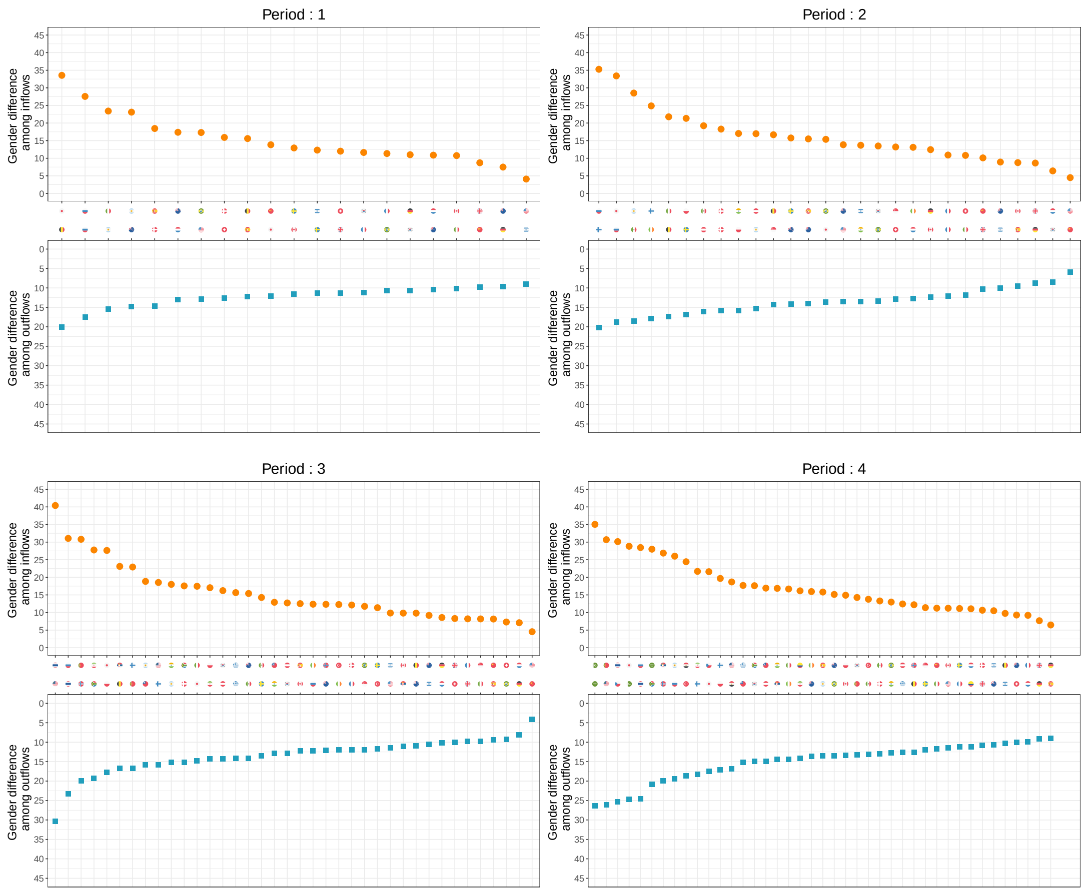


Figure B.4: Gender difference by destination and origin country in four periods by inflows (top part of each panel) and outflows (bottom part of each panel).

B.6 Gender disparities among preferred destinations

To investigate how gender differences in preferences for destinations of scholarly migration varied by country over time, Fig. B.5 showed the gender ratios of the shares of scholarly flows to each destination (Y-axis) from a given country (X-axis) over the four periods. The value in each cell can be calculated by (B.3). Taking the migration from the US (i) to China (j) as an example, in the first period, the proportion of female outgoing flows from the US to China among all female emigration flows from the US was 1.1 times the proportion of male flows from the US to China among all male emigration flows from the USA. This indicated China can be considered a destination for migrations from the US that was preferred slightly more by females than by males. In the latest period, the gender ratio of scholarly mobility along the same path became 0.87, making China a destination that was favored by male mobile researchers from the US. The cells labeled zero indicates there were no detected results for such routes among female researchers in that period.

$$Gender\_OD_{ij} = \frac{EM^{ij}_{f,t}}{EM^{i}_{f,t}} / \frac{EM^{ij}_{m,t}}{EM^{i}_{m,t}} \tag{B.3}$$

The results highlighted that the gendered patterns of scholarly migration tended to be more even and diverse over time, with a considerable share of extremely gendered migration flows turning into slightly single-gender-dominated or gender-balanced flows. It can be observed that overall, the cell colors were more neutral as the cell values approached one. For example, the migration flows from Germany to Italy, which tended to be dominated by female researchers in the earlier periods, became more gender-balanced in the latest period. Nevertheless, some countries continued to attract predominantly female or predominantly male mobile researchers. Italy and Spain were more popular among female researchers throughout the periods, as indicated by the warmer colors in the rows for Italy and Spain. Conversely, the migration flows of researchers to Japan, South Korea, and most Asian countries were consistently dominated by males, regardless of their origins (as indicated by the colder colors in the corresponding rows for these countries). In contrast to the countries that were preferred mainly by either male or female researchers, the US and the UK, as the largest receiving countries and also established scientific destinations for mobile researchers (Zhao et al., 2021, 2022a), had relatively gender-balanced patterns of scholarly migration throughout the study periods.

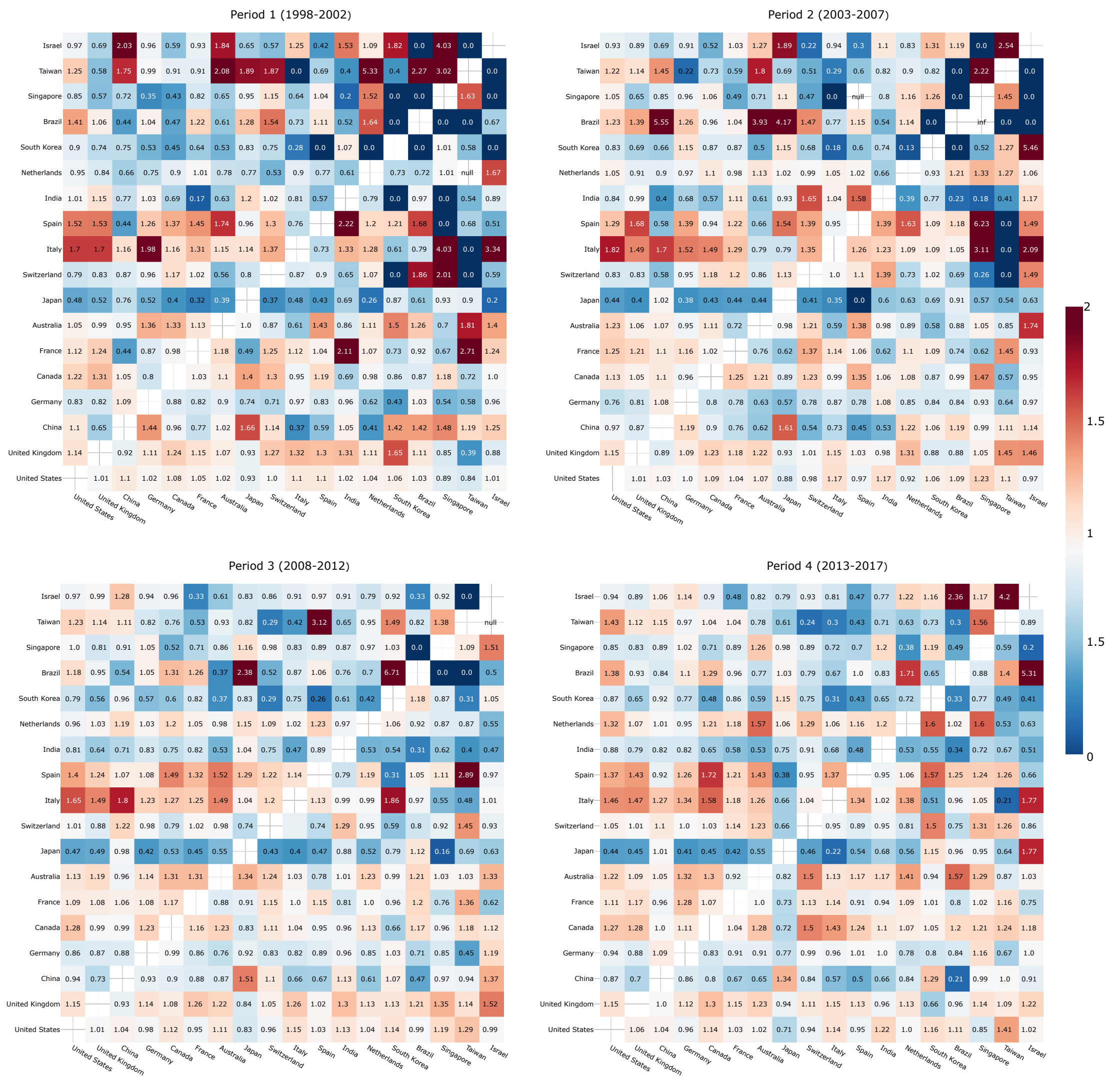


Figure B.5: Gender ratio among migrant researchers by scholarly migration path.

C
Appendix to Chapter 4

C.1 Descriptive statistics of Dataset

Table C.1: The count and percentage of the published researchers and those who received Twitter mentions and self-promoted their first publication by cohort (2012-2016), gender, and discipline.

		Number of newly-published authors		Number of online mentioned authors (% among published authors)			Number of self-promoted authors (% among published authors)		
Cohort	**Discipline**	**Female**	**Male**	**Female**	**Male**	Gender (F-M) difference	**Female**	**Male**	Gender(F-M) difference
2012	Agricultural Sciences	1320	1484	277 (20.98%)	184 (12.40%)	8.58%***	5 (0.38%)	2 (0.13%)	0.25%
	Engineering and Technology	2650	8287	125 (4.72%)	292 (3.52%)	1.20%***	7 (0.26%)	17 (0.21%)	0.05%
	Humanities	1135	1495	194 (17.09%)	233 (15.59%)	1.50%	25 (2.20%)	32 (2.14%)	0.06%
	Medical and Health Sciences	18695	20618	6365 (34.05%)	6033 (29.26%)	4.79%***	144 (0.77%)	188 (0.91%)	-0.14%
	Natural Sciences	19672	38346	3770 (19.16%)	6369 (16.61%)	2.55%***	87 (0.44%)	322 (0.84%)	-0.40%***
	Social Sciences	4391	4754	1355 (30.86%)	1192 (25.07%)	5.79%***	87 (1.98%)	145 (3.05%)	-1.07%***
	No discipline assigned	86	154	22 (25.58%)	35 (22.73%)	2.85%	1 (1.16%)	1 (0.65%)	0.51%
	Total	47949	75138	12108 (25.25%)	14338 (19.08%)	6.17%***	356 (0.74%)	707 (0.94%)	-0.20%***
2013	Agricultural Sciences	1385	1409	317 (22.89%)	253 (17.96%)	4.93%***	8 (0.58%)	11 (0.78%)	-0.20%
	Engineering and Technology	2800	8481	162 (5.79%)	432 (5.09%)	0.70%	7 (0.25%)	20 (0.24%)	0.01%
	Humanities	1126	1522	199 (17.67%)	273 (17.94%)	-0.27%	16 (1.42%)	37 (2.43%)	-1.01%*
	Medical and Health Sciences	18441	20972	7393 (40.09%)	7177 (34.22%)	5.87%***	195 (1.06%)	325 (1.55%)	-0.49%***
	Natural Sciences	19274	37207	4676 (24.26%)	7642 (20.54%)	3.72%***	177 (0.92%)	404 (1.09%)	-0.17%*
	Social Sciences	4083	4586	1336 (32.72%)	1375 (29.98%)	2.74%***	132 (3.23%)	206 (4.49%)	-1.26%***
	No discipline assigned	143	267	29 (20.28%)	32 (11.99%)	8.29%**	0 (0.00%)	3 (1.12%)	-1.12%
	Total	47252	74444	14112 (29.87%)	17184 (23.08%)	6.79%***	535 (1.13%)	1006 (1.35%)	-0.22%***
2014	Agricultural Sciences	1140	1256	301 (26.40%)	283 (22.53%)	3.87%**	9 (0.79%)	7 (0.56%)	0.23%
	Engineering and Technology	2794	8603	209 (7.48%)	486 (5.65%)	1.83%***	6 (0.21%)	30 (0.35%)	-0.14%
	Humanities	1048	1419	229 (21.85%)	293 (20.65%)	1.20%	32 (3.05%)	44 (3.10%)	-0.05%
	Medical and Health Sciences	16707	19907	7550 (45.19%)	7926 (39.82%)	5.37%***	263 (1.57%)	335 (1.68%)	-0.11%
	Natural Sciences	17956	34823	5193 (28.92%)	8732 (25.08%)	3.84%***	206 (1.15%)	469 (1.35%)	-0.20%*
	Social Sciences	3944	4156	1425 (36.13%)	1335 (32.12%)	4.01%***	158 (4.01%)	210 (5.05%)	-1.04%**
	No discipline assigned	95	173	26 (27.37%)	53 (30.64%)	-3.27%	3 (3.16%)	2 (1.16%)	2.00%
	Total	43684	70337	14933 (34.18%)	19108 (27.17%)	7.01 %***	677 (1.55%)	1097 (1.56%)	-0.01%
2015	Agricultural Sciences	1038	1186	306 (29.48%)	292 (24.62%)	4.86 %***	8 (0.77%)	17 (1.43%)	-0.66%
	Engineering and Technology	3020	8562	237 (7.85%)	590 (6.89%)	0.96 %*	8 (0.26%)	30 (0.35%)	-0.09%
	Humanities	1017	1298	239 (23.50%)	317 (24.42%)	-0.92%	46 (4.52%)	66 (5.08%)	-0.56%
	Medical and Health Sciences	15363	18273	7540 (49.08%)	8145 (44.57%)	4.51%***	293 (1.91%)	402 (2.20%)	-0.29%*
	Natural Sciences	16264	31686	5288 (32.51%)	8943 (28.22%)	4.29 %***	241 (1.48%)	492 (1.55%)	-0.07%
	Social Sciences	3760	4108	1469 (39.07%)	1365 (33.23%)	5.84 %***	196 (5.21%)	238 (5.79%)	-0.58%
	No discipline assigned	112	194	38 (33.93%)	62 (31.96%)	1.97%	1 (0.89%)	1 (0.52%)	0.37%
	Total	40574	65307	15117 (37.26%)	19714 (30.19%)	7.07 %***	793 (1.95%)	1246 (1.91%)	0.04%
2016	Agricultural Sciences	1065	1175	335 (31.46%)	278 (23.66%)	7.80 %***	21 (1.97%)	11 (0.94%)	1.03% **
	Engineering and Technology	3063	8645	298 (9.73%)	634 (7.33%)	2.40 %***	11 (0.36%)	43 (0.50%)	-0.14%
	Humanities	992	1276	265 (26.71%)	304 (23.82%)	2.89%	44 (4.44%)	67 (5.25%)	-0.81%
	Medical and Health Sciences	14625	17556	7399 (50.59%)	7956 (45.32%)	5.27 %***	359 (2.45%)	371 (2.11%)	0.34%**
	Natural Sciences	16050	30565	5611 (34.96%)	9192 (30.07%)	4.89 %***	306 (1.91%)	571 (1.87%)	0.04%
	Social Sciences	3390	3632	1413 (41.68%)	1424 (39.21%)	2.47 %**	195 (5.75%)	250 (6.88%)	1.13%*
	No discipline assigned	178	265	69 (38.76%)	92 (34.72%)	4.04%	0 (0.00%)	11 (4.15%)	-4.15%***
	Total	39363	63114	15390 (39.10%)	19880 (31.50%)	7.60 %***	936 (2.38%)	1324 (2.10%)	0.28%***
	Total	**218822**	**348340**	**71660 (32.75%)**	**90224 (25.90%)**	**6.85%***	**3290 (1.51%)**	**5380 (1.54%)**	**-0.03%**

Significance level: < 0.01 '***'; <0.05 '**' ; <0.01 '*' .

C.2 Multilevel ZINB model for measuring online mentions on Twitter

To accommodate the excess zeros and over-dispersion of Twitter mention counts, we used a zero-inflated negative binomial (ZINB) regression to model the online visibility of early-career researchers' first publication and then examine whether it differs by gender.

In addition to the variable of gender, we also considered these control variables: author's cohort (2012 is the reference level), the field of specialty (without a discipline assignment is the reference level), the number of authors in the first publication, the relative publication year since the start of the academic career (1, 2, or 3, and 1 is the reference level), the ranking quantile of the published journal in the subject area (journal rank of Q4 is the reference level), and whether the publication is a product of international collaboration or a collaboration with other universities/institutes (i.e., multiple institutions are involved). The interactions between gender and other control variables are also included in the model. We took a step-wise method by gradually adding control variables and the interaction between the control variables and gender in the base model (*Model 0*) which only includes gender. The full model is further extended to a mixed-effects version by incorporating the random effects for country variability in the intercept and the slope of gender. Thestep-wise modeling results of receving zero mentions and the number of online mentions were shown in Table. C.2 and Table. C.3.

Table C.2: Regression results of the Odds Ratios (*OR*) of *receiving zero mentions* on Twitter. Control variables are gradually added from Model 0 with only gender to Model 6. Model 7 considers the random effects of countries of affiliation of authors based on Model 6.

	Results of ZINB model comparison							
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Predictors	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***
Intercept	0.09***	0.69***	1.44**	1.79***	2.19***	2.16***	2.19***	2.18***
Gender [female]	0.00	0.79	1.05	2.09***	1.85**	1.76**	1.87**	1.50
Discipline [Agri]		0.72	0.67**	0.68**	0.70**	0.69**	0.70**	0.85
Discipline [Eng & Tech]		4.61***	4.54***	4.48***	4.65***	4.24***	4.25***	4.79***
Discipline [Hum]		0.47***	0.39***	0.14***	0.15***	0.13***	0.14***	0.35***
Discipline [Med & Hea]		0.00	0.00	0.09***	0.11***	0.10***	0.10***	0.15***
Discipline [Natur]		0.37***	0.35***	0.41***	0.49***	0.48***	0.48***	0.56***
Discipline [Soci]		0.29***	0.26***	0.19	0.20***	0.17***	0.17***	0.31***
Gender [female]*Discipline [Agri]		0.40***	0.32***	0.41***	0.46***	0.48***	0.47***	0.81
Gender [female]*Discipline [Eng & Tech]		0.93	0.96	0.91	0.86	0.88	0.87	0.89
Gender [female]*Discipline [Hum]		1.12	1.05	0.82	0.86	0.98	0.98	1.22
Gender [female]*Discipline [Med & Hea]		0.457	0.02	0.52***	0.53***	0.56***	0.57***	0.84
Gender [female]*Discipline [Natur]		0.46***	0.54***	0.43***	0.51***	0.65**	0.66**	0.75
Gender [female]*Discipline [Soci]		0.00	0.04**	0.35**	0.39***	0.41***	0.41***	0.68
Cohort [2013]			0.74***	0.74***	0.76***	0.76***	0.76***	0.78***
Cohort [2014]			0.53***	0.53***	0.53***	0.54***	0.54***	0.50***
Cohort [2015]			0.34***	0.34***	0.32***	0.33***	0.34***	0.31***
Cohort [2016]			0.34***	0.36***	0.35***	0.37***	0.36***	0.31***
Gender [female]*Cohort [2013]			0.77**	1.05	0.96	0.88	0.89	0.85**
Gender [female]*Cohort [2014]			0.58***	1.07	0.96	0.84**	0.85**	0.80***
Gender [female]*Cohort [2015]			0.88	1.59***	1.46***	1.27***	1.29***	1.19**
gender [female]*Cohort [2016]			0.70***	1.49***	1.32***	1.12	1.15	1.13
Journal rank [Q1]				0.55***	0.54***	0.58***	0.59***	0.54***
Journal rank [Q2]				1.67***	1.52***	1.54***	1.52***	1
Journal rank [Q3]				6.31***	5.33***	5.17***	5.17***	2.92**
Journal rank [Q4]				29.12***	24.36***	23.67***	23.67***	11.57***
Gender [female]*Journal rank [Q1]				0.26***	0.32*	0.40**	0.39**	0.52***
Gender [female]*Journal rank [Q2]				0.39***	0.46***	0.48***	0.47***	0.58***
Gender [female]*Journal rank [Q3]				0.91	0.94	0.87	0.84	0.88
Gender [female]*Journal rank [Q4]				0.49***	0.54***	0.49***	0.48***	0.86***
Academic age [2]					0.63***	0.67***	0.67***	0.62***
Academic age [3]					0.36***	0.39***	0.39***	0.38***
Gender [female]*Academic age [2]					1.29***	1.25***	1.25***	1.18**
Gender [female]*Academic age [3]					1.23**	1.09	1.09	1.08
Author counts						0.84**	0.85***	0.69***
Gender [female]*Author count						0.93***	0.95**	0.88***
Country-level collaboration [Y]							0.80***	1.13
Institution-level collaboration [Y]							0.97	0.86***
Gender [female]*Country-level collaboration [Y]							0.54***	0.49***
Gender [female]*Institution-level collaboration [Y]							0.87**	0.851**
Random effects								
Dispersion parameter								0.36
τ_{00}								$0.40_{country}$
τ_{11}								$0.18_{country.Gender[female]}$
ρ_{01}								$0.53_{country}$
N								197
Observations	567,162	567,162	567,162	567,162	567,162	567,162	567,162	566,022
Marginal R^2/ Conditional R^2	0.04 / NA	0.07 / NA	0.10 / NA	0.21 / NA	0.21 / NA	0.21 / NA	0.21 / NA	0.21 / 0.12
AIC	1,289,639	1,258,360	1,251,208	1,212,349	1,210,242	1,209,874	1,209,447	1,173,564

Significance level: < 0.01 '***'; <0.05 '**' ; <0.1 '*' .

Table C.3: Regression results of the Incidence Rate Ratios (*IRR*) of Twitter mention counts. Control variables are gradually added from Model 0 with only gender to Model 6. Model 7 considers the random effects of countries of affiliation of authors based on Model 6.

	Results of ZINB model comparison							
	Model 0	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Predictors	***IRR***	***IRR***	***IRR***	***IRR***	***IRR***	***IRR***	***IRR***	***IRR***
Intercept	0.96***	2.79***	1.91***	1.92***	2.04***	2.09***	2.02***	1.58***
Gender [female]	1.20***	1.22	1.30	1.13	1.12	1.08	1.13	1.14
Discipline [Agri]		0.31***	0.33***	0.41***	0.40***	0.39***	0.39***	0.43***
Discipline [Eng & Tech]		0.19***	0.19***	0.21***	0.20***	0.20***	0.20***	0.24***
Discipline [Hum]		0.25***	0.24***	0.26***	0.26***	0.26***	0.26***	0.25***
Discipline [Med & Hea]		0.53***	0.52***	0.63***	0.61***	0.59***	0.59***	0.58***
Discipline [Natur]		0.31***	0.32***	0.33***	0.32***	0.32***	0.32***	0.37***
Discipline [Soci]		0.60***	0.61***	0.66	0.66***	0.64***	0.64***	0.60***
Gender [female]*Discipline [Agri]		0.83	0.77	0.69**	0.70**	0.72	0.71**	0.72*
Gender [female]*Discipline [Eng & Tech]		0.82	0.81	0.75*	0.71**	0.75*	0.73	0.77
Gender [female]*Discipline [Hum]		0.85	0.84	0.80	0.78	0.78	0.77	0.73*
Gender [female]*Discipline [Med & Hea]		0.93	0.90	0.82	0.79	0.83	0.83	0.75
Gender [female]*Discipline [Natur]		0.84	0.81	0.77	0.74	0.80	0.80	0.75*
Gender [female]*Discipline [Soci]		0.67**	0.64***	0.68**	0.66***	0.65***	0.64***	0.627**
Cohort [2013]			1.31***	1.28***	1.28***	1.28***	1.27***	1.32***
Cohort [2014]			1.59***	1.55***	1.53***	1.52***	1.52***	1.59***
Cohort [2015]			1.59***	1.54***	1.50***	1.49***	1.49***	1.60***
Cohort [2016]			1.90***	−1.02***	−0.62***	0.62***	0.60***	1.89***
Gender [female]*Cohort [2013]			0.92***	0.96	0.94**	0.92***	0.93***	0.90***
Gender [female]*Cohort [2014]			0.87***	0.93***	0.92***	0.89***	0.89***	0.89***
Gender [female]*Cohort [2015]			1.09***	1.18***	1.18***	1.15***	1.16***	1.13***
gender [female]*Cohort [2016]			1.05**	1.13***	1.11***	1.08***	1.10***	1.08***
Journal rank [Q1]				1.30***	1.26***	1.28***	1.28***	1.22**
Journal rank [Q2]				0.63***	0.61***	0.62***	0.61***	1.21***
Journal rank [Q3]				0.54***	0.52***	0.52***	0.52***	1.42***
Journal rank [Q4]				0.66***	0.63***	0.62***	0.63***	0.83*
Gender [female]*Journal rank [Q1]				1.11***	1.11*	1.13**	1.11**	1.22***
Gender [female]*Journal rank [Q2]				1.06***	1.08	1.08	1.08	1.21***
Gender [female]*Journal rank [Q3]				1.52***	1.47***	1.35***	1.35***	1.42***
Gender [female]*Journal rank [Q4]				0.70***	0.71***	0.64***	0.64***	0.83*
Academic age [2]					1.04***	1.04***	1.05***	1.11***
Academic age [3]					0.13***	1.14***	1.14***	1.20***
Gender [female]*Academic age [2]					1.18***	1.19***	1.19***	1.14***
Gender [female]*Academic age [3]					1.16***	1.16***	1.16***	1.15***
Author counts						0.99**	0.98***	1.05***
Gender [female]*Author count						0.94***	0.95***	0.95***
Country-level collaboration [Y]							1.22***	1.02
Institution-level collaboration [Y]							1.12***	1.19***
Gender [female]*Country-level collaboration [Y]							0.83***	0.88**
Gender [female]*Institution-level collaboration [Y]							0.91***	0.90***
Random effects								
Dispersion parameter								0.36
τ_{00}								$0.21_{country}$
τ_{11}								$0.06_{country.Gender[female]}$
ρ_{01}								$-0.37_{country}$
N								197
Observations	567,162	567,162	567,162	567,162	567,162	567,162	567,162	566,022
Marginal R^2/ Conditional R^2	0.04 / NA	0.07 / NA	0.10 / NA	0.21 / NA	0.21 / NA	0.21 / NA	0.21 / NA	0.21 / 0.12
AIC	1,289,639	1,258,360	1,251,208	1,212,349	1,210,242	1,209,874	1,209,447	1,173,564

Significance level: < 0.01 '***'; <0.05 '**' ; <0.1 '*' .

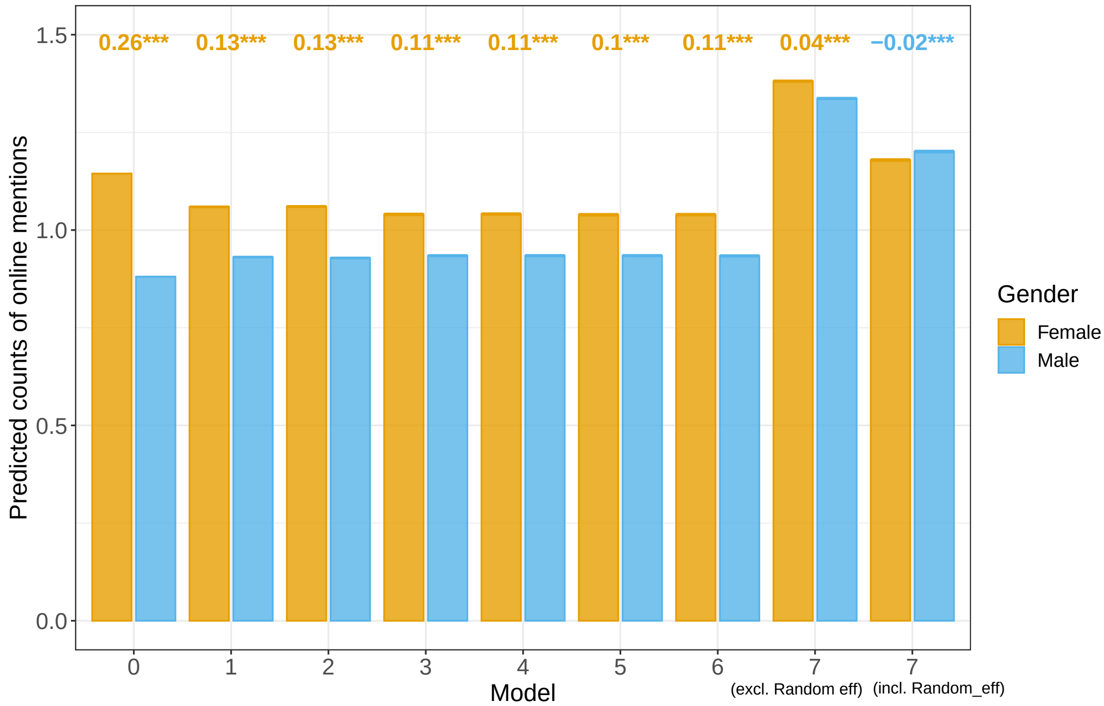


Figure C.1: Predicted number of Twitter mentions on early-career female and male researchers' first publication by model 0–7. Model 7 includes the predicted Twitter counts without and with random effects. The numbers at the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.1 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women.

C.3 Country-level heterogeneity in the gender differences of online dissemination

Fig. C.2 showed the heterogeneity in the gender differences by country. We found that male researchers in the countries with the largest number of researchers, such as China, the US, the UK, and Australia, receive higher online visibility on Twitter than females. It contributes to the overall pattern of online visibility disfavoring female researchers in terms of tweet counts.

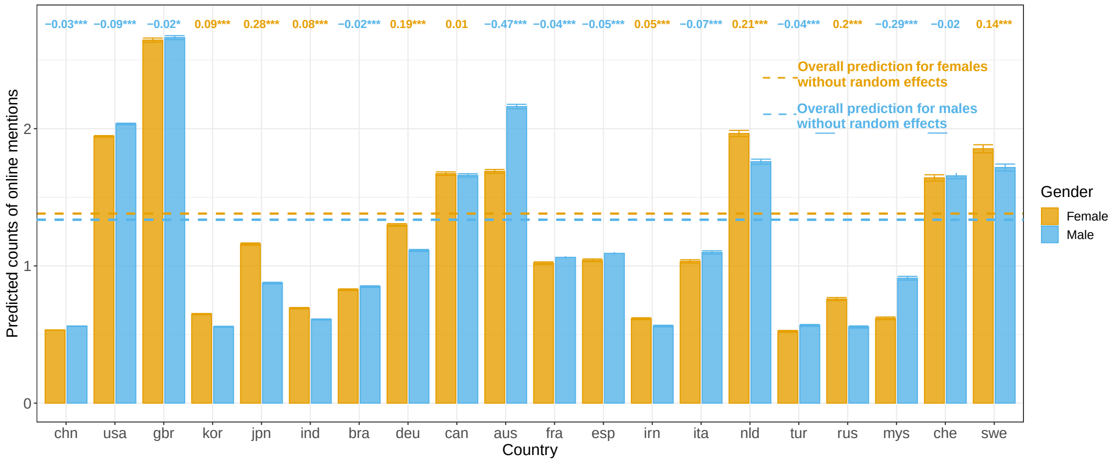


Figure C.2: Predicted number of Twitter mentions on early-career female and male researchers' first publication by affiliated country. The selected countries are the top 20 countries with the highest number of early-career researchers from the 2012–2016 cohort, in descending order of the number of researchers. The dashed lines in the plot show the overall predicted number of Twitter mentions on early-career female and male researchers' first publications, not conditioned on random-effect variances. The numbers at the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.1 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women (chn: China; usa: The U.S.; gbr: The U.k.; kor: South Korea; jpn: Japan; ind: India; bra: Brazil; deu: Germany; can: Canada; fra: France; esp: Spain; irn: Iran; ita: Italy; nld: Netherland; tur: Turkey; rus: Russia; mys: Malaysia; che: Switzerland; swe: Sweden).

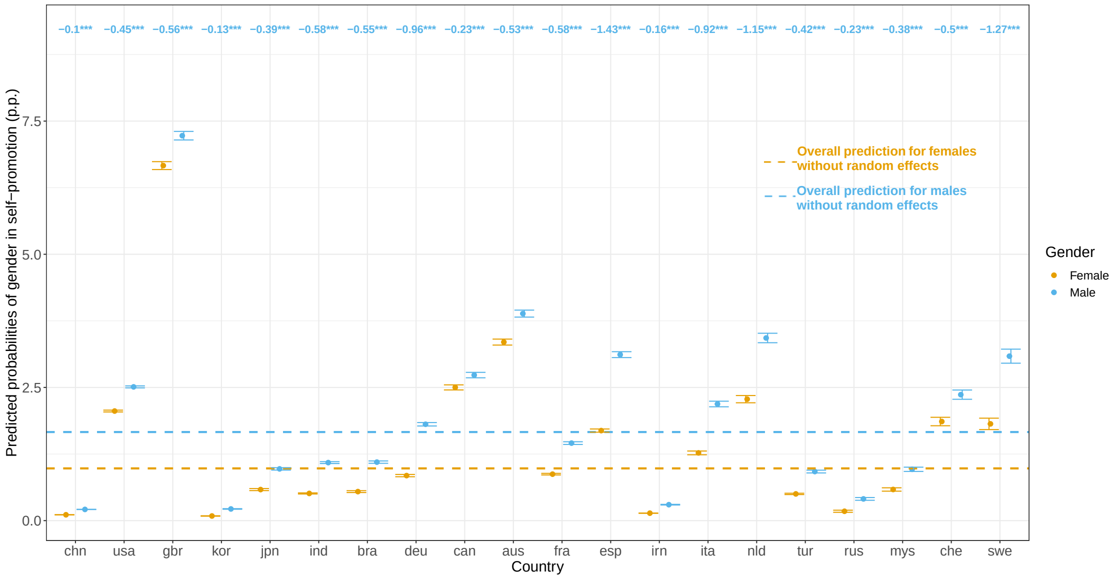


Figure C.3: Predicted probabilities of early-career female and male researchers self-promoting their first publication. The selected countries are the top 20 countries with the highest number of early-career researchers from the 2012–2016 cohort, in descending order of the number of researchers. The dashed lines in the plot show the overall predicted probabilities of self-promotion among early-career female and male researchers, not conditioned on random-effect variances. The numbers at the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.1 '*'). Negative values (printed in blue text) correspond to a gender gap disfavouring women (chn: China; usa: The U.S.; gbr: The U.k.; kor: South Korea; jpn: Japan; ind: India; bra: Brazil; deu: Germany; can: Canada; fra: France; esp: Spain; irn: Iran; ita: Italy; nld: Netherland; tur: Turkey; rus: Russia; mys: Malaysia; che: Switzerland; swe: Sweden).

C.4 Logistic model for measuring self-promotion on Twitter

We also employed the step-wise method by gradually adding control variables in the base model (*Model 0*) which only includes gender. The results of step-wise modeling of self-promotion were shown in Table. C.4.

Table C.4: Logistic regression results of the Odds Ratios (OR) of self-promotion on Twitter. Control variables are gradually added from Model 0 with only gender to Model 7. Model 8 considers the random effects of countries of affiliation of authors based on Model 7.

	Results of logistic model comparison								
	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Predictors	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***	***OR***
Intercept	0.02***	0.02***	0.01**	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
Gender [female]	0.98	0.47	0.36**	0.34**	0.34**	0.33**	0.33**	0.27**	0.25**
Discipline [Agri]		0.43***	0.45***	0.47***	0.46***	0.46***	0.46***	0.56**	0.47**
Discipline [Eng & Tech]		0.19***	0.19***	0.19***	0.19***	0.17***	0.17***	0.32***	0.31***
Discipline [Hum]		2.09***	2.20***	2.40***	2.42***	1.74**	1,76**	2.22***	1.39
Discipline [Med & Hea]		0.97	1.02	1.05	1.01	1.07	1.07	0.96	0.71
Discipline [Natur]		0.76	0.80	0.74	0.71	0.69	0.69	0.87	0.79
Discipline [Soci]		2.99***	3.16***	3.32***	3.29***	2.54***	2.55***	2.65***	1.89**
Gender [female]*Discipline [Agri]		2.47*	2.60*	2.56*	2.54*	2.58*	2.59*	3.55**	2.94*
Gender [female]*Discipline [Eng & Tech]		1.75	1.79	1.85	1.84	1.90	1.89	2.43	2.46
Gender [female]*Discipline [Hum]		1.84	1.92	1.94	1.93	1.92	1.91	2.77*	2.33
Gender [female]*Discipline [Med & Hea]		1.90	2.01	2.02	2.01	2.01	2.01	2.91*	2.29
Gender [female]*Discipline [Natur]		1.84	1.93	1.99	1.97	2.05	2.04	2.85*	2.54
Gender [female]*Discipline [Soci]		1.67	1.75	1.78	1.77	1.78	1.78	2.56	2.03
Cohort [2013]			1.45***	1.47***	1.47***	1.48***	1.48***	1.35***	1.39***
Cohort [2014]			1.70***	1.73***	1.72***	1.76***	1.75***	1.49***	1.61***
Cohort [2015]			2.08***	2.16***	2.14***	2.21***	2.20***	1.80***	2.02***
Cohort [2016]			2.33***	2.44***	2.41***	2.52***	2.51***	1.95***	2.24***
Gender [female]*Cohort [2013]			1.07	1.07	1.07	1.06	1.07	1.12	1.11
Gender [female]*Cohort [2014]			1.25***	1.26***	1.26***	1.25***	1.26***	1.34***	1.33**
Gender [female]*Cohort [2015]			1.30***	1.30***	1.30***	1.29***	1.30***	1.36**	1.36**
gender [female]*Cohort [2016]			1.44***	1.44***	1.44***	1.43***	1.43***	1.55***	1.55**
Journal rank [Q1]				1.42***	1.40***	1.50***	1.50***	1.22**	1.17
Journal rank [Q2]				0.71***	0.70***	0.70***	0.72***	0.78	0.84**
Journal rank [Q3]				0.42***	0.42***	0.42***	0.42***	0.55**	0.63**
Journal rank [Q4]				0.28***	0.28***	0.28***	0.28***	0.42***	0.54***
Gender [female]*Journal rank [Q1]				1.02	1.02	1.04	1.03	0.97	0.90
Gender [female]*Journal rank [Q2]				1.14	1.15	1.15	1.15	1.05	1.00
Gender [female]*Journal rank [Q3]				1.03	1.03	1.03	1.03	0.94	0.96
Gender [female]*Journal rank [Q4]				0.73	0.73	0.73	0.72	0.65*	0.69
Academic age [2]					1.05	1.15***	1.15***	1.09**	
Academic age [3]					1.35***	1.51***	1.51***	1.37***	
Gender [female]*Academic age [2]					0.96	0.96	0.96	0.97	0.99
Gender [female]*Academic age [3]					1.06	1.06	1.06	1.07	1.09
Author counts						0.77**	0.76***	0.74***	0.86***
Gender [female]*Author count						0.98	0.98	0.99	0.98
Country-level collaboration [Y]							1.29***	1.25***	0.87**
Institution-level collaboration [Y]							1.14***	1.10**	1.28***
Gender [female]*Country-level collaboration [Y]							1.06	1.09	1.18
Gender [female]*Institution-level collaboration [Y]							0.89*	0.90	0.88*
Tweeted by others [Y]								3.65***	2.88***
Counts of others' tweets								1.36**	1.31***
Gender [female]*Tweeted by others [Y]								0.75***	0.70***
Gender [female]*Counts of others' tweets								1.03	1.04**
Random effects									
τ_{00}									$0.42_{country}$
τ_{11}									$0.10_{country.Gender[female]}$
ρ_{01}									$0.45_{country}$
N									197
Observations	567,162	567,162	567,162	567,162	567,162	567,162	567,162	567,162	566,022
Marginal R^2/ Conditional R^2	0.00 / NA	0.10 / NA	0.13 / NA	0.19 / NA	0.19 / NA	0.20 / NA	0.21 / NA	0.26 / NA	0.29 / 0.18
AIC	89,762	86,978	86,046	84,293	84,192	83,741	83,693	78,911	74,390

Significance level: < 0.01 '***'; <0.05 '**' ; <0.1 '*' .

C.5 Propensity score estimation (PSM) for matching researchers

For (*Matching 1*), the $Outcome_i$ presents the propensity (i.e., probability) of online exposure on Twitter for each researcher i, from the pool of all early-career researchers. We considered gender, academic cohort, discipline, the journal rank of their first publication, affiliated country, the number of authors, and whether the publication is a product of international collaboration as the covariates ($cov_{k,i}$) to generate the propensity score for each researcher. The researchers with Twitter mentions were then paired with the ones without mentions but sharing similar propensity scores using the method of *nearest neighbor matching*. Fig. C.4 and the result of the balance test in Table C.5 showed the distributions of baseline variables are similar between the treatment and control groups in *Matching 1*.

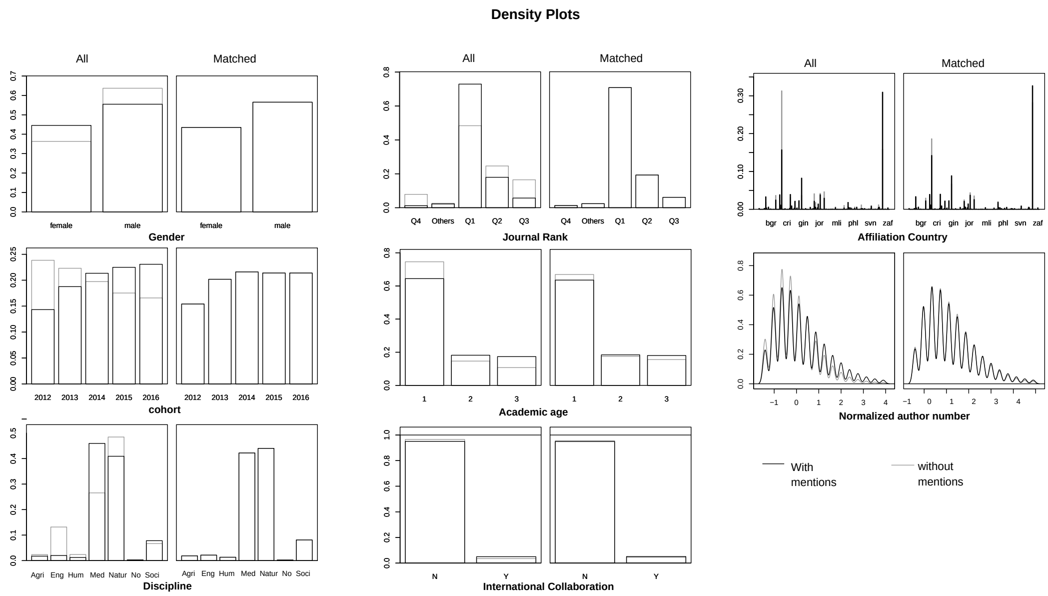


Figure C.4: Density plot of the distributions of covariables in the treatment groups (with Twitter mentions) and control groups (without Twitter mentions)

Table C.5: Balance test for *Matching 1* (Treatment group: with Twitter mentions; Control group: without Twitter mentions)

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
Gender							
Female	0.4349	0.4349	0.0000	NA	0.0000	0.0000	0.0000
Male	0.5651	0.5651	0.0000	NA	0.0000	0.0000	0.0000
Academic age							
1	0.6354	0.6693	-0.0708	NA	0.0339	0.0339	0.4435
2	0.1844	0.1750	0.0242	NA	0.0093	0.0093	0.4725
3	0.1802	0.1557	0.0648	NA	0.0246	0.0246	0.3940
Cohort							
2012	0.1542	0.1542	0.0000	NA	0.0000	0.0000	0.0000
2013	0.2018	0.2018	0.0000	NA	0.0000	0.0000	0.0000
2014	0.2161	0.2161	0.0000	NA	0.0000	0.0000	0.0000
2015	0.2140	0.2140	0.0000	NA	0.0000	0.0000	0.0000
2016	0.2140	0.2140	0.0000	NA	0.0000	0.0000	0.0000
Discipline							
Agricultural Sciences	0.0189	0.0189	0.0000	NA	0.0000	0.0000	0.0000
Engineering and Technology	0.0217	0.0217	0.0000	NA	0.0000	0.0000	0.0000
Humanities	0.0133	0.0133	0.0000	NA	0.0000	0.0000	0.0000
Medical and Health Sciences	0.4223	0.4223	0.0000	NA	0.0000	0.0000	0.0000
Natural Sciences	0.4401	0.4401	0.0000	NA	0.0000	0.0000	0.0000
No discipline assigned	0.0025	0.0025	0.0000	NA	0.0000	0.0000	0.0000
Social Sciences	0.0812	0.0812	0.0000	NA	0.0000	0.0000	0.0000
Journal rank							
Q1	0.7085	0.7085	0.0000	NA	0.0000	0.0000	0.0000
Q2	0.1931	0.1931	0.0000	NA	0.0000	0.0000	0.0000
Q3	0.0614	0.0614	0.0000	NA	0.0000	0.0000	0.0000
Q4	0.0130	0.0130	0.0000	NA	0.0000	0.0000	0.0000
Others	0.0240	0.0240	0.0000	NA	0.0000	0.0000	0.0000
Country							
aus	0.0337	0.0314	0.0130	NA	0.0023	0.0023	0.2259
bra	0.0225	0.0297	-0.0463	NA	0.0072	0.0072	0.1251
can	0.0396	0.0355	0.0210	NA	0.0040	0.0040	0.2366
che	0.0120	0.0105	0.0142	NA	0.0015	0.0015	0.1539
chn	0.1424	0.1865	-0.1211	NA	0.0441	0.0441	0.1882
deu	0.0398	0.0402	-0.0017	NA	0.0003	0.0003	0.2202
esp	0.0218	0.0224	-0.0044	NA	0.0006	0.0006	0.1731
fra	0.0222	0.0237	-0.0099	NA	0.0015	0.0015	0.1730
gbr	0.0888	0.0671	0.0787	NA	0.0217	0.0217	0.3261
ind	0.0212	0.0251	-0.0267	NA	0.0039	0.0039	0.0825
irn	0.0051	0.0064	-0.0169	NA	0.0012	0.0012	0.0452
ita	0.0139	0.0155	-0.0133	NA	0.0016	0.0016	0.1472
jpn	0.0375	0.0439	-0.0334	NA	0.0064	0.0064	0.1758
kor	0.0259	0.0355	-0.0569	NA	0.0096	0.0096	0.1228
mys	0.0034	0.0039	-0.0087	NA	0.0005	0.0005	0.0510
nld	0.0193	0.0154	0.0294	NA	0.0039	0.0039	0.1926
rus	0.0023	0.0025	-0.0042	NA	0.0002	0.0002	0.0385
swe	0.0090	0.0084	0.0059	NA	0.0006	0.0006	0.1234
tur	0.0049	0.0062	-0.0191	NA	0.0014	0.0014	0.0571
usa	0.3268	0.2824	0.0961	NA	0.0444	0.0444	0.4753
Author number	0.1876	0.1461	0.0370	1.0866	0.0073	0.0143	0.5863

Only the top 20 countries are shown in the table.

For (*Matching 2*), the $Outcome_i$ represents the propensity (i.e., probability) of self-promotion from the pool of the early-career researchers with Twitter mentions on their first publications. The number of total Twitter mentions each researcher obtained on their first publications was taken into account in matching process, to find the pair of researchers with similar characteristics (i.e., propensity score) but differences in self-promotion behaviors. Fig. C.5 and the result of balance test in Table C.6 showed the distributions of baseline variables are similar between the treatment and control groups in *Matching 2*.

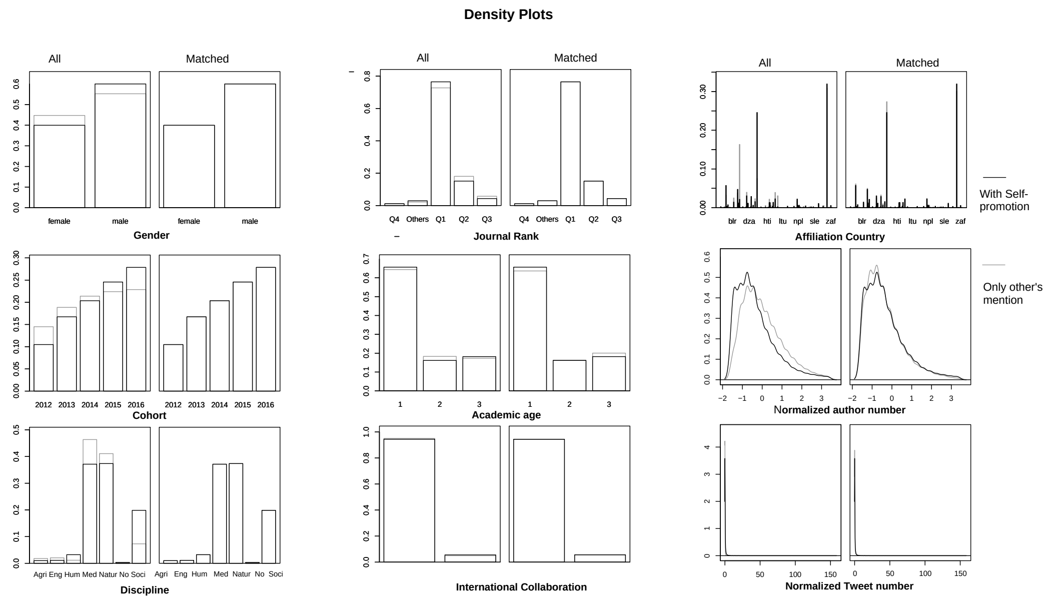


Figure C.5: Density Plot of the distributions of covariables in the treatment groups (with self-promotions) and control groups (with only others' promotions)

Table C.6: Balance test for *Matching 2* (Treatment group: with self-promotion; Control group: without self-promotion)

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
Gender							
Female	0.3998	0.3998	0.0000	NA	0.0000	0.0000	0.0000
Male	0.6002	0.6002	0.0000	NA	0.0000	0.0000	0.0000
Academic age							
1	0.6566	0.6364	0.0425	NA	0.0202	0.0202	0.4670
2	0.1615	0.1631	-0.0044	NA	0.0016	0.0016	0.3836
3	0.1819	0.2005	-0.0482	NA	0.0186	0.0186	0.4225
Cohort							
2012	0.1048	0.1048	0.0000	NA	0.0000	0.0000	0.0000
2013	0.1673	0.1673	0.0000	NA	0.0000	0.0000	0.0000
2014	0.2035	0.2035	0.0000	NA	0.0000	0.0000	0.0000
2015	0.2457	0.2457	0.0000	NA	0.0000	0.0000	0.0000
2016	0.2786	0.2786	0.0000	NA	0.0000	0.0000	0.0000
Discipline							
Agricultural Sciences	0.0100	0.0100	0.0000	NA	0.0000	0.0000	0.0000
Engineering and Technology	0.0111	0.0111	0.0000	NA	0.0000	0.0000	0.0000
Humanities	0.0323	0.0323	0.0000	NA	0.0000	0.0000	0.0000
Medical and Health Sciences	0.3713	0.3713	0.0000	NA	0.0000	0.0000	0.0000
Natural Sciences	0.3736	0.3736	0.0000	NA	0.0000	0.0000	0.0000
No_discipline_assigned	0.0032	0.0032	0.0000	NA	0.0000	0.0000	0.0000
Social Sciences	0.1984	0.1984	0.0000	NA	0.0000	0.0000	0.0000
Journal rank							
Q1	0.7647	0.7647	0.0000	NA	0.0000	0.0000	0.0000
Q2	0.1512	0.1512	0.0000	NA	0.0000	0.0000	0.0000
Q3	0.0433	0.0433	0.0000	NA	0.0000	0.0000	0.0000
Q4	0.0113	0.0113	0.0000	NA	0.0000	0.0000	0.0000
Others	0.0296	0.0296	0.0000	NA	0.0000	0.0000	0.0000
Country							
aus	0.0573	0.0614	-0.0174	NA	0.0040	0.0040	0.2105
bra	0.0144	0.0118	0.0217	NA	0.0026	0.0026	0.1248
can	0.0473	0.0499	-0.0122	NA	0.0026	0.0026	0.2191
che	0.0123	0.0110	0.0117	NA	0.0013	0.0013	0.1320
chn	0.0215	0.0212	0.0022	NA	0.0003	0.0003	0.0334
deu	0.0305	0.0246	0.0347	NA	0.0060	0.0060	0.1606
esp	0.0289	0.0328	-0.0231	NA	0.0039	0.0039	0.2101
fra	0.0157	0.0134	0.0182	NA	0.0023	0.0023	0.1119
gbr	0.2458	0.2738	-0.0649	NA	0.0279	0.0279	0.2149
ind	0.0131	0.0111	0.0171	NA	0.0019	0.0019	0.1052
irn	0.0008	0.0002	0.0227	NA	0.0006	0.0006	0.0227
ita	0.0136	0.0132	0.0028	NA	0.0003	0.0003	0.1340
jpn	0.0228	0.0213	0.0097	NA	0.0015	0.0015	0.1223
kor	0.0024	0.0018	0.0131	NA	0.0006	0.0006	0.0263
mys	0.0034	0.0024	0.0167	NA	0.0010	0.0010	0.0833
nld	0.0223	0.0241	-0.0120	NA	0.0018	0.0018	0.1937
rus	0.0013	0.0005	0.0225	NA	0.0008	0.0008	0.0495
swe	0.0108	0.0121	-0.0125	NA	0.0013	0.0013	0.1686
tur	0.0032	0.0018	0.0256	NA	0.0015	0.0015	0.0882
usa	0.3197	0.3008	0.0405	NA	0.0189	0.0189	0.2920
Tweet number	0.5606	0.3000	0.0705	1.8791	0.0148	0.2192	0.1346
Author number	-0.3321	-0.3650	0.0338	1.1306	0.0095	0.0226	0.5435

Only the top 20 countries are shown in the table.

C.6 Robustness analysis of the impact of online visibility

We made a robustness analysis on the marginal effects of Twitter mentions on citations by limiting the treatment group to receiving only one mention. This leads to 61,671 pairs (123,342 individuals) of researchers. The results shown in Fig. C.6, further confirmed that female researchers generally gain higher citation counts when mentioned on Twitter.

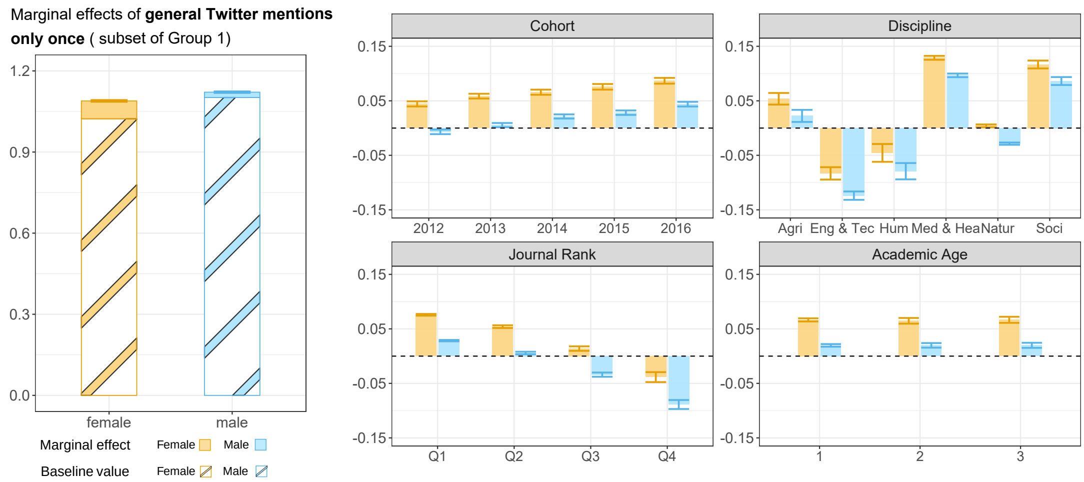


Figure C.6: Marginal effects of Twitter mention (**only once**) in the 5-year cumulative *discipline-normalized citation scores* (DNCS) among early-career female and male researchers (Matching 1), at the overall level (left panel) and disaggregated by cohort, discipline, and the journal rank (right panel).

D

Appendix to Chapter 5

Table D.1: Summary statistics of the respondent researchers by gender and birth cohort

		before 1960		**1960-1964**		**1965-1969**		**1970-1974**		**1975-1978**	
		Female	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
Ph.D. recipient	Y	521	683	499	466	813	634	1228	895	1177	740
		(92.05%)	(95.13%)	(93.45%)	(94.30%)	(93.34%)	(95.48%)	(94.46%)	(96.24%)	(93.64%)	(96.35%)
	N	45	35	35	23	58	30	72	35	80	28
		(7.95%)	(4.87%)	(6.55%)	(4.70%)	(6.66%)	(4.52%)	(5.54%)	(3.76%)	(6.36%)	(3.65%)
Number of Children	1	130	123	135	104	240	144	353	200	367	177
		(22.97%)	(17.13%)	(25.28%)	(21.27%)	(27.55%)	(21.69%)	(27.15%)	(21.51%)	(29.20%)	(23.05%)
	2	307	378	293	240	488	359	737	487	687	404
		(54.24%)	(52.65%)	(54.87%)	(49.08%)	(56.03%)	(54.07%)	(56.69%)	(52.37%)	(54.65%)	(52.60%)
	3+	129	217	106	145	143	161	210	243	203	187
		(22.79%)	(30.22%)	(19.85%)	(29.65%)	(16.42%)	(24.25%)	(16.15%)	(26.13%)	(16.15%)	(24.35%)
Research Field	Arts and Humanities	58	37	54	32	73	49	101	71	73	49
		(22.79%)	(30.22%)	(19.85%)	(29.65%)	(16.42%)	(24.25%)	(16.15%)	(26.13%)	(16.15%)	(24.35%)
	Health Sciences	214	186	170	127	287	145	393	217	420	206
		(22.79%)	(30.22%)	(19.85%)	(29.65%)	(16.42%)	(24.25%)	(16.15%)	(26.13%)	(16.15%)	(24.35%)
	Natural Sciences	118	304	138	229	242	280	294	361	341	313
		(22.79%)	(30.22%)	(19.85%)	(29.65%)	(16.42%)	(24.25%)	(16.15%)	(26.13%)	(16.15%)	(24.35%)
	Social Sciences	176	191	172	101	269	190	512	281	423	200
		1(22.79%)	(30.22%)	(19.85%)	(29.65%)	(16.42%)	(24.25%)	(16.15%)	(26.13%)	(16.15%)	(24.35%)
Region	US	331	387	254	198	391	256	570	366	568	313
		(55.30%)	(53.90%)	(47.57%)	(40.49%)	(44.89%)	(38.55%)	(43.85%)	(39.35%)	(45.19%)	(40.76%)
	UK	52	65	54	49	73	88	100	81	92	48
		(9.19%)	(9.05%)	(10.11%)	(10.02%)	(8.38%)	(13.25%)	(7.69%)	(8.71%)	(7.32%)	(6.25%)
	Canada	32	19	28	30	62	34	101	54	91	44
		(5.65%)	(2.65%)	(5.24%)	(6.14%)	(7.12%)	(5.12%)	(7.77%)	(5.81%)	(7.24%)	(5.73%)
	Western Europe	33	75	46	47	75	85	131	115	124	105
		(5.83%)	(10.45%)	(8.61%)	(9.61%)	(8.61%)	(12.80%)	(10.08%)	(12.37%)	(9.86%)	(13.67%)
	Northern Europe	28	35	25	28	42	33	61	48	62	52
		(4.95%)	(4.87%)	(4.68%)	(5.73%)	(4.82%)	(4.97%)	(4.69%)	(5.16%)	(4.93%)	(6.77%)
	Southern Europe	23	31	29	42	67	43	106	89	68	69
		(4.06%)	(4.32%)	(5.43%)	(8.59%)	(7.69%)	(6.48%)	(8.15%)	(9.57%)	(5.41%)	(8.98%)
	Eastern Europe	17	20	21	11	27	20	44	28	61	38
		(3.00%)	(2.79%)	(3.93%)	(2.25%)	(3.10%)	(3.01%)	(3.38%)	(3.01%)	(4.85%)	(4.95%)
	Oceania	23	27	30	26	40	29	58	33	46	31
		(4.06%)	(3.76%)	(5.62%)	(5.32%)	(4.59%)	(4.37%)	(4.46%)	(3.55%)	(3.66%)	(4.04%)
	Asia	22	36	23	37	53	42	77	63	90	45
		(3.89%)	(5.01%)	4.31%)	(7.57%)	(6.09%)	(6.33%)	(5.92%)	(6.77%)	(7.16%)	(5.86%)
	Latin America	15	12	16	13	26	22	33	39	36	17
		(2.65%)	(1.67%)	(3.00%)	(2.66%)	(2.99%)	(3.31%)	(2.54%)	(4.19%)	(2.86%)	(2.21%)
	Africa	8	11	8	8	15	12	19	14	19	6
		(1.41%)	(1.53%)	(1.50%)	(1.64%)	(1.72%)	(1.81%)	(1.46%)	(1.51%)	(1.51%)	(0.78%)
Partner in academia	Y	152	184	169	136	257	187	360	282	322	210
		(26.86%)	(25.63%)	(31.65%)	(27.81%)	(29.51%)	(28.61%)	(27.69%)	(30.32%)	(25.62%)	(27.34%)
	N	327	477	284	325	515	444	824	601	861	532
		(57.77%)	(66.43%)	(53.18%)	(66.46%)	(59.13%)	(66.87%)	(63.38%)	(64.62%)	(68.50%)	(69.27%)
	Missing	87	57	81	28	99	33	116	47	74	26
		(15.37%)	(7.94%)	(15.17%)	(5.73%)	(11.37%)	(4.97%)	(8.92%)	(5.05%)	(5.89%)	(3.39%)
	Total	**566**	**718**	**534**	**489**	**871**	**664**	**1300**	**930**	**1257**	**768**

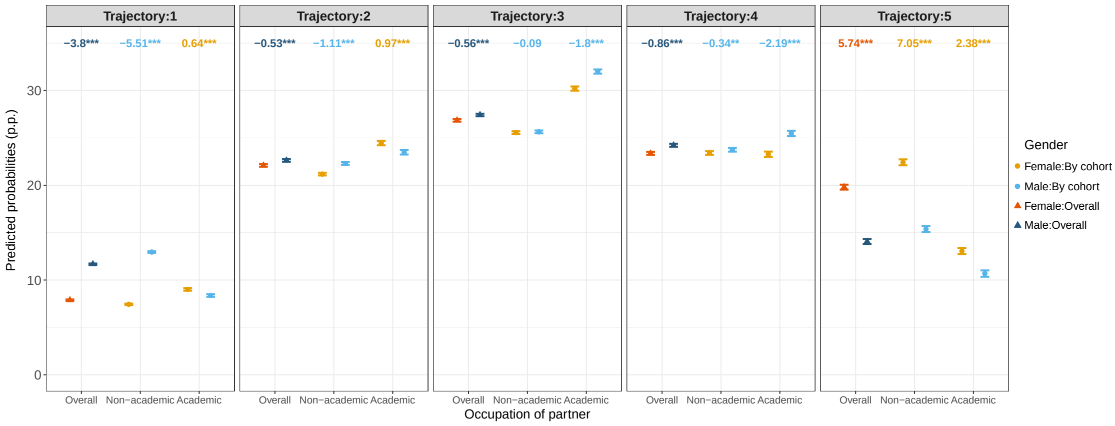


Figure D.1: Predicted probability of each trajectory by gender across the full sample and disaggregated by the occupation of partner (Non-academic/Academic). The numbers on the top show the marginal effects of gender on the probabilities in each trajectory with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavoring women.

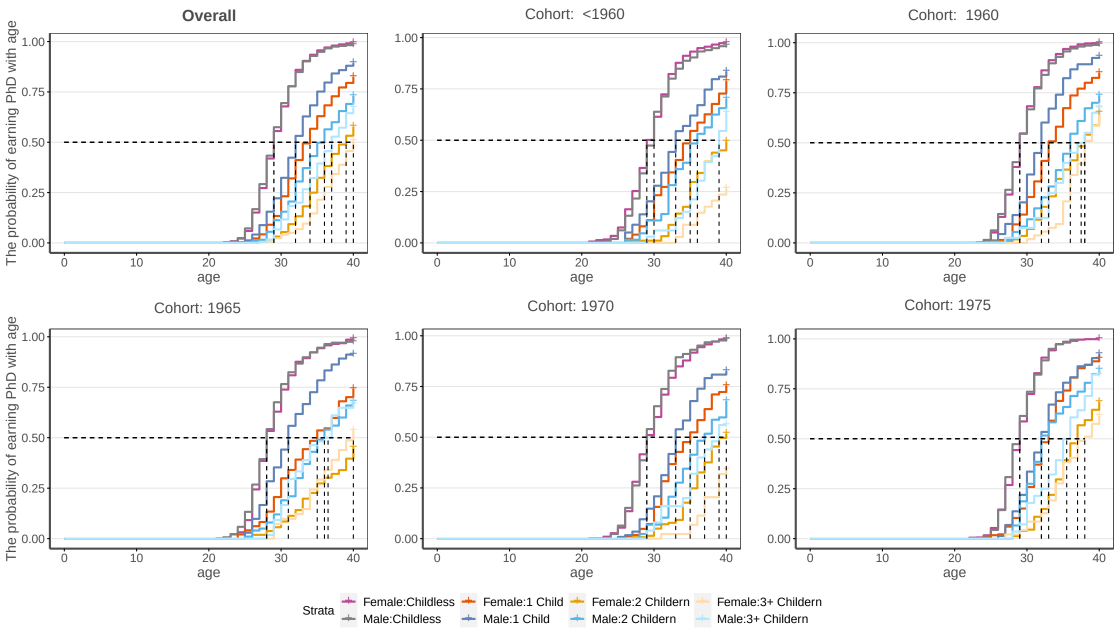


Figure D.2: Survival curves of time to obtain a Ph.D. for researchers separated by the interaction of gender and the number of children (0,1,2, and 3+) across cohort

Table D.2: Cox regression results of obtaining a Ph.D. among researchers. Control variables are gradually added from Model 0 with only gender to Model 5.

	Cox Model 0	Cox Model 1	Cox Model 2	Cox Model 3	Cox Model 4	Cox Model 5
Predictors	*Parameter estimates*					
Gender (Ref: Male)						
Female	0.83***	1.01	0.89	0.87	0.88	0.84
Number of Children (Ref: 0)						
1		0.41***	0.41***	0.40***	0.41***	0.41***
2		0.22***	0.21***	0.22***	0.21***	0.21***
3+		0.18***	0.18***	0.18***	0.18***	0.18***
Female × 1		0.41***	0.41***	0.40***	0.41***	0.41***
Female × 2		0.22***	0.21***	0.22***	0.21***	0.21***
Female × 3+		0.18***	0.18***	0.18***	0.18***	0.18***
Cohort (Ref: <1960)						
1960			0.71***	0.73***	0.73***	0.73***
1965			0.78***	0.81***	0.83***	0.83***
1970			0.91*	0.96	1	1
1975			1.12**	1.19***	1.19***	1.21***
Female × 1960			1.19*	1.17*	1.18*	1.18*
Female × 1965			1.07	1.04	1.02	1.02
Female × 1970			1.19**	1.15*	1.14*	1.15*
Female × 1975			1.12	1.08	1.06	1.07
Region (Ref: Eastern Europe)						
Africa				0.66**	0.66**	0.66**
Asia				0.89	0.91	0.91
Canada				0.86	0.94	0.94
Latin America				0.69***	0.68***	0.68***
Northern Europe				0.96	1.03	1.04
Oceania				0.81	0.84	0.84
Southern Europe				0.92	0.90	0.90
UK				1.38***	1.48***	1.48***
US				1.25**	1.37***	1.37***
Western Europe				1.01	1.06	1.06
Female × Africa				1.37	1.47	1.50
Female × Asia				1.24	1.29	1.27
Female × Canada				1.09	1.03	1.02
Female × Female × Latin America				1.07	1.08	1.07
Female × Northern Europe				1.13	1.10	1.10
Female × Oceania				1.15	1.16	1.18
Female × Southern Europe				1.14	1.16	1.15
Female × UK				0.79	0.78	0.79
Female × US				0.98	0.95	0.94
Female × Western Europe				1.16	1.15	1.14
Discipline (Ref: Natural Sciences)						
Arts and Humanities					0.69***	0.69***
Health Sciences					0.92*	0.92*
Social Sciences					0.66***	0.66***
Female × Arts and Humanities					1.20*	1.20*
Female × Health Sciences					1.05	1.06
Female × Social Sciences					1.07	1.08
Partner Occupation (Ref: Non-academics)						
Academics						1.01
Female × Academics						1.14**
R^2	0.007	0.350	0.363	0.377	0.392	0.393

Significance level: < 0.01 '***'; <0.05 '**' ; <0.1 '*' .

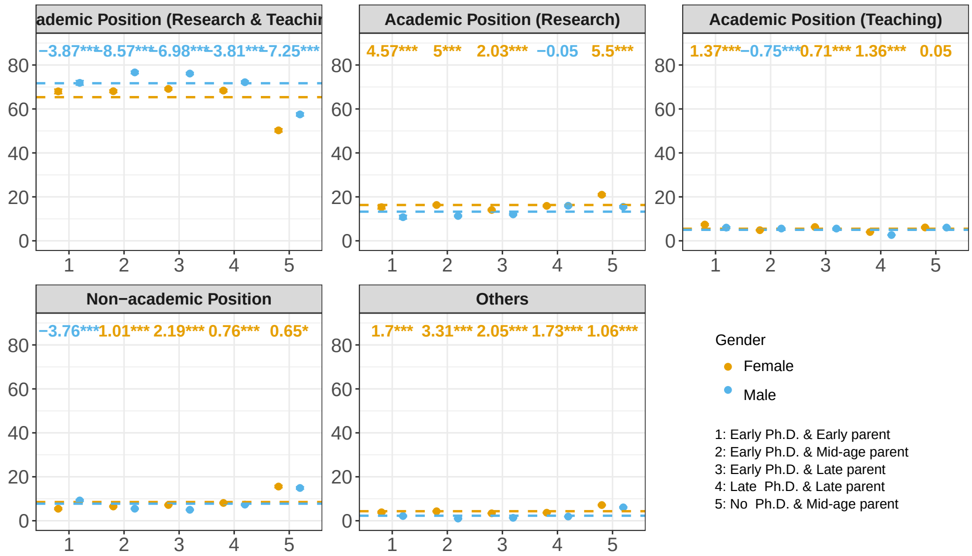


Figure D.3: Predicted probabilities of employment sectors across trajectories by gender. The numbers on the top show the marginal effects of gender (reference is male) with statistical significance (Significance level: < 0.01 '***'; <0.05 '**'; <0.01 '*'). Negative values (printed in blue text) correspond to a gender gap disfavoring women. The horizontal lines with annotated values on the right mean the average predicted probabilities in each employment sector across trajectories by gender

Bibiography

Abramitzky, R., Boustan, L., & Eriksson, K. (2019). To the new world and back again: Return migrants in the age of mass migration. ILR Review, *72*(2): 300–322. http://doi.org/10.1177/0019793917726981

Ackers, L., & Gill, B. (2005). Attracting and retaining 'early career' researchers in english higher education institutions. Innovation: The European Journal of Social Science Research, *18*(3): 277–299. http://doi.org/10.1080/13511610500186649

Adsera, A., Nikolka, T., & Poutvaara, P. (2022). Cognitive abilities, gender, and migration aspirations. European Population Conference 2022.

Ahmed, W., Bath, P., & Demartini, G. (2017). Chapter 4 using twitter as a data source: An overview of ethical, legal, and methodological challenges. In K. Woodfield (Ed.), *The ethics of online research (advances in research ethics and integrity, vol. 2)* (pp. 79–107). Emerald Publishing Limited. https://doi.org/10.1108/S2398-601820180000002004

Akbaritabar, A., & Squazzoni, F. (2021). Gender patterns of publication in top sociological journals. Science, Technology, & Human Values, *46*(3): 555–576. https://doi.org/10.1177/0162243920941588

Alfasoft. (2024). Exploring bibliometric methods: Citation analysis in research. https://alfasoft.com/blog/products/scientific-writing-and-publishing/exploring-bibliometric-methods-citation-analysis-in-research/

Altenburger, K., De, R., Frazier, K., Avteniev, N., & Hamilton, J. (2017). Are there gender differences in professional self-promotion? an empirical case study of LinkedIn profiles among recent MBA graduates. Proceedings of the International AAAI Conference on Web and Social Media, *11*(1): 460–463. https://doi.org/10.1609/icwsm.v11i1.14929

Altmertic. (2023). X (formerly twitter). https://help.altmetric.com/support/solutions/articles/6000235926-twitter/

Aman, V. (2018a). Does the Scopus author ID suffice to track scientific international mobility? A case study based on Leibniz laureates. Scientometrics, *117*(2): 705–720. https://doi.org/10.1007/s11192-018-2895-3

Aman, V. (2018b). Transfer of formal knowledge through international scientific mobility-introduction of a network-based bibliometric method. STI 2018 Conference Proceedings: 545–552. http://doi.org/10.1162/qss_a_00028

Andall, J. (2013). Gendered mobilities and work in europe: An introduction. Journal of Ethnic and Migration Studies, *39*: 525–534. https://doi.org/10.1080/1369183X.2013.745229

Anders, S. M. v. (2004). Why the academic pipeline leaks: Fewer men than women perceive barriers to becoming professors. Sex Roles, *51*(9–10): 511–521. http://dx.doi.org/10.1007/s11199-004-5461-9

Andrey, L., & Elena, A. (2019). Bibliometric analysis of publications from postsoviet countries in psychological journals in 1992–2017. Scientometrics, *119*: 1157–1171. https://doi.org/10.1007/s11192-019-03087-y

Appelt, S., van Beuzekom, B., Galindo-Rueda, F., & de Pinho, R. (2015). Chapter 7 - which factors influence the international mobility of research scientists? In A. Geuna (Ed.), *Global mobility of research scientists*. Academic Press. http://doi.org/10.1016/B978-0-12-801396-0.00007-7

Aref, S., Zagheni, E., & West, J. (2019). The demography of the peripatetic researcher: Evidence on highly mobile scholars from the Web of Science. International Conference on Social Informatics: 50–65. https://doi.org10.1007/978-3-030-34971-4_4

Auschra, C., Bartosch, J., & Lohmeyer, N. (2022). Differences in female representation in leading management and organization journals: Establishing a benchmark. Research Policy, *51*. https://doi.org/10.1016/j.respol.2021.104410

Azur, M., Stuart, E., Frangakis, C., & Leaf, P. (2011). Multiple imputation by chained equations: What is it and how does it work? International Journal of Methods in Psychiatric Research, *20*(1): 40–49. https://doi.org/10.1002/mpr.329

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quantitative Science Studies, *1*(1): 377–386. https://doi.org/10.1162/qss_a_00019

Bahous, S. (2022). Lack of gender equality in science is holding back solutions. https://www.springernature.com/gp/advancing-discovery/springboard/blog/blogposts-sustainability-inclusion/gender-equality-science-iwd/20200002

Bailey, A., & Mulder, C. (2017). Highly skilled migration between the global north and south: Gender, life courses and institutions. Journal of Ethnic and Migration Studies, *43*: 2689–2703. https://doi.org/10.1080/1369183X.2017.1314594

Bank, T. W. (2021). World Bank Country and Lending Groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups/

Bapna, S., & Funk, R. (2021). Interventions for Improving Professional Networking for Women: Experimental Evidence from the IT Sector. Management Information Systems Quarterly, *45*(2): 593–636.

Bardin, A. (2016). The Problem of Migration in German Academic Discourse. Political Studies, (6): 183–188. https://10.17976/jpps/2016.06.13

Baruffaldi, S. H., & Landoni, P. (2012). Return mobility and scientific productivity of researchers working abroad: The role of home country linkages. Research Policy, *41*(9): 1655–1665. https://doi.org/10.1016/j.respol.2012.04.005

Bauder, H. (2015). The international mobility of academics: A labour market perspective. International Migration, *53*(1): 83–96. https://doi.org/10.1111/j.1468-2435.2012.00783.x

BibliometriX. (2023). Bibliographic databases supported by bibliometrix and biblioshiny: Characteristics and difference. https://www.bibliometrix.org/home/index.php/blog/134-bibliographic-databases-supported-by-bibliometrix-and-biblioshiny-characteristics-and-differences

Bilecen, B., & Van Mol, C. (2013). Introduction: International academic mobility and inequalities. journal of ethnic and migration studies. Journal of Ethnic and

Migration Studies, *43*(8): 1241–1255. https://doi.org/10.1080/1369183X.2017.1300225
Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. Quantitative Science Studies, *1*(1): 363–376. http://dx.doi.org/10.1162/qss_a_00018
Björk, B., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. Journal of Informetrics, *7*(4): 914–923. https://doi.org/10.1016/j.joi.2013.09.001
Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, *55*(4): 77–84. https://doi.org/10.1145/2133806.2133826
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, *3*: 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993
Boekhout, H. D., Traag, V. A., & Takes, F. W. (2021). Investigating scientific mobility in co-authorship networks using multilayer temporal motifs. Network Science, *9*(3): 354–386. http://dx.doi.org/10.1017/nws.2021.12
Bornmann, L., & Haunschild, R. (2018). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data (P. Dorta-González, Ed.). PLOS ONE, *13*(5): e0197133. https://doi.org/10.1371/journal.pone.0197133
Brandth, B., Eydal, G. B., Lammi-Taskula, J., Gíslason, I. V., Duvander, A.-Z., & Rostgaard, T. (2012). *Parental leave, childcare and gender equality in the nordic countries*. Nordic Council of Ministers. http://dx.doi.org/10.6027/tn2011-562
Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2023). How effective are female role models in steering girls towards stem? Evidence from french high schools. The Economic Journal, *133*(653): 1773–1809.
Broderick, N. A., & Casadevall, A. (2019). Gender inequalities among authors who contributed equally. eLife, *8*.
Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. Annual Review of Sociology, *34*(1): 319–337.
Burnham, J. F. (2006). Scopus database: A review. Biomedical Digital Libraries, *3*(1): 1. https://doi.org/10.1186/1742-5581-3-1
Cao, C., Baas, J., Wagner, C., & Jonkers, K. (2020). Returning scientists and the emergence of China's science system, science and public policy. Science and Public Policy, *47*(2): 172–183. https://doi.org/10.1093/scipol/scz056
Cardoso, R. (2022). PhD parenting: The good, the bad, and the unspoken. https://researcherblogs.ki.se/2022/05/29/phd-parenting-the-good-the-bad-and-the-unspoken/
Caritas Internationalis. (2012). *The female face of migration* (tech. rep.). http://www.caritas.org/includes/pdf/backgroundmigration.pdf
Chen, K.-h., Tang, M.-c., Wang, C.-m., & Hsiang, J. (2014). Exploring alternative metrics of scholarly performance in the social sciences and humanities in Taiwan. Scientometrics, *102*(1): 97–112. https://doi.org/10.1007/s11192-014-1420-6
Chou, M. (2020). Sticky and slippery destinations for academic mobility: The case of singapore. Higher Education, *82*: 749–764. https://doi.org/10.1007/s10734-020-00574-7
Christian, K. (2018). Gender inequality in the sciences: Why is it still with us? http://blogs.nature.com/naturejobs/2018/04/09/gender-inequality-in-the-sciences-why-is-it-still-with-us/

Cole, J., & Zuckerman, H. (1984). The productivity puzzle. Advances in Motivation and Achievement: 217–258.

Collins, K., Shiffman, D., & Rock, J. (2016). How are scientists using social media in the workplace? (S. Goffredo, Ed.). PLOS ONE, *11*(10): e0162680. https://doi.org/10.1371/journal.pone.0162680

Conchi, S., & Michels, C. (2014). *Scientific mobility: An analysis of Germany, Austria, France and Great Britain* (Working Paper). Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis. https://www.econstor.eu/handle/10419/94371

Corona-Sobrino, C., Garcıa-Melón, M., Poveda-Bautista, R., & González-Urango, H. (2020). Closing the gender gap at academic conferences: A tool for monitoring and assessing academic events. PLoS One, *15*(12): e0243549.

Costas, R., Mongeon, P., Ferreira, M. R., Honk, J. v., & Franssen, T. (2020). Large-scale identification and characterization of scholars on twitter. Quantitative Science Studies, *1*(2): 1–21. https://doi.org/10.1162/qss_a_00047

Costas, R., van Honk, J., & Franssen, T. (2017). Scholars on twitter: Who and how many are they? arXiv preprint arXiv:1712.05667. https://arxiv.org/abs/1712.05667

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. Journal of the Association for Information Science and Technology, *66*(10): 2003–2019. https://doi.org/10.1002/asi.23309

Czaika, M. (2018). *High-skilled migration: Drivers and policies.* Oxford University Press.

Czaika, M., & de Haas, H. (2014). The globalization of migration: Has the world become more migratory? International Migration Review, *48*(2): 283–323. https://doi.org/10.1111/imre.12095

Czaika, M., & Orazbayev, S. (2018). The globalisation of scientific mobility, 1970-2014. Applied Geography, *96*: 1–10. https://doi.org/10.1016/j.apgeog.2018.04.017

Czaika, M., & Parsons, C. R. (2017). The gravity of high-skilled migration policies. Demography, *54*(2): 603–630. https://doi.org/10.1007/s13524-017-0559-1

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. Social Network Analysis and Mining, *9*(1): 24. https://doi.org/10.1007/s13278-019-0568-8

D'Angelo, C. A., & van Eck, N. J. (2020). Collecting large-scale publication data at the level of individual researchers: A practical proposal for author name disambiguation. Scientometrics, *123*: 883–907. https://doi.org/10.1007/s11192-020-03410-y

Débarre, F., Rode, N. O., & Ugelvig, L. V. (2018). Gender equity at scientific events. Evolution Letters, *2*(3): 148–158.

Dehdarirad, T. (2020). Could early tweet counts predict later citation counts? A gender study in life sciences and biomedicine (2014–2016) (A. Abbasi, Ed.). PLOS ONE, *15*(11): e0241723. https://doi.org/10.1371/journal.pone.0241723

Derrick, G. E., Chen, P.-Y., Leeuwen, T. v., Larivière, V., & Sugimoto, C. R. (2022). The relationship between parenting engagement and academic performance. Scientific Reports, *12*(1). https://doi.org/10.1038/s41598-022-26258-z

Desai, T. A., Eniola-Adefeso, O., Stevens, K. R., Vazquez, M., & Imoukhuede, P. (2021). Perspectives on disparities in scientific visibility. Nature Reviews Materials, *6*(7): 556–559. https://doi.org/10.1038/s41578-021-00329-5

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: 4171–4186.

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. Political Analysis, *26*(3): 312–327. https://doi.org/10.1017/pan.2018.12

Directorate-General for Research and Innovation. (2021a). *Gender gap at education level is shrinking, but women are still under-represented in research and innovation* (tech. rep.). European Commission.

Directorate-General for Research and Innovation. (2021b). *She figures 2021: Gender in research and innovation: Statistics and indicators* (tech. rep.). European Commission.

Docquier, F., Lowell, B. L., & Marfouk, A. (2009). A gendered assessment of highly skilled emigration. Population and Development Review, *35*(2): 297–321. https://doi.org/10.1111/j.1728-4457.2009.00277.x

Docquier, F., & Marfouk, A. (2004, August 19). *Measuring the international mobility of skilled workers (1990–2000): Release 1.0*. The World Bank. Retrieved December 10, 2020, from https://doi.org/10.1596/1813-9450-3381

Donato, K., & Gabaccia, D. (2015). *Gender and international migration: From the slavery era to the global age*. Russell Sage Foundation. https://www.journals.uchicago.edu/doi/10.1086/691431

Donato, K., & Gabaccia, D. (2016). *The global feminization of migration: Past, present, and future* (tech. rep.). Migration Policy Institute. https://www.migrationpolicy.org/article/global-feminization-migration-past-present-and-future

Donner, P. (2021). Citation analysis of Ph.D. theses with data from Scopus and Google Books. Scientometrics, *126*(12): 9431–9456. http://dx.doi.org/10.1007/s11192-021-04173-w

Dorling, D., Barford, A., & Newman, M. (2006). Worldmapper: The world as you've never seen it before. IEEE Transactions on Visualization and Computer Graphics, *12*(5): 757–764. https://doi.org/10.1109/TVCG.2006.202

Drake, I., & Svenkerud, S. W. (2023). Career ambitions of women academics. are women willing *and* able to rise to the top in higher education institutions? Studies in Higher Education: 1–12. https://doi.org/10.1080/03075079.2023.2272742

Drivas, K., Economidou, C., Karamanis, D., & Sanders, M. (2018). *Mobility of highly skilled individuals and local innovation activity* (Working Paper). Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/88883/

Dumont, J.-C., Martin, J. P., & Spielvogel, G. (2007). *Women on the move: The neglected gender dimension of the brain drain (ssrn scholarly paper id 1001216)* (tech. rep.). Social Science Research Network. https://doi.org/10.2139/ssrn.1001216

Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: Quality measurements of scientific publication. Radiology, *255*(2): 342–351. http://dx.doi.org/10.1148/radiol.09090626

Düvell, F. (2019). *Germany: Selected Migration Policies* (tech. rep. SEO Report No. 2019-24). Dutch Ministry of Foreign Affairs.

El-Ouahi, J., Nicolas, R.-G., & Rodrigo, C. (2021). Analyzing scientific mobility and collaboration in the Middle East and North Africa. Quantitative Science Studies: 1–30. https://doi.org/10.1007/s11192-019-03087-y

Elsevier. (2020a). *The researcher journey through a gender lens* (tech. rep.). https://doi.org/10.17632/ww6g4t2r32.2

Elsevier. (2020b). Scopus content coverage guide. https://www.elsevier.com/solutions/scopus/how-scopus-works/content

Elsevier. (2020c). *Scopus content coverage guide* (tech. rep.). https://www.elsevier.com/solutions/scopus/how-scopus-works/content

Encinas-Martín, M., & Cherian, M. (2023). *Gender, education and skills: The persistence of gender gaps in education and skills*. OECD.

Eule, T. G. (2016). *Inside Immigration Law: Migration Management and Policy Application in Germany*. Routledge. https://doi.org/10.4324/9781315588728

European Commission & Directorate-General for Research and Innovation. (2018). *More3 study: Support data collection and analysis concerning mobility patterns and career paths of researchers*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/710643

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. Journal of Medical Internet Research, *13*(4): e123. https://doi.org/10.2196/jmir.2012

Falco, V., Cuntrera, D., & Attanasio, M. (2023). Gender differences in career advancements in italian universities over the last 20 years. Genus, *79*(1).

Fedasiuk, R. (2020). *The China Scholarship Council: An Overview* (tech. rep.). Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/the-china-scholarship-council-an-overview/

Fernández-Zubieta, A., Geuna, A., & Lawson, C. (2015). Chapter 1 - What do we know of the mobility of research scientists and impact on scientific production. In A. Geuna (Ed.), *Global mobility of research scientists* (pp. 1–33). Academic Press. https://doi.org/10.1016/B978-0-12-801396-0.00001-6

Findlay, A. (2010). An assessment of supply and demand-side theorizations of international student mobility. International Migration, *49*(2): 162–190. https://doi.org/10.1111/j.1468-2435.2010.00643.x

Fiske, E. B. (2012). *World atlas of gender equality in education*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000215522

Fortin, J., Bartlett, B., Michael, Kantar, M. T., & Mehrabi, Z. (2021). Digital technology helps remove gender bias in academia. Scientometrics, *126*(5): 4073–4081. https://doi.org/10.1007/s11192-021-03911-4

Franzoni, C., Scellato, G., & Stephan, P. (2014). The mover's advantage: The superior performance of migrant scientists. Economics Letters, *122*(1): 89–93. https://doi.org/10.1016/j.econlet.2013.10.040

Franzoni, C., Scellato, G., & Stephan, P. (2015). Chapter 2 - International mobility of research scientists: Lessons from GlobSci. In A. Geuna (Ed.), *Global mobility of research scientists* (pp. 35–65). Academic Press. https://doi.org/10.1016/B978-0-12-801396-0.00002-8

Fresnoza-Flot, A. (2022). *Gender gaps in migration studies: Recent developments and prospects*. Brill Nijhoff.

Gabaccia, D. R. (2016). Feminization of migration. In *The wiley blackwell encyclopedia of gender and sexuality studies* (pp. 1–3). John Wiley; Sons, Ltd. https://doi.org/10.1002/9781118663219.wbegss732

Gabadinho, A., Ritschard, G., S. Müller, N., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. Journal of Statistical Software, *40*(4). https://doi.org/10.18637/jss.v040.i04

Gangl, M., & Ziefle, A. (2009). Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. Demography, *46*(2): 341–369. https://doi.org/10.1353/dem.0.0056

Gauthier, J.-A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. Sociological Methodology, *40*(1): 1–38. https://doi.org/10.1111/j.1467-9531.2010.01227.x

GenderAPI.io. https://www.genderapi.io/

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. Science Advances, *4*(7): eaaq1360. https://doi.org/10.1126/sciadv.aaq1360

Goldin, C. (2021). *Career and family: Women's century-long journey toward equity.* Princeton University Press. Retrieved March 21, 2024, from http://www.jstor.org/stable/j.ctv1kbgt9p

Gouws, A. (2010). The feminisation of migrations. Africa Insight, *40*(1): 169–180. https://www.africabib.org/rec.php?RID=339376384

Government of Japan, G. E. B. C. O. (2019). *Women and men in japan 2019* (tech. rep.). https://www.gender.go.jp/english_contents/pr_act/pub/pamphlet/women-and-men19/index.html

Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., Mendoza, A., & de Moya-Anegón, F. (2021). Comparative analysis of the bibliographic data sources dimensions and scopus: An approach at the country and institutional levels. Frontiers in Research Metrics and Analytics, *5*. http://dx.doi.org/10.3389/frma.2020.593494

Guthrie, S., Lichten, C. A., Harte, E., Parks, S., & Wooding, S. (2017a). *International mobility of researchers: A survey of researchers in the uk.* RAND Corporation. https://doi.org/10.7249/RR1991

Guthrie, S., Lichten, C., Harte, E., Parks, S., & Wooding, S. (2017b, May 4). *International mobility of researchers: A survey of researchers in the UK*. RAND Corporation.

Haase, I., & Müller, A. (2020). Fighting for attention: Early career researchers and the online scientific community. In *How to make your doctoral research relevant: Insights and strategies for the modern research environment* (pp. 130–141). Edward Elgar Publishing Ltd. Retrieved February 10, 2023, from https://www.elgaronline.com/display/edcoll/9781788977609/9781788977609.00025.xml

Habicht, I. M. (2022). Do mothers get lost at the postdoc stage? Event history analysis of psychologists at German universities (1980–2019). Higher Education. https://doi.org/10.1007/s10734-022-00949-y

Halevi, G. (2019). Bibliometric studies on gender disparities in science. In W. Glänzel, H. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of science and technology indicators* (pp. 563–580). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_21

Haustein, S., Costas, R., & Larivière, V. (2015). Correction: Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. PLOS ONE, *10*(5): e0127830. https://doi.org/10.1371/journal.pone.0127830

Hayum, O. (2023). Do scientists still love 'X'? exploring the platform's modest but dedicated user base. https://www.missionnorth.com/dispatches/do-scientists-still-love-x-exploring-the-platforms-modest-but-dedicated-user-base

Henley, M. M. (2015). Women's success in academic science: Challenges to breaking through the ivory ceiling. Sociology Compass, *9*(8): 668–680. https://doi.org/10.1111/soc4.12291

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. Quantitative Science Studies, *1*(1): 387–395. http://dx.doi.org/10.1162/qss_a_00020

Holloway, S., O'Hara, S., & Pimlott-Wilson, H. (2012). Educational mobility and the gendered geography of cultural capital: The case of international student flows between central asia and the uk. Environment and Planning A, *44*: 2278–2294. https://doi.org/10.1068/a44655

Houle, B., Holt, J., Gillespie, C., Freedman, D. S., & Reyes, M. (2009). Use of density-equalizing cartograms to visualize trends and disparities in state-specific prevalence of obesity: 1996–2006. American Journal of Public Health, *99*(2): 308–312. https://doi.org/10.2105/AJPH.2008.138750

Howoldt, D., Kroll, H., Neuhäusler, P., & Feidenheimer, A. (2022). Understanding researchers' twitter uptake, activity and popularity—an analysis of applied research in germany. Scientometrics, *128*(1): 325–344. https://doi.org/10.1007/s11192-022-04569-2

Huang, C.-K. (, Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. Quantitative Science Studies: 1–34. http://dx.doi.org/10.1162/qss_a_00031

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. Proceedings of the National Academy of Sciences, *117*(9): 4609–4616. https://doi.org/10.1073/pnas.1914221117

Huang, W., Wang, P., & Wu, Q. (2018). A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals (P. van den Besselaar, Ed.). PLOS ONE, *13*(4): e0194962. http://dx.doi.org/10.1371/journal.pone.0194962

Huyer, S. (2015). Is the gender gap narrowing in science and engineering. UNESCO science report: towards 2030: 85–103.

International Organization for Migration and Organisation for Economic Co-operation and Development (OECD). (2014). *Harnessing knowledge on the migration of highly skilled women* (tech. rep.). International Organization for Migration (IOM).

Jonkers, K., & Cruz-Castro, L. (2013). Research upon return: The effect of international mobility on scientific ties, production and impact. Research Policy, *42*: 1366–1377. https://doi.org/10.1016/j.respol.2013.05.005

Jöns, H. (2011a). Transnational academic mobility and gender. Globalisation Societies and Education, *9*: 183–209. https://doi.org/10.1080/14767724.2011.577199

Jöns, H. (2011b). Transnational academic mobility and gender. Globalisation, Societies and Education, *9*(2): 183–209.

Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M., & Strohmaier, M. (2016). Inferring gender from names on the web: A comparative evaluation of gender detection methods. Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion. http://dx.doi.org/10.1145/2872518.2889385

Kashyap, R., Fatehkia, M., Tamime, R. A., & Weber, I. (2020). Monitoring global digital gender inequality using the online populations of facebook and google. Demographic Research, *43*: 779–816. https://doi.org/10.4054/demres.2020.43.27

Kawashima, H., & Tomizawa, H. (2015). Accuracy evaluation of Scopus Author ID based on the largest funding database in Japan. Scientometrics, *103*(3): 1061–1071. https://doi.org/10.1007/s11192-015-1580-z

Kennelly, I., & Spalter-Roth, R. M. (2006). Parents on the job market: Resources and strategies that help sociologists attain tenure-track jobs. The American Sociologist, *37*(4): 29–49. http://dx.doi.org/10.1007/BF02915066

Kerr, S. P., Kerr, W., Özden, Ç., & Parsons, C. (2016). Global talent flows. Journal of Economic Perspectives, *30*(4): 83–106. https://doi.org/10.1257/jep.30.4.83

Kidambi, M. (2024). Since Twitter became X... https://www.altmetric.com/blog/since-twitter-became-x/

Kirloskar, P., & Inamdar, N. (2021). Shifting international student mobility directions and factors influencing students' higher education destination choices. Journal of Higher Education Policy And Leadership Studies, *2*(3). https://doi.org/10.52547/johepal.2.3.160

Klar, S., Krupnikov, Y., Ryan, J. B., Searles, K., & Shmargad, Y. (2020). Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work (J. Jankowski, Ed.). PLOS ONE, *15*(4): e0229446. https://doi.org/10.1371/journal.pone.0229446

Kompetenznetzwerk Bibliometrie. (2021). Kompetenznetzwerk bibliometrie (competence network for bibliometrics). https://bibliometrie.info/

Korhonen, V. (2023). Doctoral degrees earned in the united states 1950-2032, by gender. https://www.statista.com/statistics/185167/number-of-doctoral-degrees-by-gender-since-1950/

Kotsemir, E., M.and Dyachenko, & Nefedova, A. (2022). Mobile young researchers and their non-mobile 'twins': Who is winning the academic race? Scientometrics, *127*: 7307–7332. https://doi.org/10.1007/s11192-022-04488-2

Kozlowski, D., Larivióre, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. Proceedings of the National Academy of Sciences, *119*: 2. https://doi.org/10.1073/pnas.2113067119

Kulis, S., & Sicotte, D. (2002). Women scientists in academia: Geographically constrained to big cities, college clusters, or the coasts? Research in Higher Education, *43*: 1–30. https://link.springer.com/article/10.1023/A:1013097716317

Kulp, A. M. (2016). The effects of parenthood during graduate school on PhD recipients' paths to the professoriate: A focus on motherhood. New Directions for Higher Education, *2016*(176): 81–95. https://doi.org/10.1002/he.20211

Kulp, A. M. (2019). Parenting on the path to the professoriate: A focus on graduate student mothers. Research in Higher Education, *61*(3): 408–429. https://doi.org/10.1007/s11162-019-09561-z

Lantsoght, E. O. L., Tse Crepaldi, Y., Tavares, S. G., Leemans, K., & Paig-Tran, E. W. M. (2021). Challenges and opportunities for academic parents during COVID-19. Frontiers in Psychology, *12*. http://dx.doi.org/10.3389/fpsyg.2021.645734

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. Nature News, *504*: 211–213. https://doi.org/10.1038/504211a

Laudel, G. (2002). What do we measure by co-authorships? Research Evaluation, *11*(1): 3–15. https://doi.org/10.3152/147154402781776961

Leahey, E. (2007). Not by productivity alone: How visibility and specialization contribute to academic earnings. American Sociological Review, *72*(4): 533–561. https://doi.org/10.1177/000312240707200403

Leal Filho, W., Kovaleva, M., Tsani, S., Țîrcă, D.-M., Shiel, C., Dinis, M. A. P., Nicolau, M., Sima, M., Fritzen, B., Lange Salvia, A., Minhas, A., Kozlova, V., Doni, F., Spiteri, J., Gupta, T., Wakunuma, K., Sharma, M., Barbir, J., Shulla, K., … Tripathi, S. (2022). Promoting gender equality across the sustainable development goals. Environment, Development and Sustainability, *25*(12): 14177–14198.

Lee, J., & Haupt, J. (2021). Scientific globalism during a global crisis: Research collaboration and open access publications on covid-19. Higher Education volume, *81*: 949–966. https://doi.org/10.1007/s10734-020-00589-0

Leemann, R. (2010). Gender inequalities in transnational academic mobility and the ideal type of academic entrepreneur. Discourse: Studies in the Cultural Politics of Education, *31*(5): 605–625. https://doi.org/10.1080/01596306.2010.516942

Leggon, C., McNeely, C., & Yoon, J. (2015). Advancing women in science: Policies for progress. Advancing Women in Science: An International Perspective: 307–340. https://doi.org/10.1007/978-3-319-08629-3_10

Liang, L., & Acuna, L. (2021). Demographicx: A python package for estimating gender and ethnicity using deep learning transformers [python] [Science of Science]. https://github.com/sciosci/demographicx?tab=readme-ov-file

Liu, F., Rahwan, T., & AlShebli, B. (2023). Non-white scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. Proceedings of the National Academy of Sciences, *120*(13). https://doi.org/10.1073/pnas.2215324120

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W., Kidd, C., Marcelle, E., … Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. Neuron, *109*(13): 2047–2074. https://doi.org/10.1016/j.neuron.2021.06.002

López-Goñi, I., & Sánchez-Angulo, M. (2017). Social networks as a tool for science communication and public engagement: Focus on twitter. FEMS Microbiology Letters, *365*(2). https://doi.org/10.1093/femsle/fnx246

Lutter, M., & Schröder, M. (2019). Is there a motherhood penalty in academia? The gendered effect of children on academic publications in german sociology. European Sociological Review, *36*(3): 442–459. https://doi.org/10.1093/esr/jcz063

Macaluso, B., Larivière, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is science built on the shoulders of women? a study of gender differences in contributorship.

Academic Medicine, *91*(8): 1136–1142. https://doi.org/10.1097/ACM.0000000000001261
Macháček, V., Srholec, M., Ferreira, M. R., Robinson-Garcia, N., & Costas, R. (2022). Researcher's institutional mobility: Bibliometric evidence on academic inbreeding and internationalization. Science and Public Policy, *49*(1): 85–97. https://doi.org/10.1093/scipol/scab064
Macindoe, H., & Abbott, A. (2004). Sequence analysis and optimal matching techniques for social science data. In *Handbook of data analysis* (pp. 387–406). SAGE Publications, Ltd. https://doi.org/10.4135/9781848608184.n17
Madgavkar, A., Manyika, J., Krishnan, M., Ellingrud, K., Yee, L., Woetzel, J., Chui, M., Hunt, V., & Sruti, B. (2019). *The future of women at work: Transitions in the age of automation* (tech. rep.). McKinsey Globa Institutel. http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/10850/1/MGI-The-future-of-women-at-work-Report-July-2019%20%281%29.pdf
Madsen, S. R. (Ed.). (2017). *Handbook of research on gender and leadership*. Edward Elgar Publishing.
Madsen, S. R., Ngunjiri, F. W., Longman, K. A., & Cherrey, C. (Eds.). (2015). *Women and leadership around the world*. Information Age Publishing.
Maier, G., Kurka, B., & Trippl, M. (2007). Knowledge spillover agents and regional development: Spatial distribution and mobility of star scientists. Papers DYNREG17.
Malakha, I. (2002). On "Brain Drain" in russia during the second half of the 1990s. In I. Vladimir (Ed.), *Scientific series 'international migration of population: Russia and the contemporary world'*. MAX Press.
Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. International Organization, *67*(4): 889–922. http://dx.doi.org/10.1017/S0020818313000209
Marcus, J. (2007). Helping academics have families and tenure too: Universities discover their self-interest. Change: The Magazine of Higher Learning, *39*(2): 27–32. https://doi.org/10.3200/chng.39.2.27-32
Marwell, G., Rosenfeld, R., & Spilerman, S. (1979). Geographic constraints on women's careers in academia. Science, *205*(34): 1225–1231.
Mason, M. A. (2009a). Is tenure a trap for women? https://www.law.berkeley.edu/files/Is_Tenure_a_Trap_for_Women.pdf
Mason, M. A. (2009b). Why so few doctoral-student parents? https://www.law.berkeley.edu/files/Why_So_Few_Doctoral-Student_Parents.pdf
Mason, M. A., & Goulden, M. (2004). Marriage and baby blues: Redefining gender equity in the academy. The ANNALS of the American Academy of Political and Social Science, *596*(1): 86–103. http://dx.doi.org/10.1177/0002716204268744
Mason, M. A., Wolfinger, N. H., & Goulden, M. (2013a). *Do babies matter? gender and family in the ivory tower*. Rutgers University Press. https://doi.org/10.36019/9780813560823
Mason, M. A., Wolfinger, N. H., & Goulden, M. (2013b). *Do babies matter?: Gender and family in the ivory tower*. Rutgers University Press. http://www.jstor.org/stable/j.ctt5hj7tp
Mason, S., Bond, M., & Ledger, S. (2023). How motherhood enhances and strains doctoral research/ers. Journal of Further and Higher Education, *47*(8): 1087–1105. https://doi.org/10.1080/0309877x.2023.2218274

McPherson, J. M., & Smith-Lovin, L. (1986). Sex segregation in voluntary associations. American Sociological Review, *51*(1): 61. https://doi.org/10.2307/2095478

Melin, G., & Janson, K. (2006). What skills and knowledge should a PhD have? Changing preconditions for PhD education and post doc work. In *The formative years of scholars, wenner-gren international series* (pp. 105–118). Portland Press.

Menéndez, D. A. *Damegender: Writing and comparing gender detection tools*. EasyChair Preprint no. 3288. 2020. https://ceur-ws.org/Vol-2754/paper3.pdf

Miranda-González, A., Aref, S., Theile, T., & Zagheni, E. (2020). Scholarly migration within mexico: Analyzing internal migration among researchers using scopus longitudinal bibliometric data. EPJ Data Science, *9*(34). https://doi.org/10.1140/epjds/s13688-020-00252-9

Mody, L., Griffith, K. A., Jones, R. D., Stewart, A., Ubel, P. A., & Jagsi, R. (2021). Gender differences in work-family conflict experiences of faculty in academic medicine. Journal of General Internal Medicine, *37*(1): 280–282. http://dx.doi.org/10.1007/s11606-020-06559-7

Moed, H. F., Aisati, M., & Plume, A. (2013). Studying scientific migration in Scopus. Scientometrics, *94*(3): 929–942. https://doi.org/10.1007/s11192-012-0783-9

Moed, H. F., & Halevi, G. (2014). A bibliometric approach to tracking international scientific migration. Scientometrics, *101*(3): 1987–2001. https://doi.org/10.1007/s11192-014-1307-6

Mongeon, P., Bowman, T. D., & Costas, R. (2023). An open dataset of scholars on Twitter. Quantitative Science Studies, *4*(2): 1–18. Retrieved March 11, 2023, from https://doi.org/10.1162/qss_a_00250

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. Scientometrics, *106*(1): 213–228. https://doi.org/10.1007/s11192-015-1765-5

Moreau, M.-P., & Kerner, C. (2013). Care in academia: An exploration of student parents' experiences. British Journal of Sociology of Education, *36*(2): 215–233. http://dx.doi.org/10.1080/01425692.2013.814533

Morgan, A. C., Way, S. F., Hoefer, M. J. D., Larremore, D. B., Galesic, M., & Clauset, A. (2021a). The unequal impact of parenthood in academia. Science Advances, *7*(9). https://doi.org/10.1126/sciadv.abd1996

Morgan, A. C., Way, S. F., Hoefer, M. J., Larremore, D. B., Galesic, M., & Clauset, A. (2021b). The unequal impact of parenthood in academia. Science Advances, *7*(9): eabd1996. https://doi.org/10.1126/sciadv.abd1996

Morrison, E., Rudd, E., & Nerad, M. (2011). Onto, up, off the academic faculty ladder: The gendered effects of family on career transitions for a cohort of Social Science Ph.D.s. The Review of Higher Education, *34*(4): 525–553. https://doi.org/10.1353/rhe.2011.0017

Mosse, J. C. (1993). *Half the world half a chance: An introduction to gender and development*. Oxfam; Reprint Edition. http://hdl.handle.net/10546/122709

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences, *109*(41): 16474–16479. https://doi.org/10.1073/pnas.1211286109

Muchomba, F. M., & Kaushal, N. (2022). Gender differences in immigrant assimilation activities in the us: Evidence from time-use data. Feminist Economics, *28*: 1–28. https://doi.org/10.1080/13545701.2021.2015538

Muñoz Rojas, C. (2023). *Public policies for gender equality in science, technology, engineering and mathematics (stem): Challenges for the economic autonomy of women and transformative recovery in latin america* (tech. rep.). Economic Commission for Latin America and the Caribbean (ECLAC). https://repositorio.cepal.org/server/api/core/bitstreams/6b7dd3c2-8639-46c9-b7db-2cab26fab283/content

Nair-Bedouelle, S. (2023). The lack of gender equality in science is everyone's problem. https://www.un.org/en/un-chronicle/lack-gender-equality-science-everyone%E2%80%99s-problem#:~:text=Their%20work%20is%20underrepresented%20in,national%20science%20academies%20are%20women.

Namesor. https://namsor.app/

Nascia, L., Pianta, M., & Zacharewicz, T. (2021). Staying or leaving? Patterns and determinants of italian researchers' migration. Science and Public Policy, *48*(2): 200–211. https://doi.org/10.1093/scipol/scab007

National Center for Science and Engineering Statistics, National Science Foundation. (2023). Survey of earned doctorates. https://ncses.nsf.gov/pubs/nsf24300/data-tables

Nejad, M. N., & Young, A. T. (2014). *Female brain drains and women's rights gaps: A gravity model analysis of bilateral migration flows [discussion paper]* (tech. rep.). IZA Institute of Labour Economics.

Netz, N., & Jaksztat, S. (2014). Mobilised by mobility? Determinants of international mobility plans among doctoral candidates in Germany [Academic mobility].

Netz, N., & Jaksztat, S. (2017). Explaining scientists' plans for international mobility from a life course perspective. Research in Higher Education, *58*(5): 497–519. https://doi.org/10.1007/s11162-016-9438-7

Nicholas, D., Rodrıguez-Bravo, B., Watkinson, A., Boukacem-Zeghmouri, C., Herman, E., Xu, J., Abrizah, A., & Świgoń, M. (2017). Early career researchers and their publishing and authorship practices. Learned Publishing, *30*(3): 205–217. https://doi.org/10.1002/leap.1102

Nicolas, G. (2020). Using twitter data in research guidance for researchers and ethics reviewers. https://alfasoft.com/blog/products/scientific-writing-and-publishing/exploring-bibliometric-methods-citation-analysis-in-research/

Nittrouer, C. L., Hebl, M. R., Ashburn-Nardo, L., Trump-Steele, R. C., Lane, D. M., & Valian, V. (2018). Gender disparities in colloquium speakers at top universities. Proceedings of the National Academy of Sciences, *115*(1): 104–108.

Organisation for Economic Co-operation and Development (OECD). (2007). *Revised field of science and technology (fos) classification in the frascati manual (classification, field of science and technology classification, fos, frascati, methodology, research and development)* (tech. rep.). https://www.oecd.org/science/inno/38235147.pdf

Organisation for Economic Co-operation and Development (OECD). (2008). *The global competition for talent: Mobility of the highly skilled.* OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2013). *International migration outlook 2013* (tech. rep.). https://doi.org/10.1787/1999124x

Organisation for Economic Co-operation and Development (OECD). (2015). *Talent abroad: A review of german emigrants.* OECD Publishing. https://doi.org/10.1787/9789264231702-en

Organisation for Economic Co-operation and Development (OECD). (2021). Gross domestic spending on R&D (indicator). https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm
Pain, E. (2021). How to navigate authorship of scientific manuscripts. Science, *0*. https://doi.org/10.1126/science.caredit.abj3459
Paksi, V., Nagy, B., & Király, G. (2016). The timing of motherhood while earning a PhD in engineering. International Journal of Doctoral Studies, *11*: 285–304. http://dx.doi.org/10.28945/3544
Palamar, J. J., & Strain, E. C. (2021). News and social media coverage is associated with more downloads and citations of manuscripts that focus on substance use. Drug and Alcohol Dependence, *218*: 108357. https://doi.org/10.1016/j.drugalcdep.2020.108357
Parey, M., Ruhose, J., Waldinger, F., & Netz, N. (2017). The selection of high-skilled emigrants. Review of Economics and Statistics, *99*(5): 776–792. https://doi.org/10.1162/REST_a_00687
Paturi, M., & Loktev, A. (2020). The best gets better: Scopus data quality, measured. https://www.brighttalk.com/webcast/13819/456949
Paul-Hus, A., Sugimoto, C., Haustein, S., & Larivière, V. (2015). Is there a gender gap in social media metrics? 15th International Conference on Scientometrics & Informetrics.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, *12*: 2825–2830.
Peng, H., Teplitskiy, M., Romero, D. M., & Horvát, E.-Á. (2022). The gender gap in scholarly self-promotion on social media. arXiv preprint arXiv:2206.05330. https://arxiv.org/abs/2206.05330
Peters, S. A. E., & Norton, R. (2018). Sex and gender reporting in global health: New editorial policies. BMJ Global Health, *3*(4): e001038. http://dx.doi.org/10.1136/bmjgh-2018-001038
Petersen, A. M., Vincent, E. M., & Westerling, A. L. (2019). Discrepancy in scientific authority and media visibility of climate change scientists and contrarians. Nature Communications, *10*(1). https://doi.org/10.1038/s41467-019-09959-4
Priem, J., Piwowar, H., & Orr, R. (2022). Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. https://arxiv.org/abs/2205.01833
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, *155*(2): 945–959. https://www.genetics.org/content/155/2/945
Raffo, J. (2016). Worldwide Gender-Name Dictionary (WGND). In W. —Economics & S. Division (Eds.), *Wipo economics and statistics related resources*. https://ideas.repec.org/c/wip/eccode/10.html
Rafnsdóttir, G. L., & Heijstra, T. M. (2011). Balancing work–family life in academia: The power of time. Gender, Work, and Organization, *20*(3): 283–296. http://dx.doi.org/10.1111/j.1468-0432.2011.00571.x
Reilly, J. J., Armstrong, J., Dorosty, A. R., Emmett, P. M., Ness, A., Rogers, I., Steer, C., & Sherriff, A. (2005). Early life risk factors for obesity in childhood: Cohort study. BMJ, *330*(7504): 1357. https://doi.org/10.1136/bmj.38470.670903.E0

Directorate-General for Research and Innovation (European Commission). (2019). *She figures 2018* (tech. rep.). Publications Office of the European Union. https://data.europa.eu/doi/10.2777/936

Directorate-General for Research and Innovation (European Commission). (2021a). Horizon europe guidance on gender equality plans. https://op.europa.eu/en/publication-detail/-/publication/ffcb06c3-200a-11ec-bd8e-01aa75ed71a1

Directorate-General for Research and Innovation (European Commission). (2021b). *She figures 2021: The path towards gender equality in research and innovation (R&I)* (tech. rep.). Publications Office of the European Union. https://data.europa.eu/doi/10.2777/759686

Rexrode, K. M. (2016). The gender gap in first authorship of research papers. BMJ: i1130.

Rivera, L. A. (2017). When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring. American Sociological Review, *82*(6): 1111–1138.

Robinson-Garcia, N., Sugimoto, C. R., Murray, D., Yegros-Yegros, A., Larivière, V., & Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. Journal of Informetrics, *13*(1): 50–63. https://doi.org/10.1016/j.joi.2018.11.002

Rosenfeld, R., & Jones, J. (1987). Patterns and effects of geographic mobility for academic women and men. The Journal of Higher Education, *58*(5): 493–515. https://doi.org/10.2307/1981784

Rothman, K. J. (2012). *Epidemiology: An introduction.* Oxford University Press. https://doi.org/10.1093/aje/kwf028

Ruggieri, R., Pecoraro, F., & Luzi, D. (2021). An intersectional approach to analyse gender productivity and open access: A bibliometric analysis of the italian national research council. Scientometrics, *126*(2): 1647–1673. http://dx.doi.org/10.1007/s11192-020-03802-0

Schaer, M., Dahinden, J., & Toader, A. (2017). Transnational mobility among early-career academics: Gendered aspects of negotiations and arrangements within heterosexual couples. Journal of Ethnic and Migration Studies, *43*(8): 1292–1307.

Schiller, D., & Cordes, A. (2016). Measuring researcher mobility. OECD Blue Sky Forum: 24. https://www.slideshare.net/innovationoecd/schiller-measuring-researchers-mobility

Science|Business. (2022). Scientists leave the UK as China overtakes US as most favoured destination. https://sciencebusiness.net/news-byte/scientists-leave-uk-china-overtakes-us-most-favoured-destination/

SCImago. SJR — SCImago Journal & Country Rank [Portal]. http://www.scimagojr.com

Sebo, P. (2022). How accurate are gender detection tools in predicting the gender for chinese names? a study with 20,000 given names in pinyin format. Journal of the Medical Library Association, *110*(2): 205–211. https://doi.org/10.5195/jmla.2022.1289

Sebo, P. (2021a). Performance of gender detection tools: A comparative study of name-to-gender inference services. Journal of the Medical Library Association, *109*(3). http://dx.doi.org/10.5195/jmla.2021.1185

Sebo, P. (2021b). Using genderize.io to infer the gender of first names: How to improve the accuracy of the inference. Journal of the Medical Library Association, *109*(4). http://dx.doi.org/10.5195/jmla.2021.1252

Seguino, S. (2000). Gender inequality and economic growth: A cross-country analysis. World Development, *28*(7): 1211–1230.

Sharafizad, F., Brown, K., Jogulu, U., & Omari, M. (2020). Letting a picture speak a thousand words: Arts-based research in a study of the careers of female academics. Sociological Methods & Research, *52*(1): 438–479. https://doi.org/10.1177/0049124120926206

Shauman, Y., K. A.and Xie. (1996). Geographic mobility of scientists: Sex differences and family constraints. Demography, *33*(4): 455–468. https://doi.org/10.2307/2061780

Shinozaki, K. (2017). Gender and citizenship in academic career progression: An intersectional, meso-scale analysis in german higher education institutions. Journal of Ethnic and Migration Studies, *43*: 1325–1346. https://doi.org/10.1080/1369183X.2017.1300309

Shu, F., & Haustein, S. (2017). On the citation advantage of tweeted papers at the journal level. Proceedings of the Association for Information Science and Technology, *54*(1): 366–372. https://doi.org/10.1002/pra2.2017.14505401040

Shu, F., Lou, W., & Haustein, S. (2018). Can twitter increase the visibility of chinese publications? Scientometrics, *116*(1): 505–519. https://doi.org/10.1007/s11192-018-2732-8

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics, *126*(6): 5113–5142. http://dx.doi.org/10.1007/s11192-021-03948-5

Skiles, M., Yang, E., Reshef, O., Muñoz, D. R., Cintron, D., Lind, M. L., Rush, A., Calleja, P. P., Nerenberg, R., Armani, A., M. Faust, K., & Kumar, M. (2022). Conference demographics and footprint changed by virtual platforms. Nature Sustainability, *5*: 149–156. Retrieved December 20, 2021, from https://www.nature.com/articles/s41893-021-00823-2

Solga, H., Rusconi, A., & Netz, N. (2023). Professors' gender biases in assessing applicants for professorships. European Sociological Review, *39*(6): 841–861. http://dx.doi.org/10.1093/esr/jcad007

Sondhi, G., & King, R. (2017). Gendering international student migration: An indian case–study. Journal of Ethnic and Migration Studies, *43*(8): 1308–1324. https://doi.org/10.1016/j.jacr.2016.08.026

Song, Q., & Liang, Z. (2019). New emigration from China: Patterns, causes and impacts. Modern China studies, *26*(1): 5–31. https://pubmed.ncbi.nlm.nih.gov/34377583/

Sougou, N., Ndiaye, M., O.and Folayan, Sarr, E., S.C.and Mbaye, & Martinez-Pérez, G. (2022). Barriers of west african women scientists in their research and academic careers: A qualitative research. PLoS ONE, *17*(13): e0265413. https://doi.org/10.1371/journal.pone.0265413

Staniscuaski, F., Reichert, F., Werneck, F. P., de Oliveira, L., Mello-Carpes, P. B., Soletti, R. C., Almeida, C. I., Zandona, E., Ricachenevsky, F. K., Neumann, A., Schwartz, I. V. D., Tamajusuku, A. S. K., Seixas, A., & Kmetzsch, L. (2020). Impact of covid-19 on academic mothers (J. Sills, Ed.). Science, *368*(6492): 724–724. http://dx.doi.org/10.1126/science.abc2740

Subbotin, A., & Aref, S. (2021). Brain drain and brain gain in Russia: Analyzing international mobility of researchers by discipline using Scopus bibliometric data 1996-2020. Scientometrics, *126*(9): 7875–7900. https://doi.org/10.1007/s11192-021-04091-x

Sugimoto, C. R., & Larivière, V. (2023). *Equity for women in science: Dismantling systemic barriers to advancement*. Harvard University Press. http://dx.doi.org/10.4159/9780674292918

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. Journal of the Association for Information Science and Technology, *68*(9): 2037–2062. https://doi.org/10.1002/asi.23833

Surawicz, C. M. (2016). Women in leadership: Why so few and what to do about it. Journal of the American College of Radiology, *13*(12): 1433–1437. http://dx.doi.org/10.1016/j.jacr.2016.08.026

Teele, D. L., & Thelen, K. (2017). Gender in the journals: Publication patterns in political science. PS: Political Science & Politics, *50*(2): 433–447. https://doi.org/10.1017/S1049096516002985

Teich, E. G., Kim, J. Z., Lynn, C. W., Simon, S. C., Klishin, K. P., Andrei A. andSzymula, Srivastava, P., Bassett, L. C., Zurn, P., Dworkin, J. D., & Bassett, D. S. (2022). Citation inequity and gendered citation practices in contemporary physics. Nature Physics, *18*(10): 1161–1170. https://doi.org/10.1038/s41567-022-01770-1

Teichler, U. (2015). Academic mobility and migration: What we know and what we do not know. European Review, *23*: S6–S37. https://doi.org/10.1017/S1062798714000787

The Intergovernmental Support Division of UN Women. (2019). *A short history of the commission on the status of women* (G. Luchsinger, Ed.). United Nations Division on the Advancement of Women. https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2019/A-short-history-of-the-CSW-en.pdf

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLOS ONE, *8*(5): e64841.

Tokunaga, T., & Makoto, I. (1994). Text categorization based on weighted inverse document frequency. Special Interest Groups and Information Process Society of Japan (SIG-IPSJ): 33–39. https://doi.org/10.1.1.49.7015

Torche, F. (2014). Intergenerational mobility and inequality: The latin american case. Annual Review of Sociology, *40*: 619–642. https://doi.org/10.1146/annurev-soc-071811-145521

Tower, L. E., & Latimer, M. (2016). Cumulative disadvantage. Affilia, *31*(3): 317–330. https://doi.org/10.1177/0886109915622527

Turney, A. (2019). Opportunities for foreign researchers in Switzerland. https://www.academics.com/guide/research-switzerland (accessed: 01.2019)

Van der Lee, R., & Ellemers, N. (2015). Gender contributes to personal research funding success in the netherlands. Proceedings of the National Academy of Sciences, *112*(40): 12349–12353.

Vásárhelyi, O., Zakhlebin, I., Milojevic, S., & Horvát, E.-Á. (2021). Gender inequities in the online dissemination of scholars' work. Proceedings of the National Academy of Sciences, *118*: e2102945118. https://doi.org/10.1073/pnas.2102945118

Venkatraman, V. (2014). Elite male faculty employ fewer women. https://www.science.org/content/article/elite-male-faculty-employ-fewer-women

Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. Scientometrics, *121*(3): 1803–1813. http://dx.doi.org/10.1007/s11192-019-03264-z

Vidal Valero, M. (2023). Thousands of scientists are cutting back on twitter, seeding angst and uncertainty. Nature, *620*(7974): 482–484. http://dx.doi.org/10.1038/d41586-023-02554-0

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Quantitative Science Studies, *2*(1): 20–41. http://dx.doi.org/10.1162/qss_a_00112

Wang, B. (2020). A temporal gaze towards academic migration: Everyday times, lifetimes and temporal strategies amongst early career Chinese academic returnees [Publisher: SAGE Publications Ltd]. Time & Society, *29*(1): 166–186.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, *58*(301): 236–244. https://doi.org/10.1080/01621459.1963.10500845

Weert, E. d. (2013). *Support for continued data collection and analysis concerning mobility patterns and career paths of researchers*. European Commission, Research Directorate-General. Directorate B - European Research Area.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship (L. Hadany, Ed.). PLoS ONE, *8*(7): e66212.

Wiesel, T. (2014). Fellowships: Turning brain drain into brain circulation. Nature, *510*: 213–214. https://doi.org/10.1038/510213a

Williams, A. M. (2007). International labour migration and tacit knowledge transactions: A multi-level perspective. Global Networks, *7*(1): 29–50. https://doi.org/10.1111/j.1471-0374.2006.00155.x

Winkelmann, M. (2022). List of first names, genders and country-specific frequencies. Retrieved December 5, 2022, from https://github.com/MatthiasWinkelmann/firstname-database

Wolfinger, N. H., Mason, M. A., & Goulden, M. (2009). Stay in the game: Gender, family formation and alternative trajectories in the academic life course. Social Forces, *87*(3): 1591–1621. https://doi.org/10.1353/sof.0.0182

Wolfinger, N. H., Mason, M. A., & Goulden, M. (2008). Problems in the pipeline: Gender, marriage, and fertility in the ivory tower. The Journal of Higher Education, *79*(4): 388–405. https://doi.org/10.1080/00221546.2008.11772108

Wolf-Wendel, L., & Ward, K. (2014). Academic mothers: Exploring disciplinary perspectives. Innovative Higher Education, *40*(1): 19–35. https://doi.org/10.1007/s10755-014-9293-4

World Economic Forum. (2024). The gender health gap: It's more than a women's issue. here's why. https://www.weforum.org/agenda/2024/02/women-health-gender-gap-equality/

Yammine, S. Z., Liu, C., Jarreau, P. B., & Coe, I. R. (2018). Social media for social change in science. Science, *360*(6385): 162–163.

Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F., & Uzzi, B. (2022). Gender-diverse teams produce more novel and higher-impact scientific ideas. Proceedings of the National Academy of Sciences, *119*(36).

Yoon, H., & Kim, H. (2018). Career-family strategies and spatial mobility among south korean women scholars. Asian and Pacific Migration Journal, *27*(1): 101–120. https://doi.org/10.1177/0117196818760863

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. Psychology of Women Quarterly, *37*(3): 283–292.

Yu, H., Xiao, T., Xu, S., & Wang, Y. (2019). Who posts scientific tweets? An investigation into the productivity, locations, and identities of scientific tweeters. Journal of Informetrics, *13*(3): 841–855.

Zhang, M.-Z., Wang, T.-R., Lyu, P.-H., Chen, Q.-M., Li, Z.-X., & Ngai, E. W. (2024). Impact of gender composition of academic teams on disruptive output. Journal of Informetrics, *18*(2): 101520.

Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2021). International migration in academia and citation performance: An analysis of German-affiliated researchers by gender and discipline using Scopus publications 1996-2020. Proceedings of the 18th International Conference on Scientometrics and Informetrics (ISSI 2021).

Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2022a). Return migration of German-affiliated researchers: Analyzing departure and return by gender, cohort, and discipline using Scopus bibliometric data 1996-2020. Scientometrics, *127*. https://doi.org/10.1007/s11192-022-04351-4

Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars [Publisher: Proceedings of the National Academy of Sciences]. Proceedings of the National Academy of Sciences, *120*(10): e2214664120. Retrieved February 27, 2023, from https://www.pnas.org/doi/10.1073/pnas.2214664120

Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2022b). Dataset of international migration among German-affiliated researchers in Scopus over 1996-2020. https://doi.org/10.6084/m9.figshare.18433139

Zhou, X., & Mollo, V. (2023). Parenting doctoral students: Who are they, and how are they doing? Journal of Diversity in Higher Education. http://dx.doi.org/10.1037/dhe0000540

Zhu, Y., Kim, D., Yan, E., Kim, M., & Qi, G. (2020). Analyzing China's research collaboration with the United States in high-impact and high-technology research. Quantitative Science Studies, *2*: 1–13. https://doi.org/10.1162/qss_a_00098

Zippel, K. (2017). *Women in global science: Advancing academic careers through international collaboration*. Stanford University Press. https://doi.org10.1515/9781503601505