



## IMPRS-PHDS 2022 course (IDEM187) on Topics in Digital and Computational Demography – Day 4 (November 14<sup>th</sup> 2022)

### BiblioDemography:

Using large-scale bibliometric data for demographic research;  
Advantages and pitfalls of using Scopus data to trace internal and international scholarly migration worldwide

Aliakbar Akbaritabar<sup>1</sup>

<sup>1</sup> Max Planck Institute for Demographic Research (MPIDR)  
Akbaritabar@demogr.mpg.de

## The main **goals** for this course (and day 4) are:

- To introduce students to **core demographic and social science methods** that are essential to interpret **digital trace data**;
- To introduce students to **core data science methods** that are key to advance our understanding of population processes in the context of the increasing heterogeneity of data sources useful for demographic research.
- To introduce students to **recent substantive advances** in the field of Digital and Computational Demography, with emphasis on fostering critical thinking about modern demographic analysis and (big) data-driven discovery.
- To help students **identify research questions** in their own area of substantive interest that could be addressed with innovative data sources, and support them in the process of devising an appropriate research plan.

Please tweet with hashtag  
#BiblioDemography,  
a tribute to James W. Vaupel (1945-2022).  
Thanks Ilya and Jonas for bringing up Jim's  
labeling idea!



## AGENDA

### 1. Introduction (15 minutes, [video 1])

- Welcome and introduction
- What is bibliometric data?
- What type of questions can be studied using bibliometric data?
- What type of demographic questions can be studied using bibliometric data?
- Limitations and pitfalls of using bibliometric data.

### 2. Data Science skills to use bibliometric data (45 minutes, [videos: 2\_1, 2\_2])

- [video 2\_1] Introduction to parallelised analysis of large-scale bibliometric, text and network data (with Dask in Python, DuckDB and DBeaver in SQL)
- [video 2\_2] Hands-on example of parallelised analysis of bibliometric data
- [video 2\_2] Hands-on example on use of text and network analysis

### 3. Example empirical study using bibliometric data (30 minutes, [video 3])

- Internal and international migration of scholars worldwide: Trends, patterns, and inter-relationships

Please tweet with hashtag  
#BiblioDemography,  
a tribute to James W. Vaupel (1945-2022).  
Thanks Ilya and Jonas for bringing up Jim's  
labeling idea!



AGENDA

1. Introduction (15 minutes, [video 1])

- Welcome and introduction
- What is bibliometric data?
- What type of questions can be studied using bibliometric data?
- What type of demographic questions can be studied using bibliometric data?
- Limitations and pitfalls of using bibliometric data.

2. Data Science skills to use bibliometric data (45 minutes, [videos: 2\_1, 2\_2])

- [video 2\_1] Introduction to parallelised analysis of large-scale bibliometric, text and network data (with Dask in Python, DuckDB and DBeaver in SQL)
- [video 2\_2] Hands-on example of parallelised analysis of bibliometric data
- [video 2\_2] Hands-on example on use of text and network analysis

3. Example empirical study using bibliometric data (30 minutes, [video 3])

- Internal and international migration of scholars worldwide: Trends, patterns, and inter-relationships

# Materials publicly available



## Related links:

- Interview and references on using bibliometric data for demographic research:  
[https://www.demogr.mpg.de/en/news\\_events\\_6123/news\\_press\\_releases\\_4630/news/how\\_to\\_use\\_bibliometric\\_data\\_for\\_demographic\\_research\\_10784](https://www.demogr.mpg.de/en/news_events_6123/news_press_releases_4630/news/how_to_use_bibliometric_data_for_demographic_research_10784)

akbaritabar / BiblioDemography\_IMPRS\_PHDS\_2022\_IDEM187

main 1 branch 0 tags

File/Folder	Description	Time
akbaritabar	PDF file of course instructions (readme) added	16aa349 4 minutes ago 3 commits
0_code	Course materials added	24 minutes ago
1_data	Course materials added	24 minutes ago
2_presentations	Course materials added	24 minutes ago
3_readings	Course materials added	24 minutes ago
98_outputs	Course materials added	24 minutes ago
99_images	Download guide photo added	18 minutes ago
.gitignore	Course materials added	24 minutes ago
LICENSE	Course materials added	24 minutes ago
Readme.md	PDF file of course instructions (readme) added	4 minutes ago
Start here and read this document fir...	PDF file of course instructions (readme) added	4 minutes ago

Readme.md

### Materials for day 4 of IMPRS-PHDS 2022 course (IDEM187) on Topics in Digital and Computational Demography, 14-18 November 2022

NOTE: Please read this document first, before watching videos, or checking other directories.

- Course description:  
[https://www.demogr.mpg.de/en/career\\_6122/international\\_advanced\\_studies\\_in\\_demography\\_6682/courses\\_6931/topics\\_in\\_digital\\_and\\_computational\\_demography\\_11309](https://www.demogr.mpg.de/en/career_6122/international_advanced_studies_in_demography_6682/courses_6931/topics_in_digital_and_computational_demography_11309)
- This repository is for: Day 4, November 17th 2022
- Topics covered in day 4: Using large-scale bibliometric data for demographic research: Advantages and pitfalls

Releases: No releases published. Create a new release

Packages: No packages published. Publish your first package

Languages: Python 98.6%, TeX 1.4%

**Materials:** [https://github.com/akbaritabar/BiblioDemography\\_IMPRS\\_PHDS\\_2022\\_IDEM187](https://github.com/akbaritabar/BiblioDemography_IMPRS_PHDS_2022_IDEM187)

# What is bibliometric data?



an open access  journal



Citation: Akbaritabar, A. (2021). A quantitative view of the structure of institutional scientific collaborations using the example of Berlin. *Quantitative Science Studies*, 2(2), 753–777. [https://doi.org/10.1162/qss\\_a\\_00131](https://doi.org/10.1162/qss_a_00131)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00131](https://doi.org/10.1162/qss_a_00131)

Peer Review:  
[https://publons.com/publon/10.1162/qss\\_a\\_00131](https://publons.com/publon/10.1162/qss_a_00131)

Received: 2 September 2020  
Accepted: 2 April 2021

Corresponding Author:  
Aliakbar Akbaritabar  
[akbaritabar@demogr.mpg.de](mailto:akbaritabar@demogr.mpg.de)

Handling Editor:  
Ludo Waltman

## RESEARCH ARTICLE

# A quantitative view of the structure of institutional scientific collaborations using the example of Berlin

Aliakbar Akbaritabar<sup>1,2</sup> 

<sup>1</sup>Max Planck Institute for Demographic Research (MPIDR),  
Laboratory of Digital and Computational Demography, Rostock, Germany

<sup>2</sup>German Centre for Higher Education Research and Science Studies (DZHW), Berlin, Germany

**Keywords:** Berlin, Berlin University Alliance, bipartite community detection, coauthorship network analysis, disambiguation, internationalization

## ABSTRACT


This paper examines the structure of scientific collaborations in Berlin as a specific case with a unique history of division and reunification. It aims to identify strategic organizational coalitions in a context with high sectoral diversity. We use publications data with at least one organization located in Berlin from 1996–2017 and their collaborators worldwide. We further investigate four members of the Berlin University Alliance (BUA), as a formerly established coalition in the region, through their self-represented research profiles compared with empirical results. Using a bipartite network modeling framework, we move beyond the uncontested trend towards team science and increasing internationalization. Our results show that BUA members shape the structure of scientific collaborations in the region. However, they are not collaborating cohesively in all fields and there are many smaller scientific actors involved in more internationalized collaborations in the region. Larger divides exist in some fields. Only Medical and Health Sciences have cohesive intraregional collaborations, which signals the success of the regional cooperation established in 2003. We explain possible underlying factors shaping the intraregional groupings and potential implications for regions worldwide. A major methodological contribution of this paper is evaluating the coverage and accuracy of different organization name disambiguation techniques.





# What is bibliometric data?

## Metadata of scientific publications

- Authors' name
- Affiliation addresses
- Publication type (article, review, conference proceedings)
- Subject classification (often based on journal assignment)
- Title, abstract and keywords (for text analysis)
- Reference list (for citation analysis)
- Acknowledgements (for funding information)
- Open Access information



an open access  journal



Citation: Akbaritabar, A. (2021). A quantitative view of the structure of institutional scientific collaborations using the example of Berlin. *Quantitative Science Studies*, 2(2), 753–777. [https://doi.org/10.1162/qss\\_a\\_00131](https://doi.org/10.1162/qss_a_00131)

DOI: [https://doi.org/10.1162/qss\\_a\\_00131](https://doi.org/10.1162/qss_a_00131)

Peer Review: [https://publons.com/publon/10.1162/qss\\_a\\_00131](https://publons.com/publon/10.1162/qss_a_00131)


Received: 2 September 2020  
Accepted: 2 April 2021

Corresponding Author:  
Aliakbar Akbaritabar  
[akbaritabar@demogr.mpg.de](mailto:akbaritabar@demogr.mpg.de)

Handling Editor:  
Ludo Waltman

RESEARCH ARTICLE

## A quantitative view of the structure of institutional scientific collaborations using the example of Berlin

Aliakbar Akbaritabar<sup>1,2</sup> 

<sup>1</sup>Max Planck Institute for Demographic Research (MPIDR),  
Laboratory of Digital and Computational Demography, Rostock, Germany  
<sup>2</sup>German Centre for Higher Education Research and Science Studies (DZHW), Berlin, Germany

**Keywords:** Berlin, Berlin University Alliance, bipartite community detection, coauthorship network analysis, disambiguation, internationalization

---

**ABSTRACT**

This paper examines the structure of scientific collaborations in Berlin as a specific case with a unique history of division and reunification. It aims to identify strategic organizational coalitions in a context with high sectoral diversity. We use publications data with at least one organization located in Berlin from 1996–2017 and their collaborators worldwide. We further investigate four members of the Berlin University Alliance (BUA), as a formerly established coalition in the region, through their self-represented research profiles compared with empirical results. Using a bipartite network modeling framework, we move beyond the uncontested trend towards team science and increasing internationalization. Our results show that BUA members shape the structure of scientific collaborations in the region. However, they are not collaborating cohesively in all fields and there are many smaller scientific actors involved in more internationalized collaborations in the region. Larger divides exist in some fields. Only Medical and Health Sciences have cohesive intraregional collaborations, which signals the success of the regional cooperation established in 2003. We explain possible underlying factors shaping the intraregional groupings and potential implications for regions worldwide. A major methodological contribution of this paper is evaluating the coverage and accuracy of different organization name disambiguation techniques.

## Example science of science questions

- How do scientists do science?
- Who **collaborates** with whom?
- What **kind of teams** develop and produce more scientific breakthroughs?
- In these cases, bibliometric data is used **to trace scientists' behaviour**. Research questions can also be broader to investigate **academia and the science system in general**, for instance gender disparities.
- In addition, some research questions focus on the **content of scientific publications** to summarize or systematically review them. They aim to see how specific fields are developing or specific topics are being studied.



## Example demographic questions

- The usage in **demographic research** can roughly be divided into two groups.
- One group looks at the **population of scholars** and uses bibliometric data to study the composition of different national science systems, and their development over time. This application is similar to the research question on gender disparities.
- Another group of studies **re-purposes bibliometric data and treats it like digital traces** to study, such as scholars' mobility and migration.
- These studies investigate **academic affiliation addresses** and use them as **proxy for academic's residential addresses**. Simply put, changes in these addresses could signal a mobility event, which helps us study scientific and scholarly migration worldwide.



# Data gathering techniques in studying scholars as a population

- **Surveys of scholars** (Cañibano et al., 2020; Franzoni et al., 2012a, 2012b, 2014, 2015; Gibson & McKenzie, 2014; Guthrie et al., 2017; Jöns, 2011; Nascia et al., 2021; Netz & Jaksztat, 2017; Scellato et al., 2015, 2017; Wilson & Gaston, 1971)
- **Interviews** (Cole & Zuckerman, 1987; Jonkers, 2011; Schaer, 2021; Schaer et al., 2017, 2020; Shen et al., 2017)
- **Administrative and census data** (Fenton et al., 2000; ippedico, 2021; Shauman & Xie, 1996)
- **Online sources** (e.g., LinkedIn data or website of universities) (Park et al., 2019; Yuret, 2017).
- **Collection of CVs and ORCID** (Aman, 2018b; Bohannon, 2017; Cañibano et al., 2008; Lawson & Shibayama, 2015; Li & Tang, 2019; Mejias & Eyre, 2019; Yan et al., 2020; Z. Zhao et al., 2020)

# Two groups of literature in scientific mobility (scholarly migration)



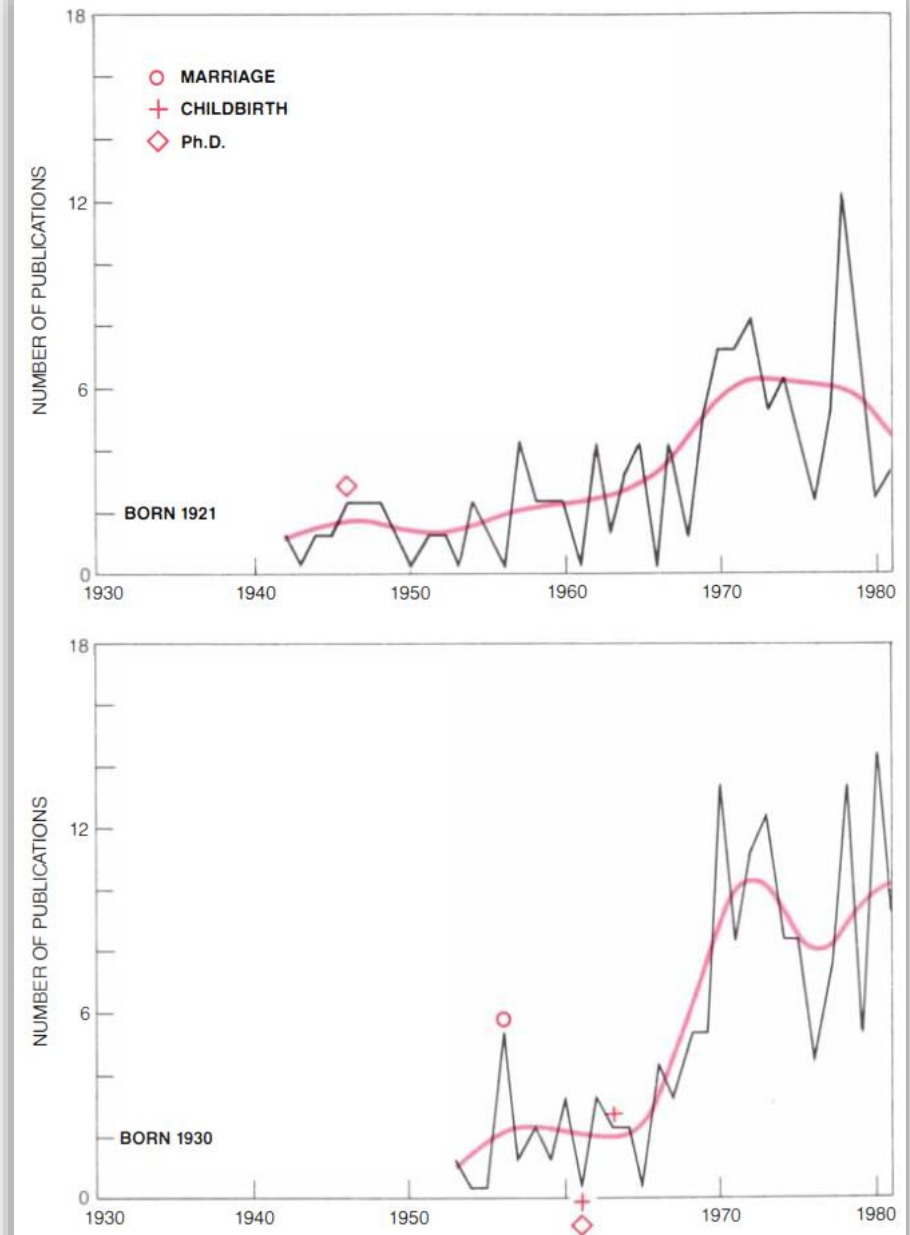
- **Bibliometric data** (Alburez-Gutierrez et al., 2019, Kashyap et al. 2022, Aman, 2018a; Laudel, 2003; Moed et al., 2013; Moed & Halevi, 2014)
- Divided based on the **study focus** to two groups.
  - Focuses on the **geographic scale of academic mobility**.
    - Internal or international migration to/from a country (see the case of Russia (Subbotin & Aref, 2021), Mexico (Miranda-González et al., 2020), and Germany (X. Zhao et al., 2021))
    - Others have focused on the global mobility of scholars (Chinchilla-Rodríguez, Miao, et al., 2018; Czaika & Orazbayev, 2018; Robinson-Garcia et al., 2019).
  - Focuses on **(dis)advantages of mobility**, or its contributions to the field, knowledge transfer, institutional, national or global productivity and innovation.
  - Some research has focused on the performance and impact of mobile scientists or the so called “**mover’s advantage**” (Aman, 2020; Bernstein et al., 2018; Franzoni et al., 2014; Halevi et al., 2016; J. A. Jacobs & Mizrachi, 2020). **Downsides** of scholarly mobility and costs that academics bear by leaving an academic context for another (Ackers & Gill, 2005; Schaer et al., 2017, 2020) or the potentials for (in)stability of scientific collaborations or difficulties of finding a job during or after mobility (Baruffaldi & Landoni, 2012; Z. Zhao et al., 2020).
  - **Policy changes** and how they can inspire (ippedico, 2021) or inhibit mobility of the general population and more specifically academics (Chinchilla-Rodríguez, Bu, et al., 2018; Sugimoto et al., 2017).

## A very innovative interview set-up

- Using bibliometric data alongside demographic life events

Cole, J. R., & Zuckerman, H. (1987). Marriage, Motherhood and Research Performance in Science. *Scientific American*, 256(2), 119–125. JSTOR.

<https://doi.org/10.1038/scientificamerican0287-119>



LOWER RATES OF PUBLICATION in the early part of a career are characteristic of both married men and single women. The publication profile of a distinguished woman biologist (*top*) who never married shows the same pattern of oscillations and an overall increase as the graphs of women who married and had children. The same pattern can be seen in the profile of an eminent male chemist (*bottom*). He published at a much slower pace when his children were young, although his domestic responsibilities were minimal.

# Bibliometric data sources and software packages

- **Clarivate's Web of Science (WoS)**
- **Elsevier's Scopus**
- **Dimensions**
- **Google Scholar**
- **Crossref**
- **Unpaywall**
- **Semantic scholar**
- **Microsoft Academic→ OpenAlex**
- **JSTOR Data for Research**
- **PubMed**

## API, software and packages:

- Scopus (rscopus package r),
- WOS (wosR package in R),
- OpenAlex (based on MAG),
- ORCID API,
- Dimensions (API),
- Google Scholar (publish perish, scholar package R),
- Bibliometrix (package in R),
- VOSviewer,
- CitenetExplorer
- CiteSpace

<sup>1</sup>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. [https://doi.org/10.1162/qss\\_a\\_00112](https://doi.org/10.1162/qss_a_00112)

# Calculation and use of demographic indicators using bibliometric data

- **Measures:** Net Migration Rate, Crude Migration Intensity, Migration Effectiveness Index, Aggregate Net Migration Rate<sup>5,10</sup>:



$$NMR_{i,t} = 1000 \times \frac{I_{i,t} - E_{i,t}}{N_{i,t}}$$

$$CMI_{i,t} = 100 \times \frac{M_{i,t}}{\sum_i N_{i,t}}$$

$$MEI_{i,t} = 100 \times \frac{\sum_i |I_{i,t} - E_{i,t}|}{\sum_i (I_{i,t} + E_{i,t})}$$

$$ANMR_{i,t} = 100 \times \frac{0.5 \sum_i |I_{i,t} - E_{i,t}|}{\sum_i N_{i,t}}$$

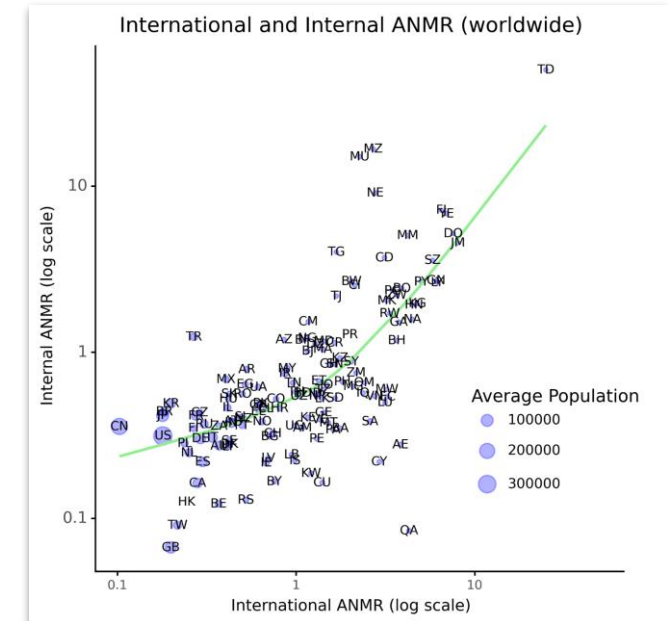
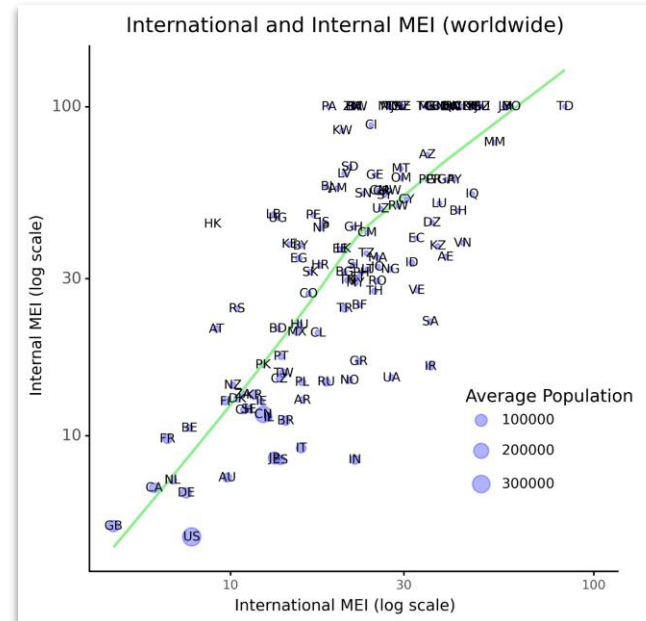
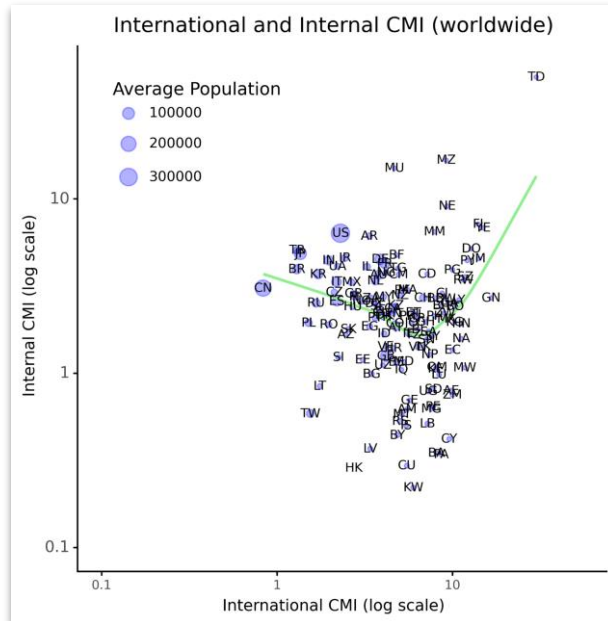
$$CMI_{i,t} = 100 \times \frac{ANMR_{i,t}}{MEI_{i,t}}$$

$I_{i,t}$  = Incoming scholars

(to a given region/country in a given year)

$E_{i,t}$  = Outgoing scholars (same as  $M_{i,t}$ )

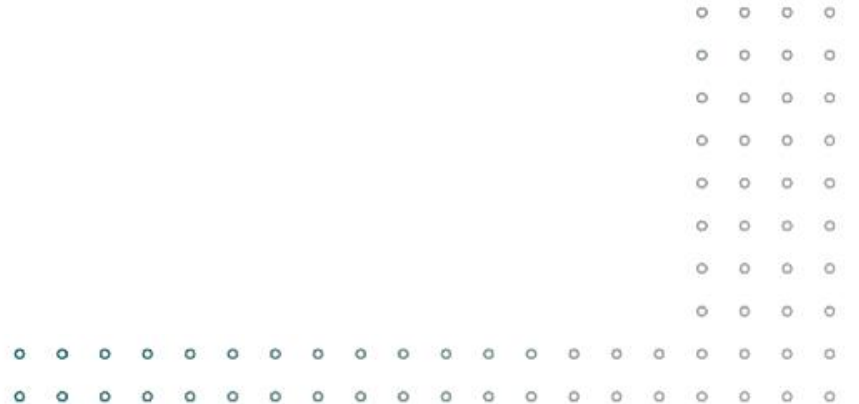
$N_{i,t}$  = Population of scholars



## Other relevant demographic and [only a few] scientometrics measures

- N and population of active scholars
- N of institutions (how large is the science system of a country)
- Stock and flows of scholars
- Academic birth
- Exiting academia (e.g., mortality and prospects of survival)
- Academic kinship (mentorship ties and supervision, collaboration)
- N of publications,
- N of co-authors,
- Citations,
- 3-year citations,
- FSS (fractional scientific strength)



[illegible]

# Data quality and need for organisation name disambiguation



authorsid	authorswithaffiliations	year	country
23494432900	Zagheni, E., Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany	2020	DEU
23494432900	Zagheni, E., Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany	2020	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research Rostock, Germany	2020	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Germany	2020	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Germany	2020	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Rostock, Germany	2020	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Rostock, Germany	2019	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Konrad-Zuse-Straße 1, Rostock, 18057, Germany	2019	DEU
23494432900	<b>Zagheni, E.</b>	<b>2019</b>	
23494432900	Zagheni, E., Department of Sociology, University of Washington, Seattle, United States, Max Planck Institute for Demographic Research, Rostock, Germany	2019	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Germany	2019	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Rostock, Germany	2019	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Germany	2019	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, Rostock, 18057, Germany	2019	DEU
23494432900	<b>Zagheni, E., Qatar Computing Research Institute, Doha, Qatar</b>	<b>2019</b>	<b>QAT</b>
23494432900	<b>Zagheni, E., Max Planck Institute for Demographic Research</b>	<b>2019</b>	
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Rostock, Germany	2018	DEU
23494432900	Zagheni, E., Max Planck Institute for Demographic Research, Rostock, Germany	2018	DEU
23494432900	<b>Zagheni, E., University of Washington, Max Planck Institute for Demographic Research, United States</b>	<b>2018</b>	<b>USA</b>
23494432900	<b>Zagheni, E., University of Washington and Max Planck Institute for Demographic Research, United States</b>	<b>2018</b>	<b>USA</b>
23494432900	<b>Zagheni, E.</b>	<b>2017</b>	
23494432900	Zagheni, E., Department of Sociology, University of Washington, Seattle, 211 Savery Hall, Box 353340, Seattle, WA 98195-3340, United States	2017	USA
23494432900	Zagheni, E., University of Washington, Seattle, United States	2017	USA
23494432900	Zagheni, E., University of Washington, United States	2017	USA
23494432900	Zagheni, E., University of Washington, Seattle, United States	2016	USA
23494432900	Zagheni, E., Department of Sociology, University of Washington at Seattle, 211 Savery Hall Box 353340, Seattle, WA 98195-3340, United States	2016	USA
23494432900	Zagheni, E., University of Washington, Seattle, WA, United States	2015	USA
23494432900	Zagheni, E., Department of Sociology, University of Washington, United States	2015	USA
23494432900	Zagheni, E., University of Washington, Seattle, United States	2015	USA
23494432900	Zagheni, E., Department of Sociology, University of Washington, Seattle, United States	2015	USA
23494432900	Zagheni, E., Department of Sociology, University of Washington, Seattle, WA, United States	2015	USA

# A publication on publisher's website (left) and Scopus (right)



Leaving Home and Entering the Housing Market

William A V Clark, Clara H Mulder

First Published See <https://doi.org/10.1068/a3315>

Article information

Abstract

We use a multinomial choice model of owning a home, owning a trailer, or renting to examine the housing-market entry of young adults in the USA after they have left the parental home. We also model the choice between renting independently and sharing with roommates. We show that the likelihood of becoming an independent actor in the housing market is closely related to the size and regional location of the housing market. The young adult's resources are an important influence on housing-market entry. Parents' resources seem to be less important as a factor in housing selection. Whereas trailer ownership is more common among the less well educated, couples, and those leaving home to live in the South or in rural areas, sharing is typical for the younger nest leavers, singles, and those leaving home to live in the cities.

References

Avery, R, Goldscheider, F, Speare, A, 1992, "Feathered nest/gilded cage: Parental income and leaving home in the transition to adulthood" *Demography* 29 375-388  
[Google Scholar](#) | [Crossref](#) | [Medline](#) | [ISI](#)

Börsch-Supan, A, 1986, "Household formation, housing prices, and public policy impacts" *Journal of Public Economics* 30 145-164  
[Google Scholar](#) | [Crossref](#) | [ISI](#)

Buck, N, Scott, J, 1993, "She's leaving home: But why? An analysis of young people leaving the parental home" *Journal of Marriage and the Family* 55 863-874  
[Google Scholar](#) | [Crossref](#) | [ISI](#)

4 056 Impact Factor  
5-Year Impact Factor 4.934  
Journal Indexing & Metrics

Altmeteric 6

Recommendations

PDF

Help

Explore More

Download PDF

Privacy

Scopus - Document details - Leaving Home and Entering the Housing Market

Search

Author profile preview

Mulder C.H.

Univ. of California, Los Angeles, Department of Geography, Los Angeles, CA 90095-1524, 1255 Bunche Hall, Box 951524, United States

View full profile Save to list

Recent documents

Family Complexity and Parents' Migration: The Role of Repartnering and Distance to Non-Resident Children  
*European Journal of Population*, 2021

Measuring migration motives with open-ended survey data: Methodological and conceptual issues  
*Population, Space and Place*, 2021

Migration after union dissolution in the United States: The role of non-resident family  
*Social Science Research*, 2021

View all documents

Metrics

126 Documents 42 h-index 4,345 Citations by 2357 documents

Leaving home and entering

Clark W.A.V.\* Mulder C.H.\*

Save all to author list

\* Univ. of California, Los Angeles, Department of Geography, Los Angeles, CA 90095-1524, United States

50 Citations in Scopus 25 Views count



<https://orcid.org/0000-0002-7260-524X>

Is this you? [Sign in to start editing](#)

Name  
**Jan Mendling**

Biography

This author profile is generated by Scopus [Learn more](#)

**Mendling, Jan**

Mendling, Jan ; Jan, Mendling ; Mendling, J.  
Wirtschaftsuniversität Wien, Vienna, Austria

#### Affiliation history

- 2004 - 2022 [Wirtschaftsuniversität Wien](#), Vienna, Austria
- 2008 - 2022 [Humboldt-Universität zu Berlin](#), Berlin, Germany
- 2013 - 2022 [Univerza v Ljubljani](#), Ljubljana, Slovenia
- 2021 [KU Leuven](#), 3000 Leuven, Belgium
- 2008 - 2020 [Universität Wien](#), Vienna, Austria
- 2016 - 2017 [Institute for Information Business](#), Vienna, Austria
- 2016 [Vrije Universiteit Amsterdam](#), Amsterdam, Netherlands
- 2007 - 2009 [Queensland University of Technology](#), Brisbane, Australia

#### Subject Areas

Computer Science • Business, Management and Accounting • Mathematics • Decision Sciences • Engineering • Arts and Humanities • Biochemistry, Genetics and Molecular Biology • Psychology • Social Sciences • Economics, Econometrics and Finance • Medicine • Chemical Engineering • Nursing

[Show less author info](#)

 6503908779  <https://orcid.org/0000-0002-7260-524X>  Is this you? [Connect to Mendeley account](#)

# Need for author name disambiguation

Out[43]:

	AUTHOR_ID	PK_AUTHORS	LAST_NAME	FIRST_NAME	INDEXED_NAME
342531	6503908779	5159491198	Jan	Mendling	Jan M.
1230731	6503908779	12326655929	Jan	Mendling	Jan M.
8960149	6503908779	15956230109	Jan	Mendling	Jan M.
12752651	6503908779	20665874236	Jan	Mendling	Jan M.
13345007	6503908779	5817946947	Jan	Mendling	Jan M.

In [44]: papers\_new[papers\_new["AUTHOR\_ID"]==6503908779][["AUTHOR\_ID","PK\_AUTHOR"]]

Out[44]:

	AUTHOR_ID	PK_AUTHORS	LAST_NAME	FIRST_NAME	INDEXED_NAME
62161	6503908779	15391353604	Mendling	NaN	Mendling J.
82568	6503908779	20744642022	Mendling	Jan	Mendling J.
122612	6503908779	2725380135	Mendling	Jan	Mendling J.
128880	6503908779	5107144514	Mendling	Jan	Mendling J.
159362	6503908779	11448344549	Mendling	Jan	Mendling J.

Same author's first and last names are swapped in Scopus while author identification number is similar!

On Scopus' author disambiguation, see: [https://doi.org/10.1162/qss\\_a\\_00019](https://doi.org/10.1162/qss_a_00019)

# DATA QUALITY

- **Author disambiguation in Scopus:**  
98.1% of author profiles do not include publications written by someone else\*\*;  
94.4% of author profiles include all publications written by the author.
- **Organization name disambiguation:**  
Research Organization Registry (ROR) API\*\*\*
- **Subset of data with the highest quality:**  
Period:1996-2020; Type of publications: Articles and Reviews;  
→ 36+ Million publications for 16+ Million unique authors

\*\*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](https://doi.org/10.1162/qss_a_00019)

\*\*\* See for comparison and description of ROR: Akbaritabar, A. (2021). A quantitative view of the structure of institutional scientific collaborations using the example of Berlin. *Quantitative Science Studies*, 2(2), 753–777. [https://doi.org/10.1162/qss\\_a\\_00131](https://doi.org/10.1162/qss_a_00131)



## Limitations in use of bibliometric data (for demographic research)

- Data quality
  - e.g., Scientific entity (e.g., authors, or institutions) **name disambiguation** (Tekles & Bornmann, 2020; Wu & Ding, 2013, Akbaritabar, 2021).
- Higher level **epistemic questions** need be addressed while repurposing these data for demographic research (Laudel, 2003; Moed et al., 2013; Moed & Halevi, 2014)
  - e.g., assigning the **country of affiliation in the first publication as the country of origin** for academic mobility is prone to error since that could simply be the country of graduation.
  - First publication year as **academic birth**
- There is a **publication delay** that can hinder proper identification of the mobility period.
- These data are limited to only those scholars who have **actively published**
- In indexed scholarly journals, so **coverage** may be incomplete (and over-represented by **WEIRD** countries).

## Future directions of research

- New services and methods to prepare **cleaner data**
- Increased availability through initiatives for **open access** to data.

### Potential future questions:

- How much of the talent circulation has happened “**within**” national borders versus “**between**” nations?
- Are there **migration corridors** connecting specific regions globally, for example between two specific regions across countries or in the same country, or systems of circulation that involve several countries or subregions?
- Do **scientific collaborations** among scholars facilitate their future mobilities?
- Do scholars have **different probabilities of being mobile** based on the trajectory of their collaborations during their scientific career?
- Complex interactions between processes related to migration of scientists and scholarly collaborations as well as **institutional settings and policies**.
- Finding **migration hubs or regions** with high concentration of academic labour or high attractiveness for future mobility that can inform policy.
- Evaluate theories **explaining migration through network tie formation** (Massey et al., 1993).



## Course assignments, live session and evaluation!

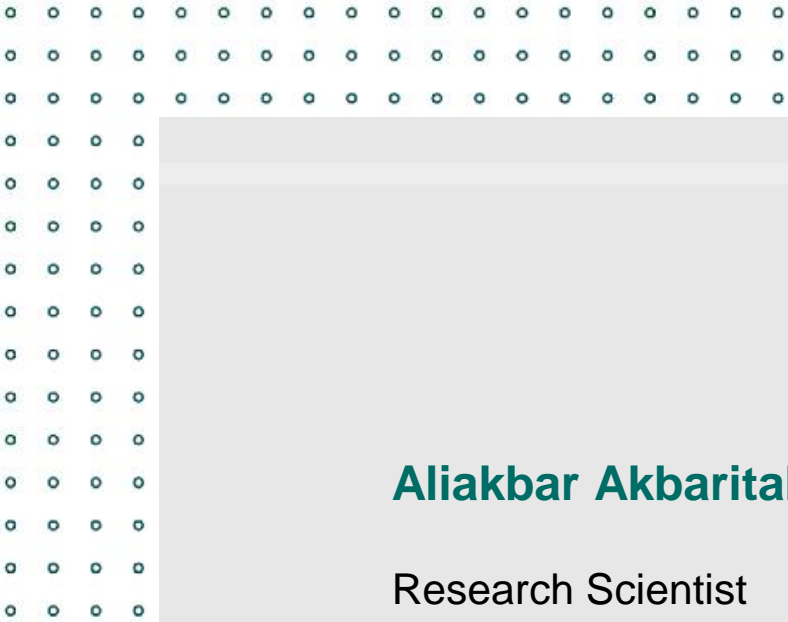
- Please read the instructions on the **ReadMe** file of the repository (the first page of the repository: [https://github.com/akbaritabar/BiblioDemography\\_IMPRS\\_PHDS\\_2022\\_IDEM187](https://github.com/akbaritabar/BiblioDemography_IMPRS_PHDS_2022_IDEM187)) carefully for instructions.
- You can **choose the tool and language with which you are familiar**, and the **topic** which is most interesting to you and do a **minimum of 1 assignment**. Bring your responses to the live session.
- **Live Q&A and discussion session (3 hours, 14:00-17:00 CET, hybrid format)**

We will have a brief discussion (and interactive quiz) on the reading materials. Then we will go over the scripts to solve the assignments together. Follow the instructions to setup one of Python/R/SQL/Excel on your laptop and bring your responses to the assignments.

### - **Course evaluation**

Evaluation of the course will be through a combination of a final exam (with two multiple choice questions), responses to assignments, and activity during the live session's discussions.

Please tweet with hashtag  
#BiblioDemography,  
a tribute to James W. Vaupel (1945-2022).  
Thanks Ilya and Jonas for bringing up Jim's  
labeling idea!



**Aliakbar Akbaritabar**

Research Scientist

[Akbaritabar@demogr.mpg.de](mailto:Akbaritabar@demogr.mpg.de)

<https://akbaritabar.github.io/>

<https://twitter.com/Akbaritabar>



**THANK YOU!**

