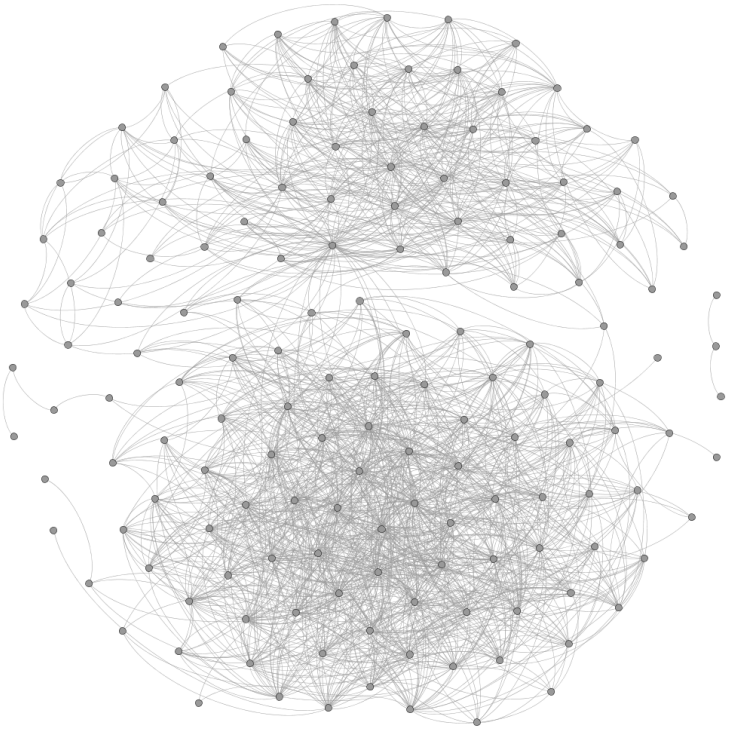


Bas Hofstra

Assistant Professor

Radboud University

Department of Sociology/ICS

bas.hofstra@ru.nl

twitter.com/bashofstra

www.bashofstra.com

Radboud University

Webscraping science and science network data

Bas Hofstra

May 2023

Radboud University

Content

Webscraping science and science network data

- Theory *(15 mins)*
  - Science data
  - Networks
  - Sources
  - Tutorial intro
- Tutorial

Science data

Why webscrape science data?

- Science is an inherently social enterprise
  - "Science" == co-evolution of *social aspects* and *production*
    - Social dynamics == collaboration, topic choices, academic job placements
      - I.e., "humans" in science
    - Production dynamics == publication, innovation, grants, patents
      - I.e., "knowledge" in science

Science data

Studied through “science data”

ScienceAdvances

Current Issue | First release papers | Archive | About

HOME > SCIENCE ADVANCES > VOL. 1, NO. 1 > SYSTEMATIC INEQUALITY AND HIERARCHY IN FACULTY HIRING NETWORKS

RESEARCH ARTICLE | NETWORK SCIENCES

Systematic inequality and hierarchy in faculty hiring networks

AARON CLAUSET, SAMUEL ARBESMAN, AND DANIEL B. LARREMORE Authors Info & Affiliations

SCIENCE ADVANCES • 12 Feb 2015 • Vol 1, Issue 1 • DOI: 10.1126/sciadv.1400005

Science

Current Issue | First release papers | Archive | About

HOME > SCIENCE > VOL. 359, NO. 6379 > SCIENCE OF SCIENCE

REVIEW

Science of science

SANTO FORTUNATO, CARL T. BERGSTROM, KATY BÖRNER, JAMES A. EVANS, DIRK HELBING, STAŠA MILOJEVIĆ, ALEXANDER M. PETERSEN, FILIPPO RADICCHI, ROBERTA SINATRA, [...], AND ALBERT-LÁSZLÓ BARABÁSI +4 authors Authors Info & Affiliations

nature

Explore content | About the journal | Publish with us

nature > letters > article

Letter | Published: 13 February 2019

Large teams develop and small teams disrupt science and technology

Lingfei Wu, Dashun Wang & James A. Evans

Nature **566**, 378–382 (2019) | Cite this article

PNAS

ARTICLES | FRONT MATTER | AUTHORS | TOPICS +

Access Provided By Radboud University

RESEARCH ARTICLE | SOCIAL SCIENCES

The Diversity–Innovation Paradox in Science

Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, +2, and Daniel A. McFarland Authors Info & Affiliations

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved March 16, 2020 (received for review September 5, 2019)

April 14, 2020 | 117 (17) 9284-9291 | https://doi.org/10.1073/pnas.1915378117

Radboud University

Science data

What is science data?

- …all data related to science

Science data

What is science data?

- …all data related to science

ELSEVIER

Research Policy
Volume 51, Issue 1, January 2022, 104411

Gendered knowledge in fields and academic careers

Lanu Kim [a], Daniel Scott Smith [b], Bas Hofstra [c], Daniel A. McFarland [b]

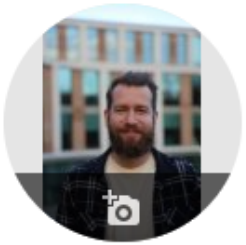
[a] School of Humanities and Social Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Daejeon 34141, South Korea
[b] Graduate School of Education, Stanford University, 520 Galvez Mall, Stanford, CA 94305, USA
[c] Department of Sociology, Radboud University, Thomas van Aquinostraat 4, 6525 GD Nijmegen, the Netherlands

Received 29 June 2020, Revised 4 October 2021, Accepted 18 October 2021, Available online 2 November 2021, Version of Record 2 November 2021.

Staff
Research Institute RSCR
The Netherlands' Life Course Survey
Family Survey Dutch Population
Socio-Cultural Developments
Norface Mifare Project
NIS2NL Project

Staff

Staff:	Expertise:
Batenburg, prof. dr. R. (Ronald)	Healthcare, labour market and healthcare professions and training
Begall, dr. K.H. (Katia)	Family, life course, labour market participation, division of household tasks and gender norms
Bekhuis, dr. H. (Hidde)	Welfare state, nationalism and sports
Berg, dr. L. van den (Lonneke)	Family, life course and transition to adulthood
Blommaert, dr. L. (Lieselotte)	Discrimination and inequality on the labour market
Eisinga, prof. dr. R.N. (Rob)	Methods of research and statistics
Gesthuizen, dr. M.J.W. (Maurice)	Poverty en social cohesion
Geurts, dr. P.G. (Nella)	Integration and migration
Glas, dr. S. (Saskia)	Islam, gender attitudes and sexuality
Hek, dr. M. van (Margriet)	Educational inequality, gender inequality, organizational sociology and culture
Hoekman, dr. R. H. A.(Remco)	Sports and policy sociology
Hofstra, dr. B. (Bas)	Diversity, inequality and innovation
Koops, dr. J. (Judith)	Family sociology and demography
Kraaykamp, prof. dr. G.L.M. (Gerbert)	Educational inequality, culture and health
Meuleman, dr. (Roza)	Culture and nationalism
Savelkoul, dr. M.J. (Michael)	Ethnic diversity, prejudice and social



Bas Hofstra

FOLLOWING

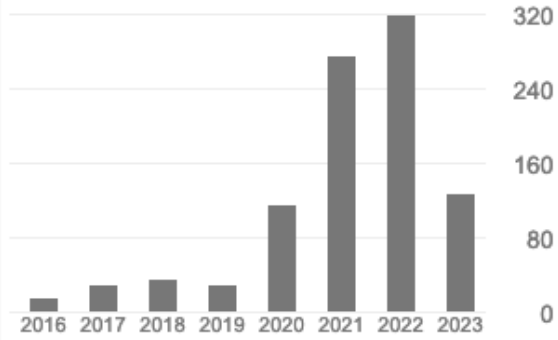
Assistant Professor, Radboud University
Verified email at ru.nl - Homepage

Sociology of science | Social networks | Online networks | Computational social science | Big data

TITLE	CITED BY	YEAR
The diversity–innovation paradox in science B Hofstra, VV Kulkarni, S Munoz-Najar Galvez, B He, D Jurafsky, ... Proceedings of the National Academy of Sciences 117 (17), 9284-9291	583	2020
Sources of Segregation in Social Networks: A Novel Approach Using Facebook B Hofstra, R Corten, F van Tubergen, NB Ellison American Sociological Review, 625-656	142	2017
Understanding the privacy behavior of adolescents on Facebook: The role of peers, popularity and trust B Hofstra, R Corten, F Van Tubergen Computers in Human Behavior 60, 611-621	68	2016
Does unemployment lead to isolation? The consequences of unemployment for social networks JJ Rözer, B Hofstra, ME Brashears, B Volker Social Networks 63, 100-111	32	2020
Predicting ethnicity with first names in online social media networks B Hofstra, NC de Schipper Big Data & Society 5 (1), 2053951718761141	32	2018
Who was first on Facebook? Determinants of early adoption among adolescents B Hofstra, R Corten, F Van Tubergen new media & society 18 (10), 2340-2358	27	2016
Learning in social networks: Selecting profitable choices among alternatives of uncertain profitability in various networks B Hofstra, R Corten, V Buskens Social Networks 43, 100-112	23	2015

Cited by

	All	Since 2018
Citations	966	897
h-index	10	9
i10-index	10	9



Public access

VIEW ALL

0 articles	9 articles
not available	available

Based on funding mandates

Co-authors

EDIT

Daniel McFarland
Professor of Education, Stanford...

Rense Corten
Associate Professor, Sociology, ...

PNAS
Vol. 117 | No. 17

Significance
Abstract
Acknowledgments
Supporting Information
References

References

1 R. K. Merton, *The Sociology of Science* (University of Chicago Press, Chicago, IL, 1973).
Google Scholar | Open URL

2 M. S. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
Crossref | Google Scholar | Open URL

3 R. S. Burt, Structural holes and good ideas. *Am. J. Sociol.* **110**, 349–399 (2004).
Crossref | Google Scholar | Open URL

4 M. W. Nielsen et al., Opinion: Gender diversity leads to better science. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1740–1742 (2017).
Crossref | PubMed | Google Scholar | Open URL

5 S. E. Page, *The Diversity Bonus: How Great Teams Payoff in the Knowledge Economy* (Princeton University Press, Princeton, NJ, 2009).
Google Scholar | Open URL

6 S. T. Bell, A. J. Villado, M. A. Lukasik, L. Belau, A. L. Briggs, Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *J. Manage.* **37**, 709–743 (2011).
Google Scholar | Open URL

7 M. de Vaan, D. Stark, B. Vedres, Game changer: The topology of creativity. *Am. J. Sociol.* **120**, 1144–1194 (2015).
Crossref | Google Scholar | Open URL

boud University

Networks in science data

What are networks in science data?

- Many!

Networks in science data

Co-author networks

ELSEVIER

Research Policy

Volume 51, Issue 1, January 2022, 104411

Gendered knowledge in fields and academic careers

Lanu Kim [a], Daniel Scott Smith [b], Bas Hofstra [c], Daniel A. McFarland [b]

[a] School of Humanities and Social Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Daejeon 34141, South Korea

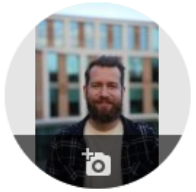
[b] Graduate School of Education, Stanford University, 520 Galvez Mall, Stanford, CA 94305, USA

[c] Department of Sociology, Radboud University, Thomas van Aquinostraat 4, 6525 GD Nijmegen, the Netherlands

Received 29 June 2020, Revised 4 October 2021, Accepted 18 October 2021, Available online 2 November 2021, Version of Record 2 November 2021.

Networks in science data

Co-author networks



Bas Hofstra FOLLOWING

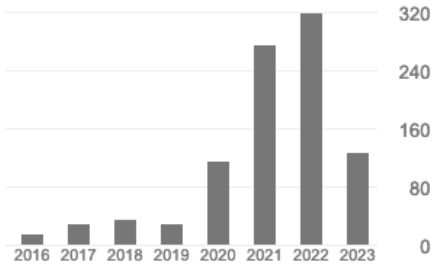
Assistant Professor, Radboud University
Verified email at ru.nl - Homepage

Sociology of science | Social networks | Online networks | Computational social science | Big data

TITLE	CITED BY	YEAR
The diversity–innovation paradox in science B Hofstra, VV Kulkarni, S Munoz-Najar Galvez, B He, D Jurafsky, ... Proceedings of the National Academy of Sciences 117 (17), 9284-9291	583	2020
Sources of Segregation in Social Networks: A Novel Approach Using Facebook B Hofstra, R Corten, F van Tubergen, NB Ellison American Sociological Review, 625-656	142	2017
Understanding the privacy behavior of adolescents on Facebook: The role of peers, popularity and trust B Hofstra, R Corten, F Van Tubergen Computers in Human Behavior 60, 611-621	68	2016
Does unemployment lead to isolation? The consequences of unemployment for social networks JJ Rözer, B Hofstra, ME Brashears, B Volker Social Networks 63, 100-111	32	2020
Predicting ethnicity with first names in online social media networks B Hofstra, NC de Schipper Big Data & Society 5 (1), 2053951718761141	32	2018
Who was first on Facebook? Determinants of early adoption among adolescents B Hofstra, R Corten, F Van Tubergen new media & society 18 (10), 2340-2358	27	2016
Learning in social networks: Selecting profitable choices among alternatives of uncertain profitability in various networks B Hofstra, R Corten, V Buskens Social Networks 43, 100-112	23	2015

Cited by

	All	Since 2018
Citations	966	897
h-index	10	9
i10-index	10	9



Public access VIEW ALL

0 articles | 9 articles
not available | available

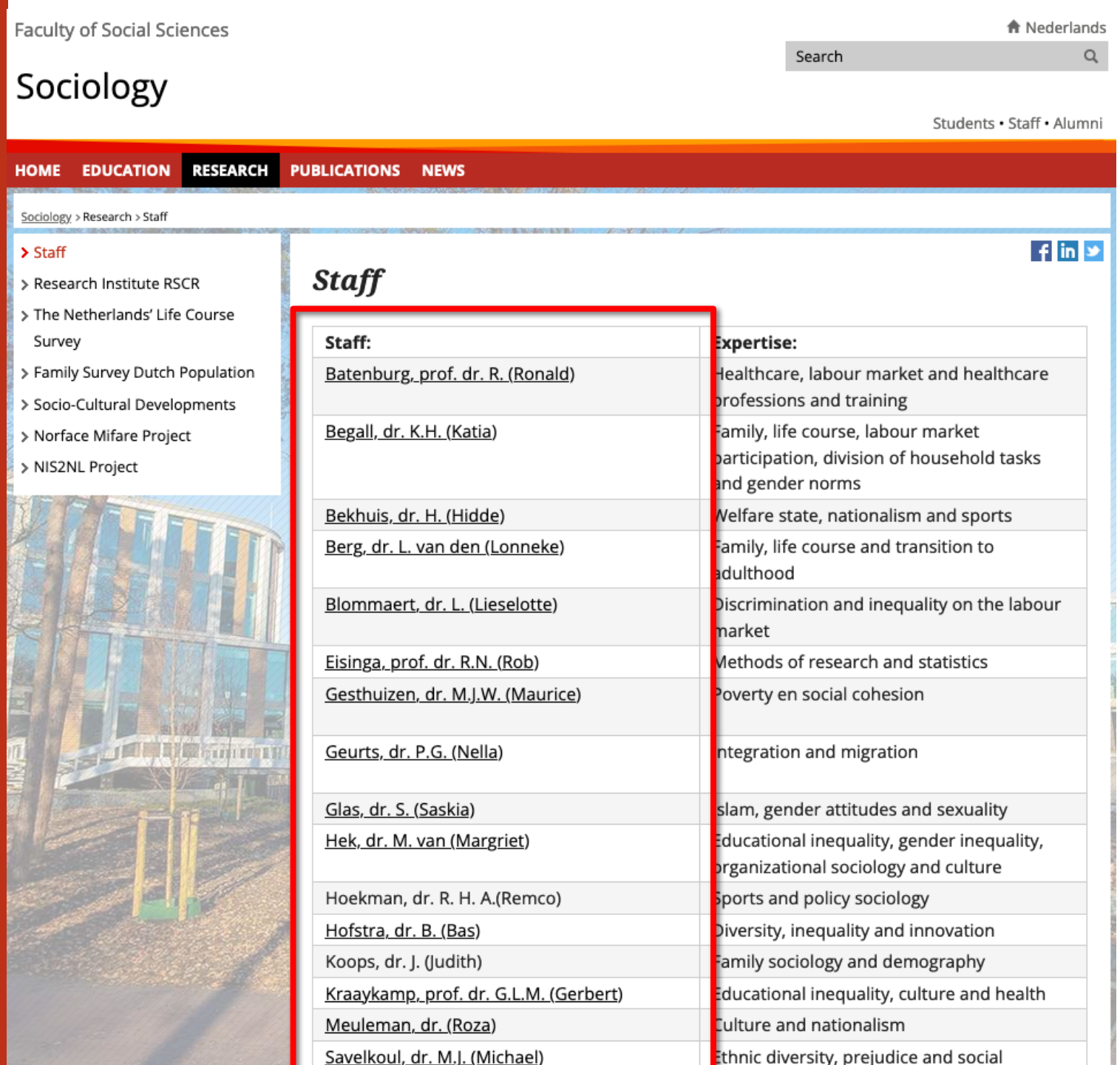
Based on funding mandates

Co-authors EDIT

Daniel McFarland
Professor of Education, Stanford...

Rense Corten
Associate Professor, Sociology, ...

Radboud University



Departmental networks

Radboud University

PNAS
Vol. 117 | No. 17

Significance
Abstract
Acknowledgments
Supporting Information
References

References

1 R. K. Merton, *The Sociology of Science* (University of Chicago Press, Chicago, IL, 1973).
Google Scholar | Open URL

2 M. S. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
Crossref | Google Scholar | Open URL

3 R. S. Burt, Structural holes and good ideas. *Am. J. Sociol.* **110**, 349–399 (2004).
Crossref | Google Scholar | Open URL

4 M. W. Nielsen et al., Opinion: Gender diversity leads to better science. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1740–1742 (2017).
Crossref | PubMed | Google Scholar | Open URL

5 S. E. Page, *The Diversity Bonus: How Great Teams Payoff in the Knowledge Economy* (Princeton University Press, Princeton, NJ, 2009).
Google Scholar | Open URL

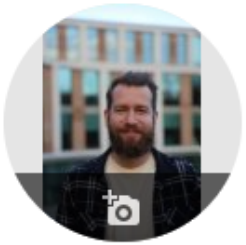
6 S. T. Bell, A. J. Villado, M. A. Lukasik, L. Belau, A. L. Briggs, Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *J. Manage.* **37**, 709–743 (2011).
Google Scholar | Open URL

7 M. de Vaan, D. Stark, B. Vedres, Game changer: The topology of creativity. *Am. J. Sociol.* **120**, 1144–1194 (2015).
Crossref | Google Scholar | Open URL

Citation networks

boud University

Hiring networks



Bas Hofstra

FOLLOWING

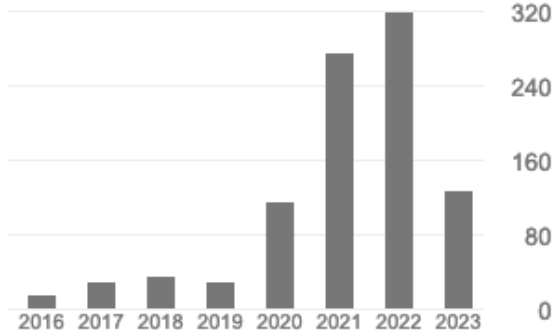
Assistant Professor, Radboud University
Verified email at ru.nl - Homepage

Sociology of science | Social networks | Online networks | Computational social science | Big data

TITLE	CITED BY	YEAR
The diversity–innovation paradox in science B Hofstra, VV Kulkarni, S Munoz-Najar Galvez, B He, D Jurafsky, ... Proceedings of the National Academy of Sciences 117 (17), 9284-9291	583	2020
Sources of Segregation in Social Networks: A Novel Approach Using Facebook B Hofstra, R Corten, F van Tubergen, NB Ellison American Sociological Review, 625-656	142	2017
Understanding the privacy behavior of adolescents on Facebook: The role of peers, popularity and trust B Hofstra, R Corten, F Van Tubergen Computers in Human Behavior 60, 611-621	68	2016
Does unemployment lead to isolation? The consequences of unemployment for social networks JJ Rözer, B Hofstra, ME Brashears, B Volker Social Networks 63, 100-111	32	2020
Predicting ethnicity with first names in online social media networks B Hofstra, NC de Schipper Big Data & Society 5 (1), 2053951718761141	32	2018
Who was first on Facebook? Determinants of early adoption among adolescents B Hofstra, R Corten, F Van Tubergen new media & society 18 (10), 2340-2358	27	2016
Learning in social networks: Selecting profitable choices among alternatives of uncertain profitability in various networks B Hofstra, R Corten, V Buskens Social Networks 43, 100-112	23	2015

Cited by

	All	Since 2018
Citations	966	897
h-index	10	9
i10-index	10	9



Public access

VIEW ALL

0 articles	9 articles
not available	available

Based on funding mandates

Co-authors

EDIT

Daniel McFarland
Professor of Education, Stanford...

Rense Corten
Associate Professor, Sociology, ...

Networks in science data

What are networks in science data?

- Many:
  - Co-author networks
  - Project networks
  - Faculty hiring networks
  - Migration networks
  - Citation networks
  - Semantic networks
  - And many more

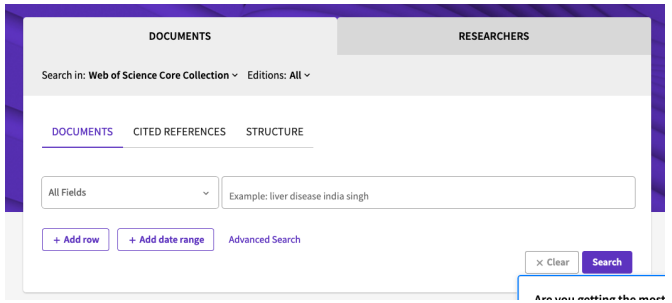
Sources: science data

Google Scholar

Articles Case law

ProQuest
Part of Clarivate
Products & Services
Librarians

Dissertations & Theses



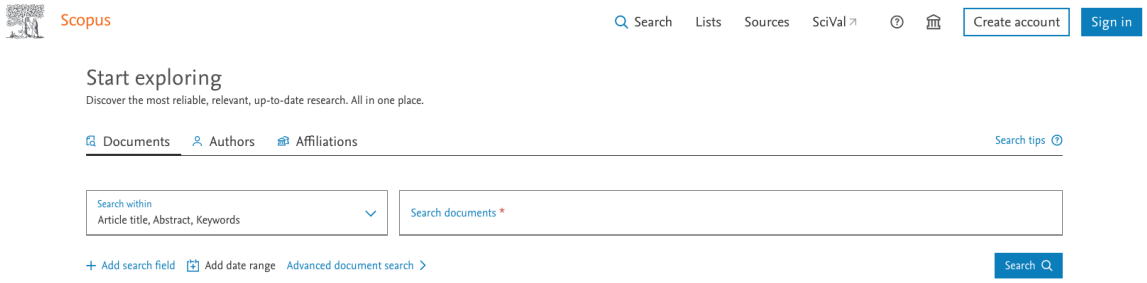
WoS

OpenAlex

Look up entities:

Works (like papers and datasets) are products of research.
Authors create works.
Sources (like journals and repositories) host works.
Institutions (like universities) are affiliated with works.
Concepts describe the topics of works.
Publishers organizations that distribute works.

SEARCH I'M FEELING LUCKY



Scopus

Radboud University

Tutorial: departments & co-authors

- Which colleagues in academic departments co-author and why?

- Let's built a dataset from scratch to answer this Q!

- What do we need?
  - Departments?
  - Papers?
  - Co-authorships?
  - Matching?