

# SCIENCE INEQUALITIES

Diego Kozlowski

Faculty of Science, Technology and Medicine, University of Luxembourg  
May, 2021

## AIM AND SCOPE

- ▶ As a social activity, scientific output is not independent from the society in which it is produced [1].
- ▶ Inequality plays a major role in the distribution of prestige, resources and impact among scientists.
- ▶ This project aims to a better understanding of biases and inequalities in science.

## METHODOLOGY

This project is **data-driven**, which means that we aim at a **quantitative** understanding, without losing sight of the **context of data**.

We leverage on:

- ▶ Bibliometric databases (Web of Science, Scopus)
- ▶ Natural Language Processing, to do large-scale analysis over scientific communications (Word embeddings, Topic Modelling)
- ▶ Network analysis, to understand collaboration and citation patterns (Graph Neural Networks, Social Network Analysis)
- ▶ A commitment for clear and engaging scientific communication (Data visualization, Shiny Apps)

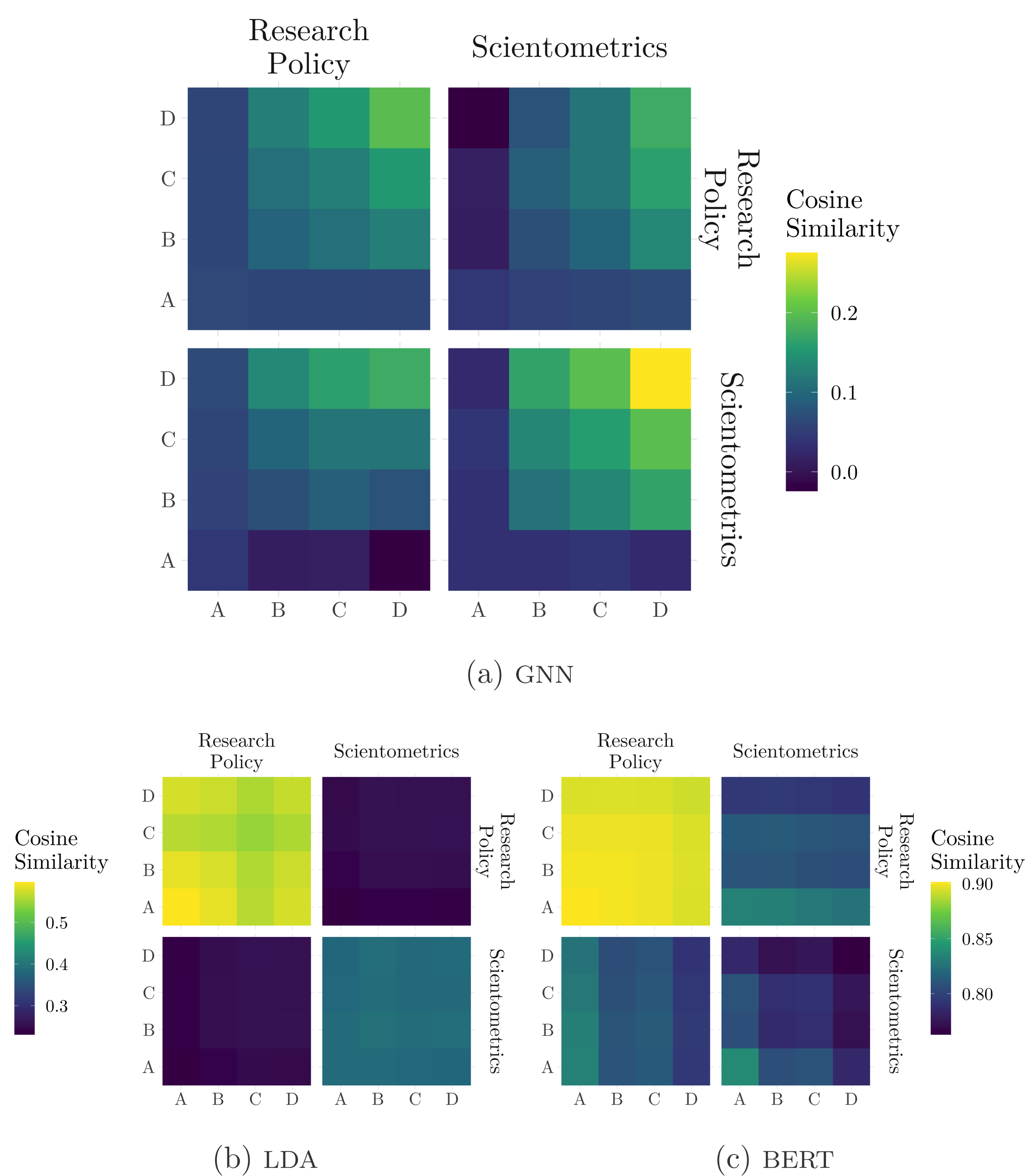
Nevertheless, we prioritize the best tool for answering each research question, regardless if it is the newest method or an old proven technique.

## Lines Of Work

### 1. EMBEDDED BIAS

- ▶ We problematize the use of Deep Learning embeddings for the study of science. We focus on how different models encode science inequalities, and should be used/avoided for specific research questions. [2]

FIGURE 1



We studied how Graph Neural Networks and text-embeddings capture the centrality of international collaborations

## SUPERVISORY TEAM

Main advisor: Prof. Andreas Zilian; co-advisors: Dr. Jun Pang, Dr. Jennifer Dusdal

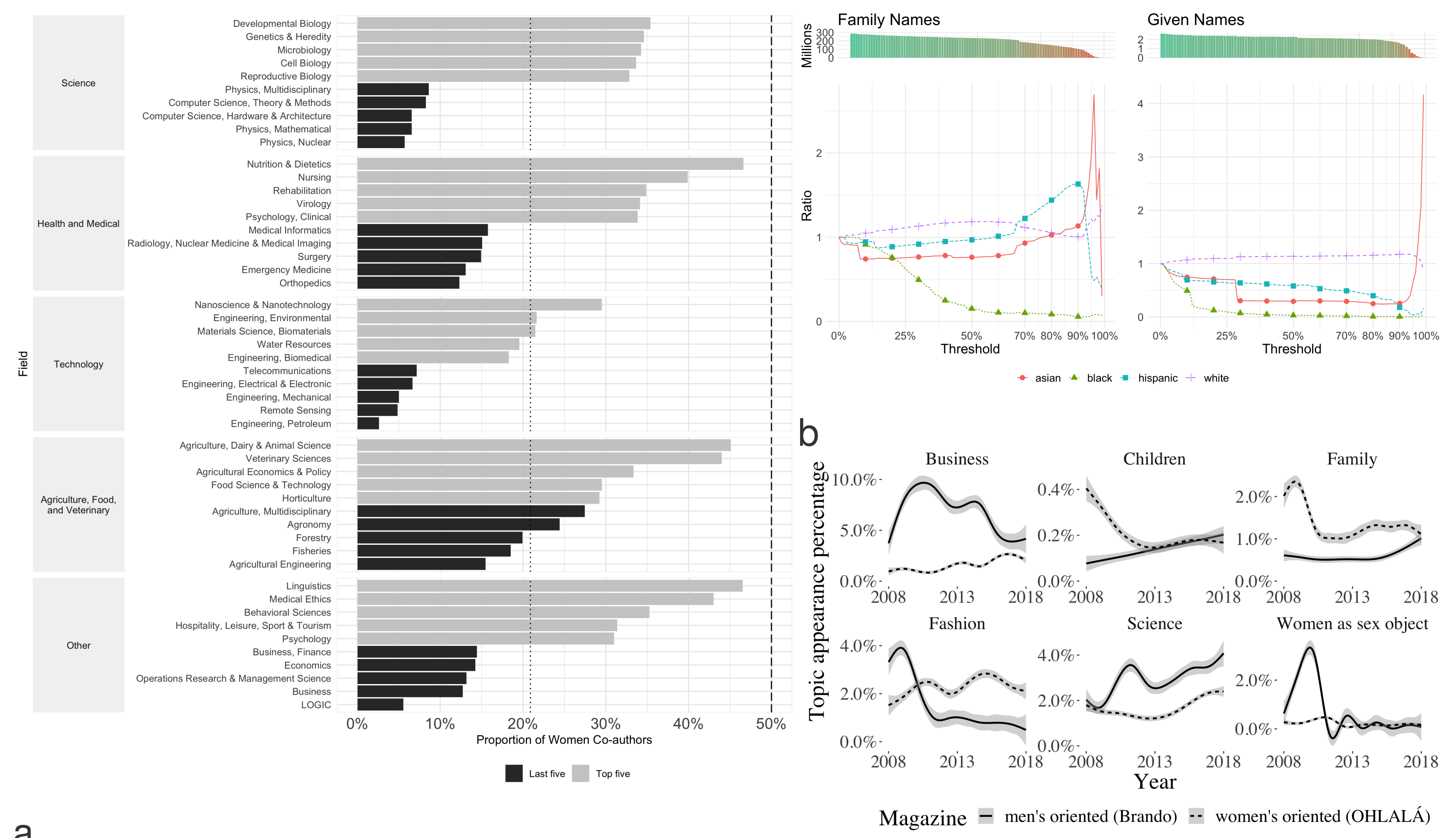
## REFERENCES

- [1] Cassidy R. Sugimoto. "Scientific success by numbers". In: *Nature* 593.7857 (May 2021), pp. 30–31. doi: 10.1038/d41586-021-01169-7. URL: <https://doi.org/10.1038/d41586-021-01169-7>.
- [2] Diego Kozlowski, Jennifer Dusdal, Jun Pang, and Andreas Zilian. "Semantic and relational spaces in science of science: deep learning models for article vectorisation". In: *Scientometrics* (May 2021). doi: 10.1007/s11192-021-03984-1.
- [3] Diego Kozlowski, Dakota S. Murray, Alexis Bell, Will Hulsey, Vincent Larivière, Thema Monroe-White, and Cassidy R. Sugimoto. "Avoiding bias when inferring race using name-based approaches". In: (Apr. 14, 2021). arXiv: 2104.12553 [cs.CL].
- [4] Diego Kozlowski, Gabriela Lozano, Carla M. Felcher, Fernando Gonzalez, and Edgar Altszyler. "Gender bias in magazines oriented to men and women: a computational approach". In: (Nov. 24, 2020). arXiv: 2011.12096 [cs.CL].
- [5] Diego Kozlowski, Viktoriya Semeshenko, and Andrea Molinari. "Latent Dirichlet allocation model for world trade analysis". In: *PLOS ONE* 16.2 (Feb. 2021), e0245393. doi: 10.1371/journal.pone.0245393.

## 2. RACE & GENDER INEQUALITIES

- ▶ We study scientific publication patterns by race and gender intersection in US research community.
- ▶ As we work with bibliometric databases, we first need to infer author's race to do our analysis. The first piece of this larger project focuses on the bias introduced by different types of name-based racial inference methods. [3]
- ▶ In a second project on this track, we study gender inequalities in Germany's scientific publications. (work in progress, with Jennifer Dusdal and Justin Powell)
- ▶ On a third track, we study non-scientific publications. We want to understand the variation of discourse in men-oriented and women-oriented magazines. [4]

FIGURE 2



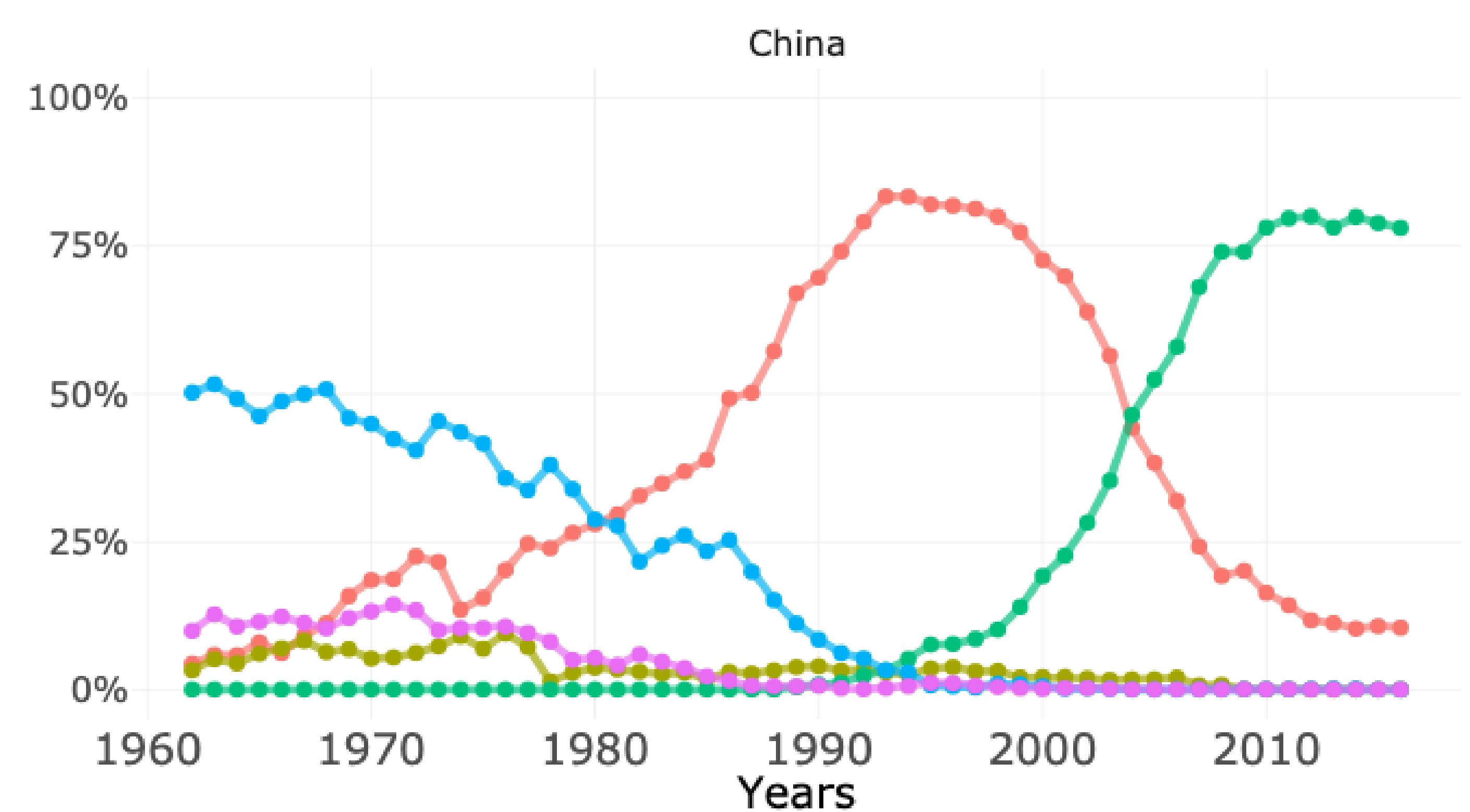
a

a. The proportion of women authors by Discipline and Field in Germany, b. The effect of thresholding on name-based estimations, c. The topics of discussion in women-oriented and men-oriented magazines

## 3. INTERNATIONAL DIVISION OF SCIENCE

- ▶ The international division of scientific labor is deeply related with economic inequality around the globe. Therefore, a comprehensive analysis of science inequality should consider this aspect.
- ▶ A first step was towards a better understanding of the international division of labor, through finding the latent dimensions of world trade [5]
- ▶ next steps will focus on the relation between economic inequality and scientific inequality.

FIGURE 3



The evolution of China's exports, from rice & cotton, to lightweight industry, and to microprocessors

## ACKNOWLEDGEMENT

The Doctoral Training Unit **Data-driven computational modelling and applications** (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781). <https://driven.uni.lu>