Informetric methods for studying diversity of the scientific workforce

Towards a State-of-the-Art

Nicolas Robinson-Garcia
Carmen Corona-Sobrino
Zaida Chinchilla-Rodríguez
Daniel Torres-Salinas
Rodrigo Costas

Background and rationale

★ Genuine interest since the conception of scientometrics on the quantitative study of researchers' activity

- Productivity inequality - Lotka, 1927
- Social stratification - Cole & Cole, 1973
- Hiring & Promotion - Gingras, 2020

Information Studies

↓

Sociology of Science

↓

Evaluative Bibliometrics

★ Cautionary remarks for individual metrics in evaluation

- Partial indicators - Martin & Irvine, 1983
- Unit of analysis - Glänzel & Wouters, 2013
- Conceptual flaws - Waltman & van Eck, 2012
- Coverage - van Leeuwen et al, 2001

Background and rationale

★ **Diversity**, **inclusion** and **openness** are the new buzzwords in science policy

- Diversity of outputs and outcomes
- Diversity of knowledges
- Diversity of people/researchers

I believe in a research culture that recognises a diversity of contributions to science and society; that celebrates high quality and impactful research; and that values sharing, collaboration, integrity and engagement with society, transmitting knowledge from generation to generation.

Mariya Gabriel

Commissioner for Innovation, Research, Culture, Education and Youth

- Acknowledge and valorise the diversity in research roles and careers, including roles outside academia. Value the skills (including open science skills), competences and merits of individual researchers, but also recognise team science and collaboration.
- Ensure gender equality, equal opportunities and inclusiveness. Consider gender balance, the gender dimension, and take into account diversity in the broader sense (e.g. racial or ethnic origin, sexual orientation, socio-economic, disability) in research teams at all levels, and in the content of research and innovation.

DORA

CoARA

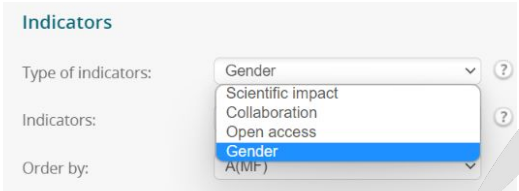
Background and rationale

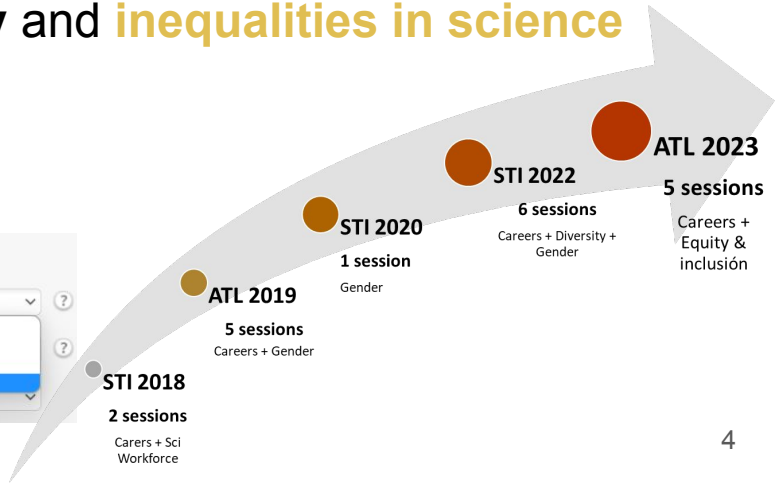
★ **Diversity**, **inclusion** and **openness** are the new buzzwords in science policy

- Diversity of outputs and outcomes
- Diversity of knowledges
- Diversity of people/researchers

★ There is growing sensitivity towards **diversity** and **inequalities in science**

- Introduction of new metrics
- New studies quantifying inequalities





Scope of the review

- ★ Many reviews on individual level indicators, but **focused on performance rather than context**
- ★ The raise of **big data** and **machine learning algorithms** provides a great opportunity to study the dynamics of science
- ★ The launch of **new data sources** provides new perspectives and ways into which science can be studied and understood

What information can metrics provide to study individuals' diversity in science?

Scope of the review

CRITERIA

- ★ Individual level metrics
- ★ Contextualize rather than assess - *No rankings or league tables!*

ASPECTS CONSIDERED

- ★ Data sources
- ★ Personal characteristics
- ★ Context
- ★ Team dynamics

Data sources

Technological and technical developments have been key to the expansion of individual level metrics. We highlight 4 main developments:

1. **Author name identification**
   - Databases' author profiles - Scopus AuthorID, WoS Researcher Profiles, Google Scholar Profiles, Dimensions IDs, CWTS disambiguation algorithm
   - Author registries - ORCID, CRIS Author profiles
2. **Metadata quality**
   - Linkages between fields
   - Interoperability between journals and author IDs
3. **Introduction of author-level features**
   - Funding
   - Contribution statements
   - Academic status
   - Affiliation history
4. **Gender / ethnicity identification algorithms**

Individual characteristics

Career stage	• 1st publication year a *de facto* standard to calculate academic age • PhD Year as an alternative when available to calculate academic age • Differences by fields noted, also career gaps not considered
Gender	• Binary approaches based on the use of gender identification algorithms with many approaches used here • Attempts made at measuring gender inequality at the publication level
Ethnicity/nationality	• Most studies use affiliation data to identify nationality or academic origin, mainly used in mobility studies • Surnames used to infer ethnicity, less literature in this area in Europe
Cultural identity	• Area less developed • Studies focus on language and country of origin • Publication and collaboration patterns by fields (esp. in SSH)

Context

Trajectory and career	• Indicators on geographic mobility types of scholars • Multiple affiliations with private sector • Transitions between sectors
Open practices	• Open Access • Data sharing • Open code/software

Funding	• Studies using funding information at the individual level from:
	  ○ ORCID
	  ○ Official data sources

Societal relevance	• Altmetrics - Discussions in social media, policy, etc. • Inclusion of non-academic outputs

Team dynamics

Author order	• 1st and last positions as *proxies* for scientific leadership • Papers using corresponding authors, but units of analysis ≠ individual • Studies on middle authors as *proxy* for technical roles
Collaboration	• Measured mainly through co-authorship • Indicators characterizing researchers based on the collaboration practices (e.g., churning of co-authors, academic age)
Disciplinary composition	• Studies measuring team interdisciplinarity as proxy to context or IDR outputs • Number of authors influenced by field ← Team Science
Contributions	• Types of researchers based on contribution statements • Number of contributions by author • Changing roles in international collaboration (see van Schalkwyk)

Conclusions

- ★ Computational developments have been key on their expansion
- ★ Unbalance on the aspects under study and sometimes lack of technical questioning (e.g., accuracy and biases in algorithms)
- ★ Metadata quality but also data access is key for the expansion of these metrics

Next steps | **Towards a research agenda on the study of diversity of the scientific workforce**

COMPARE
UGR PROJECT

Thank you for your attention

Questions, comments?

Project PID2020-117007RA-I00 funded by:

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

More on the COMPARE project at

https://compare-project.eu

UNIVERSITAS GRANATENSIS
PLUS ULTRA
1531