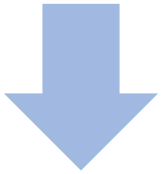
The effects of specialization on research careers

Nicolas Robinson-Garcia
Rodrigo Costas
Cassidy R. Sugimoto
Vincent Larivière
Tina Nane

Background

Increase of middle authors (Mongeon et al. 2016)

Shortening of their career length (Milojević et al. 2018)



Evaluation schemes

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Research questions

1. Is there a diversity of profiles in research careers?

2. Are scientists' career trajectories affected by their type of profile?
   – Impact and productivity
   – Gender

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Rationale

1. Modelling for predicting contribution data based on bibliometric variables
2. Prediction of authors' contributions at the paper level throughout their research career
3. Distinction of career stages
4. Identification of scientist archetypes at each career stage
5. Flows of scientists from one stage to the other by archetype
6. Distribution of productivity, impact and gender by archetype and career stage

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Data and Methods. Seed data set

Dataset:

- 70,694 publications from Plos journals from Medical and Life Sciences
- For each publication:
  - Contribution data
  - Bibliometric variables
- 347,136 distinct authors
  (Caron & van Eck, 2014)

Acronym	Definition	Source
WR	Wrote the manuscript	Plos data
AD	Analyzed the data	Plos data
CE	Conceived the experiments	Plos data
CT	Contributed with tools	Plos data
PE	Performed the experiments	Plos data
NC	Number of contributions	Plos data
PO	Author position in the paper	WoS data
AU	Total number of authors in the paper	WoS data
DT	Document type. Only articles and reviews are included	WoS data
YE	Number of years active as an academic based on year of first publication	WoS data
CO	Number of countries to which all authors are affiliated in the paper	WoS data
IN	Number of institutions to which all authors are affiliated in the paper	WoS data
PU	Average number of publications per year of the author at the time of publication of the paper	WoS data

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Data and Methods. Modelling

Definition:

*Bayesian Networks are **directed acyclic graphical models** where the nodes represent random variables and directed edges capture their dependence. Briefly speaking, a **Bayesian network offers a simple** and convenient **way of representing** a factorization of a **joint probability** mass function or density function **of a vector of random variables.***

Choi, 2015

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Modelling

Acronym	Definition	Source
WR	Wrote the manuscript	Plos data
AD	Analyzed the data	Plos data
CE	Conceived the experiments	Plos data
CT	Contributed with tools	Plos data
PE	Performed the experiments	Plos data
NC	Number of contributions	Plos data
PO	Author position in the paper	WoS data
AU	Total number of authors in the paper	WoS data
DT	Document type. Only articles and reviews are included	WoS data
YE	Number of years active as an academic based on year of first publication	WoS data
CO	Number of countries to which all authors are affiliated in the paper	WoS data
IN	Number of institutions to which all authors are affiliated in the paper	WoS data
PU	Average number of publications per year of the author at the time of publication of the paper	WoS data

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Modelling

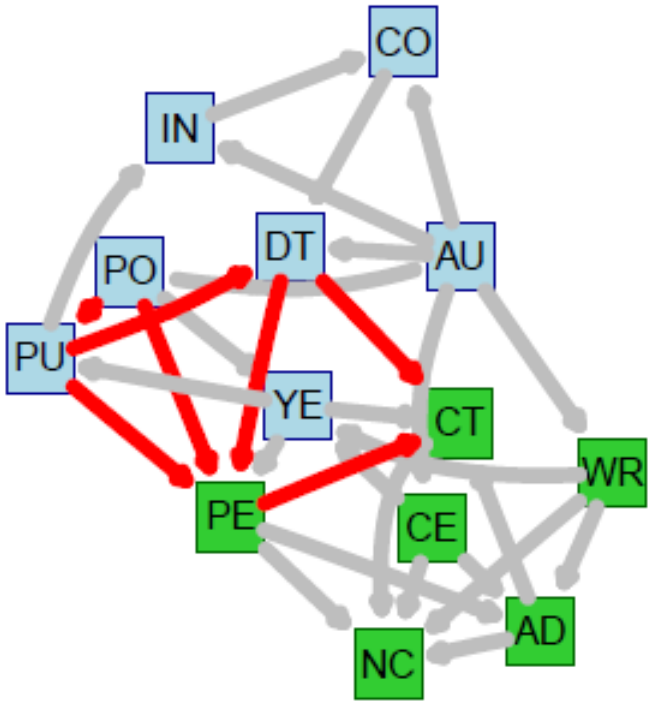


Table 2: Classification error ratios from cross-validation.

	Min.	Median	Mean	Max.
Wrote the manuscript	0.067	0.069	0.069	0.070
Analyzed the data	0.065	0.068	0.067	0.069
Performed the experiments	0.073	0.074	0.074	0.076
Conceived experiments	0.064	0.065	0.066	0.068
Contributed with tools	0.076	0.078	0.078	0.079
Number of contributions	0.073	0.073	0.073	0.073

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Data and Methods. Predictions

Dataset:

- 222,925 unique authors and 6,236,239 publications
- Criteria:
  1. Gender is clearly determined (using Gender API, Genderize.io and Gender Guesser - 90% accuracy threshold)
  2. First paper from 1980 onwards
  3. Letters excluded
  4. ≥ 5 publications
- Distribution by career length:
  - Junior (< 5 years). 222,925 authors
  - Early-career (< 15 years). 205,309 authors
  - Mid-career(< 30 years). 99,972 authors
  - Full career (≥ 30 years). 27,923 authors

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Predictions

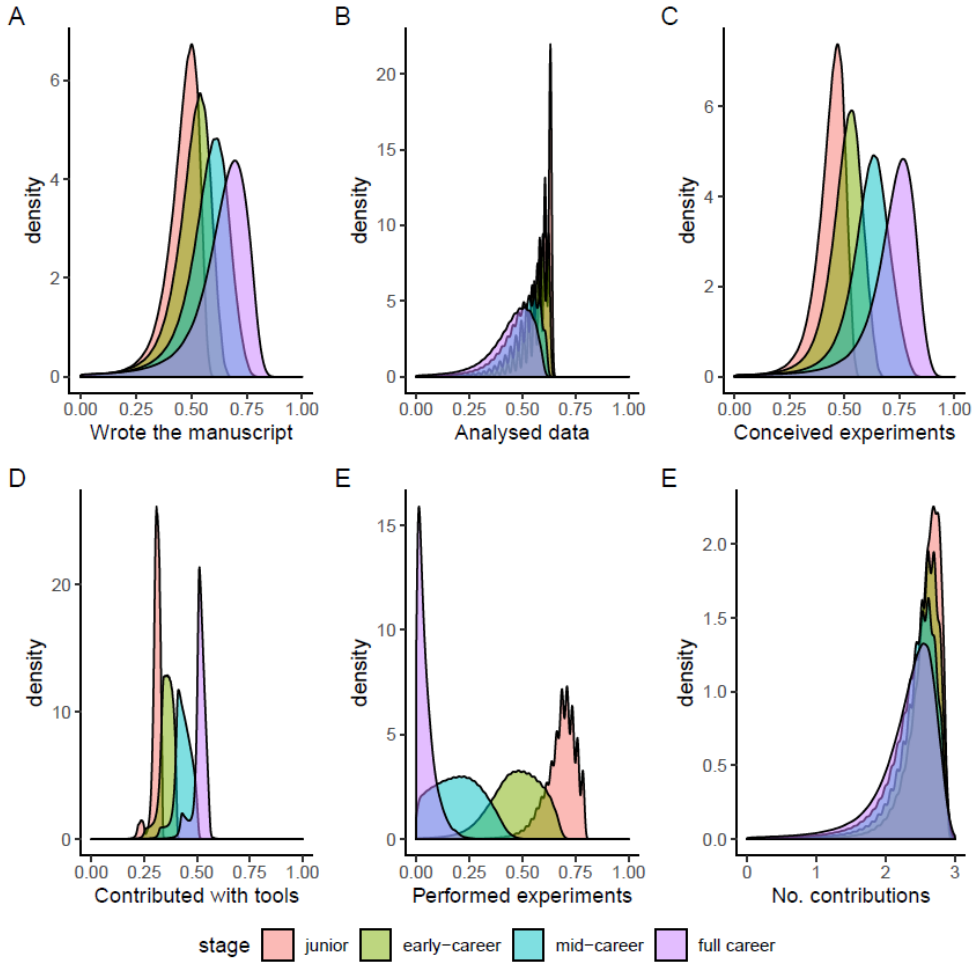
Distribution of probabilities contribution type and career stage

Junior ≥ 0 & < 5 years since 1st publication

Early-career ≥ 5 & < 15 years since 1st pub.

Mid-career ≥ 15 & < 30 years since 1st pub.

Full career ≥ 30 years since 1st pub.



[Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists]

Data and Methods. Archetypal analysis

ARCHETYPAL ANALYSIS

- Statistical data representation technique to characterize multivariate data sets (Cutler & Braiman, 1994)
- First used in scientometrics in 2013 (Seiler & Wohlrabe, 2013)
- It defines ***archetypes*** of individuals based on **extreme values of one or more variables**
- Individuals are then assigned to each archetype

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Data and Methods. Archetypal analysis

Method I

Aggregation method

Maximum value by contribution type for each stage

Archetypal assignment

Assignment to maximum α value

Method II

Aggregation method

Median value by contribution type for each stage

Archetypal assignment

Assignment to maximum α value

Method III

Aggregation method

Maximum value by contribution type for each stage

Archetypal assignment

Assignment to maximum ranked α value

Method IV

Aggregation method

Median value by contribution type for each stage

Archetypal assignment

Assignment to maximum ranked α value

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Data and Methods. Archetypal analysis

Method I

Aggregation method

Maximum value by contribution type for each stage

Archetypal assignment

Assignment to maximum α value

Method II

Aggregation method

Median value by contribution type for each stage

Archetypal assignment

Assignment to maximum α value

Method III

Aggregation method

Maximum value by contribution type for each stage

Archetypal assignment

Assignment to maximum ranked α value

Method IV

Aggregation method

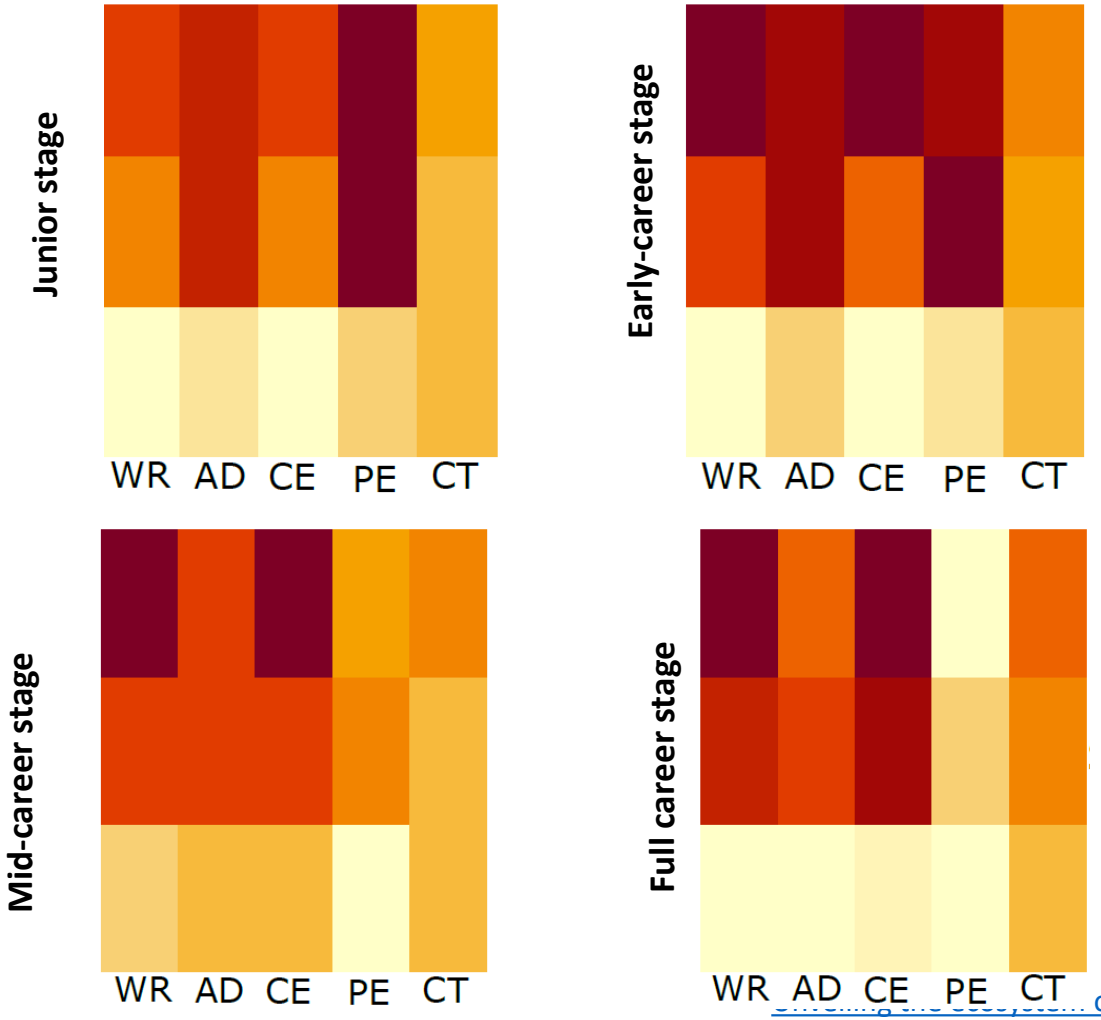
Median value by contribution type for each stage

Archetypal assignment

Assignment to maximum ranked α value

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Archetypal

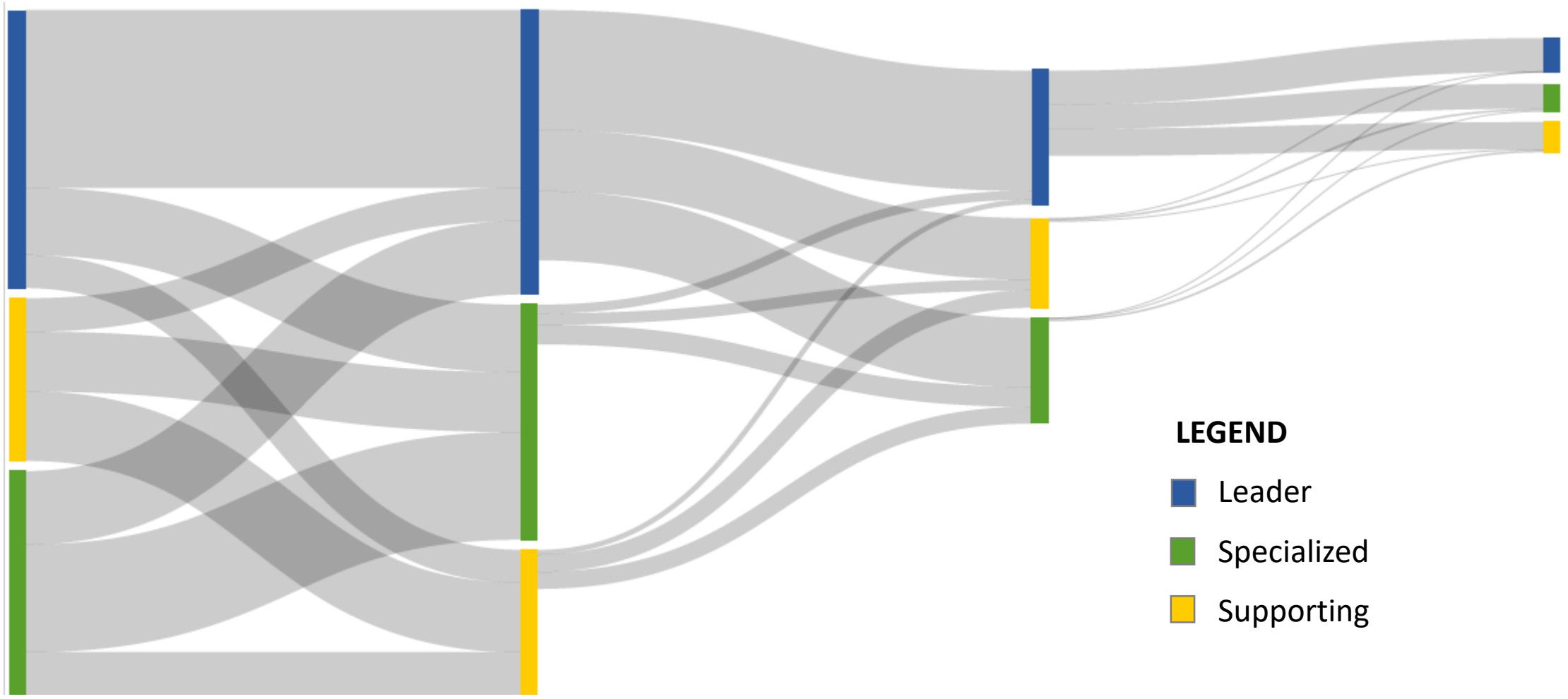


Three profiles relatively similar in different stages:

- **Profile 1** related with leadership and multitasking – **Leader**
- **Profile 2** related with specialized knowledge – **Specialized**
- **Profile 3** related with specific contributions related with tools – **Supporting**

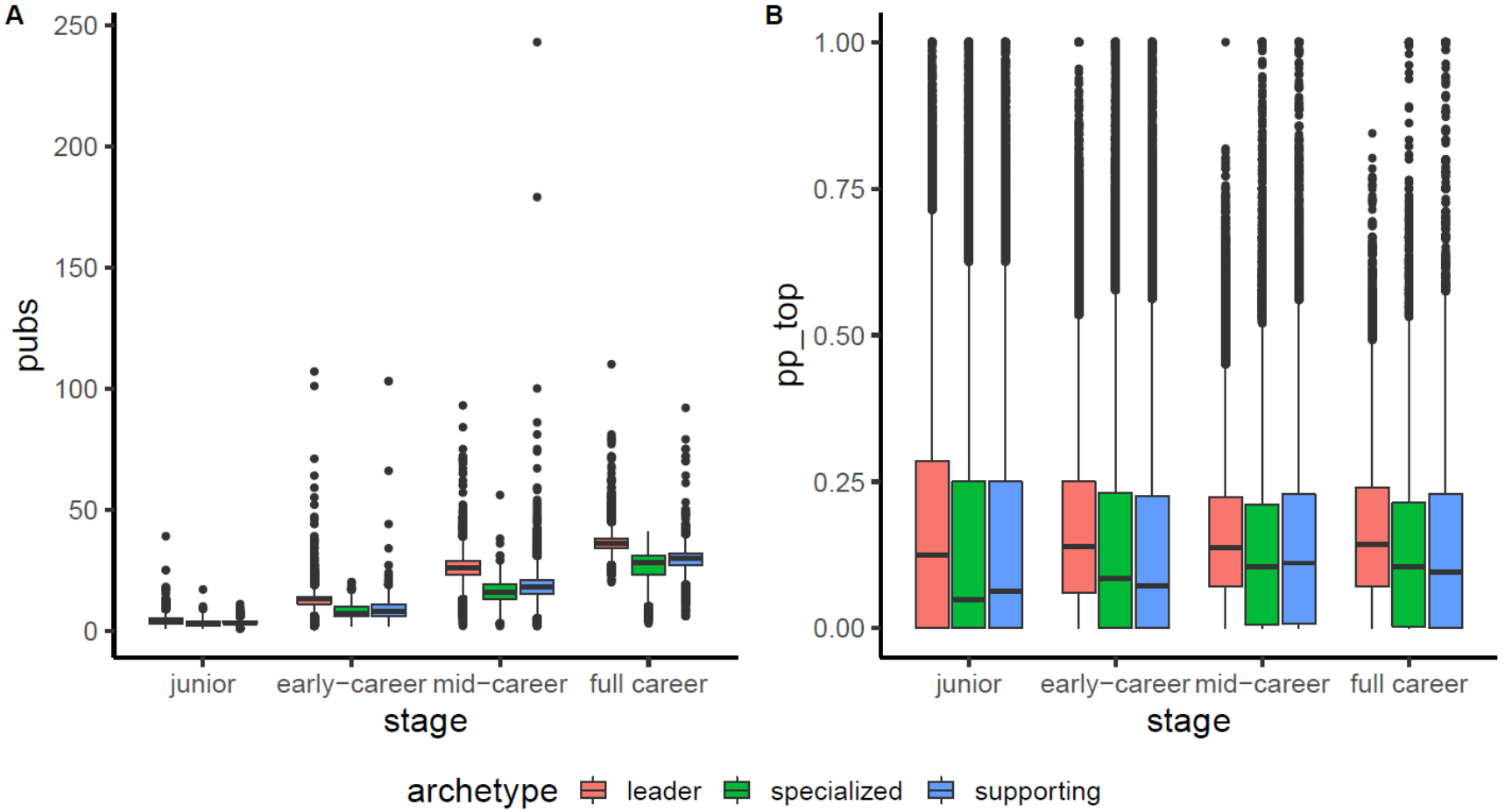
[...] of science: A contextual perspective of the many roles of scientists

Results. Career length



Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Productivity and impact



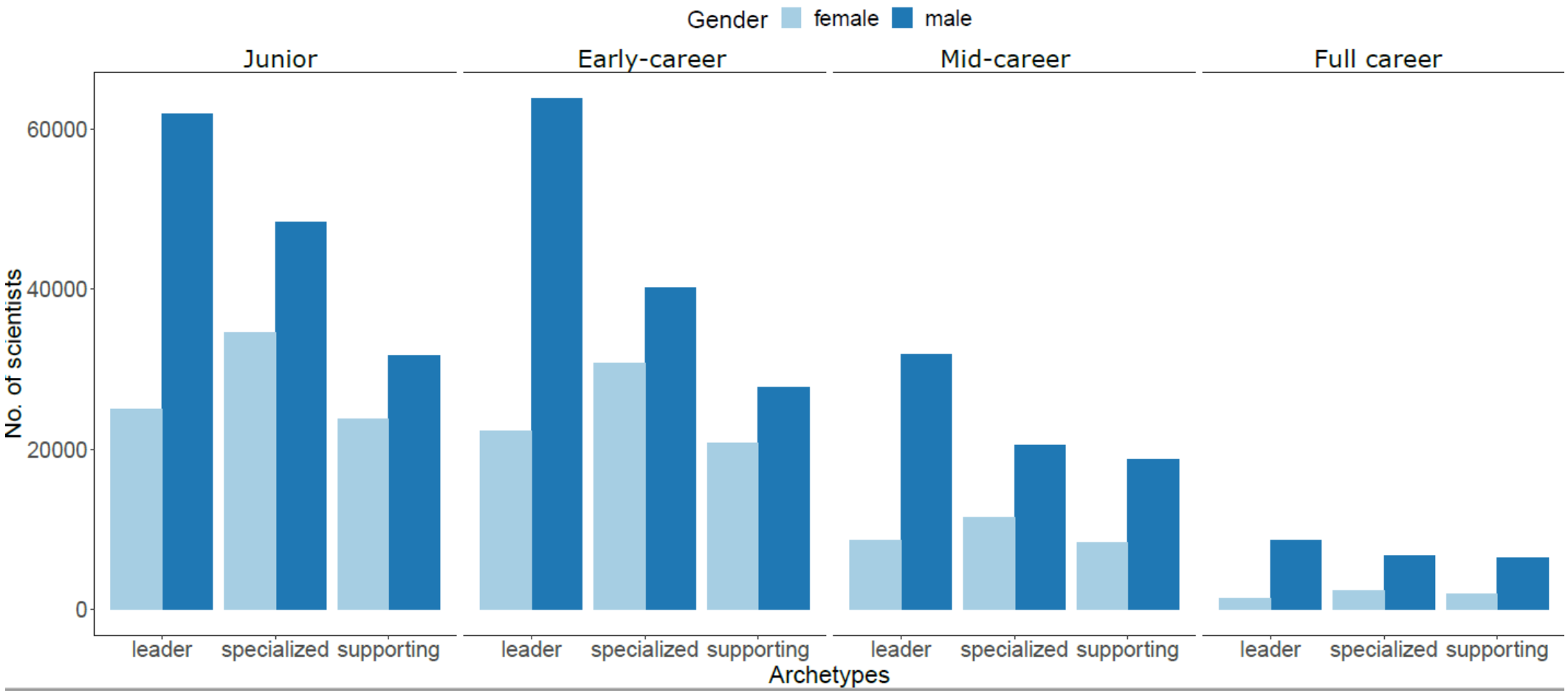
A Total number of publications

B Share of top 10% highly cited publications

- Differences on productivity b2n the leader profile and the other two
- No significant differences on citation impact, only for lower part of the leader distro

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Results. Gender distro by career stage



Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Limitations of the study

- Representativeness of the sample of scientists
- Identification of scientists
- Taxonomy of contributorships

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists

Discussion

1. We do observe an heterogeneity of scientific profiles when looking at contribution statements
2. Stability of profiles by career stage, but many career paths
3. Leader profiles seem to have longer career trajectories
4. It seems easier to shift profiles from Leader to others than viceversa
5. No significant differences in terms of productivity and impact between profiles
6. There is a consistent unbalance on the distribution by gender and career stage of profiles

Many thanks

TU Delft

Questions?

Unveiling the ecosystem of science: A contextual perspective of the many roles of scientists