

# Investigating scientific collaboration through the sequence of authors in the publication bylines and the diversity of collaborators

**Authors:** Yi Bu, Yong Huang, Cassidy R. Sugimoto, and Zaida Chinchilla-Rodríguez

**Presenter:** Yi Bu, School of Informatics, Computing, and Engineering, Indiana University

# Investigating **scientific collaboration** through the **sequence of authors** in the publication bylines and the **diversity of** **collaborators**

**Authors:** Yi Bu, Yong Huang, Cassidy R. Sugimoto, and Zaida Chinchilla-Rodríguez

**Presenter:** Yi Bu, School of Informatics, Computing, and Engineering, Indiana University

# Scientific collaboration

- Scientometrics, science of science, quantitative science studies...
- Scientometricians have made great efforts to understanding scientific collaborations from different perspectives (e.g., Katz, 1994; Lu et al., 2018; Newman, 2001; Wu et al., 2019; Zhang et al., 2018)
  - Co-authorship as a proxy of measuring scientific collaborations (Milojević, 2010)

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.

Lu, C., Ding, Y., Zhang, Y., Bu, Y., & Zhang, C. (2018). Types of scientific collaborators: A perspective of author contribution network. In *Proceedings of iConference 2018*, March 25-28, 2018, Sheffield, U.K.

Milojević, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. *Journal of the American Society for Information Science and Technology*, 61(7), 1410-1423.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 404-409.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams have developed science and technology; small teams have disrupted it. *Nature*.

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology*, 69(1), 72-86.

# Diversity of collaborators

- Among these, one branch of study in scientific collaboration focuses on the diversity of collaborators, a.k.a., members in a research team.
  - Bu et al. (2018): Relationships between an author's impact and his/her collaborators' diversity (**impact** diversity and **research topic** diversity)
  - Zhang et al. (2019): Diversity on team members' **productivity** and **scientific ages** will increase the team performance

# Why care about diversity of collaborators?

- Diversity and collaboration go hand in hand
- Diversity can be wielded as a powerful tool
- More diverse is better? (take research topic diversity as an example)
  - ✓ Bright side: “handle high levels of complexity, tap otherwise isolated sources of local knowledge, foster transformative thinking, and enhance legitimacy” (Xu et al., 2015, p. 2)...
  - ✓ Dark side: cause inter-personal friction and require extra resource and time investment (Schaltegger et al., 2013), and is often confronted with tremendous practical barriers such as communication among members due to different jargons (Institute of Medicine, 2000)...

# Authors' sequences in publications' bylines

- The sequence of co-authors identifies details on “who is accountable for the integrity of the reported study and who deserves what amount of credit for the work” (p. 359), as well as their contributions (He et al., 2012)
- Previous studies on diversity of collaborators ignore authors' sequences in publications' bylines when research performance is analyzed (Chinchilla-Rodríguez et al., 2019)

# Objectives

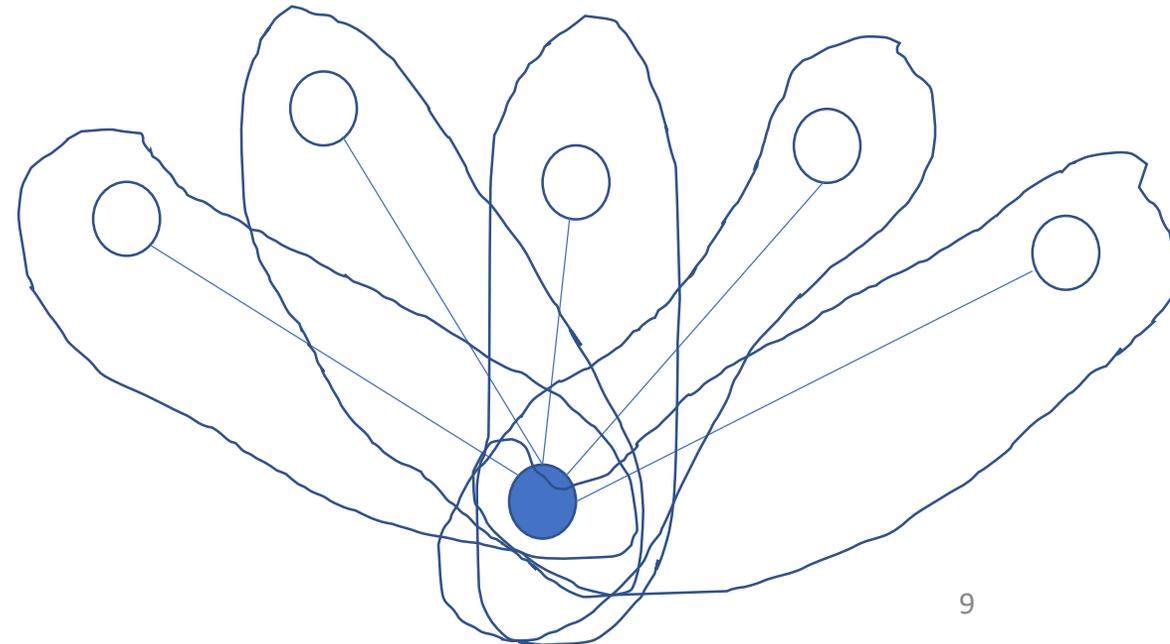
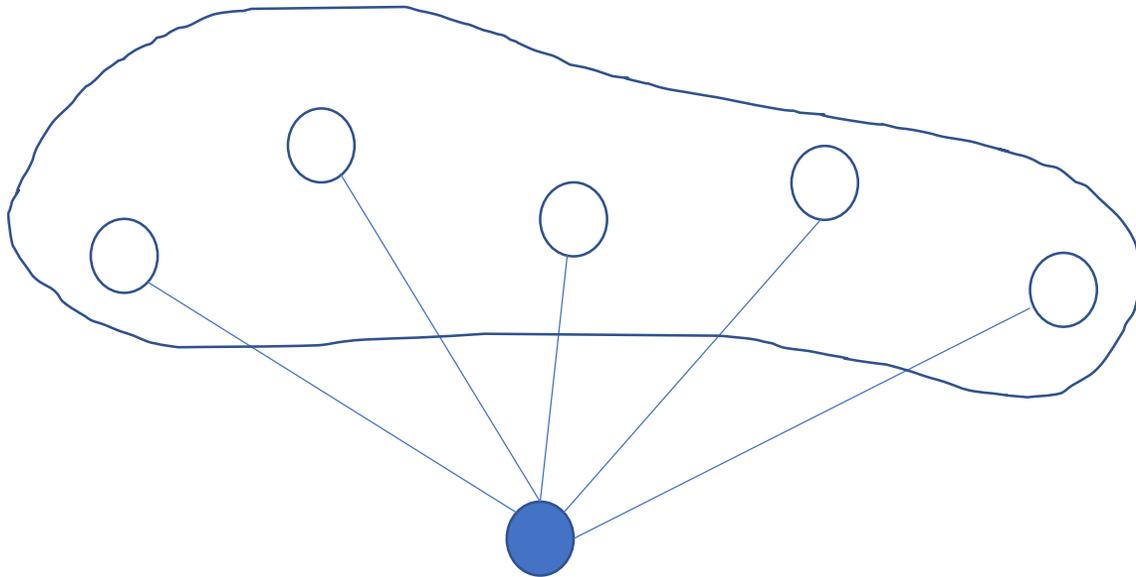
- We investigate the relationship between **the sequence of authors in publication bylines** and **the diversity of their collaborators**.
  - How do the sequence of authors and the diversity of collaborators relate to the quality of their co-authored publications? (3-dimensional analysis)
    - ❖ What patterns (e.g., more tendencies to work as first authors + more topic diversity) relate to team outputs that receive more citations?”
  - Diversities: research topic, citation impact, gender

# Data

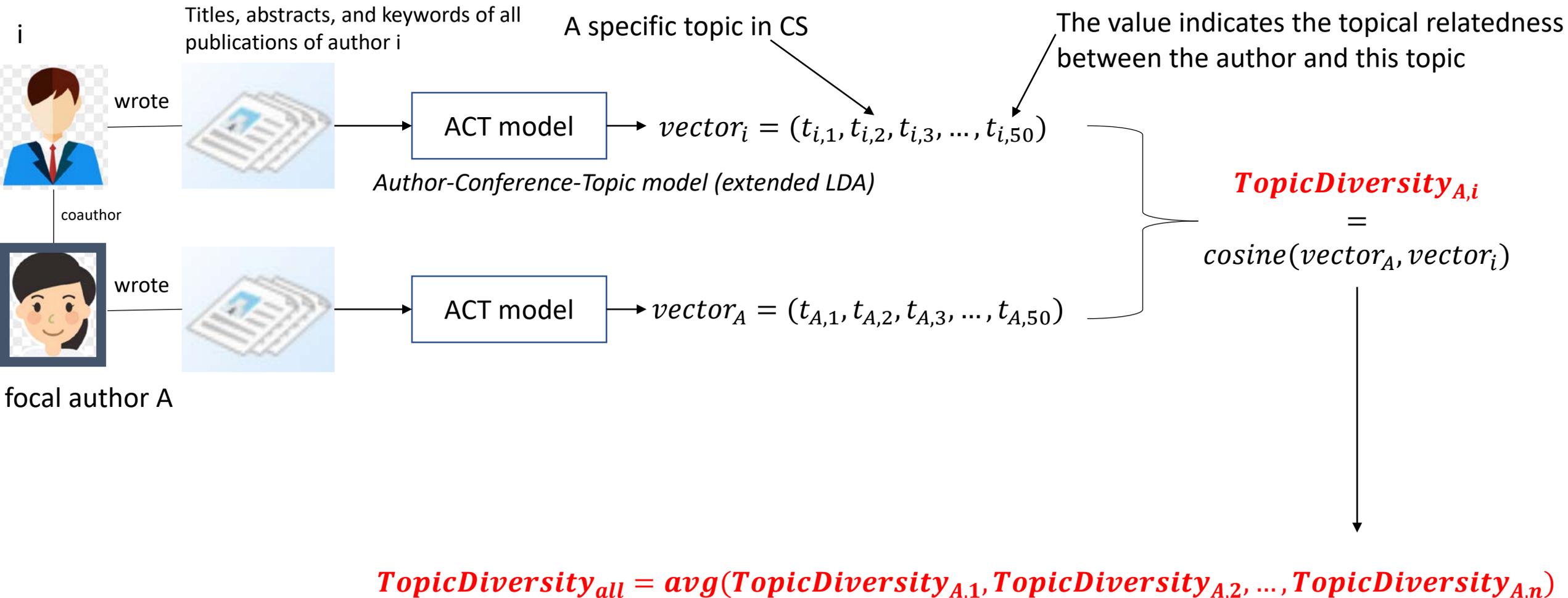
- ArnetMiner: 2M publications, 1.2M authors, and ~8M citing relations between these publications (1936-2014). (Tang et al., 2008)
- 2001-2010 publications (450K papers, 885K authors, 4M collaboration pairs, and 606K local citation relations)
- Author name disambiguation (Tang et al., 2012).
- Issues of local vs. global citations

# “Diversity between collaborators” is confusing?

- Diversity among collaborators of an authors (the left figure)
- **Diversity between an author and his/her collaborator (the right figure)**



# Topic diversity of collaborators of an author



# (Citation) Impact diversity of collaborators of an author

n: total number of collaborators of A

Citation count of the i-th collaborator

number of collaborations between A the his/her i-th collaborator

the impact diversity of collaborators of the author A

$$ID(A) = \frac{\sum_{i=1}^n (h_i x_i - avg(A))^2}{\sum_{i=1}^n x_i}$$
$$avg(A) = \frac{\sum_{i=1}^n h_i x_i}{\sum_{i=1}^n x_i}$$

Essentially: Normalized Standard Deviation

# Gender diversity of collaborators of an author

- Gender = {Male, Female}
- Gender identifier: <https://gender-api.com/>
- Manually test the accuracy with a small sample (1000 authors): ~91%
- Chowdury (2005) measured gender diversity in a straightforward measure as the percentage of the smaller gender representation

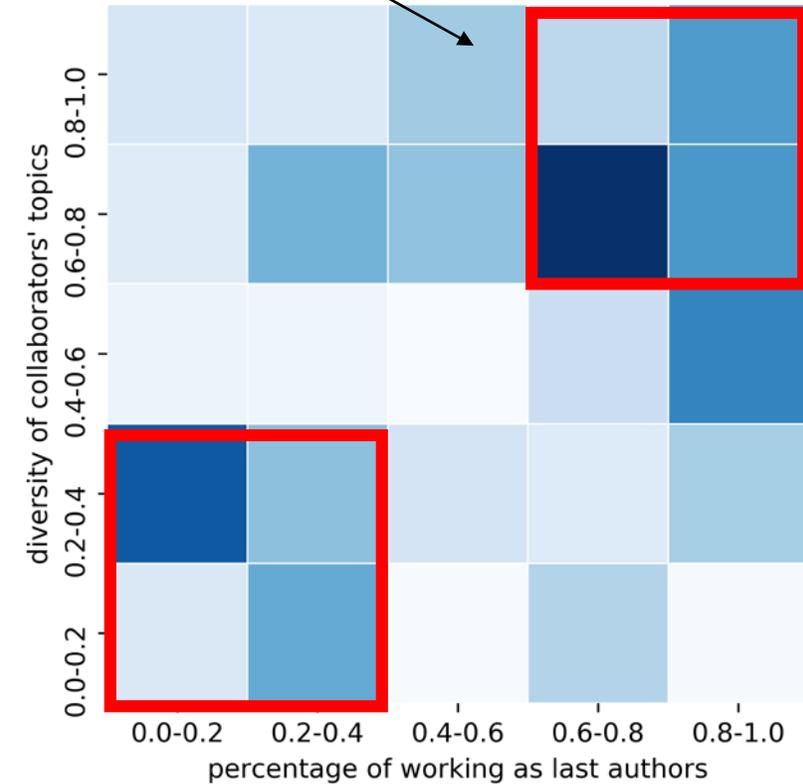
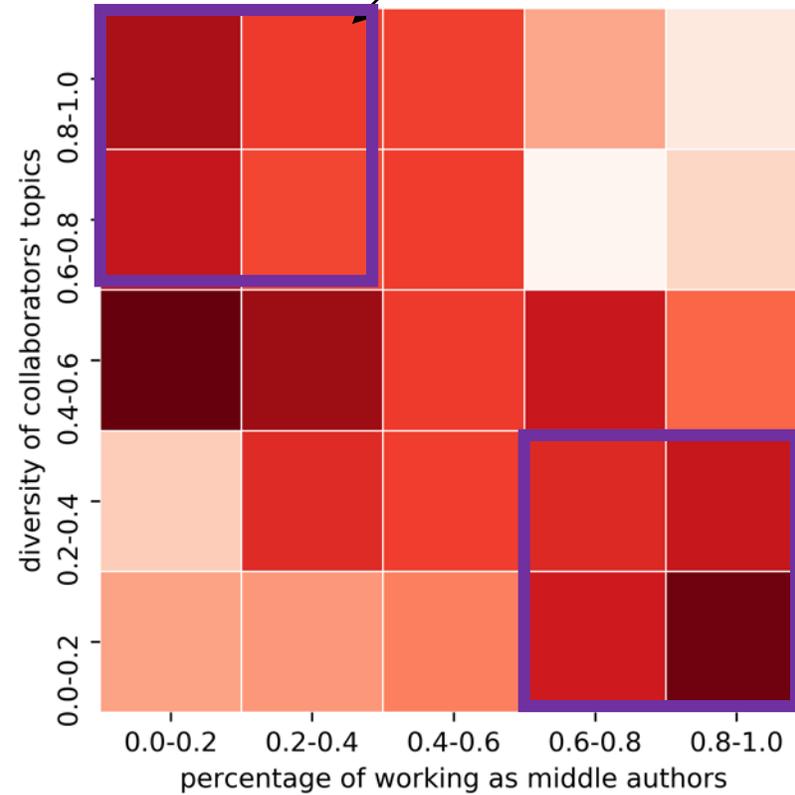
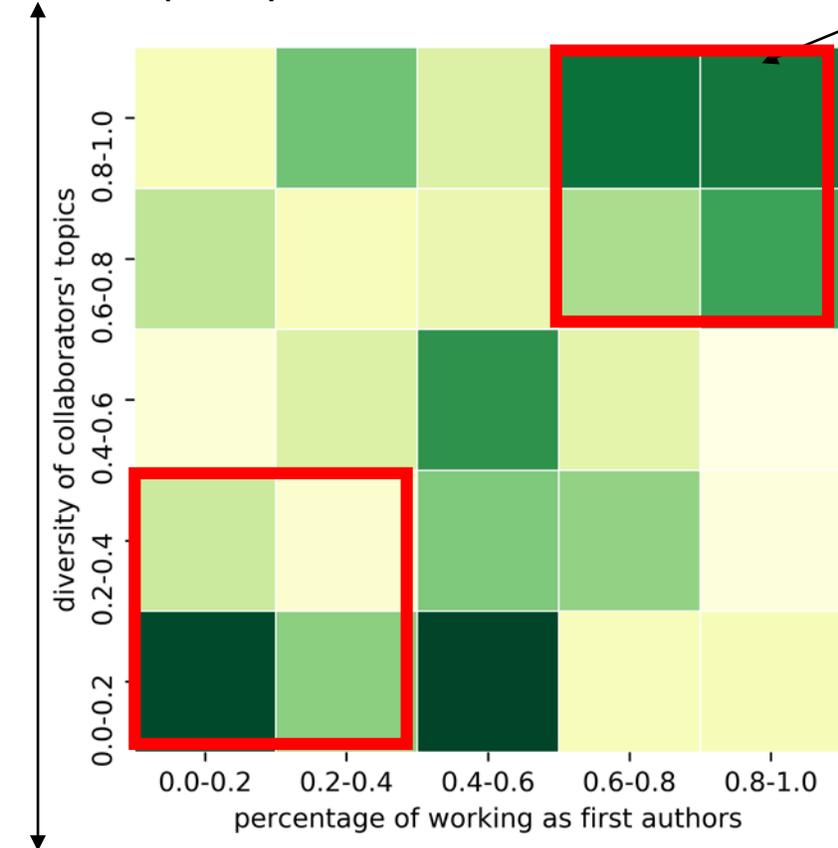
# Measurement on “quality of team outputs”

- Publications as a proxy of team outputs
- Quality of their outputs: Scientific, economic, societal, ...
- Citations as a measurement on their scientific impact (visibility, influence)
- Relative number of citations of the co-authored publication compared with the number of citations of articles published in the same year under the same topic.
- Example:
  - ✓ A certain publication in artificial intelligence published in 2010
  - ✓ Relative number of citations = its raw citation count / average number of citations of all publications in 2010 under the topic of artificial intelligence

# Author sequence & topic diversity of collaborators

Darkness of cells is proportional to the average number of citations of publications authored by a scholar with the corresponding x and y axis values

more topically diverse collaborators

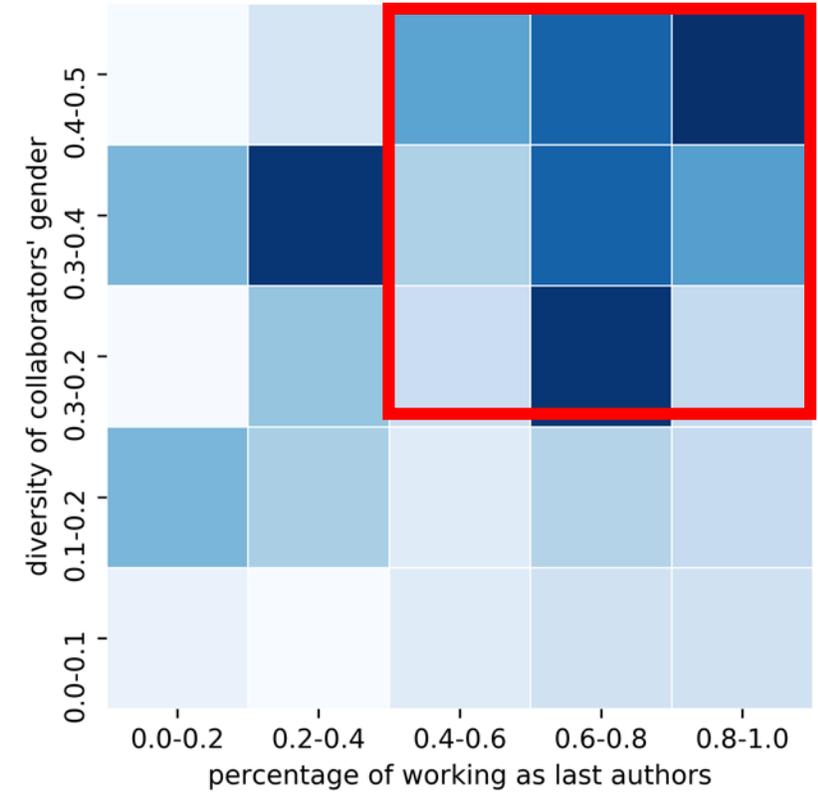
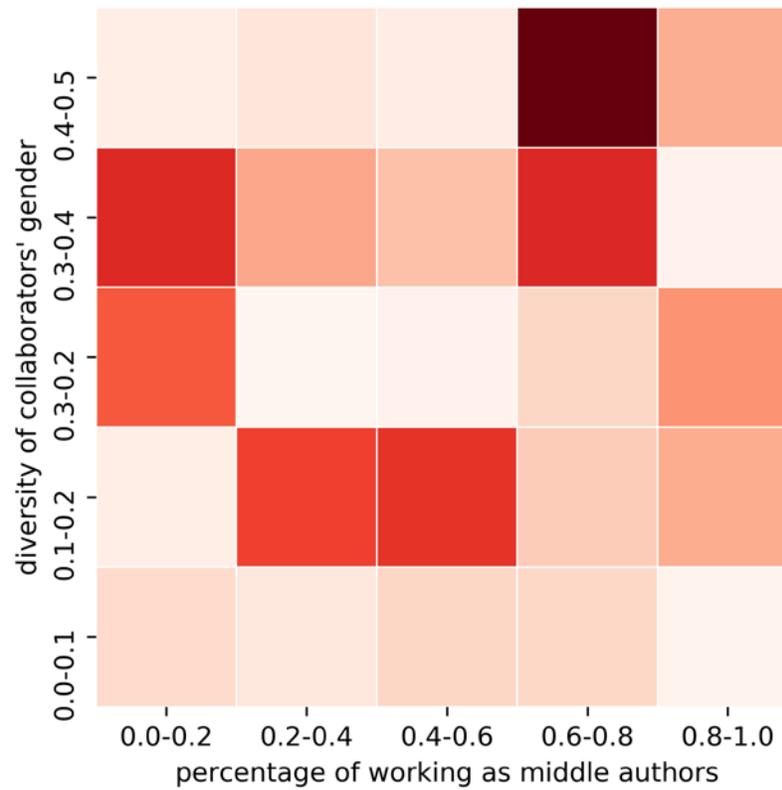
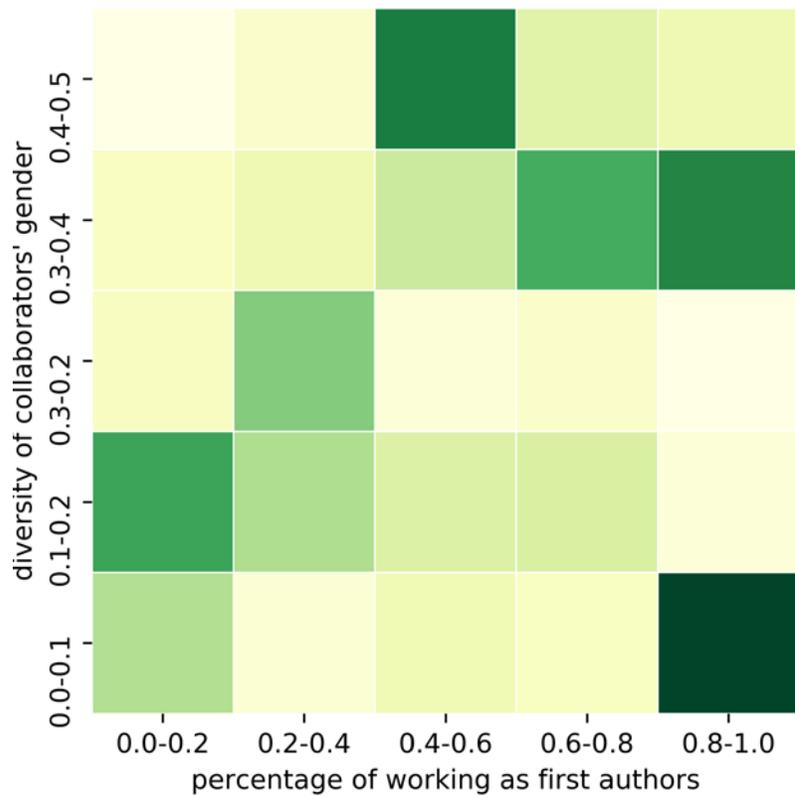


less topically diverse collaborators



# Author sequence & gender diversity of collaborators

more gender-diverse collaborators



less gender-diverse collaborators

# Objectives

- We investigate the relationship between **the sequence of authors in the publication bylines** and **the diversity of their collaborators**.
  - How do the sequence of authors and the diversity of collaborators relate to their co-authored publications? (3-dimensional analysis)
  - How do the patterns differ for authors before and after their Ph.D. graduation?
  - Diversities: research topic, citation impact, gender

# In different career stages, scholars may have different collaboration patterns

- Costas & Bordons (2011): Young scholars during their doctoral studies tend to have more first-authored publications.
- Raelin (1985): Between Ph.D. graduation and getting a tenure position, scholars may also begin to experience stress related to productivity, promotion, and tenure—being pushed towards activities that more directly benefit their institutions.
- Gingras et al. (2008): Senior researchers have a smaller ratio of first-authored articles, indicating that they might be managing research teams rather than executing the work.
- Bu et al. (2018): High-impact CS authors tend to work with “giants” early in their career, but not in the later stages; collaboration is important but the size of collaborations also matters.

# Before and after PhD graduation?

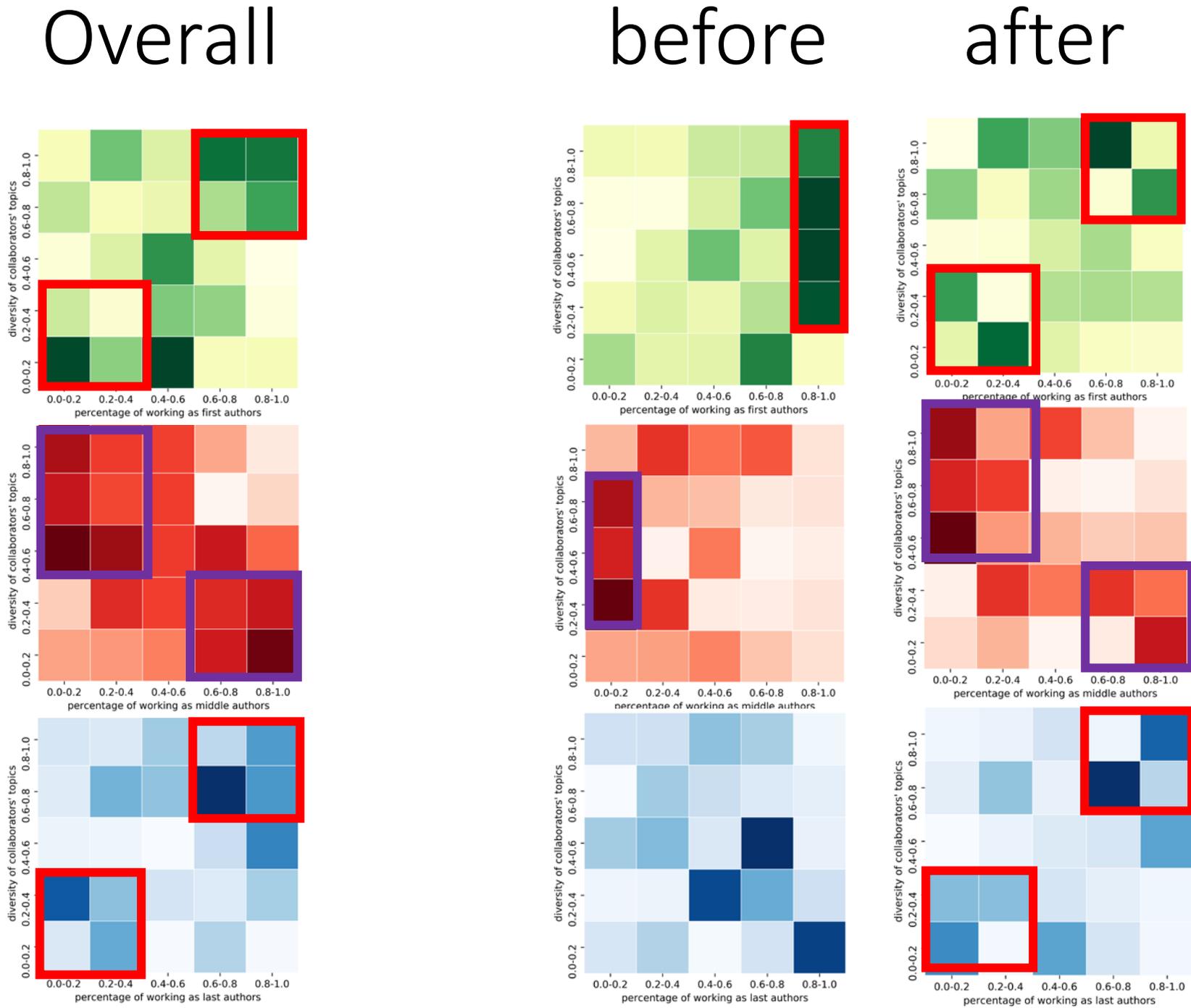
	First-authored publication	Last-authored publications	Middle-authored publications	Research topic diversity	Impact diversity	Gender diversity
Before Ph.D. graduation	73.8%	8.2%	18.0%	0.48	2.45	0.10
After Ph.D. graduation	21.4%	39.5%	39.1%	0.60	2.86	0.17

# Topic diversity

more topic-diverse collaborators



less topic-diverse collaborators



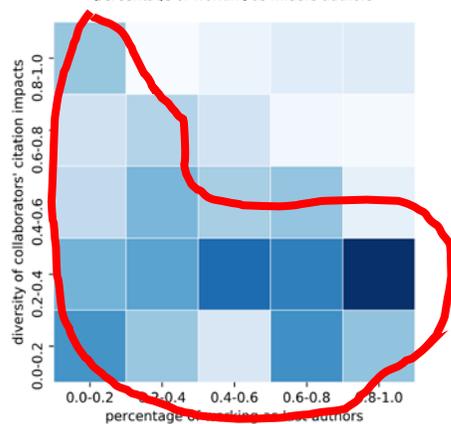
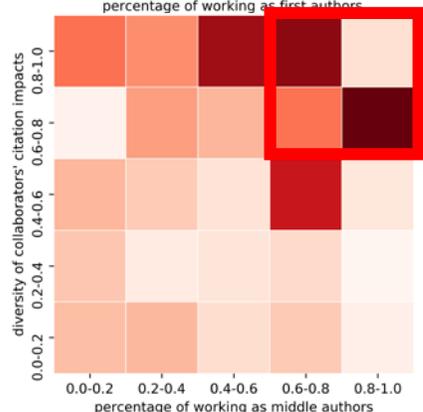
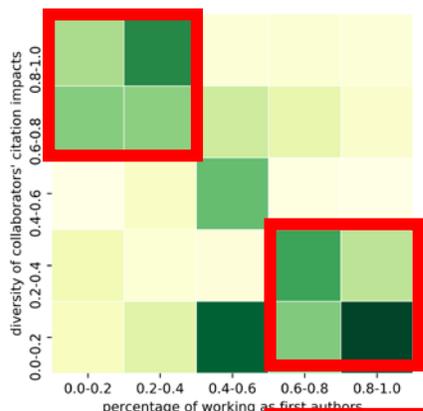
# Impact diversity

more impact-diverse collaborators

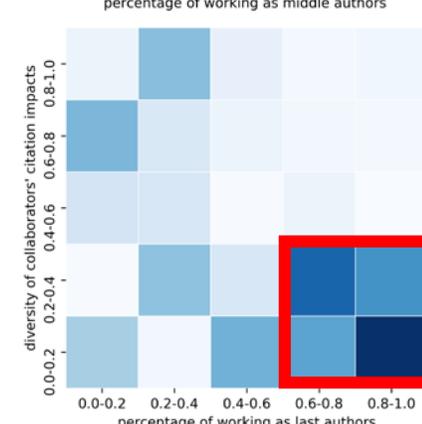
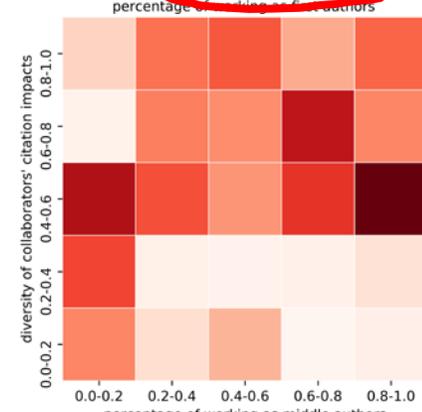
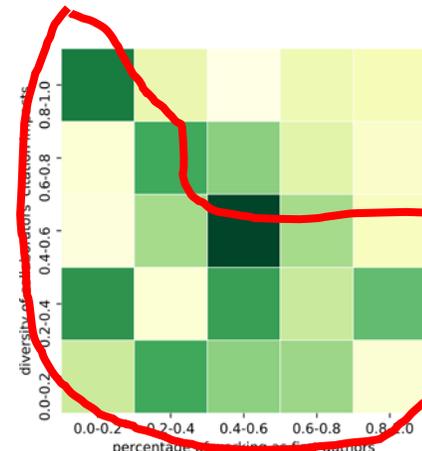


less impact-diverse collaborators

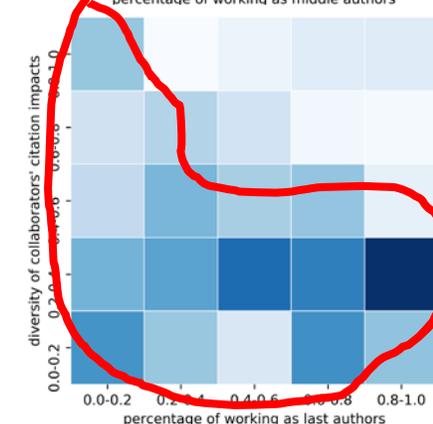
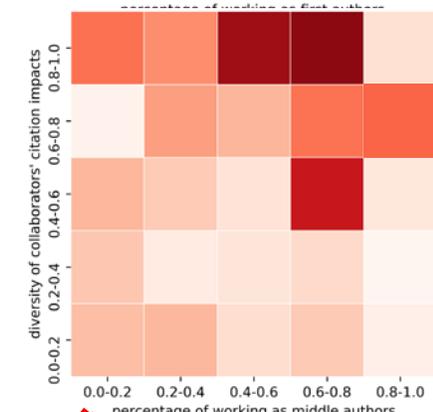
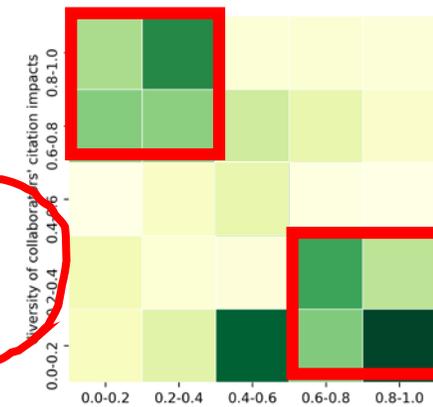
## Overall



## before



## after



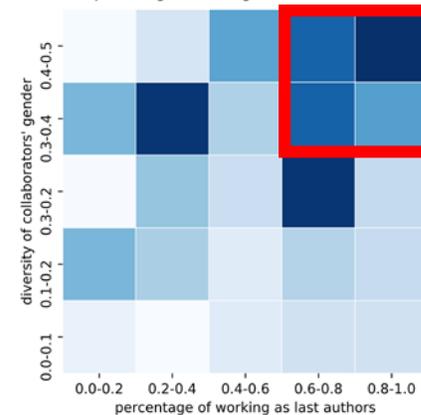
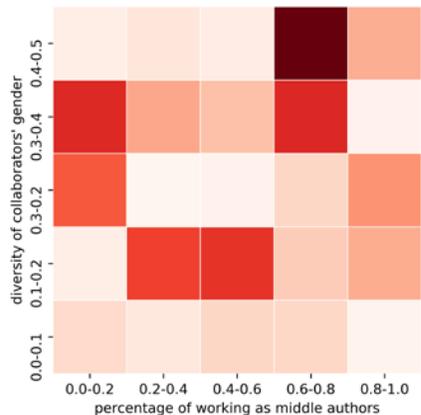
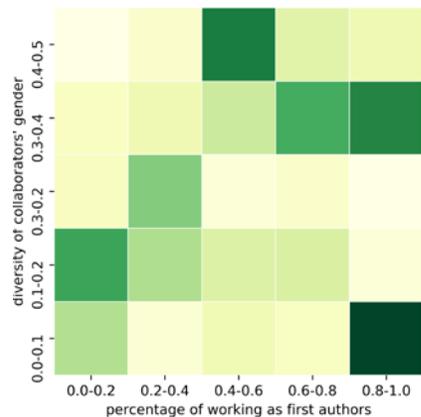
# Gender diversity

more gender-diverse collaborators

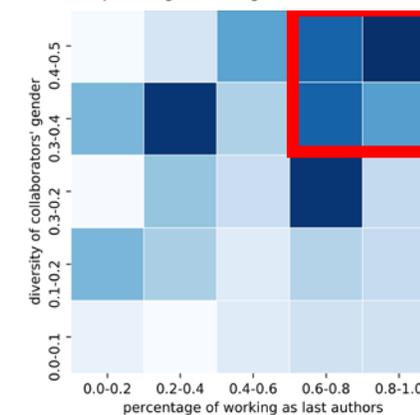
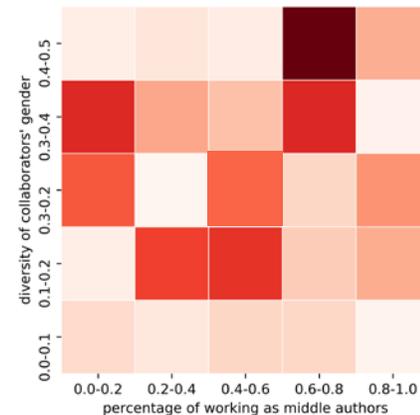
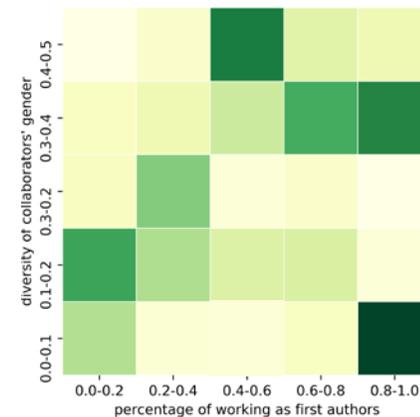


less gender-diverse collaborators

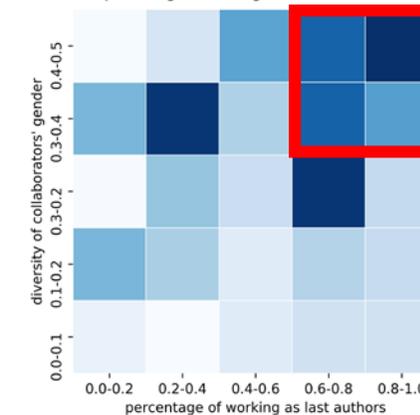
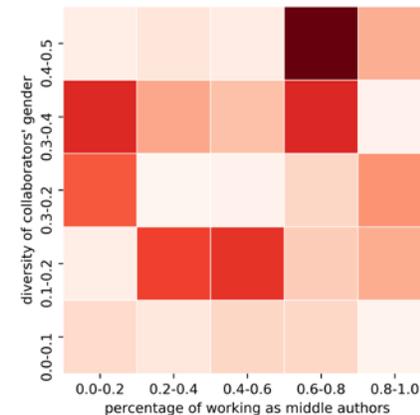
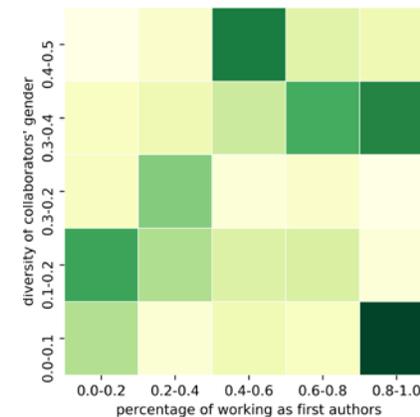
## Overall



## before



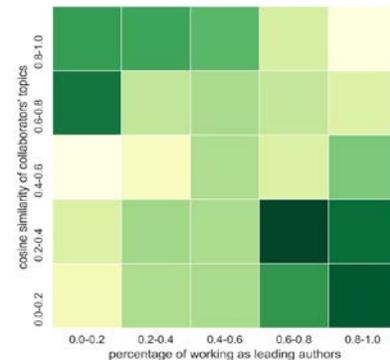
## after



# Summary of preliminary results: These combinations relate to high-quality outputs of teams

- Overall
  - ✓ Topic diversity: Greater (less) diversity of collaborators + more (less) % of first/last authors;
  - ✓ Impact diversity: Any combination except greater diversity of collaborators + greater % last authors; and
  - ✓ Gender diversity: Greater diversity of collaborators + more % of last authors.
- Before Ph.D. graduation
  - ✓ Topic diversity: Greater diversity of collaborators, more % of first authors;
  - ✓ Impact diversity: Any combination except greater diversity of collaborators + greater % first authors; and
  - ✓ Gender diversity: Greater diversity of collaborators + more % of last authors.
- After Ph.D. graduation: Quite similar to “Overall” because of dominant productivity of researchers

# Implications



- Interdisciplinarity doesn't always lead to a great number of citations.
- While research evaluation criteria are institutionalized based (mostly) on scientific impact (number of citations received and the impact factors of journals), interdisciplinarity requires some conditions and time which are not reflected just in number of citations but in teams' skills, capacities, and innovative and social impact.
- High topic diversity of collaborators is not always accompanied by high impact diversity of collaborators, either in first and last positions. However, middle positions play an important role in both dimensions.
- This approach and findings, even when they are very preliminary, could provide a tentative framework for analyzing patterns of interdisciplinarity and collaboration work in order to inform policy recommendations.

# Limitations

- Ph.D. graduation (now only 1K authors)
- Discipline (now only CS)
- Leading author  $\neq$  corresponding author  $\neq$  first author + last author, even in CS
- Local citations
- Temporal analysis following my previous study (Bu et al., 2018)?
  - ❖ -4-0 years, 1-5 years, 6-10 years, 11-15 years, ... (until his/her career ends)
  - ❖ How do the diversities (in various dimensions) change over time?
  - ❖ How do such patterns relate to his own productivity, scientific impact, and collaborators'?
  - ❖ ...

# THANKS

**Authors:** Yi Bu, Yong Huang, and Cassidy R. Sugimoto, Zaida Chinchilla-Rodríguez

**Presenter:** Yi Bu, School of Informatics, Computing, and Engineering, Indiana University

05/22/2019