The association between early career informal mentorship in academic collaborations and junior author performance

Bedoor Al Shebli, Kinga Makovi, Talal Rahwan

NYU Abu Dhabi

WiSE Journal club

Dec 16th, 2020

Goal of paper

- Premise: Mentorship is important to academia, but difficult to quantify
  - Advances individual careers
  - Perpetuates culture, best practices and inner workings of professions
  - Alleviates barriers of entry for under-represented minorities
- Goal: Study the association between mentorship and scientific impact
  - Hypothesis: Quality mentorship leads to higher protégé scientific impact

- Academic age: # of years since first publication
- Junior scientist: academic age <= 7
- Senior scientist: academic age > 7
- Protégé – mentor pairs: “Whenever a junior scientist publishes a paper with a senior scientist, we consider the former to be a protégé, and the latter to be a mentor, as long as they coauthored at least one paper with 20 or less co-authors and share the same discipline and US-based affiliation”

Dataset

- Microsoft Academic Graph (MAG)
  - ”…containing scientific publication records, citation relationships between those publications, as well as authors, institutions, journals, conferences, and fields of study.”

Name disambiguation – correctly linking an author to a paper

MAG methods

- Author affiliation + publication venue + co-author network -> ML algorithms -> define authors
- Data is mined from authors’ websites and other publicly available data -> ML algorithms -> cross check MAG -> define authors

Authors’ extension on MAGs approach

- Iterative merging of authors by set criteria
- Ended up with 215 million putative unique authors

False positive rate: 0%
False negative rate: 2%

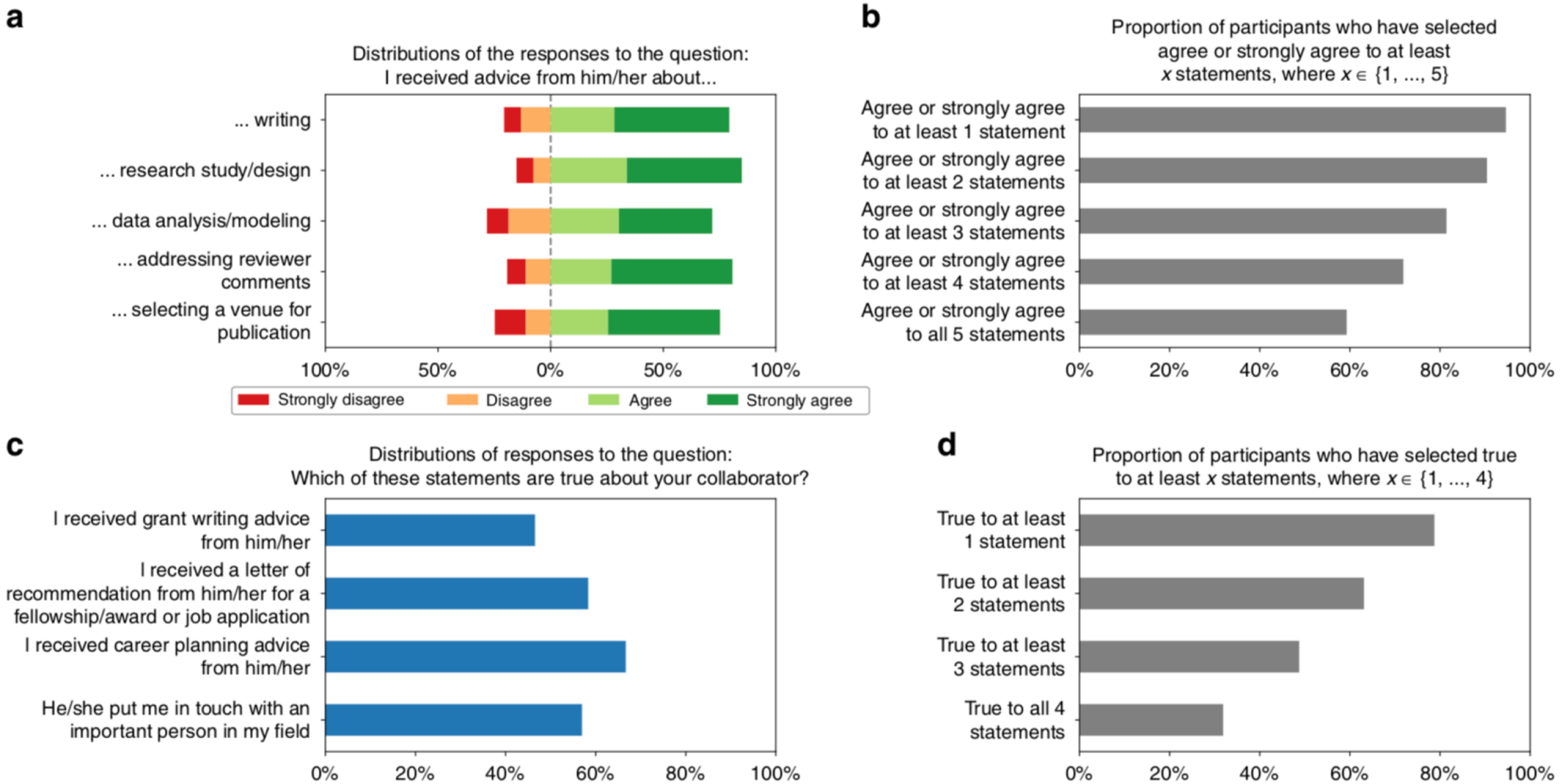
Determined from a sample of 400 pairs of papers.

Methods: Assigning Protégé-mentor pairs

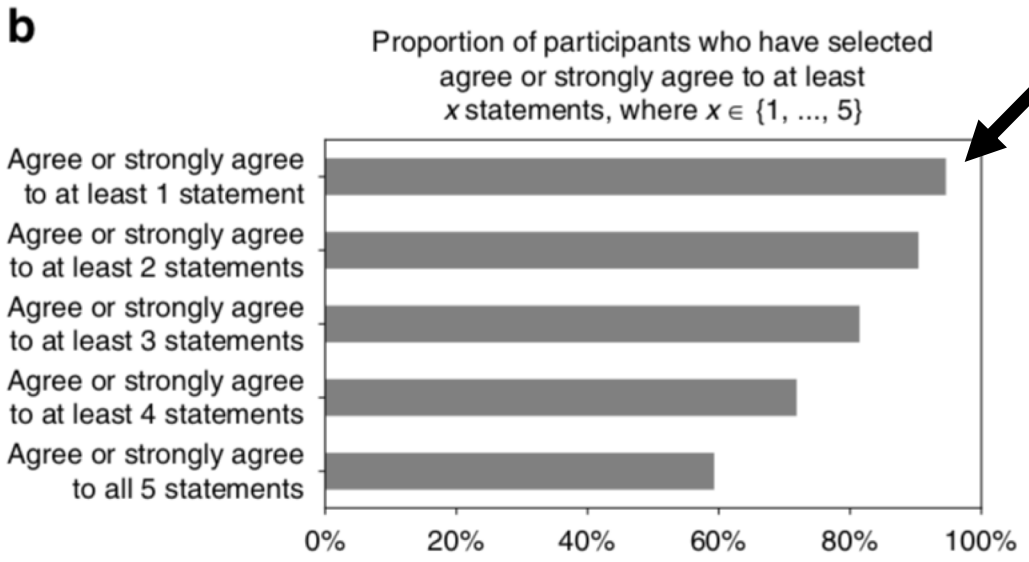
- Assigning gender
  - Used a tool called Genderize.io, classifies by first name
  - Only kept the ~90 million scientists that were classified with at least 90% confidence

- Protégé-mentor pairs
  - End up with 3 million unique pairs
  - “they authored at least one paper with 20 or less co-authors and share the same discipline and US-based affiliation “

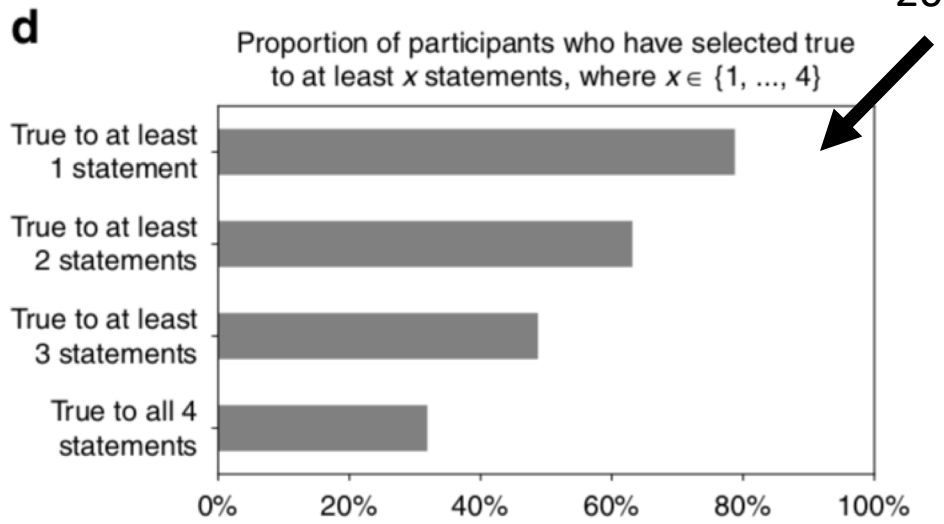
Figure 1: evidence for mentorship



5% reported no mentorship



20% reported no mentorship



- Survey was sent to 2000 protégés, was completed by 167
- Claiming a sample of 167 is representative of the 3 million protégé-mentor dataset
- If survey is actually representative, then between 5% to 20% of their protégé-mentor pairs had zero mentorship by their metrics

Methods: Measuring Mentorship quality

Mentors

- Bigshot effect
  - “average number of citations per annum up to the year of their first publication with the protégé “
  - Averaged across all mentors

- Hub experience
  - “the average degree of the mentors prior to mentorship, where the degree of each mentor is calculated in the network of scientific collaborations up to the year of their first publication with the protégé “
  - Averaged across all mentors

Methods: Measuring Protégé impact

Protégé

- “we conceptualize as the scientific impact of the protégé during their senior years without their mentors”
- Average impact of published papers
  - “they were published when the academic age of the protégé was greater than 7 years”
  - ”the authors include the protégé but none of the scientists who were identified as their mentors”

- C5 – Number of citations in 5 years post publication

Independent variables

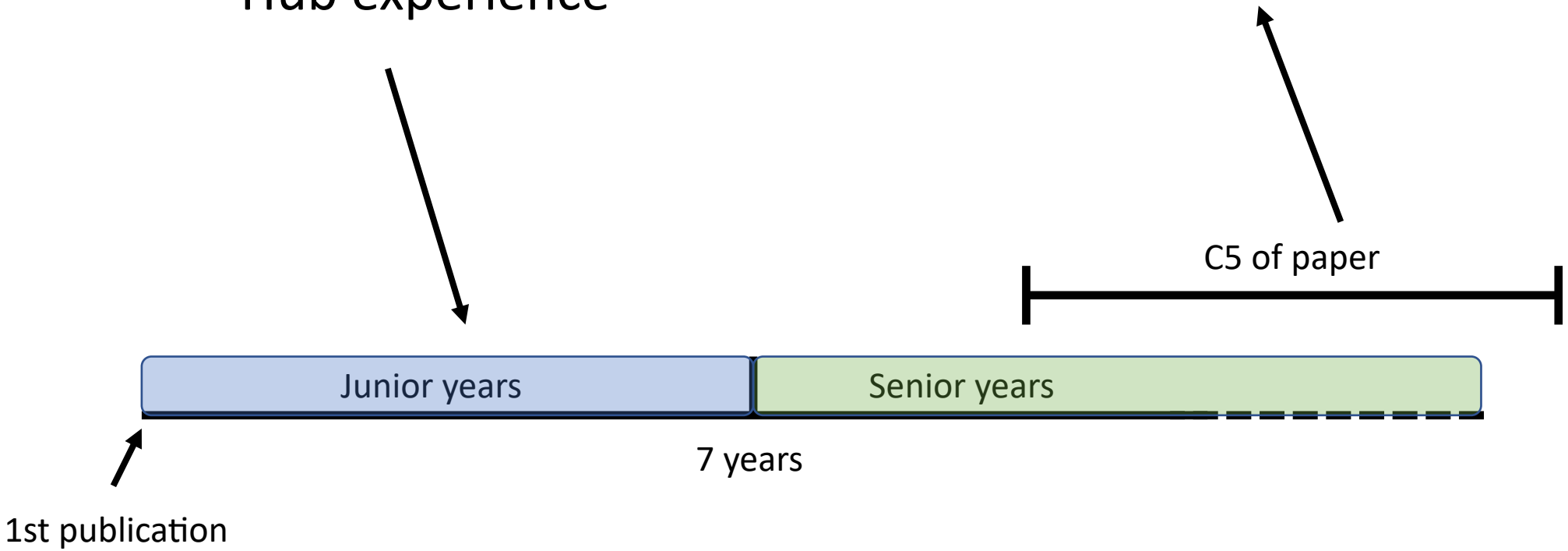
Mentorship quality

- Bigshot effect
- Hub experience

Dependent variable

Protégé Impact

- C5



Methods: Choosing protégés for comparisons

Coarsened Exact Matching

- Matches protégés by metadata

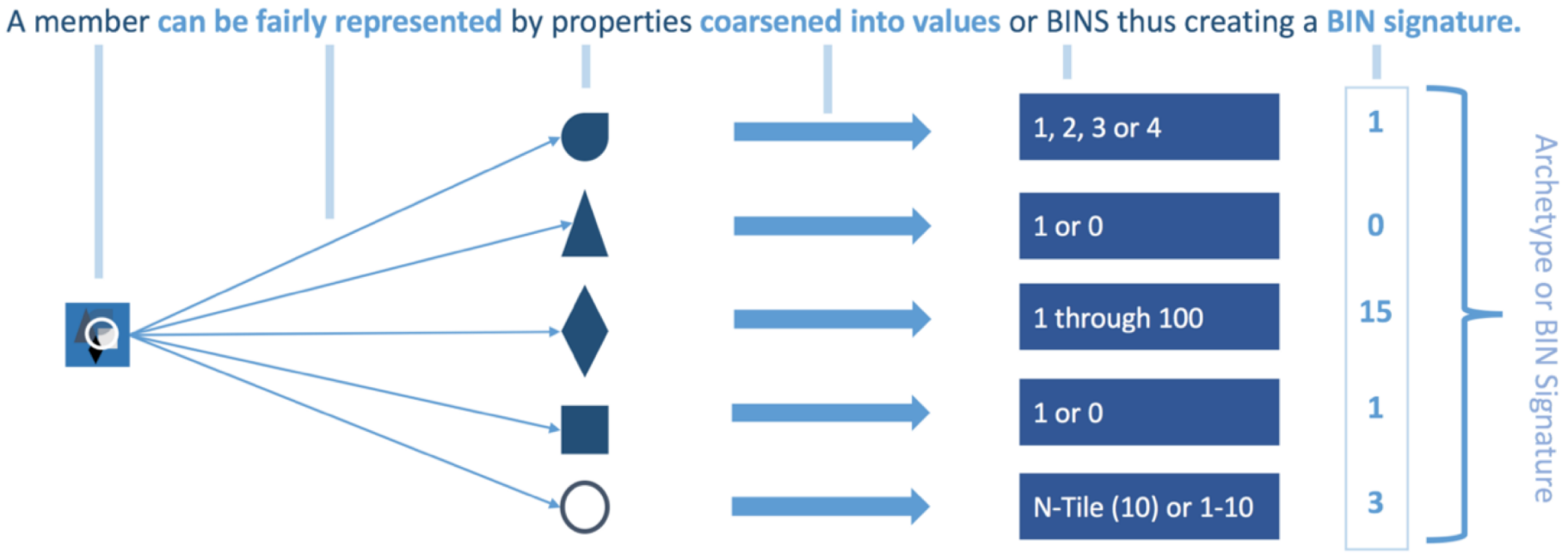
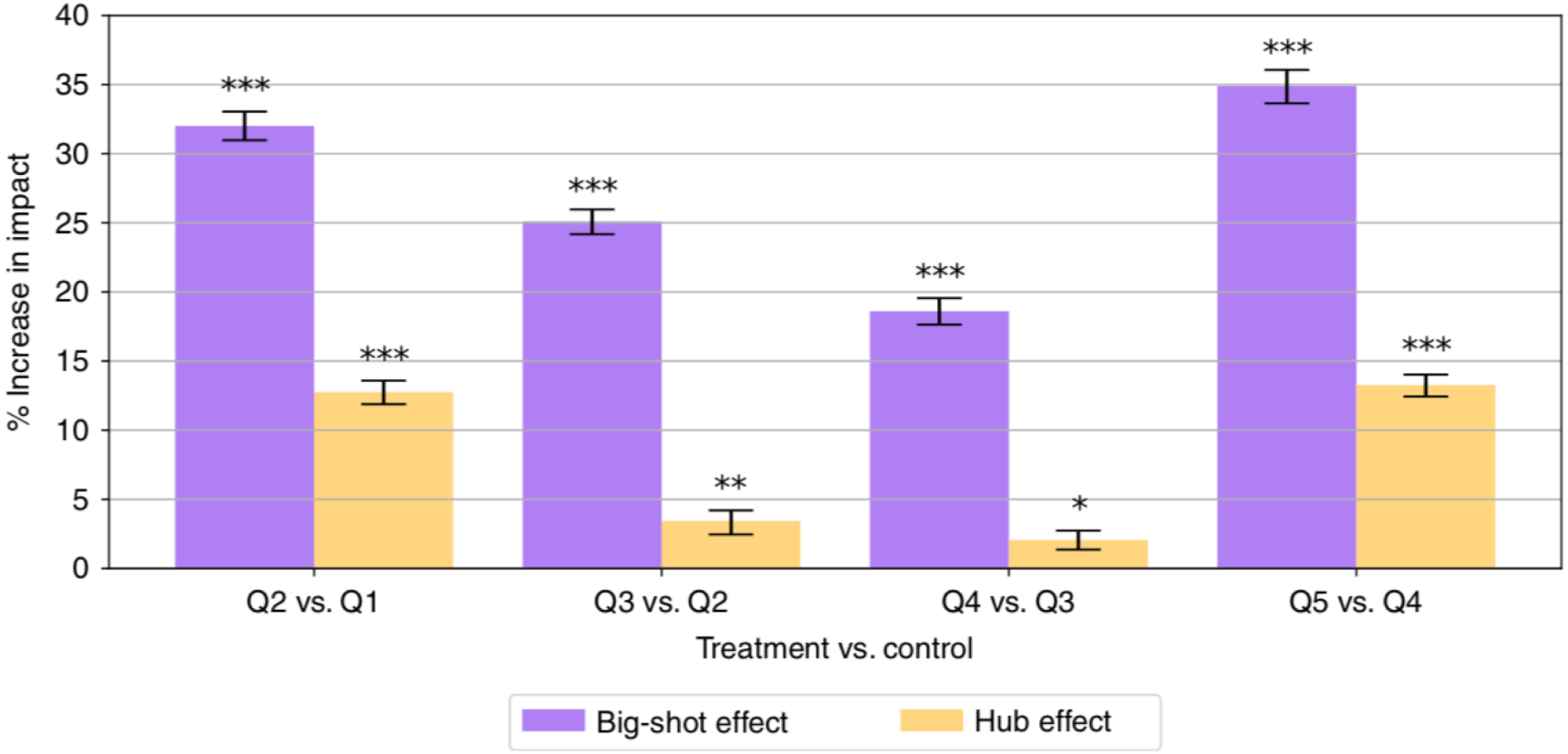


Figure 2: Mentorship quality is associated with protégé Impact

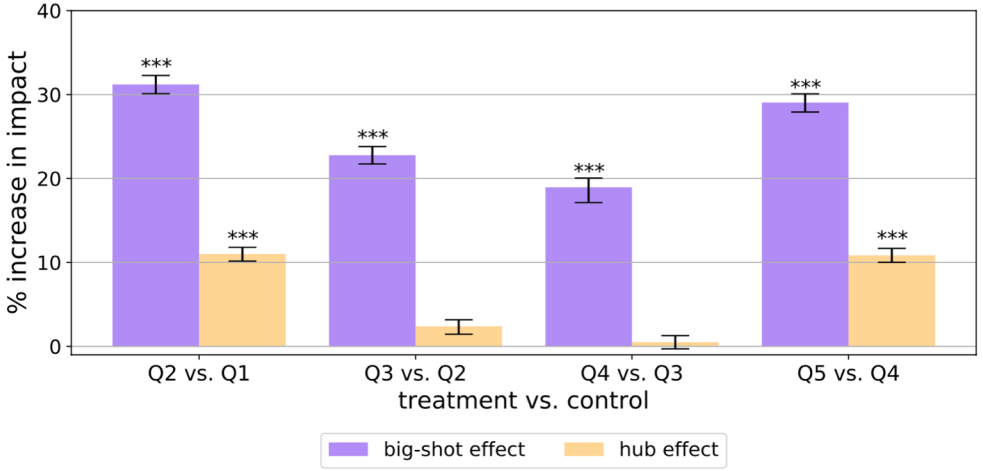


- Split mentors into quintiles
- Build CEMs of protégés across quintiles (Q1 vs Q2, ect)
- Compare protégé impact across the quintiles (t-test)

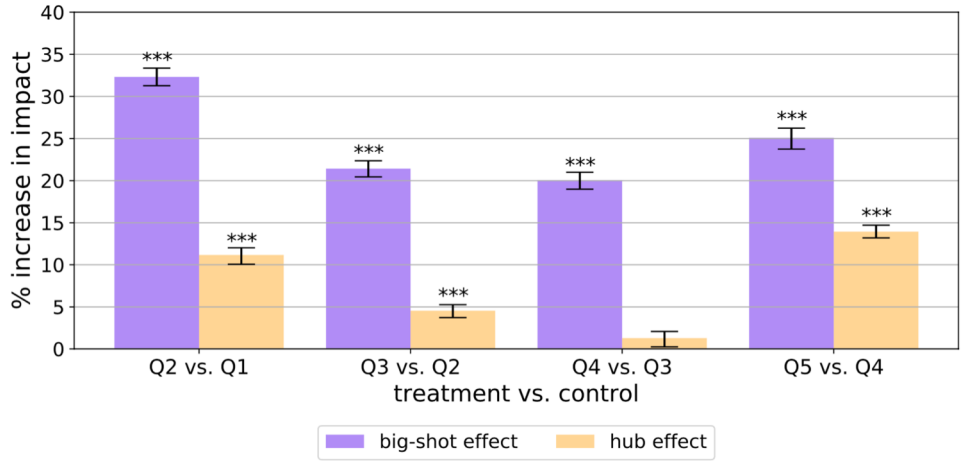
$$\frac{\text{Avg(protégé impact Q2) - Avg(protégé impact Q1)}}{\text{Avg(protégé impact Q1)}} \text{ X } 100$$

Supplemental figures: Claiming robustness of results

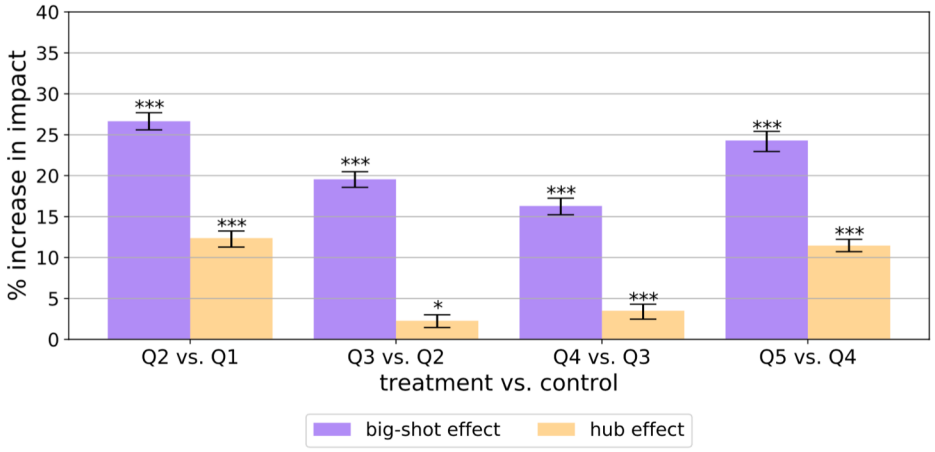
C10 instead of C5



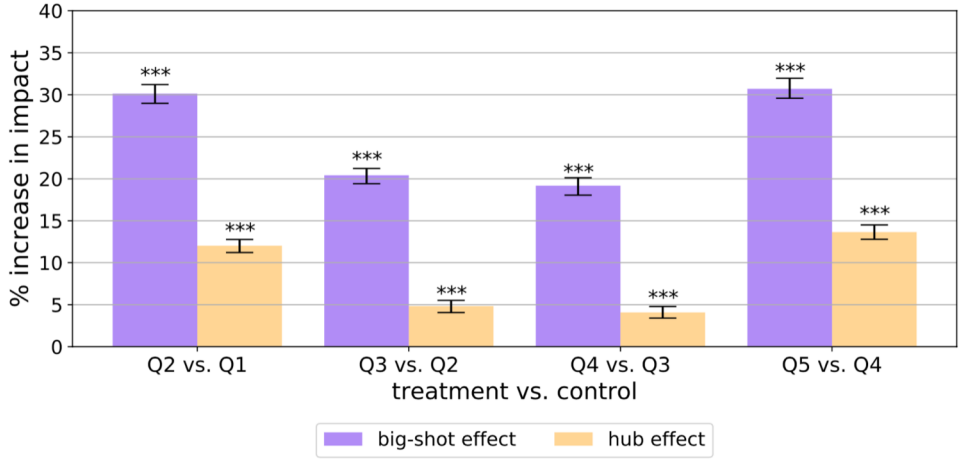
Max of Bigshot and Hub instead of average



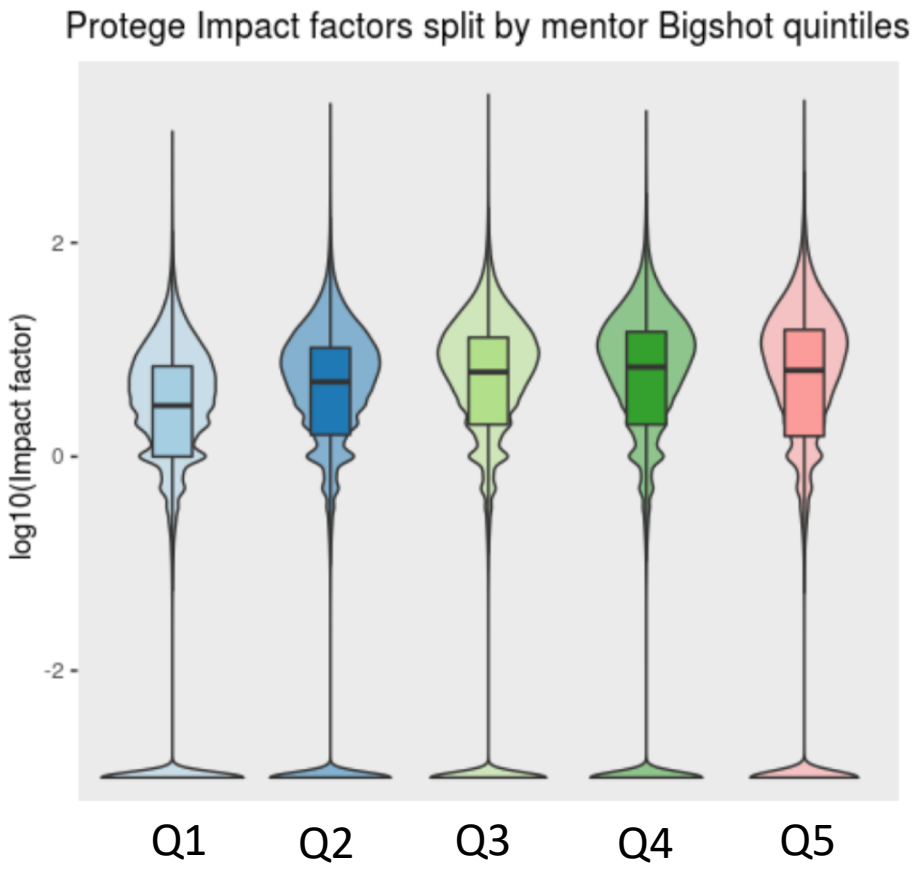
Median of Bigshot and Hub instead of average

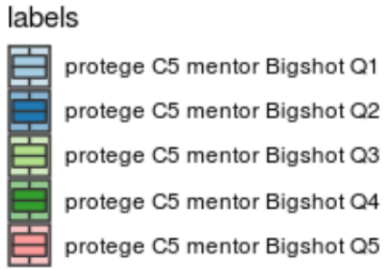


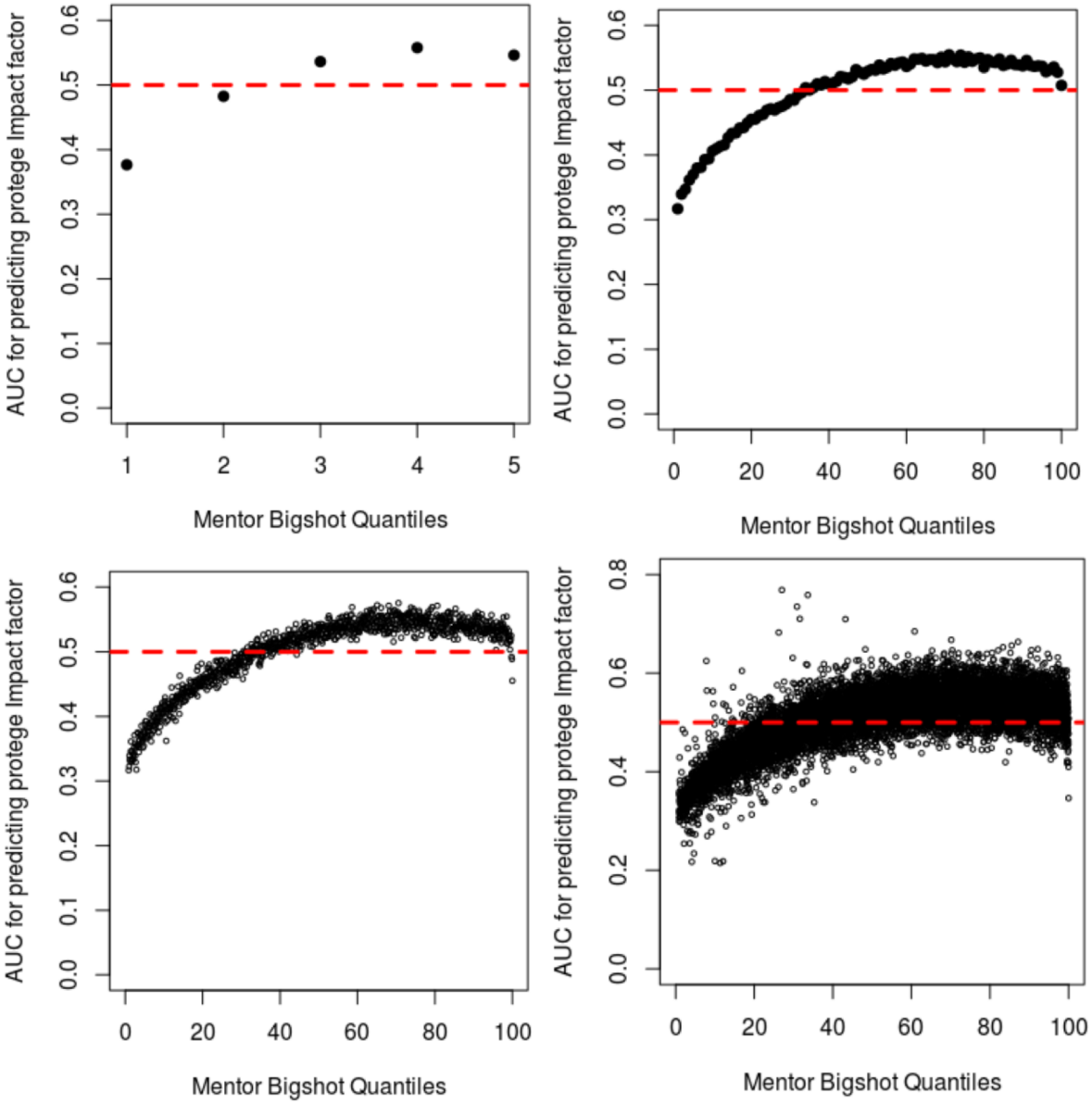
Larger gap in academic age between protégés and mentors



Mentorship quality does not predict protégé impact



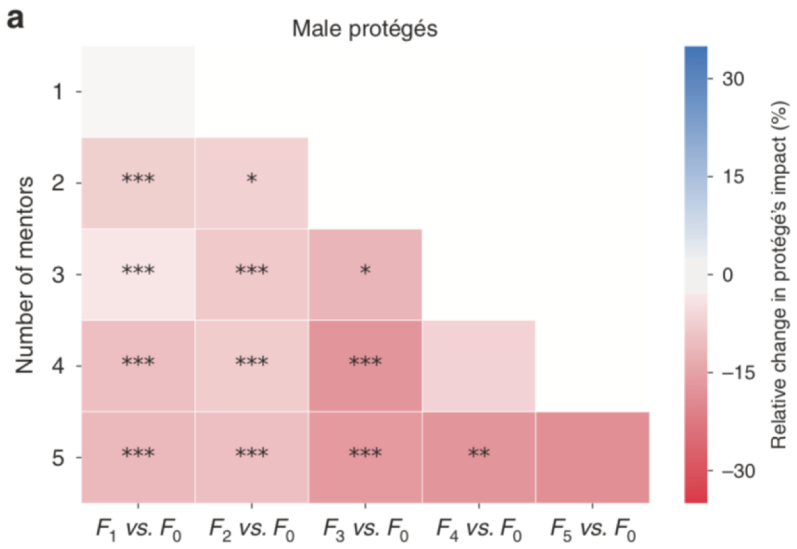


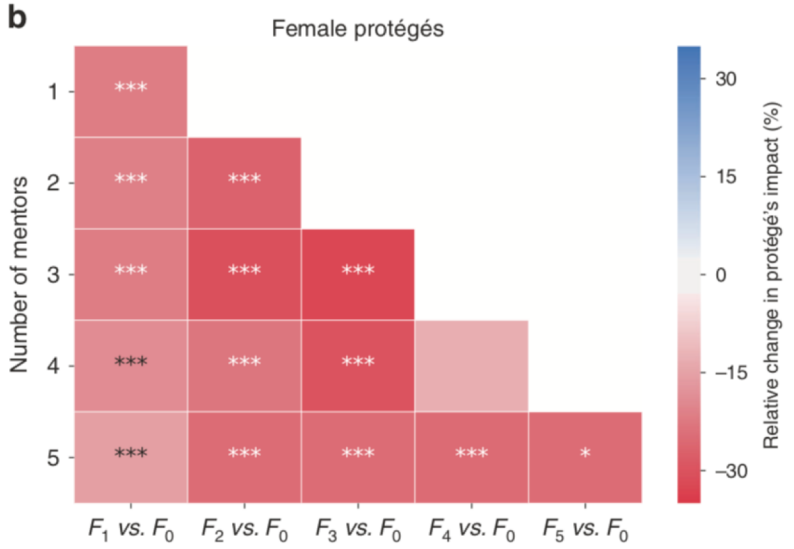


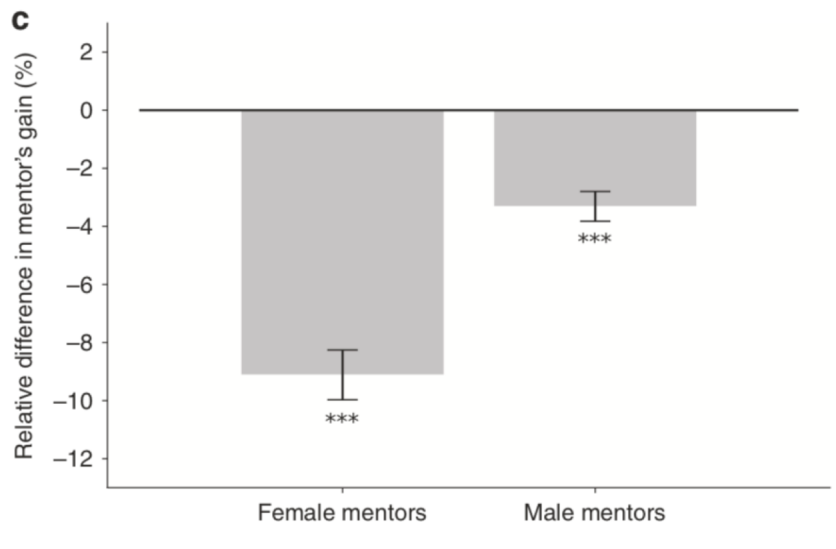
~1 million data points, but with duplicates
~800,000 unique protégé-mentor pairs

Where's the rest of the data?

Results Figure 3







Just demonstrates a general decrease in citations for female authored papers, in my opinion.

C5 of papers mentors published with protégés during the mentorship period

Compare (*f*,F) to (*m*,F)

Compare (*f*,M) to (*m*,M)

Conclusion

- Pros:
  - Attempts to quantify mentorship
  - Aims to use a rather large dataset

- Cons:
  - Too many unsubstantiated assumptions
    - Co-authorship does not equal mentorship
    - Mentor impact factors and collaboration networks do not reflect mentorship quality
  - Complete lack of discussion on historical and systemic bias against women in science
    - Which is a necessity if you're using citations as a measure of 'quality' or 'impact'