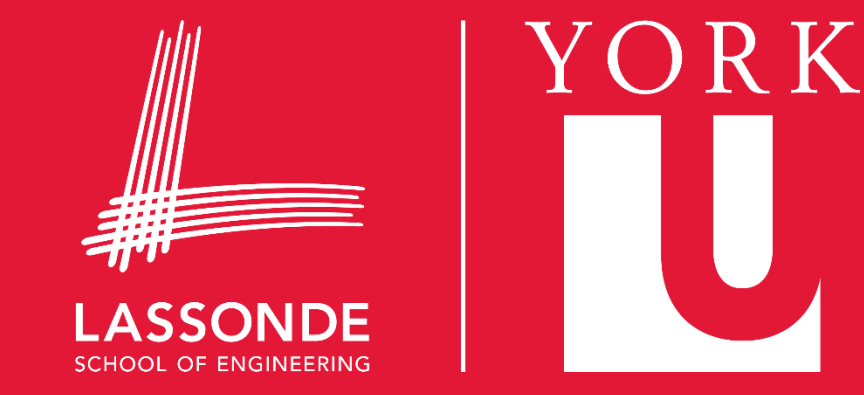


Gender (in)equity in citation practices in software engineering: myth or reality?

Karolina Tchilinguirova and Dr. Alvine Boaye Belle

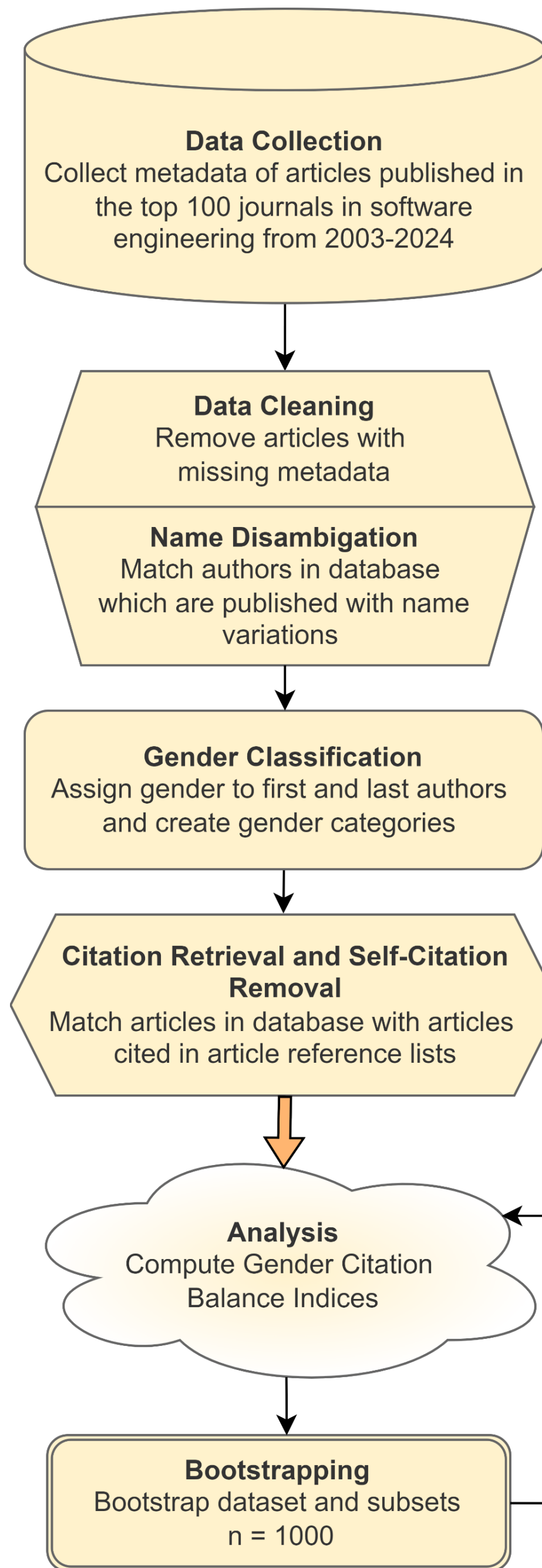
York University, DARE! Research Group



INTRODUCTION

- Citations play a critical role in determining funding allocations, recruitment, promotion decisions, and awards, among other key outcomes.
- Recent studies across various scientific disciplines (e.g., physics, neuroscience) have shown that researchers from certain socio-cultural groups, including women are often cited less frequently than those from dominant groups.
- A citation bias occurs when authors of a given paper decides to either include or exclude a reference based on considerations that are not related to the reference's relevance and/or quality.
- Citation bias can lead to unfair and inaccurate assessments of an article's impact.
- In this paper, we present what we believe to be the first study analyzing gendered citation practices within the software engineering field.

METHODOLOGY



Our process consisted of 6 main steps. Steps 2, 3, and 4 were conducted by leveraging Dworkin et al.'s open-source R code. Steps 5 and 6 were processed using our own Python code.

Step 1 – Downloaded the metadata from the top 100 software engineering journals as ranked by eigenfactor on Web Of Science.

Step 2 – Articles with missing DOIs and names are removed from our collection.

Authors who are published under multiple name variations were identified and linked together.

Step 3 – The US Social Security Administration and Gender-API name databases were used to assign gender based on first names. If both first and last authors met the probability threshold of 0.7, the article was grouped into 1 of 4 categories.

METHODOLOGY

- MM = First and last authors are men
- MW = First author is a man, and last author is a woman
- WM = First author is a woman, and last author is a man
- WW = First and last authors are women

Step 4 – DOIs from reference lists were retained if those references also existed in our database. Self-citations were identified as a cited paper whose first or last author is the same as the first or last author of the citing article, then removed.

Step 5 – Gender Citation Balance Indices were computed:

$$\text{Gender Citation Balance Index} = \frac{\text{observed proportion} - \text{expected proportion}}{\text{expected proportion}}$$

$$\text{expected proportion} = \frac{\sum \text{articles (gender category)}}{\sum \text{articles}} \times 100$$

$$\text{observed proportion} = \frac{\sum \text{citations (gender category)}}{\sum \text{citations}} \times 100$$

Step 6 – 95% confidence intervals of the computed indices were calculated by bootstrapping dataset 1000 times.

RESULTS

Figure 1 illustrates the expected and observed proportion of each gender category.

- MM category dominates the authorship and citations while the WW category is underrepresented.
- Women involved categories are growing.
- Woman involved citations are growing at a rate of 3.6% slower than the rate of woman involved authorship.**

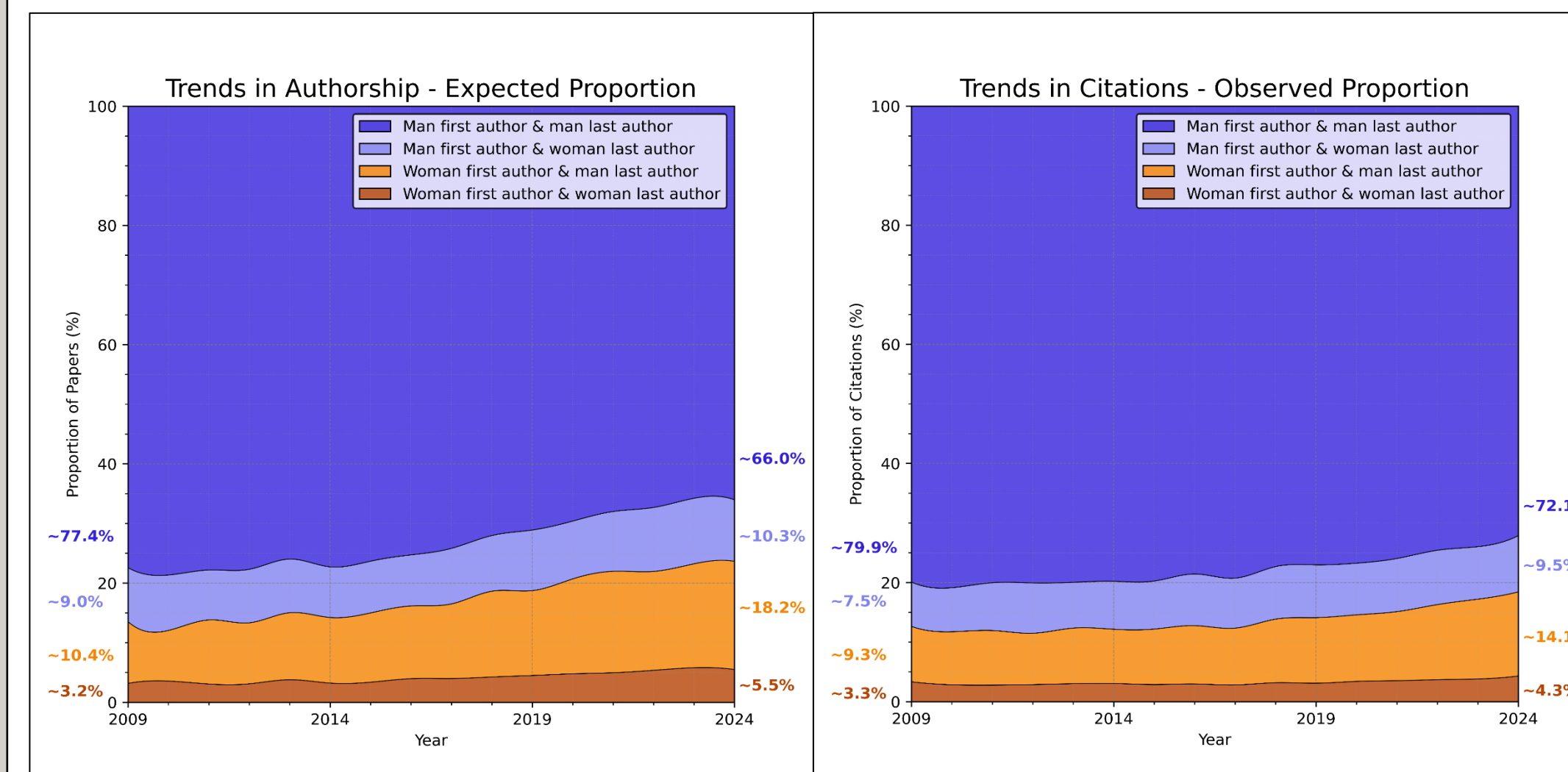


Fig. 1. Comparison of expected and observed proportions of gender through time in SWE literature.

RESULTS

Figure 2. illustrates the Gender Citation Balance Index equation applied to the full analysis dataset.

- MM gender category is over-cited.**
- MW, WM and WW categories are significantly under-cited, with **WW being the most under-cited.**

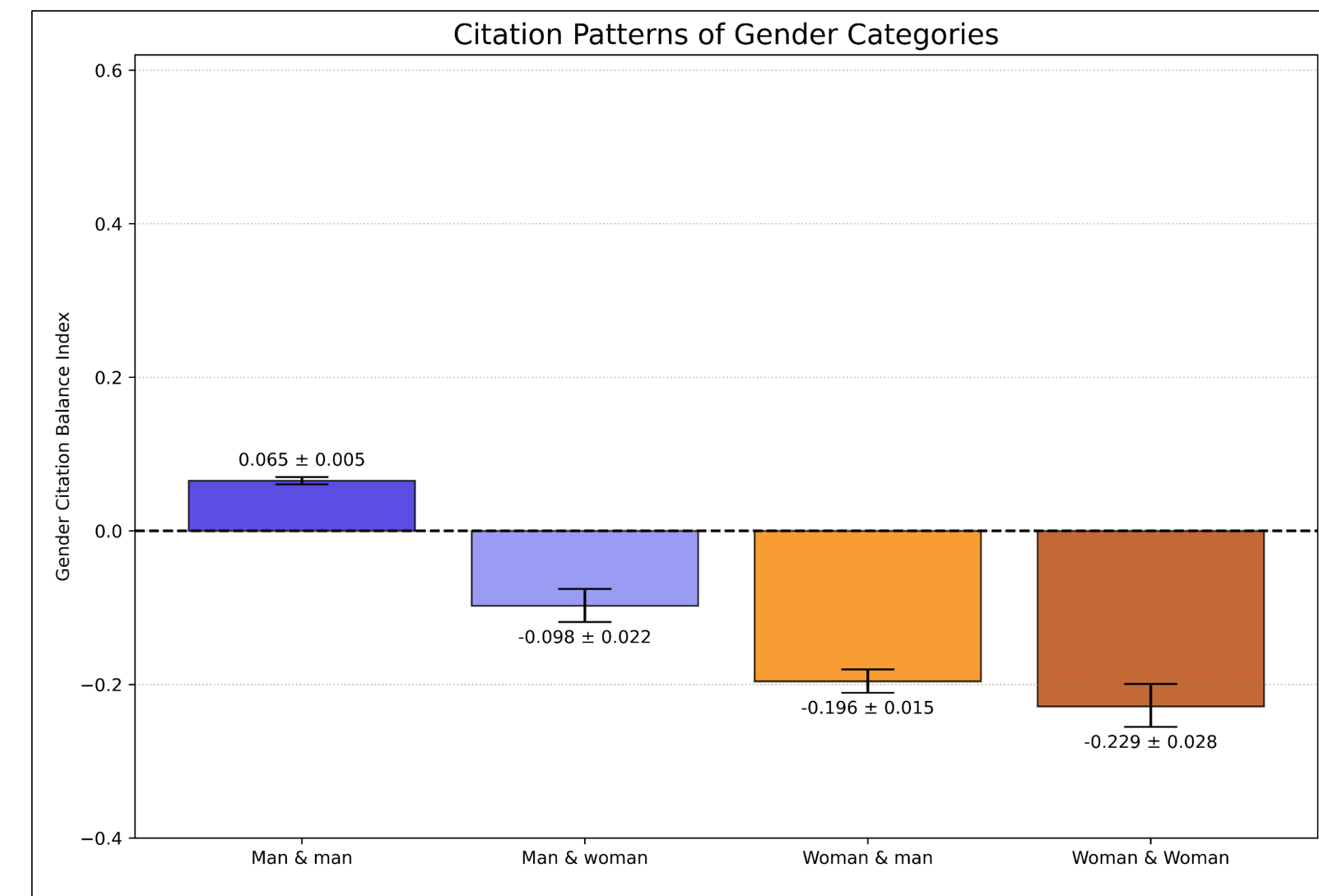
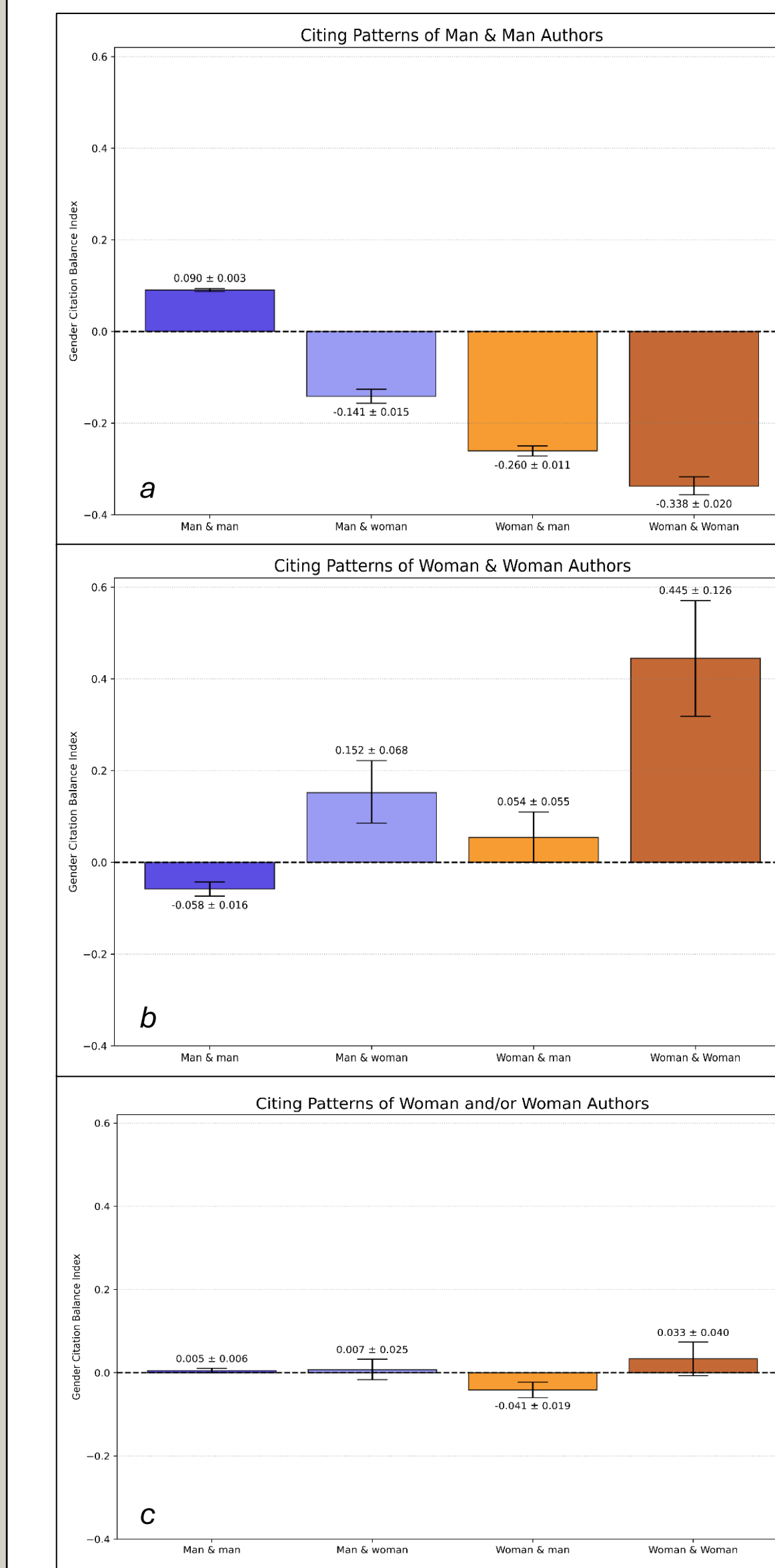


Fig. 2. Gender Citation Balance Indices of the 4 gender categories.



Men usually tend to cite men more than women.

Figure 3a shows that the MM category over-cites the MM category.

Women usually tend to significantly cite women over men.

Figure 3b shows that the WW category over-cites the WW category.

When women are included, citation bias decreases.

Figure 3c shows that citation bias is nearly nonexistent in the WUW category.

Fig. 3. Gender Citation Balance Indices of each category as cited by respective subsets.

RESULTS

Figure 4. illustrates the Gender Citation Balance Index equation applied year by year with respect to the MM and WuW categories.

- This plot shows that the **disparity of citations between man and woman authored articles is increasing through time.**

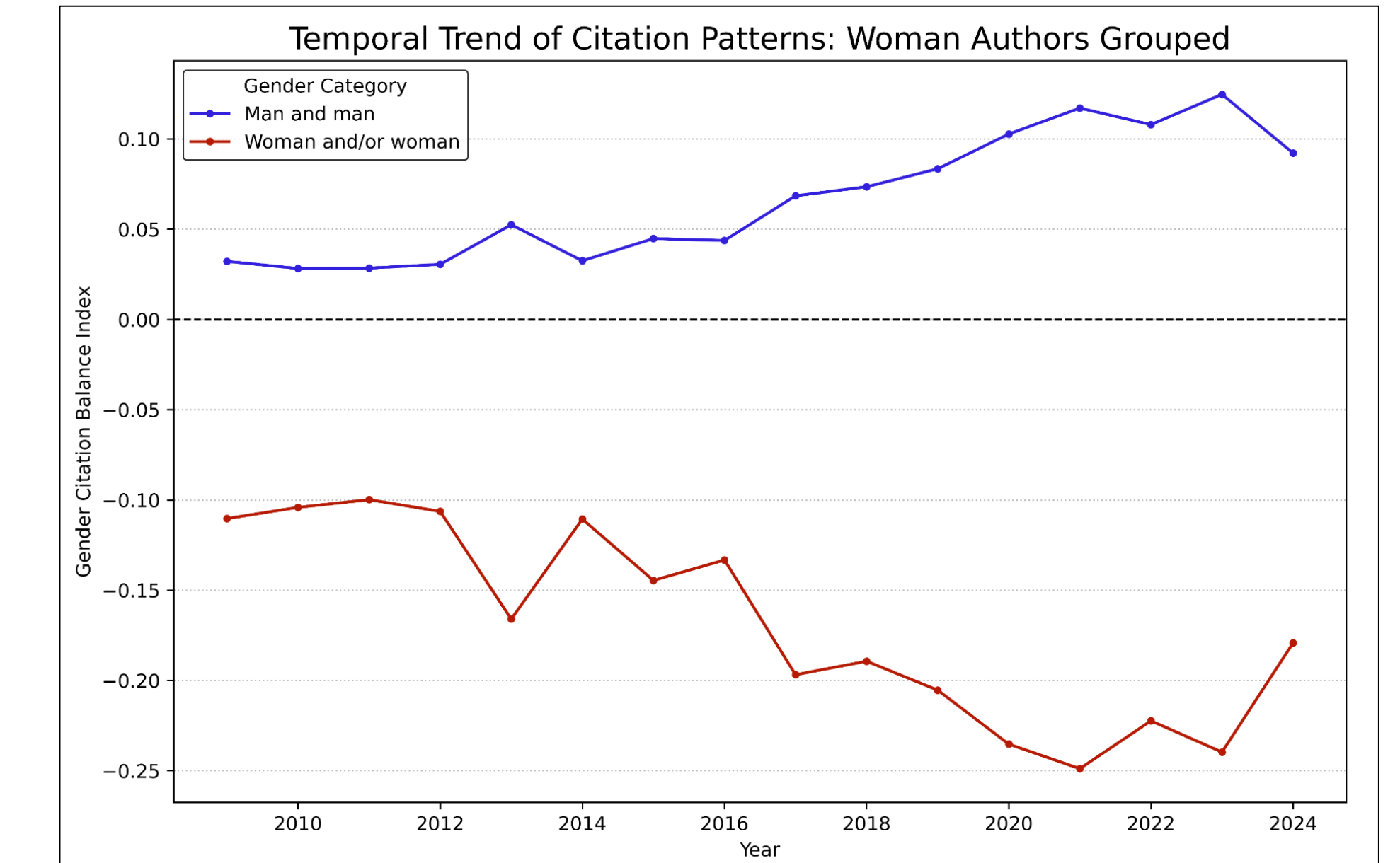


Fig. 4. Temporal evolution of the Gender Citation Balance Index for the WuW and MM category (2009 - 2024).

CONCLUSION

Our study examined potential biases in citation practices within the software engineering field, particularly concerning female authors.

- Our results are consistent with findings in other fields, such as neuroscience, physics, and communication**
- This indicates that gender disparities are widespread in scientific research, including computing fields like software engineering.**

To address this issue, it is crucial for journal editors and conference chairs to recommend the inclusion of **citation diversity statements (CDSs)**.

Efforts to improve fairness in citation practices may also involve **redefining power dynamics** across scientific communities, industry, and academia, fostering international collaboration among female researchers.

Future work will:

- Explore additional attributes that may contribute to citation bias in software engineering literature
- Study the broader impact of citation practices on researchers' careers, including recruitment, promotion, leadership, awards, publication rates, and opportunities for collaboration and tenureship.

REFERENCES

- Fulvio, J. M., Akinola, I., & Postle, B. R. (2021). Gender (im) balance in citation practices in cognitive neuroscience. *Journal of Cognitive Neuroscience*, 33(1), 3-7.
- Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature neuroscience*, 23(8), 918-926.
- Knobloch-Westervick, S., & Glynn, C. J. (2013). The Matilda effect—Role congruity effects on scholarly communication: A citation analysis of Communication Research and Journal of Communication articles. *Communication Research*, 40(1), 3-26.
- Rowson, B., Duma, S. M., King, M. R., Efimov, I., Saterbak, A., & Chesler, N. C. (2021). Citation diversity statement in BMES journals. *Annals of biomedical engineering*, 49, 947-949.
- Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political analysis*, 26(3), 312-327.