



# Gendered citation patterns among the scientific elite

Kristina Lerman<sup>a,1</sup>, Yulin Yu<sup>b</sup>, Fred Morstatter<sup>a</sup>, and Jay Pujara<sup>a</sup>

Edited by Susan Fiske, Princeton University, Princeton, NJ; received April 8, 2022; accepted August 22, 2022

**Diversity in science is necessary to improve innovation and increase the capacity of the scientific workforce.** Despite decades-long efforts to increase gender diversity, however, women remain a small minority in many fields, especially in senior positions. The dearth of elite women scientists, in turn, leaves fewer women to serve as mentors and role models for young women scientists. To shed light on gender disparities in science, we study prominent scholars who were elected to the National Academy of Sciences. We construct author citation networks that capture the structure of recognition among scholars' peers. We identify gender disparities in the patterns of peer citations and show that these differences are strong enough to accurately predict the scholar's gender. In contrast, we do not observe disparities due to prestige, with few significant differences in the structure of citations of scholars affiliated with high-ranked and low-ranked institutions. These results provide further evidence that a scholar's gender plays a role in the mechanisms of success in science.

gender | bibliometrics | science of science | gender disparities

Gender disparities persist in many fields of science. Despite long-running efforts to increase women's representation in the scientific workforce, they continue to face barriers to advancement. Women are less likely than their male peers to be mentored by eminent faculty (1) and to be hired and promoted (2, 3). Women publish in less prestigious journals (4), have fewer collaborators (5), and are underrepresented among journal reviewers and editors (6), and their papers receive fewer citations (7, 8). The multifaceted gender disparities create a "glass ceiling," an invisible barrier that fundamentally limits professional recognition for even the best women scientists (9). As a result, the share of women in higher academic positions decreases steadily (3), with relatively few becoming full professors or receiving prestigious awards. For example, among physics faculty in 4-y colleges and universities, women represent 23% of assistant professors and 18% of associate professors but 10% of full professors (10). Similarly, women represent just 1.8% of Nobel laureates in physics, 3.7% in chemistry, and 2.2% in economics. The dearth of prominent women leaves fewer potential mentors and role models for the next generation of women scientists.

We study scholars elected to the National Academy of Sciences (NAS). Created by US Congress in 1863, NAS is one of the oldest and most prominent professional science organizations. New members are elected by current members based on a distinguished record of scientific achievement. We hypothesize that complex gender differences will be visible within this group of elite scientists. Our study has four main findings. First, we confirm that women are a minority of NAS members in seven fields that include sociobehavioral and physical sciences. Second, we construct citation networks that capture the structure of recognition of each NAS member among their peers. After accounting for field-specific variation in citing, we identify gender differences in the structure of citations networks. Third, we show that these differences are systematic enough to allow us to accurately classify the member's gender based on their citation network alone. Finally, in contrast to gender, we do not observe many disparities due to the prestige of a member's institutional affiliation. Although members affiliated with less prestigious institutions are a minority of NAS members and, similar to women, receive fewer citations than members with more prestigious affiliations, there are few significant differences in the structure of their citation networks.

Gender disparities among elite scientists extend beyond the number of citations: women who join one of the most prestigious scientific organizations differ from similarly achieving men in the structure of peer recognition within their research communities. These results provide further evidence about the importance of gender in the mechanisms of success in science.

## Results

**Membership Gender Gap.** Fig. 1*A* shows the election year of current members of the NAS. Although women have made gains in recent decades, they remain a small minority of members. Among the fields we study, psychology has the highest share of women (42%),

Author affiliations: <sup>a</sup>Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292; and <sup>b</sup>School of Information, University of Michigan, Ann Arbor, MI 48109

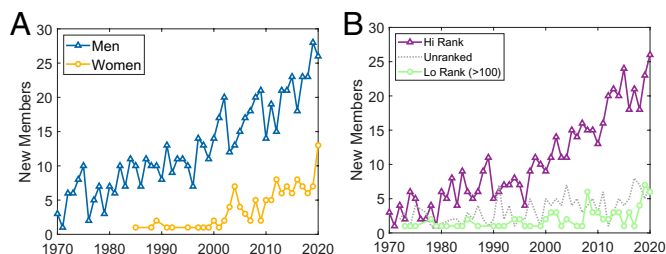
Author contributions: K.L., F.M., and J.P. designed research; K.L. and Y.Y. performed research; Y.Y. analyzed data; and K.L., Y.Y., F.M., and J.P. wrote the paper.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

<sup>1</sup>To whom correspondence may be addressed. Email: lerman@isi.edu.

Published September 26, 2022.



**Fig. 1.** Number of members elected to the NAS split by year and (A) gender or (B) prestige of the member's institutional affiliation. Only members active in seven fields as of 2021 are considered.

followed by sociology (22%), astronomy (18%), chemistry (13%), economics (12%), computer science (11%), and physics (8%).

Academy membership is also highly skewed toward prestigious institutions, which we define as being among the top 100 institutions according to the 2015 Times Higher Education World University Ranking (*Materials and Methods*). Relatively few NAS members come from lower-ranked or unranked institutions (Fig. 1B).

**Gendered Citation Patterns.** We construct an author citation ego network for each NAS member, i.e., ego (*Materials and Methods*). Fig. 2 A–C show the networks of three psychologists (ego shown in red). The ego networks capture the structure of attention among the member's peers, or who knows who within the ego's research community.

Men have more lifetime citations, on average, than women (34,880 vs. 21,062), a large gap that remains significant ( $p < 10^{-5}$ ) after standardizing to account for variation in citation across fields. However, we also find differences in the structure of citation networks, which we characterize with network features (*Materials and Methods*). We see gender disparities in the distribution of network features, which create statistically significant gaps in their means (Fig. 2D). Women reciprocate a significantly higher share of citations than men (ego mutual edges). This also applies to the ego's peers: compared to men, a higher share of women's peers cite researchers who cite them (mutual edges). Women's ego networks have higher average degree, edge density, and clustering coefficient. Together, these features suggest that women are more tightly embedded within their research communities. This is consistent with previous findings that women tend to gravitate to certain communities (11). Women have fewer peers than men, but these peers are more productive (publish more papers) and receive more citations. Finally, women NAS members have more women among their peers.

In contrast, citation networks of members from prestigious institutions are not substantially different from their counterparts with less prestigious affiliations (Fig. 2E). Although the citation gap between the former and the latter is 33,375 vs. 24,720, on average (and remains large and significant after standardization), the only significant differences in ego networks are for average degree, shortest path, and the number of papers published by peers.

**Gender Classification.** We trained a classifier to use ego network features to predict the ego's gender. On balanced data, the classifier achieved good performance as measured by area under the receiver operating curve (AUC) of 0.78. In contrast, using network features to predict the prestige of the ego's affiliation resulted in AUC of 0.48, no better than random guess. Thus, the structure of peer recognition is informative about researcher's gender but not the prestige of their institutional affiliation.

## Discussion

Despite decades-long efforts to improve the climate for women in science, many barriers remain to their professional advancement. As a result, the share of women in senior positions remains low. Previous works have shown that papers published by women receive systematically fewer citations than papers written by men (7, 8). We showed that gender disparities extend beyond the citation count. We compared citation networks of distinguished scholars who were elected to the NAS and identified gender-based differences in their structure. Women are embedded within more tightly knit research communities: they reciprocate their peers' citations (mutual citations) at a higher rate than men, and their peers are more productive, on average, receive more recognition, and cite others within the research community more (higher clustering coefficient). Women also count more women among their peers. These structural differences are strong enough to accurately classify the scholar's gender.

Inclusion in NAS is one marker of scientific success. Differences in citation network structure suggest that there are multiple pathways to success. However, as we show, gender is a key factor differentiating structural features. Although women are cited less than men, their close-knit peers may offer benefits, such as social capital, that compensate for lower citations. These differences signal that societal forces may shape the pathways to success available to scientists based on their gender. However, our study does not elucidate the societal forces responsible for this differentiation in career development, nor how early within the career it begins. Identifying causes for this critical difference in citations merits further research.

Diversifying the scientific workforce by increasing the share of women in science is a vital societal goal. Demographically diverse groups produce more innovative and equitable scholarship (12, 13), which is necessary to address today's complex challenges, such as climate change and emerging infectious diseases. Our work provides a framework to better understand—and then mitigate—gender disparities in the mechanisms of success in science.

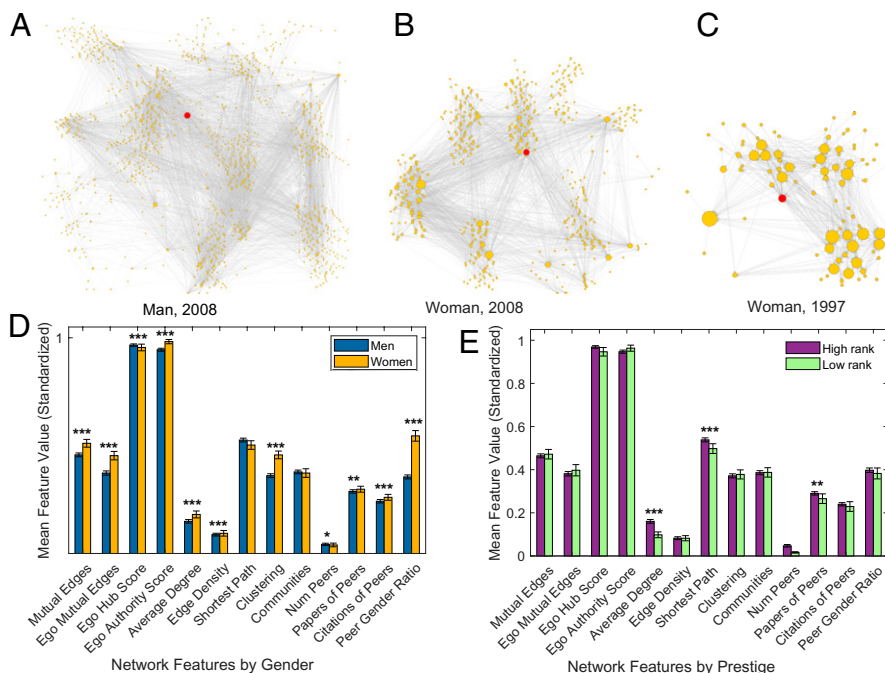
## Materials and Methods

**Data.** We scraped the NAS website in 2021 for the list of members (<http://www.nasonline.org/member-directory/>). We collected the name, affiliation, biographical sketch, and the year of election for members whose primary section was Astronomy, Physics, Chemistry, Computer and Information Sciences (CS), Psychological and Cognitive Sciences (Psychology), Social and Political Sciences (Sociology), or Economic Sciences (Economics).

We used Microsoft Academic Graph (MAG) (14) to collect information about members' papers and citations. MAG contains metadata about more than 150 million scientific publications. We matched members' names to a MAG author ID by searching for the member's name and extracting ID associated with most citations among multiple matching results, for a total of 766 scholars (120 women).

We use Genderize.io to scale up gender classification. This service creates a proxy of gender from the author's name. While the imperfect proxy ignores nuances of gender identity, we believe it is appropriate for this study as names provide signals about gender that citing authors use if they do not know the author. We checked gender labels of all NAS members against pronouns used in their biosketches or web pages. We found that very few names were misclassified (1%), but 9% had missing gender (e.g., non-Western or ambiguous names). We manually corrected all errors.

We used the 2015 Times Higher Education World University Ranking (THE-WUR; <https://www.timeshighereducation.com/world-university-rankings/2015/world-university-rankings>) (15) to rank institutional affiliations of NAS members. A lower number in the ranking indicates a more prestigious affiliation. Note that THE-WUR does not rank nonacademic institutions.



**Fig. 2.** Citation ego networks and features. (A–C) Ego networks of three psychologists elected during the specified year. Only edges representing three or more cited papers are included. The nodes are sized by centrality, with the ego shown in red. Comparison of mean ego network features split by ego's (D) gender and (E) institutional prestige. Statistically significant differences in the means are marked by asterisks: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .

**Citation Ego Networks.** In an author citation network, a directed edge between authors  $u$  and  $v$  with weight  $w$  exists when author  $u$  cites  $v$ 's work in  $w$  of his or her papers. We construct an author citation ego network for each NAS member, i.e., the ego. We first identify the papers published by the ego during the year prior to election. We include in the ego network all the authors the ego cites in these papers, plus all the authors who cited these papers in their own publications, and citation edges among all authors. We genderized peers' names. A random spot check of 200 names showed that 127/128 were correct and 72/200 were missing labels.

**Network Features.** We used the features below to capture the structure of ego networks. Except for the first four features that are calculated on the directed network that includes the ego, we calculate the features of the author citation ego network after removing the ego (and its edges) and symmetrizing edges to ignore their direction.

- Mutual edges are the fraction of edges that are bidirectional.
- Ego mutual edges are the fraction of bidirectional ego's edges.
- Ego hub/authority score is the ego's percentile centrality ranking.
- Average degree is average node degree in the ego network.
- Edge density is the fraction of all possible edges that exist.
- Shortest path is the average shortest path in the ego network.
- Clustering is the fraction of a node's neighbors that are connected.
- Communities are the number of communities identified by Blondel's method (16).

- Peers are nodes in the ego network not counting the ego.
- Papers of peers are the average number of papers published by peers.
- Citations of peers are the average citations received by peers.
- Peer gender ratio is the fraction of peers with women's names.

To partly control for different citation rates across fields, we standardized network features based on the ego's primary section: we rescaled the feature value by its range across all members of that section.

**Gender Classification from Citation Networks.** We used ego network features to predict the ego's gender or affiliation prestige. For each classification task, we created a balanced dataset containing all egos from the minority class (women or low-ranked affiliations) and the same number of egos chosen at random from the majority class (men or high-ranked affiliations). We trained a random forest classifier on a random 75% subset of the data and tested on the remaining 25% of the data. The classifier parameters were set to their default values in sklearn.

**Data, Materials, and Software Availability.** Previously published data were used for this work (14). Data, code and analyses are available on <https://github.com/yulin-yu/NASGender> (17).

**ACKNOWLEDGMENTS.** This work was supported, in part, by the Defense Advanced Research Projects Agency (contract W911NF192027) and the Air Force Office of Scientific Research (contract FA9550-17-1-0327).

1. J. M. Sheltzer, J. C. Smith, Elite male faculty in the life sciences employ fewer women. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10107–10112 (2014).
2. S. F. Way, D. B. Larremore, A. Clauset, "Gender, productivity, and prestige in computer science faculty hiring networks" in *Proceedings of the 25th International Conference Companion on World Wide Web*, I. Horrocks, B. Zhao, Eds. (Association for Computing Machinery [ACM], 2016), pp. 1169–1179.
3. J. Huang, A. J. Gates, R. Sinatra, A. L. Barabási, Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4609–4616 (2020).
4. C. O. Ross, A. Gupta, N. Mehrabi, G. Muric, K. Lerman, The leaky pipeline in physics publishing. *arXiv [Preprint]* (2020). <https://doi.org/10.48550/arXiv.2010.08912> (Accessed 23 October 2021).
5. L. Ductor, S. Goyal, A. Prummer, Gender and Collaboration. *Rev. Econ. Stat.*, 1–40 (2021).
6. M. R. Berenbaum, Speaking of gender bias. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8086–8088 (2019).
7. E. G. Teich *et al.*, Citation inequity and gendered citation practices in contemporary physics. *arXiv [Preprint]* (2021). <https://doi.org/10.48550/arXiv.2112.09047> (Accessed 26 December 2021).
8. M. L. Dion, J. L. Sumner, S. M. Mitchell, Gendered citation patterns across political science and social science methodology fields. *Polit. Anal.* **26**, 312–327 (2018).
9. C. Avin *et al.*, "Homophily and the glass ceiling effect in social networks" in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, T. Roughgarden, Ed. (ACM, 2015), pp. 41–50.
10. A. M. Porter, R. Ivie, Women in physics and astronomy, 2019 report. AIP Statistical Research Center (2019). <https://eric.ed.gov/?id=ED594227>. Accessed 20 March 2022.
11. M. O. Jackson, Inequality's economic and social roots: The role of social networks and homophily (2021). <https://ssrn.com/abstract=3795626>. Accessed 20 March 2022.
12. A. Joshi, By whom and when is women's expertise recognized? The interactive effects of gender and education in science and engineering teams. *Adm. Sci. Q.* **59**, 202–239 (2014).
13. L. G. Campbell, S. Mehtani, M. E. Dozier, J. Rinehart, Gender-heterogeneous working groups produce higher quality science. *PLoS One* **8**, e79147 (2013).
14. A. Sinha *et al.*, "An overview of Microsoft Academic Service (MAS) and applications" in *Proceedings of the 24th International Conference on World Wide Web*, A. Gangemi, S. Leonardi, A. Panconesi, Eds. (ACM, 2015), pp. 243–246.
15. Times Higher Education. 2015 World University Rankings. *The Times Higher Education Supplement*. <https://www.timeshighereducation.com/world-university-rankings/2015/world-university-rankings>. Accessed 20 October 2021.
16. V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
17. K. Lerman, Y. Yu, F. Morstatter, J. Pujara, Data, code and analyses for "Gendered citation patterns among the scientific elite." GitHub. <https://github.com/yulin-yu/NASGender>. Deposited 31 August 2022.