Supplementary Materials for "First-mover advantage explains gender disparities in physics citations"

Hyunsik Kong[1], **Samuel Martin-Gutierrez**[1], **and Fariba Karimi**[1,*]

[1]Networked Inequality group, Complexity Science Hub, Josefstaedter Strasse 39, Vienna, 1080, Austria
[*]corresponding author: karimi@csh.ac.at

1 Author name disambiguation

We used a preprocessed version of the APS dataset based on Sinatra et al.[1] where a name disambiguation method was applied to correctly match every author to her papers. With this method, summarized in the flow chart of Figure 1, 237,000 different authors were identified.

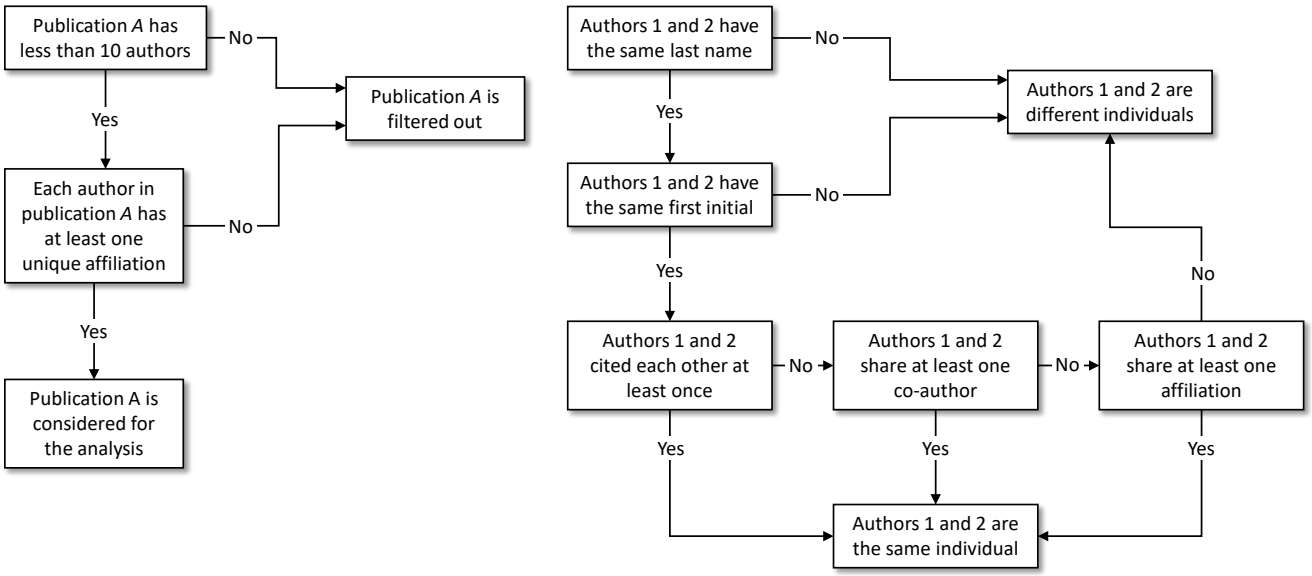

Figure 1. Author name disambiguation algorithm. This flow chart schematizes the author name disambiguation algorithm that Sinatra *et al.* used[1]. The algorithm first decides whether a publication is considered in the analysis. Then, for any two author names 1 and 2, it decides whether they are the same individual or two different authors.

2 Gender detection

In order to detect the gender from authors' names, the first step is to remove those authors whose first name is not mentioned and initialized. No existing name-based gender inference techniques can tackle those cases. For those authors that we had first names available, we first use the application *Genderize*[2]. Then, for the names whose gender this application is unable to infer, we use the picture-based gender inference technique *Face++*[3]. In this second step, we perform a Google image search with the author's first name and family name, and feed the resulting images to *Face++*. This methodology was developed by Karimi *et al.*[4], who compared it with commonly used dictionary-based gender detection techniques and showed that it consistently achieves high accuracy for names of different nationalities. The results they obtained for a random sample of researchers whose names and genders are known are shown in Table 1.

As a preliminary step to use the gender detection technique we performed a thorough standardization of names to avoid issues with the use of special characters. We followed the rules from the Program for Cooperative Cataloguing of the Library of Congress (NACO)[5]. Supplementing the NACO normalization by translating accented characters and other special characters accordingly improves the overall query matching by 63%. Using this methodology we were able to detect the gender of 124,000 authors.

	Sample Size	*SSA*	*IPUMS*	*Sexmachine*	*Genderize*	*Face++*	*Genderize & Face++*
United States	419	82%	76%	84%	83%	91%	91%
China	113	20%	11%	67%	28%	65%	50%
United Kingdom	96	94%	92%	92%	94%	81%	98%
Germany	82	87%	88%	96%	94%	87%	96%
Italy	75	93%	92%	94%	98%	79%	99%
Canada	60	87%	77%	86%	91%	90%	96%
France	58	93%	92%	80%	96%	81%	97%
Japan	56	79%	70%	100%	90%	62%	91%
Brazil	44	29%	29%	15%	44%	81%	90%
Spain	39	96%	92%	92%	100%	92%	100%
Australia	31	89%	89%	90%	86%	86%	94%
India	29	67%	17%	71%	78%	83%	83%
South Korea	27	4%	0%	58%	11%	74%	37%
Switzerland	25	78%	70%	56%	83%	88%	90%
Turkey	21	43%	14%	79%	81%	86%	100%

Table 1. Comparison of gender detection techniques. This table compares the accuracy of various gender inference methodologies for names from 16 different countries. The dictionary-based methods are respectively based on the US Social Security Administration (SSA), the survey from Integrated Public Use Microdata Series (IPUMS), and the publicly available list of names *Sexmachine*. This table is a modified version of Table 2 of[4].

3 Similarity measures for publications

The main objective of this paper is to compare pairs of similar papers in an unbiased fashion. The similarity analysis is based on the concept of bibliographic coupling strength N_{ij} of pairs of articles (i, j), which is defined as the number of common articles cited by both i and j[6,7]. But using N_{ij} without further considerations can lead to misleading results. For example, the similarity between two papers that include each 20 and 25 citations and share $N_{ij} = 5$ of them should not be the same as the similarity between two papers that also share 5 references but respectively cite 65 and 82 publications. On the other hand, within subfields there are usually a handful of very popular publications that are cited in most works (such as review papers), so their inclusion in two different papers may not indicate actual similarity. In order to obtain meaningful measures of similarity, several normalization approaches have been explored.

A widely used measure that addresses the first kind of the issues described above is the Jaccard index. The Jaccard index is computed as the quotient of the cardinality of the intersection and the cardinality of the union of the sets of cited publications by the two papers under consideration. One of the problems of this method is that it considers the weight of all citations to be identical and therefore does not take the significance of each paper into account[8,9]. In addition, narrowing our analysis to counting the common articles may not lead to an accurate interpretation due to the massive differences between male and female sample sizes. The reason is that, if the sizes of the sets of citations of the two papers are very different, their similarity is primarily determined by the size of the smallest one, as their intersection is bounded by the size of the smallest set[10].

Fractional counting is another common normalization technique for bibliographic coupling. In this case, instead of normalizing by the outgoing citations of the two papers of interest, each commonly cited reference contributes to the similarity score with a weight inversely proportional to its number of incoming citations[11,12]. Therefore, fractional counting addresses the second kind of situation discussed above by compensating the disproportionate influence of very popular publications in the similarity score. However, unlike the Jaccard index, it does not take into account the relative size of the sets of outgoing citations.

To overcome the issues of the Jaccard index and fractional counting, we identify couples of similar papers by looking both at the outgoing references of the pair and the incoming citations of the articles they cite. In particular, we perform a statistical test using the hypergeometric distribution as a null model and detect pairs of papers whose set of common outgoing citations has a very low probability of having been generated by chance[13]. In Figure 2 we present a diagram of this methodology, which is explained in Methods in detail.

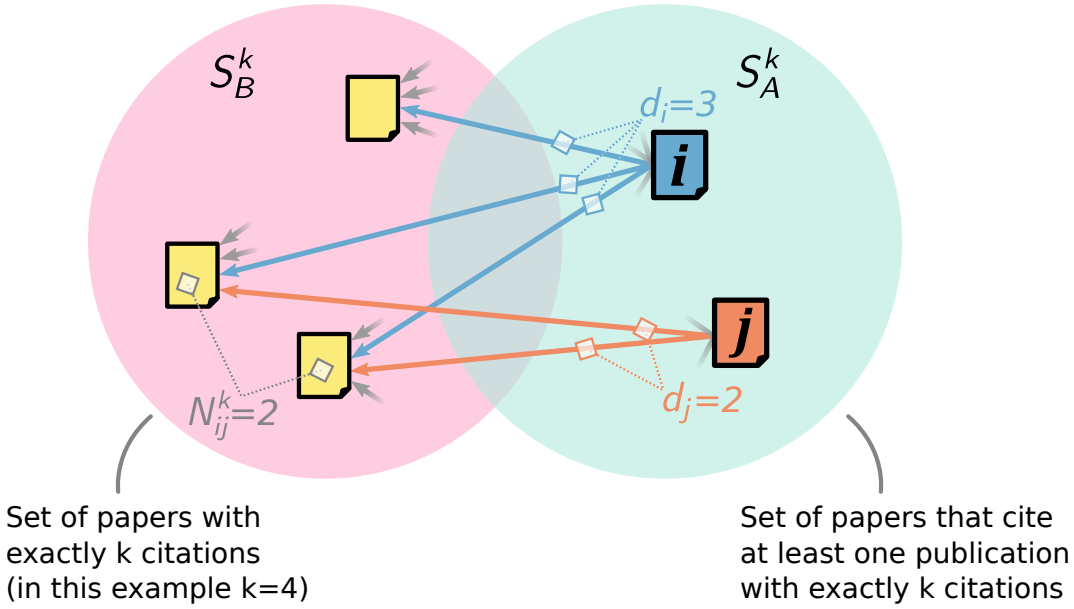

Figure 2. Setting up for the similarity algorithm. This figure sketches the variables involved in the computation of the hypergeometric distribution function used to obtain the paper similarity measure[10].

4 Additional supplementary figures

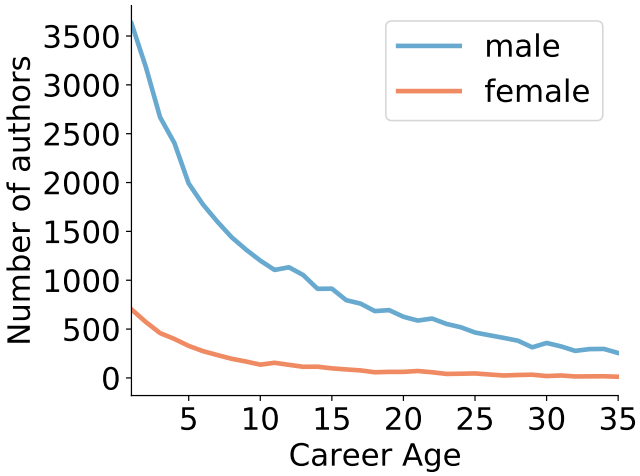

Figure 3. Number of authors by career age. Number of male and female authors by their career age. As mentioned in the main part of the paper, authors with career age of 0 were exempt from this analysis.

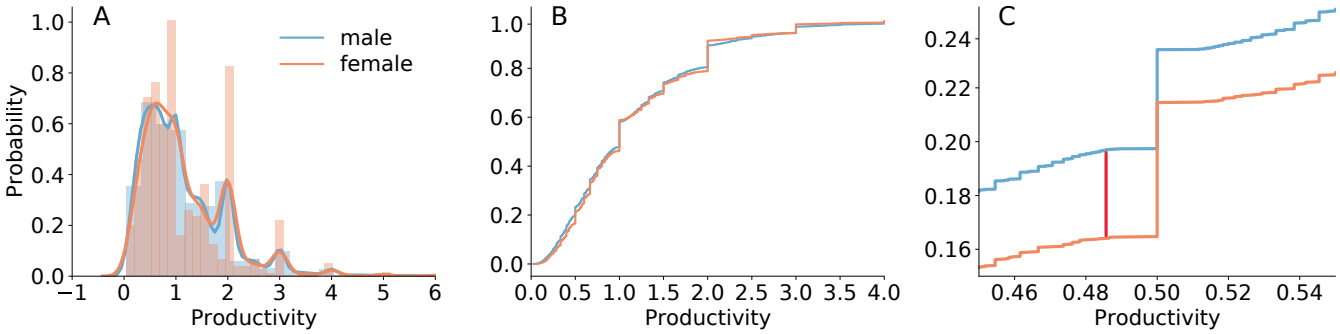

Figure 4. Productivity distribution of authors by gender. A: Probability density functions (PDF) of the productivity distributions of male and female APS authors. B: Corresponding empirical cumulative distribution functions (CDF) of the productivity distributions. C: Same as B but focused on the productivity interval with the maximum difference between the two distributions, indicated with a vertical red line.

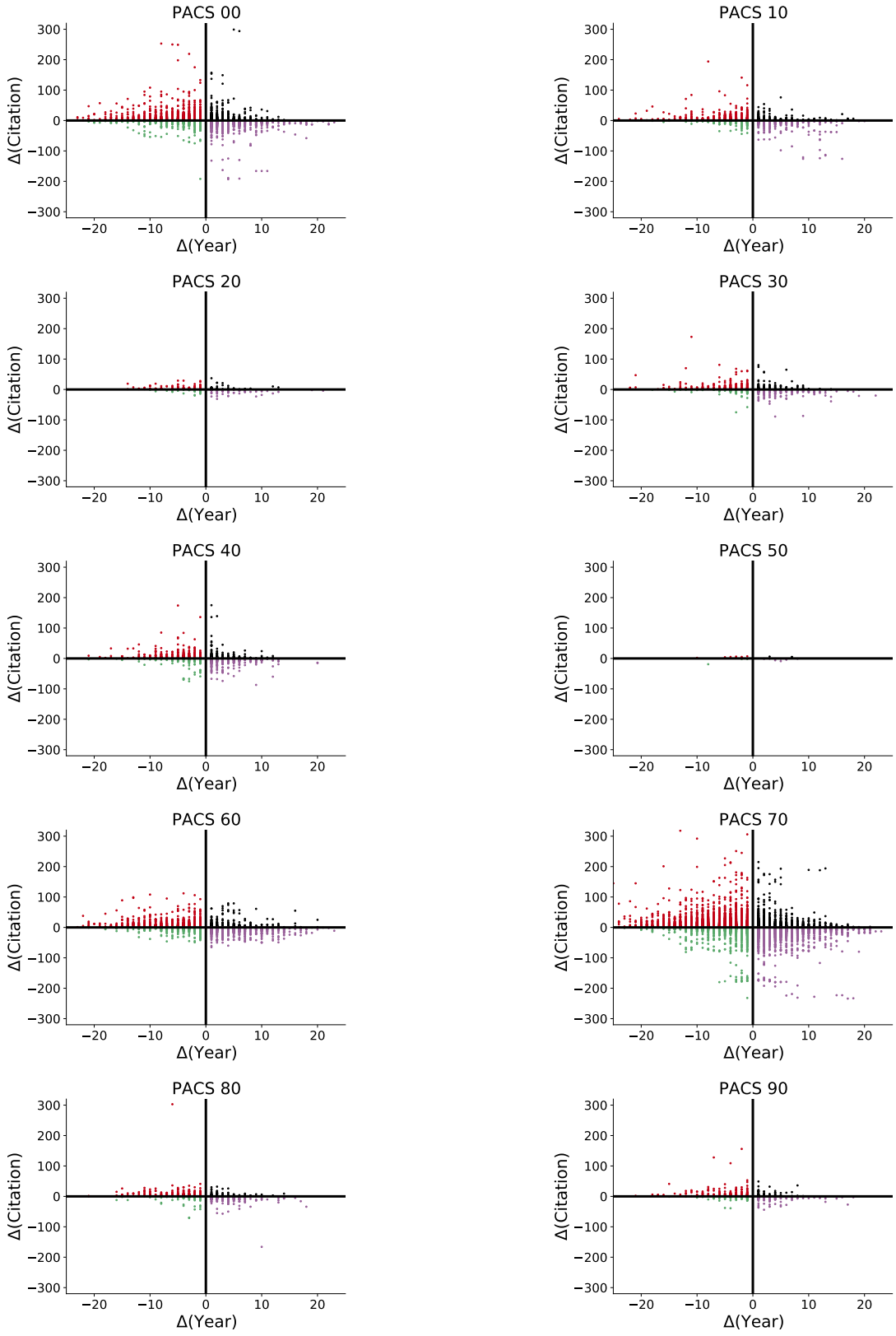

Figure 5. Centrality and year difference for similar pairs of papers. As in Figure 3B of the main document, these scatterplots display the centrality difference and the year difference between similar male-female pairs of papers in each subfield.

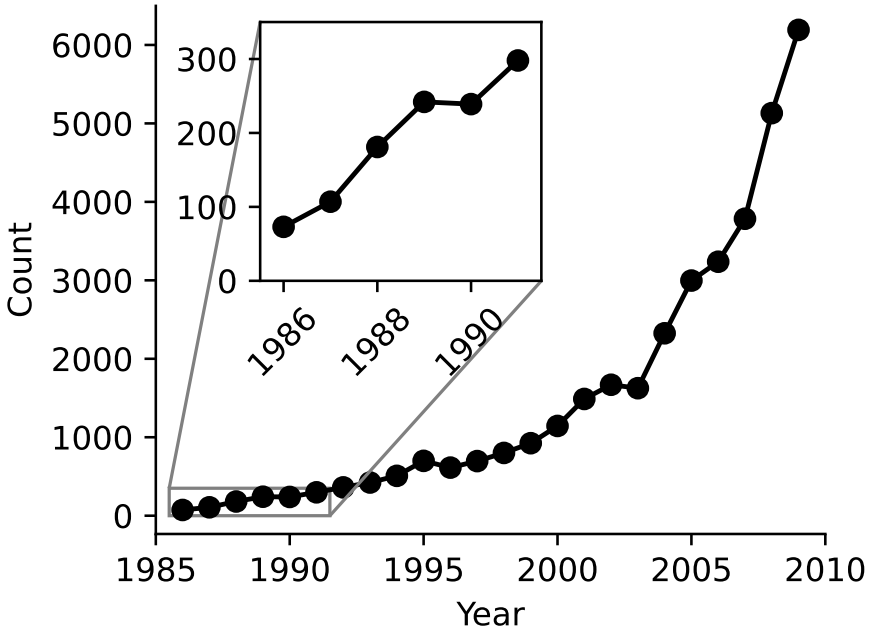

Figure 6. Number of sampled similar pairs of papers by publication year. This figure shows the number of sampled similar pairs for the centrality difference analysis between similar pairs per year (see Figure 3C in the main document).

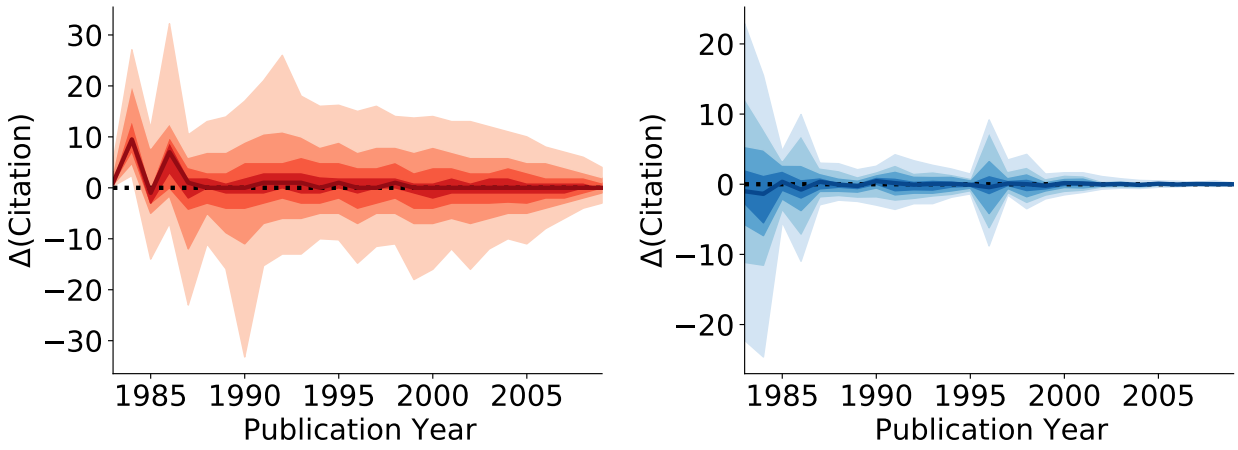

Figure 7. Percentile plots of centrality difference by year. Percentile plots showing the evolution of the distribution of centrality differences for similar male-female (left) and male-male (right) pairs over the years. The mean and standard errors of these distributions are shown in Figure 3C of the main document. Percentiles 10% to 90% are shown in different shades of red (male-female) and blue (male-male) in steps of 10%. The two papers within each pair are published no more than 3 years from each other, and their citation difference is assigned to the year when the latter paper is published. We performed a robustness check by assigning different p^* values and time intervals, and the resulting plots returned similar distributions.

5 Additional supplementary tables

	n	n_{male}	p_{male}	n_{female}	p_{female}
Total # of citations	9,384,218	8,516,293	0.9075	867,925	0.0925
Self-citations	564,630	524,788	0.9294	39,572	0.0706
Self-citation ratio	(6.02%)	(6.16%)		(4.56%)	
Total # of observed authors	68,505	58,888	0.8596	9,617	0.1404
Self-citing authors	36,070	31,987	0.8868	4,083	0.1132
Self-citing author ratio	(52.65%)	(54.32%)		(42.46%)	

Table 2. Proportion of self-citation and self-citing authors by gender. This table shows the statistics on self-citations by gender. It illustrates the proportions of self-citation performed by male and female APS authors as well as the proportions of self-citing male and female authors.

Subfield	Total Papers	Alphabetically Ordered Papers	Alphabetical %
PACS 00	50,719	2,582	5.091
PACS 10	24,142	3,218	13.329
PACS 20	14,510	879	6.058
PACS 30	23,145	707	3.055
PACS 40	28,800	967	3.358
PACS 50	5,610	146	2.602
PACS 60	50,569	1,270	2.511
PACS 70	87,066	1,754	2.015
PACS 80	25,281	628	2.484
PACS 90	9,222	902	9.781

Table 3. Proportion of alphabetically ordered papers by subfield.

Position	$n_m + n_f$	n_m	p_m	n_f	p_f	z	p-value
First	149,627	137,223	0.2986	12,404	0.2834	**6.6346**	<0.00001
Second	87,869	80,073	0.1742	7,796	0.1781	-2.052	0.9799
Middle	115,619	104,827	0.2281	10,792	0.2466	-8.7869	>0.99999
Last	150,182	137,412	0.2990	12,770	0.2918	**3.1535**	0.0008

Table 4. Statistical tests for author order analysis. In this table every pair (publication,author) is a unique data point, so each author appears repeated the number of times he or she has published in a given position. As a result, n_f (resp. n_m) is the number of times a female (resp. male) author appears in a paper in the corresponding position. z-scores and p-values are accordingly calculated (see Methods) and are rounded up to the fourth decimal places with an exception of extreme values. n and p respectively denote sample size and proportion.

	n	n_{male}	p_{male}	n_{female}	p_{female}	z	p-value
Top 10% (511+ citations)	40	39	0.9750	1	0.0250	1.279	0.1004
Top 20% (346+ citations)	90	89	0.9889	1	0.0111	2.405	0.0081
Top 30% (288+ citations)	152	149	0.9803	3	0.0197	2.733	0.0031
Top 40% (243+ citations)	226	220	0.9735	6	0.0265	2.954	0.0016

Table 5. Statistical tests comparing degree centrality by gender in the top ranks. Comparison of the proportion of papers respectively led by male and female primary authors in the top ranks of degree centrality. For reference, the overall proportion of female led papers is 0.08. The high z-scores and low p-values corroborate the gender disparities found in Figure 3A of the main document.

PACS	Subfield	N_{mf}	p^*	$\lvert M(p^*)\rvert$	$\frac{\lvert M(p^*)\rvert}{N_{mf}}$	$d(p^*)$	z	p-value
00	General Physics	184694	0.002	9931	5.38%	-0.398	-1.181	0.238
10	Elementary Particles and Fields	49254	0.003	2833	5.75%	0.758	**2.908**	0.0036
20	Nuclear Physics	7698	0.003	385	5.00%	0.584	1.453	0.146
30	Atomic and Molecular Physics	29246	0.002	1474	5.04%	1.028	**3.058**	0.0022
40	Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, Fluid Dynamics	54621	0.0025	2525	4.62%	0.526	1.889	0.059
50	Gases, Plasmas, Electric Discharges	747	0.006	48	6.43%	-0.021	-0.032	0.974
60	Condensed Matter (CM): Mechanical, Thermal	123631	0.0018	7063	5.71%	0.432	**3.039**	0.0024
70	CM: Electrical, Magnetic, Optical	529069	0.002	28952	5.47%	0.674	**5.623**	<0.00001
80	Interdisciplinary Physics & Related Studies	29173	0.0025	1602	5.49%	-0.408	-0.860	0.390
90	Geophysics, Astronomy, Astrophysics	18760	0.006	1041	5.55%	1.603	**4.266**	0.00002

Table 6. Differences in received citations among similar pairs of publications. Gender differences in received citations among pairs of publications with validated similarity measured by z-scores. The variables of the columns are the following (more details in Methods): N_{mf} - number of all possible male-female pairs; p^* - chosen critical similarity value, the lower, the more similar; $M(p^*)$ - subset of pairs with similarity of p^* or better; $d(p^*)$ - average male-female citation difference; z - normalized difference of male-female average citations. Values of p^* are chosen to establish $\frac{\lvert M(p^*)\rvert}{N_{mf}}$ values between 4% and 7%. Significant z-scores are marked in bold.

	PACS 00		PACS 10		PACS 20		PACS 30		PACS 40		PACS 60		PACS 70		PACS 80		PACS 90	
Year Diff	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2	Q1 vs. Q3	Q4 vs. Q2
1	-1.0276	-0.4593	0.8158	**2.4389**	-0.1137	1.3310	1.8284	1.8513	1.9239	-1.0231	0.4762	-0.6276	-0.9009	-0.4480	-0.2413	0.0963	**2.4604**	1.1396
2	-0.7112	-0.6080	-0.4986	1.1457	0.1802	-0.4811	1.3864	0.3983	0.4605	-0.4848	-0.6398	0.5378	-0.5815	-0.5130	1.2425	-0.4183	0.6071	1.3616
3	0.2387	-1.2630	0.8695	0.1341	1.5063	-0.7702	-0.3992	0.7234	-1.0897	**-2.4929**	1.3901	-0.0954	-0.5551	0.4223	-1.0803	-1.0713	0.6929	-0.8376
4	0.0998	-1.4989	-1.2089	-0.0157	0.6585	0.3445	1.1498	0.3478	**-2.3946**	1.2368	1.2652	0.8284	0.1545	-1.0583	-0.6824	-1.1105	-0.2388	1.1283
5	**2.5848**	-0.2843	0.8910	0.2543	0.6547	0.9623	1.1239	0.0808	**2.0510**	1.4468	1.6584	0.1256	1.4723	-0.2557	0.2562	0.7429	-0.6698	-1.3254
6	0.2919	-1.4444	0.6168	-0.7551	-	0.5976	0.7370	0.5413	-1.2679	-0.9277	0.4748	-0.2807	-1.2351	0.7690	-0.7708	0.8244	-	0.1472
7	-0.8214	-0.2422	1.0674	-0.6662	-	**7.5951**	1.3540	0.7106	1.6427	0.1136	-1.2621	-1.2536	-0.7857	0.9365	-	-1.2055	-0.0735	0.9886
8	-1.9479	-0.2470	**-2.5156**	0.5238	-	-	-	-0.4330	-1.3060	1.8453	-0.2936	-0.6337	-1.4345	-1.1993	**-2.2968**	-0.8736	-	**2.9049**
9	-0.7831	1.1800	1.0545	-1.5591	-	-0.9384	**2.5156**	-0.8575	-	0.2941	1.0550	0.8075	-1.6362	1.7124	-	0.6258	-	**2.0403**
10	0.8334	-0.2938	0.4264	**-1.9832**	-	-0.1902	0.2774	-0.3513	1.3392	-0.6085	-0.4489	-1.3741	0.9067	0.6636	0.8944	-0.9810	-	-

Table 7. Statistical tests of gender asymmetry in the first-mover advantage. Comparison of the citation differences between quadrants $Q1/Q3$ and $Q2/Q4$ of Figure 5 for each year difference. The values shown in this table are z-scores computed according to equation (11) of the main document. Most of them lie in the range $(-2, 2)$ (not significant), indicating that there is no gender asymmetry in the advantage gained by an author by publishing earlier. Data with less than three data points do not yield meaningful statistics and therefore are excluded from our analysis (they are marked as '-'). PACS 50 has a very small sample size and hence is not analyzed.

References

1. Sinatra, R., Wang, D., Deville, P., Song, C. & Barabási, A.-L. Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
2. https://genderize.io/.
3. Zhou, E., Cao, Z. & Yin, Q. Naive-deep face recognition: Touching the limit of lfw benchmark or not? (2015). https://arxiv.org/pdf/1603.04322.pdf.
4. Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 53–54 (International World Wide Web Conferences Steering Committee, 2016).
5. NACO. Authority file comparison rules (naco normalization).
6. Kessler, M. M. Bibliographic coupling between scientific papers. *Am. documentation* **14**, 10–25 (1963).
7. Egghe, L. & Rousseau, R. Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics* **55**, 349–361 (2002).
8. Lee, S. Improving jaccard index for measuring similarity in collaborative filtering. In *International Conference on Information Science and Applications*, 799–806 (Springer, 2017).
9. Saranya, K., Sadasivam, G. S. & Chandralekha, M. Performance comparison of different similarity measures for collaborative filtering technique. *Indian journal science Technol.* **9**, 1–8 (2016).
10. Ciotti, V., Bonaventura, M., Nicosia, V., Panzarasa, P. & Latora, V. Homophily and missing links in citation networks. *EPJ Data Sci.* **5**, 7 (2016).
11. Perianes-Rodriguez, A., Waltman, L. & van Eck, N. J. Constructing bibliometric networks: A comparison between full and fractional counting. *J. Informetrics* **10**, 1178–1195, DOI: https://doi.org/10.1016/j.joi.2016.10.006 (2016).
12. Batagelj, V. On fractional approach to analysis of linked networks. *Scientometrics* **123**, 621–633, DOI: 10.1007/s11192-020-03383-y (2020).
13. Tumminello, M., Micciche, S., Lillo, F., Piilo, J. & Mantegna, R. N. Statistically validated networks in bipartite complex systems. *PloS one* **6**, e17994 (2011).