

Exploring Gender Differences in Bioinformatics Research

Jorge Saldivar¹, Nataly Buslón², María José Rementería³

Social Links Analytics Group, Life Science Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain

¹jorge.saldivar@bsc.es, ²nataly.buslon@bsc.es, ³maria.rementeria@bsc.es

Keywords— gender inequality, bioinformatics, science of science

EXTENDED ABSTRACT

Women’s underrepresentation in science is an unfair situation apparent in different research fields. This article examines the problem of gender bias in the field of Bioinformatics. Through a quantitative study of more than 40K articles, we discovered that even when the difference between women and men representation in academic publications has been slightly reduced in the last years, men still dominate the scientific activity. We also found that publications led by female researchers are less recognized in terms of citations than articles led by male scholars.

A. Introduction

The underrepresentation of women in the scientific community is noticeable across research fields [1]. Women scientists publish less than men and have fewer opportunities than men to participate in international collaborations. In developed countries, articles with women in leading authorship positions receive fewer citations than those with men in the same positions [2]. By leveraging quantitative methods, this article explores aspects of gender differences in bioinformatics research with the aim to unveil unfair situations of gender inequalities as well as study whether already published gender biases in other fields of research are also present in bioinformatics.

B. Methods

Five representative journals in the field of Bioinformatics were chosen for the study, namely Oxford Bioinformatics, Plos Computational Biology, Nucleic Acids Research, BMC Bioinformatics, and BMC Genomics. Meta-data of articles (e.g., title, author names, publication year, abstract, number of citations, PubMed identifier) published in these journals were extracted from Scopus—one of today’s most complete repositories of scientific manuscripts. The data collection process was run in August 2019.

The search engine of Scopus was queried to find articles published between 2005 and 2017 in the selected journals. The lower-bound limit in the range of years was chosen because the journal Plos Computational Biology appeared in 2005. Also, we decided to limit our search to 2017 because we wanted to leave enough time (one full year) to ensure that Scopus already indexed the majority, if not all, of articles of the selected journals by the moment of the search.

In total, meta-data of 47,427 articles and their corresponding authors were collected. After getting rid of duplicates (194) and entries without DOI (401), 46,832 records were considered for further analysis. Table 1 summarizes the data collected per journal.

The gender of the 143,960 authors who produced the 46,832 articles was inferred using the services NamSor and gender-guesser. We hit the API of NamSor asking for the

gender of authors by providing the authors’ first and last names. NamSor API returns male, female, or unknown. In case NamSor fails to identify the gender, the python package gender-guesser was employed as a backup to find out the gender of authors. The gender of 27,705 authors (19%) could not be detected by either any of the two gender identification tools. In 10% of the cases, gender could not be identified because we could not get the author’s name. For the rest, we empirically found that, as also reported in [3], Asian names are challenging for gender identification services. After removing authors with unidentified gender, 116,255 authors left in the dataset.

TABLE I
SUMMARY OF THE DATA COLLECTED PER JOURNAL

Journal	ISSN	No. of articles
Oxford Bioinformatics	1460-2059	8,466
Plos Computational Biology	1553-734X	5,121
Nucleic Acids Research	1362-4962	15,489
BMC Bioinformatics	1471-2105	7,703
BMC Genomics	1471-2164	10,053
Total articles		46,832

We found that 1,382 of the 46,832 articles either do not provide information about their authors or all of their authors have unknown gender. These records were removed from the dataset, remaining 45,450 articles for the analysis.

C. Results

A mean of 4.8 of people authored articles published in the journals of interest, and this average shows a moderate increase from 4.1 in 2005 to 5.1 in 2017. As expected and in line with several previous research [4][5], the majority of the 116,255 authors are male (64.7%), while the rest 35.2% are female.

a. Gender distribution

In 60% of the articles (29,926), male researchers dominate the list of authors. More than one-third of the articles were authored only by men (15,583) while articles published exclusively by women represent 3.91% of the total. Female researchers lead 26.26% of articles, whereas males appear as first authors in almost 60% of articles. Percentages do not sum up to 100 because, due to limitations in the method used to infer authors’ gender, we do not have gender information about all first authors. The underrepresentation of female researchers is even more noticeable when analyzing the last

authors of articles. In this case, the proportion of female researchers that participate as last authors is less than 20%.

Results show that gender differences gradually decrease over time. Figure 1 shows how the representation of female authors within the total authors increases from 2005 to 2017. The proportion of total authors passed from .27 in 2005 to .32 in 2017. The different actions taken by governments, research centers, universities, and civil institutions to promote the participation of women in science might have impacted the increment. However, male researchers are still clearly dominating the group of authors.

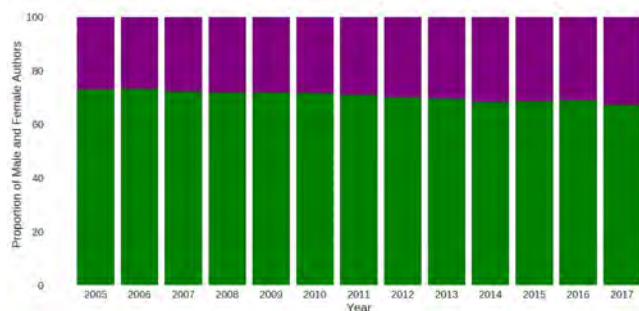


Fig. 1 Evolution of the proportion of male and female authors over time. Women proportion is shown in purple, whereas men proportion is depicted in green.

b. Productivity and citations

In general, the 45,450 articles received a median of 28 citations. The top ten most-cited authors are all males, and they accounted for almost 2% of the total citations (223,053 out of 13,368,610).

The average citation by paper within the group of male authors is 46.2, while in female researchers is 43.12. Articles authored exclusively by women received on average fewer citations (33) than articles in which only men participated (54), and the difference is statistically significant (p -value < .001, α =.05). We found a similar situation when analyzing articles led by a female researcher and articles in which a woman appears as the last author. The difference between the average citations in both cases showed to be statistically significant (p -value < .001, α =.05).

Authors have, in general, a mean of productivity of 1.89 articles. In the case of male researchers, they have productivity of 2.05 articles while females 1.60 articles. As in the case of citations, the top-ten most productive authors are all male. The first woman appears in the position 13th in the productivity ranking (i.e., sorting authors by their number of articles in descending order).

Two females are part of the list of the top ten authors that participated in articles as the last author. As found in previous research [6], articles with females as last authors are less frequently cited (mean=38.50) than articles with males as last authors (mean=50.64), and the difference is statistically significant (p -value < .001, α =0.05). Contrary to what Bendels et al. [6] have found, the difference does not get larger as more people authored articles. Similarly, we discovered that articles led by female researchers got significantly fewer citations on average than articles with male scholars as first authors (p -value < .001, α =.05). Besides, our analyses show that articles authored only by women

received significantly fewer citations on average than articles published only by male authors (p -value < .001, α =.05).

D. Conclusion and Future Enhancement

Although the last years show a reduction in the inequalities between men and women in scientific production, our results confirm that still there is a noticeable gender difference in favor of male researchers. In future work, we will deep dive into the presented results to further explore other dimensions of the gender bias in bioinformatics.

E. ACKNOWLEDGEMENTS

European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement H2020-MSCA-COFUND-2016-754433.

References

- [1] Bonham, Kevin S., and Melanie I. Stefan. "Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology." *PLoS computational biology* 13, no. 10 (2017).
- [2] L. Vincent, C. Ni, Y. Gingras, B. Cronin, and C. R. Sugimoto. "Bibliometrics: Global gender disparities in science." *Nature News* 504, no. 7479 (2013): 211.
- [3] K. Fariba, C. Wagner, F. Lemmerich, M. Jadidi, and M. Strohmaier. "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*, pp. 53-54. 2016.
- [4] H. Luke, D. Stuart-Fox, and C. E. Hauser. "The gender gap in science: How long until women are equally represented?." *PLoS biology* 16, no. 4 (2018): e2004956.
- [5] H. Junming, A. Gates, R. Sinatra, and A. Barabási. "Historical comparison of gender inequality in scientific careers across countries and disciplines." *Proceedings of the National Academy of Sciences* 117, no. 9 (2020): 4609-4616.
- [6] B. Michael, R. Müller, D. Brueggmann, and D. Groneberg. "Gender disparities in high-quality research revealed by Nature Index journals." *PloS one* 13, no. 1 (2018).

Author biography



Jorge Saldivar is a doctor in Information and Communication Technologies by the University of Trento, Italy. He is a postdoctoral researcher in the Social Link Analytics group at the BSC. Jorge has a broad interest across data science, especially in research that brings together quantitative and qualitative methods. Previously, he was a Data Scientist Fellow at the University of Chicago, USA, and worked as a Postdoctoral Research Fellow in the Department of Electronics and Informatics at the Catholic University of Asuncion, Paraguay.