2022 Smart Information Systems and Technologies (SIST)
28-30 April, 2022, Nur-Sultan, Kazakhstan

Impact of Gender on Publication Productivity and Scientific Collaboration

Andrii Biloshchytskyi
Vice-Rector for Science and Innovation
Astana IT University
Kazakhstan, Nur-Sultan,
Kyiv National University of Construction and Architecture
Kyiv, Ukraine
ORCID: 0000-0001-9548-1959

Alexander Kuchansky
Dept. of information systems and techn.
Taras Shevchenko Nat. University of Kyiv
Kyiv National University of Construction and Architecture.
Kyiv, Ukraine
kuczanski@gmail.com

Yurii Andrashko
Dept. of system analisys and optim. theory
Uzhhorod National University
Uzhhorod, Ukraine
yurii.andrashko@uzhnu.edu.ua

Myroslava Gladka
Dept. of information systems and techn.
Taras Shevchenko Nat. University of Kyiv
Kyiv, Ukraine
miragladka@gmail.com

***Abstract* —The article describes the task of researching the impact of gender on scientific productivity and collaboration. The contribution of publications to the scientific productivity of authors with different gender compositions is analyzed. A modification of the PR method for teams of authors from different countries was used. Based on the Citation Network Dataset (version 13) (out of more than 5 million scientific publications and 48 million citations), there was a significant gap in the contribution of male authors' articles from those published by female authors. The advantage of articles with male composition over articles with female composition by more than 1.6 times was revealed. The mixed composition of authors does not allow increasing the contribution of publications to scientific productivity, and their contribution is, on average, 1.1 times less than articles with a male composition. The gender impact on scientific cooperation and the productivity of scientific activity has been revealed. This leads to a significant impact on the career path of women in academia. The results obtained confirm the presence of gender inequality in science. The results are intended to intensify the discussion on ensuring equality of rights and gender equality at the state level.**

Keywords—scientific collaboration, gender, citation impact, publication productivity

I. INTRODUCTION

The development of scientific cooperation is an essential engine for creating innovations, producing new knowledge and ideas. The results of scientific cooperation are reflected in scientific publications, joint projects, and activities. Identifying the influence of demographic, social characteristics, and gender on publication productivity and forms of scientific cooperation is essential for scientometrics. Understanding scientific cooperation as a joint activity of scientists to produce new knowledge determines the influence of social mechanisms on this process.

Papers [1, 2] determined that the form and intensity of scientific cooperation directly impact publication productivity and innovation [3]. The work [4] describes the influence of gender diversity in the work of scientific groups on publication activity and citation. In [5], it was found that gender-heterogeneous working groups allow producing scientific results of the highest quality. However, this is complicated by the natural gender homophily, which determines the inclination to interact with their kind [6]. This property also manifests itself in quoting. It was found in [7] that scientists tend to cite publications by authors of the same gender as themselves. The work [8] noted that the impact of gender diversity on the productivity of scientific work in modern conditions is decreasing, especially among young scientists. The trend towards promoting gender equality in developed societies is mitigating the gap in scientific performance.

The analysis carried out in [8] argues that gender differences in the productivity of scientific activity disappear over time. In the older generation, men outnumbered women in scientific publications and citations, but this is no longer the case in the younger generation. In addition, the dominant trend remains that fewer woman than men hold the highest rank in academia. However, in [9], it is indicated that women with high scientific abilities in the scientific group significantly impact its growing productivity. Including because women have higher emotional intelligence than men, which is described in [10]. In general, the vision of gender diversity policy in universities is described in [11].

An actual direction is the analysis of the continuity of research in inter-gender scientific cooperation, which is described in [12]. Known methods for studying the patterns of scientific cooperation choosing scientists for organizing projects [13, 14] can also be used to study the gender impact on scientific interaction. Also, to assess the performance of scientific activities, management, and competence-based choice of project executors using a gender approach, the methods described in [15 - 19] can be used.

The work [20] describes a thorough study of the impact of gender on scientific performance. It has been found that the increase in women's participation in science over the past 60 years has been accompanied by an increase in gender differences in both performance and influence. Therefore, studying the impact of gender on scientific productivity and cooperation is relevant for the development of innovation and scientific production in general. The identified gender inequality in academic circles should be assessed and eliminated at the state level. The article aims to analyze the impact on the productivity of scientific collaboration of authors with different gender compositions. To assess the contribution of the publication to the productivity of the author's scientific activity, the method described in [15] for groups of authors from different countries was used.

II. Method for evaluating the contribution of a publication to the productivity of the author's scientific activity

Let $A=\{a_1,a_2,\ldots,a_n\}$ is the set of scientists, n is the number of scientists. $P=\{p_1,p_2,\ldots,p_m\}$ is the set of publications that these scientists have published, m is the number of publications. To assess the contribution of each of the publications to the productivity of the author's scientific activity, a modification of the PR method was used. According to this method, the scalar assessment of the contribution of the publication p_i, $i=\overline{1,m}$ to the productivity of scientific activity is calculated by the formula [15]:

$$q_i=\sum_{j=1}^{n}\beta_{ij}\xi_j q_j,\ i=\overline{1,n}, \qquad (1)$$

where q_i is the assessment of the contribution of the publication p_i to the productivity of scientific activity, β_{ij} is the coefficient that determines the presence of the publication p_i in the citation list of the publication p_j, ξ_j is the coefficient that ensures the existence of a non-trivial solution to the system of linear algebraic equations (1), q_j is the assessment of the contribution of the publication p_j to the productivity of the author's scientific activity. As a result of applying formula (1), a homogeneous system of linear algebraic equations is constructed:

$$Bq=0, \qquad (2)$$

where B is the matrix of coefficients of this system of the form:

$$B=E-\left\{\beta_{ij}\xi_j\right\}_{i,j=1}^{n}, \qquad (3)$$

E is the identity matrix, q is the column vector of unknown estimates:

$$w=\{q_i\}_{i=1}^{n},$$
$$q=w^T, \qquad (4)$$

In order for a nontrivial solution to the system of algebraic equations (1) to exist, the matrix B must be degenerate, i.e. $|B|=0$. The coefficients of system (1) can be determined by the formulas:

$$\beta_{ij}=\begin{cases}1, if\ p_i\in C(p_j)\\ 0, if\ p_i\notin C(p_j)\end{cases}, \qquad (5)$$

$$\xi_j=\left\|C(p_j)\right\|^{-1},\ i=\overline{1,n},\ j=\overline{1,n}, \qquad (6)$$

where β_{ij} is the indicator of the presence of the publication p_i in the list of citations of the publication p_j, ξ_j is the reciprocal of the total number of citations in the publication p_j.

After finding the estimates, it is advisable to normalize them according to the formula:

$$q'(p_i)=q_i\left(\sum_{j=1}^{n}q_j\right)^{-1},\ i=\overline{1,n},$$

where q_i is the assessment of the contribution of the publication p_i to the productivity of the author's scientific activity, $q'(p_i)$ is the assessment normalized by the sum.

III. Experimental part

The study was verified using the Citation Network Dataset version 13 [21]. This set contains 5354309 publications and 48227950 citations obtained from the DBLP [22], ACM, Microsoft Academic Graph, and other databases. Version 13 contains current data as of May 2021.

At the first stage of the study, the impact of each publication in the citation network was calculated. For this, a modification of the PR method was used to calculate the assessment of the contribution of publications to the productivity of scientific activity. To find the contribution of scientific publications to the productivity of scientific activity, it is necessary to solve a system of linear algebraic equations of high dimension (2). To find an approximate solution to the system of linear algebraic equations (2), an iterative process of the Gauss-Seidel method was applied. At the zero-step, the value of the assessment of the contribution to the productivity of scientific activity of all publications is equal to 1. At the k-th step, the value of each assessment

of the contribution of a publication to the productivity of scientific activity is found by the formula:

$$q_i^k = \sum_{j=1}^{n} \beta_{ij} \xi_j q_j^{k-1}, \; i = \overline{1,m}, \qquad (7)$$

where q_i^k is the approximate value of the assessment of the contribution of the publication p_i to the productivity of scientific activity at the *k*-th step, q_j^{k-1} is the approximate value of the assessment of the contribution of the publication p_j to the productivity of scientific activity at the *(k-1)*-th step, and the coefficients are calculated by formulas (5), (6).

After each step, starting from zero, the maximum absolute change in estimates of the contribution of the publication to the productivity of scientific activity is calculated using the formula:

$$\Delta^k = \max_{i=1,m} \left| q_i^k - q_i^{k-1} \right|, \qquad (8)$$

where Δ^k is the maximum absolute change in estimates of the contribution of the publication to the productivity of scientific activity.

Absolute maximum change in the assessment of the publication's contribution to the productivity of scientific activity is considered an upper estimate of the absolute error of the method. After six iterations of calculating the contribution of publications, the absolute error was. The authors consider this performance estimation accuracy sufficient, so the calculation process was completed.

At the second stage of the study, among all publications from the data set, data on publications from the countries the research hypothesis tested, namely Germany, the Netherlands, the USA, Ukraine, and Kazakhstan, were filtered. A publication belongs to a subset of publications from a particular country if at least one of the authors is affiliated with an institution of higher education that belongs to that country.

At the third stage, the author's gender was determined. To determine the author's gender, male and female names were compiled. Wikipedia data was used to compile the lists. If the author's name is not in the list, or only initials are given, it is considered that the author's gender identity could not be established. As a result of research, for 53.1% of publications, the gender identity of all authors was established. For 46.9% of publications, it was impossible to establish the gender of at least one of the authors (Table 1).

At the fourth stage of the study, from each subset of publications of the indicated countries, a subset of publications with two or more authors whose gender identity was established was identified. Further, the subset of publications with the known gender of the authors is divided into three subsets. The first subset of FF includes publications, all of whose authors are women. The second subset of MM includes publications, all of whose authors are men. The third subset of FM includes articles with mixed authors, that is, there are both women and men among the authors.

Among articles with identified gender of all authors, Germany and the Netherlands have the most significant number of articles with male authors (more than 50%). The USA, Ukraine, and Kazakhstan have more articles with a mixed-gender composition of authors. The sample of articles for Kazakhstan and Ukraine is not representative, but the general trend of the gender of the authors is visible.

TABLE I.
GENDER DISTRIBUTION OF AUTHORS OF SCIENTIFIC PUBLICATIONS

#	*Country*	*m*	*FF*	*MM*	*FM*
1	Germany	189896	858	65796	30721
2	The Netherlands	56624	363	21951	18393
3	USA	467522	3585	94858	143318
4	Ukraine	2561	55	95	851
5	Kazakhstan	952	14	6	20

For each subset, the average contribution of publications to the productivity of scientific activity was calculated using the modified PR method [15]. The results are shown in table 2 and fig. 1.

TABLE II.
CONTRIBUTION OF PUBLICATIONS WITH DIFFERENT GENDER COMPOSITION TO AUTHORS' SCIENTIFIC PRODUCTIVITY

#	*Country*	*FF*	*MM*	*FM*
1	Germany	1,4352	2,4796	2,1182
2	The Netherlands	1,7837	3,0571	2,7067
3	USA	2,3963	3,8665	3,5875
4	Ukraine	0,4745	0,7716	0,6785
5	Kazakhstan	0,0550	0,2500	0,0857

For comparison, normalization was carried out according to the maximum contribution of publications to the productivity of scientific activity of authors from the above countries. The results after normalization are presented in Table 3.

The results of assessing the contribution of publications to the scientific productivity of authors from these countries indicate the advantage of the contribution to the productivity of publications with male composition over publications with female composition.

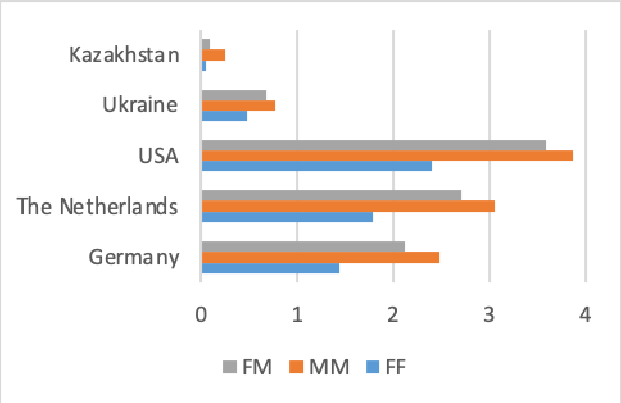

Figure 1. The contribution of publications with different gender composition to the productivity of scientific activity of authors

TABLE III.
THE ABSOLUTE CONTRIBUTION OF PUBLICATIONS WITH

DIFFERENT GENDER COMPOSITION TO AUTHORS' SCIENTIFIC PRODUCTIVITY

#	*Country*	*FF*	*MM*	*FM*
1	Germany	0,5788	1	0,8543
2	The Netherlands	0,5835	1	0,8854
3	USA	0,6197	1	0,9278
4	Ukraine	0,6150	1	0,8793
5	Kazakhstan	0,2200	1	0,3428

The contribution to the productivity of scientific activity of publications from the MM subset is 1.6–1.72 times greater than the contribution of publications from the FM subset for all the countries represented, except for Kazakhstan, for which this number is 4.5. The contribution of FM subset publications compared to MM subset publications is lower by 1.07–1.17 for all countries except for Kazakhstan, for which this number is 2.9. This indicates the presence of significant gender inequality in academic circles, which is paradoxical at the present stage of the development of society.

IV. CONCLUSION

The paper analyzes the impact of publications on the productivity of scientific collaboration of authors with different gender compositions. A modified PR method was used for teams of authors from different countries to assess the performance of the scientific activity. The study was verified using the Citation Network Dataset (ver. 13) with more than 5 million scientific publications and 48 million citations. For 53.1% of publications, the gender identity of all authors was established. The assessment results of the contribution of scientific publications indicate the advantage of the contribution of articles with male composition over articles with female composition (more than 1.6 times) or mixed (more than 1.07 times). This indicates that articles with male contributors are better cited than articles with mixed or female contributors. Among articles with identified gender of all authors, Germany and The Netherlands have the most significant number of male authors (more than 50%). The results obtained confirm the presence of gender inequality in science, which is essential for intensifying the discussion in the direction of improving the situation.

V. ACKNOWLEDGMENT

The article was written within the framework of the state order for the implementation of a scientific program under the budget program 217 «Development of Science», IRN №AP08857218 on the topic: «Information technology for assessment of scientific activity of universities, research institutes and their subdivisions».

REFERENCES

[1] M. Franceschet, and A. Costantini, “The effect of scholar collaboration on impact and quality of academic papers”, *Journal of Informetrics*, 4(4), pp. 540–553, 2010.

[2] R. Cartes-Velasquez, and C. Manterola, “Impact of collaboration on research quality: A case analysis of dental research. International”, *Journal of Information Science and Management*, 15(1), 89–93, 2017.

[3] K. Talke, S. Salomo, and A. Kock, “Top management team diversity and strategic innovation orientation: The relationship and consequences for innovativeness and performance”, *Journal of Product Innovation Management*, 28(6), pp. 819–832, 2011.

[4] A. Maddi, and Y. Gingras, “Gender diversity in research teams and citation impact in economics and management”, *Journal of Economic Surveys,* 2021. 10.1111/joes.12420

[5] L. G. Campbell, S. Mehtani, M. E. Dozier, and J. Rinehart, “Gender-heterogeneous working groups produce higher quality science”, *PloS one*, 8(10), 2013. e79147.

[6] L. Holman, and C. Morandin, “Researchers collaborate with same-gendered colleagues more often than expected across the life sciences”, *PloS one*, 14(4), 2019. e0216128.

[7] G. Ghiasi, P. Mongeon, C. Sugimoto, and V. Larivière, “Gender homophily in citations”, 3rd International Conference on Science and Technology Indicators, STI 2018, pp. 1519–1525, 2018.

[8] P. van Arensberger, I. van der Weijden, and P. van den Besselaar, “Gender differences in scientific productivity: a persisting phenomenon?”, *Scientometrics*, 93, pp. 857–868, 2012.

[9] M. Zhang, G. Zhang, Y. Liu, X. Zhai, X. Han, “Scientists’ genders and international academic collaboration: An empirical study of Chinese universities and research institutes”, *Journal of Informetrics*, 14(4), 2020. 101068.

[10] D. L. Joseph, D. A. Newman, “Emotional intelligence: An integrative meta-analysis and cascading model”, *Journal of Applied Psychology*, 95(1), 54–78. 2010.

[11] T. M. Timmers, T. M. Willemsen, and K. G. Tijdens, “Gender diversity policies in universities: A multi-perspective framework of policy measures”, *Higher Education*, 59(6), 719–735. 2010.

[12] H. Shen, J. Xie, and Y. Cheng, “The continuity and citation impact of scientific collaboration with different gender composition”, *Journal of Informetrics*, 16 (1), 2022. 101248. 10.1016/j.joi.2021.101248

[13] H. Xu, A. Kuchansky, and M. Gladka, “Devising an individually oriented method for selection of scientific activity subjects for implementing scientific projects based on scientometric analysis”, *Eastern-European Journal of Enterprise Technologies*, 6(3 (114)), pp. 93–100, 2021. DOI: 10.15587/1729-4061.2021.248040

[14] F. Fu, C. Hauert, M. A. Nowak, and L. Wang, L, “Reputation-based partner choice promotes cooperation in social networks”, *Physical Revie0w* E 78, 2008. 026117. DOI: 10.1103/PhysRevE.78.026117

[15] A. Biloshchytskyi, A. Kuchansky, Y. Andrashko, S. Biloshchytska, O. Kuzka, and O. Terentyev, “Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations”, *Eastern-European Journal of Enterprise Technologies*, 3(2(87), pp. 4–10. 2017. DOI: 10.15587/1729-4061.2017.103651

[16] A. Biloshchytskyi, S. Biloshchytska, A. Kuchansky, O. Bielova, Y. Andrashko, “Infocommunication system of scientific activity management on the basis of project-vector methodology”, *14th Intern. Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering*, pp. 200-203, 2018. DOI: 10.1109/TCSET.2018.8336186

[17] A. Biloshchytskyi, A. Kuchansky, S. Paliy, S. Biloshchytska, S. Bronin, Y. Andrashko, Y. Shabala, and V. Vatskel, “Development of technical component of the methodology for projectvector management of educational environments”, *Eastern-European Journal of Enterprise Technologies*, 2 (2-92), pp. 4-13. 2018. DOI: 10.15587/1729-4061.2018.126301

[18] O. Mulesa, F. Geche, A. Batyuk, and I. Myronyuk, “Using a systematic approach in the process of the assessment problem analysis of the staff capacity within the health care institution”, *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, 1, pp. 177–180, 2018.

[19] D. Bushuyev, V. Bushuieva, B. Kozyr, and A. Ugay, “Erosion of competencies of innovative digitalization projects”, *Scientific Journal of Astana IT University*, (1), pp. 70-83, 2020.

[20] J. Huang, A. J. Gates, R. Sinatra, and A.-L. Barabasi, “тьл”, PNAS, 117 (9), pp. 4609-4616 2020. DOI: 10.1073/pnas.1914221117

[21] Citation Network Dataset: DBLP+Citation, ACM Citation network. URL: https://www.aminer.org/citation

[22] DBLP. Computer science bibliography. URL: https://dblp.org/