https://doi.org/10.51408/issi2025_122

Distribution Differences of Knowledge Diversity among Authors in Different Contributor Roles—Evidence from 101014 PLOS ONE Articles

Jingyuan Li [1], Yi Zhao [2], Jiaqi Zeng [3], Chengzhi Zhang [4]

[1] *lijingyuan2002@njust.edu.cn*, [2] *yizhao93@njust.edu.cn*, [3]*zengjq@njust.edu.cn*, [4]*zhangcz@njust.edu.cn*
Department of Information Management, Nanjing University of Science and Technology, 210094 Nanjing (China)

Abstract

In the fast-developing academic environment, the composition and structure of research teams are becoming more diversified and complex. Authors with different roles in the team also show obvious differences in knowledge diversity. Understanding of these differences not only helps to dissect the laws of academic development, but also effectively promotes individual career development and teamwork. Therefore, based on 101,014 papers published in PLOS ONE (2017–2023), author knowledge diversity is calculated using pre-publication academic outputs from the OpenAlex dataset. Additionally, we explore the distribution patterns of knowledge diversity among authors in different research roles. The results of the study show that organizational roles such as Funding Acquisition are more likely to be undertaken by academics with a high degree of knowledge diversity. Technical roles such as Data Curation and Investigation can be finished by authors with relatively lower knowledge diversity. In addition, the study reveals gender differences in knowledge diversity and role taking. Male authors focus on overall design role and female authors are more involved in experiment. This study not only provides a strong empirical basis for the promotion of interdisciplinary collaboration and the development of innovation ability, but also provides a new theoretical perspective for a deeper understanding of the career development of researchers.

Introduction

With the rapid development of current scientific research, single-discipline knowledge become inadequate to solve the complex scientific challenges (Guimerà et al., 2005). Multidisciplinary knowledge reserve has become essential, offering foundational bases and novel perspectives for scientific research. Authors' knowledge diversity, or their interdisciplinary knowledge reserves, significantly impacts their ability to deal with complex problems. Consequently, knowledge diversity has emerged as a key metric for evaluating authors' learning and innovation capabilities.

Knowledge diversity measures the engagement breadth of authors across different disciplines (Chang, 2012). However, current research primarily focuses on team-level knowledge diversity, which fails to accurately capture individual diversity. Existing metrics, such as Rao-Stirling index (Stirling, 2007), explore the link between team interdisciplinarity and research impact. Yet, these indicators emphasize differences among team members rather than individual knowledge diversity across disciplines. Moreover, they rely on post-publication data analysis, resulting in a lag in information acquisition (Zheng et al., 2022).

The roles authors assume in research teams reflect their specific contributions to a project. With the standardization of author contribution statements, the investment of authors in knowledge, skills and labour can be more precisely quantified, providing new ways of thinking about analyzing their actual role contributions (Clement, 2014). Authors' knowledge diversity is closely tied to their roles in research programs, as individuals with varying levels of diversity tend to take on different roles and make distinct contributions (Yang et al., 2022). Consequently, knowledge diversity across roles may exhibit significant differences.

In summary, standardized author contribution statements enable the study of roles and labor division within research teams. While progress has been made in analyzing team-level knowledge diversity, research on individual author knowledge diversity and its distribution across roles remains limited. This gap hinders a deeper understanding of team knowledge structures and the enhancement of research efficiency and innovation. To address this, our study calculates author knowledge diversity using data from PLOS ONE journals (2017–2023) and the OpenAlex platform. Besides, we explore how knowledge diversity is distributed among authors in different roles within research teams.

Related Work

Knowledge Diversity in Research Teams

Research teams are usually organised in terms of outputs, and all authors of a paper are considered as a whole (Zhang & Guo, 2019). Team knowledge diversity can be divided into team shared knowledge diversity and individual author knowledge diversity in the team. The former focuses on the overall knowledge composition of the team, while the latter focuses on the degree of cross-disciplinary of individual members (Chang, 2012). The current research mainly focuses on the knowledge diversity at the team level, while less attention is paid to the knowledge diversity of individual authors in the team. For example, Chowdhary et al. (2024) found that knowledge diversity in enduring collaborative teams has a positive influence on productivity but a negative influence on its impact. Zheng et al. (2022) showed that teams with high expertise diversity do not have a significant effect on their impact in the short term but attract more interdisciplinary citations in the long term. Zhang and Guo (2019) argued that knowledge diversity has a double-edged effect in cross-functional teams. Knowledge leaders can modulate its impact on team performance through the interactive memory system.

Role and Contribution of authors

Scientific collaborations increasingly favour multi-authorship, with a declining proportion of sole-authored papers (Wuchty et al.,2007). Contributions usually refer to the division of labour among co-authors (Rahman et al., 2020), while roles reflect the specific contributions of authors in scientific research. Therefore, clarifying roles and contributions is crucial for improving research efficiency and quality (Yang et al., 2022). Earlier studies measured contribution based on the order of attribution, with the value of contribution decreasing with the order of attribution (Das & Das, 2020). However, studies have found that in some fields, the first and last authors contribute more and the middle authors contribute less (Sundling, 2023). In order to clarify the contribution of authors, many journals use classification systems (Larivière et al., 2020). Among them, some of the journals under the PLOS initially used a five-role taxonomy before fully introducing the Contributor Role Taxonomy (CRediT) in 2016(McNutt et al., 2018). Based on this, Li et al. (2023) developed mapping schemes to analyze the differences in the distribution of author contributions under different systems. Macaluso et al. (2016) found that females were more inclined to experimental work, while males were more likely to take on other roles. These studies highlight the complexity of role division in research teams and offer new insights into understanding author contributions.

Data and Methodology

Data collection and pre-processing

PLOS ONE [1] is an international, peer-reviewed journal that publishes multidisciplinary and interdisciplinary research. Since 2016, it has adopted the CRediT system, a 14-category role framework. This study uses OpenAlex[2] dataset, a global knowledge graph, which provides real-time, multi-dimensional academic data through algorithms and data mining. Additionally, Genderize.io, a widely recognized gender identification tool based on author names, is used to determine gender accurately. This study selects data from PLOS ONE papers from 2017 to 2023, which is conducted with OpenAlex dataset and the Genderize.io[3] tool.

Data collection and processing involve four steps. Firstly, we collected and preprocess metadata and contribution statements from PLOS ONE. Secondly, we extracted author contributions using two formats: line-break-separated text (requiring rule-based abbreviation matching) and JSON text (directly parsed). Both methods ensure accurate mapping of authors to their contributions. Thirdly, author publication counts and concept scores were retrieved to calculate knowledge diversity by utilizing DOIs to connect to OpenAlex. Finally, Genderize.io is utilized to determine author gender, while unidentifiable data is excluded. The final dataset comprises 101,014 articles and 405,766 authors from PLOS ONE journals.

After pre-processing the data, knowledge diversity trends are analyzed, and gender differences are compared. Additionally, role participation rates and gender disparities are examined using contribution statements. Finally, the percentage of authors in each role type within specific diversity intervals is analysed.

Measurement of the authors' knowledge diversity

Knowledge diversity measures the interdisciplinary scope of authors. A lower value indicates a more focused field, while a higher value reflects broader disciplinary involvement and balanced expertise.

[1] https://journals.plos.org/plosone/
[2] https://openalex.org/
[3] https://genderize.io/

OpenAlex defines 19 core top-disciplines. It predicts the topics to which papers belonged from information such as their titles and abstracts, assigning concept score (0-1) of the 19 disciplines for each paper (Priem et al., 2022). For this study, author annual knowledge diversity is calaulated by a 19-dimensional vector. And each dimension reflects the average concept scores of their pre-year papers in each discipline. After normalizing, knowledge diversity is quantified using Equation 1:

$$KD_j = -\sum_{i=1}^{19} p(x_{ij})\, log_2\, p(x_{ij}) \tag{1}$$

Where KD_j denotes the author knowledge diversity in year j; *and* $p(x_{ij})$ denotes the normalised value of the concept score in subject i in year j. To ensure comparable results, the final normalisation was done again using $log_2(19)$. A value of 0 indicates single-topic focus, while 1 represents a balanced knowledge structure across all disciplines.

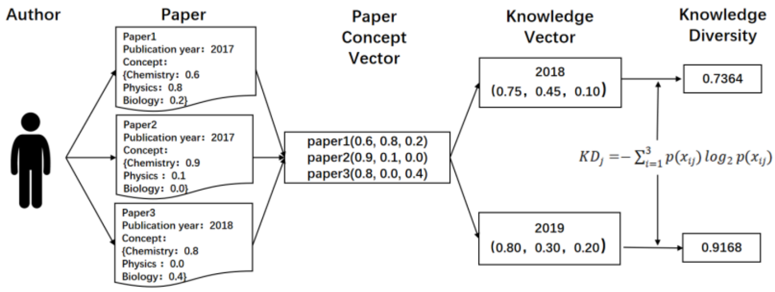

Figure 1. Example of knowledge diversity calculation.

For example, an author published three papers. The publication year and concept scores provided by OpenAlex for Chemistry, Physics and Biology are presented in Fig.1. Firstly, we construct paper concept vectors for paper1, paper2 and paper3.The vector values represent the concept scores of the three disciplines given by OpenAlex. Secondly, we calculate the annual knowledge vectors of the author. The values of each dimension of the vector represent the average concept scores of all papers published by the author before that year in each discipline. For example, the score in Chemistry in 2018 is the average of the concept scores in Chemistry of paper1 and paper2, i.e., (0.6+0.9)/2=0.75, and the same for other disciplines, which ultimately leads to the knowledge vector of the author in 2018 as (0.75, 0.45, 0.10). After normalisation, we calculated its knowledge diversity in 2018 as 0.7364 using Equation 1. Similarly, the knowledge vector and knowledge diversity in 2019 can be calculated using the concept scores of paper1, paper2 and paper3(Fig.1).

Result

Trends in knowledge diversity and gender differences

As shown in Figure 2, the average annual knowledge diversity of authors remains stable, ranging between 0.215 and 0.260. From 2017 to 2022, knowledge diversity shows minimal fluctuation but rises significantly from 2022 to 2023, peaking in 2023. This increase may be driven by the growing use of tools like large models, which have broadened research horizons and enhanced interdisciplinary knowledge integration. For example, large models in medicine have boosted transfer learning, interdisciplinary collaboration, and educational training. It allows authors to integrate multi-disciplinary expertise (Karabacak & Margetis, 2023). From a gender perspective, men's knowledge diversity is significantly higher than women's in every year, although the gap narrows between 2021 and 2023. This may be influenced by the fact that female academics are, on average, younger than their male counterparts (McChesney & Bichsel, 2020).

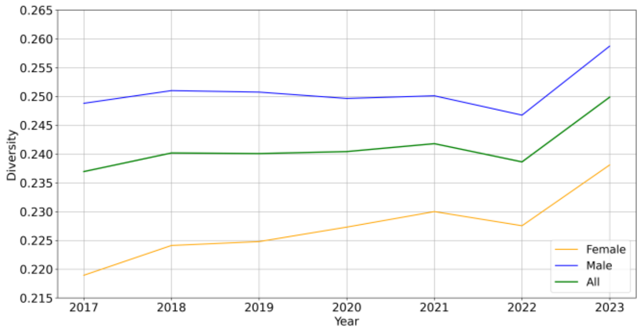

Figure 2. Average annual distribution of knowledge diversity.

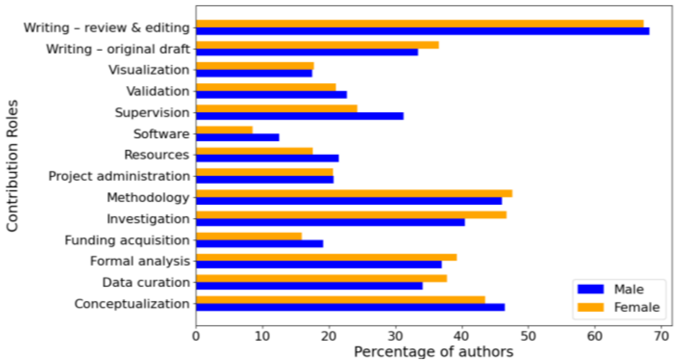

Figure 3. Gender share of roles by contribution type.

Frequency of authors' participation in roles and gender differences

We examine gender differences in research roles by calculating the participation rates of male and female authors in each role (Fig. 3). The results reveal significant variations in role participation frequencies. Writing - review & editing is the most common role, with over 60% participation for both genders, while Software had the lowest, at less than 15%. In terms of gender differences, male are more often involved in conceptual tasks such as Funding acquisition and Supervision. In contrast, female are more often involved in experiment roles such as Investigation, Data curation (Larivière et al., 2020).

Differences in the distribution of knowledge diversity among authors in different roles

The overall distribution of knowledge diversity ranges from 0.0 to 0.7 and its main part is concentrated in the interval of 0.15 to 0.4. Figure 4 depicts the distribution of various roles on the knowledge diversity dimension. It can be observed that most of the roles exhibit a high frequency distribution in the interval of medium knowledge diversity (0.2-0.4), while the frequency in the interval of high knowledge diversity (0.6-0.7) is extremely low, with a frequency close to 0. Particularly noteworthy are the frequency peaks in the intersection of certain roles with knowledge diversity zones, which are significantly higher than in other zones. Funding Acquisition, for example, has a higher distribution of knowledge diversity in the medium-high range than the other roles. It suggests that this role is more likely to be taken on by members with a broader knowledge background. And it is generally performed by leaders with deeper and broader knowledge (Chinchilla-Rodríguez et al., 2019). In other practice-specific roles like Data Curation and Investigation, authors with relatively low knowledge diversity can still perform the work. This suggests these tasks rely less on broad knowledge and more on deep expertise in a specialized area.

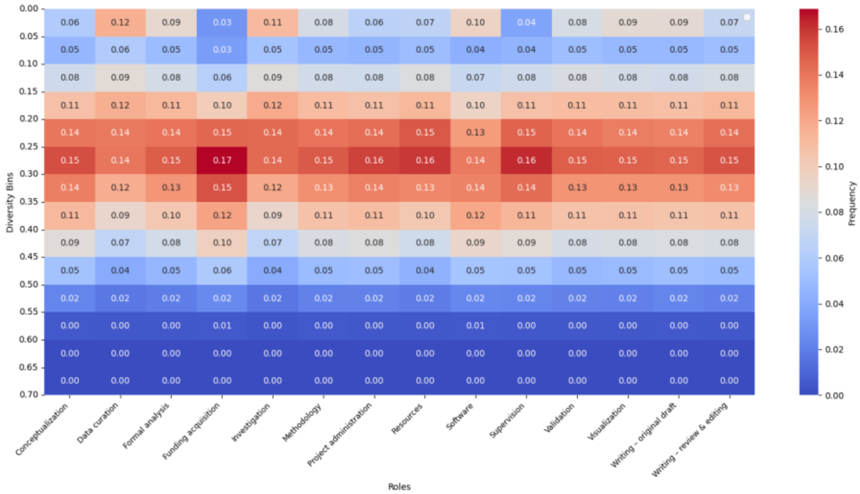

Figure 4. The frequency of distribution of knowledge diversity across contributing roles.

Conclusion

This study analyzes data from PLOS ONE journals (2017–2023), revealing gender differences in knowledge diversity and role distribution. Male authors tend to engage more in conceptual roles, while female authors are more involved in experiment-related roles. Over time, the gap in knowledge diversity between genders narrowed. Additionally, roles like Funding Acquisition require higher knowledge diversity, whereas technical roles (e.g., Data Curation, Investigation) need lower requirements. Despite these insights, the study has limitations. The data, limited to PLOS ONE journals (2017–2023), may lack generalizability despite the journal's interdisciplinary scope. Future research could expand to other journals and extend the time frame to validate findings. Advanced statistical methods, like causal and correlation analysis, can better examine the link between knowledge diversity and role contributions. This provides refined insights for optimizing research team structures and improving scientific efficiency.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (Grant No.72074113).

References

Chang, W. J. (2012). Differential effects of knowledge diversity on team innovation: An agent-based modeling. 2012 International Conference on Innovation Management and Technology Research. https://doi.org/10.1109/ICIMTR.2012.6236384

Chinchilla-Rodríguez, Z., Sugimoto, C. R., & Larivière, V. (2019). Follow the leader: On the relationship between leadership and scholarly impact in international collaborations. PLOS ONE, 14 (6), e0218309. https://doi.org/10.1371/journal.pone.0218309

Chowdhary, S., Gallo, L., & Musciotto, F. (2024). Team careers in science: Formation, composition and success of persistent collaborations. arXiv preprint. https://doi.org/10.48550/arXiv.2407.09326

Clement, T. P. (2014). Authorship matrix: A rational approach to quantify individual contributions and responsibilities in multi-author scientific articles. Science and Engineering Ethics, 20 (2), 345-361. https://doi.org/10.1007/s11948-013-9454-3

Das, N., & Das, S. (2020). 'Author contribution details' and not 'authorship sequence' as a merit to determine credit: A need to relook at the current Indian practice. The National Medical Journal of India, 33 (1), 24-27. https://doi.org/10.4103/0970-258X.308238

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. Science, 308 (5722), 697-702. https://doi.org/10.1126/science.1106340

Karabacak, M., & Margetis, K. (2023). Embracing large language models for medical applications: Opportunities and challenges. Cureus, 15 (3), e39305. https://doi.org/10.7759/cureus.39305

Larivière,V., Pontille, D. , & Sugimoto, C. R. . (2020). Investigating the division of scientific labor using the contributor roles taxonomy (credit). Quantitative Science Studies, 2(9070), 1-24. https://doi.org/10.1162/qss_a_00097

Li, K., Zhang, C., & Larivière, V. (2023). Are research contributions assigned differently under the two contributorship classification systems in PLOS ONE? arXiv preprint. https://doi.org/10.48550/arXiv.2310.11687

Macaluso, B., Larivière, V., & Sugimoto, T. (2016). Is science built on the shoulders of women? A study of gender differences in contributorship. Academic Medicine, 91 (8), 1136-1142. https://doi.org/10.1097/ACM.0000000000001261

McChesney, J., & Bichsel, J. (2020). The aging of tenure-track faculty in higher education: Implications for succession and diversity. Knoxville, TN: College and University Professional Association for Human Resources. https://doi.org/10.13140/RG.2.2.18555.95521

McNutt, M., Bradford, M., & Drazen, J. (2018). Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. OSF Preprints. https://doi.org/10.31219/osf.io/asywp
Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint. https://doi.org/10.48550/arXiv.2205.01833
Rahman, M. T., Regenstein, J. M., & Kassim, N. L. A. (2020). Contribution based author categorization to calculate author performance index. Accountability in Research, 27 (1), 1-20. https://doi.org/10.1080/08989621.2020.1860764
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. Journal of The Royal Society Interface, 4 (15), 707-719. https://doi.org/10.1098/rsif.2007.0213
Sundling, P. (2023). Author contributions and allocation of authorship credit: Testing the validity of different counting methods in the field of chemical biology. Scientometrics, 128 (3), 2737-2762. https://doi.org/10.1007/s11192-023-04680-y
Wuchty, Stefan, Jones, Benjamin, F., & Uzzi, et al. (2007). The increasing dominance of teams in production of knowledge. Science. https://doi.org/10.1126/science.1136099
Yang, S., Xiao, A., & Nie, Y. (2022). Measuring coauthors' credit in medicine field - Based on author contribution statement and citation context analysis. Information Processing & Management, 59 (3), 102924. https://doi.org/10.1016/j.ipm.2022.102924
Zhang, L., & Guo, H. (2019). Enabling knowledge diversity to benefit cross-functional project teams: Joint roles of knowledge leadership and transactive memory system. Information and Management, 56 (8), 103156. https://doi.org/10.1016/j.im.2019.03.001
Zheng, H., Li, W., & Wang, D. (2022). Expertise diversity of teams predicts originality and long-term impact in science and technology. arXiv preprint. https://doi.org/10.48550/arXiv.2210.04422