Understanding Partnership in Scientific Collaborations: A Preliminary Study from the Paper-level Perspective

Chao Lu[1,*], Mengting Li[1], and Chenyu Zhou[1]

[1] *Hohai University, 8 West Focheng Road, Nanjing, China, 211000*

Abstract

Scientific collaboration is more and more common in scientific knowledge production. It has been widely investigated through quantitative and qualitative ways recently. However, most quantitative methods purely based on co-author information usually fail to dig deeper into the internal interaction between collaborators as contributors, which fails to observe internal interactions between collaborators. In this study, we investigated how collaborators in teams work together to perform their research by understanding how two collaborators work together as partners which the traditional collaborative network usually overlooked naturally. By collecting author information from Scopus and author contribution statements from PLoS, we take the biology subject as an example and have examined more than 120,000 research articles and found that division of labor is quite common in scientific collaboration; that partnership as a form of division of labor is widely observed in our dataset; and that the diversity in contributing tasks between partners is generally mild. This study will shed light on understanding the mechanism in scientific collaboration via division of labor that co-authorship studies widely overlook. It helps us create research teams with higher levels of engagement and communication.

Keywords

Scientific Collaboration, Author Contribution Statement, Natural Language Processing.

1 Introduction

Scientific collaboration is more and more common in scientific knowledge production. It has been widely investigated through quantitative and qualitative ways [1, 2] recently. However, there is still more to be investigated especially when more data are disclosed on interactions between collaborators in each team, i.e., author contribution statement [1, 3–5] while most quantitative methods based purely on co-author information usually fail to dig deeper on the internal interaction between collaborators as contributors[6, 7]. Contributorship other than authorship especially pay attention to the actual contributions made by each scientific collaborator. Studies suggest that contributorship provides us with new perspectives to understand scientific collaborations, such as division of labor and team role differentiation[3, 8]. Recently, co-contributorship[1] as a type of partnership in scientific collaborations drew our research interest. Given that research teams consist of not only individual building blocks but living collaborators, we want to investigate how this partnership exists in scientific collaboration and how this close relationship in scientific collaboration influences scientific performance in the future.

Thus, in this preliminary study, we collected author information from Scopus and author contribution statements from PLoS, we took the biology subject as an example and examined more than 120,000 research articles to examine partnership in scientific collaboration from three perspectives: ratio, strength, and diversity. This study and the study to come will

Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents and the 4th AI + Informetrics (EEKE-AII2024), April 23~24, 2024, Changchun, China and Online

**Corresponding author.EMAIL: luchao91@hhu..edu.cn*

help shed light on understanding the mechanism in scientific collaboration via division of labor that studies via co-authorship widely overlooked. It might help us create research teams with higher levels of engagement and communication.

2 Data and Methods

2.1 Data

To examine the phenomenon of partnership in scientific collaborations, we collected 126,894 articles in the Biology domain from PLoS (Public Library of Science) from 2006 to 2020 with their author contribution statements. The yearly distribution of the articles we collected is shown in Fig.1. The plot suggests that the distribution generally followed an increasing trend except in 2015 and 2016, we double-checked the data and found that the PLoS journals did not label subject information for their papers in these two years, so we failed to include the biology papers in these years from the whole collections.

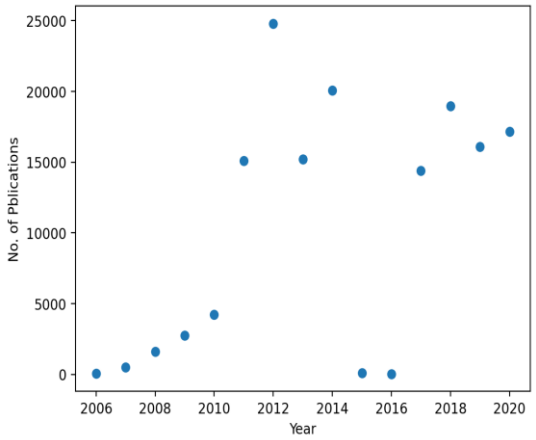

Fig. 1. Yearly distribution of biology publications from PLoS journals collected in this study

2.2 Methods

Following previous studies[1, 9], we process the author contribution statements and link the author names to their tasks in each paper using Python scripts. Using Scopus API, we can disambiguate author names for this study. In total, we have collected 574,979 pieces of disambiguated author information and 2,831,375 author-task pairs.

Given that PLoS did not adopt the CRediT1 taxonomy until 2016, we manually labeled around 99.5% of author-defined tasks according to the taxonomy with detailed definition for each role. For each piece of author-defined tasks, we might assign more than one standard contribution role(see Table 1). For those contributions that cannot be standardized using the taxonomy, we label them as "Other". The rest, around 0.5 percent, of author-defined tasks, we automatically label them as "UNKNOWN" for we did not mannually label them. Considering the amount of this part of data is quite small, the potential side effect of them on the whole study could be ignored. With the multi-labeling tactic, we can expand our data, resulting in 5,154 more author-task pairs.

Table 1. Annotation sample for author-defined contribution standardization using CrediT

Author-defined Task	Contribution Role
Participated in critical discussion of the draft's initial findings and revision of the manuscript	Writing – review & editing
Statistically analyzed the data	Formal analysis
Contributed to the design and development of the project	Conceptualization, Methodology

Then we construct paper-level co-authorship networks(CAN for short) and co-contributorship networks (CCN for short) as proposed in [9] for each paper.

We are to analyze the partnership from three perspectives: partnership ratio; partnership strength, and partnership diversity. The formulas for the three measurements are as follows:

$$PR = \frac{number\ of\ CCN\ edges}{maximum\ number\ of\ edges} \quad (1)$$

$$PS = \frac{total\ weight\ of\ CCN\ edges}{total\ weight\ of\ CAN\ edges} \quad (2)$$

$$PD = \frac{number\ of\ unique\ contributor\ roles}{number\ of\ all\ contributor\ roles} \quad (3)$$

3 Preliminary Findings

Fig. 2 plots the partnership ratio in our dataset. It suggests that generally in each team exists some level of partnership, which results in some degree of division of labor in scientific collaboration. Specifically, more than 60,000 teams all collaborators are engaged in at least one collaborative task.

Fig.3 shows the partnership strength distribution, which indicates how closely collaborators in a team

[1] http://credi.niso.org/

are connected when doing research via the number of tasks two collaborators collaborated in a study. The figure demonstrates that on average the total edge weights of CCNs are 2.09, which means on average, two collaborators collaborate two divided tasks in each study. It also suggests that some collaborators in teams might be more involved in collaboration than others, indicating the existance of the partnership in scientific collaboration. Given that the average weight of CCNs is as double as those measured in CANs. And CCNs are naturally sparser than CANs as suggested by[9], the figure implies that partnership plays a role in scientific collaboration.

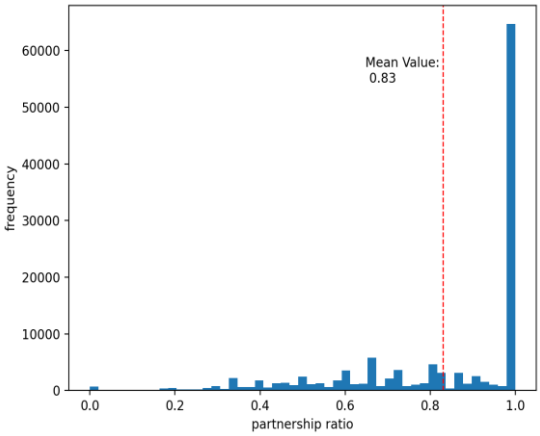

Fig. 2. The distribution of partnership ratio in our study

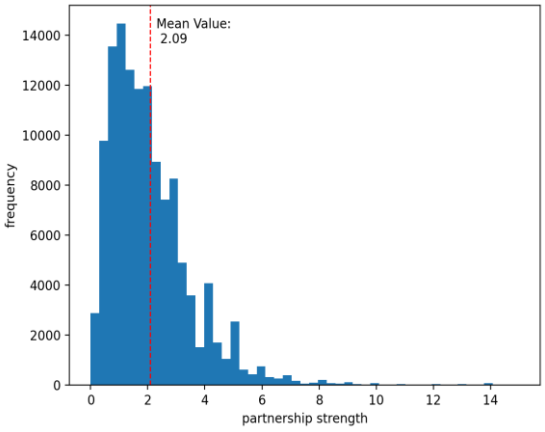

Fig. 3. The distribution of partnership strength in our study

Fig.4 shows the partnership diversity in scientific collaboration, which generally reflects how diverse it can be when two collaborators work as partners on the same tasks. It shows that generally partners usually perform 2.35 different contributor roles, on average. Given that there are 14 different contributor roles theoretically that two partners can work on, the diversity of the partnership remains quite mild.

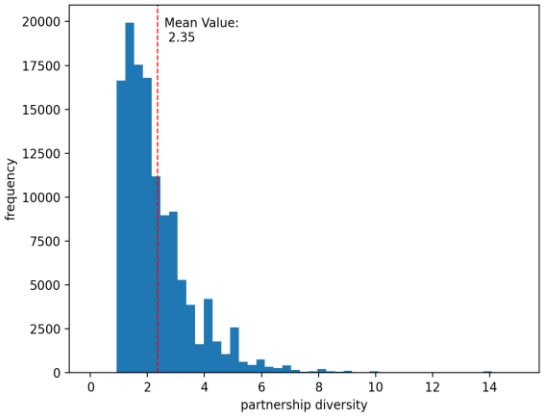

Fig. 4 The distribution of partnership diversity in our study

4 Conclusion and Future Work

In this preliminary study, we investigated how collaborators in teams work together to perform their research by understanding how two collaborators work together as partners which the traditional collaborative network usually overlooked naturally. By collecting author information from Scopus and author contribution statements from PLoS, we take the biology subject as an example and have examined more than 120,000 research articles and found that division of labor is quite common in scientific collaboration; that partnership as a form of division of labor is widely observed in our dataset; and that the diversity in contributing tasks between partners is generally mild. This study will shed light on understanding the mechanism in scientific collaboration via division of labor that coauthorship studies widely overlook. It helps us create research teams with higher levels of engagement and communication.

Acknowledgements

This article is an outcome of the youth project "Study of Scientific Collaborators' Scientific Effectiveness from The Perspective of Division of Labor" (ID: 72004054) supported by the National Natural Science Foundation of China and the project "The Causal Effect of Team Diversity on Team Performance" supported by the Fundamental Research Funds for the Central Universities (ID: B220201058).

References

[1] Lu C, Zhang C, Xiao C, Ding Y (2022) Contributorship in Scientific Collaborations: The Perspective of Contribution-based Byline Orders. Information Processing & Management 59:.

[2] García-Sánchez P, Díaz-Díaz NL, De Saá-Pérez P (2019) Social capital and knowledge sharing in academic research teams. International Review of Administrative Sciences 85:191–207.

[3] Haeussler C, Sauermann H (2020) Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. Research Policy 49:103987.

[4] Larivière V, Desrochers N, Macaluso B, et al (2016) Contributorship and division of labor in knowledge production. Soc Stud Sci 46:417–435.

[5] Allen L, Scott J, Brand A, et al (2014) Publishing: Credit where credit is due. Nature News 508:312.

[6] Devine EB, Beney J, Bero LA (2005) Equity, accountability, transparency: Implementation of the contributorship concept in a multi-site study. Am J Pharm Educ 69:61. https://doi.org/10/dqjngb

[7] Rennie D, Yank V, Emanuel L (1997) When Authorship Fails: A Proposal to Make Contributors Accountable. JAMA 278:579–585.

[8] Xu F, Wu L, Evans J (2022) Flat teams drive scientific innovation. Proc Natl Acad Sci USA 119:e2200927119.

[9] Lu C, Zhang Y, Ahn Y-Y, et al (2020) Co-contributorship Network and Division of Labor in Individual Scientific Collaborations. Journal of the Association for Information Science and Technology 71:1162–1178.