Types of Scientific Collaborators: A Perspective of Author Contribution Network

Chao Lu[1,2], Ying Ding[2], Yingyi Zhang[1], Yi Bu[2], Chengzhi Zhang[1]

[1] Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China
[2] Indiana University, Bloomington, IN 47408, USA
zhangcz@njust.edu.cn

Abstract. The purpose of this study is to investigate interaction between collaborators within individual studies by measuring how they made contributions to their studies. Author contribution network is constructed based on the author contribution statements of 140,000 full-text articles in *PloS* by viewing every collaborator as a node and a shared contribution as an edge. Three types of contributors are identified: general team-players, factotums, and mavericks. The preliminary result suggests that division of labor widely exists in scientific research and the latter two types of collaborators are common in small teams.

Keywords: Scientific collaboration, Author contribution, Network analysis.

1 Introduction

Collaboration between scholars has been reported to boost scientific research quantitatively and qualitatively [1-3]. However, bibliometric studies have usually employed co-authorship to indicate scientific collaboration, which means further interactions between co-authors are unclear and results in author credit assignment problems [3]. This study aims to parse author contribution statement of studies and construct a network for each research team to understand how team members finish their tasks and accomplish their studies.

2 Data and method

2.1 Data

Nearly 170,000 full-text articles published in *PLoS*[1] between 2006 and 2015 are collected in XML formats. The author contribution statements of these papers are extracted and parsed using natural language processing techniques assisted by necessary manual correction as exemplified in Table 1. Only those statements that are completely and correctly parsed are kept. Eventually, 147,707 articles comprise our final data set.

[1] https://www.plos.org/

Table 1. An author contribution parsed sample from our dataset[2].

Id	Authors	Task
1	EG; ES; JD	Conceived and designed the experiments
2	ES; JD; MH; JP; MS	Performed the experiments
3	EG; ES; FC; JD; JP; MS	Analyzed the data
4	ES; JD; MH; JP; MS	Contributed reagents
5	ES; JD; MH; JP; MS	Contributed materials
6	ES; JD; MH; JP; MS	Contributed analysis tools
7	EG; ES	Wrote the paper

2.2 Network Construction

Weighted undirected network model is adopted here to construct the author contribution network for every study. One *node* in an author contribution network denotes a collaborator. Every e*dge* in the network is a task shared by co-author(s). If one task is shared by more than two collaborators, these nodes share an edge; if one task is taken up by only one collaborator, this node is given a self-looped edge. The *weight* of a given edge is the number of tasks two authors share in one single study. For example, in Fig. 1-A, there are four collaborators in the study. The weight of the edge (C_2, C_3) is three, which means authors C_2 and C_3 worked together on three different tasks. The weight for C_1 is 2, meaning that C_1 independently finished two different tasks.

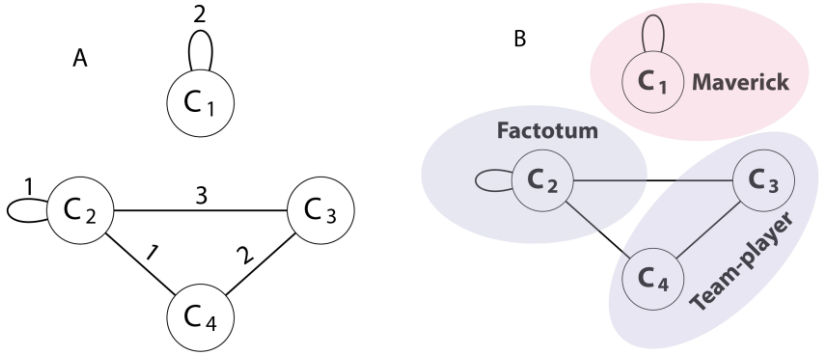

Fig. 1. Author contribution network and collaborator types based on network attributes.

2.3 Three types of collaborators

Based on the constructed author contribution network, three types of collaborators are exemplified in Fig. 1-B. *Team-players*: those who work with other collaborators in an individual study. In an author contribution network, team-players are the nodes only linked with other node(s) via an edge (e.g., C_3 and C_4). *Mavericks*: those who finish their tasks on their own in an individual study. In an author contribution network, mav-

[2] See original text: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000042#authcontrib

ericks are the nodes with only self-looped edge(s) (e.g., C_1). Factotums. *Factotum* usually means some employee who can tackle any tasks with many skills; while in this study, factotums are those who are not only recognized as team-players but also mavericks. In an author contribution network, factotums, thus, are the nodes with both self-looped edge(s) and edge(s) connected to other node(s) (e.g., C_2).

3 Preliminary Results

3.1 Overview

Team size

Fig. 2-A demonstrates the team size distribution. The most frequent team size (18,894 articles) in the data set is five. The articles with top 20 frequent team sizes make up 99% in our data set [2]. 87% articles are collaborated by no more than 10 authors.

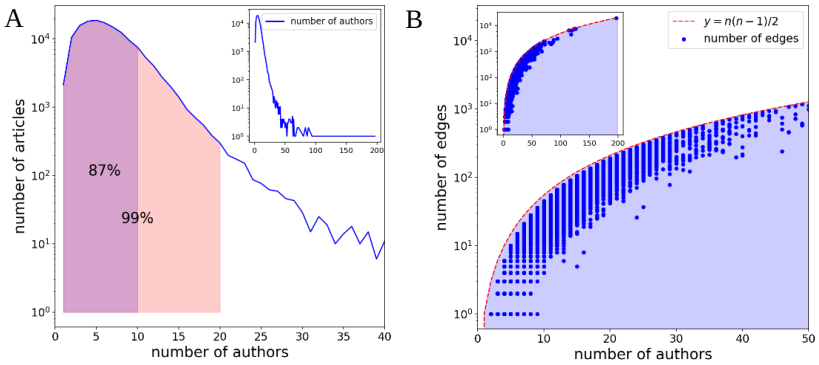

Fig. 2. Author number distribution (A) and their edge distribution (B, self-loops excluded).

Teamwork

We also plot the scatter of team size and number of edges. Self-looped edges are removed, for we are more concerned about teamwork among collaborators in this section. The red dash line shows the maximum edges given number of authors ($C_n^2 = \frac{n(n-1)}{2}$). The plot implies that teamwork among authors are more than common. However, we still observe that only a few articles reach the red line, showing these collaborations consist of not only teamwork but individual performance, which means some degree of division of labor exist in smaller teams [4].

3.2 Team-players, Factotums, and Mavericks

We plot the bar-chart of the three types of collaborators in Fig. 3. It also suggests that most collaborators in research teams are team-players, who share tasks; then follow the factotums; and, lastly mavericks are least common collaborators. We also observed that when teams grow larger, the last two types of collaborators become rare. However,

given most studies are collaborated by no more than ten authors, the two types of collaborators become interesting to investigate.

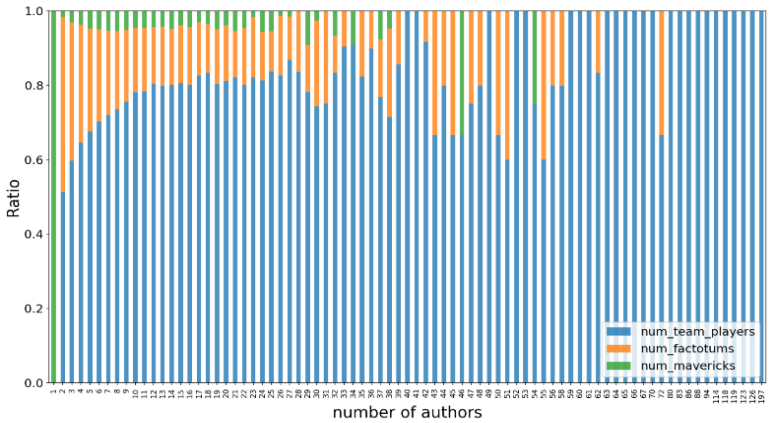

Fig. 3. Distribution of the three types of collaborators based on team size.

4 Conclusion

This study utilizes the author contribution statements embedded in full-text articles to build an author contribution network for every single study. The network structure helps us identify three types of collaborators: team-players, mavericks, and factotums. The results suggest that most authors are team-players, contributing tasks together; while there still remain some mavericks and factotums, especially in smaller teams. Given that most articles (87%) are collaborated by less than ten collaborators in our data set, the collaborations among authors seem to show some degree of division of labor, and may also imply that exiting co-authorship network analysis can be improved with author contribution networks. However, this preliminary study remains some space to improve. The next step is to closely investigate the behavior of these factotums and mavericks and find out why they work individually in a team.

References

1. Dong, Y., Ma, H., Shen, Z., & Wang, K.: A century of science: Globalization of scientific collaborations, citations, and innovations. arXiv preprint arXiv:1704.05150 (2017).
2. Wuchty, S., Jones, B. F., & Uzzi, B.: The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036-1039 (2007).
3. Katz, J. S., & Martin, B. R.: What is research collaboration? Research policy, 26(1), 1-18 (1997).
4. Corrêa, E. A., Jr., Silva, F. N., Costa, L. D. F., & Amancio, D. R.: Patterns of authors contribution in scientific manuscripts. Journal of Informetrics, 11(2), 498-510 (2017).