Understanding Roles of Collaborators from Their Byline Orders and Affiliations

Chao Lu[1], Chengwei Zhang[2], Ying Ding[2], Dandan Ma[3], Yingyi Zhang[4],

[1] luchaoink@gmail.com, [2] zhang334@indiana.edu, [3]dingying@indiana.edu, [3]jane.ma1030@gmail.com, [4]yingyizhang@njust.edu.cn

[1] Hohai University, Business School, 8 West Focheng Road, Nanjing, Jiangsu, 210098 (China)

[2] Indiana University, School of Informatics, Computing, and Engineering, 700 N Woodlawn Ave, Bloomington, IN 47408 (USA)

[3] Nanjing University of Finance & Economics, School of Information Engineering, 3 Wenyuan Rd, Nanjing, Jiangsu, 210046 (China)

[4] Nanjing University of Science & Technology, School of Management and Economics, 200 Xiaolingwei St, Nanjing, Jiangsu, 210094 (China)

Introduction

Collaborations are prevailing in science currently (Wuchty, Jones, & Uzzi, 2007). It is believed to have incomparable advantages such as bringing diverse ideas to breed innovations and sharing various facilities and equipment to enrich scientific practice. Collaborations have been encouraged in many disciplines. However, little is known how a team really functions from the detailed division of labor within the team. Here, we continue our study on scientific collaboration and division of labor within individual scholarly articles (Lu et al, 2018) by analyzing the relationship between collaborators' roles and their byline orders and affiliations.

Data and Method

Author Information Parsing

Nearly 170,000 full-text articles published in *PLoS*[i] from 2006 to 2015 are collected in XML formats with their metadata, including author information. First, 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, leaving us 138,787 articles correctly parsed. Then, authors' full names, byline orders, and affiliation information are extracted from the authors' full names to match the author name abbreviations in contribution statements, byline orders and affiliations (Here, we removed 33,595 articles where author names cannot be completely matched). Then, we remove 1,331 single-authored articles, which leads to articles excluded from our initial data set. So our final data set contains 103,861 articles with their author contribution statements parsed to identify their roles in collaboration and their author byline orders and affiliations to assist us understand their roles in collaborations.

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

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

Types of collaborators

Network Construction

Weighted undirected network model is adopted here to construct an author co-contributorship network for every study using the parsed author contribution statements. One *node* in the network denotes a collaborator. Every e*dge* in the network represents task(s) shared by co-author(s). A self-looped edge indicates task(s) performed independently. The *weight* of an edge is the number of tasks performed by author(s). Then we can identify three types of collaborators from the network: Specialists, Versatiles, and Team-players (Lu et al, 2018).

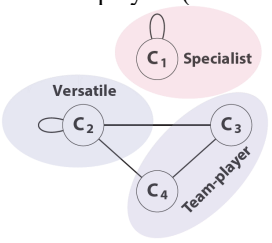

Figure 1. Types of Collaborators edited from (Lu *et al,* 2018).

Byline orders

Byline order is the order where authors' names are assigned in their publications, usually demonstrating some degree of their contributions to their work or teams (Corrêa Jr, Silva, Costa, & Amancio, 2017). We use Formula (1) to calculate each author's

normalized byline order in every single article and mitigate the effect caused by different numbers of authors among articles:

$$NB_i^j = \frac{B_i^j - 1}{N_i - 1}, N_i \geq 2, 0 \leq NB_i^j \leq 1 \quad (1)$$

Where B_i^j is one author's byline order in his/her collaborated article i and N_i denotes the total number of authors in the article i and NB_i^j represents the normalized byline order of the author, which is in the range of [0,1].

Affiliation index

we proposed AFI (Affiliation Index) to depict the disparity between one author's affiliations and the affiliations of the whole team. We use Formula (2) to calculate AFI index of each author:

$$\text{AFI(k)} = \frac{N_c^k}{N-1}, N \geq 2 \quad (2)$$

In the formula, N_c^k stands for the number of author k's colleagues within the team (sharing same affiliations); N denotes the total number of authors in the study. For instance, one paper is collaborated by three authors (i.e., A, B, and C): A is affiliated with aff_1 and aff_2; B belongs to aff_2 and aff_3; and C's affiliations is aff_3. The AFI for author A is $\frac{1+0}{2} = 0.5$. Intuitively, when AFI is 1, it means the author is a colleague of the rest of authors; when AFI is 0, it means the author is affiliated with a different organization from other collaborators.

Result

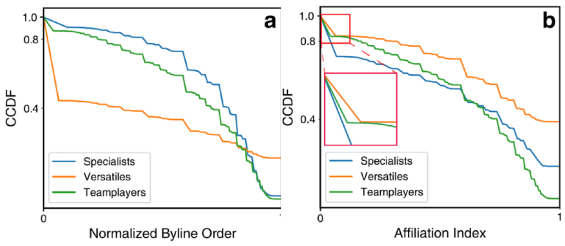

Figure 2. CCDFs for Byline orders (a) and AFIs (b) of collaborators ($p<0.0001$ in both Kolmogorov–Smirnov test between groups).

Review of our former study

Given the co-contributorship network of a paper, we defined three types of contributors: Specialists, Team-players, and Versatiles (in Figure 1). Specialists are those who contribute to all their tasks alone; team-players are those who contribute to every task with other collaborators; and versatiles are those who do both. We found that team-players are the majority and tend to contribute to the five most common tasks as expected, such as "data analysis" and "performing experiments". The specialists and versatiles are more prevalent than expected from random-graph null models. Versatiles tend to be senior authors associated with funding and supervisions. Specialists are associated with two contrasting roles: the supervising role as team leaders or marginal and specialized contributions.

Byline Order

Figure 2(a) plots the CCDF (complementary cumulative distribution function) for the three types of collaborators. In the plot, versatiles usually demonstrate their leading positions in collaborations among authors, which takes accords with our observations; while specialists usually sign their names at the end of their bylines, suggesting their more marginal contributions to teams. In between lie the team-players, who usually perform the common tasks within a team; their names are more frequently placed in the middle. However, versatiles can also occasionally appear at the end of bylines, indicating their authorities in research as corresponding authors.

Affiliation index

A larger affiliation index value of an author usually indicates one collaborates with his/her colleagues in a single study. Versatiles demonstrate their much stronger connections with other collaborators than those of the other two types of collaborators, confirming their core role in communication and coordination. Team-players, as the main labor source, tend to have a larger affiliation index value than specialists. However, specialists take over the leading position when affiliation value exceeds 0.5. It might suggest that specialists can also partake the role of communication and coordination as versatiles within teams and thus, confirm our former findings.

Conclusion

In this study, we use authors' byline orders and affiliations to understand different types of collaborators. The results extend our former findings about different types of collaborators and their roles in study and also imply the usefulness of affiliation index to identify author roles in scientific collaborations.

References

Corrêa Jr, E. A., Silva, F. N., Costa, L. D. F., & Amancio, D. R. (2017). Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics*, *11*(2), 498-510.

Lu, C., Ding, Y., Zhang, Y., Bu, Y., & Zhang, C. (2018). Types of Scientific Collaborators: A Perspective of Author Contribution Network. *iConference 2018 Proceedings*.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039.

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

[ii] See original text: http://t.cn/EMSVcGp