Does Collaboration Bring High-impact Studies? A Preliminary Study

Chao Lu[1,2] **Yi Bu[2]** **Chenwei Zhang[2]** **Ying Ding[2]** **Vetle I. Torvik[3]** **Chengzhi Zhang[1,*]**

[1]School of Economics and Management, Nanjing University of Science and Technology

[2]School of Informatics and Computing, Indiana University

[3]School of Information Sciences, University of Illinois

luchaonjust@gmail.com buyi@iu.edu Zhang334@iu.edu dingying@indiana.edu vtorvik@illinois.edu zhangcz@njust.edu.cn

ABSTRACT

Team formation becomes important when scientific collaboration become common and important to scholars. This study initially tries to understand how research team size may influence the impact of a study measured by number of citations. Thus, publications in Biology domain are specifically investigated by correlation analysis between team size and scientific impact. The weak correlation suggests that larger team size can positively increase the impact of scientific publications on average; however, it cannot literally bring more highly-cited articles.

Keywords

Scientific collaboration; scientific impact; team size.

INTRODUCTION

Nowadays, scientific collaborations become common and important to scholars, especially when its positive effects have been well investigated among researches (Wuchty, Jones, & Uzzi, 2007). This preliminary study is to discover whether larger team size can bring more high-impact studies in order to better understand scientific success for scholars.

METHODOLOGY

To investigate the relationship between team size and scientific impact, we follow the road map as shown in Figure 1.

Data

Structured fulltext data of roughly 170,000 scientific articles are collected from *PLoS* journals. To keep the consistence and prevent potential biases caused by different subjects, we simply select the articles in Biology (49,350) as our final data set. Then their authors' information is extracted, including author's first and last names and their rankings on the author list.

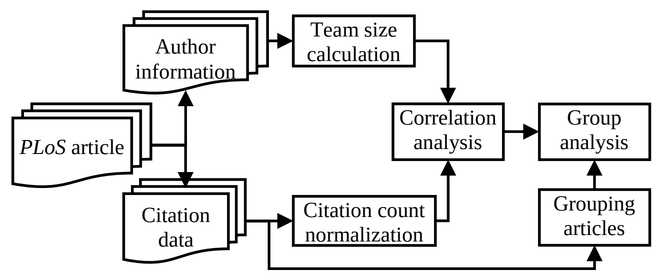

Figure 1. Road map of this study.

Using the APIs[1] developed by *PLoS*, we access the citation count of each paper in *PLoS*. The citation data for these 49,350 articles are obtained from February 3rd to 6th in 2016, a very short time period, so that we can reduce the possible limitations in citation data caused by different harvest timelines.

Method

With these data, we calculate some variables to measure the team size and scientific impact. First, we use the number of authors in the author list to measure the *team size* of each study shown as follows:

$$TS = N_{authors} \quad (1)$$

Then, we calculate the *normalized citation count* by month to mitigate the negative effects of paper age difference:

$$NCC = \frac{Citation\ count}{Month(Harvest\ date - Publish\ date)} \quad (2)$$

The function $month()$ is to calculate the paper's age in month.

Correlation Analysis

We also conduct a correlation analysis among variables. The Kolmogorov-Smirnov (KS) tests suggest that both variables

*Corresponding author: zhangcz@njust.edu.cn

[1]https://github.com/PLOS?page=1

are not normally distributed (p<0.001), so the Spearman test is adopted to discover the potential relationships.

Article Group Assignment

To better understand the relationship between team size and scientific impact, we also divide the articles into three groups according to their NCC ranking: 493 high-impact articles (top 1%), 4,935 medium-impact articles (9%), and 43,922 low-impact articles according the Bradford distribution (Leimkuhler, 1967)

PRELIMINARY RESULTS

Table 1 shows the result of correlation analysis, where the team size shows a weak positive correlation with the normalized citation count (NCC). That is to say, as the size of a team grows, the impact of a study is relatively higher—more citations can be received. To further discover the relationship between them. We plot the average NCC grouped by the team size in 22 intervals (Figure 2).

	Team size	NCC
Team size	1.000	
NCC	.156**(.000)	1.000

**. Correlation is significant at the 0.01 level (2-tailed); and p-values are given in parentheses.

Table 1. Results of correlation analyses.

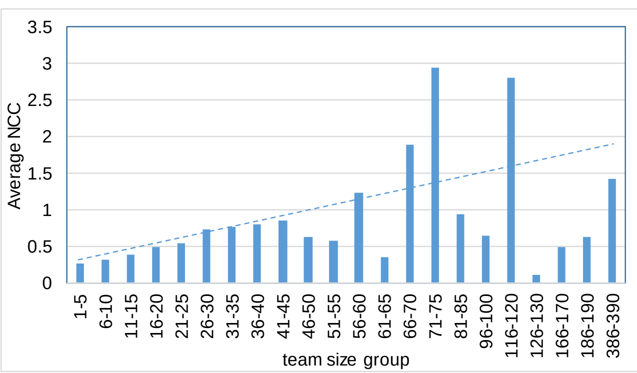

Figure 2. Average NCC grouped by team size.

From Figure 2 we can see that in Biology researches when the team size is smaller than 45, the average NCC is more strongly correlated with team size; when the team size is larger than 45 the samples become sparse, which cause more random distributions in Figure 2 (see in Figure 3 (A)). To discover the potential reasons, we further divide the articles into three categories. We find that the reduction of low-impact research (Figure 3 (B)) may explain the positive effect of team size on the scientific impact of a study that larger teams can help reduce the percentage of low-impact studies instead of producing more high-impact ones.

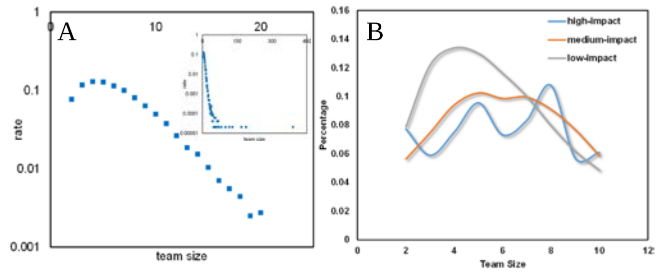

Figure 3. Team size distribution (A) and article percentage by group (B)

CONCLUSION

This preliminary study tries to understand whether an increasing team size will affect the impact of a scientific paper. Research teams in Biology are investigated to examine the relationship between team size and scientific impact. The Spearman correlation tests suggest that the increasing team size can boost the impact of a study. We suggest that less low-impact studies in larger teams can explain why their average scientific performances are much higher than those in smaller teams on average.

This study also remains some limitations. For example, this data set only comprises publications in Biology; we believe that more domains should be taken into consideration. Additionally, the distribution within smaller slides can be examined to fully understand the effect of team size on the impact of scientific publications.

ACKNOWLEDGEMENT

This work is supported by Major Projects of National Social Science Fund (No. 16ZAD224), Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (No. MJUKF201704) and Qing Lan Project, and CSC scholarship (File NO. 201606840093).

REFERENCE

Leimkuhler, F. F. (1967). The Bradford distribution. *Journal of documentation*, *23*(3), 197-207.

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

ADDRESS INFORMATION

The addresses of authors are given as follows:

1, 6: 200 Xiaolingwei Street, Xuanwu District, Nanjing, Jiangsu 210094, China;

2, 3, 4: 919 E. 10th Street, Bloomington, IN 47408, U.S.A.;

5: 501 E. Daniel St, Room 221, Champaign, IL 61820, U.S.A.