Differential Analysis on Performance of Scientific Collaborations with the Evolution of Entity Popularity

Fang Tan[†]
School of Information Management
Sun Yat-sen University
Guangzhou Guangdong China
cathytf@163.com

Tongyang Zhang
School of Information Management
Sun Yat-sen University
Guangzhou Guangdong China
tzhang39@163.com

Jian Xu [*]
School of Information Management
Sun Yat-sen University
Guangzhou Guangdong China
issxj@mail.sysu.edu.cn

ABSTRACT

In order to investigate the impact of research topic selection time on output performance of scientific collaborations, the aim of this study is to develop a differential analysis framework of scientific collaboration performance at different stages of entity popularity. The framework consists of three main sections: (1) data acquisition and processing; (2) stage division of entity popularity; (3) differential analysis on performance of scientific collaborations at different stages of entities popularity. Our findings show that the popularity stage that research topics are going through can play a role in the collaboration output performance.

CCS CONCEPTS

• Applied computing → Bioinformatics • Applied computing → Computing in government • Information systems → Information retrieval

KEYWORDS

Entitymetrics; Research popularity; Scientific collaboration; Collaboration performance

1 Introduction

The ultimate success of scientific collaborations depends on a number of factors, among which the importance of identifying promising research topics as a key success factor should not be underestimated [1]. The selection of a promising research topic can not only help the scientific collaboration develop a reputation for having an acute sense of active research domain, but also encourage the process of scientific discovery scientists and promote the development of the whole research field [2]. However, research on selecting topics behavior of scientists seldom consider about the timing issue of topic selection, and most of them only focus on the development law that underlie an individual's behavior. With more and more collaborative research studies, scientific collaborations have gradually replaced individuals as the mainstream research unit. The study regards bio-entities as research topics to analyze the evolution of topic popularity in biomedicine related research from the perspective of entitymetrics. Through analyzing the effect mechanisms of entity popularity on performance of scientific collaborations, we provide a theoretical reference for relevant decision makers in research topic selection and the management of scientific research project.

2 Methodology

The framework of differential analysis on performance of scientific collaborations in different stages of entity popularity is shown in Figure 1.
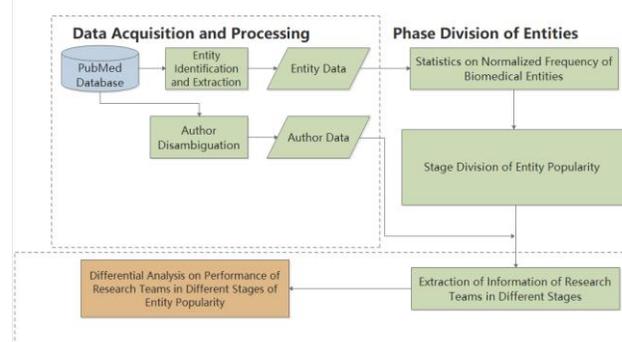
Figure 1: Research Framework of the Differential Analysis on Performance of Scientific Collaborations with the Evolution of Entity Popularity

In Figure 1, three functional modules of the analysis framework and concrete work done in modules are as follow.

Data acquisition and processing. By using BERT [3] and Bio BERT [4], we collect 317 Gene/Protein entities from the title and abstract of 1,899,671 articles between 1988 and 2017 in PubMed with author names disambiguated. All cited information of articles is obtained from Web of Science (WOS).

Stage division of entity popularity. After the normalization processing of entity frequency, we deal with the division of entity popularity stages based on the model tree proposed by Ma [5] As seen in Figure 2, k is less than or equal to -0.05 when the entity's popularity stage is descending (short for "descending stage"); k is greater than or equal to 0.05 when the entity's popularity stage is ascending (short for "ascending stage").

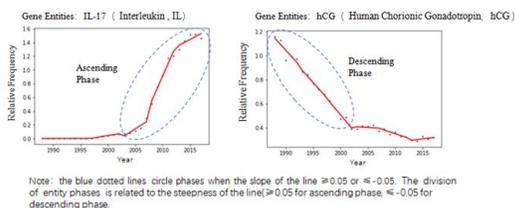
[*] Corresponding author.

Figure2: Example for the Division of Entity Popularity Stage.

Differential analysis on performance of scientific collaborations in various stages of entity popularity. Through comparing statistics on the number of published periodical articles and citations of teams of each scientific collaboration in the ascending and descending stages, the impact of entity popularity stage that a collaboration is going through when selecting a research topic on its performance and the mechanism behind is discussed. For scientific collaborations studying entities in various popularity stages, we recognize authors of the same article as a research team. Teams in which the number of authors is less than three or over 10, and groups in which ages of authors are all over 45 are ruled out. As long as the target entities appear in the title of abstract of an article, we consider the team has studied the entities. The index of normalized citations [5] can be calculated as below.

$$SC=(PC-AC)/SDC \quad \#(1)$$

where, *PC* denotes the absolute citation of a certain article, AC denotes the average absolute citation of all articles in the publication year of the article, SDC denotes the standard deviation of the absolute citation of all articles in the same year, and SC denotes the normalized citation (SC).

3 Preliminary Results

3.1 Overview of Experimental Data

Figure 2 shows the evolution of the number of articles of teams in the ascending and descending stages by year. After the 21st century, the number of teams conducting research in the ascending stage of entities popularity has experienced rapid growth while the number of teams on the other side shows a slightly declining trend.
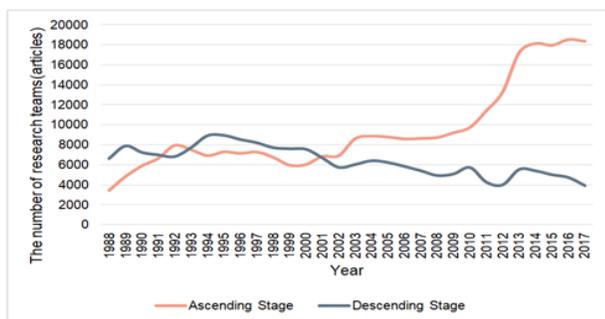

Figure 3: Yearly Distribution of the Number of Research Teams in the Ascending Stage and Descending Stage

3.2 Differential Analysis on Performance of Scientific Collaboration Research

As shown in Figure 3, either the overall average normalized citations or the annual average normalized citations of teams in the ascending stage remains significantly higher than that of teams in the descending stage. For research outputs of teams in the ascending stage, the earlier the published year, the more normalized citations compared to teams in the descending stage. It illustrates that team research in the ascending stage is more likely to have a far-reaching academic influence than that in the descending stage.
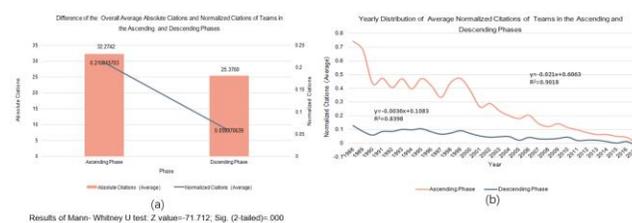

Figure 4: Difference of the Average Normalized Citations and Absolute Citations of Teams in the Ascending Stage and Descending stage

4 CONCLUSION AND FUTURE WORK

4.1 Conclusion

The study proposes a preliminary research design, applying the idea of entitymetrics to team research performance and engaged in preliminary differential analysis.

Results show that: The popularity stage that research topics are going through can play a role in the research performance of scientific collaborations

Compared with the descending stage, the ascending stage puts more positive impacts on collaboration research performance.

For different scientific collaboration modes, the study can be used as a reference in choosing research topics. For instance, when selecting topics for research, authors can implement the strategy of choosing a topic the popularity of which is in its ascending stage as a way to moderate the negative influence of descending stage on collaborative performance.

4.2 Future work

In the future, differences in other aspects such as personnel composition of a team will be considered in the study. Furthermore, an important question waiting to be answered is how sustentation funding, one of the most important external resources to encourage team research, distributes in the two types of teams? Therefore, we will further study other aspects of characteristics of teams at different stages of topic selection in the future.

ACKNOWLEDGMENTS

REFERENCES

[1] Ran Xu, Arash Baghaei Lakeh and et al. 2020. Examining the characteristics of impactful research topics: A case of three decades of HIV-AIDS research. J*ournal of Informetrics*, 15(1): 101122. DOI: https://doi.org/10.1016/j.joi.2020.101122.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI: https://doi.org/10.1145/567752.567774.

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI: https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY.

[5] J. Guan, Y. Yan, J. J. Zhang. 2017. The impact of collaboration and knowledge networks on citations. *Journal of Informetrics*, 11(2): 407-422.