

# Measuring interdisciplinary interactions using citation analysis and semantic analysis

Lu Huang<sup>1</sup>, Xingxing Ni<sup>1</sup>, Xiang Chen<sup>1✉</sup>, Yi Zhang<sup>2</sup>

<sup>1</sup> School of Management and Economics, Beijing Institute of Technology, China

<sup>2</sup> Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

bjchenxiang@hotmail.com

**Abstract.** Interdisciplinary interactions and integrations have become a major feature of the current development of science and technology. How to measure the strength of interdisciplinary interactions between two disciplines is a crucial issue. In our study, we propose a novel measurement framework based on both citation analytics and semantic analytics, which integrates three indicators - direct citation, bibliographic coupling and research content. Especially, LDA model is incorporated with a word embedding model to create a semantic solution that effectively constructing discipline-keyword vectors based on bibliometric data. At last, entropy method is applied with these three indicators to assess the interdisciplinary interactions strength. The interactions between information science & library science and other six subjects are analyzed as the case study to demonstrate the reliability of the methodology, with subsequent empirical validations.

**Keywords:** interdisciplinary interactions · citation analysis · semantic analysis · word embedding

## 1 Introduction

The importance of accelerating interdisciplinary interactions among disciplines is increasingly recognized by people [1]. For example, 2017 Nobel Prize in chemistry was awarded to physicists for solving biological problems. Cross combination of information, methods, techniques, tools, perspectives, concepts and/or theories among different disciplines or bodies of specialized knowledge has been promoted to form interdisciplinary [2], which enables to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice[3]. While interdisciplinary scientific research is increasingly concerned by science and technology policy and management departments, people gradually began to think about how to measure the strength of interdisciplinary interactions. Measuring interdisciplinary interactions has been considered as a critical issue for the management practice of interdisciplinary in scientific research management departments [4], which

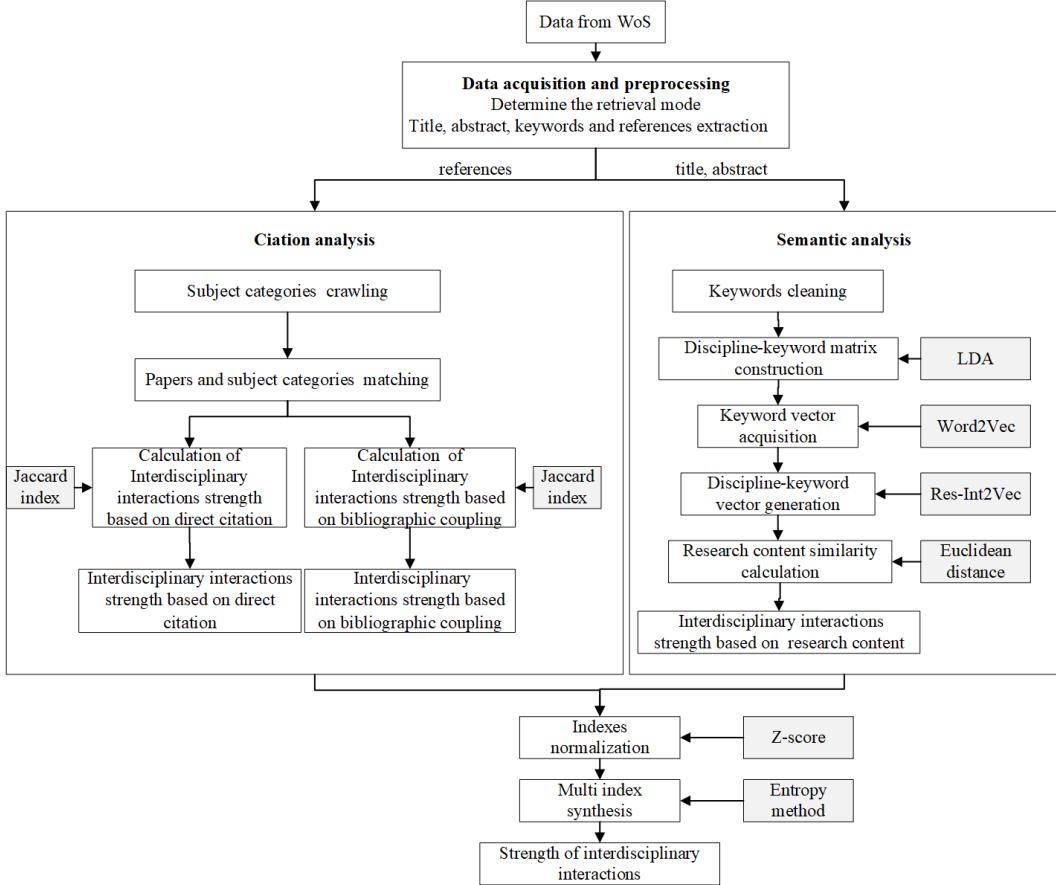
is conducive to evaluate the degree of interdisciplinary and grasp the current developing situation of the discipline, and also optimize the disciplinary layout in the future [5].

Many studies have been devoted to how to measure the interdisciplinary nature of basic research, i.e., citation analysis [6], co-author analysis [7], subject categories (SCs) and journal disciplines [8]. As citation analysis could trace the cited literature and identify the learning and referring relationships between disciplines, it has been widely applied [9]. Complementary to these literatures, some researchers are from the perspective of research content analysis, which is more microcosmic and specific and can reveal the specific integration and diffusion process of knowledge, exploring the development and change of disciplinary knowledge structure [10], for example, Xu et al. have explored the interdisciplinary of the topics based on co-word analysis [2]. However, co-word analysis ignores terminological variations (e.g., “data mining” and “data analytics”) and semantic relationships between terms from disciplines [11].

In this paper, we propose a novel framework of measuring the strength of interdisciplinary interactions between two disciplines based on citation analysis and semantic analysis from cognitive dimension. Especially, an LDA model is incorporated with a word embedding model to construct discipline-keyword vectors, which could explore the semantic and contextual relationships in order to capture their intersections. We demonstrate our method via a case study of interdisciplinary interaction measurements between “Information science & Library science” and other six disciplines, i.e., “Education & Educational research”, “Computer science, Information systems”, “Management”, “Economics”, “Mathematics, Applied”, and “Psychology, Applied”.

## 2 Methodology

The framework of measuring interdisciplinary interactions is shown in Fig. 1.



**Fig. 1.** The framework of interdisciplinary interactions measurements

## 2.1 Data acquisition and preprocessing

The full records and references of research articles of two specific disciplines (Subject Category) are acquired from the Web of Science (WoS) as the input, which include data such as titles, abstracts, keywords and references. Then, we preprocess the data, including the following work:

### Subject classification matching of references.

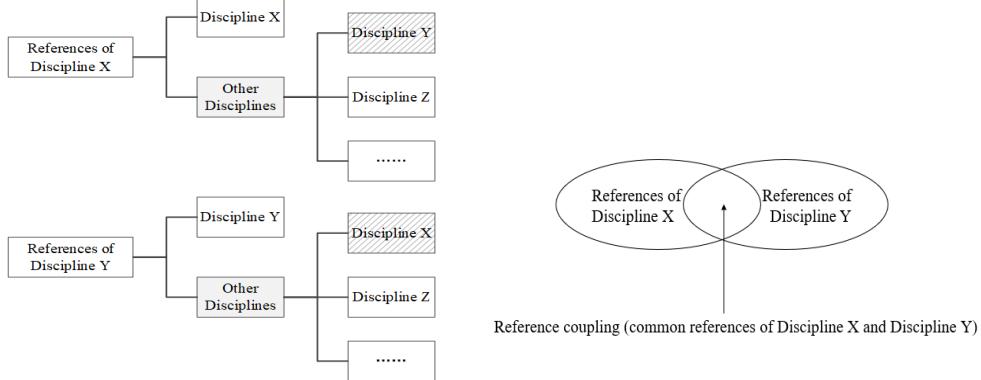
Because the downloaded citation information is only the journal to which the citation belongs, and there is no discipline (Subject Category) to which the journal belongs, we need to use Python to obtain the Subject Category information of all journals on the Journal Citation Reports (JCR) website of WoS. In addition, the Journal of the reference in the citation information of some discipline downloaded from WoS uses the abbreviation of the journal, so the full journal title should be obtained from JCR database. Finally, we construct a comparison table of the abbreviation-full journal title-Subject Category, and match the Subject Category of references according to this.

### Term clumping.

A natural language processing (NLP) technique is applied to retrieve key terms from the titles and abstracts, and a term clumping process removes noise, consolidates terms, and identifies core terms [12]. We call the key terms after term clumping by keywords.

### 2.2 Citation analysis

The aim of this part is to measure the interdisciplinary degree between two disciplines from the perspective of citation analysis. In this paper we provide two types of citation analysis: one is direct citation, which reflects a two-way interactive relationship and the most direct knowledge exchange between two disciplines [13]; the other is bibliographic coupling, which reflects the situation that the two disciplines cite other literatures together [4]. Jaccard similarity coefficient [14] is applied for both direct citation and bibliographic coupling, and the higher the value is, the stronger the strength of interdisciplinary interactions is.



**Fig. 2(1).** Subject classification structure of references of Discipline X and Discipline Y

**Fig. 2(2).** Schematic diagram of reference coupling calculation

**Fig.2.** Citation relationship between Discipline X and Discipline Y

### Interdisciplinary interactions strength based on direct citation.

The subject classification structure of references of Discipline X and Discipline Y is shown in Fig.2 (1). Here, we denote  $IJ_{xy}^{dc}$  as the strength of interdisciplinary interactions between Discipline X and Discipline Y based on direct citation. The direct citation relationship between discipline X and Y focus on two sets (diagonal line sections): the references of Discipline X belonged to Discipline Y, and the references of Discipline Y belonged to Discipline X. Following Jaccard's calculation formula, the numerator is the intersection of the above two parts, which is the minimum reference number of the two sets; while the denominator is the sum number of references belonged to other disciplines (shadow sections) minus the numerator. Therefore,  $IJ_{xy}^{dc}$  can be represented as:

$$IJ_{xy}^{dc} = \frac{\min\{j_{xy}, j_{yx}\}}{k_x + k_y - \min\{j_{xy}, j_{yx}\}} \quad (1)$$

Where  $j_{xy}$  is the number of references of Discipline X belonged to Discipline Y,  $j_{yx}$  is the number of references of Discipline Y belonged to Discipline X,  $k_x$  represents the number of references of Discipline X belonged to disciplines other than Discipline X,  $k_y$  represents the number of references of Discipline Y belonged to disciplines other than Discipline Y.

#### **Interdisciplinary interactions strength based on bibliographic coupling.**

As shown in Fig.2 (2), bibliographic coupling relationship between Discipline X and Discipline Y focus on the common references. Here, we denote  $II_{xy}^{bc}$  as the strength of interdisciplinary interactions between Discipline X and Discipline Y based on bibliographic coupling. Following Jaccard's calculation formula, the numerator is the number of common references of Discipline X and Discipline Y, and the denominator is the number of references union of Discipline X and Discipline Y. Therefore, and  $II_{xy}^{bc}$  can be represented as:

$$II_{xy}^{bc} = \frac{o_{xy}}{q_x + q_y - o_{xy}} \quad (2)$$

Where  $q_x$  is the number of references of Discipline X,  $q_y$  is the number of references of Discipline Y,  $o_{xy}$  represents the number of common references of Discipline X and Discipline Y.

In this part, we finally generate two indicators:  $II_{xy}^{dc}$  and  $II_{xy}^{bc}$ .

### **2.3 Semantic analysis**

The purpose of this part is to measure the interdisciplinary interactions strength between two disciplines by exploring semantic relationship, which is reflected by the keywords [15]. With the development of disciplines, due to the interaction between disciplines, the overlapping of disciplines can be expected in some areas of knowledge [4]. Therefore, discipline-keyword vectors are constructed to calculate the similarity of research content of disciplines.

First, Latent Dirichlet Allocation (LDA) model, which is a probabilistic topic model and defines a global hierarchical relationship from words to a topic and then from topics to a document [16], is applied to obtain keyword distribution of both Discipline X and Discipline Y. Specifically, we synthesize the keywords which generated by cleaning the paper data (titles and abstracts) of Discipline X and Discipline Y to their own keyword documents. Through LDA model, discipline documents are represented as topic probability distribution, and topics are represented as keyword probability distribution. Then, the discipline-keyword matrix could be obtained by multiplying discipline-topic matrix and topic-keyword matrix. We denote A (m, p) as the discipline-keyword matrix of Discipline m for keyword p.

Second, Word2Vec model is used to generate keyword vectors. Word2vec is a word embedding model to represent keywords as word vectors, which could capture context semantic information [17]. In our study, skip-gram modules is applied, since it has proven to have a tiny advantage with bibliometric data [18]. The inputs are word sequences generated from the text in the abstracts and titles. Keywords of Discipline X and Discipline Y are then mapped as vectors originating from a point in a multi-dimensional semantic space.

Then, the research content of disciplines could be converted into vector representation by loading the keyword vectors created in previous step into matrix A (m, p), and we denote  $V_{mp}$  as a discipline-keyword vector. It can be represented as:

$$V_{mp} = \sum A(m, p) * V_p \quad (3)$$

where  $V_p$  denotes the vector of keyword p.

Lastly, the similarity between discipline-keyword vectors of disciplines is calculated according to the Euclidean distance. We denote  $II_{xy}^{rc}$  as the strength of interdisciplinary interactions between Discipline X and Discipline Y based on research content, and it can be represented as:

$$II_{xy}^{rc} = \frac{1}{\rho_{xy}} \quad (4)$$

where  $\rho_{xy}$  is the Euclidean distance between the vectors of Discipline X and Discipline Y.

In this part, we finally generate indicator  $II_{xy}^{rc}$ .

## 2.4 Multi-index Synthesis

At this step, three indicators  $II_{xy}^{dc}$ ,  $II_{xy}^{bc}$ ,  $II_{xy}^{rc}$  are standardize with the Z-score method. In order to integrate the strength of interdisciplinary interactions of the three aspects more reasonably, it is necessary to set the weight of each index. There are two ways to determine the index weight: subjective weight and objective weight. We use the objective weighting method, because it can overcome the randomness of subjective weighting, and more objectively represent the importance of the weight. By comparing various objective weighting methods, we decided to use entropy weight method to calculate.

Entropy weight method is an objective method to determine the index weight based on mathematical statistics and the basic principle of information theory [19]. It can effectively consider the variation degree of indicators of the strength of interdisciplinary interactions. In this paper, the entropy weight of each index is defined as  $W_\beta$ . The calculation method is shown in formula (5), (6) and (7).

$$f_{\alpha\beta} = \frac{q_{\alpha\beta}}{\sum_\alpha q_{\alpha\beta}} \quad (5)$$

Where,  $f_{\alpha\beta}$  is the characteristic specific gravity of the index,  $q_{\alpha\beta}$  is the value of each indicator.

$$E_\beta = \frac{\sum_\alpha f_{\alpha\beta} \ln(f_{\alpha\beta})}{\ln(N)} \quad (6)$$

Where,  $E_\beta$  is called information entropy. N is the number of indicators. If the information entropy of an index is smaller, it means that the variation degree of the index value is greater, the amount of information covered is more, and its influence ability in the overall evaluation is greater, so it has a greater weight.

$$W_\beta = \frac{1 - E_\beta}{M - \sum_\beta E_\beta} \quad (7)$$

The comprehensive strength of interdisciplinary interactions between Discipline X and Discipline Y  $H_{xy}$  could be calculated as:

$$H_{xy} = W_1 * H_{xy}^{dc} + W_2 * H_{xy}^{bc} + W_3 * H_{xy}^{rc} \quad (8)$$

Where,  $W_1$ ,  $W_2$ ,  $W_3$  are the weights of the three indicators which calculate by the entropy method [19].

### 3 Case study

We chose Information science & Library science (LIS) as the major discipline and other six disciplines to test our framework, i.e., “Education & Educational research”, “Computer science, Information systems”, “Management”, “Economics”, “Mathematics, Applied”, and “Psychology, Applied”. Because LIS combines basic research, like mathematics, computer, and physics, with the real-world needs of social sciences.

#### 3.1 Data acquisition and preprocessing

The research papers and references of seven disciplines from Web of Science (WOS) in the year of 2019 are the data in this study. Search strategies include “WC=Information Science & Library Science”, “WC=Education & Educational Research”, “WC=Computer Science, Information Systems”, “WC=Management”, “WC=Economics”, “WC=Mathematics, Applied”, “WC=Psychology, Applied”. The SCI-EXPANDED, SSCI in the Web of Science selects and makes use of subject categories in the Web of Science via these search strategies, which selected the article type to retrieve the articles in English. We retrieved 127235 papers and 1505717 references in total (Table 1). We download the full records and references of each discipline.

**Table 1.** Number of papers and references of seven disciplines

| Subject Category                            | Papers | References |
|---|--------|------------|
| Information science & Library science (LIS) | 4423   | 75908      |
| Education & Educational Research (Edu)      | 15590  | 167748     |
| Computer Science, Information Systems (Com) | 35308  | 369168     |
| Management (Mag)                            | 13941  | 231237     |
| Economics (Eco)                             | 24158  | 279902     |
| Mathematics, Applied (Mat)                  | 29042  | 274609     |
| Psychology, Applied (Psy)                   | 4773   | 107145     |
| Total                                       | 127235 | 1505717    |

Then, We use Python to download the journals provided by JCR and their subject category information to construct the journal-Subject Category comparison table.

Then, using the full journal title obtained from JCR database by python, the abbreviation-full journal title comparison table is established. Finally, the abbreviation-full journal title -Subject Category comparison table is obtained, including 11375 journals included in WoS, and finally 17961 journal subject category

mapping results are obtained, that is, an average journal corresponds to 1.58 subject categories. The partial results is shown as Table 2.

**Table 2.** the abbreviation-full journal title -Subject Category comparison table (Partial)

| the abbreviation     | full journal title                              | Subject Category                      |
|----------------------|---|---------------------------------------|
| INT J INFORM MANAGE  | INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT | Information science & Library science |
| EDUC PSYCHOL         | EDUCATIONAL PSYCHOLOGIST                        | Education &Educational Research       |
| COMPUT NETW          | Computer Networks                               | Computer Science, Information Systems |
| ACAD MANAG ANN       | Academy of Management Annals                    | Management                            |
| Q J ECON             | QUARTERLY JOURNAL OF ECONOMICS                  | Economics                             |
| APPL MATH LETT       | APPLIED MATHEMATICS LETTERS                     | Mathematics, Applied                  |
| J OCCUP HEALTH PSYCH | Journal of Occupational Health Psychology       | Psychology, Applied                   |

The NLP process retrieved 13186 terms from the titles and abstracts of the papers. After term clumping [12], 12298 distinct terms remained.

### 3.2 Interdisciplinary interactions measurement based on citation analysis

Table 3 shows the specific data of direct citation and bibliographic coupling of six disciplines with Information Science & Library Science. Follow the design in Section 2.2, the strength of interdisciplinary interactions based on both direct citation and bibliographic coupling could be generated in Table 4.

**Table 3.** Citation relationship of six disciplines with Information Science & Library Science

| Subject Category                            | Direct Citation | Bibliographic Coupling |
|---|-----------------|------------------------|
| Education &Educational Research (Edu)       | 2190            | 13467                  |
| Computer Science, Information Systems (Com) | 4784            | 47752                  |
| Management (Mag)                            | 5273            | 38897                  |
| Economics (Eco)                             | 1581            | 14926                  |
| Mathematics, Applied (Mat)                  | 219             | 1502                   |
| Psychology, Applied (Psy)                   | 738             | 11041                  |
| Total                                       | 14785           | 127585                 |

**Table 4.** Strength of interdisciplinary interactions between six disciplines and LIS

| Subject Category | Based on direct citation (%) | Based on bibliographic coupling (%) |
|------------------|------------------------------|-------------------------------------|
|------------------|------------------------------|-------------------------------------|

|   |        |         |
|---|--------|---------|
| Education & Educational Research (Edu)      | 0.9070 | 5.8504  |
| Computer Science, Information Systems (Com) | 1.0866 | 12.0184 |
| Management (Mag)                            | 1.7468 | 14.5004 |
| Economics (Eco)                             | 0.4463 | 4.3786  |
| Mathematics, Applied (Mat)                  | 0.0625 | 0.4304  |
| Psychology, Applied (Psy)                   | 0.4048 | 6.4187  |

### 3.3 Interdisciplinary interactions measurement based on research content

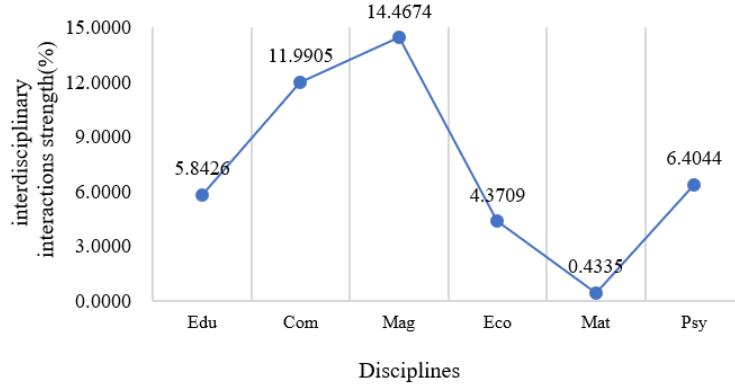
Follow Section 2.3, discipline-keyword matrix of each discipline was generated by LDA model, which includes 7 subjects and 12298 keywords. Then, Word2vec model was applied to map keywords into dense word vectors to capture semantic information of keywords. Since higher dimensions have been shown to capture better semantics [20], we set the number of dimensions for the vectors to 450, and the keywords of 7 disciplines were converted into semantic-level vectors by the trained model.

Furthermore, discipline-keyword vector could be generated following formula (3). According to Euclidean distance, we could finally generate the interdisciplinary interactions strength between six disciplines and Information Science & Library Science, and the results are shown in Table 5.

**Table 5.** Strength of interdisciplinary interactions between six disciplines and Information

| Subject Category                            | Based on research content (%) |
|---|-------------------------------|
| Education & Educational Research (Edu)      | 28.3940                       |
| Computer Science, Information Systems (Com) | 19.8481                       |
| Management (Mag)                            | 21.7641                       |
| Economics (Eco)                             | 16.5775                       |
| Mathematics, Applied (Mat)                  | 16.9179                       |
| Psychology, Applied (Psy)                   | 14.7513                       |

Z-score method was used to standardize three indicators, and entropy method was applied to calculate the index weight. Finally, the strength of interdisciplinary interactions between six disciplines and Information Science & Library Science was obtained, as shown in Fig. 3.



**Fig. 3.** Final results of the strength of interdisciplinary interactions between six disciplines and LIS

There are some observations based on above analysis results:

- 1) The Strengths of interdisciplinary interactions between LIS and other six disciplines are between 0.4335% and 14.4674%. It shows that the diversity of knowledge sources of LIS is not high, which is consistent with the conclusions of Shao et al. [21].
- 2) Among six disciplines, LIS focus more on similar disciplines for interdisciplinary interactions with Computer science & information system and Management. Li et al. proved that the speed of knowledge diffusion between LIS and Management shows a continuous growth trend [22]; Shi et al. stated that both LIS and Computer Science, Information Systems involve information science, especially in system design, technology research, and algorithm optimization [23].
- 3) The strength of interdisciplinary interactions of LIS and Mathematics, Applied is very low. Because Mathematics, Applied is a very professional discipline, while LIS only uses mathematical knowledge in scientific metrology, information retrieval and other research branches.

### 3.4 Validation

We conducted validation to prove the accuracy of our model: the comparison with the mainstream interdisciplinary index. The main indicators include Salton coefficient [24], Rao-Striling coefficient [25] and ID value [3]. The results are shown in Table 6.

**Table 6.** Comparison of the results between our method and mainstream indicators

|              | Edu      |       | Com      |       | Mag      |       | Eco      |       | Mat     |       | Psy      |       |
|--------------|----------|-------|----------|-------|----------|-------|----------|-------|---------|-------|----------|-------|
|              | Value    | #Rank | Value    | #Rank | Value    | #Rank | Value    | #Rank | Value   | #Rank | Value    | #Rank |
| Our method   | 5.8426%  | #4    | 11.9905% | #2    | 14.4674% | #1    | 4.3709%  | #5    | 0.4335% | #6    | 6.4044%  | #3    |
| Salton       | 11.9343% | #4    | 28.5257% | #2    | 29.3592% | #1    | 10.2399% | #5    | 1.0403% | #6    | 12.2427% | #3    |
| Rao-striling | 0.4215%  | #3    | 0.4376%  | #2    | 0.5221%  | #1    | 0.0032%  | #5    | 0.0021% | #6    | 0.0042%  | #4    |

| ID | 3134 | #4 | 121 | #2 | 54 | #1 | 10043 | #5 | 103092 | #6 | 2037 | #3 |
|----|------|----|-----|----|----|----|-------|----|--------|----|------|----|
|----|------|----|-----|----|----|----|-------|----|--------|----|------|----|

It can be seen that: 1) The ranking of strength of interdisciplinary interactions calculated by our method is similar to that of other mainstream methods, which proves the effectiveness of this method. 2) There is a large gap between the minimum and maximum of the strength of interdisciplinary interactions calculated by other mainstream indicators. For example, in the calculation results of Salton coefficient, the maximum value is 35.8766%, and the minimum value is only 1.1561%. The strength of interdisciplinary interactions between LIS and Computer Science, Information Systems is too high and Inconsistent with the actual situation. And the ID value is too big to understand. 3) The difference between the results calculated by other mainstream indicators is not obvious. For example, in the Rao-Striling calculation results, the strength of interdisciplinary interactions between LIS and Edu is 0.4215%, and is 0.4376% between IS and Com. The strength of intersection between LIS and these two subjects are too close. It shows the superiority of our method.

Therefore, compared with other models, the model proposed in this paper is more realistic, and distinguishable, which performs well.

## 4 Conclusion

In this paper, we propose a measurement model of interdisciplinary interactions strength between two specific disciplines, which takes the reference relations between disciplines and the semantic relations of research contents into account. For semantic analysis, the combination of word2vec and LDA can build a more multi-dimensional discipline- keyword vectors, which could accurately explore the similarity of research content between two disciplines.

We believe our method which integrating semantic analysis into citation analysis not only shows a fresh perspective and thought for measuring interdisciplinary interactions, but also other quantitative bibliometric problems. In addition, the method could be applied to disclose the dynamics of interdisciplinary research on a larger sample of disciplines.

Several future directions of research would address the limitations of this study. First, we only selected the data from Web of Science database in the year of 2019, which may not truly reflect the relations of two disciplines based on one year data. Second, this paper only considers the simple citation relationship of interdisciplinary references, ignoring the relevance of citation content.

In future research, we can combine text analysis method with citation content to explore a deeper interdisciplinary relationship.

## Acknowledgements

This work was supported by the National Science Foundation of China Funds [Grant No. 71774013] and the Australian Research Council under Discovery Early Career Researcher Award DE190100994.

## References

1. Porter, A. L., Cohen, A. S., Roessner, J. D., & Perreault, M.: Measuring researcher interdisciplinarity. *Scientometrics*, 72(1), 117-147 (2007).
2. Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S.: Interdisciplinary topics of information science: a study based on the terms interdisciplinarity index series. *Scientometrics*, 106: 583-601 (2016).
3. Zhang L, Rousseau R, Glänzel W.: Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology*, 67(5): 1257-1265 (2016).
4. Karunan, K., Lathabai, H. H., & Prabhakaran, T.: Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics*, 113, 335-367 (2017).
5. Chi, R., & Young, J.: The interdisciplinary structure of research on intercultural relations: a co-citation network analysis study. *Scientometrics*, 96(1), 147-171 (2013).
6. Bjorn, H.: Interdisciplinarity and the intellectual base of literature studies: Citation analysis of highly cited monographs. *Scientometrics*, 86(3), 705–725 (2010).
7. Zhang, H. L., Wei, J. X., Du, Z. D., Liu, X., Yan, S., Feng, Z., et al.: Interdisciplinary research based on social complex network. *Journal of Intelligence*, 30(10), 25–29 (2011).
8. Porter, A. L., & Rafols, I.: Is science becoming more interdisciplinary? measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719 (2009).
9. Rafols, I., & Meyer, M.: How cross-disciplinary is bionanotechnology? explorations in the specialty of molecular motors. *Scientometrics*, 70(3), 633-650 (2007).
10. Xu, S., Lu, C., & Zhang, C.: Measurement of Interdisciplinary Research from the Perspective of Terminology Citation: Six Disciplines on PLOS ONE. *Journal of the China Society for Scientific and Technical Information*, 36(8): 809-820 (2017).
11. Wang, Z., Ma, L., & Zhang, Y.: A Hybrid Document Feature Extraction Method Using Latent Dirichlet Allocation and Word2Vec. 2016 IEEE First International Conference on Data Science in Cyberspace (DSC). IEEE (2016).
12. Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C.: “Term clumping” for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change*, 85, 26-39 (2014).
13. Ma, R., Yan X., & Shen, N.: Direct Measurement of the Degree of Interdisciplinarity. *Journal of the China Society for Scientific and Technical Information*, 38(7), 688-696 (2019).
14. Leydesdorff, L.: On the normalization and visualization of author co-citation data salton's cosine versus the jaccard index. *Journal of the American Society for Information Science & Technology*, 59(1), 77-85 (2010).
15. Wang, L., Notten, A., & Surpatean, A.: Interdisciplinarity of nano research fields: a keyword mining approach. *Scientometrics*, 94(3), 877-892 (2013).
16. Blei, D. M., Ng, A., & Jordan, M. I.: Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3(4-5), 993-1022 (2003).
17. Mikolov, T., Chen, K., Corrado, G., & Dean, J.: Efficient Estimation of Word Representations in Vector Space. *Computer Science*. arXiv preprint arXiv:1301.3781 (2013).
18. Zhang, C., Huang, C., & Yu, L.: Camel: Content-Aware and Meta-path Augmented Metric Learning for Author Identification. *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, 709-718 (2018).
19. Wang, F., & Li, H.: On the use of the maximum entropy method for reliability evaluation involving stochastic process modeling. *Structural Safety*, 88, 102028 (2021).

20. Wang, Y., Liu, Z., & Sun, M.: Incorporating linguistic knowledge for learning distributed word representations. *PLOS ONE*, 10 (2015).
21. Shao, R., Li, L., & Liu, M.: Research on Relationship Between Interdisciplinary Degree and Academic Impact of Papers —— Taking the Library and Information Science (LIS) as an Example. *Journal of Intelligence*, 37(3), 146-151 (2018).
22. Li, L., Li, X., Liu, C., & Zhao, S.: Research on Trade Dynamic Impact and Diffusion Model of Cross Disciplinary Knowledge: A Case Study of Library and Information Science and Management. *Journal of Intelligence*, 36(02), 182-186+158 (2017).
23. Shi, S., Li, X., Song, C., & Xie, R.: Interdisciplinary Knowledge Exchange Based on CTM: Taking Information Science & Library Science (ISLS) and Computer Information System (CIS) as Examples. *Information Studies: Theory & Application*, 41(7), 99-104 (2018).
24. Salton, G., McGill, M. J.: Introduction to modern information retrieval. New York: McGraw-Hill Book Co (1983).
25. Stirling, A.: A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15): 707-719 (2007).