

CollabXplorer: Interdisciplinary Collaboration Discovery and Recommendation

Yitong Hu, Yizhe Wang, Zixuan Zhu, Junxiang Wang, Yi Wang, Gaowei Zhang, Jingping Hou

Beijing University of Posts and Telecommunications, Beijing, China

{huyt, anonymity, ternura, 1034696975, yiwang, zhanggaowei, houjingping}@bupt.edu.cn

Abstract—The paper introduces CollabXplorer, a system designed to facilitate interdisciplinary collaborations and the discovery of potential collaborators. Global semantic information and social influence are utilized to create a scholar embedding model, which converts academic information into vector representations to identify potential collaborators from various scholarly fields. To suggest suitable collaborators, the system applies a recommendation algorithm based on user prompts and scholars' influence. User-friendly visual interface aids in presenting the system's outcomes. We also conducted evaluation and testing experiments, which show that the system provides a good scholar vector representation and high recommendation accuracy, thus, making it suitable for large-scale applications.

Index Terms—CollabXplorer, interdisciplinary collaboration, scholar embedding, collaborator recommendation, information visualization, user interface design.

I. INTRODUCTION

Interdisciplinary collaboration has become a crucial aspect of academic research in recent years [1], [2]. With globalization and advancements in information technology, more innovative and comprehensive research findings are created by scholars collaborating across different fields [3], [4], [5].

The convergence of diverse fields offers a promising avenue for innovation and scientific progress, where transformations and breakthroughs are highly plausible. To facilitate this advancement, it is essential for scholars from various domains to work together collaboratively [6]. However, to achieve effective interdisciplinary collaboration, a systematic approach is necessary to identify and propose prospective interdisciplinary collaborators. This entails identifying and assessing prevailing interdisciplinary networks and suggesting appropriate collaborators based on the unique characteristics and qualities of scholars to inspire and promote efficient interdisciplinary collaboration [7].

Although interdisciplinary collaboration has numerous benefits, it also poses several challenges, such as subject specialization, knowledge boundaries, communication barriers, and data limitations [8], [9]. Therefore, there is a pressing need to develop interdisciplinary collaborative discovery and recommendation systems that offer high accuracy and real-time performance. Such systems can help to facilitate efficient knowledge exchange among scholars from different fields.

Traditional methods of discovering and recommending collaborative opportunities among scholars mainly rely on the direction and substance of their studies within a specific given field [6], making it challenging to identify, making it challeng-

ing to leverage the potential of scholars from different subject areas for collaborative purposes. This approach frequently depends on various data, such as research areas, academic titles, academic accomplishments, work experiences, and other relevant details that are disseminated across different sources, including the researchers' institutions, personal web pages, and multiple academic databases. These data necessitate significant manual annotation and routine maintenance, leading to challenges in ensuring data integrity and real-time performance. Additionally, integrating this data is a major challenge due to the disparate data formats of different institutions and scholar databases, which are often not entirely accessible to the public.

Additionally, certain approaches [10], [11], [12] rely on multimodal data, including courseware and classroom recording data, which are prohibitively expensive and not widely accessible. Consequently, conventional methods of discovering and recommending collaborative opportunities among scholars fall short in promoting an in-depth examination of collaborative relationships among scholars and fail to address the challenges introduced by the swift advancements in interdisciplinary collaboration.

To tackle the aforementioned challenges, this paper presents CollabXplorer, a system designed to facilitate interdisciplinary collaboration discovery and recommendation. It is imperative to evaluate the research directions of scholars appropriately to recommend suitable interdisciplinary collaborators. Hence, this paper emphasizes the quantification and representation of scholars by embedding them into a lower dimensional continuous vector space, which is achieved by using a novel scholar embedding model that acquires vectors for author representation using global semantic information and social influence. The semantic input pertains to the abstracts of all papers of the scholars and represents their academic achievements in various disciplines in a fine-grained semantic dimension that cuts across interdisciplinary barriers. On the other hand, the social influence input reflects a comprehensive measure of scholars' expertise and collaboration in interdisciplinary subjects within their respective disciplines. Both of these inputs are computed based on abstracts, number, and citations of papers in real-time, enabling us to monitor the research dynamics of scholars in real-time.

After successfully obtaining the scholar vector, CollabXplorer recommends appropriate interdisciplinary collaborators to users according to their research needs and preferences. The algorithm behind the recommendation engine lists candidates

by considering the suitability among scholars, the social influence of collaborators, and users' input prompt. Users can interact with the system through a user-friendly interface to explore and connect with collaborators in their respective research fields. To ensure optimal performance and stability, the system is supported by a multi-layered architecture and contemporary software technologies. The design, implementation, and experimental evaluation of CollabXplorer are described in detail in this paper. We invited experts and scholars from the Beijing University of Posts and Telecommunications (BUPT) to test and evaluate this tool, which successfully confirmed its high accuracy and effectiveness.

Overall, this paper proposes an effective approach for discovering and taking advantage of interdisciplinary collaboration opportunities. The methods and algorithms of CollabXplorer offer practical and efficient tools to encourage interdisciplinary cooperation in academia and foster creative research outcomes.

II. RELATED WORK

Recommendation systems have garnered considerable attention from researchers and have proven to be useful in a variety of fields. Similarly, recommendation systems are also crucial in the field of interdisciplinary collaboration.

We have seen different solutions to the problem of interdisciplinary collaboration. Among them, [10] focus on recommending the most suitable interdisciplinary major, simplifying and correcting the work of decision-makers, and promoting interdisciplinary major reform. However, the practical problem, in this case, is that there is no more consideration of the specific cooperation needs of scholars in these fields, and blind field crossover may lead to an insufficient collaboration of scholars in scientific research and a waste of scientific research resources. On the other hand, the system requires multimodal data input of papers, and course recordings, and requires a large amount of manually labelled data, which is not suitable for large-scale application in global institutions and subject fields.

In addition, according to [11], scholars in the same field are recommended based on author and relevant publication information. The rankings are determined by evaluating factors such as paper content, journal quality, and others. Evaluation scores are then generated, and a scholar similarity is calculated after weight distribution to perform scholar matching. However, this approach has a significant limitation for interdisciplinary collaboration because it only provides recommendations for scholars in similar fields, thus lacking the potential for collaboration across fields.

The core advantage of CollabXplorer is analyzing the entire publications of scholars at the granularity of words and considering the needs and preferences of users when recommending collaborators. These two innovations make CollabXplorer more applicable to a wide range of disciplines and better serve the needs of interdisciplinary cooperation between scholars.

The contributions of this paper are summarized as follows:

- CollabXplorer focuses on the summarization of scholars' all publications and considers the social influence and collaboration network of scholars to construct a more comprehensive vector representation.
- CollabXplorer doesn't rely on scattered structured data about academic fields, job titles, academic achievements, work histories, etc., for feature extraction because these data require extensive manual maintenance, are difficult to integrate, and their integrity and real-time performance cannot be guaranteed. Instead, CollabXplorer directly embeds scholars into a low-dimensional vector space based on the content of their publication for subsequent analysis.
- CollabXplorer's embedded model can be used directly for collaborator recommendation without constructing scholarly evaluation and recommendation methods for specific subject areas.
- CollabXplorer also doesn't need a lot of manually annotated data to train the recommendation mode. CollabXplorer only requires users to input the prompt text, and then it recommends the (interdisciplinary) cooperators that match the description and have a high fit for the user.

III. COLLABXPLOER

This section describes the system design, implementation, and interaction design of CollabXplorer.

The part of system design and implementation focuses on the system architecture, data sources, scholar embedding model, scholar recommendation algorithm and the implementation of key components of CollabXplorer. The user interaction experience part mainly covers the interface design, function and user experience of the system.

The main goal of the system is to find meaningful interdisciplinary cooperation potential from a large number of unstructured academic data and present it to users in a visual way through the recommendation algorithm.

A. System Design and Implementation

The architecture of this system, shown in Figure 1, mainly includes the following layers: infrastructure layer, data layer, algorithm layer and application layer. The data layer mainly includes data storage, data acquisition and data preprocessing. The algorithm layer covers feature engineering, recommendation engine and visualization of the collaboration network for this task. The application layer contains system functions and user interfaces such as account system, user preference, academic/recommendation information statistics, prompt text input, collaborator recommendation, cooperation network interaction and site interaction.

Regarding the infrastructure, we use Docker container technology on the Alibaba Cloud Elastic Computing Service platform to package each system component into images for easy deployment and management. And the Kubernetes cluster management tool is used to realize the system's automatic deployment, maintenance and expansion.

The system adopts the development mode of front-end and back-end separation, in which the front-end is developed by the Vue.js framework, and the back-end is implemented by Python's Flask framework. A typical system workflow for a recommendation task based on user prompts can be summarized as follows.

- Front-end:
 - Implement user login, registration, authentication, etc. by calling the backend API.
 - Pass the user input to the backend via the backend API.
 - Call the backend API, pass the user input prompt as a parameter to the backend, and block the process until the backend returns the result.
- Backend:
 - Compute vectors for all scholars within all databases based on the user's prompt text and recommendation preferences.
 - Compute the suitability of all scholars concerning the current user and sort the recommended top-ranked scholars in descending order.
 - Wrap the recommendation in JSON format and return it to the front end, persisting the record.
- Front-end:
 - Receive and parse JSON data, refresh page recommendation statistics and list of recommendations.
 - Render interdisciplinary collaboration networks locally.
 - Listen for user interactions, such as clicking on scholars in the recommendation list, clicking on nodes in the interdisciplinary collaboration network, etc., and calling back-end APIs to persist user behaviour records.

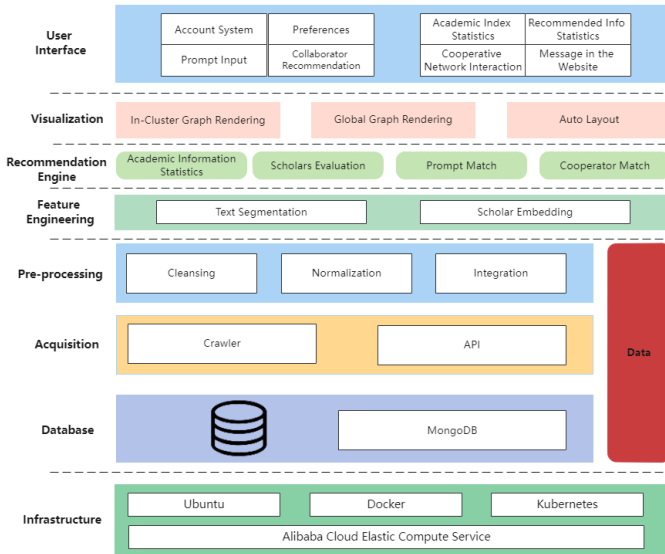


Fig. 1. System structure

1) *Data Layer:* The academic data of scholars is an important consideration when designing this system because the quality of data is the key to the recommendation algorithm. Figure 2 describes the design and implementation of the data layer.

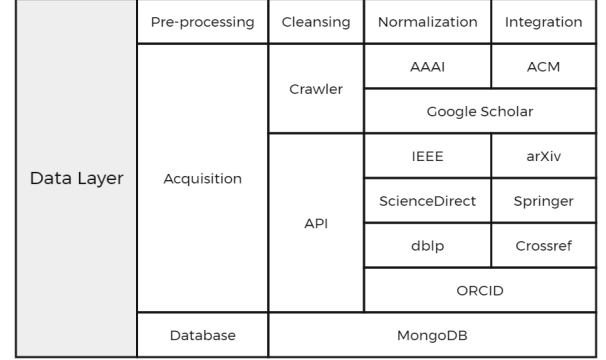


Fig. 2. Design and implementation of the data layer

a) *Data Acquisition:* We use ORCID as the unique identification of the scholar. ORCID is an open, free, and growing scholar identification system that provides each scholar with a unique, persistent, and verifiable scholar identifier. However, ORCID does not directly provide metadata of papers, such as abstracts, keywords, etc., so we need to obtain academic information of scholars through other channels. Crossref provides metadata for most of the world's scholarly publications. According to the API provided by Crossref, using the scholar's ORCID as the parameter query, we can obtain the metadata of all the papers of the scholar, including all the authors, titles, abstracts, keywords and other information. For scholars who are not queried on ORCID, our fallback databases are dblp and Google Scholar. By querying the name of the scholar, we can obtain the URL of each paper of that scholar. We wrote the corresponding web crawler or used the API provided by the publisher to obtain the metadata of the paper based on the URL.

We choose a more flexible non-relational database MongoDB as data storage. MongoDB is a database based on distributed file storage. Its data model is document-oriented, and its data is stored in JSON form, so it is very suitable for storing the academic information of scholars.

b) *Data Preprocessing:* In the next step, we need to preprocess the academic information of the scholars to ensure the quality of the data. Our preprocessing mainly consists of the following steps:

1. Data cleaning: Remove outliers such as missing abstracts, keywords, co-authors, etc.
2. Data Normalization: Remove stop words. Stop words are words that have no real meaning in natural language processing due to their frequent occurrence. For example: 'the', 'a', 'an', 'of', 'in', 'on', etc. These words

occur frequently in the text but do not contribute much to the meaning of the text, so we need to remove these stop words. We used a list of stop words from the NLTK library to normalize the data.

3. Data integration: The academic information of scholars from different data sources will be integrated, including conflict merging and deduplication. We set a timed task to automatically update academic information of scholars from data sources such as Crossref, dblp and Google Scholar every day to ensure the real-time nature of the data.

2) *Embedding*: To extract all the features of a scholar and then perform scholar recommendation, it is necessary to convert the scholar's academic information into a vector, and we call this process scholar embedding. Figure 3 describes the workflow of scholar embedding.

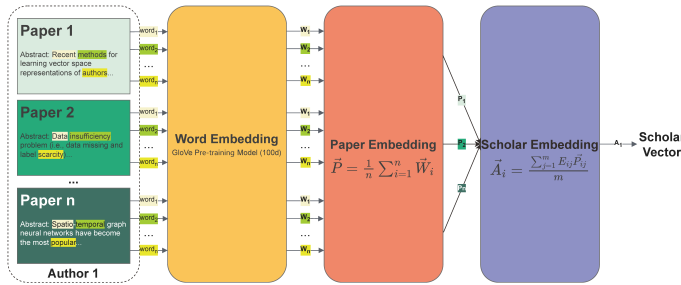


Fig. 3. Scholar embedding model

Word embedding, also known as distributed representation, is a widely used technique in natural language processing (NLP) that enables machines to capture the meaning and context of words. global Vectors for Word Representation (GloVe) is one such approach that is popular for its ability to produce high-quality word embeddings.

Pennington et al. [13] proposed the GloVe, a method for training word embeddings based on the global co-occurrence statistics of words in a corpus. The method consists of decomposing the word co-occurrence count matrix using singular value decomposition (SVD) and learning word vectors that capture statistical relationships between words.

The key formulation used in GloVe is the objective function:

$$J = \sum_{i,j=1}^V f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log(X_{ij}))^2 \quad (1)$$

where V is the vocabulary of words, X_{ij} is the number of co-occurrence of words i and j , w_i and w_j are word vectors, b_i and b_j are biased words, and $f(X_{ij})$ is a weighting function that assigns smaller weights to rare word pairs. The goal is to minimize J by adjusting the word vectors and biases. Another important formula in word embedding is cosine similarity:

$$\text{similarity}(w_i, w_j) = \cos(\theta) = \frac{w_i^T \cdot w_j}{|w_i| \cdot |w_j|} \quad (2)$$

It measures the similarity between word vectors w_i and w_j by calculating the cosine of the angle between them. High cosine similarity values indicate that words are semantically similar and that cosine similarity is superior when performing similarity calculations for searches.

When applying GloVe to the problem of this article, i.e., constructing a semantic and combined influence-based cooperative network, the corresponding principle is to transform the abstracts of scholars' papers into vectors using GloVe's pre-trained model, a process called paper embedding. Then, based on the combined influence of each paper, a weighted average operation is performed on these paper vectors to obtain the scholar vectors, a process called scholar embedding.

a) *Paper Embedding*: We define the vector $\vec{P}$ of the paper:

$$\vec{P} = \frac{1}{n} \sum_{i=1}^n \vec{W}_i \quad (3)$$

where $\vec{W}_i$ is the word vector of the i th word in the abstract of the paper and n is the total number of words in the abstract of the paper. Here we use the mean value to represent the thesis vector, which is because we believe that the topic of the paper should be determined by all the words in the abstract together.

For the word vector $\vec{W}$, we use the NLTK library to perform text segmentation on the abstracts of papers in the database and then map the resulting words, i.e., tokens, into GloVe's pre-trained model to obtain a 100-dimensional word vector for each word. In this paper, we choose GloVe's pre-trained model (6B tokens, 400K vocab, uncased, 100d vectors), which is trained on the Wikipedia corpus.

b) *Scholar Embedding*: First, we define some notation. x_{ij} represents the total number of co-authors in paper j of scholar i . In general, for a paper, the higher the author's ranking, the greater his contribution to that paper. Therefore, we can assume that the scholar's contribution is non-uniformly proportional to his ranking in that paper, using an exponential function e in $[0, 1]$ to represent.

Based on the assumptions above, the contribution degree E_{Rij} of scholar i to paper j is expressed as follows:

$$E_{Rij} = \begin{cases} e, & k = 1 \\ e^{-\frac{x_{ij}-k}{x_{ij}}}, & \text{others} \end{cases} \quad (4)$$

where k denotes the rank of the scholar among all co-authors in the paper.

In addition, the number of citations in the paper reflects the influence of the paper, and usually, the more citations, the higher the influence of the paper. Therefore, we define the impact factor C_{ij} of scholar i 's paper j as follows:

$$C_{ij} = \begin{cases} 0, & n_{ij} = 0 \\ 1, & n_{ij} > 0 \end{cases} \quad (5)$$

where n_{ij} is the number of times the paper is cited. And combined with the impact factor C_{ij} of the paper, we can get the impact E_{Cij} of paper j , which is defined as follows:

$$E_{Cij} = C_{ij} e^{\frac{n_{ij}}{n_{imax}}} \quad (6)$$

where n_{ij} is the number of citations of paper j , n_{imax} is the highest single citation number of scholar i .

The above two steps present the contribution of scholars to the paper and the influence of the paper, respectively. Next, we improve the above equation based on the idea of PageRank, an algorithm proposed by Sergey Brin and Lawrence Page et al. [14] that takes into account the influence of other factors based on the influence of a single point of the network, and we adapt the model:

$$E_{ij} = \lambda E_{Rij} + (1 - \lambda) E_{Cij} \quad (7)$$

E_{ij} represents the combined influence of paper j , here we consider two factors: the contribution of scholar i to paper j E_{Rij} and the influence of paper j E_{Cij} , where the weight parameter λ determines the influence of both E_{Rij} and E_{Cij} on E_{ij} . with values in the range $[0, 1]$.

With the above steps we obtain the influence E_{ij} of scholar i in paper j , and next multiply it with the vector $\vec{P}_{ij}$ of paper j of scholar i extracted by GloVe and perform the weighted average operation to obtain the vector $\vec{A}_i$ of scholar i , which is defined as follows:

$$\vec{A}_i = \frac{\sum_{j=1}^m E_{ij} \vec{P}_{ij}}{m} \quad (8)$$

where m denotes the total number of papers written by scholar i . By this method, we can obtain a vector representation of each scholar, and we call the above process author embedding.

3) *Collaborator Recommendation*: For a recommendation system, we must consider the scholars' needs. In this paper, we allow the central scholar (for whom the system will recommend collaborators) to enter a prompt to describe the collaborator he expects to seek. Thus, our goal is to find collaborators that match the prompt and have high suitability with the central scholar. In short, we use prompts to modulate the vectors of all scholars in the database, compute the suitability of the new vectors of the central scholar and other scholars, and recommend the top-ranked scholars as collaborators. The workflow of our recommendation engine is shown in Figure 4:

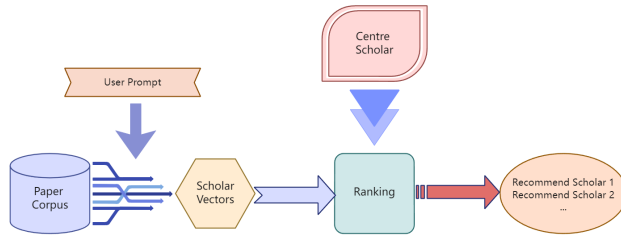


Fig. 4. Collaborator recommendation model

Similarly to the Formula 3, we define the vector form $\vec{p}$ of the user input prompt.

$$\vec{p} = \frac{1}{n} \sum_{i=1}^n \vec{W}_i \quad (9)$$

where $\vec{W}_k$ refers to the GloVe vector representation of the k th word in prompt.

Next, we compute the similarity s_{ij} between the vector $\vec{P}_{ij}$ of scholar i 's paper j and the vector representation $\vec{p}$ of prompt to obtain the similarity s_{ij} between paper j and prompt, which is calculated as follows:

$$s_{ij} = \cos(\vec{p}, \vec{P}_{ij}) \quad (10)$$

Due to $s_{ij} \in [-1, 1]$, $E_{ij} \in [1, e]$, to change the range of s_{ij} from $[-1, 1]$ to $[1, e]$, we modify the formula in the following way:

$$s'_{ij} = \frac{e-1}{2}(s_{ij} + 1) + 1 \quad (11)$$

where s'_{ij} is the transformed value of s_{ij} .

In turn, we obtain the weights W_{ij} of scholar i 's paper j , consisting of the similarity and the combined influence E_{ij} Formula 7 components.

$$W_{ij} = \beta s'_{ij} + (1 - \beta) E_{ij} \quad (12)$$

here we introduce variable coefficients β to weigh the importance of s_{ij} and E_{ij} . The β can vary according to the needs of the central scholar. Specifically, a larger β is taken if the central scholar values the suitability with the collaborator more, and a smaller β is taken if the social influence of the collaborator is more important.

At this point, we have the vector representation of scholar A_i modulated by the prompt:

$$\vec{A}_i = \frac{\sum_{j=1}^m W_{ij} \vec{P}_{ij}}{m} \quad (13)$$

where m denotes the total number of papers written by scholars i .

On this basis, we calculate the suitability S_i of central scholars and scholars A_i .

$$S_i = \cos(\vec{U}, \vec{A}_i) \quad (14)$$

where $\vec{U}$ refers to the vector representation of the central scholars.

We sort $S_1 \sim S_{126}$ in descending order and recommend the top-ranked scholars as potential collaborators of the central scholars.

B. Interaction Design

Users can use this system after authenticating their academic identity and choosing to allow this system to include their academic data. After successful login, the user will enter the dashboard shown in Figure 5, which is divided into three parts: a sidebar console, a display interface and a message notification interface. Here we will mainly introduce the display interface.

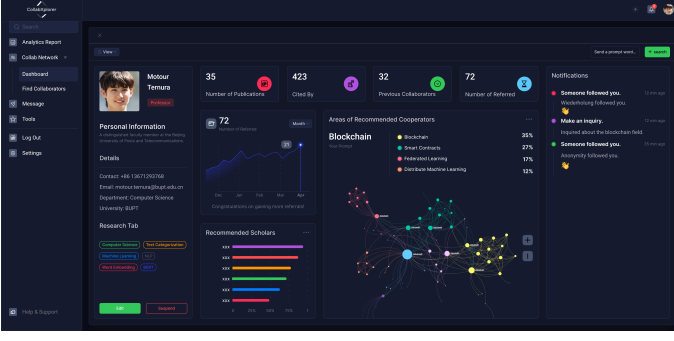


Fig. 5. Dashboard of CollabXplorer

The following are described in order from left to right, top to bottom. The first is the personal information section, where the user's personal information will be displayed, such as name, organization, email, etc., and the subject and specific subdivision tags can be added by the user. The above information can be modified, saved and revoked.

The top four rectangular boxes are the number of papers published by an individual, the number of citations of his papers, the number of scholars he has collaborated with, and the number of times he has been recommended to other scholars by the system. These indicators will help users evaluate their own academic research level and cooperation level in a quantitative form.

Of the four partitions below, the first is the top-left line chart area, which in the current plot refers to the number of times per month you are referred to other scholars when they make a query. The bottom-left bar area refers to the scholars recommended to you based on your prompt, ranked from top to bottom in terms of potential collaboration possibilities.

Through the search box on the top right, users can enter the unstructured prompt text that describes the research content of the collaborators they expect to seek. Based on the prompt text, the system will recommend collaborators with high suitability to the user. It is worth noting that the social influence of collaborators is also one of the recommendation metrics. Users can go to the "Setting" option to adjust their personal preference, that is, suitability first or influence first, corresponding to the size of the parameter β in Equation 12. The system will analyze all the recommended collaborators, classify and colour them according to the research field. Relevant statistics are presented in the "Areas of Recommend Cooperators" section.

The relationship network diagram at the bottom right is the key to our recommendation system. In this diagram, you can visually observe the potential cooperation potential between you and the recommended person. The dots in the graph represent scholars, and their sizes change accordingly with the change in cooperation potential. Wired means that two scholars have previously appeared as co-authors on the same paper; The colour of the points indicates the different fine molecular fields under the current search domain, thus helping the user to make a better choice.

In summary, this system shows the user's personal informa-

tion, the level of academic research, and the functionality of the recommendation system. Through various graphs and visualizations, users can gain insight into their research area, co-authors, and potential collaboration possibilities to better select partners and directions. This system provides a platform for academic researchers to communicate and cooperate, which helps to promote the progress and development of academic research.

IV. EVALUATION

To evaluate our system, we generated scholar documents mentioned in Section III-A1a for 126 scholars from BUPT and wrote them to the database. This work was supported by the BUPT Personnel Office and the library, and all data were manually proofread to ensure the quality of the data. For this system, we designed two experiments, which are used to evaluate the effect of author embedding and recommendation respectively.

A. Embedding Evaluation

To evaluate the effect of author embedding, we conduct experiments on the author similarity task. Pennington [13] et al. evaluated GloVe for word similarity, and accordingly, we design the author similarity task referring to their work, this is because our embedding is based on GloVe pre-trained model. Specifically, we invited experts from the Beijing University of Posts and Telecommunications to evaluate the collaborative relationships of 126 scholars in our database. The cooperative relationship is divided into 5 levels, which are 1, 2, 3, 4 and 5 respectively, and the corresponding author similarity value is 0.2, 0.4, 0.6, 0.8 and 1. Where level 1 indicates no cooperation at all and level 5 indicates close cooperation. We denote these evaluation results as t_{ij} , where i and j denote the two scholars, respectively.

Through author embedding, we can obtain the vector representation of the scholar. We define the similarity between scholars s_{ij} :

$$s_{ij} = \frac{\vec{A}_i \cdot \vec{A}_j}{|\vec{A}_i| \cdot |\vec{A}_j|} \quad (15)$$

Here, $\vec{A}_i$ denotes the vector of scholar i , and $\vec{A}_j$ denotes the vector of scholar j .

According to the above method, the similarity between scholar i and all other scholars $\vec{s}_i$ that the expert evaluation result in $\vec{t}_i$ can be obtained in the same way. We choose the cosine similarity between $\vec{s}_i$ and $\vec{t}_i$ as the evaluation metric for the author similarity task, where closer to 1 indicates that our author embedding is more accurate. We define the accuracy of the author similarity task:

$$Accuracy = \frac{\vec{s}_i \cdot \vec{t}_i}{|\vec{s}_i| \cdot |\vec{t}_i|} \quad (16)$$

1) *Results:* We show the similarity between the two sets of data in Figure 6. We chose to calculate the average accuracy for each of the 10 data points in the steps of 10 scholars. You can see that the average accuracy is roughly (0.75, 0.95). The average accuracy of 126 data points is **0.91**, which indicates that our scholar embedding model can accurately reflect the collaboration between scholars.

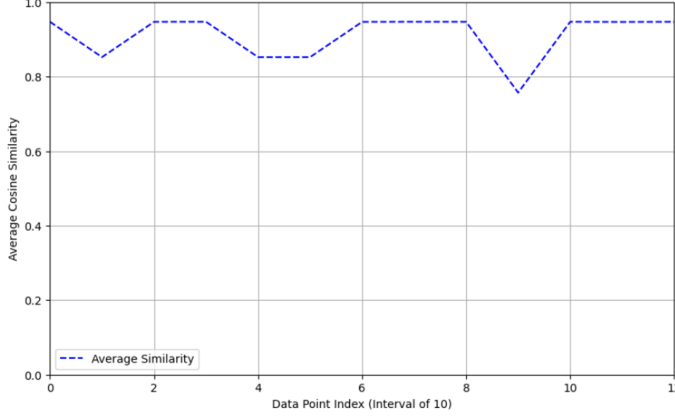


Fig. 6. Accuracy of scholar similarity task

B. Recommendation Evaluation

To evaluate the effectiveness of this recommendation system, we take a manual scoring approach and invite 20 researchers from BUPT to rate the recommendation results.

1) *Experiment Design:* We allowed the 20 raters to freely choose the prompt and designated the central scholar. This allows evaluators to evaluate the accuracy of the recommendation system in their familiar domain, based on their expertise and experience.

According to the central scholar and the prompt of the evaluator, the system recommends the top 5 collaborators, that is, generates a recommendation record. The evaluator assigns satisfaction scores to each of the five collaborators, where the score $s_j \in [0.00, 1.00]$.

To keep our sample diverse, we asked each evaluator to generate and rate five different recommendations, which meant that each evaluator had to rate 25 collaborators. Therefore, we received the scoring results of 100 recommendation records in total, containing satisfaction ratings for 500 collaborators recommended by the system.

In this evaluation, the recommendation preference of the system is suitability first. The Table I shows an example of 5 ratings for recommendation records.

After obtaining the ratings of each rater for each recommendation, we define three evaluation metrics as follows:

$$\begin{aligned} T_1 &= s_1 \\ T_3 &= 0.7s_1 + 0.2s_2 + 0.1s_3 \\ T_5 &= 0.5s_1 + 0.3s_2 + 0.1s_3 + 0.05s_4 + 0.05s_5 \end{aligned} \quad (17)$$

where T_i represents the rating of the first i recommendations, and s_j represents the satisfaction rating of the evaluator for the j recommended collaborator.

2) *Experiment Results:* We use box plots to describe 100 data points (recommendation records) statistically, to show the performance of the recommendation system more intuitively. The horizontal axis of the box plot is the three evaluation metrics, the vertical axis is the rating, the upper and lower boundaries of the box plot are the 75% and 25% quantiles, the middle line is the median, and the upper and lower edges of the box plot are the maximum and minimum values, respectively. See the following Figure 7:

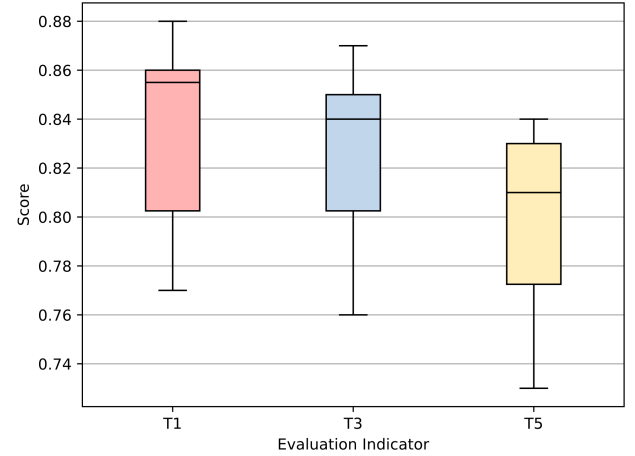


Fig. 7. Scores of the recommended collaborators

The median and mean values of T_1 are 0.85 and 0.84, respectively, which are higher than those of T_3 and T_5 , indicating that the evaluation is generally more satisfied with the first collaborator recommended by the system. This indicates that under the recommendation preference of suitability priority, the fit between the first collaborator recommended by the system and the scholars of the centre can meet the needs of carrying out cooperation.

The gap between the median and mean values of T_3 and T_1 is around 1%, which can be considered as the recommendation quality is similar. On the other hand, T_3 has the smallest variance, indicating that the top 3 recommended by the system are the relatively stable fit between the collaborators and the centre scholars.

T_5 has the lowest median and mean, as well as the largest variance, indicating a decline in recommendation quality and stability. This is also in line with expectations, because among the top 5 recommended collaborators, the fit between the collaborators and the centre scholars is gradually decreasing, and some collaborators may affect the overall recommendation quality of T_5 . However, to ensure the diversity of recommendations, we will try to minimize such quality and stability degradation problems in our future work.

In general, the median and the mean value of $T_1 - T_5$ are all above 0.8, which indicates that the recommendation system

TABLE I
5 RATING EXAMPLES

Recommendation Information			Satisfaction					
Record ID	Center	Scholar	Prompt	1st	2nd	3rd	4th	5th
01	**		Bioscience	54%	91%	39%	89%	36%
02	**		Graph Neural Network	70%	62%	49%	37%	52%
03	**		Beamforming	85%	84%	51%	69%	26%
04	**		Multi-Agent	82%	76%	74%	71%	40%
05	**		Econometrics	95%	71%	92%	96%	30%

designed in this study has good accuracy in recommending coauthors with a high degree of relevant fit, and can recommend valuable coauthors for scholars.

V. CONCLUSION

In summary, this paper presents CollabXplorer, an innovative system for interdisciplinary collaboration discovery and recommendation in the academic community. With the drastic development of globalization and information technology, interdisciplinary collaboration has become essential for producing innovative and comprehensive research outcomes. However, disciplinary specialization, knowledge boundaries, and communication barriers remain challenges in interdisciplinary collaboration. To address these challenges, this paper proposes a systematic approach that includes discovering and evaluating existing interdisciplinary networks, constructing an author embedding model based on global semantic information and social influence, and recommending suitable partners based on scholars' specific descriptions.

The proposed system's architecture comprises infrastructure, data, algorithm, and application layers, using data acquisition sources such as the ORCID system, Crossref, dblp, and Google Scholar databases. To ensure the quality of the data, the system preprocesses academic information by cleaning, normalizing, and integrating data from different sources. Scholar vectors are constructed by analyzing the abstracts of scholars' publications using a scholar embedding method, allowing for a more comprehensive evaluation of scholars from multiple disciplines. The user interface of the system is user-friendly, allowing for easy deployment and management worldwide.

Experiments on author embedding and recommendations demonstrate that the system can accurately reflect the co-operation relationship between scholars and accurately recommend valuable collaborators with a high degree of similarity in research interest. This paper provides a practical and efficient tool for the academic community to promote interdisciplinary collaboration and creative research outcomes. Future work aims to minimize the decline in recommendation quality beyond the top 3 recommended collaborators while maintaining diversity. Overall, the proposed system provides a platform for academic communication and collaboration, helping researchers to find suitable collaboration partners and move forward in their research effectively.

REFERENCES

- [1] N. A. Sultan, B. Mahmood, K. H. Thanoon, and D. S. Khadhim, "Network centralities-based approach for evaluating interdisciplinary collaboration," *2020 6th International Engineering Conference "Sustainable Technology and Development" (IEC)*, pp. 216–221, 2020.
- [2] X. Liu, B. Yi, M. Li, and J. Li, "Is Interdisciplinary Collaboration Research More Disruptive Than Monodisciplinary Research?" *Proceedings of the Association for Information Science and Technology*, vol. 58, no. 1, pp. 264–272, 2021.
- [3] J. N. Cummings and S. Kiesler, "Collaborative research across disciplinary and organizational boundaries," *Social studies of science*, vol. 35, no. 5, pp. 703–722, 2005.
- [4] P. D'este, O. Llopis, F. Rentocchini, and A. Yegros, "The relationship between interdisciplinarity and distinct modes of university-industry interaction," *Research Policy*, vol. 48, no. 9, p. 103799, 2019.
- [5] V. Larivière, S. Haustein, and K. Börner, "Long-distance interdisciplinarity leads to higher scientific impact," *Plos one*, vol. 10, no. 3, 2015.
- [6] R. Moirano, M. A. Sánchez, and L. Štěpánek, "Creative interdisciplinary collaboration: A systematic literature review," *Thinking Skills and Creativity*, vol. 35, Mar. 2020.
- [7] R. Freeth and G. Caniglia, "Learning to collaborate while collaborating: advancing interdisciplinary sustainability research," *Sustainability science*, vol. 15, no. 1, pp. 247–261, 2020.
- [8] Z. J. Ren, "The Rewards and Challenges of Interdisciplinary Collaborations," *iScience*, vol. 20, Oct. 2019.
- [9] A. Van den Beemt, M. MacLeod, J. Van der Veen, A. Van de Ven, S. van Baalen, R. Klaassen, and M. Boon, "Interdisciplinary engineering education: A review of vision, teaching, and support," *Journal of Engineering Education*, vol. 109, no. 3, pp. 508–555, 2020.
- [10] L. Zou, "A decision support method for interdisciplinary major in colleges and universities based on natural language processing and data mining," in *2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)*, 2021, pp. 75–78.
- [11] K. Alinani, G. Wang, A. Alinani, and D. H. Narejo, "Who should be my co-author? recommender system to suggest a list of collaborators," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 2017, pp. 1427–1433.
- [12] C. Yang, T. Liu, L. Liu, and X. Chen, "A nearest neighbor based personal rank algorithm for collaborator recommendation," in *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, 2018, pp. 1–5.
- [13] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.
- [14] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, November 1999.