CollabXplorer: Interdisciplinary Collaboration Discovery and Recommendation

# Abstract

The paper introduces CollabXplorer, a system designed to facilitate the discovery and recommendation of potential interdisciplinary collaborators, as well as the interdisciplinary collaboration networks. Global semantic information and social influence are utilized to create 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 scholar influence. User-friendly visual interface aids in presenting the system's outcomes. Furthermore, we conducted evaluation and testing, which shows that the system provides good vector representation effect and high recommendation accuracy, thus, making it suitable for large-scale applications.

# Keywords

# Introduction

Nowadays, interdisciplinary collaboration has become a crucial aspect of academic research. Due to globalization and advances in information technology, an increasing number of scholars need to collaborate with those from other fields to create more innovative and comprehensive research findings \cite{}.

The convergence of diverse fields offers a promising avenue for innovation and scientific progress, where transformations and breakthroughs are highly plausible. To foster scientific advancement, collaborative efforts between scholars from various domains are indispensable. Consequently, 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\cite{}.

Despite the benefits of interdisciplinary collaboration, it presents various challenges, including subject specialization, knowledge boundaries, communication barriers, and data exchange limitations \cite{}. Thus, the development of interdisciplinary collaborative discovery and recommendation systems with high accuracy and real-time performance is imperative.

Conventional methods of discovering and recommending collaborative opportunities among scholars primarily rely on the direction and substance of their studies within a given field \cite{}, making it challenging to identify and 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 scholar databases. These data necessitate significant manual annotation and routine maintenance and pose challenges for ensuring data integrity and real-time performance. Furthermore, integrating this data is a major challenge since different institutions and scholar databases have disparate data formats, and often not entirely open to the public.

Additionally, certain approaches \cite{} 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 discipline. 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, 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 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.

# Related Work

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

We have seen different solutions to the problem of interdisciplinary collaboration. Among them, \cite{} 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 waste of scientific research resources. On the other hand, the system requires multimodal data input of papers, and courese 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 \cite{}, 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 all scholar publications and considers the social influence and collaboration network of scholars to construct a more comprehensive vector representation.
* CollabXplorer doesn't require scattered structured data about academic fields, job titles, academic achievements, work histories, etc., for feature extraction because these data require a lot of 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 according to 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.

# CollabXplorer

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.

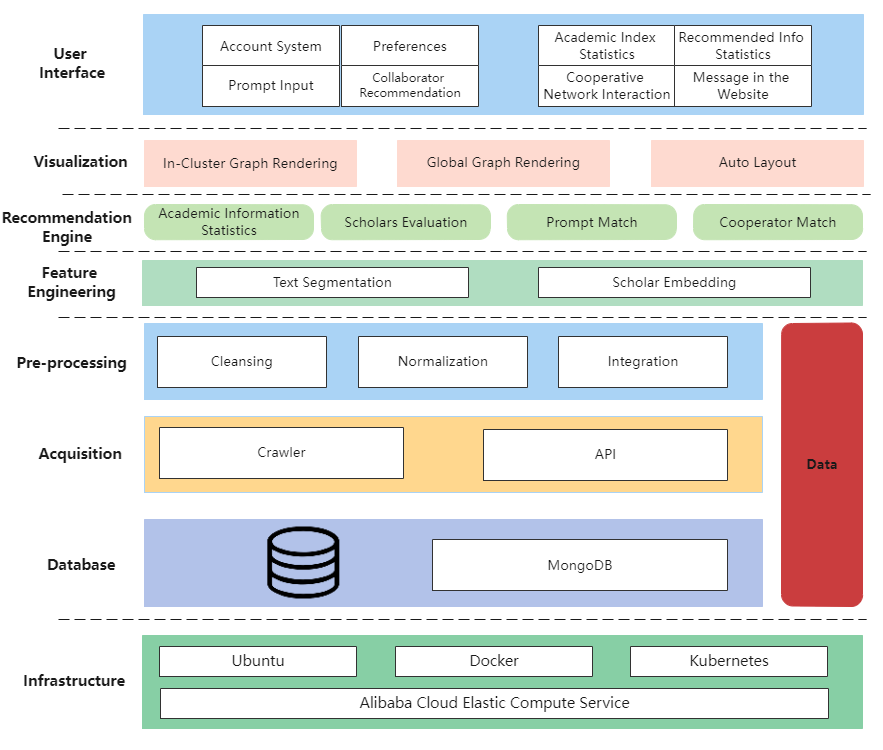
## System Design and Implementation

The architecture of this system, shown in Figure \ref{fig:framework}, 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.

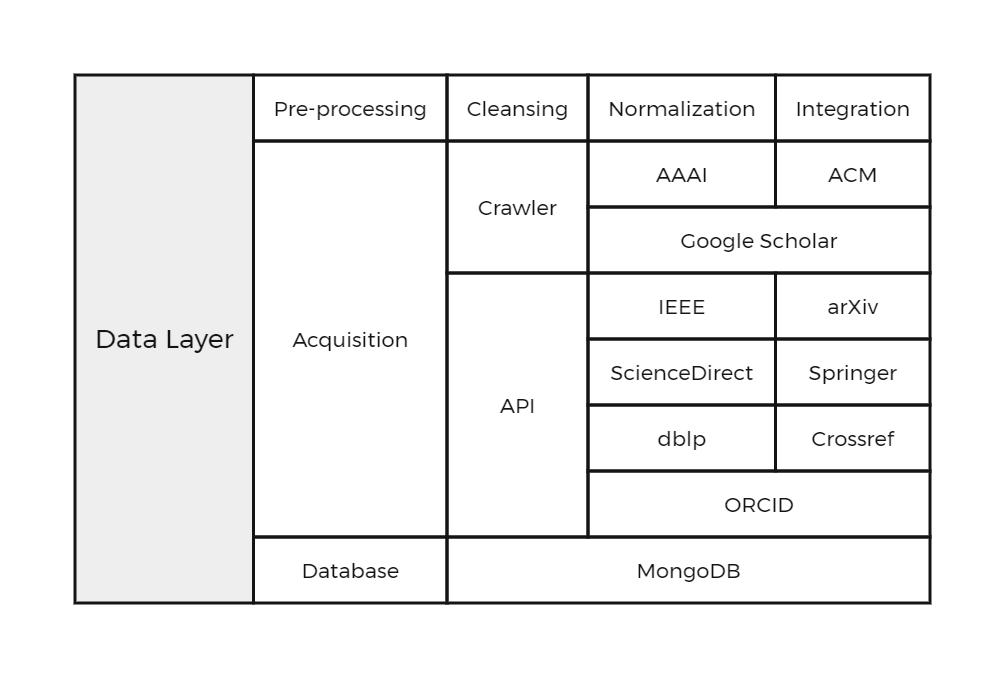
The system adopts the development mode of front-end and back-end separation, in which the front-end is developed by the Vue.js framework, and the back-end is implemented by Python's Flask framework. A typical system workflow for a recommendation task based on user prompts can be summarized as follows.

* 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 call back-end APIs to persist user behavior records.



### 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 \ref{fig:data\_layer} describes the design and implementation of the data layer.



#### 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.

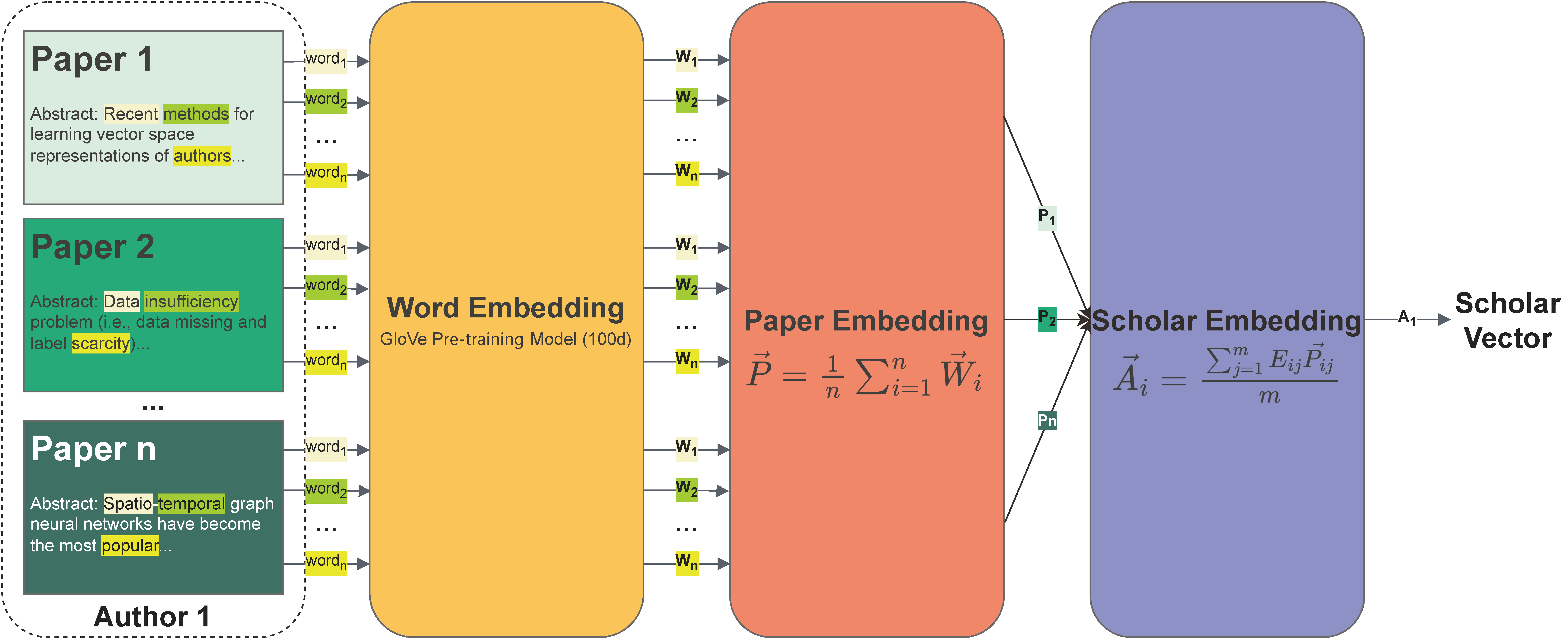
#### Data Preprocessing

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.

### 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 \ref{fig:scholar\_embedding} describes the workflow of scholar embedding.



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. \cite 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:

where is the vocabulary of words, is the number of co-occurrence of words and , and are word vectors, and are biased words, and is a weighting function that assigns smaller weights to rare word pairs. The goal is to minimize by adjusting the word vectors and biases. Another important formula in word embedding is cosine similarity：

It measures the similarity between two word vectors and 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.

#### Paper Embedding

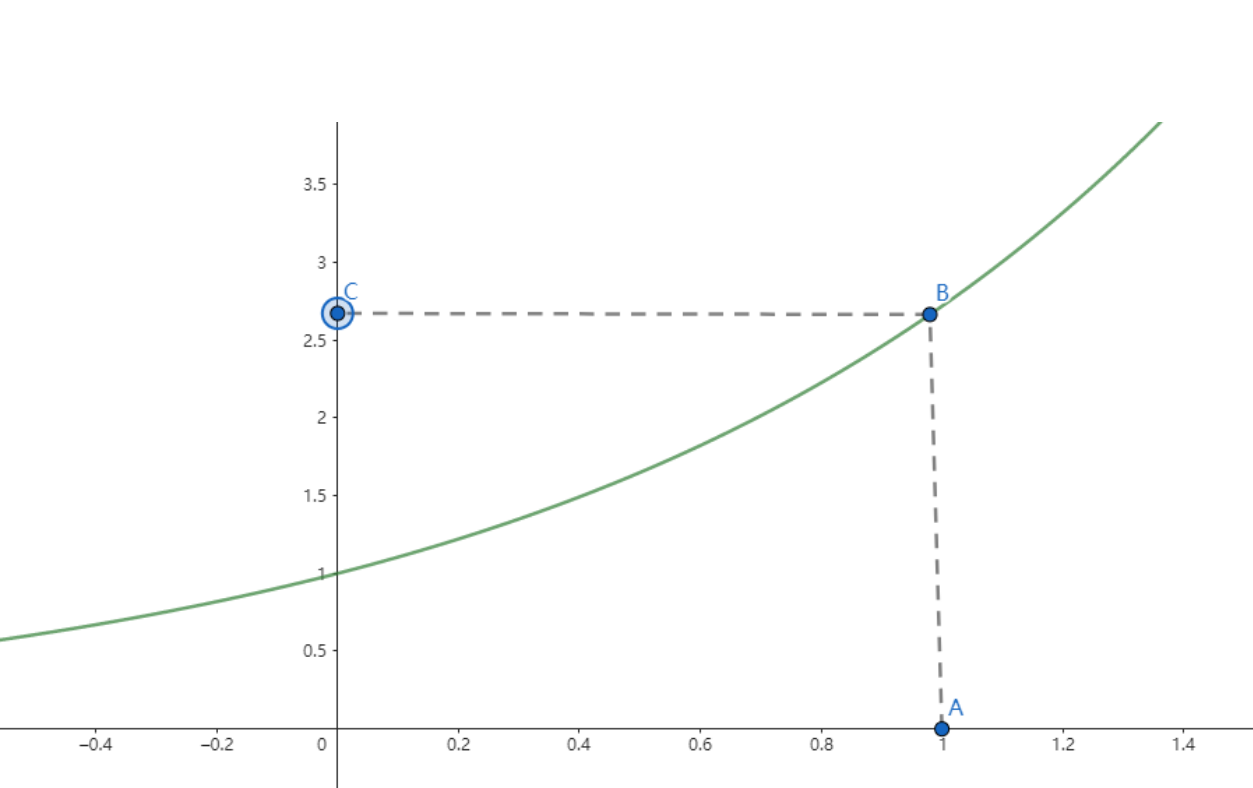
We define the vector of the paper：

where is the word vector of the word in the abstract of the paper and 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 , 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.

#### Scholar Embedding

First we define some notation. represents the total number of co-authors in paper of scholar . 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 to represent, shown in Fighure \ref{}.



Based on the above assumptions, then the contribution degree of scholar to paper is expressed as follows:

where 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 of scholar 's paper as follows:

where is the number of times the paper is cited. And combined with the impact factor of the paper, we can get the impact of paper , which is defined as follows:

where is the number of citations of paper , is the highest single citation number of scholar .

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. \cite{} 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:

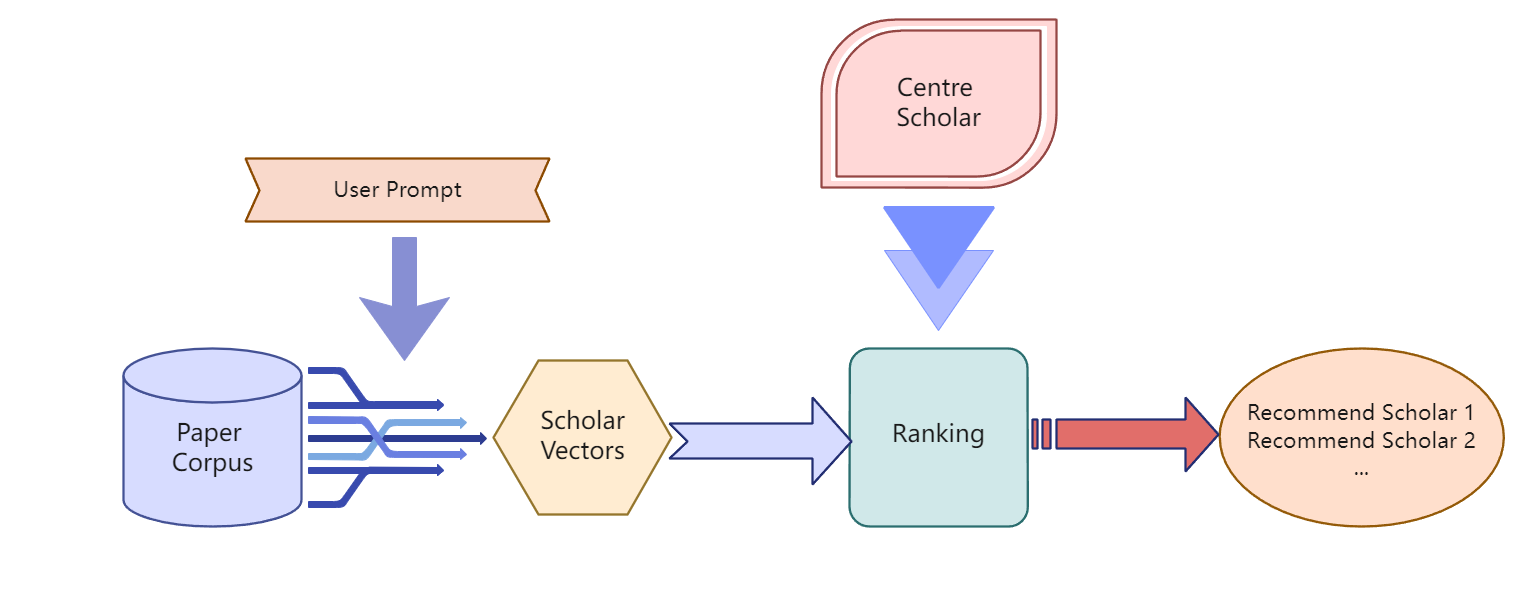
represents the combined influence of paper , here we consider two factors: the contribution of scholar to paper and the influence of paper , where the weight parameter determines the influence of both and on . with values in the range .

With the above steps we obtain the influence of scholar in paper , and next multiply it with the vector of paper of scholar extracted by GloVe and perform the weighted average operation to obtain the vector of scholar , which is defined as follows:

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

### 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 \ref{fig:flow\_chart}:



Similarly to the Formula \ref{formula}, we define the vector form of the user input prompt.

where refers to the GloVe vector representation of the word in prompt.

Next, we compute the similarity between the vector of scholar 's paper and the vector representation of prompt to obtain the similarity between paper and prompt, which is calculated as follows:

Due to , to change the range of from to , we modify the formula in the following way:

where is the transformed value of .

In turn, we obtain the weights of scholar 's paper , consisting of the similarity and the combined influence \cite formula components.

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

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

where denotes the total number of papers written by scholars .

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

where refers to the vector representation of the central scholars.

We sort in descending order and recommend the top-ranked scholars as potential collaborators of the central scholars.

## 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 dashboard shown in Figure \ref{fig:dashboard}, which is divided into three parts: sidebar console, a display interface and message notification interface. Here we will mainly introduce the display interface.

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 unstructured prompt text that describe 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 in Equation \ref{eq:}.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 information, the level of academic research, and the functionality of the recommender 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.

# Evaluation

To evaluate our system, we generated scholar documents mentioned in Section \ref{} 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 accuracy of the data. For this system, we designed two experiments, which are used to evaluate the effect of author embedding and recommendation respectively.

## Embedding Evaluation

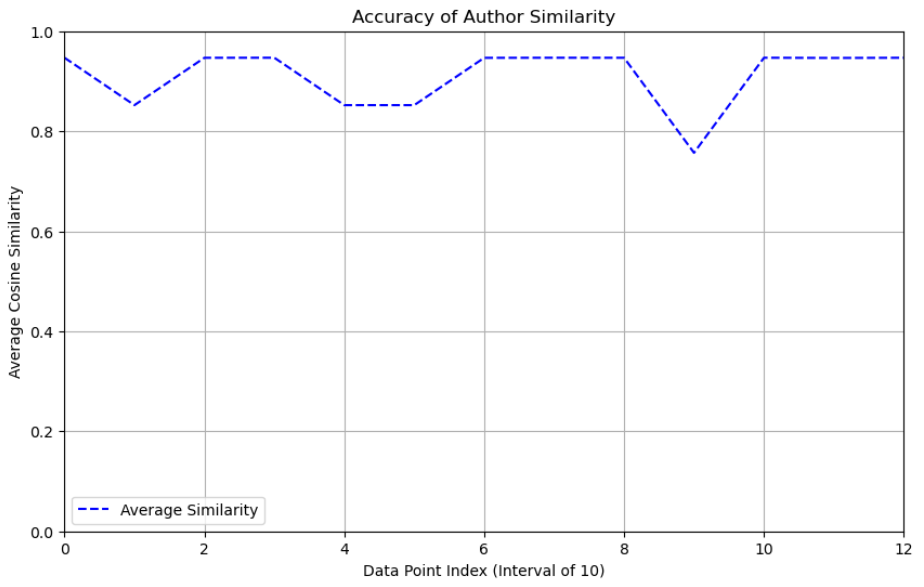
To evaluate the effect of author embedding, we conduct experiments on the author similarity task. Pennington \cite{} et al. evaluated GloVe for word similarity, and accordingly, we design the author similarity task referring to their work, this is because our embedding is based on GloVe pre-trained model. 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 , where and denote the two scholars, respectively.

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

Here, denotes the vector of scholar , and denotes the vector of scholar .

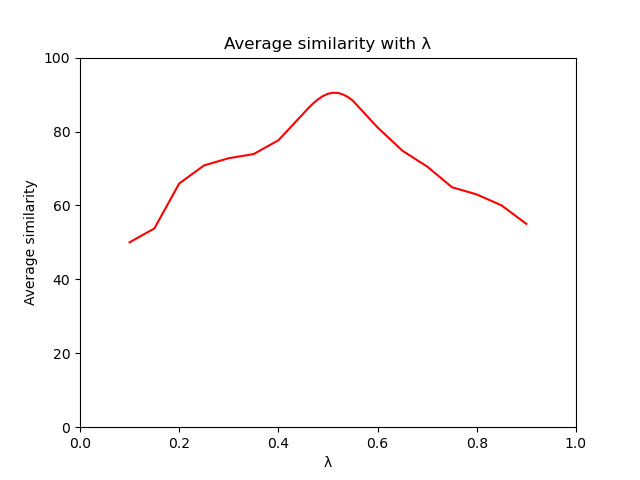
According to the above method, the similarity between scholar and all other scholars that the expert evaluation result can be obtained in the same way. We choose the cosine similarity between and as the evaluation metric for the author similarity task, where closer to 1 indicates that our author embedding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

### Results

We show the similarity between the two sets of data in Fig.. 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 . The average accuracy of 126 data points is **0.91**, which indicates that our scholar embedding model can accurately reflect the collaboration between scholars.

### Model Analysis: Selection

In Fig ?, we show that Eqn. (?), the impact of different choices of on overall model accuracy. You can see that the accuracy increases at the beginning as increases. Because when the value of is larger, the author's vector is affected by its contribution, which can better reflect the importance of the author in the paper, making the evaluation more accurate. But after a certain amount of , the accuracy starts to drop. This is because when the value of is large, the vector of authors is affected by the influence of the paper, but this influence is determined by the number of citations of the paper, and the number of citations of the paper is affected by many factors, such as the publication time of the paper, the topic of the paper, etc. Therefore, after the value of reaches a certain level, the author's vector will be affected by many irrelevant factors, making the evaluation inaccurate. For this reason, we chose a value of 0.56 for , which gives us the maximum accuracy.



## Recommendation Evaluation

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

### 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 recommender system in their familiar domain, based on their expertise and experience.  
According to the prompt of the evaluator's input and the central scholar, 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 .  
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 \ref{tb:record} shows the ratings for the top 5 recommendations.

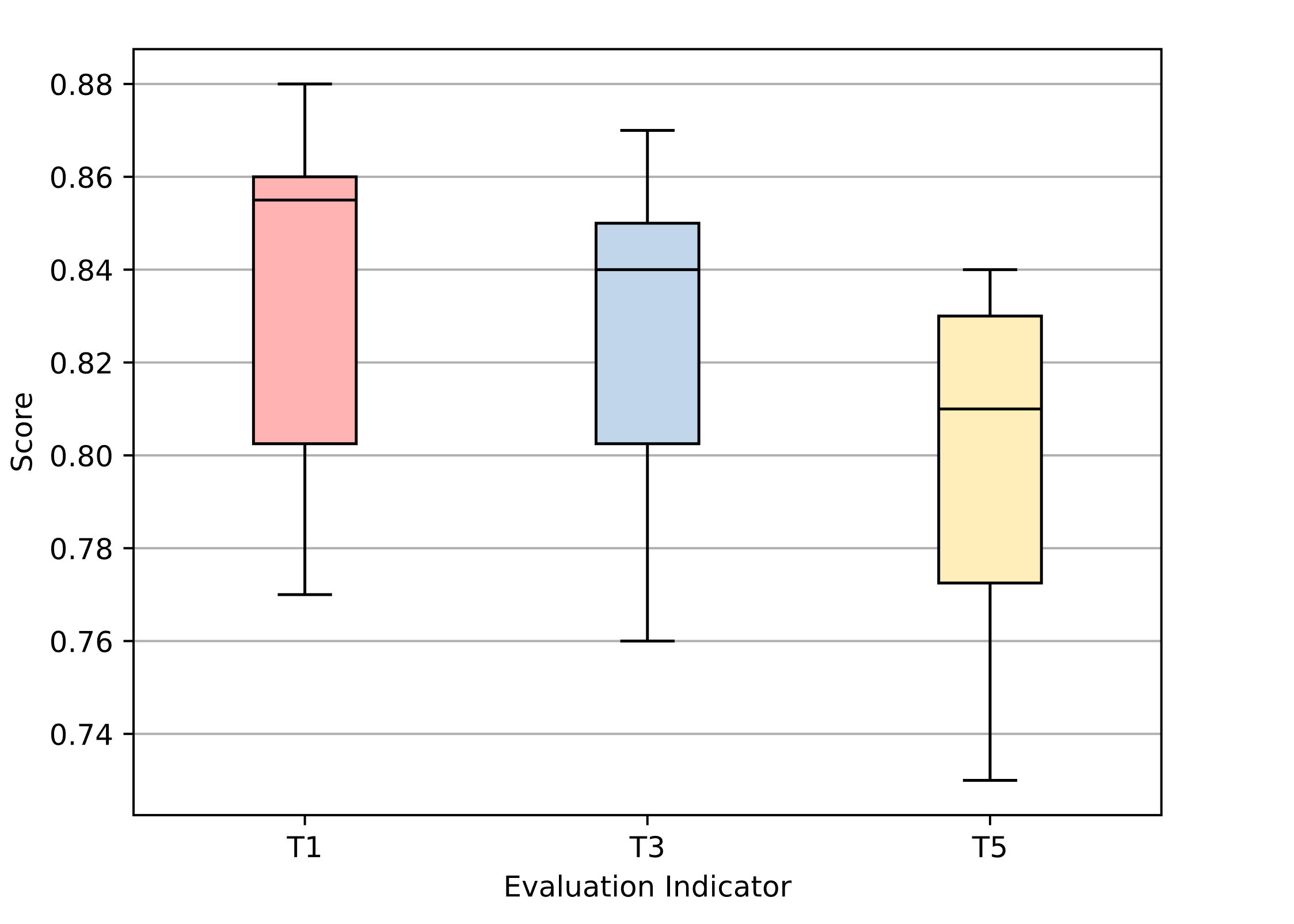
|  | Recommendation Information |  |  |  | Score |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Record ID | Center Scholar | Prompt | 1st | 2nd | 3rd | 4th | 5th |
| 01 | \*\* | bioscience | 0.54 | 0.91 | 0.39 | 0.89 | 0.36 |
| 02 | \*\* | graph neural network | 0.70 | 0.62 | 0.49 | 0.37 | 0.52 |
| 03 | \*\* | beamforming | 0.85 | 0.84 | 0.51 | 0.69 | 0.26 |
| 04 | \*\* | multi-agent | 0.82 | 0.76 | 0.74 | 0.71 | 0.40 |
| 05 | \*\* | econometrics | 0.95 | 0.71 | 0.92 | 0.96 | 0.30 |

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

where represents the rating of the first recommendations, and represents the satisfaction rating of the evaluator for the recommended collaborator.

### Experiment Results

We use box plots to describe 100 data points (recommendation records) statistically, to show the performance of the recommender 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 \ref{fig:boxplot}:



The median and mean values of are 0.85 and 0.84, respectively, which are higher than those of and , 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 and is around , which can be considered as the recommendation quality is similar. On the other hand, 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.

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 . 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 are all above 0.8, which indicates that the recommendation system designed in this study has good accuracy in recommending coauthors with a high degree of relevant fit, and can recommend valuable coauthors for scholars.

# 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 cooperation 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.

# Reference