Informatics in Medicine Unlocked 38 (2023) 101224

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Identifying effective criteria for author matching in bioinformatics

Fezzeh Ebrahimi [a], Asefeh Asemi [b,*], Andrea Ko [b]

[a] Department of Scientometrics, Faculty of Social Sciences, Yazd University, Yazd, Iran
[b] Corvinus University of Budapest, Budapest, Hungary

ARTICLE INFO

Keywords:
AHP
Bioinformatics
FDM
Fuzzy MCDM Methods
Health informatics
Potential author matches
Scientific collaboration
TFN

ABSTRACT

With the increasing development of information and scientific databases, scientific collaboration has expanded in health sciences. This study aims to prioritize the criteria that affect finding potential author matches in bioinformatics using fuzzy Multiple Criteria Decision Making (MCDM) methods such as Analytical Hierarchy Process (AHP), Fuzzy Delphi Method (FDM), and Triangular Fuzzy Numbers (TFN). To answer the research questions, a mix of documentary analysis and fuzzy methods is utilized. The documentary analysis stage involves collecting relevant documents and resources using the purposive sampling approach and ranking the effective criteria. The subsequent step involves experts determining the priorities of the effective criteria using pairwise comparisons and the Delphi questionnaire. The final weights are obtained based on the research purpose. The study shows that 79 criteria related to the research purpose can be grouped into three general categories: behavioral, topological, and content-based criteria. The most effective criteria in finding and recommending a potential author match are "journal titles", "citations", "paper titles", "affiliations", "keywords", and "abstracts". Among these criteria, citation and paper titles have a higher priority compared to others. The results indicate that content-based criteria have the most significant impact on finding potential author matches in static scholar networks and networks with text information. Furthermore, among the content-based criteria, the number of publications in common specialized journals and the number of common citations are the most sought-after criteria for finding a potential author match with the highest similarity.

1. Introduction

Identification of potential author matches in co-authorship networks is a principal issue faced by authors [1]. Regarding the increasing growth of different sciences, collaboration in producing scientific contributions has been developed [2–8]. Due to the explosion of information caused in part by the emergence of big scientific data, this trend has become much more time-consuming. Determining the best potential co-authorship leads to saving time, higher efficiency, and higher quality of research activity and knowledge development. Various criteria affect the identification of potential author matches. In the field of bioinformatics, which is an interdisciplinary field as an interaction between computer science, mathematics, statistics, physics, biology, chemistry, biochemistry, engineering, and biophysics [9], the method for identification of potential author matches to the corresponding co-authorship networks helps other authors to find their potential author matches more appropriately and successfully. Although various methods such as link prediction, text analysis, machine learning techniques, etc. have been used for this purpose [10–13], specific features have been used for these without considering the weight and priority of features. In this research, we identified and prioritized effective criteria in the field of bioinformatics using the PubMed database. We used the Analytic Hierarchy Process (AHP) technique to calculate the weights and prioritize the effective criteria for finding co-authors in bioinformatics. The main purpose of the present research is to identify and prioritize effective criteria for finding a potential author match using the AHP technique in the bioinformatics area. To achieve this, we followed the following specific objectives: 1. To determine the effective criteria for finding co-authors based on the literature, 2. To determine effective criteria for finding the author's match in bioinformatics from PubMed based on the results obtained from the previous step, step 3. To calculate the weights and prioritize the effective criteria for finding co-authors in bioinformatics by using fuzzy methods and 4. To finalize the most important criteria for finding co-authors compare the achieved results of different methods. In summary, the identification of potential author matches in bioinformatics is crucial for improving the efficiency and quality of

* Corresponding author.
E-mail addresses: ebrahimif@yahoo.com (F. Ebrahimi), asemi.asefeh@uni-corvinus.hu (A. Asemi), andrea.ko@uni-corvinus.hu (A. Ko).

https://doi.org/10.1016/j.imu.2023.101224
Received 28 October 2022; Received in revised form 16 March 2023; Accepted 16 March 2023
Available online 17 March 2023

research activities. By using the AHP technique, we can prioritize effective criteria for finding co-authors in bioinformatics. This research provides insights into the effective criteria for identifying potential co-authors and helps authors to find their potential author matches more successfully.

2. Scientific collaboration

Scientific collaboration has been expanding in recent times, thanks to the advanced web-based communication systems and comprehensive scientific databases that have become available [14]. Collaboration is now common, particularly in the sciences, as it provides a means of sharing resources, skills, and data [15]. Scientific collaboration occurs when two or more researchers work together to achieve a common goal and create new knowledge. Scientific collaboration can be viewed as the interaction of scientists in a social context where they share tasks and responsibilities to facilitate the achievement of a common goal [16]. Scientific collaborations arise in larger social contexts beyond science and involve various elements, such as peer review, reward systems, hidden colleagues, scientific concepts, national and international policies, and norms of academic discipline [17]. Scientific collaboration can take various forms, including authoring or translating a book, presenting a paper in journals and domestic and international conferences, proposing research, membership, and participation in scientific communities, and collaborating with scientific journals [18]. Collaboration may also involve scholars partnering in the supervision, consultation, and review of graduates [19]. Researchers and writers can have scientific partnerships at different levels, including organizational and inter-organizational (national and international) levels.

3. Co-authorship

Effective research on various topics often requires a suitable co-author. Data science has played a significant role in facilitating scientific collaboration [20]. Co-authorship involves two or more authors working together to produce scientific output. This process allows authors to share resources and talents, resulting in collaborative scientific work. Today, co-authorship is widely accepted in most fields of science and engineering [21]. While only those directly involved in the scientific output are listed as co-authors, other contributors who have played a role in the formation of the work are recognized in the appreciation section. The number of contributors mentioned in this section is steadily increasing [22,23]. Many countries have policies aimed at promoting university-industry relationships and increasing international cooperation to enhance scientific collaboration. This is supported by governments at various levels as the benefits of scientific collaboration outweigh the costs. However, the level of scientific collaboration varies across disciplines and regions. Collaborators who are responsible for key parts of the research process are called “scientific collaborators”, and their involvement is often critical to the success of the research. Scientific collaboration is necessary due to the increasing complexity and specialization of scientific fields, as well as the growing need for interdisciplinary and multidisciplinary studies. Trust is an essential factor in scientific collaboration, although it requires further research. Other factors that influence scientific collaboration include geographical location, social and cultural factors, individual characteristics, language, economic and political factors, and sometimes religious factors.

4. Potential author matches

Researchers and authors often seek out suitable colleagues whose expertise complements their own in order to enhance the quality of their scientific work [24]. Collaborating with like-minded individuals who possess complementary skills and interests can lead to “synergistic creativity” [15]. Co-authors may also act as “critical friends” who challenge assumptions and point out shortcomings, resulting in improved article quality. Collaboration between authors can also accelerate manuscript production for publication, especially in the case of books, where sharing content allows for completion in a shorter time frame. However, joint writing is not solely for the purpose of publication; it can also be an enjoyable and enriching scientific collaboration. More experienced writers may view it as an opportunity to mentor less experienced colleagues. Some researchers choose to collaborate with colleagues within their organization, while others prefer to work with those outside for various reasons. Organizational limitations, such as restricted access to laboratories and facilities, are often the most important reason for seeking external collaboration [25]. Overall, the benefits of collaborating with other researchers and authors, and having potential co-authors, include improving article quality, accessing valuable expertise and ideas, dividing labor, providing additional releases for promotion/tenure, receiving guidance from senior colleagues, learning from fellow authors, exploring interdisciplinary studies, and interacting with other researchers [26,27]. These reasons are driven by the complexity of human knowledge fields and the need to access a wide range of expertise, resources, facilities, and skills.

5. Co-authorship in bioinformatics

Bioinformatics is an interdisciplinary field that involves the use of advanced computational techniques for biological data analysis. It combines computer science, statistics, and engineering to interpret and analyze biological data using mathematical and statistical techniques [28,29]. As this field continues to grow, collaboration and scientific partnership are essential to progress in interdisciplinary and multidisciplinary topics. Since companies and organizations that work in bioinformatics have different resources and facilities, researchers and authors require collaboration and scientific partnerships with other companies and organizations to keep up with the latest technologies. In recent years, Social Network Analysis (SNA) has been applied to co-authorship in bioinformatics, given the potential of this field in the collaboration of authors. As shown by content analysis, there is an increasing overlap between bioinformatics journals in terms of topics, and more research groups are participating in bioinformatics research due to the similarity of the co-authorship network [30]. To provide more insight into the research published in bioinformatics, we conducted a bibliometric analysis of the publication on the Web of Science Core Collection. We retrieved 84,811 results from the publication year 1998–2022 (17th March 2023). Fig. 1 shows that the rate of increase in bioinformatics publications has increased in recent years. Fig. 2 displays the top 25 authors in the field of bioinformatics, highlighting the significant number of publications attributed to each author and emphasizing the collaborative nature of research in this field. Recent years have seen a considerable increase in collaborative efforts. To further investigate the extent of author contributions, we utilized VOSviewer software on the WoS database to analyze co-authorship patterns in 2023 for 1267 documents with 8391 authors (Fig. 3). The co-authorship analysis was based on the authors as the unit of analysis, with a maximum of 25 authors per document and a minimum of 5 documents per author. To select authors, we calculated the total strength of co-authorship links for each authors. The 78 authors with the highest link strength were chosen, and a network of 1267 connected items was formed. Some items in the network remained unconnected. Given the collaborative nature of bioinformatics research, facilitating the identification of potential co-authors may benefit researchers in this field.

6. Literature review

Several articles on co-authorship, scientific networking, scientific collaboration, content-based suggestion systems, and group refinement have been published, and this study examines some of them. Mooney and Roy [31] studied the LIBRA system for “book suggestion” using a content-based approach to learn how to classify text. Their findings

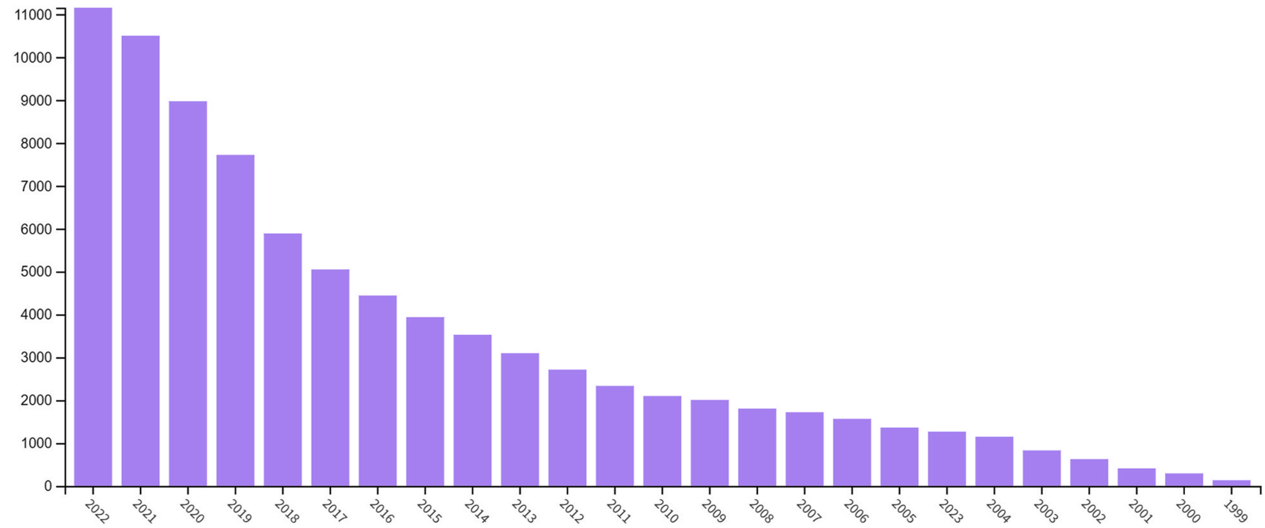

Fig. 1. Number of publications for bioinformatics (topic) in WoS per year.

Fig. 2. Number of publications for bioinformatics (topic) in WoS by each author.

indicate that this method can provide accurate book suggestions. The article uses various elements, such as the title, components, authors, abstracts, published reviews, customer reviews, related authors, related titles, and thematic terms to suggest books. The data used for this study was collected from the Amazon website, and the implementation of the LIBRA intelligent system is based on Bayesian algorithms. The evaluation of this system is based on components such as F-Measure, classification accuracy, recall, precision, and rank correlation. In another study on Scientific Literature Recommender Systems (SLRSs), Cabanac [32] presented articles to researchers based on their scientific interests. He relied on the criteria of similarity between researchers, which are typically calculated based on the content of publications (extracting topics and citations of articles). Cabanac defined the thematic similarity between the two scholars concerning their publication, regardless of the abstract and full text. Achary [33] designed a recommender system for authors using content-based refining methods. He used a combination of content-based and group-based methods on the datasets from CiteSeer[x] and BibSonomy, respectively. His results showed that by using social tagging and textual information, the quality of the recommendations improved. Sun et al. [34] predicted the co-authorship relationship in a bibliographic network using heterogeneous topological features for predicting co-authorship relationships. Their results showed that heterogeneous topological properties can improve the accuracy of link prediction and that topological feature plays a key role in deciding on future cooperation. Brandão & Moro [35] examined link prediction on an academic social network to recommend collaborations. They considered the affiliation with the institution (given by the metric called Affin) with the criteria of cooperation, solidarity, and social closeness. Ghare-Chamani [36] presented a consulting system that identifies the most important and authoritative scientific sources in a field. The proposed method has been done with the help of the scientific network of articles. Andrikopoulos et al. [37] analyzed articles in the journal

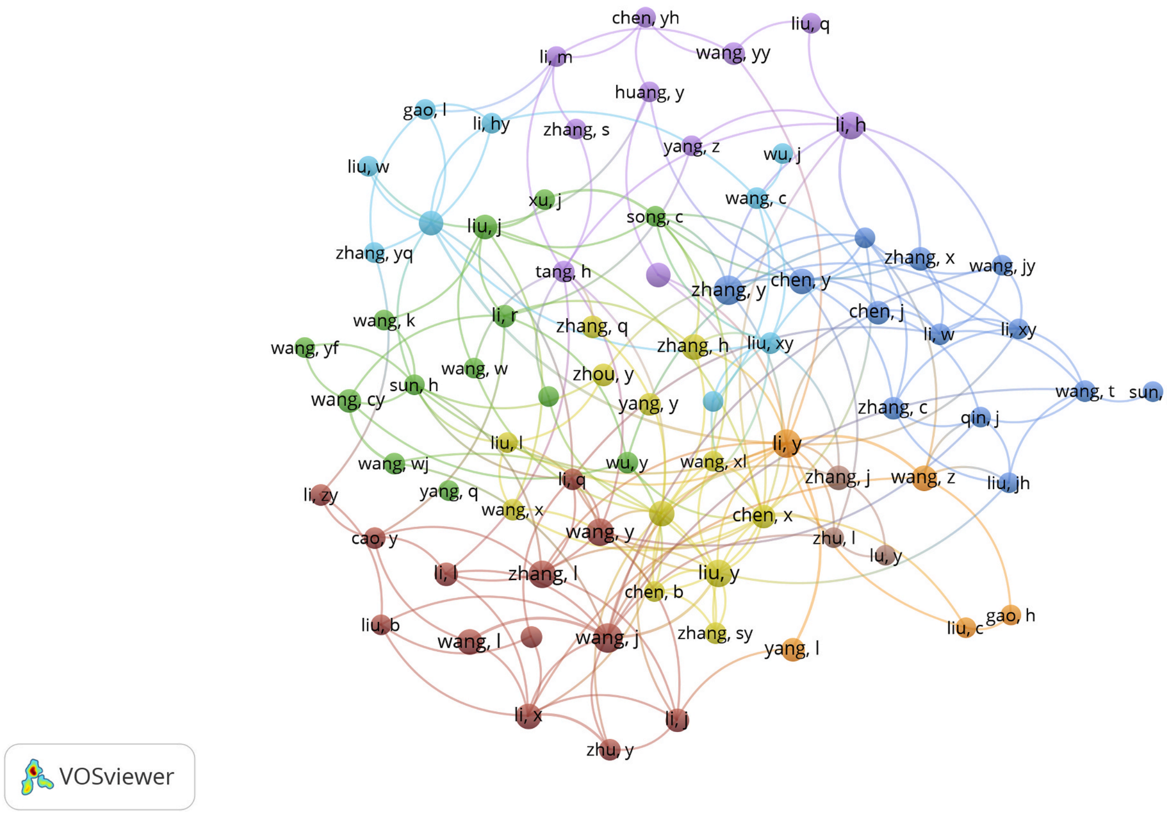

Fig. 3. Co-authorship analysis in bioinformatics (topic) based on the WoS database.

Econometrics and identified the most prolific authors, institutions, and countries in the Journal of Econometrics. The findings show that the co-authorship network is increasingly integrated, and the pattern of collaborative writing has been growing. Makarov et al. [38] proposed a system for finding collaborations according to research interests. They developed the recommendation system as a link prediction in the co-authorship network, derived from bibliographic databases and enriched with information on research papers obtained from Scopus and other journal ranking systems. Ho et al. [39] in their paper present an approach to predict the co-authorship of a bibliographic network using geographical factors and latent subject information. They used a supervised method to predict co-author relationship formation in which the combination of atypical features with different measurement coefficients. In the first stage, in addition to examining the relationships in previous research, a new relationship related to the geographical factor that plays a role as a topological feature has been exploited. Of course, the content feature is extracted based on the textual information of the keyword, titles, abstracts, or all articles. Finally, topological features and content features are integrated for predicting the common relationship. In this paper, thematically modeling is used simultaneously to estimate the similarity of textual information instead of the two existing methods, which are the number of common keywords and TF-IDF, and accuracy and ROC-AUC are used to evaluate the performance. Li et al. [40] in an article on personalized reclassification of the article proposal using content and behavioral characteristics. They believe that recommendation systems play an important complement to search engines. Proposal scenarios are divided into three categories based on timing. The first category is before the start of a search, the second category is during the search, and the third category is after the search. This research focuses on the third party. This article presents a combined content and behavior model for ranking selected articles in Science Direct. In this study, a pairwise learning model was used for reranking to propose a candidate article, which leads to improved results. Moreover, Hybrid Research Methods (HRMs) of content and user behavior and SVD algorithms and descending gradients have been used. The results show that LibFM and SVDFeature performed worse than HRM. There are some studies about the related networks such as Ji et al [41] and Ullah et al [42]. They have studied the co-citation and co-authorship networks of statisticians and analyzed interdisciplinary research using co-authorship networks. This research aimed to identify and prioritize the criteria affecting finding potential author matches in bioinformatics using fuzzy MCDM methods such as AHP, FDM, and TFN. This research aimed to identify and prioritize the criteria affecting finding potential author matches in bioinformatics using fuzzy Multiple Criteria Decision Making (MCDM) methods such as AHP, FDM, and TFN.

7. Materials and methods

This research uses a mixed quantitative and qualitative approach. In this research, the effective criteria of potential author matches are first extracted, and the most important ones in the PubMed database are then prioritized and weighted. Fig. 4 presents the conceptual model and research steps. The first step involves identifying the effective criteria for finding potential author matches based on a literature review. PubMed, a key scientific database in bioinformatics, was reviewed to identify relevant search tools for the 79 criteria obtained. Six search tools corresponding to the criteria were identified, including the journal's title, citations, article's title, affiliations, keywords, and abstracts, which were confirmed as the main criteria for finding potential author matches. A preferential judgment questionnaire was designed based on these six criteria using the qualitative approach of a focus group, which involved gathering views from participants who were not necessarily trained experts. The finalized questionnaire was then subject to a quantitative approach using Saaty's nine-point scale and a pairwise comparison matrix of expert opinions derived based on group AHP one of the MCDM techniques. The questionnaire was sent to nine authors in bioinformatics, biology, informatics, and scientometrics to assess the importance and priority of the criteria, with an inconsistency ratio calculated. Only

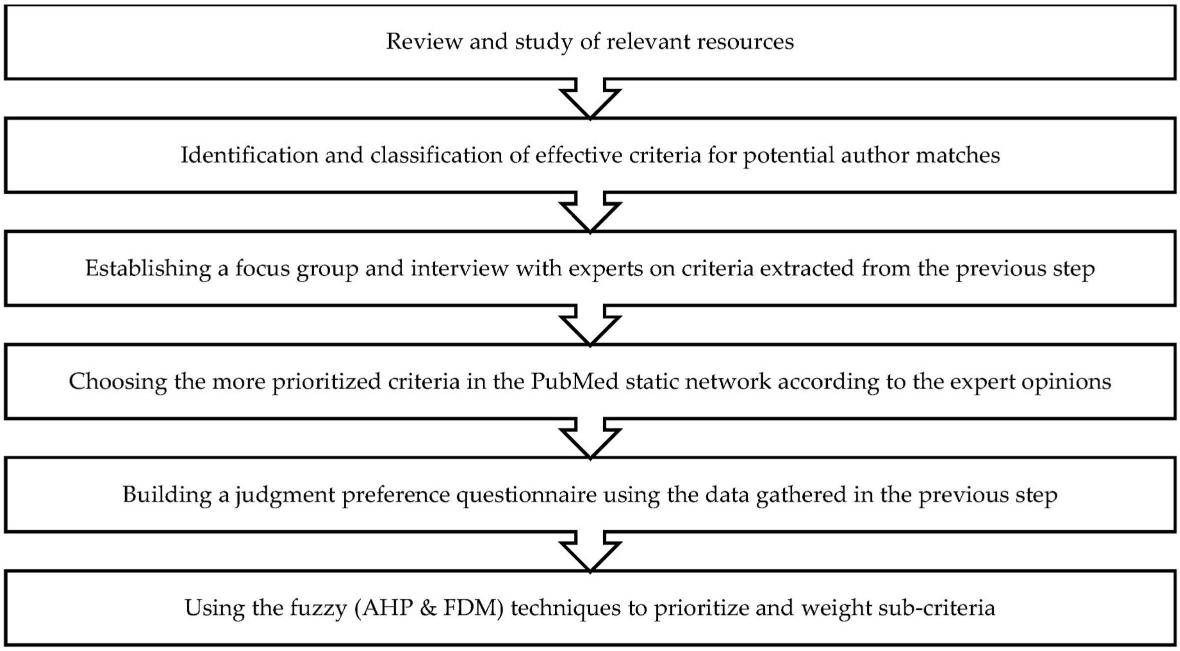

Fig. 4. Conceptual model of research.

one questionnaire had an inconsistency ratio of less than 0.1, and the associated ratios for other questionnaires were higher. As a result, the focus group was consulted again, some criteria were revisited, and experts were given clarifications about the questions orally or via phone call. The weight obtained for the criteria was 0.8 with an incompatibility ratio. The relevant matrix was then drawn by the expert Choice software and analyzed. Focus groups typically consist of 6–12 participants to ensure diverse information is obtained without making participants feel inconvenienced sharing their ideas, beliefs, and experiences [43]. In this step, the final questionnaire is designed using the quantitative approach and Saaty's nine-point scale [44], such as 1 for equal preference to 9 for infinite preference, and the pairwise comparison matrix of expert opinions is derived based on group AHP. After the effective criteria in the field of bioinformatics in the PubMed database were selected, a questionnaire for assessing preference judgment was designed based on the 6 main criteria, and it was sent to nine authors in the areas of bioinformatics, biology, informatics, and scientometrics to determine the importance and prioritize the criteria (Table 1). The inconsistency ratio was also calculated. Only one of the questionnaires had an inconsistency ratio of less than 0.1, and the associated ratio for other questionnaires was at a higher level. Accordingly, the focus group was consulted again, and some criteria were revisited. Moreover, the experts were given some clarifications about the questions orally or via phone call. The criteria weights were obtained with an inconsistency ratio of 0.8, and the matrices were formed and analyzed by the Expert Choice software. It should be noted that focus groups often consist of 6-12 participants. The logic behind this rule is that the number of members should be adequate to provide diverse information, and on the other hand, it should not be so extremely large that the members feel inconvenienced to share and express their ideas, beliefs, and experiences [45]. To collect information, we employed the Fuzzy Delphi Method (FDM), a modified and enhanced version of the classical Delphi technique that overcomes some of its limitations, such as low convergence, loss of essential information, and lengthy investigation [46]. Next, we used FDM to validate the results obtained from the AHP. Specifically, we formed triangular fuzzy numbers (TFN) in the FDM [47] to reprioritize the criteria affecting the identification of potential author matches. We provided an FDM-based questionnaire to the same experts, which was prepared based on the six finalized criteria. The experts rated the importance of each criterion

Table 1
Pairwise Comparisons members.

Field and Organization	Frequency
Scientometrics specialist, University of Tehran	1
Scientometrics specialist, Kermanshah University	1
Scientometrics specialist, Yazd university	2
Scientometrics specialist, Maritime Provinces Higher Education Commission, NB, Canada	1
Informatics specialist, University of Masaryk, Czech Republic	1
Scientometrics & Bioinformatics specialist, Corvinus University of Budapest, Hungary	1
Bioinformatics Specialist, The School of Biological Sciences at the Institute for Research in Fundamental Sciences, Tehran	2

on a five-point Likert scale, and we determined the frequency of their responses. Using Equations (1) and (2), we calculated the TFN of each criterion to determine its priority.

$$E_i = \left(e_1^{(i)}, e_2^{(i)}, e_3^{(i)}\right), i = 1, 2, 3, \ldots\ldots n \qquad \text{Eq. (1)}$$

$$E_{ave} = (m_1, m_2, m_3) = \left(\frac{1}{n}\sum_{i=1}^{n} e_1^{(i)}, \frac{1}{n}\sum_{i=1}^{n} e_2^{(i)}, \frac{1}{n}\sum_{i=1}^{n} e_3^{(i)}\right) \qquad \text{Eq. (2)}$$

This equation uses Ei to represent the expert opinion on the six effective criteria for a given expert, with E1 representing the first expert, E2 representing the second expert, and so on up to the nth expert. Here, n is a non-negative integer. The variable Eave represents the average of all experts' opinions in the present study. These variables are defined and presented as TFN in Fig. 5. Table 2 shows how the linguistic variables are converted into TFN and Crisp Numbers/values (CN). This study employed linguistic variables with five scales to survey experts, and the membership functions for these variables are illustrated in Fig. 6. The figure presents the priority of each criterion in the AHP and FDM methods. The linguistic variables were used to gauge the impact of each factor or criterion in finding potential author matches based on expert judgment. After estimating the TFN for each of the six effective criteria, the Minkowski formula (Eq. 3) was used to determine the definite fuzzy numbers for each criterion in finding potential author matches [21,22].

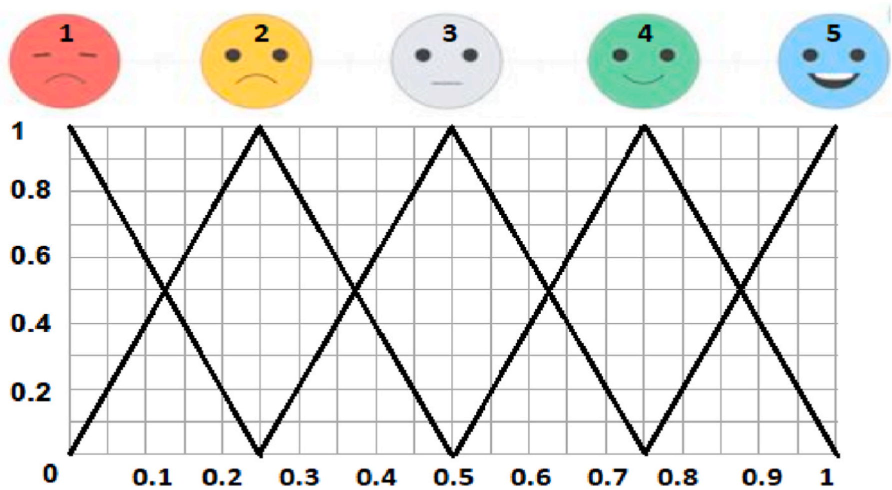

Fig. 5. Definition of linguistic variables based on a five-point Likert scale.

Table 2
TFN & CN of linguistic variables.

Linguistic Variable	TFN	Crisp Values
Extremely More Importance (EMI)	0.75	0.75
Very Strong Importance (VSI)	0.75, 0.5	0.55
Strong Importance (SI)	0.25, 0.5	0.31
Moderate Importance (MI)	0.25, 0	0.151
Equal Importance (EI)	0.25, 0	0.151

$$x = m_1 + \frac{(m_2 - m_3)}{4} \quad \text{Eq. (3)}$$

8. Findings

The research findings are presented in the following three sections: Effective criteria for finding potential author matches, Identifying the effective criteria for finding potential author matches in the bioinformatics area, and Weighting and prioritizing the effective criteria for finding potential author matches in the bioinformatics area using AHP and FDM. Finally, the results obtained from AHP and FDM were compared.

8.1. *Effective criteria for finding a potential author matches*

Based on a review of the literature and an analysis of existing domestic and international studies, a total of 79 sub-criteria were identified in response to the first research question. These sub-criteria were then categorized into three classes: behavioral criteria (15 sub-criteria), topological criteria (56 sub-criteria), and content-based criteria (8 sub-criteria), as shown in Table 3.

8.2. *Identifying the effective criteria for finding potential author matches in the bioinformatics area*

To identify the effective criteria for static social networks of PubMed,

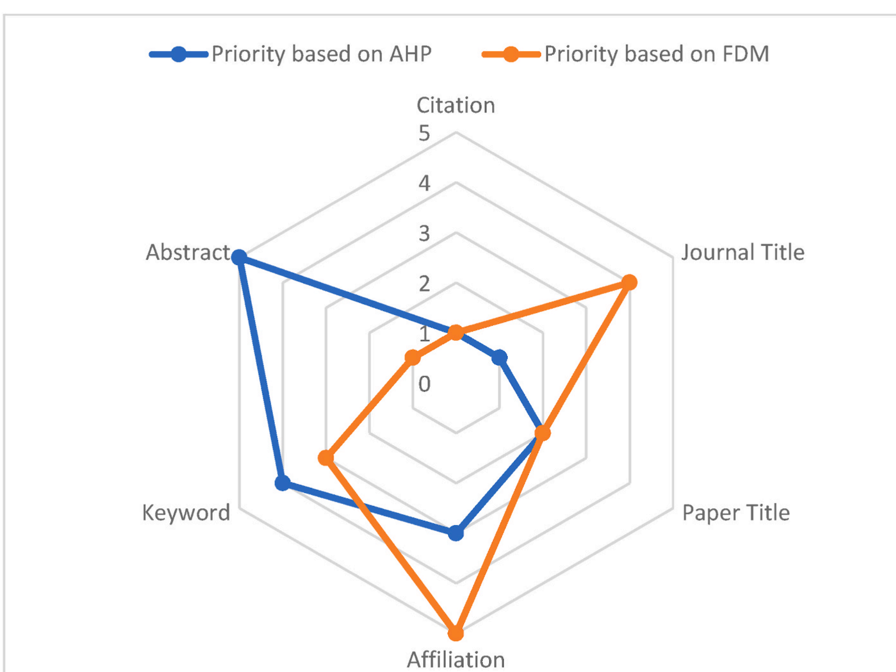

Fig. 6. The priority of each criterion in AHP and FDM methods.

Table 3
Criteria & sub-criteria in finding a potential author match.

No.	Criteria	Sub-criteria
1	**Behavior-Based Criteria**	Users click behaviors (such as clicking on the ranking results, previous clicks, etc.)
		Pages browse (such as reviewing articles in a particular journal, browsing freshly published articles, log history, etc.)
		Dialogue
		Bookmarks
		Personal web pages
		How to inform on the web
		Tweets
		Frequency of likes
		Instant availability
		Save
		Visibility rate
		Download rate
		Timestamps
		Stop time on the page
		User interest and scope of work
2	**Content-Based Criteria**	Overlap of terms in the title of the publication
		Similarities between the title of the article and the title of citations/references
		Keyword's similarity
		Keywords assigned by the author
		The similarity of subject similarity of tags
		Abstract similarity
		Body similarity
3	**Topological Criteria**	Frequency of citations
		Frequency of references
		Bibliographic coupling
		Frequency of citations of the first author
		Frequency of citations of the last author
		Co-citation
		Co-citation analysis of authors (examining the relationship between Cited authors and citer authors)
		Journal citation analysis
		Author reputation-based h-index
		Author reputation based on the Nobel prize
		Author reputation-based g-index
		Author reputation-based hg- index
		Author reputation-based m quotient
		Author reputation-based a-index
		Author reputation-based m- index
		Author reputation-based r- index
		Author reputation based ar-index
		Author reputation based on h2
		Co-author
		Physical proximity between the first and last author
		Frequency of joint co-authors
		Frequency of articles by the first author
		Frequency of articles for the last author
		Research friends
		Author impact factor
		Collaborative proximity
		Published works
		Co-authored works frequency
		Institution's frequency
		The quality of the organization of the first author
		Affiliations
		Type of title (question, descriptive, and announcement)
		Title length
		Similarity of place
		Geographical proximity
		Scientific cooperation with other countries
		Collaboration in an institution (being a faculty member)
		Collaboration of the first and last author in an institution
		Increase the average frequency of citations with the frequency of Affiliated countries
		Difference between dates
		Date of publication
		Publication venues
		Common publication
		Commonly visited venues
		Source type
		Source language
		Type of article (review, …)
		Type of access to the article (open access, etc.)
		Abstract format (structured, unstructured)
		No reference to gray sources
		Lack of self-citation
		Quality of the journal (criterion of the journal's title)
		The historical similarity of publication
		Development of countries
		Journals impact factor
		Article length

a focus group was organized consisting of experts in biology, bioinformatics, and scientometrics. After conducting interviews and discussions, 79 potential criteria were identified. From these, six criteria, including common journals, paper titles, affiliations, paper keywords, abstract similarity, and citations, were selected as the most effective criteria. These six criteria were members of the content-based class, and a preference judgment questionnaire was created to gather expert opinions on their importance.

8.3. *Weighting and prioritizing the effective criteria in finding potential author matches in the bioinformatics area by AHP*

In order to determine the relative importance and prioritization of the criteria, a matrix was formed using the Expert Choice software which included six sub-criteria: "subject terms in paper titles," "subject terms in abstracts," "subject terms in keywords," "affiliation similarity," "frequency of paper publications in common specialized journals," and "similar citations" (Table 4). The weights were then calculated using the group AHP method. The weights obtained are presented in Table 5.

8.4. *Weighting and prioritizing the effective criteria in finding potential author matches in the bioinformatics area of FDM*

In this stage, we administered the Delphi questionnaire to 40 purposefully selected experts in the field of bioinformatics, based on the 6 criteria obtained in the previous stage. The experts' responses were then analyzed to determine the frequency of their opinions on each criterion, using a Likert scale. The results of this analysis are presented in Table 5. Subsequently, we calculated the Triangular Fuzzy Mean (TFM) and the defuzzification mean (DM) for each criterion using Minkowski's formula, and the results are also presented in this Table.

8.5. *Comparative analysis between AHP and FDM methods*

Both the AHP and FDM use fuzzy set theory [48] in the decision-making process to select a potential author match. However, due to the low accuracy in judging qualitative variables in AHP, FDM in TFN format was used to better evaluate the criteria using linguistic variables. Based on the results, both AHP and FDM methods are suitable for finding a potential author match. Each method has its own

Table 4
Matrix of expert opinions about potential authors matches criteria.

Affiliation	Citation	Title	Abstract	Journal	Keyword
	6.0	**3.0**	**3.0**	**6.0**	**3.0**
		6.0	**8.0**	**1.0**	**8.0**
			3.0	**6.0**	**1.0**
				8.0	**3.0**
					8.0
Incon: 0.08					

Table 5
Findings related to AHP and FDM techniques.

Code	Criteria	Matrix of expert opinions and final weight calculation	Frequency of expert opinions on the importance of each criterion based on the Likert scale					TFM & DM of each criterion		Comparison of the Criteria priority in AHP and FDM methods	
			EMI	VSI	SI	MI	EI	TFM m1, m2, m3	DM	Priority based on AHP	Priority based on FDM
C1	Citation	0.374	28	9	3	–	–	0.91, 0.68, 0.93	0.99	1	1
C2	Journal Title	0.374	15	13	10	2	–	0.91, 0.71, 0.11	0.95	1	4
C3	Paper Title	0.091	29	8	2	1	–	0.91, 0.65, 0.99	0.98	2	2
C4	Affiliation	0.075	18	10	6	3	3	0.85, 0.61, 0.95	0.92	3	5
C5	Keyword	0.055	11	13	8	7	1	0.89, 0.64, 0.98	0.97	4	3
C6	Abstract	0.031	20	11	5	4	–	0.93, 0.68, 0.11	0.99	5	1

advantages and disadvantages, and the most appropriate method should be selected based on the problem at hand. The AHP method involves a more complex number of comparative calculations than the FDM method. In AHP, criteria comparison is done in pairs on a 9-point scale, while in FDM, criteria comparison is not done in pairs and on a 5-point scale. Furthermore, FDM has no limitation on the number of criteria that can be chosen, whereas AHP does have this limitation. The results obtained from this round of methods are shown in Table 5, and Fig. 6 shows the priority of each criterion in AHP and FDM methods. The prioritization of the criteria in Table 5 shows that the citation criteria in both methods have the 1st priority in finding a potential author match. The title of the article is the second priority in both methods. The title of the journal and the abstract of the article have completely different results, and the results obtained from the two methods are opposite. Affiliation and keywords have the same priority in finding a potential author match.

9. Discussion & conclusion

The objective of this study was to prioritize the effective criteria for identifying potential author matches. Content-based criteria were found to be the most important among the three criteria, namely, content-based, behavioral, and topological criteria, in the field of bioinformatics and PubMed static networks. The content-based criteria included six aspects, such as "subject terms in paper titles," "subject terms in abstracts," "subject terms in keywords," "affiliation similarity," "frequency of paper publications in common specialized journals," and "similar citations." It is important to note that only citations to resources in the PubMed Central database are available. The findings suggested that "frequency of paper publications in common specialized journals" and "similar citations" had the highest importance, while "subject terms in keywords" had the lowest importance. Cabanac [32] highlighted journal contents as the key factor for recommender systems of scientific texts and suggested reading journals and conference papers to continuously review relevant resources and scientific materials. This result may suggest that authors select journals based on their specific expertise. For instance, an author specializing in genomics would publish their papers in journals related to this area. Citations demonstrate the impact of cited resources on citing ones. Franke et al. [49] proposed a recommender system based on the number of citations to scientific papers, which suggests the most frequently cited or favorite papers in the relevant research area. Beel et al. [50] proposed a recommender system suggesting papers to specialists based on the number of citations and similarity between the reference lists. The highest score was obtained by the criterion of "paper titles." The title of a resource reminds readers of its identity and serves as the first image of the text. It includes the main idea of the author or authors. Although some works in the humanities may have metaphorical titles, there may be little conformity between the title and the body text. However, in the PubMed database that includes resources related to biology and medical science, the title characteristics are highly effective. In their study, Davarpanah [51] found that the consistency between the title and the body text in human science areas is lower than that in medical sciences. In the present study, "paper titles" received the highest weight in experts' views, compared with "affiliations," "keywords," and "abstracts." Nascimento et al. [52] found that the weight of a term in the title of a work is three times the weight of the same term in the context of the work. Similarly, Mooney & Roy [31] and Li et al. [40] designed a title-based recommender system to introduce an article or book to authors. Achary [33] also designed a content-based recommender system in which the author/authors' affiliation system is considered one of the important criteria [57–59]. Makarov, Bulanov, and Zhukov [38] believe that affiliation is an important criterion for scientific cooperation based on the experience of HSE University specialists. Other researchers [39,53–55] have found that keywords in scientific articles are particularly important for evaluating the similarity of the subject of articles. Aanonson [56] emphasized the importance of keywords in his study and stated that retrieval of documents using the keyword is sometimes better than using the subject. Ghare-Chamani [36] has also recommended the use of keywords for retrieving articles and has worked on this criterion. Sun et al. [34] concluded that metadata have common features in various scientific databases in articles, and the abstract of a work contains the most important part of the content of a work. Cabanac [32] found that access to full-text of articles in scientific databases has many problems for researchers, and therefore, abstract similarity was ranked fifth according to experts' views. Given the continuous growth in the volume of scientific research, finding the favorite subject and co-author is incredibly difficult. Information overload is a real phenomenon that causes delays and disruptions in most of our important decisions. Co-authorship in producing scientific contributions is growing, and the number of specialists collaborating in scientific publications is continuously increasing in all knowledge fields [4–10]. In this regard, one of the specialists' concerns is to find the best potential author match because the identification of effective authors in co-authorship networks helps other specialists to choose their specific co-authorship strategy more successfully. To address this problem, recommender systems, which include software tools and techniques that propose the best option using various knowledge types, user-related data, current items, and previous transactions, can be designed using the criteria identified in this research among three criterion types, including content-based, behavioral, and topological criteria. Content-based criteria have the most application in the PubMed information network, and among these criteria, journal titles, citations, paper titles, author affiliations, keywords, and abstracts are recognized as the most important ones. Determining a potential author's match is a complicated process that depends on various parameters, and determining the relative weight of each criterion, which is an MCDM process, is important during this process. The current study deals with the potential author match for scientific collaboration in bioinformatics, and different potential authors' match criteria were considered via three methods. The first step was the documentary method, and in the next two fuzzy methods, AHP and FDM.

Human ethics and declarations sections, including consent to participate

Not applicable.

Ethics declaration statement

Not applicable.

Consent to participate statement

Not applicable.

Consent to publication

All the authors are agreed with the publication of the manuscript.

Funding

Not applicable.

Authors' contributions

EF: Data collection and data analysis, writing some parts of the main text, AA: Supervising, data analysis, and writing the main text, AK: Editing. All authors have read and agreed to the published version of the manuscript.

Availability of data and material

One part of the data was downloaded from the site https://www.ncbi.nlm.nih.gov/pubme d/in. January 2020, other data, tools, and technical files are available on Mendeley Data [60].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Abbreviations

Term Abbreviation
Analytic Hierarchy Process AHP
Area under the ROC Curve AUC
Crisp Numbers CN
Delphi Method DM
Factorization Machine Library LibFM
Frequency–Inverse Document Frequency TF-IDF
Fuzzy Delphi Method FDM
Hybrid Research Methods MRHs
Multiple-Criteria Decision Analysis MCDM
Receiver Operating Characteristic curve ROC curve
Scientific Literature Recommender Systems SLRSs
Singular Value Decomposition SVD
Social Network Analysis SNA
Triangular Fuzzy Number TFN

References

[1] Das, K.; Samanta, S.; Pal, M. Study on centrality measures in social networks: a survey. Soc 2018 Netw. Anal. Min. 8, 13. https://doi.org/10.1007/s13278-018-0493-2.
[2] Drenth JP. Multiple authorship: the contribution of senior authors. J Am Med Assoc 1998;280:219–21. https://doi.org/10.1001/jama.280.3.219.
[3] Glanzel W. National characteristics in international scientific co-authorship relations. Scientometrics 2001;51:69–115. https://doi.org/10.1023/A:1010512628145.
[4] Weeks WB, Wallace AE, Kimberly BC. Changes in authorship patterns in prestigious US medical journals. Soc Sci Med 2004;59:1949–54. https://doi.org/10.1016/j.socscimed.2004.02.029.
[5] Levsky ME, Rosin A, Coon TP, Enslow WL, Miller MA. A descriptive analysis of authorship within medical journals, 1995-2005. South Med J 2007;100:371–5.
[6] O'Brien TL. Change in academic co-authorship, 1953–2003. Sci Technol Hum Val 2012;37:210–34. https://doi.org/10.1177/0162243911406744.
[7] Henriksen D. The rise in co-authorship in the social sciences (1980–2013). Scientometrics 2016;107:455–76. https://doi.org/10.1007/s11192-016-1849-x.
[8] Kuld L, O'Hagan J. Rise of multi-authored papers in economics: demise of the 'lone star' and why? Scientometrics 2018;114:1207–25. https://doi.org/10.1007/s11192-017-2588-3.
[9] Benton D. Bioinformatics — principles and potential of a new multidisciplinary tool. Trends Biotechnol 1996;14:261–72. https://doi.org/10.1016/0167-7799(96)10037-8.
[10] Cho H, Yu Y. Link prediction for interdisciplinary collaboration via co-authorship network. Social Network Analysis and Mining 2018;8(1):1–12.
[11] Huang L, Chen X, Zhang Y, et al. Dynamic network analytics for recommending scientific collaborators. Scientometrics 2021;126:8789–814. https://doi.org/10.1007/s11192-021-04164-x.
[12] Ghanei Rad MA. Status of the scientific community in social sciences. JNOE 2006;27:27–55.
[13] Tabarzeh F. Analysis of the scientific cooperation network of university professors in social sciences. Master's thesis. Tehran: Tarbiat Modares University; 2018.
[14] Ebrahimi F, Asemi A, Shabani A, Nezarat A. Developing a prediction model for author collaboration in bioinformatics research using graph mining techniques and big data applications. Int J Integrated Supply Manag 2021;9(2):1–18. https://ijism.ricest.ac.ir/index.php/ijism/article/view/1873.
[15] Hyland K. Academic publishing: issues and challenges in the construction of knowledge. Oxford: Oxford University Press; 2016.
[16] Sonnenwald Diane H, Scientific Collaboration DH. Annu Rev Inf Sci Technol 2008;41(1):643–81.
[17] Song M, Yang CC, Tang X, Han W. Mapping the field of Bioinformatics with a content and co-authorship analysis. IEEE International Conference on Bioinformatics and Biomedicine Workshops 2012:774–81. https://doi.org/10.1109/BIBMW.2012.6470238.
[18] Grodzinski N, Grodzinski B, Davies BM. Can co-authorship networks be used to predict author research impact? A machine-learning-based analysis within the field of degenerative cervical myelopathy research. PLoS One 2021;16(9):e0256997. https://doi.org/10.1371/journal.pone.0256997.
[19] Lande D, Fu M, Guo W, Balagura I, Gorbov I, Yang H. Link prediction of scientific collaboration networks based on information retrieval. World Wide Web 2020;23(4):2239–57.
[20] Ebrahimi F, Asemi A, Nezarat A, Ko A. Developing a mathematical model of the co-author recommender system using graph mining techniques and big data applications. Journal of Big Data 2021;8. https://doi.org/10.1186/s40537-021-00432-y.
[21] Youtie J, Bozeman B. Social dynamics of research collaboration: norms, practices, and ethical issues in determining co-authorship rights. Scientometrics 2014;101:953–62. https://doi.org/10.1007/s11192-014-1391-7.
[22] Irvanizam I, Rusdiana S, Amrusi A, Arifah P, Usman T. An application of fuzzy multiple-attribute decision-making model based on simple additive weighting with triangular fuzzy numbers to distribute the decent homes for impoverished families. J Phys: Conf Ser. 2018. https://doi.org/10.1088/1742-6596/1116/2/022016 [cited 2021 Jul 15], https://iopscience.iop.org/article/10.1088/1742-6596/1116/2/022016.
[23] Cronin B, Shaw D, La Barre K. A cast of thousands: coauthorship and sub-authorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. J Am Soc Inf Sci Technol 2003;54:855–71.
[24] Höhle U. Minkowski functionals of L-fuzzy sets. In: Wang PP, Chang SK, editors. Fuzzy sets. Boston, MA: Springer; 1980. https://doi.org/10.1007/978-1-4684-3848-2_2.
[25] Cronin B. The hand of science: academic writing and its rewards. Lanham, MD: Scarecrow Press; 2005.
[26] Yu Q, Long C, Lv Y, Shao H, He P, Duan Z. Predicting Co-author relationship in medical Co-authorship networks. PLoS One 2014;9(7):e101214. https://doi.org/10.1371/journal.pone.0101214.

[27] Roemer R, Borchardt R. Meaningful metrics: a 21st century librarian's guide to bibliometrics, altmetrics and research impact. USA: ACRL; 2015, ISBN 978-0-8389-8755-1.
[28] Yeo M, Lewis M. Co-authoring in action: practice, problems, and possibilities. Iran J Lang Teach Res 2019;7(3):109–23.
[29] Green BN, Johnson CD. Interprofessional collaboration in research, education, and clinical practice: working together for a better future. J Chiropr Educ 2015;29(1): 1–10. https://doi.org/10.7899/JCE-14-36.
[30] Song M, Yang CC, Tang X. Detecting evolution of bioinformatics with a content and co-authorship analysis. SpringerPlus 2013;2(1):186. https://doi.org/10.1186/2193-1801-2-186.
[31] Chien W, Chang C, Shih C. Patterns of international coauthor collaboration in bioinformatics. Biomed J Sci & Tech Res. 2017. https://doi.org/10.26717/BJSTR.2017.01.000548.
[32] Cronin B, Shaw D, La Barre K. Visible, less visible, and invisible work: patterns of collaboration in 20th-century chemistry. J Am Soc Inf Sci Technol 2004;55:160–8.
[33] Mooney RJ, Roy L. Content-based book recommending using learning for text categorization. Proceedings of the fifth ACM conference on Digital libraries 2000: 195–204. https://doi.org/10.1145/336597.336662.
[34] Cabanac G. Accuracy of inter-researcher similarity measures based on topical and social clues. Scientometrics 2011;87:597–620. https://doi.org/10.1007/s11192-011-0358-1.
[35] Achary R. An author recommendation system using both content-based and collaborative filtering methods. master's thesis. Department of computer engineering and computer science, California State University; 2011. Available from: ProQuest Dissertations and Theses database.
[36] Sun Y, Barber R, Gupta M, Aggarwal C, Han J. Co-author relationship prediction in heterogeneous bibliographic networks. In: Advances in social networks analysis and mining (ASONAM); 2011. p. 121–8. 00044. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5992571.
[37] Ghare-Chamani J. Provide a way to suggest referrals in the referral network. Master's thesis. Sharif University of Technology, Computer Engineering Department; 2013.
[38] Makarov I, Bulanov O, Zhukov LE. Co-Author recommender system. Cham: Technologies for Network Analysis; 2017. Paper presented at the Models, Algorithms, and.
[39] Ho T, Bui Q, Bui M. Co-Author relationship prediction in bibliographic network: a new approach using geographic factor and latent topic information. In: Proceedings of the tenth international symposium on information and communication technology (SoICT 2019). New York, NY, USA: Association for Computing Machinery; 2019. p. 69–77. https://doi-org.ezp.semantak.com/10.1145/3368926.3369668.
[40] Li Xinyi, Chen Y, Pettit B, Rijke M. Personalized reranking of paper recommendations using paper content and user behavior. ACM Trans Inf Syst 2019;37(3):23. https://doi.org/10.1145/3312528. Article 31 (March 2019).
[41] Ji PS, Jin JS, Ke ZT, Li WS. Co-Citation and Co-authorship networks of statisticians. J Bus Econ Stat 2022;40(2):499–504. https://doi.org/10.1080/07350015.2022.2055358.
[42] Ullah M, Shahid A, Muhammad Roman ID, Muhammad Fayaz MA, Hanan Aljuaid YG. Analyzing interdisciplinary research using Co-authorship networks. Complexity 2022. https://doi.org/10.1155/2022/2524491.
[43] Wilkinson s, Silverman D. Focus group research. Qualitative research: theory, method, and practice. 2004. p. 177–99.
[44] Saaty TL. The analytical Hierarchy process. New York: McGraw-Hill; 1980.
[45] Baumgartner TA, Strong CH, Hensley LD. Conducting and reading research in health and human performance. New York: Mc Graw-Hill; 2002.
[46] Saffie NAM, Shukor NM, Rasmani KA. Fuzzy Delphi method: issues and challenges. In: 2016 international conference on logistics, informatics and service sciences. LISS; 2016. p. 1–7.
[47] Hu A, Chen L-T, Hsu C-W, Ao JG. An evaluation framework for scoring corporate sustainability reports in taiwan. Environ Eng Sci 2011;28(12):843–58.
[48] Zadeh LA. Fuzzy sets as a basis for a theory of possibility. Jan 1 [cited 2021 Jul 15]; 1(1) Fuzzy Set Syst 1978:3–28. Available from: https://www.sciencedirect.com/science/article/pii/0165011478900295.
[49] Franke M, Geyer-Schulz A, Neumann AW. Recommender services in scientific digital libraries. Multimedia Services in Intelligent Environments. Springer; 2008. p. 377–417.
[50] Beel J, Gipp B, Langer S, et al. Research-paper recommender systems: a literature survey. Int J Digit Libr 2016;17:305–38. https://doi.org/10.1007/s00799-015-0156-0.
[51] Davarpanah M. Investigating the compatibility of Persian article titles with their content. IRANDOC 1996;12(2):1–12.
[52] Nascimento C, Laender AH, da Silva AS, Gonçalves MA. A source independent framework for research paper recommendation. In: Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries; 2011. p. 297–306.
[53] Yan E, Guns R. Predicting and recommending collaborations: an author-, institution-, and country-level analysis. Journal of Informetrics 2014;8:295–309. https://doi.org/10.1016/j.joi.2014.01.008.
[54] Brandão Michele A, Moro Mirella M. Affiliation influence on recommendation in academic social networks. In: Proceedings of the 6th alberto mendelzon international workshop on foundations of data management. Brazil: Ouro Preto; 2012. p. 230–4.
[55] Andrikopoulos A, Samitas A, Kostaris K. Four decades of the journal of Econometrics: coauthorship patterns and networks. J Econom 2016;195(1):23–32. https://doi.org/10.1016/j.jeconom.2016.04.018.
[56] Aanonson J. Precision, and recall in title keyword searchers. Inf Technol Libr 1987; 14(3):162–70.
[57] Ferrara F, Pudota N, Tasso C. A keyphrase-based paper recommender system. In: Italian research conference on digital libraries. Springer; 2011. p. 14–25.
[58] Jiang Y, Jia A, Feng Y, Zhao D. Recommending academic papers via users' reading purposes. In: Proceedings of the 6th ACM conference on recommender systems. ACM; 2012. p. 241–4.
[59] Sugiyama K, Kan MY. Scholarly paper recommendation via user's recent research interests. In: Proceedings of the 10th annual joint conference on digital libraries. ACM; 2010. p. 29–38.
[60] Asemi A, Ebrahimi F. Research elements for identifying effective criteria for author matching in bioinformatics. Mendeley Data; 2023. p. V1. https://doi.org/10.17632/dv4pk6tbhy.1.