WILEY | Hindawi

Research Article

Analyzing Interdisciplinary Research Using Co-Authorship Networks

Mati Ullah,[1] Abdul Shahid [ID],[1] Irfan ud Din [ID],[1] Muhammad Roman [ID],[1] Muhammad Assam,[2] Muhammad Fayaz [ID],[3] Yazeed Ghadi [ID],[4] and Hanan Aljuaid[5]

[1]*Institute of Computing, Kohat University of Science & Technology, Kohat 26000, Pakistan*
[2]*Department of Software Engineering, University of Science and Technology Bannu, Bannu, Khyber-Pakhtunkhwa, Pakistan*
[3]*Department of Computer Science, University of Central Asia, Naryn, Kyrgyzstan*
[4]*Department of Computer Science, Software Engineering, Al Ain University, Al Jimi, UAE*
[5]*Computer Sciences Department, College of Computer and Information Sciences,*
 Princess Nourah Bint Abdulrahman University (PNU), P .O. Box 84428, Riyadh 11671, Saudi Arabia

Correspondence should be addressed to Muhammad Fayaz; muhammad.fayaz@ucentralasia.org

With the advancement of scientific collaboration in the 20th century, researchers started collaborating in many research areas. Researchers and scientists no longer remain solitary individuals; instead, they collaborate to advance fundamental understandings of research topics. Various bibliometric methods are used to quantify the scientific collaboration among researchers and scientific communities. Among these different bibliometric methods, the co-authorship method is one of the most verifiable methods to quantify or analyze scientific collaboration. In this research, the initial study has been conducted to analyze interdisciplinary research (IDR) activities in the computer science domain. The ACM has classified the computer science fields. We selected the Journal of Universal Computer Science (J.UCS) for experimentation purposes. The J.UCS is the first Journal of Computer Science that addresses a complete ACM topic. Using J.UCS data, the co-authorship network of the researcher up to the 2nd level was developed. Then the co-authorship network was analyzed to find interdisciplinary among scientific communities. Additionally, the results are also visualized to comprehend the interdisciplinary among the ACM categories. A whole working web-based system has been developed, and a forced directed graph technique has been implemented to understand IDR trends in ACM categories. Finally, the IDR values between the categories are computed to quantify the collaboration trends among the ACM categories. It was found that "Artificial Intelligence" and "Information Storage and Retrieval", "Natural Language Processing and Information Storage and Retrieval", and "Human-Computer Interface" and "Database Applications" were found the most overlapping areas by acquiring an IDR score of 0.879, 0.711, and 0.663, respectively.

1. Introduction

The pattern of scientific collaboration has been increasing since the end of the 20th century [1]. Scientists are working in collaboration to address various problems, such as social, political, economic, and technological issues. The collaboration among researchers builds up a communication network where they share thoughts assets and convey new learning [2]. This scientific collaboration aids in the improvement of research findings and the expansion of research quality and variety on a particular subject [3]. On the other side, an interdisciplinary collaboration includes incorporating knowledge from multiple disciplines. Participants usually originate from diverse fields and collaborate on knowledge gained from the corresponding domains to create new knowledge. Different research support programs and science policies are increasingly paying attention toward interdisciplinary research collaboration [4].

There has been abundance of literature that addresses interdisciplinary and associated concepts. Many researchers believe interdisciplinary research positively affects information creation and creativity [5, 6]. Interdisciplinary is now

fueled by several funding instruments at the national [7], international [8], and university levels. These projects seek to enable independent researchers to collaborate to foster interdisciplinarity. Interdisciplinary research (IDR) has evolved into a wide range of brushes to demonstrate a wide range of research strategies and practices [2]. Research policies and support programs are increasingly focused on IDR, claiming that more and more research initiatives are interdisciplinary [9]. Borut et al. [10] have quantified interdisciplinarity collaboration among the country's researchers. They have used the co-authorship of researchers to quantify interdisciplinarity in research communities [10]. The co-author network is their collaboration with other researchers through their publications. Co-authors network is already used in many applications such as conflict of interest [11], knowledge diffusion [12], creating a social network of the researcher [13], and finding experts [14]. Karlovcec et al. used a graph of project collaboration and co-authorship to investigate interdisciplinarity in scientific fields and their evolution [15]. One of the essential applications of co-author networks is collaborator finding. Hence, a co-author network is an important measure used in many applications, and it is easy to compute a co-author network.

In this research, we developed, built, and deployed a co-authorship network-based solution to investigate the IDR trend in the computer science field. We developed a system to prepare and handle the J.UCS dataset, which contains over 1200 research publications published by over 2500 authors in a range of ACM categories. To study the expanding IDR patterns, we created 1st and 2nd level co-authorship networks. These trials assisted us in answering questions such as what are the most common categories of 1st-level co-authorship networks where IDR activities are carried out on a regular basis, and what are the most common categories of 2nd level co-authorship networks in which IDR activities are regularly performed? These results are discussed in detail in the Result section.

2. Related Work

Bibliometric techniques permit researchers to put together their findings concerning collected bibliographic data created by researchers working in the area and express their findings through writing, citations, and collaboration. When this information is compiled and examined, bits of knowledge into the social networks, it is possible to advance the "field's structure and topical concerns." Bibliometric techniques including bibliographic coupling cocitation analysis create underlying pictures of scientific fields using bibliographic data from publication databases. They add objectivity to the evaluation of scientific literature and can be used to differentiate between implicit research networks. For example, there are "invisible colleges" under the surface that are not formally related but having same research interests as each other. These organizations have similar scientific priorities and keep in touch through seminars, staff correspondence, and private summer schools. The "authors" judgments on the subject matter, methods, and the importance of other "authors" work are reflected in the

images cited from the study fields, which have been added over time [16].

These bibliographic methods have two main applications: results estimation and scientific visualization [17]. Performance analysis aims to evaluate the effectiveness of individual and organizational research and publication efforts. On the other hand, scientific visualization seeks to explain the nature and complexities of scientific areas. If the researcher's goal is to review a particular line of study, this insight on structure and growth may be helpful. The Bibliography method adds methodological rigor to the subjective assessment of the literature. In the review paper, they can include proof of logically derived categories.

We briefly describe the five most popular bibliographic methods in the section below. Citation analysis and cocitation analysis are the first two approaches or processes that use citation data to create measurements of effect and relations. Co-authorship data are used to assess collaboration in co-authorship research. On the other hand, the co-word study looks for links between ideas and theories that appear in document titles, keywords, or abstracts. Table 1 summarizes bibliometric methods, along with their strengths and weaknesses.

2.1. Bibliometric Methods

2.1.1. Citations. The top N list of the most cited research, author, or journal in the field of interest is typically provided by most bibliographic studies in citation analysis of research fields. Citations are used to determine the degree of impact. It is considered significant if an article is most often cited. The author references a paper relevant to his work, so this suggestion is based on that assumption. Citation review may provide information about the relative importance of publications, but it cannot understand researchers' networks [18].

2.1.2. Cocitation. Another bibliometric method, cocitation, measures how many papers are cited simultaneously in the same article. This indicator reflects the influence and impact of thematic networks and authors. However, in the final analysis, the methods of cocitation represent the responses and reactions of the scientific community to the research results. The cocitation clusters provide a complementary description of similar and related research topics and related studies measured by citations. It may also be possible to map and identify the researcher's community within a specific network. Such clusters also show how fields and subfields evolve [3].

2.1.3. Coanalysis. Co-word is a content analysis technique that builds relationships and makes the conceptual structure of the domain by using words in documents. The underlying idea behind this method is that the frequent cooccurrence of words in a document indicates a close relationship among the concepts behind words.

TABLE 1: Summary of bibliometric methods.

#	Methods	Description	Units of analysis	Strengths	Limitations
1	Citation	Citation rates evaluate the impact of documents, authors, or journals	Author document journal	Important work in the field can quickly be found	Since newer articles have little time to be referenced, citation count as a metric of impact is weighted against older publications
2	Cocitation	Connect journals, documents, and authors, based on joint presence in the reference list	Author document journal	The most widely used and validated bibliometric tool for linking authors, articles, and journals is cocitation. It is considered reliable. It filters the most important works.	Since it is conducted on cited papers, cocitation is not ideal for mapping research fronts. Since citations take time to accumulate, new publications can only be linked by knowledge base clusters. Since multiple citations are needed to map an article, it is not possible to map articles that are rarely cited.
3	Co-word	It connects keywords if they seem in a similar title, abstract, or keyword list	Word	It analyses documents based on their content. Many other methods, on the other hand, depend on metadata.	It's possible that the word will appear in multiple contexts and take on different meanings
4	Bibliographic-coupling	Connects the journals, documents, and authors based on a number of mutual references	Author document journal	.It does not need a citation to accrue. It could be used for newer publications that are not cited yet.	Can use only for a short timeframe (for the interval of five years). The most critical works are not even mentioned. Even it has trouble determining whether or not the mapped publications are relevant.
5	Co-author	When two or more authors collaborate on a document, it connects both	Author	It produces the social structure of fields and can provide an indication of collaboration	Author name disambiguation issues arise in the co-authorship network

Other methods indirectly connect the documents through coauthorships or citations, while in the case of co-word analysis, it constructs similarity measures by using the actual content of the papers. The network of themes and their relationships, which reflect the field's conceptual space, is the product of a co-word analysis. Co-word analysis may apply to entire documents, abstracts, keywords, and paper titles. However, the accuracy of the co-word analysis results depends on multiple aspects, such as the quality of keywords, the complexity of the statistical methods used for analysis, and the scope of the database used [3].

There are two possible reasons for the concern when using just the keywords for a co-word analysis. The first explanation is that often the journal's bibliographic data do not contain keywords. The second is that depending solely on keywords undergoes an indexer effect. The chart's validity depends on whether the indexer collects all the specific facets of the text. The alternative to this issue is to use entire texts or abstracts, but this indicates noise in the data as algorithms have difficulty separating the importance of terms in vast corpora of text [18].

2.1.4. Bibliographic Coupling. Bibliographic coupling is ignored, and the process is years older than cocitation [18]. Bibliographic coupling tests the compatibility between the two texts by using mutual references. Furthermore, if bibliographies overlapped in the two papers, the greater their relationship would be. Between the two articles, the number of references remained static over time, as the number of references stayed unchanged throughout the paper. However, cocitation-based affinity grows with citation trends. When citation ways shift, bibliographical coupling works well within a short timeframe [19]. The distinction between bibliographic coupling and cocitation is that a bibliographic coupling relation is formed by the paper's authors also in focus. In contrast, a cocitation connection is established by the scholar citing the works under consideration.

When two documents are heavily cocited, it suggests that each paper is highly cited independently [20]. This demonstrates that documentation chosen based on cocitation thresholds is more valuable from the researcher's perspective when quoting them. However, since bibliographic coupling cannot be used in this manner, identifying essential documents within many documents is a difficult challenge when doing bibliographic coupling. Otherwise, the bibliographic coupling is beneficial for scientific mapping boundaries and new areas lacking citation evidence or smaller subfields where cocitation analysis cannot generate accurate relations [21, 22]. In Figure 1, the distinction between bibliographic coupling and cocitation analysis is visually depicted.

2.1.5. Co-Authorship. The co-authorship measures scientific interaction and relationships amongst networks, teams, institutions, and countries. The joint publication results from a collaboration between organizations and representatives from different countries participating in a research

Citing documents (research front)
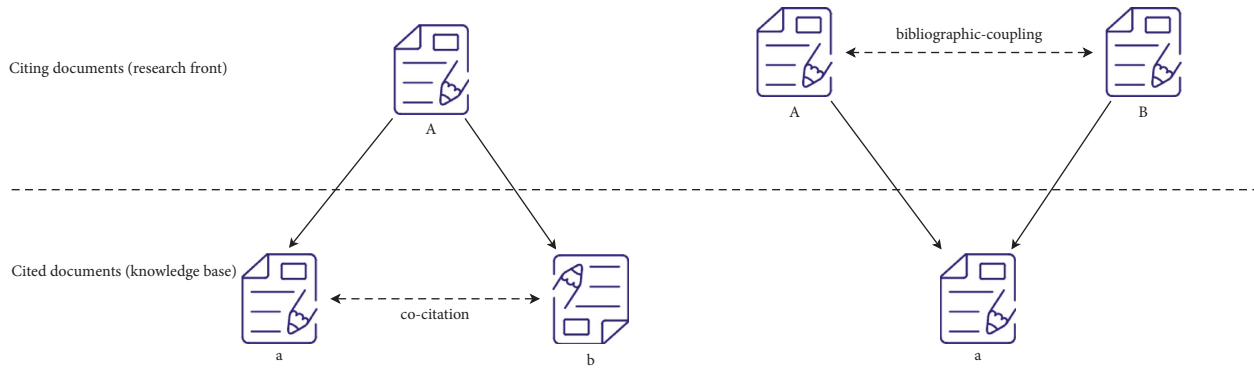
Cited documents (knowledge base)

FIGURE 1: Bibliographic coupling and co-citation analysis.

program. Such research establishes relationships between teams (scientists, laboratories, institutes, and countries) to produce a scientific article. Co-authorship can identify, measure, and display the number of links established by individual contributors. Thus, co-authorship can be used to indicate these relationships. Following this principle, you can construct a matrix where each cell shows the number of cosignatures between the author (or authors) displayed in the rows and the author (or authors) indicated in the column. This indicator can identify key research partners and describe scientific networks descriptions [3]. Coauthorship is a credible metaphor for scientific collaboration among the various bibliometric methods discussed above. The coauthorship network has many uses, and scholars have used it in their research studies. Macro and micro features of massive co-authorship networks using SNA techniques are investigated in [23]. The dynamics and evolution of co-authorship networks were studied in [4], which followed up on "Newman's 2001 work." Since then, co-authorship networks have been widely researched in various ways in both the natural and social sciences [24].

Taskn et al. [25] analyzed co-authored astrobiology papers and also analyzed journal references. By studying topological configurations of co-authorship networks, Pavlov et al. [14] discovered essential functional knowledge characteristics of the characteristics of scientific collaboration. A systematic model of cumulative benefit in terms of preferential attachment as the guiding force of co-authorship was investigated by Barabasi and Albert [26]. They noticed a common property across several large networks: the scale-free power-law distribution is followed by the node degrees of each network. The effects of scale-free distributions have been extensively used to explain scientific co-authorship networks.

Morel et al. [27] used a graph of project collaboration and coauthorship to investigate interdisciplinary scientific fields and their evolution. Porter [28] stared into the impact of collaborative research in the academic finance literature and discovered that it could result in high-impact articles even though it was found that interdisciplinary collaborations have a higher potential for fostering research outcomes. When two authors collaborate on a study, this is known as co-authorship. It is one of the most visible and well-documented forms of scientific collaboration. By

analyzing co-authorship networks using bibliometric methods, almost every aspect of scientific collaboration networks can be reliably tracked. These networks of collaborations (co-authorship) reveal research teams, as well as factors that influence the impact or output of collaborations. According to our research requirements, we have found co-authorship network methodology to be the best and most reliable method to implement in our research methodology.

Co-authorship is among the most effective methods of scientific collaboration among the various bibliographic methods mentioned above. The co-author has many applications, and many researchers have used them in their research studies. Porter et al. [28] investigated large co-author networks' macro and micro characteristics using SNA methods. Co-author networks have since been extensively studied in various ways in both the natural and social sciences. Taskn et al. [29] reviewed journal references and co-authored publications in the field of astrobiology. By studying the topological configurations of the coauthor network, Huang et al. [30] discovered essential practical knowledge features of the research collaboration method. In co-authorship, Abramo et al. [31] established small-world systems. The accepted model of cumulative benefits in terms of preferential attachment as a guiding force for co-authorship was studied by Hennemann et al. [24]. They discovered a common property of large networks and each node degree followed a scalable power less distribution. Scientific co-author networks have been analyzed thoroughly using the results of scale-less distributions. Qin et al. [32] investigated interdisciplinary research fields and their growth using the project partnership and co-authorship graph. Figg et al. [33] studied the effects of collaborative research in academic finance and discovered that collaboration results in high-impact articles.

On the other hand, interdisciplinary collaboration has been shown to have the ability to increase research outcomes. According to the study, when two authors work on research, they are known as co-authors, having the most concrete and well-documented form of research cooperation. By analyzing co-"authors" networks using bibliometric techniques, almost any component of scientific collaboration networks can be accurately examined. These collaboration networks (co-authorship) show the impact of co-authorship, research teams, and collaboration output.

FIGURE 2: The link of an author X with 1st-level co-author Y.
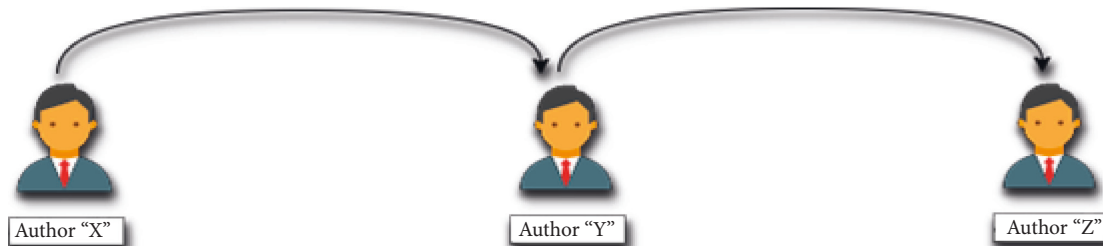

FIGURE 3: The link of an author X with 2nd level co-author Z.

According to our research requirements, we have concluded that the co-authorship network methodology is the best suited and most reliable method for implementing our research methodology. A co-authorship network could have various levels, as explained in the previous chapter. In our research, we have used a co-authorship network up to the second level described in the next section.

2.2. Levels of Co-Authorship Network. Co-authorship is viewed as a valid indicator for scientific collaboration in research publications. Since the 1960s, using co-authorship to assess research collaboration has been a hot topic. Research collaboration may accumulate numerous assistances for researchers to give scientific credibility from bringing different talents together [8]. One of the most concrete and well-known methods of scientific collaboration is co-authorship. Almost any component of research collaboration networks can be accurately traced by studying co-authorship networks. Co-authorship is a form of collaboration where two or more authors publish a paper, and these authors are connected to form a co-authorship network [9].

Thousands of authors can be linked together in co-authorship networks, with the best example being the "Paul Erdos" network, which has over 500 co-authors. An author who has published with Erdos has an Erdos number of 1. Those who have published with Erdos as a co-author have an Erdos number of 2, and so on in the "Paul Erdos" network [19]. However, this reveals that in a co-authorship network, there are many levels of co-authors.

2.2.1. First Level Co-Author. As previously mentioned, there are various levels of co-authors in a co-authorship network. Consider the following scenario: author X has co-authored a research paper with another author, Y, and Y is the first level co-author of X. The concept of a first-level co-author is clearly explained in Figure 2.

2.2.2. Second Level Co-Author. In the second level co-author, we can consider a scenario in which author X has published a research paper with another author Y. As explained above, author Y is at the 1st-level. However, if author Y has published a research paper with another author, Z, then author Z will be the 2nd level co-author of X. Figure 3 clarifies the current second-level co-author scenario.

Due to the advantages and availability of bibliometric data, co-authorship measures interdisciplinary, as co-authorship is commonly used as a metaphor for collaboration in science.

3. Proposed Methodology

The suggested approach is discussed in this section. Figure 4 illustrates the architecture of the current methodology. It consists of several steps, such as dataset selection of research articles, extraction of research publications, and formation of co-authorship networks that will be analyzed to quantify interdisciplinary scientific communities. However, the section is divided into subsections in detail for further clarification.

3.1. ACM Classification. The Association for Computing Machinery (ACM) is the world's largest computing society, putting together experts, scholars, and educators to exchange expertise, promote debate, and solve the field's challenges. ACM has over 100,000 participants all over the world. ACM provides opportunities for career development and professional networking to support the professional growth of its members. The ACM classification scheme is polyhierarchical, containing the list of topics available from topic A to topic M. The topics from A to K are ACM's classification and its subclassification, while other two topics were added to reveal the growth of the computer science discipline. The complete list of topics is given in Table 2.
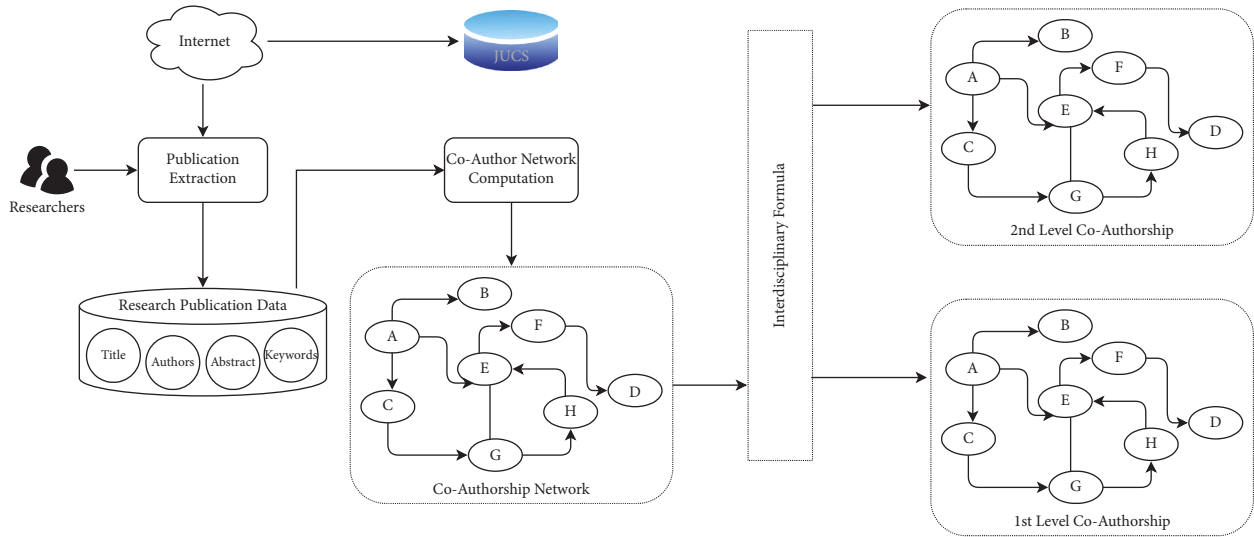
FIGURE 4: Publication information aggregation framework.

TABLE 2: Thirteen Categories following ACM classification (https://www.jucs.org/jucs_info/acm_categories).

Topic A-General Literature	Topic H-Information System
Topic B-hardware	Topic I-computing methodology
Topic C-computer system organization	Topic J-computer applications
Topic D-software	Topic K-computer milieux
Topic E-data theory	Topic L-science and technology of learning
Topic F-theory of computation	Topic M-knowledge management
Topic G-mathematics of computer	

3.2. Comprehensive Dataset Selection.

The dataset collection criteria for our proposed approach are as follows: (1) we required a large enough dataset to complete our research. The chosen dataset should cover a broad range of topics; (2) the second primary criterion for our dataset was that it should enable us to access the metadata of numerous authors' research articles and information about the authors' publication records and co-authors' information. We choose the dataset from the (J.UCS) journal of Universal Computer Science to satisfy these criteria. J.UCS is the first Journal of Computer Science that addresses various topics, where authors from different backgrounds and domains publish their research.

So, the J.UCS dataset will help us comprehensively explore our proposed research. The dataset of J.UCS is denoted at the top of Figure 4.

3.3. Extraction of Metadata.

Metadata of research papers is another source that the researchers use. Metadata is defined as data about the data. In the research articles context, the metadata could be the author, keywords, paper title, and ACM topics (if any). We believe that metadata could be divided into two categories: the traditional metadata, including the data about research articles. The second type of metadata is acquired from bookmarking and social tagging. The user could annotate the data using online services like CiteULike, which contains research articles, references, and bookmarks.

Metadata techniques have some limitations because they are usually dataset-dependent and cannot be generalized; for example, metadata of one type in a dataset may not exist in another dataset. Generally, the title of a paper, author, and publication information are available in each dataset. But, sometimes, this information is too little to compute the relatedness between the research articles. Occasionally, the free availability of metadata is not possible. In this type of situation, the metadata is automatically extracted.

Our database contains various tables, like papers and categories shown in Figure 5. Each paper in the relationship can have more than one category; therefore, we have added a third relation called "papers categories" as a join table. These tables contain metadata about papers, authors, categories, and subcategories. Our first step consists of extracting the metadata of papers, authors, categories, and subcategories using a crawler. The crawler, developed in PHP, crawls the pages of JUCS through a service. It looks for a specified structure of the content of the web page, containing the paper metadata and pdf contents. We directly store the media contents to their relevant tables in the database, whereas the pdf files are further converted to XML as it is nearly impossible to read the sections of a pdf document, explained in the next section.

3.3.1. Extraction of Authors.

The author information was extracted from XML formats of the paper containing the author's name and first- and second-level author's
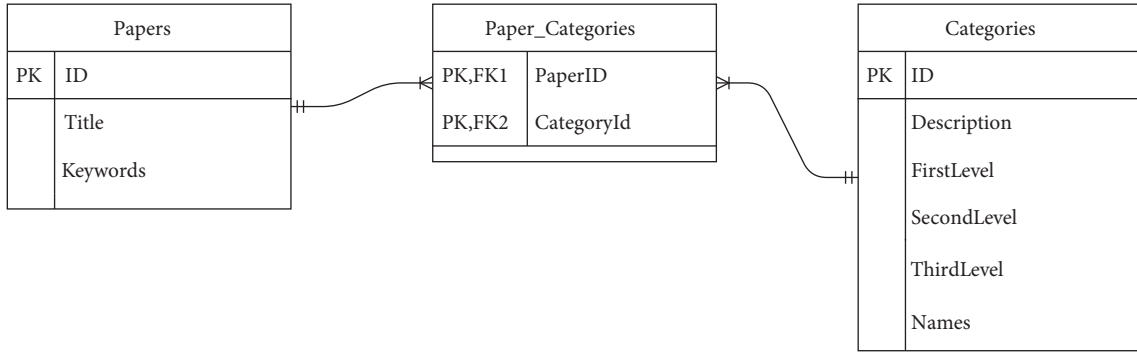
Figure 5: Pictorial presentation of a list of three tables.

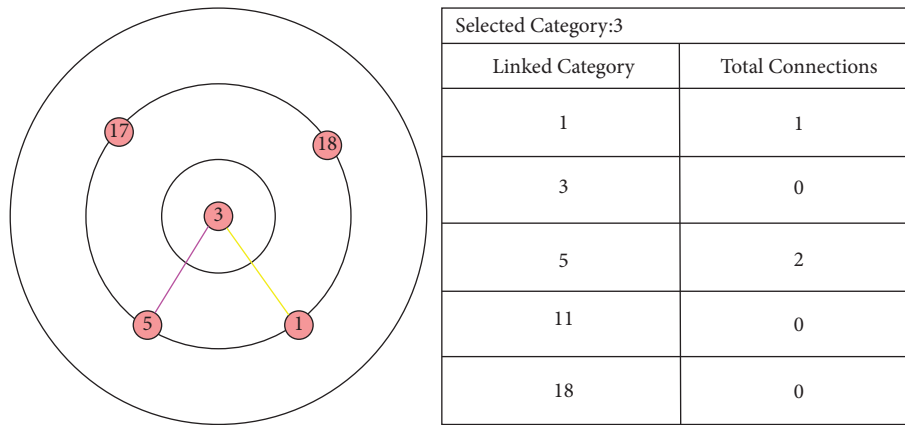| Selected Category:3	
Linked Category	Total Connections
1	1
3	0
5	2
11	0
18	0

Figure 6: A screencast of force-directed graph.

information. We have converted the pdf format of papers to XML because it is nearly impossible to extract the author's information from pdf files. A separate author's table has been formed in our database. The table contains the author's id, author's name, and his co-author's information.

3.3.2. Extraction of Papers.
The information regarding papers has been extracted from XML files and stored in organized form in a separate table. The paper table contains the paper's id, title, keyword, and the abstract. Paper information is also metadata that must be extracted as part of the researcher's publication record. Figure 6 has highlighted the papers table containing the papers' information.

3.3.3. Extraction of Papers Categories.
The metadata in our database also contain the categorized information. Thirteen categories have 420 subcategories in the database following the ACM classification scheme. We created a category table that includes information about the list of categories in our database. We also created a paper-category table that indicates the specific category of paper in which the article resides. The category information of a paper is extracted from the paper-categories table. How our database has metadata containing the complete information of papers, authors, articles, and categories is summarized. This metadata helps us in building co-authorship networks of different categories.

3.4. Building Co-Authorship Network.
The co-authorship network explains how authors have been associated with each other from various fields of research based on their published articles. The network is considered one of the most credible and concrete methods for describing the author's collaborations [8]. A co-authorship network can extract any research component by studying the links among various network nodes [9]. The Paul Erdos network is one of the examples, with over 500 co-author nodes in the network [19]. The network shows all the authors who have worked directly or indirectly with the Hungarian mathematician Paul Erdos, who wrote many research articles in mathematics. The network is based on Erdos number, which describes the collaboration with Paul Erdos. If an author has published an article with Erdos as a co-author, the assigned Erdos number is 1. Authors who have an association with the co-authors of Erdos are assigned Erdos number 2, and so on. The authors with more publications with the same Erdos number are given preference while fetching the n number of collaborations. This illustrates that in a co-authorship network, there are several levels of co-authors. As explained in the literature review section, we have constructed a co-authorship network up to the 2nd level.
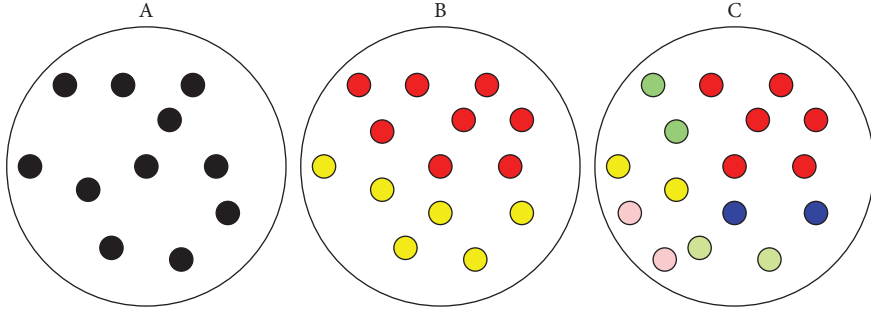
FIGURE 7: Groups of researchers bellowing to various categories represented by colors.

3.5. System Development. System development was our main focus for the completion of our methodology and for achieving our research goal. We had to construct co-authorship networks of authors belonging to different categories and then analyze the interdisciplinary nature of each category with other categories. To accomplish our goal, we have used the tool Visual studio 2019. Visual studio is an open-source IDE used to develop web apps, mobile apps, and computer programs. We used the ASP.net language, and the framework used in development was MVC. We have also created a database, and the formation of the database is done using MySQL.

Using the J.UCS dataset, we have constructed co-authorship networks of researchers belonging to any category up to the 2nd level. Two categories are considered to be connected by an edge if the researcher of one category has co-authored at least one paper with another researcher of any other category. However, the co-authorship networks amongst different categories are set to form scientific communities.

In our database, thirteen different categories have a total of 421 subcategories. These categories follow the ACM classification. Each researcher registered in a database has assigned a number from 1 to n, where n is equal to 421, the last category of our dataset.

The number assigned to any researcher indicates the category to which an author belongs. Some of the top categories are general literature, hardware, computer system organization, software, data theory, Theory of Computation, information system, computing methodology, computer applications, computer milieux, Science and Technology of Learning, and knowledge management.

To measure the interdisciplinary of each community C formed among a group of categories, we use the ACM classification scheme described above. The n-component vector IC has been allocated, with each component reflecting the fraction of researchers in one of the "n" categories. (1) defines the interdisciplinary amongst different categories in scientific communities:

$$IRC(C) = \beta \sqrt{1 - \sum_{i=1}^{n} Xi^2}. \qquad (1)$$

Here, the ith component of IC is $[\![Xi]\!]^2$ and $\beta = [1 - (1/n)] - 0.5$, which is the normalization constant that ensures that $0 \leq IDR(C) \leq 1$. Rendering to equation (1) the IDR(C) =

0 if $[\![Xi]\!]^2 = 1$ for any n components (in this case, all the other components are 0). If the value of n is equal to 7, the $[\![Xi]\!]^2$ is 1/7, then the interdisciplinary IDR(C) is 1.

Researchers belonging to various groups are illustrated in different colors in Figure 7. There are seven categories, represented by Xi, where I = 1, 2, 3, . . ., and 7. Each category is represented by different color denoted as $n = 7$ in Equation 1. In A, all the researchers belong to the same category or work in the same research area X1. The n-component vector will be IC = (1, 0, 0, 0, 0, 0, 0), as well as the community's interdisciplinary, according to equation (1), is IDR(C) = 0.

In community B, the researchers belong to two different research areas $X1$ and $X5$. So, the n-component vector will be IC = (1/2, 0, 0, 0, 1/2, 0, 0), and the interdisciplinary of such a community is, according to equation (1), is IDR(C) = 0.88. In community C, all the researchers work in different research areas from $X1$ to $X7$. So, the n-component vector will be IC = (1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7), and such a community's interdisciplinary, according to equation (1), is IDR(C) = 1.

Similarly, we have analyzed the interdisciplinary of all categories in our database with other categories up to the 2nd level. We made a co-authorship network of each category up to the 2nd level. For visualization purposes, we have used some visualization libraries that show the co-authorship network of the categories is visualized, as shown in Figure 8.

3.6. Visualization. The visual presentation of co-authorship networks of categories was also important; we used several visual libraries. There are two different graphs of our visualization process and both show the categories' relatedness, but one of the graphs has more information about the connection(s).

3.6.1. Force-Directed Graph. The force-directed graph shows the relatedness of categories and, compared to the category graph, the force-directed graph shows a table that has two columns and contains the information of linked categories and the total number of connections of authors amongst different categories. As shown in Figure 6, the force-directed graph also presents the list of categories in our database.

3.6.2. Categories Graph. The category graph, shown in Figure 9, represents the categories' network information, showing the links amongst categories. The category graph
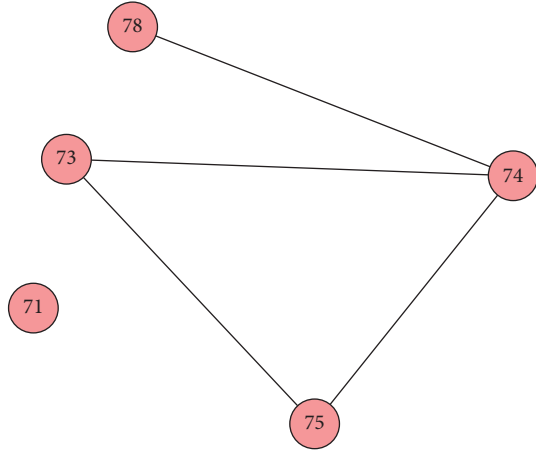
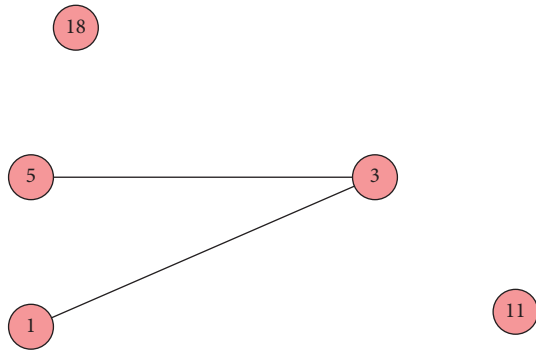Figure 8: Co-authorship network of five different categories.

Figure 9: A screencast of category graph.

Table 3: List of categories, sub-categories, authors, and paper per categories.

Categories	Number of sub-categories	Authors per category	Papers per category
General literature	5	168	80
Hardware	56	410	144
Computer system organization	29	771	300
Software	48	2567	1018
Data theory	08	271	106
Theory of computation	27	1539	809
Mathematics of computer	24	423	211
Information system	43	3370	1270
Computing methodology	75	1706	626
Computer applications	10	484	186
Computer milieux	43	1040	425
Science and technology of learning	42	103	33
Knowledge management	11	73	26
Total:	421	12,925	5,234

does not contain any information about the connection of authors amongst categories.

4. Results

This section contains the complete results and information about many categories and subcategories, paper(s) per subcategories. It also clarifies the results at different levels of the co-authorship network amongst authors. The strong bond between any two categories depends upon how many authors of one category are linked with the authors of another category. We have further analyzed the interdisciplinary value of categories in a community.

Table 3 defines the information regarding the whole scenario of our database. The table provides information about categories and subcategories. The table also provides the number of authors in a particular category and the papers in each category.

The Pie Chart in Figure 10 shows the number of categories starting from A to M, which are 13 and subcategories that are 421 in number. The different colors show the total number of categories. At the same time, the subcategories are represented in the form of a percentage.

4.1. Results of Co-Authorship Network at First Level. Our dataset contains four hundred categories, including
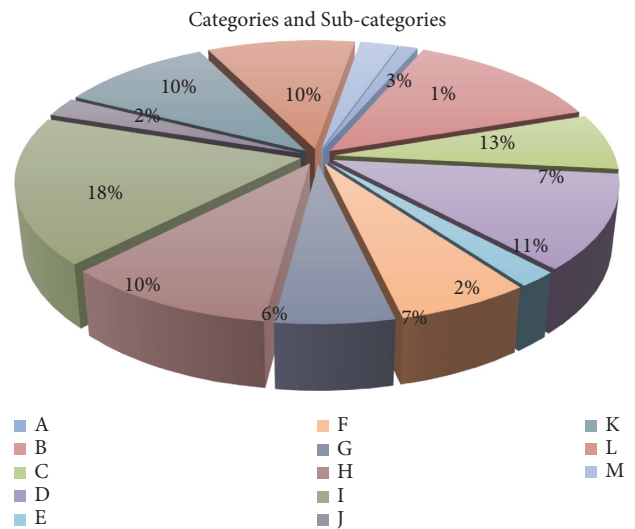

Figure 10: A pie chart of categories and subcategories.

subcategories. Twenty-one categories follow the ACM classification scheme as explained in the previous section. We have selected ten different subcategories to form a co-authorship network at first level to analyze interdisciplinarily, as shown in Figure 11. The subcategories include arithmetic and logic structure, control structure performance analysis and design aids, reliability, testing, fault tolerance, design styles, design, network architecture design, visual programming, management of computing and information systems, and storage management. The combination of linking subcategories will create a mesh connecting
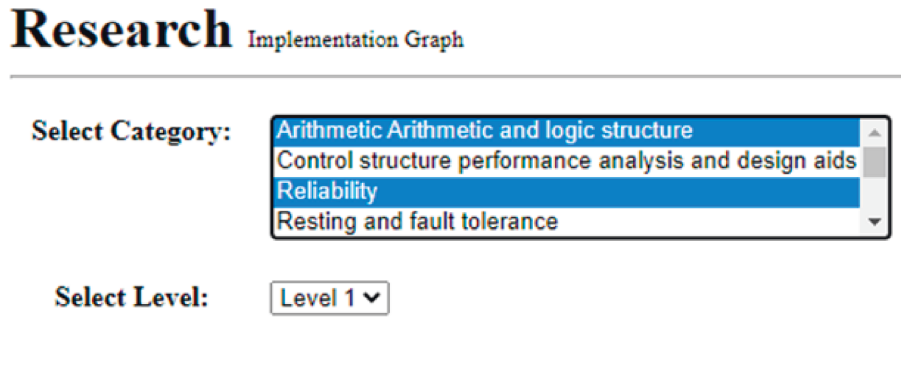
Research Implementation Graph

Select Category:
- Arithmetic Arithmetic and logic structure
- Control structure performance analysis and design aids
- Reliability
- Resting and fault tolerance

Select Level: Level 1

FIGURE 11: A screencast of selecting ten categories from the list of 421 subcategories to form a co-authorship network at the first level.
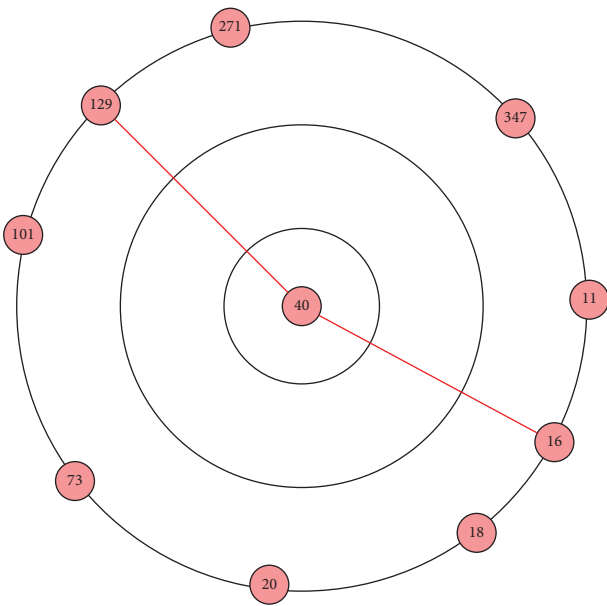

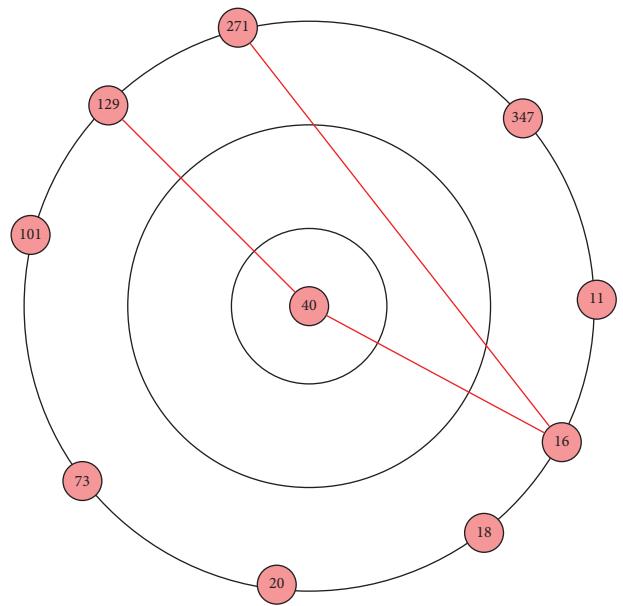FIGURE 12: A screencast of force-directed graph.

FIGURE 14: A screencast of the force-directed graph.

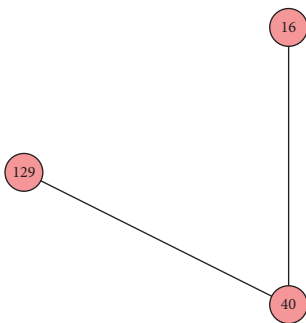FIGURE 13: A screencast of category graph.

TABLE 4: Connected categories represented in the directed graph at the 1st-level co-authorship network.

Selected category: 40	
Linked category	Total connections
11, 18, 20, 40, 73, 101, 129, 271, 347	0
16, 129	1

TABLE 5: Connected categories represented in the directed graph at 2nd-level co-authorship network.

Selected category: 40	
Linked category	Total connections
11, 18, 20, 40, 73, 101, 271, 347	0
16	5
129	2

each subcategory with every other subcategory; therefore, we have selected only ten subcategories. The pictorial presentations of the force-directed graph and category graph have been shown in Figures 12 and 13, respectively, describing the co-authorship network formed at the first level amongst these categories.

4.1.1. Force-Directed Graph Results. Fore directed graph is a particular type of graph in which nodes have forces applied to them. The connected nodes pull towards one another
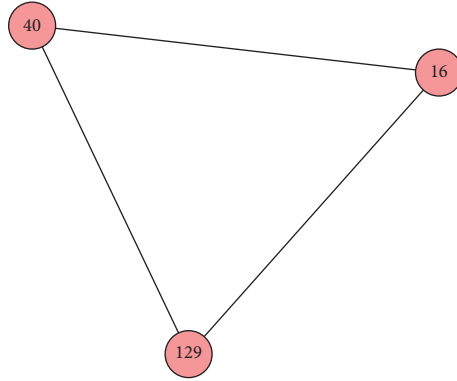
FIGURE 15: A screencast of the category graph.

TABLE 6: List of categories, sub-categories, authors, and paper per categories.

S.No	Categories from description	Categories to description	Connections	Interdisciplinarity value
1	Artificial intelligence	Information storage and retrieval	49	0.879
2	Natural language processing	Information storage and retrieval	33	0.711
3	System	Coding tools and techniques	30	0.679
4	Human-computer interface	Database applications	29	0.663
5	Logic design	Language constructs and technique	25	0.611
6	Robotics	Software engineering	21	0.572
7	Programming languages and software	Processor architectures	20	0.56
8	Computer-communication networks	Object oriented programming	18	0.543
9	System and information theory	Content analysis and indexing	16	0.521
10	Game-based learning	Software architecture	15	0.513

while nodes that are not similar repel each other. As a result, it will group nodes identical to each other based on the co-authorship, in our case. Figure 12, based on Table 4, is a force-directed graph showing the first level co-authorship network amongst ten different subcategories, where subcategory id 40 in the categories list has a connection and is linked with subcategory id 139 and 16. This means that only these three sub-categories have co-authored papers among the ten different subcategories. The table shown below represents the linked categories and total connections. The rest of the subcategories have not published a single paper with each other.

The column of total connections in the above table represents the number of co-authored papers amongst subcategories. The subcategory id # 40 has co-authored a paper only with subcategory id # 16 and subcategory id # 129. In contrast, the rest of the subcategories are unconnected because none of the remaining subcategories have co-authored a single paper with subcategory id # 40.

4.1.2. Category Graph Results. The Category graph in Figure 13 shows only the connected subcategories, leaving the other unconnected subcategories. The connection between these subcategories indicates that the authors from these three subcategories have co-authored paper(s).

4.2. Result of Co-Authorship Network at Second Level. The same ten different subcategories from the list of four hundred and twenty subcategories are selected to form

a co-authorship network at the 2nd level. The results are shown in Figure 11.

4.2.1. Force-Directed Graph Results. The force-directed graph in Figure 14 shows that category 40 is connected with categories 129 and 16, but in the second level co-authorship network, the total connection column in Table 5 shows that the authors of these categories have published more than one paper with each other.

4.2.2. Category Graph Results. The category graph only shows the related categories. Figure 15 is a co-authorship network up to the 2nd level that indicates a subcategory's connectivity with the other ten subcategories selected in the network.

We have found and shown in Table 6 the top 10 most collaborating subcategories among the group of all 421 subcategories in our database. The ten subcategories shown in the table have the highest number of connections with a specific category that indicate their scientific collaboration, as well as mark the most collaborating subcategories in a scientific community. Furthermore, the interdisciplinary values between the subcategories have been analyzed and shown in Table 6.

When examining subcategory collaboration, it was discovered that subcategories with the strongest association or the most significant number of connections had key term correlations, having the highest number of correlations among the words of these categories. As an
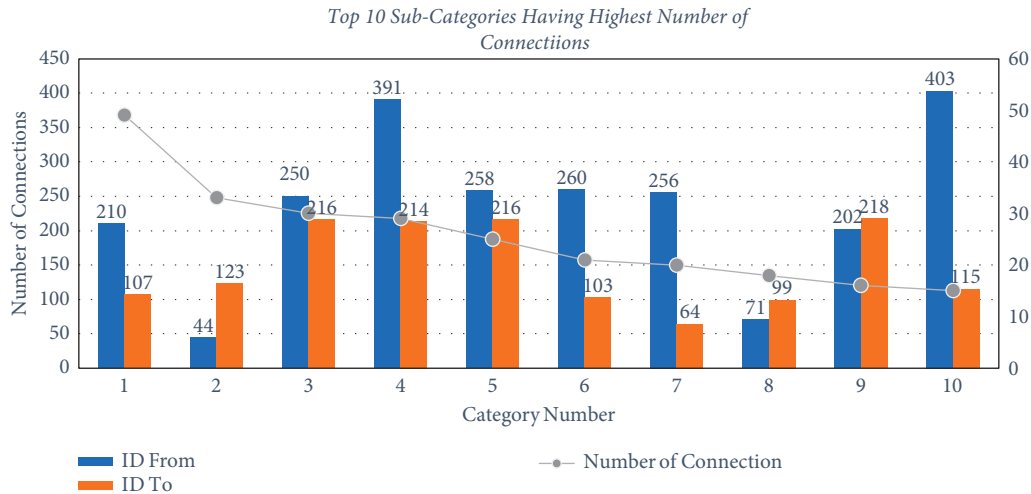
FIGURE 16: Graph of ten sub-categories having the highest number of connections.

example, the subcategory Artificial Intelligence has a close relationship with the subcategory information storage and retrieval because Artificial Intelligence has the most connections with information storage and retrieval relative to other subcategories in the co-authorship network. The same holds for all subcategories with the most links to the specific subcategory. Figure 16 displays the number of connections for each of the subcategories. It clearly shows the number of times a category is used as the source and destination.

Various methods in literature have performed the IDR analysis, but our technique has used the co-authorship network for performing IDR analysis in terms of authors' relationships from multiple disciplines.

5. Conclusion

Scientific collaboration is a dire need of today's research. Interdisciplinary collaboration is very common, and researchers from other disciplines collaborate to solve complex issues whose solutions go beyond a particular category or domain.

In this research, we designed, developed, and deployed a co-authorship network-based solution to analyze the trend of IDR in the computer science domain. We developed a system that prepared and processed the J.UCS dataset in this research. The dataset was persisted in MySQL. The dataset comprised more than 1200 research articles published in various ACM categories by authors of more than 2500. This dataset enables us to perform the experiments to answer the following research questions: Are there any IDR activities going on in various computer science fields? What are the most common categories w.r.t 1st-level co-authorship network where IDR activities are frequently conducted? and What are the most common categories w.r.t 2nd level co-authorship network where IDR activities are frequently conducted? Finally, the directed graph visualization is implemented to quickly understand the IDR activities in the computer science domain.

The growing trend of scientific collaboration has opened new avenues of research. This research may help us to understand what special training is needed for new researchers in any area. What are the different areas affecting each other? So, this kind of information can help us to design curricula for specific programs. Finally, this research may help us make a wise decision for allocating resources. In this research, we have used the J.UCS dataset related to the computer science field. In the future, it would be interesting to conduct a study that can analyze interdisciplinary phenomena among different disciplines like physical sciences, numerical sciences, and social sciences using a dataset covering a wide area of scientific disciplines.

Data Availability

The dataset was taken from the open-access journal (Journal of Universal Computer Science) https://www.jucs.org/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. J. Acedo, C. Barroso, C. Casanueva, and J. L. Galan, "Co-authorship in management and organizational studies: an empirical and network analysis," *Journal of Management Studies*, vol. 43, no. 5, pp. 957–983, 2006.

[2] B. P. Fonseca, R. B. Sampaio, M. V. Fonseca, and F. Zicker, "Co-authorship network analysis in health research: method and potential use," *Health Research Policy and Systems*, vol. 14, no. 1, pp. 34–10, 2016.

[3] L. Aldieri, M. Kotsemir, and C. P. Vinci, "The impact of research collaboration on academic performance: an empirical analysis for some European countries," *Socio-Economic Planning Sciences*, vol. 62, pp. 13–30, 2018.

[4] A. L. Razera, M. R. Errera, E. D. Dos Santos, L. A. Isoldi, and L. A. O. Rocha, "Constructal network OF scientific publications, CO-authorship and citations," *Proc Roman Acad Ser A-Math Phys Tech Sci Inform Sci*, vol. 18, pp. 105–110, 2018.

[5] S. Kumar, "Co-authorship networks: a review of the literature," *Aslib Journal of Information Management*, 2015.

[6] Y. Chen, C. Ding, J. Hu, R. Chen, P. Hui, and X. Fu, "Building and analyzing a global co-authorship network using google scholar data," in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 1219–1224, Perth, Australia, April 2017.

[7] M. Pavlov and R. Ichise, "Finding experts by link prediction in co-authorship networks," *FEWS*, vol. 290, pp. 42–55, 2007.

[8] S. Hennemann, "Hierarchies in the science and technology system of China-System reforms and their consequences for knowledge flows," *Geographische Zeitschrift*, vol. 98, no. 3, pp. 155–174, 2010.

[9] C. M. Morel, S. J. Serruya, G. O. Penna, and R. Guimarães, "Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases," *PLoS Neglected Tropical Diseases*, vol. 3, no. 8, p. e501, 2009.

[10] A. L. Porter, D. J. Roessner, and A. E. Heberger, "How interdisciplinary is a given body of research?" *Research Evaluation*, vol. 17, no. 4, pp. 273–282, 2008.

[11] I. Rafols and M. Meyer, "Diversity measures and network centralities as indicators of interdisciplinarity: case studies in bionanoscience," *Proceedings of ISSI*, vol. 2, pp. 631–637, 2007.

[12] M.-H. Huang and Y.-W. Chang, "A study of interdisciplinaryin information science: using direct citation and co-authorship analysis," *Journal of Information Science*, vol. 37, no. 4, pp. 369–378, 2011.

[13] G. Abramo, C. A. D'Angelo, and F. Di Costa, "Identifying interdisciplinarythrough the disciplinary classification of co-authors of scientific publications," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 11, pp. 2206–2222, 2012.

[14] J. Qin, F. W. Lancaster, and B. Allen, "Types and levels of collaboration in interdisciplinary research in the sciences," *Journal of the American Society for Information Science*, vol. 48, no. 10, pp. 893–916, 1997.

[15] J. W. Grossman, "Paul Erdős: The Master of Collaboration," *The Mathematics of Paul Erdős II*, Springer, Manhattan, NY, USA, pp. 489–496, 2013.

[16] A. H. Marino, K. A. Suda-Blake, and K. R. Fulton, "Innovative Collaboration Formation: The National Academies Keck Futures Initiative," *Strategies for Team Science Success*, Springer, Manhattan, NY, USA, pp. 241–250, 2019.

[17] W. D. Figg, L. Dunn, D. J. Liewehr et al., "Scientific collaboration results in higher citation rates of published articles," *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, vol. 26, no. 6, pp. 759–767, 2006.

[18] Z. Taşkın and A. U. Aydinoglu, "Collaborative interdisciplinary astrobiology research: a bibliometric study of the NASA Astrobiology Institute," *Scientometrics*, vol. 103, no. 3, pp. 1003–1022, 2015.

[19] G. Baker, R. Gibbons, and K. J. Murphy, "Subjective performance measures in optimal incentive contracts," *Quarterly Journal of Economics*, vol. 109, no. 4, pp. 1125–1156, 1994.

[20] A. Yegros-Yegros, I. Rafols, and P. D'Este, "Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity," *PloS one*, vol. 10, no. 8, Article ID e0135095, 2015.

[21] H. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents," *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 1973.

[22] A. M. Khan, A. Shahid, M. T. Afzal, F. Nazar, F. S. Alotaibi, and K. H. Alyoubi, "SwICS: section-wise in-text citation score," *IEEE Access*, vol. 7, pp. 137090–137102, 2019.

[23] B. Lepori, P. Van Den Besselaar, M. Dinges et al., "Comparing the evolution of national research policies: what patterns of change?" *Science and Public Policy*, vol. 34, no. 6, pp. 372–388, 2007.

[24] M. Bruce, L. Daly, and N. Towers, "Lean or agile: a solution for supply chain management in the textiles and clothing industry," *International Journal of Operations & Production Management*, 2004.

[25] T. Braun and A. Schubert, "A quantitative view on the coming of age of interdisciplinaryin the science," *Scientometrics*, vol. 58, no. 1, pp. 183–189, 2003.

[26] B. Lužar, Z. Levnajić, J. Povh, and M. Perc, "Community structure and the evolution of interdisciplinaryin Slovenia's scientific collaboration network," *Plos one*, vol. 9, no. 4, Article ID e94429, 2014.

[27] B. Aleman-Meza, L. Ding, M. Nagarajan, and C. Ramakrishnan, "Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection," in *Proceedings of the 15th international conference on World Wide Web*, pp. 407–416, Scotland, UK, May 2006.

[28] I. Molotov, V. Agapov, V. Titenko et al., "International scientific optical network for space debris research," *Advances in Space Research*, vol. 41, no. 7, pp. 1022–1028, 2008.

[29] M. Karlovčec and D. Mladenić, "Interdisciplinaryof scientific fields and its evolution based on graph of project collaboration and co-authoring," *Scientometrics*, vol. 102, no. 1, pp. 433–454, 2015.

[30] A. L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mechanics and Its Applications*, vol. 311, no. 3-4, pp. 590–614, 2002.

[31] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organizational Research Methods*, vol. 18, no. 3, pp. 429–472, 2015.

[32] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: review, analysis, and cooperative study among tools," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011.

[33] B. Jarneving, "A comparison of two bibliometric methods for mapping of the research front," *Scientometrics*, vol. 65, no. 2, pp. 245–263, 2005.