

Research Collaboration Influence Analysis Using Dynamic Co-authorship and Citation Networks

Sidra Razzaq¹, Ahmad Kamran Malik^{1*}, Basit Raza¹, Hasan Ali Khattak^{2*}, Giomar W. Moscoso Zegarra³, Yvan Díaz Zelada³

¹ Department of Computer Science, COMSATS University Islamabad, Islamabad (Pakistan)

² National University of Sciences and Technology (NUST), Islamabad (Pakistan)

³ Escuela de Posgrado Newman, Tacna (Peru)

Received 20 April 2021 | Accepted 20 February 2022 | Published 1 March 2022



ABSTRACT

Collaborative research is increasing in terms of publications, skills, and formal interactions, which certainly makes it the hotspot in both academia and the industrial sector. Knowing the factors and behavior of dynamic collaboration network provides insights that helps in improving the researcher's profile and coordinator's productivity of research. Despite rapid developments in the research collaboration process with various outcomes, its validity is still difficult to address. Existing approaches have used bibliometric network analysis with different aspects to understand collaboration patterns that measure the quality of their corresponding relationships. At this point in time, we would like to investigate an efficient method to outline the credibility of findings in publication—author relations. In this research, we propose a new collaboration method to analyze the structure of research articles using four types of graphs for discerning authors' influence. We apply different combinations of network relationships and bibliometric analysis on the G-index parameter to disclose their interrelated differences. Our model is designed to find the dynamic indicators of co-authored collaboration with an influence on the author's behavior in terms of change in research area/interest. In the research we investigate the dynamic relations in an academic field using metadata of openly available articles and collaborating international authors in interrelated areas/domains. Based on filtered evidence of relationship networks and their statistical results, the research shows an increment in productivity and better influence over time.

KEYWORDS

Collaborative Research Analysis, Influence Analysis, Network Centrality, Quality Of Research, Statistical Analysis.

DOI: 10.9781/ijimai.2022.03.001

I. INTRODUCTION

INTERACTION among scientists is vital for novelty and productivity in their research areas. Research collaboration (RC) is seen as a primary indicator of certain effective means of cooperation among states. Besides, by being actively involved, professionals can quickly grasp the essentials of a large research area [1]. The dramatic growth in scientific collaborations provides further insight into the evolution of social sciences research collaboration. Sharing of knowledge increases the chances of spreading the benefits of protocols and ideas from one region to another [2]. The collaborators are using this approach to study the network research work both for their own progress and to have a better understanding of networked research. This approach is required because scholars are providing transparent transformation information about careers, research areas, interests, and their key factors of productivity: their publications record and influences [13].

Researchers somehow go beyond the study of network research to favor the effects of different types of relations: co-authors, productivity, and their influence. While the fact that most of the recent studies have

been geared around quality appraisal is an unavoidable management feature at all levels. It encourages progress in growth analysis, quality, and performance of researchers for spreading interests and knowledge [5]. Such a professional assessment, which should be focused on the findings of the output of the scientist, is important not only for performance evaluation but also for gaining a high reputation in the research community [9]. To evaluate a scholar's quality of work, several studies suggest quantifying the writing practices as a good measure of a scholar's efficiency. The general idea is that if the author publishes and these articles are quoted, a scientist will receive good evaluation in the research community. In addition, the citations count characterizes the number of publications [14], consisting of various statistical methods such as co-citation, co-authors, etc. that evaluate data in the scientific corpus to provide a quantitative understanding of the growing literature and the flow of knowledge in an individual area. This allows investigators to overcome problems that hinder the achievement of progress in their profiles. Many measures including centrality have been used in the literature [18], [19], [38] to achieve these goals.

Analysis of scientific articles and citation collaboration has an extensive history of exploring explicitly the scientific outcomes; however, interests of collaborative researchers between authors and their publications and research areas are less explored. Such alliances also originated from social networks and are actively promoted

* Corresponding author.

E-mail addresses: ahmad.kamran@comsats.edu.pk (MA. Kamran), hasan.alikhattak@seecs.edu.pk (HA. Khattak).

across them because they transcend global, organizational, and administrative boundaries. The quality of community-based research is examined through the published articles and their references over years and areas [4] [3].

There has been a substantial rise in the number of partnerships among scientists in the science environment. They are exhibiting their information exchange practices by collectively publishing papers, which is an indicator of knowledge formation. Authors in [7] noticed that the development of scientific knowledge, including new theories, research problems, and research ideas are an important result of scientific partnerships. Despite all traditional analysis of researchers' communication through citation patterns, collaboration involves a rational breach to accelerate the efficiency and time needed for challenging discoveries. As a consequence, it was reported that the growing awareness of collaboration in research has contributed to a strong focus on the issue of collaboration [2].

Collaborative qualitative research is being judged for lacking transparency in scientific procedures and analysis. In fact, is it necessary to investigate the influence of how co-authorship/collaboration has developed in the social sciences? There is a demonstrable increase in the individual disciplines in which the task of scientific impact prediction is formulated as using standard methods for predicting performance. [12] [10]. However, few researchers examine or compare the factors across disciplines. Either, they focus on the development in a few fields or they investigate the main branch of research: the physical life, social sciences, and humanities which means that developments in the individual field are not visible [6].

Recent studies have also faced challenges in evaluation of correlation analysis over time [15]–[17]. The challenges involved are: increase in visibility of citation rates as research output, publications in other research areas due to less citations, and publications, impact on scholar performance. Moreover, few authors have examined or compared the factors across disciplines [39], [40].

On the contrary, it is beneficial to suggest that collaboration can be improved further by creating effective network relations in different domains. The influence of cooperation established in social sciences with the shift in research areas must be explored. Moreover, by considering the previous works, we are looking into how different research relations support each other in network language, finding their influence by considering issues faced to date. If the network responds is great; productivity and influence increase and there will be more opportunities to publish.

We aim to provide a new collaboration model to analyze the characteristics of co-authored research articles. We propose new performance metrics (active area, self and average citations, paper score and authors score) for inspecting the quantifiable analysis collaboration. In this study we measure the combinations of performance metrics to explore four types of relationship networks: publication to author, author to author, publication to publication, and publication to research areas, to find the relationship between centrality factors and g-index metrics as the key proxy in our collaboration model by considering the shortcomings of previous works.

Our main points of investigation, in this research, are the relevancy effect on scholars' articles in their research area (which was ignored by previous researchers), their affected quality of research, and the impact on scholars' articles' performance in terms of citations in their area.

The remaining sections of this work are structured as follows. Section II discusses the studies conducted previously. In section III the dataset details and mathematical formulation along with the proposed methodology are presented. Section IV demonstrates the results and their analysis and offers a discussion. Finally, the findings of this work along with future directions are presented in section V.

II. LITERATURE REVIEW

Complex research communication networks are being looked at as a new framework that is used in multiple studies and practical applications of social network analysis. Collaboration has been studied by researchers of various fields, with the aim of performing comprehensive analysis of three decades of co-authorship network data [20]. Similarly, authors have also proposed the inverted U-shaped collaboration network by considering the citations at individual level.

With the combination of cognitive and relational dimensions of social capital, a positive effect of relationships strengthens the ties of networks [8]. The higher the relationship value, the lower the biasedness, although this value makes for a mixture of strong or weak ties. Many researchers did not take notice of the ambiguity of researcher's names at the individual level, indirect ties, global networks, and other aspects of knowledge creation networks [21]. The ambiguity of author's names has been resolved with the selection of the preprocessing bibliometric method, which enables a better representation of the co-authorship collaboration network using digital bibliography and library project data sources [33], [43].

The bibliometric cooperation in social sciences has gained accolades at both domestic and international levels. Here researchers focus on two types of bibliometric methods, namely: the parametric [22]–[24] and descriptive models [25]–[28], in which they use research articles, citations, and their collaboration networks. Challenges such as measurement error, performance, scaling, dimensions, normalization, and quality were also addressed. Similarly, new strong correlation methods are proposed with slight modifications in social network algorithms [29] to influence the network and to evaluate their importance in the research community through centrality measures.

The normalized centrality measures and average ties strength have a strong effect on scientific academic performance, in terms of h-index and g-index. However, the results showed that researchers at national level perform better than those at international level in the network. As the node gains the central position in the network, it determines the opportunity to collaborate and share knowledge in terms of betweenness centrality but not of improving its performance. Usually the authors with high betweenness centrality are implied to have more importance. However, the weighted number of citations shows the influence of only significant papers and not the authors in the community [11], [30], [31]. Furthermore, studies of Italian academia [37] show that the data sources have strong influence on network analysis to test the scientific performance of the researchers. It has been observed that small-world structure, both at national and international level, characterizes the networks with three popular data sources: current index to statistics (CIS) [32], web of science, and national funded projects. However, in general, CIS is more widely used in international research topics [33].

In [34], the authors conducted social network analysis (SNA) to show that emerging scientific topics receive less attention from researchers compared to the subtopics derived from the main scientific topics. The authors discussed the co-authorship network of forest entrepreneurship, which is a new and emerging field of study. They concluded that the topic of forest entrepreneurship is understudied compared to subtopics such as innovation forest, forest industry, and the policy of forest entrepreneurship. Normally, the best method to determine the performance of a researcher is measuring centrality or community detection. However, scholars mostly work on different projects and they have different roles in each one. In [35], the authors concentrate on the overlapping of scholars in projects and the roles they play therein. Based on the comparison, in this study, it need to be analyzed whether the performance of the researchers correlates with their contribution patterns in the projects or not.

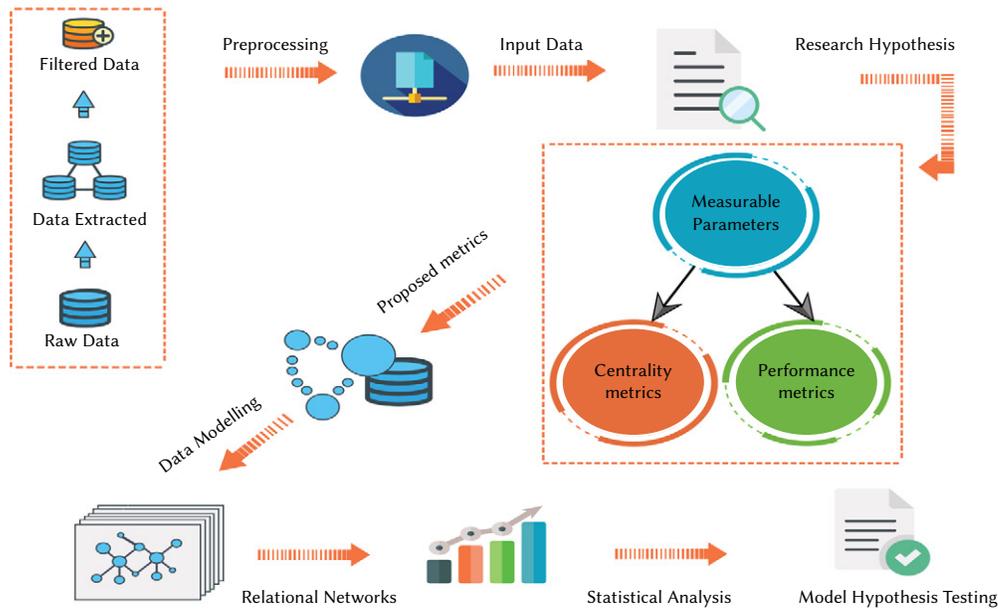


Fig. 1. Proposed collaboration model.

Psychometric rasch model has been used by authors to evaluate the researcher's performance at an individual level by considering the control variable errors of bibliometric research [36]. They achieved promising performance for percentage, fractional counting, and normalization, while ignoring the total citations, multiple co-authorship, and distribution assumptions. Moreover, in article [42], the authors proposed a method for finding a domain level influence. These researchers focus on author citations through the information linkage process. They use an assumption that author-nodes are reflected as more influential if they have a great self-influence. The authors in [44] study the benefits of collaboration in scientific studies. They show, how collaboration can affect the position of authors in a co-authorship network and increases productivity as well as the influence of the authors. The authors state that researchers have recently become more collaborative. They conclude that collaboration in scientific studies improves research as well as stabilizes the relationship between researchers. The impact of network size and correctness on researchers' productivity and influence has been studied in [45]. Assuming the number of publications as productivity and the number of citations as an influence, the researcher's analyze the times when the scholars were productive and influential.

III. RESEARCH COLLABORATION INFLUENCE ANALYSIS

In this section, we address in detail the proposed methodology of research collaboration influence analysis.

Collaboration/joint collections of networks used in different research fields are considered to be a structural network where all objects are considered to be linked and their degree value is important. For instance, authors being tied to each other result in non-redundancy of some critical information flows in their ties because of some structural holes. This may enable scholars to enhance the research to accrue potential opportunities for controlling the information flows among them. As described earlier in the introduction part, the numbers of citations and publications are positively linked, which points to quality of research work.

To address the collaborative research problems, the proposed scheme represented in Fig. 1 is used. The key role of this collaboration model is to overcome the issues of researchers' growing professional relationships and the scientific influence of their collaboration

networks in a research community. Many years ago, Garfield [41]; acknowledged that scientific enterprise has increased and become more complex. If the complexity and collaboration change, the quality and content of citations should also evolve. This paper explores the characteristics of international research articles and contributes to the enhancement of correlation analysis.

For collaborative research analysis, we extracted data via Arnet-Miner to perform the analysis of centrality measures, g-index, and introduced metrics: active area, self-citations, paper score, and authors score. Furthermore, we propose a new collaborative approach to find the better influence with performance metrics. Finally, we draw a correlation analysis to show the effectiveness of our model. We create four relationship graphs the publication–author, co–author, publication–publication and author–research area, to give leverage to the relationships of quantifiable research. In this regard, Python language, Network X, and SPSS tools are used for visualization and metrics calculation.

A. Data Pre-processing

The dataset used in this research is mostly taken from <https://www.aminer.org/lab-datasets/soinf/> provided by [42], consisting of publications, co-authors, citations, and research area. We used this raw data to extract the required graphs from Arnet-Miner to bring into form that can be used for our research analysis. The raw data contains information on 2555 publications. The information includes publication title, year of publication, conference/journal, and authors' names. The publications are from 10 different research areas of computer science. The following list shows areas of publications with the codes:

- Area 75: Information retrieval
- Area 131: Bayesian networks / Belief function
- Area 107: Web services
- Area 199: Natural language system / Statistical machine translation
- Area 16: Data mining / Association rules
- Area 24: Database systems / XML data
- Area 145: Semantic web / Description logics
- Area 144: Web mining / Information fusion
- Area 162: Machine learning
- Area 182: Pattern recognition / Image analysis

In addition, the raw data include a citation relationship among the publications. The citation relationship shows the papers that have referenced a specific paper. The dataset provides 6101 citation relationships. Furthermore, we generated our own graphs to show the relationship between publications and research areas. We have performed the preprocessing whereby we removed duplicate records, blank spaces, and other unnecessary data and organized the data according to our requirements.

B. Measurable Parameters

There are many network properties that describe how nodes are connected to each other on a network. The SNA has many measurement types to systematically characterize nodes in the networks. A few measures used in our research are described below to find which node is important in the network and to find an influence on the other nodes. We determine the importance of nodes by calculating their score and then finding their position within a network. Here, importance means any effective authors and papers with respect to citations.

1. Degree Centrality

Degree centrality is considered as the simplest and most common way of finding important nodes in a network. For example, if a vertex has five edges, then we say that it has degree 5. Furthermore, in directed graphs, there are two kinds of degrees: in-degree and out-degree. The in-degree shows the number of edges coming from one vertex to another, and out-degree is the edges originating from the vertex going outwards. Equation 1 is defined to find the degree of nodes. We are only considering in-degree. The formula for evaluating the normalized degree centrality is as follows; where $d(G)$ represents the degree of a node (like number of papers that cited this node/paper) and N represents total nodes in the network (like total number of papers).

$$C_b(G) = \frac{d(G)}{N-1} \quad (1)$$

2. Eigenvector Centrality

Eigenvector centrality is a kind of extension of degree centrality. It is particularly focused on two things: the node itself and its neighbor's. Here, we calculate values with 0 and 1 only. The closer the value to 1, the higher the centrality. We use this centrality to find which nodes take information to other nodes quickly. Equation 2 is used for finding the centrality of this measure. Here $A = (a_{v,t})$ presents the adjacency matrix, i.e. $a_{v,t} = 1$ if vertex v is linked directly to vertex t and is 0 if otherwise.

$$x_t = \frac{1}{(\lambda)} \sum_{t \in G} a_{v,t} x_t \quad (2)$$

3. Closeness Centrality

Closeness centrality indicates how close a node is to the other nodes in the entire network. It is highly effective for calculating the shortest possible paths among all nodes before assigning each node a score based on its sum of shortest paths. In equation 3, $D(y, x)$ are used to find distance between y and x .

$$C(x) = \frac{1}{\sum_{x \neq y} D(y, x)} \quad (3)$$

4. Harmonic Centrality

Harmonic centrality is a variant of closeness centrality. Instead of summing the distances of a node to all other nodes, the harmonic centrality algorithm sums the inverse of those distances. It can display interesting results, especially for the top nodes of the graphs. Simply referring to equation 4, if the value is equal to zero then there is no path between x and y .

$$H(x) = \sum_{y \neq x} \frac{1}{D(y, x)} \quad (4)$$

5. Betweenness Centrality

Betweenness centrality indicates the extent to which a node lies on the shortest path among other nodes in the entire network. It indicates nodes that can primarily act as bridges between nodes and can be used for finding the individuals who influence the flow around a system. In equation 5, σ_{st} represents a pair of vertices used to compute the shortest path from node s to node t . V is used to define the vertex in a network.

$$C_b(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (5)$$

6. Clustering Co-efficient

The clustering coefficient is a measure of how likely it is that two nodes that are connected are part of a larger highly connected group of nodes. The global version gives an overall indication of the clustering in the whole network and can be applied to both undirected and directed networks, whereas the local version gives an indication of the embeddedness of single nodes. In equation 6, clustering coefficient C_i is shown for all vertices n .

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (6)$$

C. Data Modeling (Research Collaboration)

For decades, academic science research used collaboration networks as a proxy. Apparently, the usage of bibliometric data analysis and accessibility makes the collaboration more interesting for the future research. Moreover, many scientists have been actively involved in co-authorship, for a strong relationship among researchers in their academic careers. When we use a network of researchers, they are always connected with solid connections between two or more persons, if they have a strong relation. Relational networks provide with an insight about which individual or specific node has a great influence is strongly connected. Furthermore, many activities have been performed by researchers with the assumption of receiving positive effects of profiles; the co-authorship or collaboration suggests more relevant or accurate influence and authors getting publications in many fields of related research.

If this is the case then it can be stated that they might be having different interests or maybe just trying to add co-authors as a matter of getting increased publications. What needs to be understood is whether they have any accurate citations of research conducted so far in different domains. It would definitely be worthwhile to know about researchers' performance until higher ranks, based primarily on their academic activities, research interests, and especially research area changes with the passage of time. This literature gap motivates us to study the scholar's in specific field. It also makes it feasible for readers to evaluate the credibility of publications and authors.

For this purpose, we also need to understand the influence of citations through mathematical modeling. In the research we investigate the dynamic relations in an academic field using metadata of openly available articles in inter-related areas/domains. We develop network relationship graphs based on co-authorship, publications through their number of citations (self and original), and research areas in which they are actively involved. We use centrality measures (helping us to identify the real influencer's) on introduced performance metrics to find a correlation between evaluated results and g-index. In addition, based on the lack of collaboration analysis in existing studies, this work provides research collaboration development across other scientific fields that can be effective in future research.

D. Constructed Parameters

1. Average Citations

In our filtered data, each article has one or more authors. To get better citation information, we calculated an average citation score by equation 7 for each author. Here, variable denoted by *ACS* represents "average citation score" of an author, *CPA* represents "citations of each paper authored", and *NP* represents "number of papers" published by an author, so it is interesting to compare with total citations to know whether an author's citations are from few paper or many papers.

$$CS_i = \frac{\sum_{j=1}^{NP} (CPA_j)}{NP_i} \quad (7)$$

Algorithm 1: Creating Average Citations

Input: Publications Authors Bipartite Graph

Output: Average Citations Values for each publication by an author
 find citing papers
 add all citations
 Divide total citations by total number of papers return average citations

2. Active Area of Author

The active area of an author is the research domain in which a researcher has been recently and most frequently performing research. For this purpose, we examined the publications of each researcher to determine their active areas.

Algorithm 2: Finding Active Area

Input: Average Citations Values

Output: Authors Active Areas

create an array of paper names
 Paper names exist in publication area nodes
 get paper and year from selected nodes
 subtract current year from publication year
 area score = dividing constant factor by year
 area in publication area
 find index of selected areas
 add score value to index area
 get maximum area score
 find index of active score area
 return authors active areas list

Now, owing to the change of trends, and changing areas with respect to time, the researcher's productivity could be compromised. The trending area and publication time matters in research. It could be possible that an author published papers in one area and later in another area, because of lost of interest in the first area. This can create a great variance in results when comparing productivity of authors in general and in specific areas. For this purpose, we need to take into account the score of each paper depending on its publication time using equation 8. The variable "Age" shows how old the paper is. Age represents a difference of value between the publication year and the current year. We have given $K = 0.2$ a constant value, which is taken into consideration to reduce the variations between the score of papers published in two consecutive years. This is used to avoid the counter values, that produces 20% difference value for a max of two years.

$$Score_i = \frac{1}{Age} \times K \quad (8)$$

To find the active area of a researcher, we formulated the equation 9. Here *NPA* represents "number of publications in the area" for author *j*, *AA* represents "active area" *i*, and *Score* represents the sum of scores of all papers of the author *j* in the area *i*. The area *i* having max value is selected as the authors active area.

$$AA_i = \max(NPA_j + Score_j) \quad (9)$$

3. Citations

In research, there are two types of citations. The first is original citations by a random researcher. The second type is self-citations quoted by one of the authors of a paper. We categorized the total citations of a publication into original and self-citations and analyzed the influence or connection between network nodes. It also helped us in reranking (paper score) the publications based on citations in addition to research area.

Algorithm 3: Calculating Citations

Input: Authors Active Areas

Output: Original and Self Citation Values

create an array of paper edges from active areas
 get all neighbor papers that cited an original paper
 get total citations by adding 1 for each neighbor (citing) paper
 finding original papers from neighbors
 get original citations that do not include authors
 Self citations = total citations – Original citations
 return original and self citations list

For example, an author from the database "Fuad.M.Alkoot" published a total of one paper that got five citations from which only three papers originally cited his paper and two citations were made by self-citations. To calculate the self-citations, we used equation 10, where Symbols SC_p , T_c , and O_c represent self-citations, total citations, and original citations.

$$SC_i = T_c - O_c \quad (10)$$

4. Paper/author Ranking Score

We rank or rate publications through citations and active area of the author. This helps in finding author ranking score. This ranking will help students and readers select and evaluate a publication based on its ranking and also to evaluate a researcher in a specific research area.

Algorithm 4: Creating Paper Score

Input: Original and Self Citation, active area Values

Output: Paper Ranking List

finding paper rank
 assign area score 1 or 0.7 to paper
 get paper score against original, self citations, and area score using Equation 12
 return paper ranking list

The technique used in equation 11 is giving a value of 1 in the case of author's active area and paper belonging to the same area and 0.7 in the case of different areas. Here S_a represents the score of each author against publications. It was considered, a weighted value if there were four authors on a paper, and if they were interested in the same area then we assigned 1 to all.

$$S_a = \begin{cases} 1 & \text{if authors belongs to same area} \\ 0.7 & \text{otherwise} \end{cases} \quad (11)$$

We used the categorized citations of the publication while ranking it. In equation 12, C_s represents self citations, C_o represents

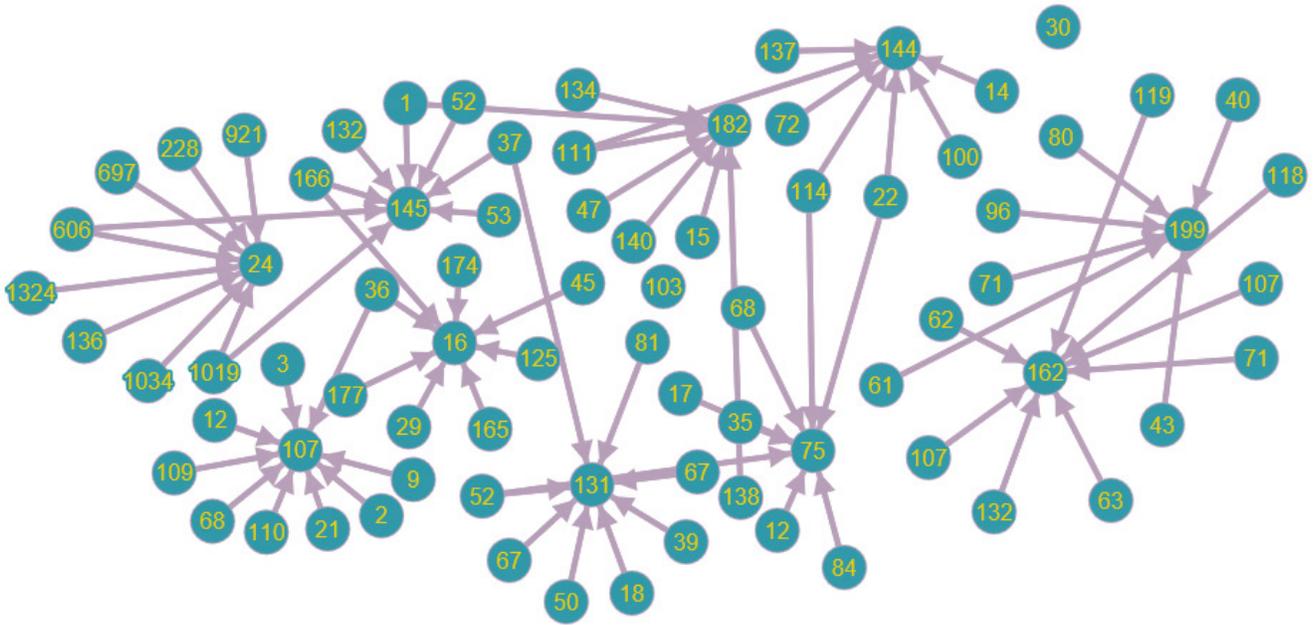


Fig. 4. Publications to area-id network.

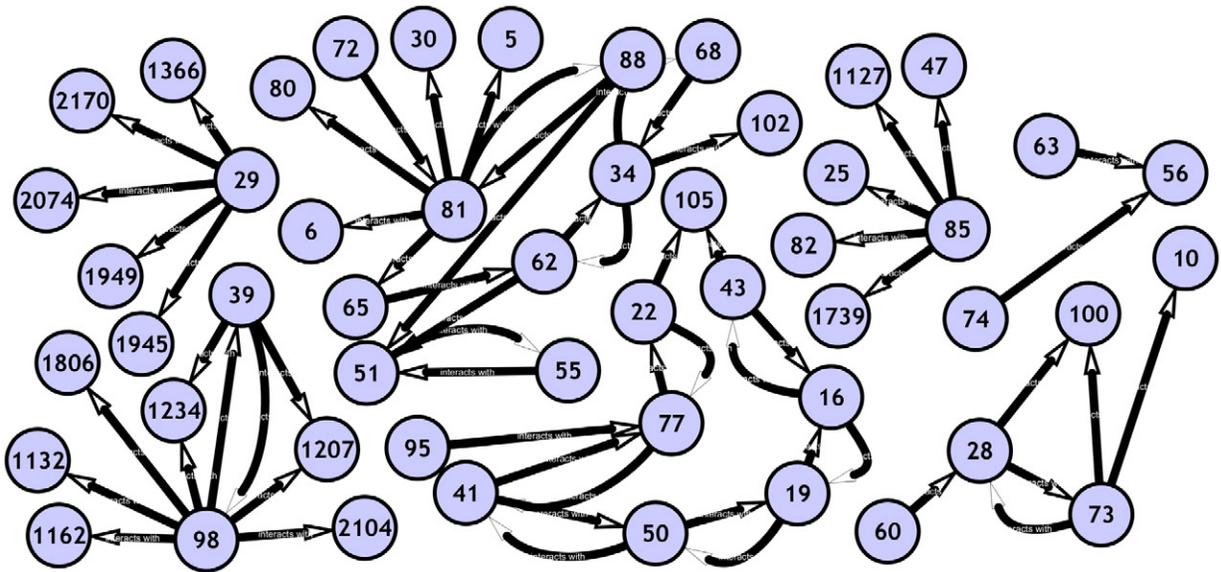


Fig. 5. Publications to publications network.

authors who are directly interacting in an article in different domains can be found in the graph. This is effective for all co-authors in calculating collaborative work scores, for example, if any author of the article has original citations it directly gives benefits to all co-authors.

There is another directed relationship graph which is shown in Fig. 4, presenting the relationship between the authors and their interest area in research publications, which is called the publication to active area relation network. It shows all publications related to similar area collectively. It creates a research area node to which all related nodes are connected. This graph is used to find research area of the publications.

Another directed network relationship graph shows relations with publications cited by another publication which is also called co-citation network. It represents citations that each paper has in other papers. Each publication node generates a directed edge to the corresponding node cited in that publication. Fig. 5, represents the

graph; in which each vertex represents publication id and its in-degree can be used to find which publication has what number of citations.

B. Data Analysis

In this section, we present the results evaluated based on our proposed approach. As mentioned in the research methodology, our main goal was to find the paper score and author influence by comparing the network relationships in which the number of papers, authors, research area, and citations play a major role.

We are focused on creating the ranking score for authors and papers. In our data, each publication has one or more authors from different countries; we performed an analysis to know their progress in different research domains. It is observed that researchers change their research interest and with the passage of time they start publishing in another related research area. After this observation, we found which area the researcher had adopted or changed up to now. The active area of

an author is the research area in which the researcher has been most recently and most frequently performing research. Because of rapid changes in research trends, researchers willingly change their minds, make collaborations, and publish in trending topics in their academic career to gain advanced knowledge of growing technology and information; however, this may affect the researcher’s performance and profile ranking in specific research areas. With the consideration of these facts, we evaluated the active area of authors. Specifically, we depend on data collected from co-authored articles of international collaborators to evaluate the aforementioned factors. We found the research area in which each author has a special interest, as shown in table I, and in which they have recently been active and are making progress.

TABLE I. SELECTED RESEARCHERS PUBLISHING IN SPECIFIC RESEARCH AREAS

Authors	Active
Byron Dom	16
Jonas S. Karlsson	24
Lyman Do	24
Younkyung Cha Kang	24
Fegaras	24
Regis Sabbadin	131
Sara Comai	144
Klaus Schild	145
Andrea Schaefer	145
Alessandro Artale	145
Quentin Elhaik	145
Pavel Paclik	182
C. Fairhurst	182
Miles Osborne	199
H. Gregory Silber	199

Because authors tend to work in different areas over a period, the quality of their research in one area may differ from that in another area. The quality of a publication depends on the area of expertise of its authors. If a researcher is not involved greatly in a certain research area, it is possible that his/her interest might have changed and also become focused to work on a few articles instead of many to meet the standards and reputation of an academic field.

Moreover, in the case of citations of an author, Google Scholar shows the sum of citations of all the publications. However, to have a better idea, we calculated different types of citation scores in addition to author’s active area and used them for calculating paper score. It is difficult to evaluate a researcher only by the citations, as some publications might have high citations while some may have very low citations, including self and original citations. To, build a researcher’s

profile, we need to calculate the average citations as well. The Fig. 6 shows the visual representation of researchers performance in terms of their total citations, average citations and number of published papers.

Researchers having many papers and many citations will have lower average citations than the one having small number of papers and many citations. Furthermore, we evaluated authors who received a maximum number of papers but fewer citations as having less influence. Besides, researchers with fewer publications and many citations show highly influential results. This shows, the effectiveness and quality of researchers’ publications. An increased number of papers does not help in increasing profile productivity. Instead, a high original citations value boosts one’s profile ranking. Furthermore, we have used the active areas and evaluated citation results for the evaluation of researchers’ profiles and the quality of their research.

In collaborative research, the concept of types of citations (self/original) becomes influenced in two ways. We consider citations with respect to number and also with respect to research areas. In this research, we focused particularly on two types of citations for finding the effects of citations rate on paper quality. First, the genuine citations by a random researcher, and second self-citations are those cited by one of the authors of papers. This method also helped us in ranking the publications. Because we have created the profiles of researchers based on their active areas, the approach helps other researchers, students, and readers to select and evaluate authors and their publications in selected research areas of computer science.

With the help of collected data and generated graphs, we evaluated the originality score of each paper. Table II, shows the originality score of some publications based on research areas and all citations factors. An author’s score can be calculated by summing up all his publication scores. The rank that an author achieves through publications presents the influence of articles on the researcher’s profile. The paper score is mainly based on citations and active research area of the author. The higher the number of original citation the higher the paper score. Also, if the paper is from active research area of an author, it increases the paper score. Moreover, researchers having most publications from active area with many citations, but with less self-citations, will get high profile ranking.

As mentioned in the earlier discussion, more self-citations affect the quality of a paper. Fig. 7 shows, the breakdown of citations into self-citations and original citations from total-citations of published papers along with their paper score. With the help of more citations, there is a chance of getting a high-ranking score. However, in many cases the reason behind a lower-than-average score is that self-citation directly affects the quality of articles. If a research is not quoted by other researchers, it will lose its worth and self-citations simply increase

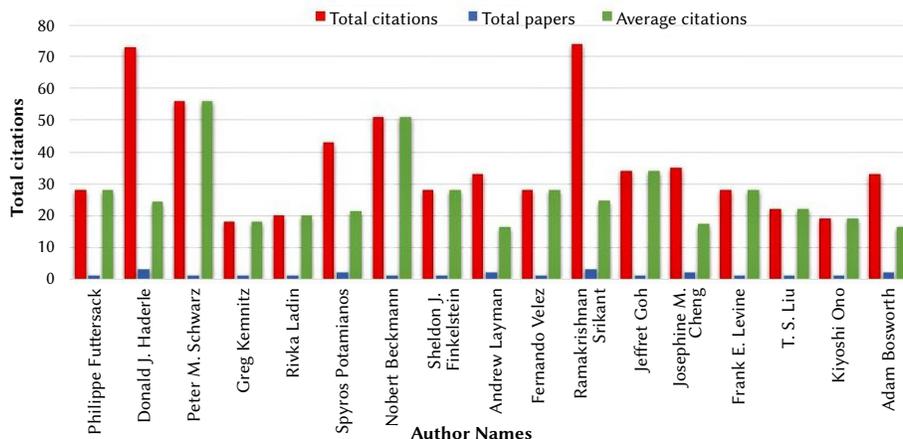


Fig. 6. A comparison of authors total and average citations depending on number of publications.

TABLE II. COMPARISON OF DIFFERENT TYPES OF CITATIONS AND CALCULATED PAPER SCORE

Paper Title	Total Citations	Original Citations	Self Citations	Paper Score
The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles	51	37	14	28.8
Fast Algorithms for Mining Association Rules in Large Databases	55	53	2	33.6
The hB-Tree: A Multiattribute Indexing Method with Good Guaranteed Performance	15	10	5	9.5
ARIES: A Transaction Recovery Method Supporting Fine-Granularity Locking and Partial Rollbacks Using Write-Ahead Logging	56	36	20	30.6
ARIES/KVL: A Key-Value Locking Method for Concurrency Control of Multiaction Transactions Operating on B-Tree Indexes	26	12	14	8
Mining Association Rules between Sets of Items in Large Databases	32	25	7	18.8
An Approximate Analysis of the LRU and FIFO Buffer Replacement Schemes	6	0	6	3.4
An Effective Hash Based Algorithm for Mining Association Rules	23	20	3	14.2
Efficient and Effective Clustering Methods for Spatial Data Mining	30	23	7	17.6
Recovery and Coherency-Control Protocols for Fast Intersystem Page Transfer and Fine-Granularity Locking in a Shared Disks Transaction Environment	23	14	9	13

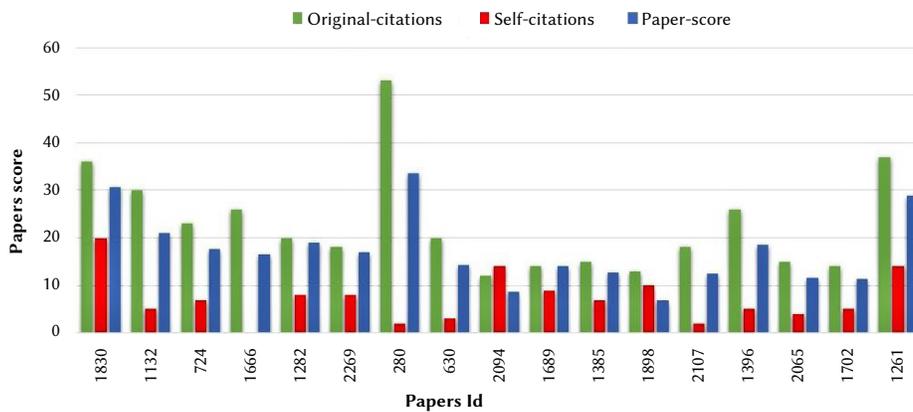


Fig. 7. Comparison of Original Citations, Self Citations, and Paper Score.

the citations rate, which indirectly affects the researcher’s profile and paper score. From this representation, we conclude that most of the publications having a high original citations score get more paper score if they are not out of the active research area of authors. The paper-id 280 has a total of 53 original citations with two self-citations getting a high paper score of 33.6. However, paper-id 2094 has more self-citations (14) than original citations (12), which results in a low paper score of 13.8. Some papers having self-citations get a reasonable paper score because they also have either more original citations or the paper is from active research area of authors. We can conclude that researchers try to raise their profile ranking by making more self-citations, directly affecting their performance and academic career because it is not effective to achieve the higher rank by self-citations.

Author’s research area is an important factor to be considered especially in the era of Internet as many researchers are doing collaborative research and some of them change their active research area to get high research rank through another popular research area. In Fig. 8, the circle are created based on area-id and the numbers inside each circle represents the author’s name and paper score. Size of the circles represent author score which is collective score of authors’ papers ranking, which represents their performance score in the interest area over time. Through all results, we can conclude that citations and change of area plays a major role in scholar profile performance and research productivity throughout their academic career. If the authors remain focused on incrementing their citations, then it would eventually affect their paper quality and their profile.

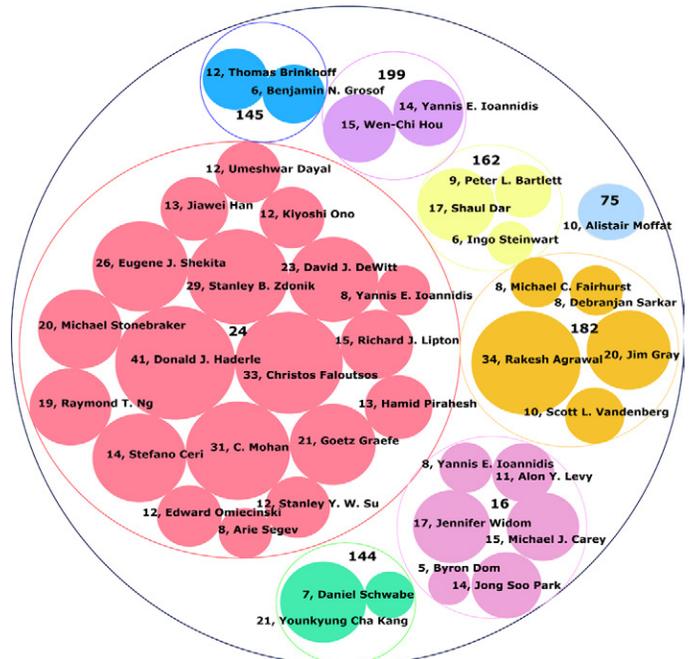


Fig. 8. Circles represent research areas in which inner bubbles are labeled with authors name, paper score, and the size of different inner bubbles represent author’s score.

C. Results Analysis

To test and analyze the data, we used paired difference sample statistical tests, which evaluate the difference among groups of means. We evaluated our networks through paired difference analysis with g-index using IBM SPSS. As shown in Fig. 9, it calculates the mean difference value of independent ratio data by the standard error of each value, the freedom degree (df), mean difference, and 95% confidence interval between specified significant value and significant p-value. According to the mentioned analysis, co-authorship relation shows that except for the clustering co-efficient, all other parameters are statistically significant. The reason for not evaluating the value is that results have zero square difference error. This happened because we assigned the weighted values in which the correlation of clustering coefficients had negative results against that relationship. Although the results could be quite reasonable if the values of nodes were assigned with relatively similar information. Other factors mentioned in the figure are strongly correlated with this relationship.

Moreover, in Fig. 10, only one factor is not supporting this relationship with a significant value that is between author-g-index. Scientifically, many collaborations are performed with different

scholars or showing strong connections with co-authors, which leads to the high significant value. Thus, in this context it shows negative results. As in the publication area relation graph, the authors were not directly involved, which naturally negatively influences g-index. Furthermore, in Fig. 11 publications to author relation, clustering coefficients and eigenvector centrality are factors that are not effective to support the network relation. The same case occurred, which shows the extent to which the clustering coefficient drives the behavior of instances and results in negative influence. It investigates the structure of scientific collaboration in a whole network, which is why it is scored against the global network graph. Apparently, closeness, harmonic, and clustering co-efficient are strong measures that show positive influence as represented in Fig. 12 except others. This relational network has directed edges and integer values in a global network which makes factors not suitable.

Based on the aforementioned results, it is suggested that the clustering coefficient is the major factor that does not fit in for finding the influence of aforementioned relational networks. The reason behind is that clusters need their neighbor bodies strong for all actors in a network. Two nodes attain the same probability if they

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Closeness Centrality - G - index	-26.9388350	.0469028	.0046903	-26.9481416	-26.9295285	-5743.541	99	.000
Pair 2	Degree Centrality - G - index	-26.9850833	.0286307	.0028631	-26.9907643	-26.9794024	-9425.225	99	.000
Pair 3	Betweenness Centrality - G - index	-26.9990826	.0063113	.0006311	-27.0003349	-26.9978303	-42779.274	99	.000
Pair 4	Harmonic Centrality - G - index	-19.450	6.147	.615	-20.670	-18.230	-31.642	99	.000
Pair 5	Eigenvector Centrality - G - index	-26.9593602	.0880870	.0088087	-26.9768385	-26.9418818	-3060.538	99	.000
Pair 7	Edge Centrality - G - index	-26.9980579	.001561	.0001561	-26.9983675	-26.9977482	-172994.051	99	.000

Fig. 9. Paired difference analysis of co-authorship network.

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Authors - g - index	15.667	49.810	7.189	1.203	30.130	2.179	47	.034
Pair 2	Degree Centrality - g - index	-26.8909575	.1425336	.0205729	-26.9323449	-26.8495700	-1307.103	47	.000
Pair 3	Betweenness Centrality - g - index	-26.9762951	.1136740	.0164074	-27.0093026	-26.9432876	-1644.151	47	.000
Pair 4	Harmonic Centrality - g - index	-2.5138750	3.5720863	.5155863	-3.5511008	-1.4766492	-4.876	47	.000
Pair 5	Eigenvector Centrality - g - index	-26.8855014	.0888129	.0128190	-26.9112900	-26.8597128	-2097.311	47	.000
Pair 6	Clustering Coefficient - g - index	-26.9267231	.1415273	.0204277	-26.9678183	-26.8856278	-1318.147	47	.000
Pair 7	Edge Centrality - g - index	-26.9782408	.0169456	.0024459	-26.9831613	-26.9733203	-11030.033	47	.000

Fig. 10. Paired difference analysis of publication to area network.

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Closeness Centrality - g - index	-26.8744917	.0732103	.0163703	-26.9087551	-26.8402282	-1641.660	19	.000
Pair 2	Degree Centrality - g - index	-26.9407408	.0773169	.0172886	-26.9769262	-26.9045554	-1558.297	19	.000
Pair 3	Betweenness Centrality - g - index	-26.9927351	.0287211	.0064222	-27.0061769	-26.9792932	-4203.013	19	.000
Pair 4	Harmonic Centrality - g - index	-23.9750	2.4946	.5578	-25.1425	-22.8075	-42.981	19	.000
Pair 6	Edge Centrality - g - index	-26.9838625	.0109549	.0024496	-26.9889895	-26.9787354	-11015.713	19	.000

Fig. 11. Paired difference analysis of publication to author network.

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Closeness Centrality - g - index	-26.989	.104	.011	-27.011	-26.968	-2483.000	91	.000
Pair 4	Harmonic Centrality - g - index	8.000	8.557	.892	6.228	9.772	8.967	91	.000
Pair 6	Edge Centrality - g - index	-26.870	.339	.035	-26.940	-26.940	-761.086	91	.000

Fig. 12. Paired difference analysis of publication to publications network.

are adjacent to each other. At the same time, it shows a significant influence on the publication citation network. The negative (greater) result of the eigenvector and other parameters not mentioned in the publication citation network are somehow counter-intuitive. Their relationship with g-index could be expected to be better. Another variable could be affected as a result of excessive data values of the comparison variable. It is also not possible for all publications to be connected because of self-citations. We can say that all other measures actively affect researchers' performance with respect to g-index. However, the betweenness variable does not differ largely as compared to other variables. So, with respect to other measures except for the betweenness centrality, it is stated that represented parameters are more effective in the context of all relationships.

Furthermore, we have performed correlation analysis t-test between SNA measures and g-index using the Python language NetworkX tool. To remove the biasedness, all data values were taken randomly for calculation and analysis. These results have slightly different values. In Fig. 13 we calculate the relationship between co-authors. In this analysis, all variables show a strong influence against performance measure except the relations between betweenness centrality and g-index. Based on this, we can say; that in this context, the mentioned significant variables are more effective with respect to the researcher's performance.

Furthermore, in Fig. 14, the relation between publications that are citations of researchers shows that the betweenness centrality, harmonic centrality, eigenvector centrality, edge centrality, and clustering coefficient are significant values. On the other hand, only the degree of centrality is not significant in this scenario. All significant variables exert a positive impact on researchers' performance with g-index. Likewise, we performed the test on publication –area and publication –author relationships as shown in Figs. 15 and 16. Here, all variables are statistically significant except degree and betweenness centrality and not effective to use in both scenarios. We could say that all variables have a positive impact on researchers' performance and profiles with a 95% confidence interval.

Based on all analysis performed, we can conclude that betweenness centrality is the only variable that is not significant in two scenarios, publication to area and publication to author and does not show a positive impact on researcher's performance with respect to g-index. Betweenness centrality is increased by 0.01, which also increases the g-index value. But if we add more authors, it may change centrality measures values. In addition, the reason of betweenness not showing the significant value is the dominance of high values of researchers. For this reason, smaller node values around it take higher betweenness value and give other nodes the lower values.

Co-authorship (T-test results)

	0	1	2	3	4	5	6	7
0	Closeness Centrality		0.00	0.00	0.00	0.00	0.00	0.00
1	Degree Centrality			0.00	0.00	0.05	0.00	0.00
2	Betweenness Centrality				0.00	0.00	0.13	0.02
3	Harmonic Centrality					0.00	0.00	0.00
4	Eigon Centrality						0.00	0.00
5	Clustering Centrality							0.00
6	Edge Centrality							0.00
7	G-index	Degree	Betweenness	Harmonic	Eigon	Clustering	Edge	G-index

P-value < 0.05

Fig. 13. Correlation analysis for co-authorship relation.

Publications to Citations (T-test results)

	0	1	2	3	4	5	6	7
0	Closeness Centrality		0.00	0.00	0.00	0.00	0.00	0.00
1	Degree Centrality			0.00	0.00	0.98	0.00	0.00
2	Betweenness Centrality				0.00	0.00	0.00	0.00
3	Harmonic Centrality					0.00	0.00	0.00
4	Eigon Centrality						0.00	0.00
5	Clustering Centrality							0.00
6	Edge Centrality							0.00
7	G-index	Degree	Betweenness	Harmonic	Eigon	Clustering	Edge	G-index

P-value < 0.05

Fig. 14. Correlation analysis of publications to publications network.

Publications to Area (T-test results)

	0	1	2	3	4	5	6	7
0	Closeness Centrality		0.00	0.00	0.00	0.00	0.00	0.00
1	Degree Centrality			0.00	0.00	0.82	0.22	0.00
2	Betweenness Centrality				0.00	0.00	0.06	0.52
3	Harmonic Centrality					0.00	0.00	0.00
4	Eigon Centrality						0.09	0.00
5	Clustering Centrality							0.00
6	Edge Centrality							0.00
7	G-index	Degree	Betweenness	Harmonic	Eigon	Clustering	Edge	G-index

P-value < 0.05

Fig. 15. Correlation analysis of publications to area network.

Publications to Author (T-test results)

	0	1	2	3	4	5	6	7
0	Closeness Centrality		0.00	0.00	0.01	0.00	0.00	0.00
1	Degree Centrality			0.01	0.01	0.01	0.00	0.05
2	Betweenness Centrality				0.00	0.00	0.23	0.12
3	Harmonic Centrality					0.00	0.00	0.00
4	Eigon Centrality						0.00	0.00
5	Clustering Centrality							0.00
6	Edge Centrality							0.00
7	G-index	Degree	Betweenness	Harmonic	Eigon	Clustering	Edge	G-index

P-value < 0.05

Fig. 16. Correlation analysis of publications to author network.

V. CONCLUSION AND FUTURE WORK

The study of research cultural growth has demonstrated an increased interest in publications and citations rate. The previous literature has involved the use of SNA techniques, significant testing strategies, and randomization to learn and find the influence of network models under some hypothesis with respect to different countries, institutions, and disciplines. Although these models are appropriate according to the network structure assumptions, they may not be the most informative for analysis according to the research area of interest. Presently, no general structural methodology exists to evaluate the network relationship graph or research area over particular measures.

In our collaboration method, we analyzed the citation scores of publications, author's progressive path of an interesting area/domain, author's score, and paper score in the computer science academic field. Fortunately, the study has shown promising results and productivity on the mentioned metrics. We created the following four collaboration networks and used them for our analysis.

- publications to authors
- authors to authors
- publications to publications
- publications to research areas.

The outcomes exhibit the productivity of authors regarding their academic progress in the selected field (computer science) and portrays it as an energetic field with the performed experiments that validate our research work. The significance of this approach is that it provides a template for future perspectives to measure the importance of each researcher in terms of different relations. This paper provides analysis on specific domains of computer science. In future, this can be extended to various research areas and can be analyzed with different datasets. Such future works can give better insights in the influence of publication and citations in various research domains.

REFERENCES

- [1] Iglıc, H., Doreian, P., Kronegger, L., & Ferligoj, A. (2017). "With whom do researchers collaborate and why?," *Scientometrics*, 112(1), 153-174.
- [2] Abramo, G., D'Angelo, A. C., & Murgia, G. (2017). "The relationship among research productivity, research collaboration, and their determinants," *Journal of Informetrics*, 11(4), 1016-1030.
- [3] de Moya-Anegon, F., Guerrero-Bote, V. P., Lopez- Illescas, C., & Moed, H. F. (2018). "Statistical relationships between corresponding authorship, international co-authorship and citation impact of national research systems," *Journal of Informetrics*, 12(4), 1251-1262.
- [4] Katz, J. S., & Martin, B. R. (1997). "What is research collaboration?," *Research policy*, 26(1), 1-18.
- [5] Adams, J. (2012). "Collaborations: The rise of research networks," *Nature*, 490(7420), 335.
- [6] Anwaar, F., Iltaf, N., Afzal, H. and Nawaz, R., 2018. "HRS-CE: A hybrid framework to integrate content embeddings in recommender systems for cold start items," *Journal of computational science*, 29, pp.9-18.
- [7] Ductor, L. (2015). "Does co-authorship lead to higher academic productivity?," *Oxford Bulletin of Economics and Statistics*, 77(3), 385-407.
- [8] Waheed, Hajra, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman, Salem Alelyami, and Raheel Nawaz. "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior* 104 (2020): 106189.
- [9] Stvilia, Besiki, et al. "Toward collaborator selection and determination of data ownership and publication authorship in research collaborations," *Library & Information Science Research* 39.2 (2017): 85-97.
- [10] Kwiek, Marek. "International research collaboration and international research orientation: Comparative findings about European academics," *Journal of Studies in International Education* 22.2 (2018): 136-160.
- [11] Thompson, P., Nawaz, R., Korkontzelos, I., Black, W., McNaught, J. and Ananiadou, S., 2013, October. News search using discourse analytics. In 2013 Digital Heritage International Congress (DigitalHeritage) (Vol. 1, pp. 597- 604). IEEE.
- [12] Chen, Kaihua, Yi Zhang, and Xiaolan Fu. "International research collaboration: An emerging domain of innovation studies?," *Research Policy* 48.1 (2019): 149-168.
- [13] Adams, J., Gurney, K., Hook, D., & Leydesdorff, L. (2014). "International collaboration clusters in Africa," *Scientometrics*, 98(1), 547-556.
- [14] Pritchard, A. (1969). "Statistical bibliography or bibliometrics," *Journal of documentation*, 25(4), 348-349.
- [15] Leung, X. Y., Sun, J., & Bai, B. (2017). "Bibliometrics of social media research: a co-citation and coword analysis," *International Journal of Hospitality Management*, 66, 35-45.
- [16] Bordons, M., Aparicio, J., González-Albo, B., & Díaz-Faes, A. A. (2015). "The relationship between the research performance of scientists and their position in co-authorship networks in three fields," *Journal of Informetrics*, 9(1), 135-144.
- [17] Csomós, G., & Lengyel, B. (2018). "Mapping the efficiency of international scientific collaboration between cities worldwide," *arXiv preprint arXiv:1808.03730*.
- [18] Wai-Chan, S. (2017). International research collaboration creates higher impact.
- [19] Freshwater, D., Sherwood, G., & Drury, V. (2006). "International research collaboration: Issues, benefits and challenges of the global network," *Journal of Research in Nursing*, 11(4), 295-303
- [20] Türker, İlker, and Abdullah Çavuşoğlu. "Detailing the co-authorship networks in degree coupling, edge weight and academic age perspective," *Chaos, Solitons & Fractals* 91 (2016): 386-392.
- [21] Wang, Jian. "Knowledge creation in collaboration networks: Effects of tie configuration," *Research Policy* 45.1 (2016): 68-80.
- [22] Morris, Steven A., and Michel L. Goldstein. "Manifestation of research teams in journal literature: A growth model of papers, authors, collaboration, coauthorship, weak ties, and Lotka's law," *Journal of the American Society for Information Science and Technology* 58.12 (2007): 1764-1782.
- [23] Kundra, Ramesh, et al. "Studies in Co-authorship Pairs Distribution: Part-2: Co-author pairs' frequencies distribution in journals of gender studies," *COLLNET Journal of Scientometrics and Information Management* 2.1 (2008): 63-71.

- [24] Egghe, Leo. "A model for the size-frequency function of coauthor pairs," *Journal of the American Society for Information Science and Technology* 59.13 (2008): 2133- 2137.
- [25] Naldi, Fulvio, et al. "Scientific and technological performance by gender," *Handbook of quantitative science and technology research*. Springer, Dordrecht, 2004. 299-314.
- [26] Carr, Phyllis L., et al. "Collaboration in academic medicine: reflections on gender and advancement," *Academic Medicine* 84.10 (2009): 1447-1453.
- [27] Pepe, Alberto, and Marko Rodriguez. "Collaboration in sensor network research: an in-depth longitudinal analysis of assortative mixing patterns," *Scientometrics* 84.3 (2009): 687-701.
- [28] Kretschmer, Hildrun, Bulent Ozel, and Theo Kretschmer. "Who is collaborating with whom? Part I. Mathematical model and methods for empirical testing," *Journal of Informetrics* 9.2 (2015): 359-372.
- [29] Liu, Jie, et al. "A new method to construct coauthor networks," *Physica A: Statistical Mechanics and its Applications* 419 (2015): 29-39.
- [30] Abbasi, Alireza, Jörn Altmann, and Liaquat Hossain. "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures," *Journal of Informetrics* (2011): 594- 607.
- [31] Bordons, María, et al. "The relationship between the research performance of scientists and their position in co-authorship networks in three fields," *Journal of Informetrics* 9.1 (2015): 135-144.
- [32] Maggioni, Mario A., and Teodora Erika Uberti. "Networks and geography in the economics of knowledge flows," *Quality & quantity* 45.5 (2011): 1031- 1051.
- [33] De Stefano, Domenico, et al. "The use of different data sources in the analysis of co-authorship networks and scientific performance," *Social Networks* 35.3 (2013): 370-381.
- [34] Mourao, Paulo Reis, and Vítor Domingues Martinho. "Forest entrepreneurship: A bibliometric analysis and a discussion about the co-authorship networks of an emerging scientific field," *Journal of Cleaner Production* 256 (2020): 120413.
- [35] Jeon, H. J., O. J. Lee, and J. J. Jung. "Is Performance of Scholars Correlated to Their Research Collaboration Patterns," *Front. Big Data* 2: 39. (2019).
- [36] Mutz, Rüdiger, and Hans-Dieter Daniel. "The bibliometric quotient (BQ), or how to measure a researcher's performance capacity: A Bayesian Poisson Rasch model," *Journal of Informetrics* 12.4 (2018): 1282- 1295.
- [37] Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Flavia Di Costa. "The effect of multidisciplinary collaborations on research diversification," *Scientometrics* 116.1 (2018): 423- 433.
- [38] Clemente-Gallardo, J., Ferrer, A., Íñiguez, D., Rivero, A., Ruiz, G., & Tarancón, A. (2019). "Do researchers collaborate in a similar way to publish and to develop projects?," *Journal of Informetrics*, 13(1), 64-77.
- [39] Kwiek, Marek. "International research collaboration and international research orientation: Comparative findings about European academics," *Journal of Studies in International Education* 22.2 (2018): 136-160.
- [40] Kwiek, M. (2018). "High research productivity in vertically undifferentiated higher education systems: Who are the top performers?," *Scientometrics*, 115(1), 415-462.
- [41] Garfield, Eugene. "Is citation analysis a legitimate evaluation tool?," *Scientometrics* 1.4 (1979): 359-375.
- [42] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2009).
- [43] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008).
- [44] Kong, Xiangjie, et al. "How does collaboration affect researchers' positions in co-authorship networks?," *Journal of Informetrics* 13.3 (2019): 887-900.
- [45] Hayat, Tsahi, Dimitrina Dimitrova, and Barry Wellman. "The differential impact of network connectedness and size on researchers' productivity and influence," *Information, Communication & Society* 23.5 (2020): 701-718.



Sidra Razzaq

Sidra Razzaq has recently completed her Masters degree in Computer Science from COMSATS University at Islamabad (CUI), Islamabad, Pakistan. Her Research Interests include data science, social network analysis, and access control.



Ahmad Kamran Malik

Ahmad Kamran Malik received the Ph.D. degree from the Vienna University of Technology (TU-Wien), Vienna, Austria. He is currently an Assistant Professor with COMSATS University at Islamabad (CUI), Islamabad, Pakistan. He has published a book and many research papers in reputed international journals and conferences. His current research interests include data science, social network analysis, and access control.



Basit Raza

Basit Raza received the master's degree in computer science from the University of Central Punjab, Lahore, Pakistan, and the Ph.D. degree in computer science from International Islamic University Islamabad and the University of Technology Malaysia, in 2014. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University at Islamabad (CUI), Islamabad, Pakistan. He has authored several articles in refereed journals. His research interests include database management systems, data mining, data warehousing, machine learning, deep learning, and artificial intelligence. He has been serving as a Reviewer for prestigious journals, such as Applied Soft Computing, Swarm and Evolutionary Computation, Swarm Intelligence, Applied Intelligence, IEEE Access, and Future Generation Computer Systems.



Hasan Ali Khattak

Hasan Ali Khattak received his Ph.D. in Electrical and Computer Engineering degree from Politecnico di Bari, Bari, Italy on April 2015, Master's degree in Information Engineering from Politecnico di Torino, Torino, Italy, in 2011, and B.CS. degree in Computer Science from the University of Peshawar, Peshawar, Pakistan in 2006. He is currently serving as Associate Professor - School of Electrical Engineering and Computer Science, National University of Sciences and Technology - Pakistan since October 2020. His current research interests focus on Future Internet Architectures such as the Web of Things and leveraging Data Sciences and Social Engineering for Future Smart Cities. Along with publishing in good research venues and completing successful funded National and International funded projects, he is also serving as a reviewer in reputed venues such as IEEE Access, IEEE Network Magazine, IEEE Consumer Electronics, Hindawi, SAI, IET, and a few national publishers. He is currently involved in several funded research projects in various domains such as Healthcare Information Management, Semantic Web of Things, and Fog Computing while exploring Ontologies and other Semantic Web Technologies. He has worked on Contiki OS, NS 2/3, and Omnet++ frameworks. His perspective research areas are the application of Machine Learning and Data Sciences for improving and enhancing Quality of life in Smart Urban Spaces especially in the Healthcare and Transportation domain through Knowledge management, predictive analysis, and visualization. He is an active Senior Member of IEEE, a professional member of ACM, and a member of societies such as IEEE ComSoc, IEEE VTS, and Internet Society.



Giomar W. Moscoso Zegarra

Prorrector and associate professor at Escuela de Posgrado Newman. RENACYT researcher qualified by CONCYTEC. Candidate for a Doctorate in Accounting and Finance, Master of Science with a mention in Auditing, Master of Business Administration, Commercial Engineer from the Tarapacá University of Chile, Certified Public Accountant, Bachelor of Administration from the Tarapacá University of Chile. Experience in financial, research and academic work at management level. Head of licensing teams and processes related to SUNEDU. Former University Defender at Escuela de Posgrado Newman and General Director of the scientific journal Iberoamerican Business Journal.



Yvan Díaz Zelada

Consultant with experience in business organization and management, market research, formulation and evaluation of projects and design of business models. He is a researcher recognized by CONCYTEC in the María Rostworowski II category of RENACYT. Doctoral candidate in Administration, graduated with maximum distinction as an MBA from the Master in Business Administration and Management from the University of Tarapacá (Chile), Master in Business Administration from Escuela de Posgrado Newman (Peru) and graduated from the Banking Administration Program by Catholic Center (Peru). Studies in commercial engineering at the Tarapacá University of Chile, and collegiate systems engineer (CIP 112273). He is currently Academic Director of Escuela de Posgrado Newman. He served as director of the e-Learning center for Escuela de Posgrado Newman. He has been a business officer for Banco de Credito del Peru (BCP), Mi Banco and Banco Solventa. He teaches at Escuela de Posgrado Newman, at Universidad San Pablo (Arequipa) and Universidad Continental (Arequipa). He has taught at different Peruvian universities in the south of Peru (Tacna and Arequipa) in the lines of business, innovation, administration and technology.