

RESEARCH

Open Access



A new insight to the analysis of co-authorship in Google Scholar

Ghazal Kalhor, Amin Asadi Sarijalou, Niloofar Sharifi Sadr and Behnam Bahrak* 

*Correspondence:
bahrak@ut.ac.ir
School of Electrical
and Computer Engineering,
College of Engineering,
University of Tehran, Tehran,
Iran

Abstract

Google Scholar is a search engine for scholarly literature which indexes most academic papers, dissertations, and books that are available online. This paper aims to analyze the characteristics of the manually added co-authorship network (MACN), in which nodes are authors who manually add their collaborators from a list of co-authors suggested by Google scholar based on their joint scholarly work. In addition to this network, we also perform structural analysis on the authors' fields of interest network (FIN), and their affiliated institute network (AIN). We introduce a new citation metric based on the distribution of authors' citation count, which captures the position of authors in their research area and can help us in ranking the universities in each scientific field.

Keywords: Google Scholar, Manually added co-authorship network, Scientometrics

Introduction

There are several ways to find research resources, but Google Scholar is one of the most convenient tools for scholars working in various fields. In addition to the millions of documents indexed in Google Scholar, many authors have public profiles on this platform that include their publications, the number of citations of their work for each year, and their h-index. It is undeniable that the information provided by Google Scholar enables researchers to measure the impact of their studies and others' scholarly work in the scientific community.

An interesting feature of Google Scholar is a list of potential co-authors suggested by the platform to the users based on their joint scholarly work with other users. Users can select co-authors from the list and add them to their profiles. In this paper, we focus on the network of co-authors, manually added by the Google Scholar users. In this network, which is named Manually Added Co-authorship Network (MACN), nodes are authors with a public profile on Google Scholar, and edges represent the co-authorship relationships that they have manually added to their profile.

In prevalent co-authorship networks, the co-authorship relation between nodes is extracted from the author list of the papers. In other words, whenever the names of two authors appear on a paper, an edge is added between their corresponding nodes. These networks are usually undirected and weighted (weights indicate the number of papers the authors have published together). The main limitation of such networks is that in many

cases, there are authors whose names appear in the author list of a paper, but they do not have any direct research collaboration, and they may not even regard each other as co-authors. For example, there are papers with more than 5000 authors (Lobo 2021) who are unlikely to know all of their co-authors. In contrast, in MACN, the authors voluntarily add their research collaborators from a suggested list which addresses the raised issue. As a result, MACN is much sparser than an ordinary co-authorship network, and unlike the latter, it is directed and unweighted.

The inspiring questions for our study are: (i) what is the motivation of people in adding their co-authors manually? (ii) who adds more collaborators? (iii) Whom do authors prefer to add?

Our contributions can be summarized as follows:

1. We collected a dataset of 496486 Google Scholar public profiles and enriched this data by assigning standard fields of study to 99.81% of them using different clustering-based and community-based approaches.
2. We built an MACN using the collected dataset and investigated it by performing various analyses, calculating, and interpreting several metrics on this network.
3. We depicted the relationship between the number of authors, averages of citation count and h-index at each community of the MACN to check whether they are linearly correlated or not.
4. We computed assortativity coefficients of the MACN to find the attributes in which the similarity leads authors to collaborate.
5. We made an exponential random graph model (ERGM) based on the attributes of the nodes of the MACN to find the features that are statistically significant in forming the network edges.
6. We compared the MACN with an ordinary co-authorship network of Google Scholar to see whether they follow similar structures or not.
7. We built the authors' fields of interest network (FIN) and their affiliated institute network (AIN) using the dataset and calculated centrality metrics such as PageRank, betweenness, and closeness for FIN to find the most centric fields of interest.
8. We analyzed the distributions of club-coefficient to discover the trend by which distinguished authors select their co-authors.
9. We investigated the authors' citation distribution in four universities with the highest number of authors in our dataset and defined a new measure of citations for institutes that captures authors' position in their research area and can be useful in ranking the universities in each scientific field.

The remainder of this paper is structured as follows. “[Related work](#)” section provides an overview of previous related work. In “[Methodology](#)” section we describe the dataset and present the measures and methods used in this study. Then, we present and discuss our results in “[Results](#)” section. Finally “[Conclusion](#)” section concludes the paper.

Related work

Various studies have been conducted on online bibliographic platforms such as ResearchGate, Microsoft Academic Search Service, and Clarivate's Web of Science. Some of these studies mainly concentrated on the yearly evolution of the co-authorship graph. Sarigöl et al. (2014) studied time-evolving collaboration networks and citation numbers on Microsoft Academic Search Service that contains fewer documents and citations than those in Google Scholar (Ortega and Aguillo 2014). Yan and Ding (2009) calculated centrality measures on the co-authorship network of the Institute for Scientific Information. They found that centrality measures are helpful in author ranking as these metrics are strongly correlated with citation counts. Also, they measured both article impact and author's field impact. Higaki et al. (2020) built the co-authorship network from publications indexed in Clarivate's Web of Science and compared it with the equivalent Barabasi Model. They found that the co-authorship network is similar to scale-free networks. Some of the previous studies focused on specific research areas to build the co-authorship network. For instance, Newman (2004) has constructed collaboration networks between scientists in physics, biomedical engineering, and computer science. Some studies tried to assess gender parity in the co-authorship network. Bravo-Hermsdorff et al. (2019) discussed the relevance of gender in scientific collaboration patterns in the Institute for Operations Research and the Management Sciences (INFORMS). Findings of this study showed that the INFORMS society was far from gender parity in many crucial local statistics such as the number of publications, homophily, and author order.

There are few studies on Google Scholar. Chen et al. (2017) built the co-authorship network of Google Scholar based on the list of authors appearing in papers and calculated several metrics such as PageRank, clustering coefficient, and degree. They also explored the correlation between the co-authorship network metrics and citation metrics. They found that there is a strong correlation between PageRank and h-index. Ortega and Aguillo (2013) performed country and institutional level analysis on Google Scholar to find the dominant country in the research area. Tang et al. (2021) concentrated on misconfigured author profiles detection on Google Scholar and used a method for labeling these profiles. To the best of our knowledge, there is no analysis of the manually added co-authorship network. Moreover, no study has been conducted on the fields of interest graph obtained from Google Scholar.

Methodology

In this section, we introduce the dataset that we collected for this study. We also define networks that are constructed using the dataset, and discuss the metrics and techniques which are used for analyzing these networks.

Dataset

We developed a web crawler in Python for collecting the data used in this study directly from Google Scholar. First, we created a list of 65 universities in different countries and found their pages on Google Scholar. Then, we crawled the profiles of authors on the first page of each university, i.e., the most cited researchers of these institutes. Finally, we collected information about the co-authors of these authors. The data collection is stopped on March 31, 2021. Moreover, we added information about the authors' gender and country based on their names using the NameToGAN tool (<https://quecst.qcri.org/tool/Name2GAN>).

To enrich our dataset, we determined the standard field of interest for each author in our dataset from a list of 39 general fields available in Table 6. We used a 3-step approach for this goal. In the first step, we found a field of interest for authors using the name of the journals where they had published their work. We used Scimago Journal & Country Ranking (SJR) (<https://www.scimagojr.com>) to map each journal name to one of the 39 general fields of interest. Using this method, 80.83% (401311 out of 496486) of users were assigned a field of interest. In the second step, we used non-general fields of interest that authors have added to their profiles. There are 221759 different non-general fields of interest in our dataset. We used Mini-Batch K-Means clustering (Sculley 2010) to cluster them into 39 standard fields of interest. Then we assigned a standard field of interest to authors by considering the clusters of the majority of their unstandardized fields of interest. By this approach, we assigned the standard fields of interest to 77.78% (74032 out of 95175) of the remaining users. In the third step, we used the community structure of the MACN. Using the Infomap community detection algorithm (Rosvall and Bergstrom 2008) we found communities in the MACN. Then, we considered the most frequent field of interest of users in each community as the field of interest for all the users in that community. By this method, we were able to assign the standard fields of interest to 99.01% (94234 out of 95175) of the remaining users. After combining the results of these three steps, 99.81% (495545 out of 496486) of the users got a standard field of interest. Table 1 provides an overview of the features available in our dataset.

Table 1 The dataset features and their explanations

Feature	Explanation
Author URL	The URL linking us to the author's page on Google Scholar
Name	The author's full name
Institute URL	The URL linking us to the author's affiliation page on Google Scholar
Citation Count	The author's number of citations
h-index	The author's h-index
Co-author URLs	The URLs linking the author's profile to the co-authors' profiles on Google Scholar
Fields of Interest	The fields of interest added by the author
Publisher Names	The names of journals/conferences where the author has published papers
Gender	The author's gender
Country	The author's country

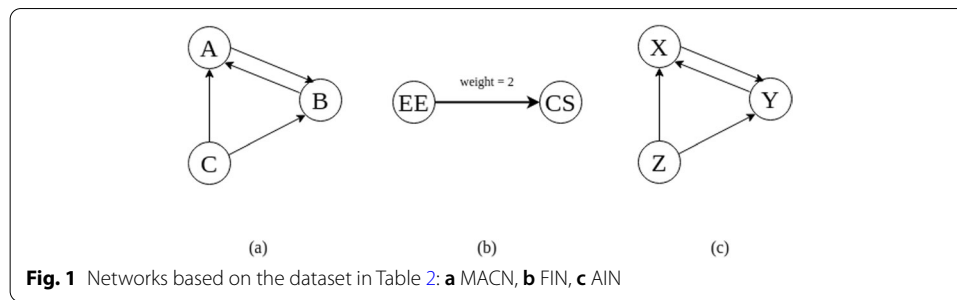


Table 2 An example of dataset of Google Scholar authors for constructing networks

Author	Institute	Field of Interest	Co-authors
A	X	CS	B
B	Y	CS	A
C	Z	EE	A, B

Manually added co-authorship network

We built a directed unweighted manually added co-authorship network (MACN) using our dataset. In this network, the set of nodes represents authors with public profiles on Google Scholar, and an edge from node u to node v means that u has added v as a co-author into her/his profile.

Field of interest network

We built a directed weighted network that represents the connections between standard fields of interest of authors in Google Scholar. In this field of interest network (FIN), the set of nodes represents the standard fields of interest of authors according to their profile on Google Scholar, and an edge of weight W from node u to node v means that there are W added co-authorship relations between authors of field u and authors of field v .

Affiliated institute network

We built a directed weighted network that represents the cooperation between institutes in Google Scholar. In this affiliated institute network (AIN), the set of nodes represents the institute of users of Google Scholar, and an edge of weight W from u to v means that there are W added co-authorship relations between authors of institute u and authors of institute v (Fig. 1).

Metrics

There are various metrics for the quantitative analysis of complex networks. We used the most common social network analysis metrics that are meaningful in the context of our dataset.

Definition 1 Transitivity of a network is defined as the fraction of all possible triangles present in the graph (Yang 2013).

$$C = \frac{\text{number of triangles} \times 3}{\text{number of connected triples of vertices}}$$

Definition 2 Edge reciprocity of a directed network is defined as the ratio of the number of edges pointing in both directions to the total number of edges in the network (Costa et al. 2007).

Definition 3 Modularity measures the strength of division of a network into modules and is defined as:

$$Q = \sum_{c=1}^n \left[\frac{L_c}{m} - \left(\frac{k_c}{2m} \right) \right]$$

where the sum iterates over all communities c , m is the number of edges, L_c is the number of intra-community links for community c , and k_c is the sum of degrees of the nodes in community c (Clauset et al. 2004).

Definition 4 Average clustering coefficient of a network measures the degree to which nodes intend to cluster together and is defined as:

$$C = \frac{1}{n} \sum_{v \in G} c_v$$

where n is the number of nodes in the network and c_v is the local clustering coefficient for node v (Kaiser 2008).

Definition 5 Assortativity coefficient of a network indicates the tendency of its nodes to attach to other nodes that are similar to them. The similarity of two nodes is usually measured using a given nodal attribute such as degree (Newman 2003).

Definition 6 Rich-club coefficient of an undirected network reflects the tendency of hubs (nodes of a higher degree) to be better connected among themselves than nodes with a smaller degree. It is defined as:

$$\phi(k) = \frac{2E_k}{N_k(N_k - 1)}$$

where N_k is the number of nodes with degrees larger than k , and E_k is the number of edges among those nodes (McAuley et al. 2007).

Definition 7 Betweenness centrality of a node measures how frequently that node lies on short paths between other pairs of nodes. It is defined as:

$$B_u = \sum_{ij} \frac{\sigma(i, u, j)}{\sigma(i, j)}$$

where $\sigma(i, u, j)$ is the number of shortest paths between vertices i and j that pass through vertex or edge u , $\sigma(i, j)$ is the total number of shortest paths between i and j , and the sum is over all pairs i, j of distinct vertices (Beveridge and Shan 2016).

Definition 8 Closeness centrality of a node is the average distance from the node to all other nodes and is defined as:

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v, u)}$$

where $d(v, u)$ is the shortest-path distance between v and u , and n is the number of nodes that can reach u (Beveridge and Shan 2016).

Definition 9 Weighted degree centrality of a node is the sum of the weights of the edges incident with that node (Beveridge and Shan 2016).

Definition 10 Eigenvector centrality computes the centrality for a node based on the centrality of its neighbors. The eigenvector centrality for node i is the i -th element of the vector x defined by the equation

$$Ax = \lambda x$$

where A is the adjacency matrix of graph G with eigenvalue λ (Yang 2013).

Definition 11 PageRank of a node in a directed graph computes a ranking of that node based on the structure of the incoming links. It is defined as:

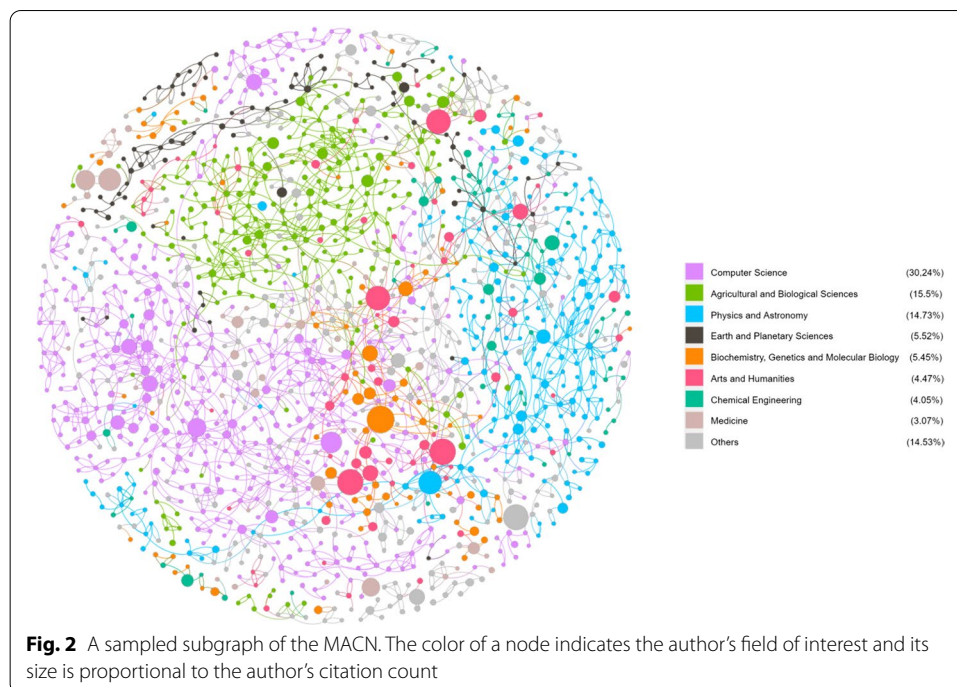
$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{d_j^{out}} + \beta$$

where α and β are positive constants and A_{ij} is an element of the adjacency matrix (Page et al. 1999).

Methods

In this section, we introduce the methods that are employed in this study.

In this study, we use exponential random graph models (ERGM), which are a family of models that are suitable to model the formation of dyads in relational data like the network datasets. They describe the local selection forces that shape the global structure of a network (Robins et al. 2007). ERGMs allow us to include and control node attributes, structural configurations, and edge attributes.



Mini-Batch K-Means clustering is a variant of the KMeans algorithm that uses mini-batches to reduce the computation time while still attempting to optimize the same objective function (Sculley 2010). We use K-Means clustering for standardizing authors' fields of interest.

Results

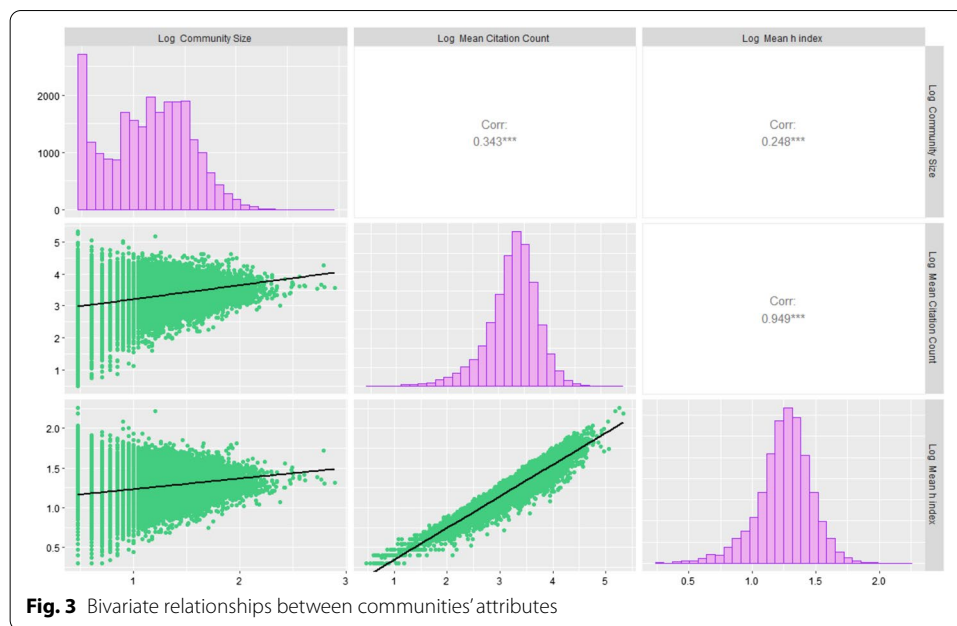
This section provides a thorough explanation of the methods used to analyze the data and provide some insight into the meaning of our results.

MACN of Google Scholar

Figure 2 shows a sample of the MACN. Each node is selected with a probability that is proportional to its degree. The color of a node and its size indicate the author's field of interest and citation count, respectively. We can see that Computer Science is the dominant field of interest in this network. This observation confirms the finding of Ortega and Aguillo (2014) that Google Scholar has a strong bias towards the Information and Computing sciences.

Analysis of communities

In this section, we perform a bivariate analysis of communities' attributes for the MACN. As mentioned earlier, we used the Infomap community detection algorithm (Rosvall and Bergstrom 2008) to detect communities for the MACN. For each community of the MACN, we calculated attributes such as the size of the community

**Table 3** Assortativity coefficients for different attributes in the MACN

Attribute	Assortativity coefficient
Field of interest	0.525
h-index	0.224
Country	0.220
Institute	0.193
Citation count	0.162
Gender	0.095
Node degree	0.088

(number of authors in that community), mean citation count, and mean h-index. The relationships between these attributes is illustrated in Fig. 3. We applied the logarithm transformation to all three variables because the distribution of community size is right-skewed. The correlation method that we used in this analysis was Pearson correlation.

As it is illustrated in Fig. 3, there is a strong positive linear relationship between mean citation count and mean h-index of communities. Our finding confirms the result of Yong (2014), where the authors argued that the h-index is approximately equal to 0.54 times the square root of citation count for typical scientists. However, the correlations between the community size and the rest of the variables are weak.

Table 4 Coefficients for ERGM of the MACN

Factor	Estimate	Standard error	<i>p</i> value
Edges	− 3.871	0.103	< 0.01***
Field of interest	2.744	0.070	< 0.01***
Country	1.187	0.078	< 0.01***
Institute	4.442	0.090	< 0.01***
Gender	1.265	0.074	< 0.01***
Citation count	− 0.012	0.002	< 0.01***
h-index	0.384	0.023	< 0.01***

Table 5 Structural metrics of the MACN

Metric	Measurement
Reciprocity	0.311
Clustering coefficient	0.188
Average shortest path length	6.384

Interpretation of assortativity coefficients

One of the important metrics usually calculated on social networks is assortativity. This metric shows whether nodes tend to communicate with other nodes which are similar to them or not. In this section, we calculate assortativity based on different measures of similarity (using various attributes) for the MACN. You can see the result in Table 3. Based on the results, all the assortativity coefficients are positive. A positive value of assortativity for a similarity measure based on an attribute indicates a tendency between nodes (authors) with a similar value for that attribute to make a connection (add each other as a co-author). Due to the positive value for the h-index attribute in Table 3, it is more probable that two people from the same field add each other as co-authors and form an edge in the MACN.

As it can be seen, the highest value of this metric is for the field of interest of authors. Thus, we can deduce that authors with the same field of interest tend more to collaborate. Additionally, attributes such as country, institute, and citation count have also relatively large positive assortativity values in this network. Therefore, the users that are co-authors based on the MACN are more probable to be from the same country, from the same institute, or to have close citation counts. Besides, gender and node degree have small assortativity coefficients; thus, these attributes do not affect the connection between nodes in the MACN.

Interpretation of ERGM coefficients

In this section, we calculated ERGM for the MACN. Table 4 shows the result of this study. The coefficient for the Edges term is negative, indicating that the density of the network is below 50%. A negative Edges coefficient is a typical feature of an observed

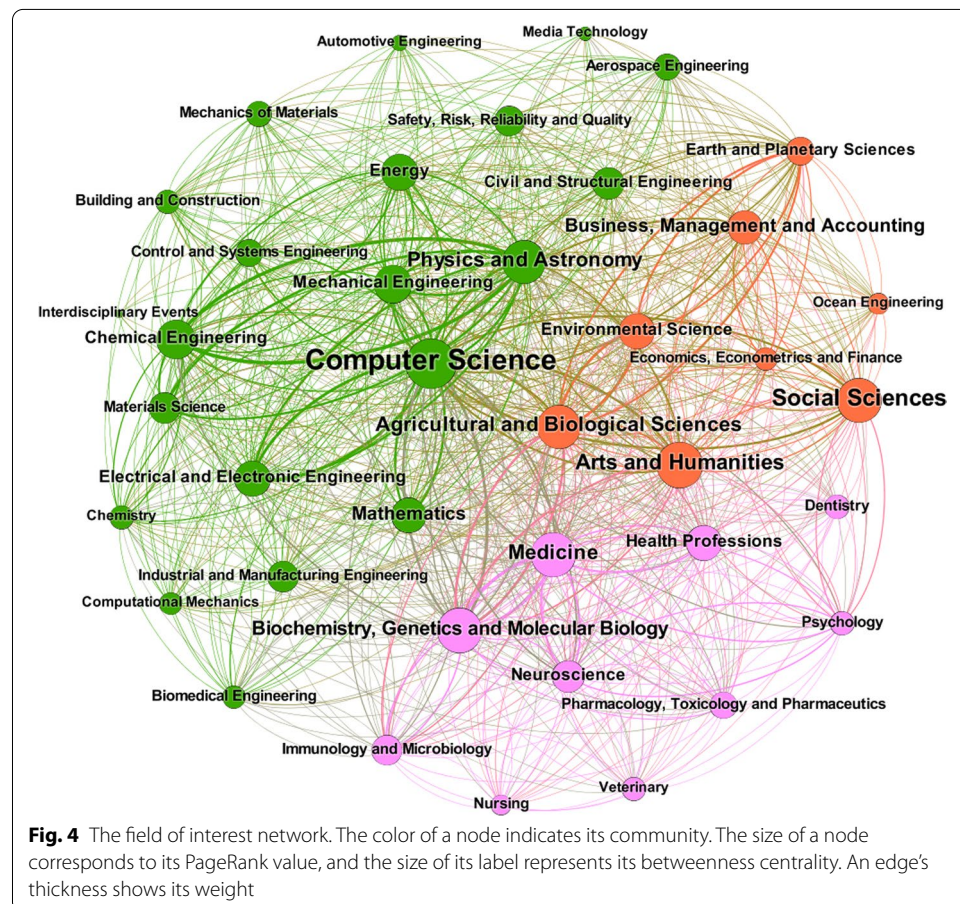
network; very few observed networks have a density of 0.5 or higher. Most network models will contain negative Edges terms (Harris 2018).

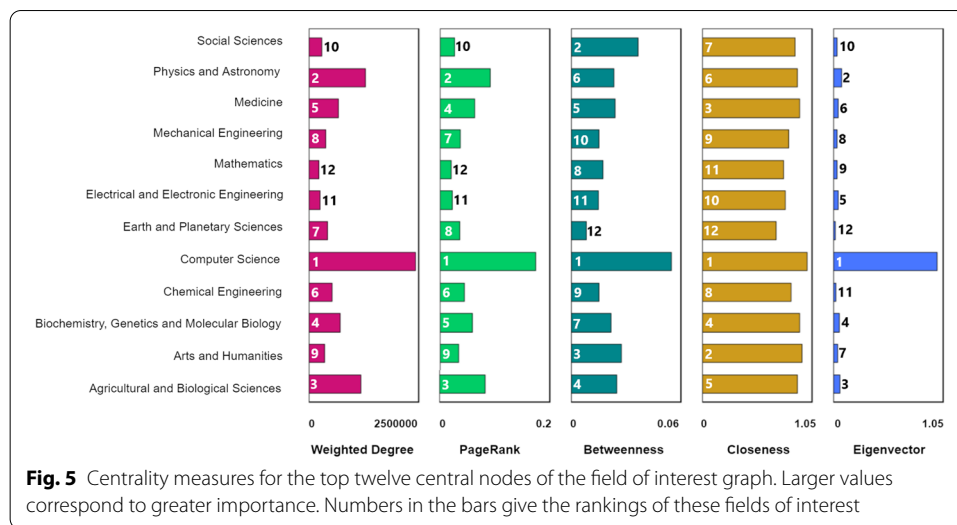
All estimates are statistically significant. Significant positive coefficients for attributes such as country show that two users with the same value for these attributes are more likely to be connected. In other words, they are more likely to co-author a paper. A significant negative coefficient for citation count means that two users with different values for this metric are more likely to be connected.

We should also note that the significance of the gender attribute might be due to our unbalanced dataset, in which most users (70.10%) are male.

Interpretation of structural characteristics

We investigated the structural characteristics of the MACN using reciprocity, clustering coefficient, and average shortest path length. Table 5 represents the values of these metrics. The value of edge reciprocity shows that only 31.10% of the edges in the MACN are bidirectional. Therefore, we can observe that many authors do not tend to add all of their co-authors to their profiles. For all unidirectional edges in the MACN, we calculated the difference between the endpoints (authors)' h-index (the h-index of head minus the h-index of tail). The average of these values was 2.302, a large positive number, which





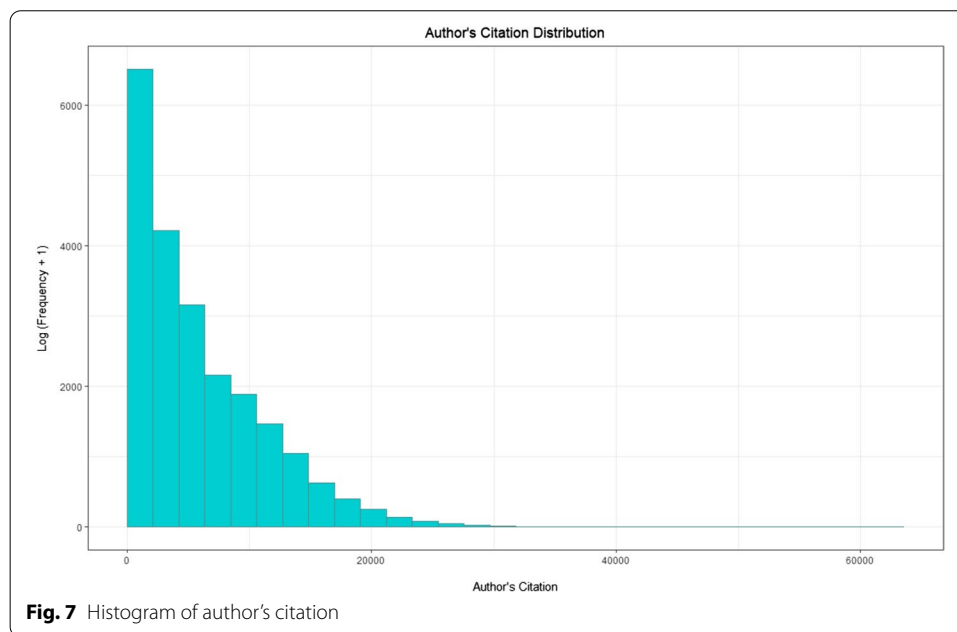
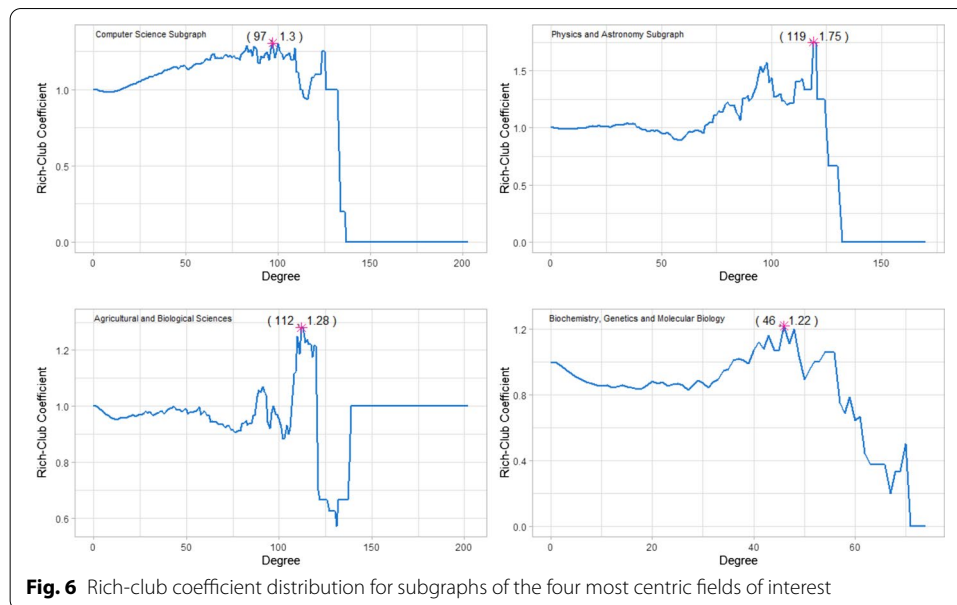
indicates that users usually tend to add co-authors whose h-index are higher than their h-index, i.e. more well-known researchers.

We considered the largest weakly connected component of the MACN to calculate clustering coefficient and average shortest path length. Chen et al. (2017) computed these two metrics for the largest connected component of the co-authorship network of Google Scholar constructed from the author list of papers. Based on their results, the value of the clustering coefficient was 0.30 and the value of the average shortest path length was 5.96. As you can see in Table 5, these values are different for the MACN. The lower average clustering coefficient for the MACN means that there are more weak ties in this network compared to the ordinary network. In other words, the probability that the co-authors of a user collaborate is less than that of the ordinary network. In addition, based on the higher value of the average shortest path in the MACN, more steps are needed to get from a randomly chosen author to another. This result is in agreement with our previous finding that the MACN is not a well-connected network.

Centrality metrics of the FIN

Using the disparity filter algorithm (Serrano et al. 2009) as a sparsification technique, we reduced the density of the FIN from 0.947 to 0.639 in order to remove edges with low weight and make the values of centrality metrics more meaningful. Then, we calculated centrality metrics to find the most critical (or centric) fields of interest in the FIN. Since there are numerous centrality measures, we calculated five of them that are more meaningful in this context. Figure 4 shows an overview of the field of interest network. The size of a node and the size of its label represent the value of its PageRank and betweenness, respectively. The nodes are colored based on their community which is detected using the Louvain community detection algorithm (Blondel et al. 2008).

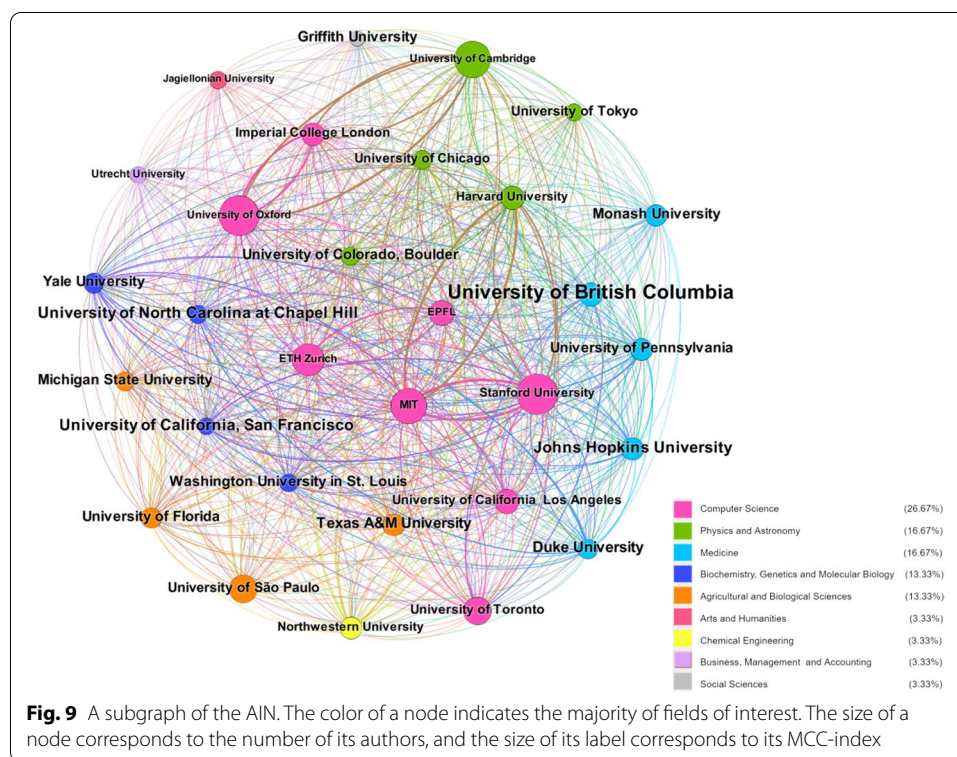
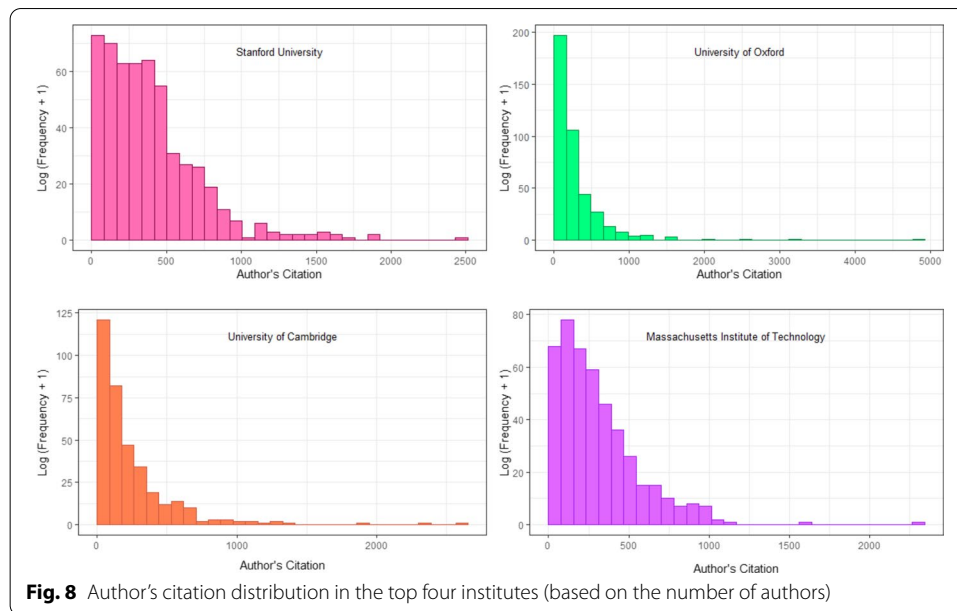
Additionally, values of calculated centrality metrics for the top twelve central fields of interest can be seen in Fig. 5. According to this figure, the PageRank works almost the same as the weighted degree centrality for this network.



This led us to analyze other centrality metrics which have a more global view of the graph. For this goal, we considered closeness and betweenness centralities. Ordering for betweenness centrality is also different from the ordering of other metrics. However, Computer Science (CS) keeps its rank as the first.

The eigenvector values for our list of most centric fields of interest is about zero, except for the faraway Computer Science.

Besides, in this network, three fields of interest stand out consistently: Computer Science, Physics and Astronomy, and Agricultural and Biological Sciences. Computer Science is a highly interdisciplinary scientific domain having significant overlaps with



mathematics, physics, and even biology (Fiala and Tutoky 2017). CS research combines aspects of engineering and natural sciences (in Systems) as well as mathematics (Meyer et al. 2009). Our findings confirm the results of Fiala and Tutoky (2017) and Meyer et al. (2009).

Interpretation of rich-club coefficients

We studied the subgraphs in which all authors have the same fields of interest, separately. Then, we marked the degree to which nodes have the highest value for the rich-club coefficient. For this goal, we considered the four most centric fields of interest in the FIN to separate their corresponding subgraphs from the MACN. Each of the plots in Fig. 6 shows the rich-club coefficient for different possible degrees in each of these subgraphs. As it is presented, the highest value for this metric appears in high degrees. This confirms the claim of Colizza et al. (2006) that in science, influential researchers of some research areas tend to form collaborative groups and publish papers together.

Introducing a new citation metric

Figure 7 shows the histogram of the distribution of users' citation counts. As we can see, the distribution is right-skewed.

Moreover, we saw the same pattern for this distribution among institutes with the highest number of authors. Figure 8 indicates the citation distribution in four of these institutes. Therefore, the median can be a robust measure of center for this distribution. Based on these observations, we defined a new citation metric to evaluate academic institutes. We named this metric MCC-index (MCC stands for Median Citation Count) which is the median of the authors' number of citations of an institute.

Figure 9 illustrates an overview of a subgraph of the AIN. The size of a node (an institute) and the size of its label represent the number of its authors and its MCC-index, respectively. Besides, node color indicates the majority of fields of interest of its authors.

Conclusion

In this paper, we analyzed the manually added co-authorship network (MACN) of Google scholar users. We discussed the effect of various attributes on forming the links between users and observed that the field of interest plays the most crucial role. We showed that there is a linear correlation between h-index and citation count. Moreover, we studied different structural properties of the MACN, such as reciprocity, clustering coefficient, and average shortest path length. Then we compared these values with those of the ordinary co-authorship network of Google Scholar, and the result of this comparison shows that the two networks do not have similar structures, and the MACN is not as connected as the ordinary network. Also, we found that there is a tendency to collaborate and publish papers together among influential authors. The observed results for both the field of interest and institute networks confirm the prominent position of Computer Science in the scholar space. We also defined a new citation metric called MCC-index to assess institutes in the specific research areas.

There is a limitation to this study. Due to the large size of the MACN and the complexity of the calculations of structural terms in ERGM, we had to exclude those terms from the model and only focus on nodal attributes.

Appendix

See Table 6.

Table 6 Standard fields of interest in alphabetical order

Field of interest
Aerospace Engineering
Agricultural and Biological Sciences
Arts and Humanities
Automotive Engineering
Biochemistry, Genetics and Molecular Biology
Biomedical Engineering
Building and Construction
Business, Management and Accounting
Chemical Engineering
Chemistry
Civil and Structural Engineering
Computational Mechanics
Computer Science
Control and Systems Engineering
Dentistry
Earth and Planetary Sciences
Economics, Econometrics and Finance
Electrical and Electronic Engineering
Energy
Environmental Science
Health Professions
Immunology and Microbiology
Industrial and Manufacturing Engineering
Interdisciplinary Events
Materials Science
Mathematics
Mechanical Engineering
Mechanics of Materials
Media Technology
Medicine
Neuroscience
Nursing
Ocean Engineering
Pharmacology, Toxicology and Pharmaceuticals
Physics and Astronomy
Psychology
Safety, Risk, Reliability and Quality
Social Sciences
Veterinary

Author contributions

All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 26 November 2021 Accepted: 28 March 2022

Published online: 08 April 2022

References

- Beveridge A, Shan J (2016) Network of thrones. *Math Horizons* 23(4):18–22
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bravo-Hermesdorff G, Felso V, Ray E, Gunderson LM, Helander ME, Maria J, Niv Y (2019) Gender and collaboration patterns in a temporal scientific authorship network. *Appl Netw Sci* 4(1):1–17
- Chen Y, Ding C, Hu J, Chen R, Hui P, Fu X (2017) 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
- Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. *Phys Rev E* 70(6):066111
- Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nat Phys* 2(2):110–115
- Costa LdF, Rodrigues FA, Travieso G, Villas Boas PR (2007) Characterization of complex networks: a survey of measurements. *Adv Phys* 56(1):167–242
- Fiala D, Tutok G (2017) Computer science papers in web of science: a bibliometric analysis. *Publications* 5(4):23
- Harris B (2018) Building a useful exponential random graph model. *Sage Publications*
- Higaki A, Uetani T, Ikeda S, Yamaguchi O (2020) Co-authorship network analysis in cardiovascular research utilizing machine learning (2009–2019). *Int J Med Inform* 143:104274
- Kaiser M (2008) Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. *New J Phys* 10(8):083042
- Lobo MP (2021) Uma revista aberta
- McAuley JJ, da Fontoura Costa L, Caetano TS (2007) Rich-club phenomenon across complex network hierarchies. *Appl Phys Lett* 91(8):084103
- Meyer B, Choppy C, Staunstrup J, van Leeuwen J (2009) Viewpoint research evaluation for computer science. *Commun ACM* 52(4):31–34
- Name2GAN. <https://quecst.qcri.org/tool/Name2GAN>
- Newman ME (2004) Who is the best connected scientist? A study of scientific coauthorship networks. In: *Complex networks*. Springer, pp 337–370
- Newman ME (2003) Mixing patterns in networks. *Phys Rev E* 67(2):026126
- Ortega JL, Aguillo IF (2013) Institutional and country collaboration in an online service of scientific profiles: Google Scholar citations. *J Informetr* 7(2):394–403
- Ortega JL, Aguillo IF (2014) Microsoft academic search and Google Scholar citations: comparative analysis of author profiles. *J Am Soc Inf Sci* 65(6):1149–1156
- Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical report, Stanford InfoLab
- Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p^*) models for social networks. *Soc Netw* 29(2):173–191
- Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118–1123
- Sarigöl E, Pfitzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scientific success based on coauthorship networks. *EPJ Data Sci* 3:1–16
- Scimago Journal & Country Rank. <https://www.scimagojr.com>
- Sculley D (2010) Web-scale k-means clustering. In: *Proceedings of the 19th International Conference on World Wide Web*, pp 1177–1178
- Serrano MA, Boguna M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci* 106(16):6483–6488. <https://doi.org/10.1073/pnas.0808904106>
- Tang J, Chen Y, She G, Xu Y, Sha K, Wang X, Wang Y, Zhang Z, Hui P (2021) Identifying mis-configured author profiles on google scholar using deep learning. *Appl Sci* 11(15):6912
- Yan E, Ding Y (2009) Applying centrality measures to impact analysis: a coauthorship network analysis. *J Am Soc Inf Sci Technol* 60(10):2107–2118
- Yang S (2013) *Networks: an introduction* by MEJ Newman. Oxford University Press, Oxford, 720 pp, \$85.00. Taylor & Francis
- Yong A et al (2014) Critique of Hirsch's citation index: a combinatorial fermi problem. *Not AMS* 61(9):1040–1050

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.