International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 11, Issue – 01, January 2026, PP – 26-32

Mining Academic and Research Collaboration Networks: The Mendeley, Scopus, and Google Scholar Case Study

JRG Pulido, José Román Herrera-Morales and Armando Román Gallardo

Faculty of Telematics – University of Colima, Colima, Mexico

Abstract: The exponential growth of indexed scientific publications has made the automatic identification of complete and accurate metadata increasingly challenging. In particular, inconsistencies in author name representation—such as the use of abbreviations in some sources and full names in others—often lead to ambiguity and misattribution. Consequently, precise identification of authors -their academic profiles, scientific output such as institutional affiliations, and collaboration relationships- is essential.By addressing these identification issues, it becomes possible to generate relevant insights through automated data mining processes. This paper presents a methodological framework that leverages information management services to identify authors and collaboration networks. Data mining techniques were applied to Mendeley, Scopus, and Scholar platforms, thenby extracting and consolidating bibliographic data from these repositories, a structured set of metadata was obtained.The results demonstrate that collaboration networks within scientific publications can be effectively identified through author-based data mining approaches. These networks can be used to generate precise quantitative indicators, which may serve as valuable metrics for evaluating academic influence, collaboration patterns, and the prestige of research communities.

Introduction

Through knowledge management, researchers aim to add value to the information they generate. Scientific publications constitute the fundamental core for evaluating research activity and, according to García-Peñalvo (2018), provide quality indicators.As researchers increasingly adopt web-based platforms for scientific communication, information sources regarding research impact expand globally with greater coverage and transparency than previously available. Information is not a static object but rather a dynamic entity produced and shared through technology. Therefore, long ago, most researchers migrated their activities to the web, employing social networks to enhance their academic influence compared to traditional publication environments. Even more, online bibliographic reference managers have been adoptedto save time, facilitate bibliographic resource management, and prevent errors in the manual composition of scientific works.

Related work

Professional information services index numerous scientific publications without automatically identifying all work-related information, such as author names appearing with abbreviations in some publications and full names in others, causing reference errors (Moncada-Hernández, 2014). A potential solution involves creating academic profiles through the association of research groups using academic collaboration networks. The importance and difficulty of maintaining a quality academic profile offers numerous benefits, including enhanced presence and visibility while avoiding ambiguities in indexing and scientific information retrieval services (García-Peñalvo, 2018). In a study examining scientific publication authors, Torres-Salinas and Milanés-Guisado (2014) found that 77% maintained public profiles on Google Scholar Profiles, 70% on LinkedIn and Mendeley, 55% on Twitter, and 47% on Slideshare, demonstrating the complexity of author-level identification. Similarly, Fernández-Marcial and González-Solar (2015) conducted research on digital identity status among the research community at the University of Coruña, examining their presence across Orcid, ResearcherID, Scopus Author, Google Scholar Citations, ResearchGate, and Mendeley, concluding that greater efforts in authorship identification are required. A more recent effort using Scopus, Mendeley, and Scholar was carried out to determine the readership impact on these platforms (Naudé and Kroeze, 2025).

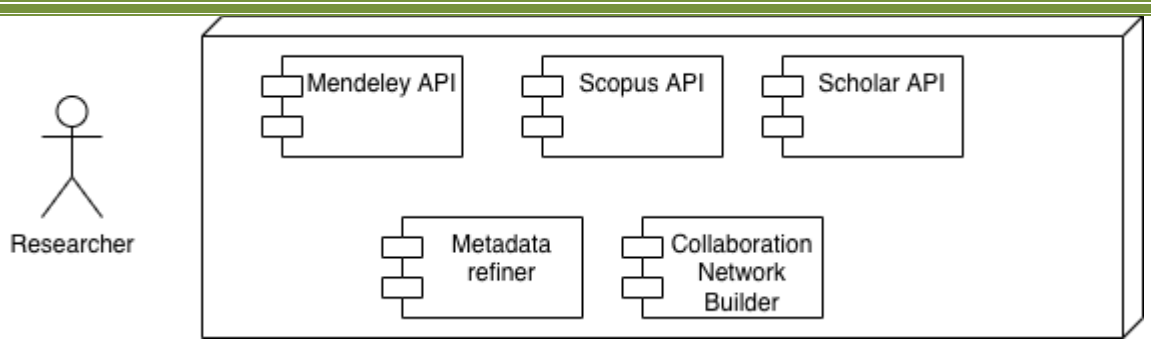

Figure 1: The overview of our software architecture

Recent bibliometric studies confirm the crucial role of co-authorship networks in assessing research dynamics, often extracting data from major platforms like Scopus and Web of Science (Popescu et al., 2025, Li, X., et al. 2024, Börner et al. 2024). This is why social network-based indicators (altmetrics), when compared to journal-based ones, show a stronger correlation with citations and greater filtering power to identify highly cited publications (Waltman and Costas, 2014). Consequently, this work proposed a method for identifying collaboration networks using data mining applied to web repositories of academic publications.

Methodology

For this study, the Mendeley academic platform was initially selected due to its social network integration features and web-based infrastructure for organizing research citations and storing articles in PDF format. This platform integrates research article management with collaborative features connecting researchers locally and globally, facilitating rapid and dynamic establishment of researcher expertise before article submission for evaluation. Some other altmetrics features were also considered by confronting Mendeley against similar platforms, namely Google Scholar and Elsevier Scopus.

TABLE 1: Authors analyzed in each case and academic profile identification on Mendeley, Google Scholar, and Scopus platforms

# case	Author	Mendeley	Google Scholar	Scopus
1	Celestini Frank	yes	Yes	yes
1	Cuauhtémoc Calderón	yes	Yes	no
1	Fernanda Julia Gaspari	yes	Yes	yes
2	Alberto Jiménez-Valverde	yes	Yes	yes
2	David Posada	yes	yes	yes
2	Francisco Guinea	yes	yes	yes
2	Hermenegildo Garcia	yes	yes	yes
2	Ivan Mora-Sero	yes	yes	no
2	Pedro Jordano	yes	yes	yes
2	PilarGayan	yes	yes	yes
2	Roberto Solano	yes	no	yes
2	Xavier Querol	yes	yes	yes
3	Julio Huerta-Espino	yes	yes	yes
3	Oscar Castillo	yes	yes	yes
3	Patricia Melin	yes	yes	yes

The process carried out to identify author collaboration networks from information management services was performed using the following steps:

1. **Mendeley** Starting with a known list of authors, without modifying the their data, and regardless of the order of surname or authors, three cases were used in the experiments of our mining process.The refinement of the authors metadata was carried out by disambiguating the first name and surnames, and

abbreviations of universities for instance. This search was conducted on the Mendeley repository using its respective API (Application Programming Interface).

2. **Scopus** Specific metadata of the registered publications of each author was obtained from Mendeley. A comparative semi-automated analysis wasthen performed on the author metadata -surnames and first names-, including the acquisition of the Scopus ID Elsevier Scopus, even to retrieve extra metadata that may be useful in the creation of collaboration networks.
3. **Scholar**Accessing an author academic profilerequires to knowin advance the identifier with which the author is registered in the Google Scholar database. This identifier was obtained by injecting a search into the "citations" section, specifying two necessary parameterswith the name data of the author being searched.

A brief description of the obtained metadata is as follows:

- **AFFILIATION** the institution, department, and even the country of affiliation of the author in some cases are gathered.
- **TOPICS** the research interest of each of the authors considered are also gathered in most of the cases.
- **METRICS**some global metrics of the authors, such as total citations throughout their career, as well as their h-index and i10-index, from the different platforms, if present are as well collected.
- **ARTICLES**a list of the author's publications indexed by Google Scholar, including the title of the publication, their respective co-authors, the year of publication, the number of citations, the publisher and journal, and even the link to the full PDF text when available is also extracted.
- **CO-AUTHORS** If available, the list of people that the author has carried out collaboration with is also obtained. This list provides more precise identification and affiliation data for their collaborators.

Upon completion of the mining process, the obtained data is stored in a local repository for further an analysis. This approach -combining multiple platforms- is recognized as a robust methodology in bibliometric studies (Lazzari & Castelli, 2024; Bridgeman, 2023, Naudé and Kroeze, 2025). Different programming languages have been used for these tasks, but mainly Python and its scrappinglibraries were at hand.

TABLE 2: Academic productivity analysis for authors in each case

# case	Author	No. articles
1	Celestini Frank	20
1	Cuauhtémoc Calderón	20
1	Fernanda Julia Gaspari	18
2	Alberto Jiménez-Valverde	18
2	David Posada	20
2	Francisco Guinea	20
2	Hermenegildo Garcia	18
2	Ivan Mora-Sero	19
2	Pedro Jordano	19
2	PilarGayan	19
2	Xavier Querol	21
3	Julio Huerta-Espino	20
3	Oscar Castillo	20
3	Patricia Melin	20
	TOTAL	272

Results

We are at a very early stage of our research, nonetheless the preliminary results are promising. We have been able to produce lists of authors, co-authors, institutions, and author areas of interest, among other data. Similarly, we have obtained results from cross-referencing this data between platforms – Mendeley, Scopus, Scholar. The case studies we have analyzed are as follows:

- **INEGI** authors of different nationalities who commonly utilized references to publications from Mexico's National Institute of Statistics and Geography (INEGI) to support their work. All authors were originally obtained through the Mendeley API. Three were selected as samples: Fernanda Julia Gaspari, Cuauhtémoc Calderón, and Frank Celestini, as they maintained Google Scholar profiles and their co-authors could be detected.

- **SPAIN** considering Mendeley profiles of nine influential Spanish scientists that also areincluded in the Highly Cited Researchers report (Clarivate). This lists+3,000 most highly cited scientists worldwide.
- **MEXICO** for this case, and again from mendeley profiles, a few authors with Mexican affiliations appearing also in the Highly Cited Researchers report were considered.

As mentioned,case 1 includes 3 authors, case 2 includes 9 Spanish authors, and case 3 includes 3 Mexican authors.As can be seen in Table 1,the total number of articles retrieved accounts for 272. It also displays whether academic profiles on Mendeley, Google Scholar, and Scopus for each authorwere identified through automated processes.

The Google Scholar academic profile identification is very important for the subsequent mining process stages, as it enables the intelligent extraction process for author's publications and collaborator lists. Table 2 presents a summary of articles retrieved per author and their respective co-author counts. From this table can also be seen that some authors lack Google Scholar profiles. A full list of articles of one of the authors is presented in Table 5. Bear in mind that, publication metadata scrapping faces some platform-imposed restrictions, for instance limiting retrieval results to a maximum of 20 publications.

After performing the automated Google Scholar profile processing to analyze author collaborators, validated collaborator lists are produced. It can be seen from Table 3that amongst the 15 authors with Google Scholar profiles, only 9 authors indicated with whom they maintain direct collaboration. Again, one important scrapping issue here is the limitation of 20 co-authors as maximum,for close collaborators marked by the researchers themselves.

After the corresponding scrapping process, an interactionnetwork has been identified as is shown in Table 4. Two authors from case 3, corresponding to Mexican researchers, were sampled, with collaboration networks identified across internal scope -work networks within the same institution- networks with other institutions with in the same country, and networks with collaborators from institutions in other countries.

Discusión

This ongoing research demonstrates that online resources from this kind of platforms can be used to extract even more specific metadata of interest from the author profiles. Some mechanisms to avoid author ambiguities have been created, namely ORCID, DOI, but they have not yet been automated at all.By following the list of profiles and publications of the researchers, a list of possible collaborators

TABLE 3: Analysis of collaborators registered in Google Scholar profiles

#case	Author	no_collaborators
1	Celestini Frank	20
1	Cuauhtémoc Calderón	13
2	Alberto Jiménez-Valverde	20
2	David Posada	7
2	Francisco Guinea	20
2	Ivan Mora-Sero	20
2	Pedro Jordano	20
3	Oscar Castillo	20
3	Patricia Melin	8

TABLE 4: Collaboration network identified – Affiliation and Type - case 3 Mexico

Author	Collaborators' affiliation	Type of collaboration network	Country	No. collaborators
Oscar Castillo	Tijuana Institute of Technology	national internal	Mexico	11
	Universidad Autónoma of Baja California	national inter-institutional	Mexico	4
	Instituto Politécnico Nacional, CITEDI	national inter-institutional	Mexico	1
	Systems Research Institute, Polish Academy of Sciences	international	Poland	1
	Unknown	na	na	3
Patricia	Tijuana Institute of Technology	national internal	Mexico	3

Melin				
	Madero Institute of Technology	national inter-institutional	Mexico	1
	Instituto Politécnico Nacional, CITEDI	national inter-institutional	Mexico	2
	Systems Research Institute, Polish Academy of Sciences	international	Poland	1
	Unknown	Na	na	1

Can be obtained. Knowing this information is highly important when developing future projects, as collaboration networks can provide recommendations for researchers who have similar research interest to one's own network.

Collaboration networks allow us to clearly identify clusters of people, pinpoint common interests, and as mentioned reveal new possibilities for collaboration. Automate extraction of collaboration patterns – including even topological and semantic features - from full-text articles is still an important avenue of research. Here we are not interested in predicting emerging collaborations, but to provide proof-of-concepr software tools for researchers pursuing to create a solid collaboration network.

In this work, we have shown that by using diverse heterogeneous sources of information, it is feasible to create academic and research interaction networks, ranging from a single classroom to national collaboration networks, and extending to multidisciplinary international networks.

Conclusions

TABLE 5: Full list of articles of one of the authors as extracted

Ivan Mora-Sero	
1	Characteristics of high efficiency dye-sensitized solar cells
2	Characterization of nanostructured hybrid and organic solar cells by impedance spectroscopy
3	Low-Temperature Processed Electron Collection Layers of Graphene/TiO2 Nanocomposites in Thin Film Perovskite Solar Cells
4	Recombination in quantum dot sensitized solar cells
5	Mechanism of carrier accumulation in perovskite thin-absorber solar cells
6	General working principles of CH3NH3PbX3 perovskite solar cells
7	Modeling high-efficiency quantum dot sensitized solar cells
8	Slow dynamic processes in lead halide perovskite solar cells. Characteristic times and hysteresis
9	High-efficiency “green” quantum dot solar cells
10	Role of the selective contacts in the performance of lead halide perovskite solar cells
11	Improving the performance of colloidal quantum-dot-sensitized solar cells
12	Photoinduced giant dielectric constant in lead halide perovskite solar cells
13	Titanium dioxide nanomaterials for photovoltaic applications
14	CdSe quantum dot-sensitized TiO2 electrodes: effect of quantum dot coverage and mode of attachment
15	Cyclic voltammetry studies of nanoporous semiconductors. Capacitive and reactive properties of nanocrystalline TiO2 electrodes in aqueous electrolyte
16	Core/shell colloidal quantum dot exciplex states for the development of highly efficient quantum-dot-sensitized solar cells
17	Simulation of steady-state characteristics of dye-sensitized solar cells and the interpretation of the diffusion length
18	Breakthroughs in the development of semiconductor-sensitized solar cells
19	Electron lifetime in dye-sensitized solar cells: theory and interpretation of measurements

The use of social features in platforms like Mendeley, Scopus, and Scholar continue to be essential for understanding scholarly collaboration.It is possible to identify collaboration networks by analyzing the authors in scientific publication platforms.

Our future work will aim to automate the processing of data cross-referencing to develop the interaction networks. This includes using visualization libraries to analyze and profit from the found networks and the vast, rich information they can provide, and to develop software that integrates the different components into a web application.

Using open scientific databases such as OpenAlex and Semantic Scholar APIs would provide access to richer metadata and citation networks at scale, while enabling cross-validation with extra data sources.

TABLE 6: Full list of collaborators of one of the authors as extracted

1	AméricaIvonne Zamora Torres
2	Ana CA Veloso
3	Anna Tykhonenko
4	Carlos Albero Calderon
5	Cristian Ramirez
6	Daniela M Correia
7	Eduardo Mendoza
8	Fernando Morales
9	Gabriela Punin
10	R Ponce
11	Teresa Dias
12	Thomas M Fullerton
13	Veronica Barros

We are to explore the incorporation of generative AI and large language models (LLMs) through frameworks like LangChain could automate the extraction and structuring of collaboration information from full-text articles, acknowledgment sections, and funding statements—expanding beyond traditional metadata-based approaches. Also our research may benefit from incorporating transformer models -SciBERT, Scholar BERT- to enable semantic analysis of research topics, automatic classification of collaboration types, and identification of interdisciplinary partnerships based on content similarity rather than solely on co-authorship patterns.

Finally, the integration of these components into unified web applications with real-time analytics capabilities would provide researchers and institutions with an accessible platform for monitoring academic collaboration dynamics and supporting evidence-based decision-making in research policy and team formation.

References

[1]. Afolabi, I. T., Ayo, A., & Odetunmibi, O. A. (2021).Academic Collaboration Recommendation for Computer Science Researchers Using Social Network Analysis.*Wireless Personal Communications*, 118(1), 487–501. https://doi.org/10.1007/s11277-021-08646-2

[2]. Alderson, D. (2016). How to critically appraise a research paper, *Paediatrics and Child Health*, 26(3), 110–113. https://doi.org/10.1016/j.paed.2015.09.007

[3]. Bridgeman, M. (2023). Building Your Digital Presence on Social Media. In: Dreker, M.R., Downey, K.J. (eds) Building Your Academic Research Digital Identity. Springer, Cham. https://doi.org/10.1007/978-3-031-50317-7_8

[4]. Fernández-Marcial, V., & González-Solar, L. (2015). Promoción de la investigación e identidad digital: el caso de la Universidade da Coruña. *El profesional de la información*, 24(5), 656-664. http://dx.doi.org/10.3145/epi.2015.sep.14

[5]. García-Peñalvo, F.-J. (2018). Digital Identity as Researchers. The Evidence and Transparency of Scientific Production. *EKS*, 19(2), 7–28. https://doi.org/10.14201/eks2017182717

[6]. Gunn, W. (2014). On numbers and freedom. *El profesional de la información*, 23(5), 463-466. https://doi.org/10.3145/epi.2014.sep.02

[7]. Henning, V., & Reichelt, J. (2008).Mendeley-A Last.fm for research?. En: *Fourth IEEE International Conference on eScience*, pp. 327–328. https://doi.org/10.1109/eScience.2008.128

[8]. Holt, Z., West, R., Tateishi, I., & Daniel, R. (2011). Mendeley: Creating Communities of Scholarly Inquiry through Research Collaboration. *Tech Trends*, 55(1), 32–36.

[9]. Jeng, W., He, D., & Jiang, J. (2015).User participation in an academic social networking service: A survey of open group users on Mendeley. *Journal of the Association for Information Science and Technology*, 66(5), 890–904. http://dx.doi.org/10.1002/asi.23225

[10]. Li, X., Wang, M., & Liu, X. (2024). Predicting collaborative relationship among scholars by integrating scholars' content-based and structure-based features. *Scientometrics*, 129(6), 3225–3244. https://doi.org/10.1007/s11192-024-05012-4

[11]. LópezCarreño, M. (2014). Análisiscomparativo de los gestoresbibliográficossocialesZotero, Docear y Mendeley: características y prestaciones. *Cuadernos de Gestión de Información*, 4, 51-66.

[12]. Mohammadi, E., Thelwall, M., & Kousha, K. (2015).Can Mendeley bookmarks reflect readership? A survey of user motivations.*Journal of the Association for Information Science and Technology*, 67(5), 1198-1209.http://dx.doi.org/10.1002/asi.23477

[13]. Moncada-Hernández, S.-G. (2014). Cómorealizarunabúsqueda de informacióneficiente.Foco en estudiantes, profesores e investigadores en el áreaeducativa. *Investigación en Educación Médica*, 3(10), 106-115.

[14]. Naudé and Kroeze (2025). A 45-year review of the South African Computer Journal (1979–2023). South African Computer Journal 37(2), 5-36. DOI: 10.18489/sacjv37i2/21924

[15]. Nederhof, A. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, 66(1), 81–100. https://doi.org/10.1007/s11192-006-0007-2

[16]. Popescu et al. (2025). European Union Machine Learning Research: A Network Analysis of Collaboration in Higher Education (2020–2024). *Electronics*, 14(7), 1248. https://doi.org/10.3390/electronics14071248

[17]. Salija, K., Hidayat, R., & Patak, A.-A. (2016). Mendeley Impact on Scientific Writing: Thematic Analysis. *International Journal of Advanced Science Engineering Information Technology*, 6(5), 657-662. http://dx.doi.org/10.18517/ijaseit.6.5.1140

[18]. Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

[19]. Torres-Salinas, D., & Milanés-Guisado, Y. (2014). Presencia en redessociales y altmétricas de los principalesautores de la revista El profesional de la información. *El profesional de la información*, 23(4), 367-372. https://doi.org/10.3145/epi.2014.jul.04

[20]. Van Noorden, R. (2014). Online collaboration: Scientists and the social network, *Nature*, 512(7513), 126–129. http://dx.doi.org/10.1038/512126a

[21]. Waltman, L., & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison with Citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445. http://dx.doi.org/10.1002/asi23040

[22]. Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, 68(10), 2511–2521. http://dx.doy.org/10.1002/asi.23883

[23]. Zaugg, H., West, R., Tateishi, I., & Daniel, R. (2011). Mendeley: Creating Communities of Scholarly Inquiry through Research Collaboration. *Tech Trends*, 55(1), 32–36.

