Mapping Scientific Collaboration Networks

Teresa Codoñer Esparza
Data Science student 3rd Year , Universitat Politècnica de València
Valencia, Spain
tcodesp@etsinf.upv.es

Laura Jiménez Navarro
Data Science student 3rd Year , Universitat Politècnica de València
Valencia, Spain
ljimnav@etsinf.upv.es

Eva Martín Iglesias
Data Science student 3rd Year , Universitat Politècnica de València
Valencia, Spain
emarigl@etsinf.upv.es

Pilar Mas Sarrión
Data Science student 3rd Year , Universitat Politècnica de València
Valencia, Spain
pmassar@etsinf.upv.es

ABSTRACT

Many researchers, especially early-career or underrepresented groups, face challenges in identifying relevant collaborators or understanding their position within the academic landscape. This project focuses on building and analyzing a scientific collaboration network to better understand how researchers connect and how these structures can inspire new approaches to education.

Using OpenAlex data, a graph was constructed where nodes represent authors and edges represent co-authorships. Basic graph theory techniques were applied to identify influential figures, collaboration groups, and emerging research areas. An interactive prototype was also developed to explore the network dynamically, allowing filtering by field, country, and gender. The study shows how collaboration networks can support personalized and evidence-based educational strategies.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Human-centered computing** → *Collaborative and social computing*.

KEYWORDS

Collaboration Networks, Graph Theory, Educational Innovation, Open Science, Network Visualization

ACM Reference Format:
Teresa Codoñer Esparza, Laura Jiménez Navarro, Eva Martín Iglesias, and Pilar Mas Sarrión. 2025. Mapping Scientific Collaboration Networks . In . ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Scientific research is increasingly characterized by collaboration across disciplines, institutions, and countries. Understanding how researchers connect provides valuable insights into both the organization of science and opportunities for improving education through collaborative approaches.

However, existing collaboration networks often reflect structural inequalities that limit the visibility of women and early-career researchers. Traditional metrics may overlook their contributions, reinforcing gender and seniority biases. Promoting inclusive collaboration patterns is essential to support emerging talent and ensure that diverse voices contribute to scientific progress.

Building on previous work modeling scientific networks through graph theory [2, 3], this project analyzes a real-world co-authorship network with a focus on equity and inclusion. Using data from OpenAlex [1], we constructed a graph to identify major collaboration hubs and offer a flexible platform for demographic-aware educational analysis.

2 DATA COLLECTION AND NETWORK CONSTRUCTION

The data used for this study were extracted from OpenAlex, focusing on peer-reviewed articles published between 2018 and 2023 in Social Sciences, Computer Science, and Multidisciplinary Science. After filtering, the sample included 55,875 authors and 5,551 articles.

The network construction process involved:

- Representing each author as a node.
- Creating an undirected edge between authors who had collaborated on at least one paper.
- Assigning weights to edges based on the number of co-authored papers.

Data cleaning steps included standardizing institution names and resolving author ambiguities using ORCID identifiers. Authors' countries and inferred genders were also recorded to allow further demographic analysis.

3 NETWORK ANALYSIS

Once the network was built, basic graph analysis techniques were applied:

3.1 Centrality Measures

Several centrality metrics were calculated to identify influential figures:

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- **Degree Centrality** highlighted authors with many collaborations.
- **Betweenness Centrality** identified authors who served as bridges connecting different parts of the network.
- **Eigenvector Centrality** revealed authors with influential neighbors.

3.2 Community Detection

Communities of researchers were detected by identifying groups of authors with dense internal connections. These communities typically reflected shared research interests or institutional collaborations.

One notable observation was the low representation of female researchers in computer science communities compared to other fields. This underrepresentation limits diversity of thought and innovation in an area that increasingly shapes society. Encouraging female participation in computing-related research is crucial not only for equity, but also for producing more inclusive and socially aware technologies. By visualizing collaboration patterns, tools like the one developed in this project can help surface hidden contributions and promote greater visibility for women in tech.

3.3 Connected Components

Unlike typical large-scale collaboration networks, no single giant component dominated the structure. Instead, the network was composed of many smaller components. Within these, certain authors acted as important connectors, linking otherwise separate groups.

Studying these bridging figures provides opportunities to better understand the characteristics of researchers who facilitate broader collaboration and interdisciplinary exchange.

4 FINDINGS

The analysis revealed a decentralized structure with numerous independent clusters. Although there were no very large components, some authors demonstrated a high capacity to connect different groups.

These connectors tended to have high betweenness centrality scores and were often engaged in interdisciplinary work. Their role suggests that monitoring and supporting such researchers could be important for fostering collaboration across academic fields and promoting innovative educational programs.

The project highlights the potential of collaboration networks not just to map existing relationships, but also to identify individuals who drive connectivity and innovation.

5 PROTOTYPE APPLICATION AND FUTURE WORK

An interactive prototype was developed to allow dynamic exploration of the network.

Features include:

- Filtering by field, country, or gender.
- Highlighting researchers based on centrality measures.
- Exploring connections between communities.

One key future objective is to track the evolution of these bridging authors over time. Analyzing how their roles develop could provide valuable insights into how scientific fields grow and how interdisciplinary collaboration can be encouraged.

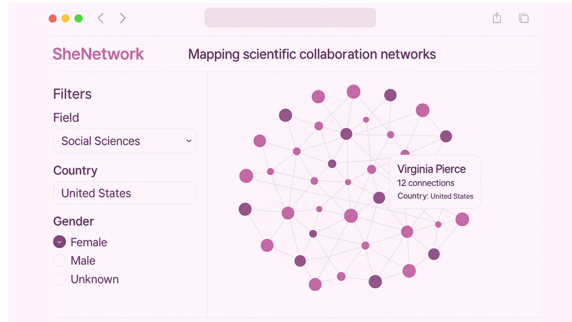


Figure 1: Prototype for visualizing collaboration networks

Planned improvements include:

- Scaling the network to millions of nodes.
- Integrating live updates from OpenAlex.
- Applying predictive models to forecast emerging research trends and collaboration opportunities.

6 CONCLUSION

Building and analyzing a scientific collaboration network offers a practical approach to understanding how research communities form and evolve. By focusing on the role of influential connectors, it becomes possible not only to visualize collaboration patterns but also to support educational initiatives aimed at enhancing interdisciplinary work.

This solution can be particularly impactful for young women researchers, who often face greater challenges in gaining visibility, building academic networks, and accessing leadership roles in STEM. By highlighting overlooked collaboration patterns and surfacing influential female figures, the platform fosters a sense of belonging and recognition. Moreover, it helps early-career women identify potential mentors and collaboration opportunities, ultimately contributing to more diverse and inclusive research ecosystems.

The project shows that even relatively simple graph techniques, applied to open data, can provide powerful insights into the scientific ecosystem. Future work will focus on scaling the platform and using these insights to inform smarter educational policies and programs.

REFERENCES

[1] Jason Priem and Heather Piwowar. 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *Quantitative Science Studies* 3, 3 (2022), 1–22.

[2] M. E. J. Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* 98, 2 (2001), 404–409.

[3] Albert-László Barabási, Hawoong Jeong, Zoltán Neda, Erzsébet Ravasz, András Schubert, and Tamás Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications* 311, 3–4 (2002), 590–614.

Mapping Scientific Collaboration Networks

Teresa Codoñer Esparza, Laura Jiménez Navarro, Eva Martín Iglesias, Pilar Mas Sarrión

acm's women in computing acm-w Europe

Introduction

Scientific research increasingly depends on collaboration between researchers, institutions, and countries.

By building a scientific collaboration network, it is possible to better understand how knowledge spreads and how these patterns can inspire innovative educational strategies.

This project uses data from OpenAlex to build a real-world co-authorship network.

Data and Methods

Data was collected from OpenAlex, focusing on articles published between 2018 and 2023.

Fields selected: Social Sciences, Computer Science, Multidisciplinary Science.

A network was built where each node represents an author and each edge represents a co-authorship.

Attributes such as country and gender were recorded to allow further exploration.

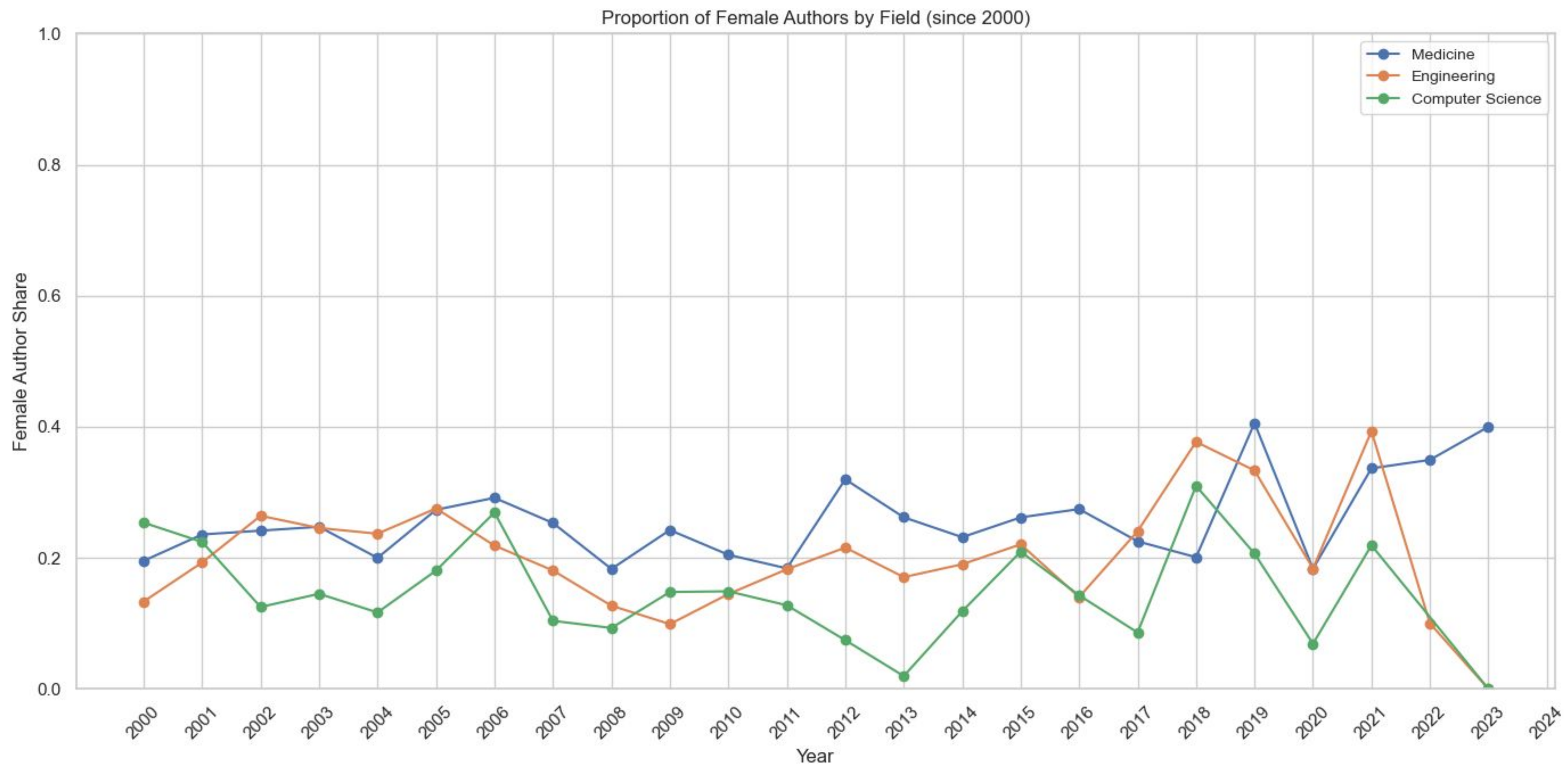
Network Analysis

Centrality measures (degree, betweenness, eigenvector) were calculated to detect key researchers.

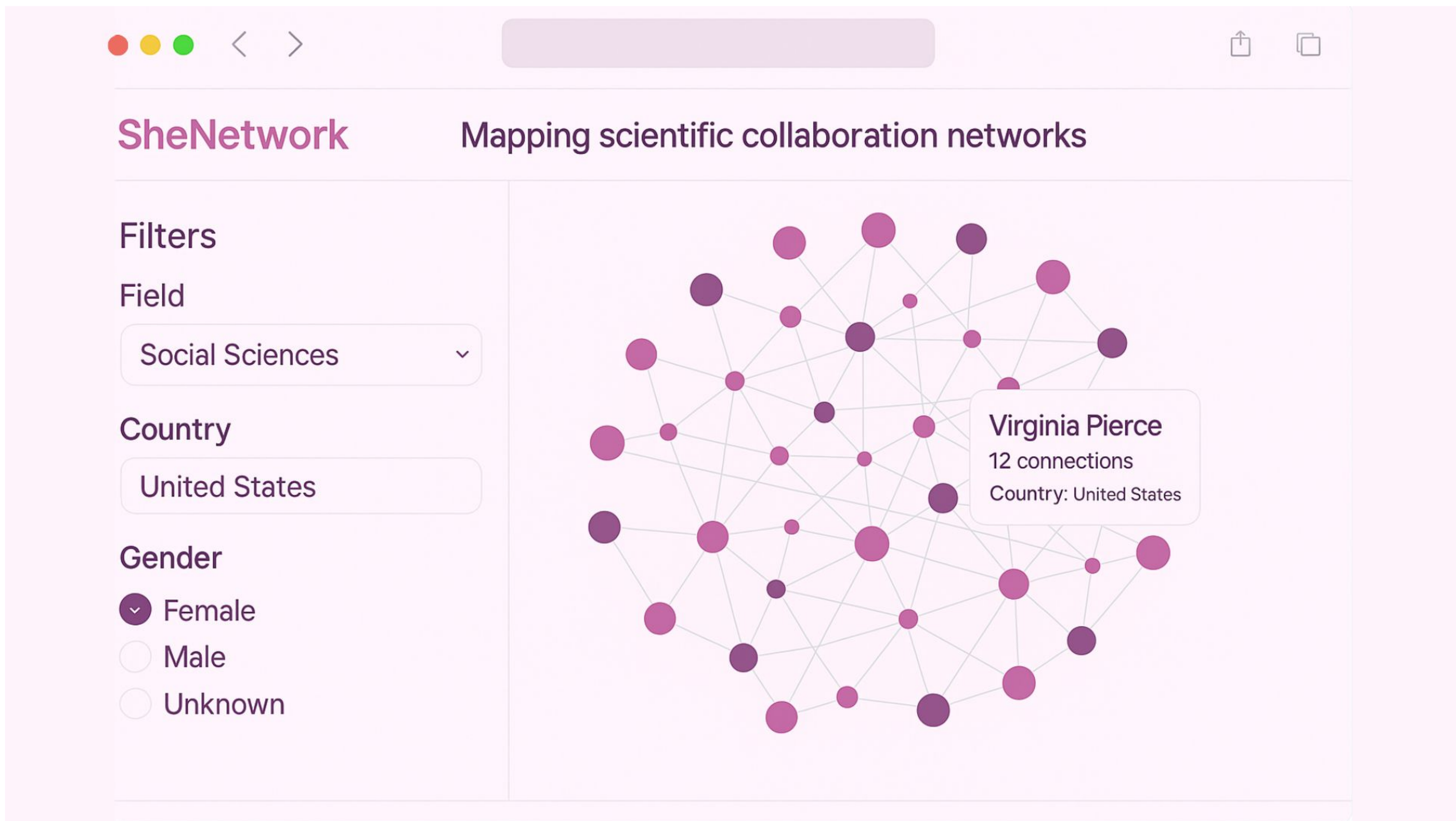
Groups of closely collaborating authors were identified through community detection.

The network was composed of multiple small components, with no giant component.

Certain influential researchers acted as connectors between different groups.



Proportion of female authors by field



Web prototype

Prototype-Future Work

A web-based prototype was created to explore the collaboration network dynamically.

Users can filter by field of study, country, and gender.

Future work will include expanding the dataset, refreshing the network in real time, and analyzing the evolution of key researchers over time.

The long-term goal is to support the design of smarter, data-driven educational strategies.

References

1. Jason Priem and Heather Piwowar. OpenAlex: A fully-open index of scholarly works, Quantitative Science Studies, 2022.
2. M. E. J. Newman. The structure of scientific collaboration networks, PNAS, 2001.
3. Albert-László Barabási et al. Evolution of scientific collaborations, Physica A, 2002.