Co-Authorship Networks: Investigating Collaborative Patterns in Academic Fields

Class Group Number: Group 2

Contributors:

Daniella Smith: *Investigating Centrality Distributions Between Academic Fields*
Brooke Feinberg: Exploring Connectivity Dynamics in Feminised Academic Fields
Saachi Sabnis: Understanding Structural Differences Among Academic Networks

Word Count: 2495

Introduction

Empirical research has demonstrated how occupations undergoing feminisation (experiencing an inflow of female workers) are devalued through a decline in wages and status. The devaluation hypothesis proposes a causal link between feminisation and devaluation (Busch, 2017). Academia has undergone significant structural changes over the last 50 years, with female student enrollment increasing by almost 500% worldwide (Fiske, 2012), alongside the devaluation of non-STEM subjects, particularly social sciences and humanities (Heller,2023). Scholarly devaluation increases intentions to leave academia (Settles et al., 2021) and has been posed, alongside institutional prestige, as an explanation for disparities in gender representation among tenured faculty (Spoon et al., 2023). Moreover, network effects, such as degree centrality, publishing tenure and assortativity have been identified as predictors of authors' research impact (E. Y. Li et al., 2013). These findings lead us to question whether a field's feminisation and devaluation has implications for network characteristics, and thus collaboration.

To explore this phenomenon, we used Chen et.al (2017)'s network, consisting of 402.39 thousand Google Scholar profiles. Google Scholar is a search engine that collects scholarly publications from databases across many disciplines. Co-authorship of articles documents a collaboration between two or more co-authors, with partnerships forming a co-authorship network. In this dataset, nodes refer to an author's profile, and edges denote a co-authored paper, with authors being coded as belonging to fields of computer science, biology or sociology. As such, the structure of the network provides insight into collaboration patterns within an academic community, enabling us to explore the relationship between feminisation and collaboration structure.

Computer Science is a non-feminised field: Wang et al (2010)'s predictive analysis projected that the proportion of female authors in Computer Science will not reach 50% within the next century.

In contrast, over 52% of biology doctorates in the US are awarded to women, with Feldon et al (2017) theorising it may be a unique case of a STEM field undergoing feminisation. However, these statements are brought into question by the very high proportion of male biology faculty, and ongoing gender discrimination faced by female biology students and researchers (Cheryan et al., 2017).

The social sciences have undergone significant feminisation : currently, women hold the majority of instructor and lecturer positions in psychology, anthropology, and sociology in the US and represent the majority of sociology and psychology PhDs. However, regarding the proportion of research publications and tenured faculty positions, women are underrepresented, but to a much lesser extent compared to other fields (Casad et al., 2022). For example, Bornmann et al, 2015 found, that while women author 14% of highly cited papers, they account for 31% of highly cited authors in the social sciences, compared to 3.7% in Engineering.

In this view, we aim to compare the collaboration structure across these three academic fields, viewing computer science as non-feminised, biology as undergoing feminisation and sociology as highly feminised, with the view of examining the impact of field feminisation on collaboration.

Investigating Centrality Distributions Between Academic Fields :
Student 22417
Word Count : 498

Research Question: Do mechanisms of collaboration, operationalised as degree distributions, differ between academic fields?

To examine differences in collaboration mechanisms, mean degree and eigenvector centrality were compared across three fields. Further, we explored whether cumulative distributions of node degree followed a power law by plotting them on a log-log scale.

Degree centrality is the count of edges on a node; in this network, it represents the number of co-authors an individual declares on their Google Scholar profile. Highly connected nodes are likely to have greater influence and access to information than those with few connections; a network with a higher mean degree is thus characterised by frequent partnerships and resource sharing (Newman, 2010).

Table 1.1 Degree Centrality

	Mean Degree	Median Degree	Min.Degree	Max .Degree
Whole Network	6.133	2.000	0	463
Computer Science	1.269	0	0	89
Biology	1.244	0	0	84
Sociology	0.021	0	0	2

The network of Computer Science authors has the highest average centrality (mean degree = 1.27), closely followed by the Biology network (mean degree 1.24). In contrast, the sociology network has a low mean degree (0.02).

Eigenvector centrality measures how central a node is proportional to its neighbours ; a high average eigenvector centrality indicates the prevalence of individuals holding great influence among their connections (ibid).

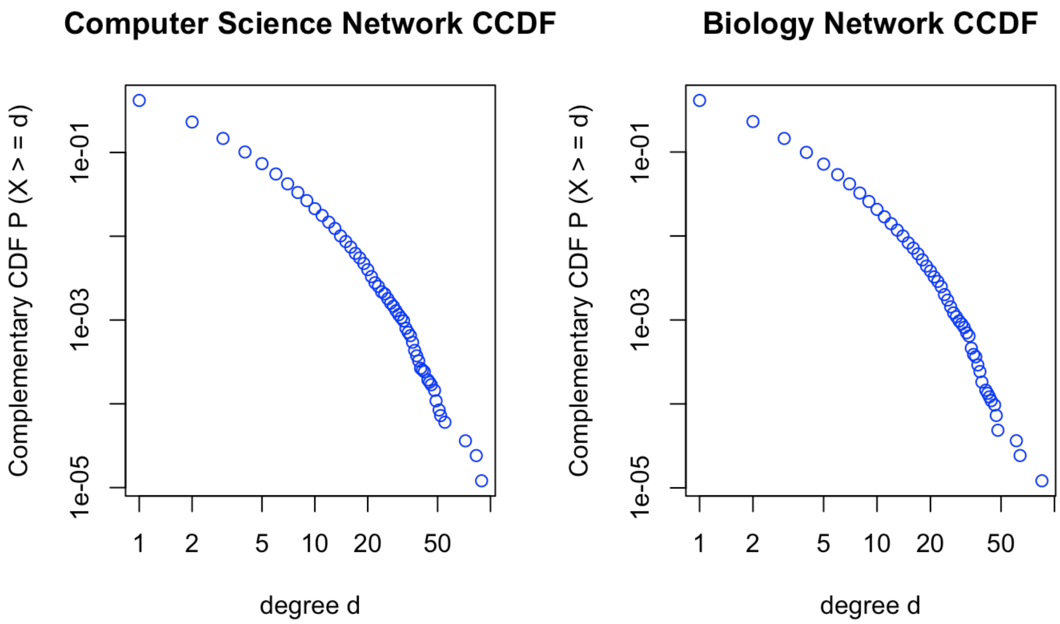
Table 1.2 : Eigenvector Centrality

	Mean E.Cent	Median E.Cent	Min E.Cent	Max E. Cent

Whole Network	0.000401	0	0	1
Computer Science	0.000508	0	0	1
Biology	0.000398	0	0	1
Sociology	0.00151	0	0	1

Sociology authors have the highest mean eigenvector centrality (0.00151), almost three times as large as the computer science network (0.000508) and almost four times as large as the biology network (0.000398).

To further investigate differences between fields, we plotted cumulative degree distributions on log-log scales for Computer Science and Biology authors. There was insufficient degree variation among sociology authors (Table 1), thus we were unable to perform this analysis for that network.



Both cumulative degree distributions have heavy tails, indicating most nodes have a low degree, but a small number are highly connected. We infer both networks have hubs - a subset of influential authors collaborating with an above-average number of individuals.

Table 3 : Fitting To Power Law

	Continuous	Alpha	Xmin	LogLik	KS.stat	KS.p
Computer Science	FALSE	4.24	18	-1360	0.0359	0.598
Biology	FALSE	3.20	8	-7190	0.0433	0.000

The cumulative degree distribution of the computer science network conforms well to a power law (KS.p > 0.05), unlike the biology network. A proposed explanation for power laws in social networks is the tendency of individuals to copy their predecessors, with actions being repeated due to their 'popularity' (Easley & Kleinberg, 2010). Prestige may be a facilitating mechanism of co-authorship in the computer science network, whereby authors tend to co-author papers with individuals who have already frequently collaborated.

We infer computer science, with the highest mean degree, low mean eigenvector centrality, and adherence to a power law, is characterised by recurrent collaboration, with reputation, but not influence, promoting co-authorship. The biology network's cumulative degree distribution does not conform to a power law, has a smaller mean degree and low eigenvector centrality, indicating lower levels of collaboration and providing no evidence of preferential attachment nor influence as network mechanisms. The low mean degree and degree range and high eigenvector centrality of the sociology network indicate collaboration is comparatively rare, but those who do collaborate, are highly influential.

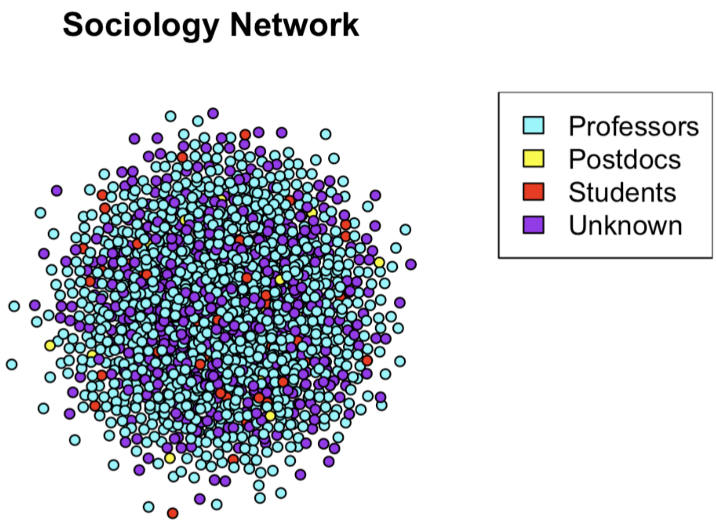
Taken together, there is greater collaboration in Biology and Computer Science, however, we have less evidence of prestige and popularity as mechanisms in Biology and Sociology, despite hubs being influential in these networks.

Exploring Connectivity Dynamics in Feminised Academic Fields: Student 34809

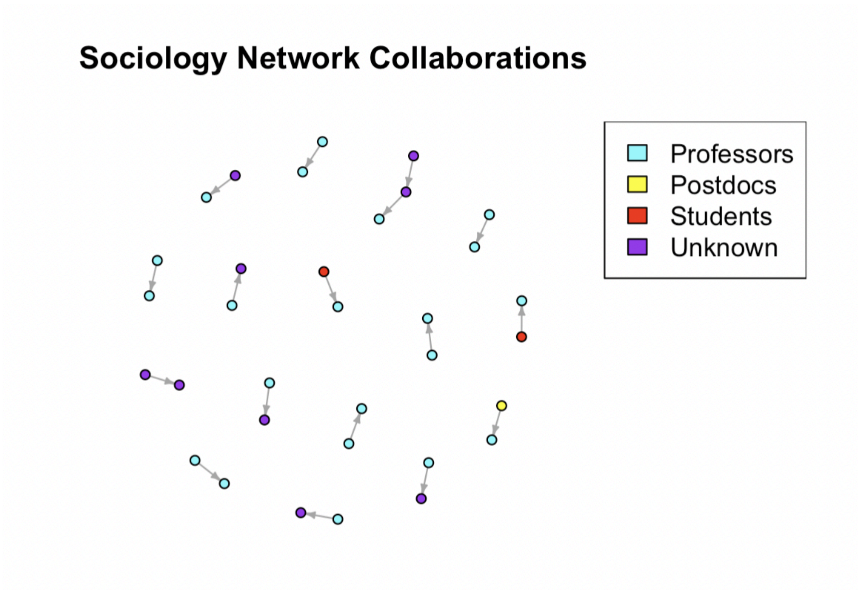
RQ: Do distinct collaborative patterns, in terms of frequency and structure, emerge in more feminised academic fields and how do these patterns affect the spread and flow of information?

To further investigate feminised academic fields, we chose to partition sociology and examine its collaborative patterns through an exponential random graph model.

Graph 2.1: Partitioned Graph of the Sociology Network



Graph 2.2: Partitioned Graph of the Sociology Network Collaborations



The sociology network consists of 1,602 nodes yet only 17 edges, which is disproportionately lower than both computer science and biology networks (80,000+). As anticipated, this disparity in network size versus connections gives rise to an extremely low density measure of 6.628192e-06, indicating the network's connectivity is inherently sparse. The existing connections are dominated by independent dyadic relationships, which restrict the flow of information and collaboration amongst sociology.

While the sociology's density provides insight into its underlying structure, its size discrepancy makes a comparative analysis between the three network densities impractical. Instead, we construct an exponential random graph model to identify the key structural patterns contributing to the sparse connectivity in the sociology network. An ERGM will assess if certain attributes such as, academic title or g-index, influence the formation of connections.

Model 2.3: Base Model

```
Call:
ergm(formula = socio_network ~ edges)

Maximum Likelihood Results:

      Estimate Std. Error MCMC % z value Pr(>|z|)
edges -11.9242     0.2425      0  -49.16   <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 3555570.5  on 2564802  degrees of freedom
 Residual Deviance:     439.4  on 2564801  degrees of freedom

AIC: 441.4  BIC: 454.2  (Smaller is better. MC Std. Err. = 0)
```

First, we fit a base model that measures the probability of edges forming between any two nodes in the network. The negative coefficient has a statistically significant p-value, indicating that edge formation in this network is very unlikely. We proceed to fit various attributes and structural dependence terms to identify the best ERGM model.

In model two we conditioned `nodefactor` on "title," however, the AIC and BIC values both increased, failing to improve the original model. In model three, we conditioned `nodemix` on "title" and observed positive coefficients in various combinations of academic titles, suggesting particular title combinations positively influence edge formation. Although, the overall AIC and BIC values also increased, failing to improve the baseline ERGM. In model 4, we tested structural dependence terms like `mutual`, `triangle`, and `gwesp` to more closely mirror a real-world social network. Although, these provided infinite coefficient estimates, reflecting the networks sparsity and lack of reciprocated edges.

Model seven conditioned `nodefactor` on "title" and "g_index," resulting in an improved AIC and worsened BIC. The improved AIC suggests that g_index –defined as the distribution of citations an individual has–impacts scholars' propensity to connect, as academics with higher g_index exhibit higher positive coefficients. However, given model seven's increased BIC, we conclude that all models failed to improve the baseline, consistently reflecting that edge formation is exceedingly rare. This is because ERGMs simulate model networks based on existing edge connections, making it challenging to fit a meaningful ERGM, given the scarcity of data available.

Given our inability to accurately model an ERGM beyond the baseline, we assert that the sparse connectivity in sociology inhibits our capacity to effectively examine the key attributes influencing edge formation and network structure. Moreover, the disproportionate ratio of edges to nodes implies that as academic disciplines, such as sociology, experience this feminization, there is a notable decline in collaboration among scholars. Thus, the observed division in connectivity, coupled with the underlying ideology of academic feminization, likely contributes to gender-based productivity gaps, as limited access to collaboration and information flows hinders academic advancement.

Understanding Structural Differences among academic networks: Student 25864

Research Q 3. How do the structural properties of co-authorship networks vary across different academic disciplines, and what does this imply about the nature of scholarly collaboration within these fields?"

To compare clustering tendencies, assortativity and transitivity (local and global) were calculated across all three academic fields, with further analysis employing edge-betweenness as a community detection algorithm to identify existing subsets of nodes within the network.

Transitivity reflects the propensity for nodes in a network to form tightly knit clusters. In this context, it describes the likelihood that, within a co-authorship network, two academics with a mutual co-author are themselves likely to have collaborated on a paper. Indirect relations in scientific creative processes are key in determining where new collaborations are formed (Inoue, 2020).

Table 3.1. Transitivity and Assortativity Values for the three networks

	Global Transitivity	**Assortativity**
Computer Science Network	0.22	0.000514
Biology Network	0.20	-0.001223
Sociology Network	0	-0.187302

The computer science network exhibited moderate global transitivity (0.22), indicating a reasonable degree of closed triads where collaborators also collaborate with each other. This was also seen in the biology network at a slightly lower intensity (0.20). Comparatively, the sociology network had a transitivity of 0 suggesting a more linear or hierarchical collaboration pattern with no mutual co-authorships forming among trio groups. Additionally, local transitivity was consistent across academic ranks, with values ranging from 0.27-0.28 in both fields, implying that academic status does not severely impact an author's ability to engage in or form clustered collaborations.

Table 3.2. Local Transitivity by academic titles for computer science and biology network

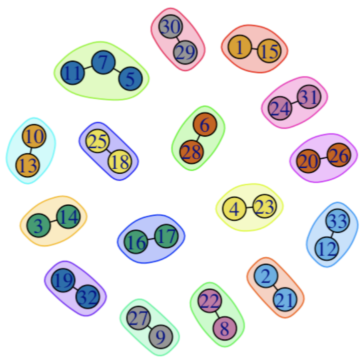
Fields	**Group**	**Average Clustering Coefficient**
Computer Science	Professors Postdocs Students	0.28 0.28 0.27
Biology	Professors Postdocs Students	0.27 0.28 0.28

Assortativity is the tendency for nodes to connect to other nodes with similar properties within a network. Here, assortativity by seniority was looked at i.e. the likelihood of authors to co-author a paper with authors with different academic titles.

The computer science network displayed a nearly neutral assortativity coefficient (0.000514), indicating no strong preference for collaboration among authors of the same academic title than authors of different titles. In the biology network, the slightly negative assortativity (-0.001223) suggests a minimal tendency to avoid collaborations with peers of the same academic level, although the effect is very slight. Comparatively, the sociology network showed significant negative assortativity (-0.187), revealing a clear pattern wherein authors in the sociology network tend to collaborate across different academic ranks rather than within the same rank.

To gain a further understanding of how these collaborations are organised and subgrouped, community detection using edge-betweenness was employed. Despite removing isolates, the biology and computer science networks remained too large to provide clear visualisations and thus, had to be left out.

Graph 3.3. Community Detection for the Sociology Network



As observed in graph X, most communities in the network are visually distinct and isolated with no visible links connecting them, which may imply a low level of inter-community collaboration. This could reflect that research in the field tends to operate within more isolated groups that do not frequently collaborate outside their immediate community.

In conclusion, the computer science network, characterised by moderate global transitivity and nearly neutral assortativity, suggests a balanced environment that supports frequent collaboration across all academic ranks, seemingly motivated more by reputation than by hierarchical influences. In contrast, the biology network's slightly lower transitivity and minimal negative assortativity, reveals a more restricted pattern of collaboration, lacking strong indicators of preferential attachment or significant rank-based collaboration. The isolated clusters, absence of transitivity and significantly negative assortativity in the sociology network demonstrates a more isolated community structure, suggesting a tendency towards specialisation and segmented collaboration within the field along with greater inter-rank collaboration.

Conclusion

Taken together, our findings provide tentative support for our hypothesis that increased feminisation of an academic field is associated with reduced collaboration.

Analyses exploring differences in centrality measures and degree distributions between fields identified a relationship between field feminisation and reduced collaboration. Although

Biology and Computer Science had similar average degrees and higher levels of collaboration compared to Sociology, only the Computer Science subgraph's degree distribution conformed to a power law. We also identified a trend where Biology authors had average node degrees smaller than Computer Science authors, but greater than Sociology ones. As such, a field's feminisation may be associated with reduced collaboration and reduced mechanisms of preferential attachment.

Moreover, as academic disciplines increasingly experience feminization, we see a decline in dynamic relationships and connectivity crucial for fostering cross-collaboration and academic achievement. In sociology, the disproportionately small ratio of nodes to edges indicates that edge formation is unlikely. Moreover, the prevalence of dyadic relationships dominates the network, impeding the spread and flow of information. Therefore, sociology, often perceived as a more feminine field, struggles to promote cross-collaboration, likely contributing to the gender-based productivity gaps impacting academia.

While examining clustering tendencies and title-based hierarchies across academic disciplines, we observe that fields experiencing increased feminization tend to show a decline in collaboration and a diminished impact of cross-hierarchical collaborations. Both the computer science and biology networks display a stronger inclination towards clustering, without exhibiting significant homophily based on academic titles, unlike the sociology network. In contrast, sociology features a more isolated structure dominated by dyadic relationships. This trend suggests that feminization within a field may lead to a less collaborative environment with a flatter hierarchical structure, which could hinder effective information transfer and networking.

This study has some limitations. As node attributes were inferred from Google Scholar profiles, representation of academic fields was biased, with many nodes being categorised as authoring in 'Other' academic fields, and a comparatively small number of nodes categorised as Sociology authors. Differences in subgraph size thus limits our capacity to draw conclusions from our comparisons of network mechanisms between fields. We hypothesise complexities in categorising authors according to their discipline, specifically, the elusiveness of sharp disciplinary boundaries in the social sciences (Fuller,1991), may have led to this surplus of authors publishing in 'Other' fields, and a paucity of Sociology authors. Future social network analyses exploring relationships between field feminisation and collaboration patterns would benefit from exploring a wider range of fields, and may benefit from avoiding vague attributional categories such as 'Other'.

Beyond expanding the number of fields explored, we propose future scholars consider a greater number of attributes in their analyses. Particularly, and in light of Spoon et al (2023) and E. Y. Li et al.,(2013)'s findings, we propose attributes such as the prestige of the author's academic institution and the length of tenure publishing in a discipline should be considered as potential moderators of the relationship between a field's feminisation, and the identified collaboration patterns.

Bibliography

Busch, F. (2017). Occupational Devaluation Due to Feminization? Causal Mechanics, Effect Heterogeneity, and Evidence from the United States, 1960 to 2010. Social Forces, 96(3), 1351–1376. https://doi.org/10.1093/sf/sox077

Bornmann, L., Bauer, J., & Haunschild, R. (2015). Distribution of women and men among highly cited scientists. *Journal of the Association for Information Science and Technology*, *66*(12), 2715–2716. https://doi.org/10.1002/asi.23583

Casad, B. J., Garasky, C. E., Jancetic, T. R., Brown, A. K., Franks, J. E., & Bach, C. R. (2022). U.S. Women Faculty in the Social Sciences Also Face Gender Inequalities. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.792756

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. https://doi.org/10.1037/bul0000052

Easley, D., & Kleinberg, J. (2010). Power laws and Rich-Get-Richer phenomena. In *Cambridge University Press eBooks* (pp. 479–494)

Feldon, D. F., Peugh, J., Maher, M., Roksa, J., & Tofel‐Grehl, C. (2017). Time-to-Credit gender inequities of First-Year PhD students in the Biological Sciences. *CBE- Life Sciences Education*, *16*(1), ar4. https://doi.org/10.1187/cbe.16-08-0237

Fiske, E. B. (2012). *World Atlas of Gender Equality in Education*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000215522

Fuller, S. (1991). Disciplinary boundaries and the rhetoric of the social sciences. *Poetics Today*, *12*(2), 301-325.https://doi.org/10.2307/1772855

Heller, N. (2023, February 27). The end of the English major. *The New Yorker*. https://www.newyorker.com/magazine/2023/03/06/the-end-of-the-english-major

Inoue, M., Pham, T., & Shimodaira, H. (2020). Joint estimation of non-parametric transitivity and preferential attachment functions in scientific co-authorship networks. *Journal of Informetrics*, *14*(3), 101042.

Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, *42*(9), 1515–1530. https://doi.org/10.1016/j.respol.2013.06.012

Li, W., Zhang, S., Zheng, Z., Cranmer, S. J., & Clauset, A. (2022). Untangling the network effects of productivity and prominence among scientists. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-32604-6

Newman, M. (2010). Measures and metrics. In *Oxford University Press eBooks* (pp. 168–234). https://doi.org/10.1093/acprof:oso/9780199206650.003.0007

Settles, I. H., Jones, M. S., Buchanan, N. T., & Brassel, S. T. (2021). Epistemic Exclusion of Women Faculty and Faculty of Color: Understanding Scholar(ly) Devaluation as a Predictor of Turnover Intentions. *Journal of Higher Education/˜the œJournal of Higher Education*, *93*(1), 31–55. https://doi.org/10.1080/00221546.2021.1914494

Spoon, K., LaBerge, N., Wapman, K. H., Zhang, S., Morgan, A. C., Galešić, M., Fosdick, B. K., Larremore, D. B., & Clauset, A. (2023). Gender and retention patterns among U.S. faculty. *Science Advances*, *9*(42). https://doi.org/10.1126/sciadv.adi2205

Wang, L. L., Stanovsky, G., Weihs, L., & Etzioni, O. (2021). Gender trends in computer science authorship. *Communications of the ACM*, *64*(3), 78–84. https://doi.org/10.1145/3430803

Yang, C., Ding, C., Hu, J., Chen, R., Hui, P., & Fu, X. (2017). Building and Analyzing a Global Co-Authorship Network Using Google Scholar Data. *Proceedings of the 26th International Conference on World Wide Web Companion*.
Link To GitHub : https://github.com/chenyang03/co-authorship-network