

Gender Gap in the Stacks: o paradoxo de género na investigação sobre Bibliotecas de Ensino Superior – uma análise bibliométrica

Ana Sofia Mota^a, Filipa Marinho^b

^a Universidade Portucalense, Portugal, amota@upt.pt

^b Universidade Portucalense, Portugal, filipamarinho@upt.pt

Resumo

A produção científica em Ciências da Informação e da Documentação (CID) evidencia desigualdades de género paradoxais numa área profissional predominantemente feminina. Este estudo apresenta uma análise bibliométrica da autoria por género na investigação sobre Bibliotecas de Ensino Superior (BES), com base num corpus bibliográfico proveniente da Scopus, da Web of Science e do OpenAlex, expandido para 33.885 ocorrências de autoria, no período 2000-2025. A inferência de género foi realizada através da plataforma Genderize.io e validada por amostragem estratificada, com concordância global de 80%. Os resultados revelam uma maioria feminina na primeira autoria (55,7%), um *glass ceiling* moderado (diferença de 4,6 pontos percentuais entre primeira e última autoria) e representação feminina maioritária em todos os quartis, incluindo o Q1 (59%). Não se verificaram diferenças estatisticamente significativas nas citações por género, nem padrões de homofilia nas redes de colaboração. A subanálise para Portugal combina o corpus indexado (52,9%) com o levantamento dos Cadernos BAD 2001–2025 (50,0% F, autores com afiliação portuguesa, validação integral), revelando que Portugal se situa abaixo da média internacional, com os *Cadernos BAD* a aproximarem-se da paridade (50,0%). Os resultados divergem de tendências frequentemente reportadas na literatura, sugerindo que a relação entre género e visibilidade científica em CID é mais complexa do que frequentemente assumida.

Palavras-chave: Desigualdade de género, Bibliometria, Ciências da Informação e da Documentação, Bibliotecas de Ensino Superior, Autoria científica.

Introdução

Nas Ciências da Informação e da Documentação (CID), uma área profissionalmente feminina em que as mulheres representam entre 72% e 80% da força de trabalho a nível internacional (Nazim et al., 2025), tendência confirmada em Portugal por Alvim e Vargues (2024), seria expectável que a produção científica refletisse essa preponderância. A literatura tem, porém, evidenciado desigualdades persistentes: sub-representação feminina na autoria sénior, nos conselhos editoriais e nos indicadores de impacto (Monroe-Gulick et al., 2024; Choji et al., 2024; McBurney et al., 2025).

Apesar do crescimento da participação das mulheres na academia nas últimas décadas, subsistem assimetrias significativas na autoria sénior, nas redes de colaboração internacional e no impacto académico das publicações (Huang et al., 2020). O relatório *She Figures 2024* documenta que a representação feminina na autoria científica decresce de 48% nas fases iniciais de carreira para 36% nas

posições sénior, padrão que os dados europeus mais recentes confirmam como persistente (European Commission, 2025).

A quase inexistência de investigação bibliométrica sobre género em CID no contexto português justifica este estudo, que analisa a produção científica sobre Bibliotecas de Ensino Superior (BES) entre 2000 e 2025, com base em 33.885 ocorrências de autoria provenientes da Scopus, Web of Science e OpenAlex, complementadas pelo levantamento dos Cadernos BAD (2001–2025).

A análise organiza-se em torno de três hipóteses: (H1) a participação feminina é maioritária na primeira autoria, mas subsistem assimetrias nos indicadores de autoria sénior e de impacto; (H2) a distribuição geográfica da autoria revela variações regionais significativas na proporção de primeiras autoras femininas; (H3) em Portugal, o desfasamento entre a predominância feminina na profissão e a visibilidade científica das autoras é mais pronunciado do que a média internacional.

Método

A recolha de dados foi realizada em três bases bibliográficas internacionais: Scopus, Web of Science Core Collection e OpenAlex, selecionadas pela sua cobertura complementar em CID (Maddi et al., 2025; Abrizah et al., 2013). A pesquisa incidu sobre o campo título, utilizando termos combinados relativos a bibliotecas académicas e de ensino superior, limitada ao período 2000–2025 e aos tipos de documento artigo, revisão e comunicação em conferência, devolvendo 9.068 registos na Scopus, 4.550 na Web of Science e 11.257 no OpenAlex. A opção por pesquisa em título visou maximizar a precisão temática e reduzir registos associados a outros domínios científicos.

O corpus principal para a análise bibliométrica relacional é constituído pelos registos Scopus e Web of Science, após deduplicação com a função `mergeDbSources` do pacote `bibliometrix` (R), totalizando 10.122 artigos únicos. O OpenAlex foi incorporado exclusivamente na construção do dataset de autoria para inferência de género, dado que os metadados revelaram lacunas em campos relacionais essenciais que dificultavam a análise bibliométrica integrada. Para a análise de género, o dataset foi expandido para uma linha por autor por artigo, totalizando 33.885 ocorrências de autoria, com identificação da posição (primeiro, último, único ou intermédio). A última autoria foi utilizada como proxy exploratória de autoria sénior, reconhecendo as variações disciplinares nas convenções de ordenação de autores em CID.

A atribuição de género foi realizada por inferência probabilística a partir do primeiro nome via `Genderize.io`, com limiar de probabilidade $\geq 0,80$. Cada autor foi classificado como feminino, masculino ou indeterminado, esta última categoria incluindo os casos abaixo do limiar e os nomes não reconhecidos (10,6% do total). A fiabilidade da inferência foi avaliada exploratoriamente por validação manual de uma amostra estratificada de 80 registos ($\kappa = 0,70$; concordância observada: 80,0%), com limitações conhecidas para nomes do Leste Asiático (VanHelene et al., 2024; Sebo, 2022).

Os indicadores organizam-se em quatro dimensões: produtividade, posição de autoria, colaboração científica e impacto académico, este último com base no SCImago Journal Rank (cobertura de 60,6% do corpus). O tratamento estatístico e as visualizações foram produzidos em R (pacotes `bibliometrix` e `ggplot2`).

A subanálise para Portugal assenta em duas fontes: os registos do corpus indexado com afiliação portuguesa ($n=1.752$) e o levantamento manual das autorias dos *Cadernos BAD* entre 2001 e 2025 (248 registos de autoria, 176 com afiliação portuguesa), com atribuição de género via `Genderize.io` e validação integral por reconhecimento onomástico (concordância de 99,6%, 247/248).

Resultados e discussão

Os resultados das quatro dimensões analíticas encontram-se sintetizados na Tabela 1.

Dimensão	Indicador	Resultado
Produtividade	Primeira autoria feminina (média 2000–2025)	55,7%
Posição de autoria	<i>Glass ceiling</i> (1. ^a vs. última autoria)	4,6 pp
Posição de autoria	Autoria única feminina	55,3%
Impacto - revistas Q1	Proporção de primeiras autoras femininas	59,0%
Impacto - citações	Média F vs. M (Mann-Whitney $p=0,332$)	Não significativa
Colaboração	Homofilia feminina observada vs. esperada	26,8% vs. 31,0%
Portugal	Primeira autoria feminina	52,9%

Tabela 1- Síntese dos principais resultados por dimensão analítica.

A análise da primeira autoria revela maioria feminina em quase todo o período 2000–2025, com uma média de 55,7% (Figura 1), resultado que diverge de padrões frequentemente reportados na literatura sobre sub-representação feminina na produção científica. A variação anual oscila entre 49,4% (2025) e 65,3% (2001), sem tendência linear de crescimento, o que sugere um padrão relativamente estável no corpus analisado.

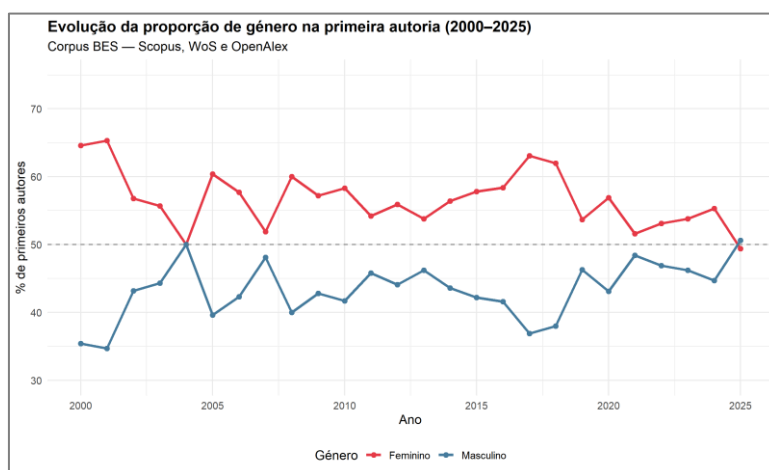


Figura 1- Evolução da proporção de género na primeira autoria (2000–2025).

A análise por posição de autoria evidencia um *glass ceiling* moderado, mas persistente: 55,7% de primeiras autorias femininas contra 51,1% de últimas autorias (Figura 2), com uma diferença de 4,6 pontos percentuais.

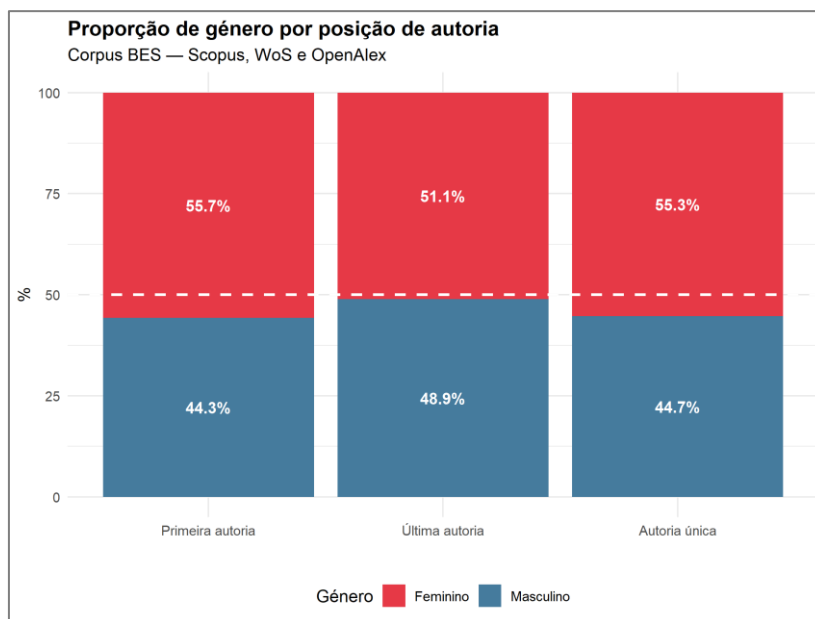


Figura 2- Proporção de género por posição de autoria.

A evolução temporal desta assimetria sugere estabilidade ao longo dos 25 anos, com picos em 2000 (~13 pp) e 2018 (~11 pp) (Figura 3), embora não se observe uma tendência clara de redução ao longo do período.

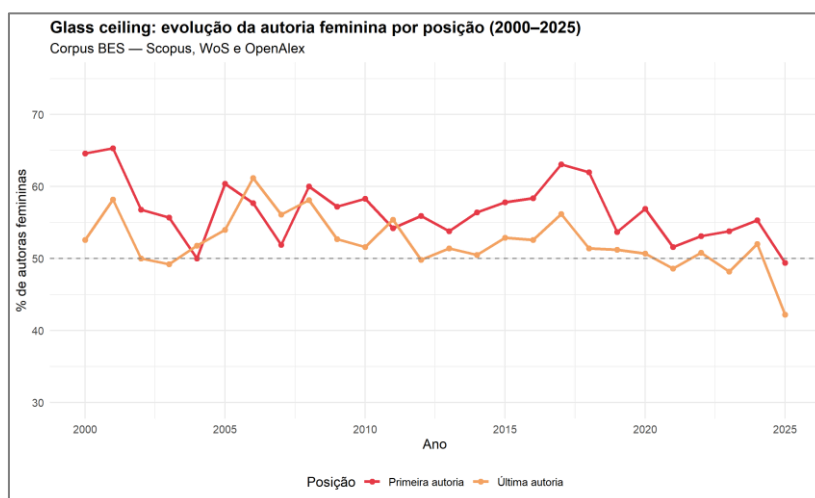


Figura 3 - Glass ceiling: diferença entre primeira e última autoria feminina (2000–2025).

As autoras são maioritárias em todos os quartis, incluindo o Q1 (59%), observando-se uma relação positiva entre o quartil da revista e a proporção feminina (Figura 4), padrão distinto do documentado noutras disciplinas (Yesmin & Tasmim, 2025; McBurney et al., 2025).

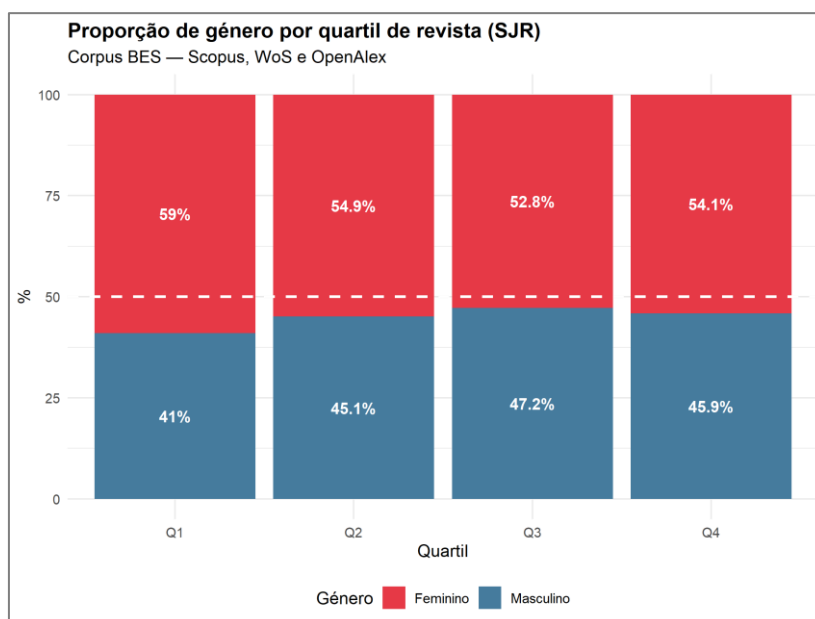


Figura 4 - Proporção de género por quartil de revista (SJR).

Não se verificaram diferenças estatisticamente significativas nas citações por género (Figura 5): a média de citações é de 9,37 para autoras femininas e 8,80 para autores masculinos, com medianas iguais (3) e um *p-value* de 0,332 no teste de Mann-Whitney.

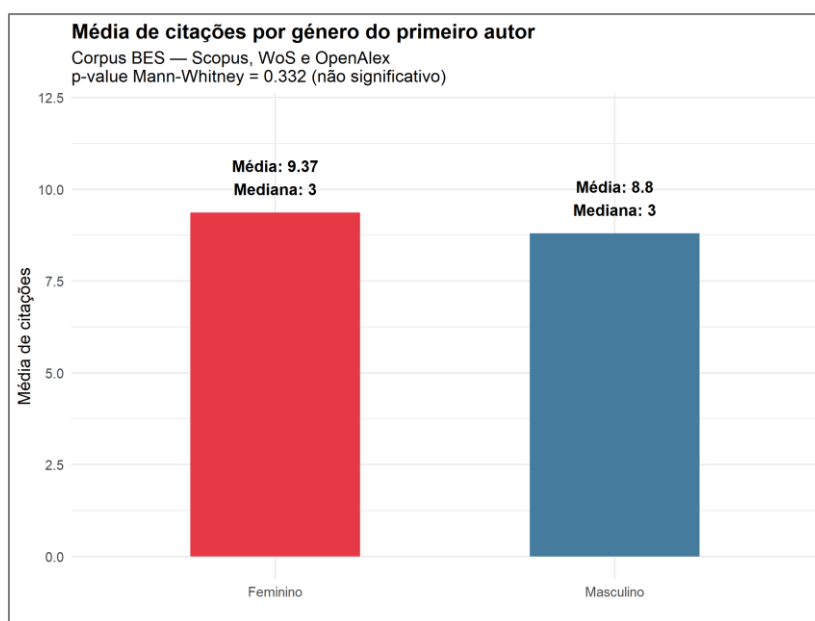


Figura 5 - Média de citações por género do primeiro autor.

A análise da colaboração não confirmou homofilia feminina (Figura 6): a colaboração em equipas exclusivamente femininas (26,8%) é inferior ao esperado pelo acaso (31%), enquanto a colaboração mista (54,3%) supera o esperado (49,4%), padrão que diverge do documentado por Wang et al. (2023) e Bunker Whittington et al. (2024) noutras disciplinas e regiões.

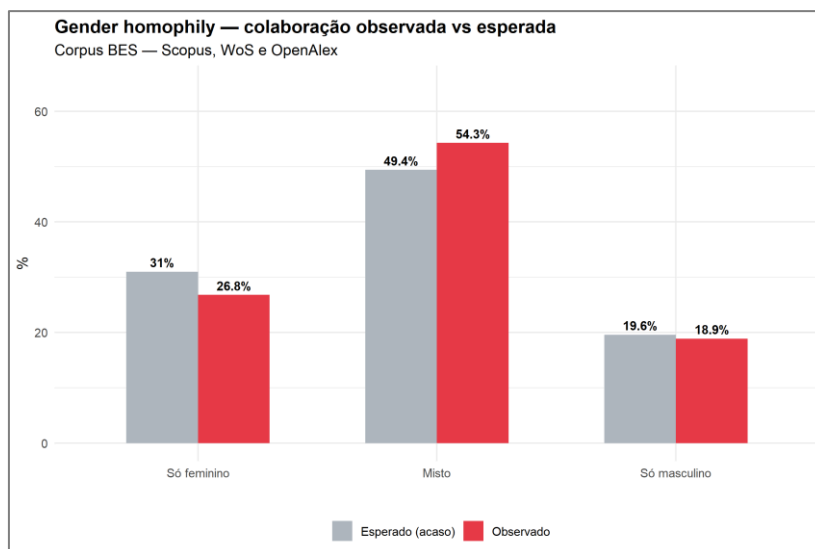


Figura 6 – Homofilia de género: colaboração observada vs. esperada.

A subanálise para Portugal revela 52,9% de primeiras autorias femininas no corpus indexado (Figura 7, n=1.752), acima da paridade, mas 2,8 pp abaixo da média internacional, resultado coerente com a literatura sobre produtividade científica feminina em Portugal (Santos et al., 2023) e com o perfil profissional documentado por Alvim e Vargues (2024).

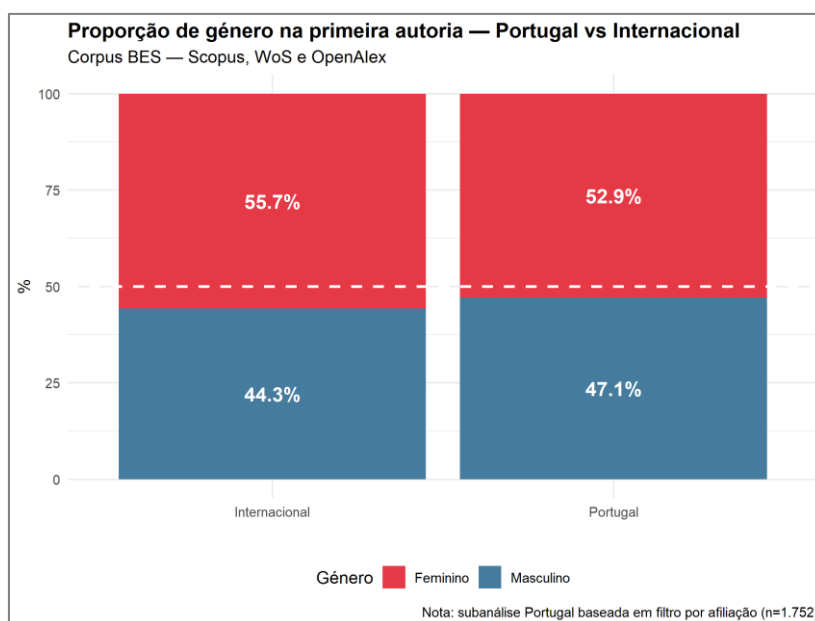


Figura 7 - Proporção de género na primeira autoria, Portugal vs. Internacional.

Nota: subanálise Portugal baseada em filtro por afiliação (n=1.752).

A subanálise dos *Cadernos BAD* (2001–2025) abrange 248 registos de autoria, dos quais 176 com afiliação em instituição portuguesa. A atribuição de género foi validada integralmente em contexto cultural português, obtendo-se uma concordância de 99,6% (247/248). Os resultados revelam paridade na autoria de primeiro autor com afiliação portuguesa: 50,0% feminino e 50,0% masculino (n=176). Este valor, observado nos *Cadernos BAD*, é inferior ao registado no corpus indexado português (52,9%) e à média internacional (55,7%).

A paridade observada nos *Cadernos BAD*, revista nacional não indexada nas bases internacionais, pode indiciar diferenças nos padrões de publicação entre circuitos nacionais e internacionais, hipótese que merece investigação futura.

Conclusões

Os resultados divergem de tendências frequentemente reportadas na literatura: a feminização da profissão associa-se a uma presença feminina maioritária na autoria científica, incluindo em revistas classificadas nos quartis superiores, sem diferenças estatisticamente significativas nas citações e sem evidência de homofilia feminina nas redes de colaboração.

A H1 encontra suporte parcial nos resultados: a participação feminina é maioritária, mas subsistem assimetrias na autoria sénior. A H2 encontra suporte quanto à variação geográfica, mas não relativamente aos padrões de colaboração, que em BES divergem do documentado noutras disciplinas. A H3 encontra suporte nos resultados: observa-se uma gradação decrescente entre o corpus internacional (55,7%), o corpus indexado português (52,9%) e os *Cadernos BAD* (50,0%), padrão que pode indiciar diferenças entre circuitos nacionais e internacionais de publicação científica.

Destacam-se três limitações: a cobertura parcial do SCImago (60,6% do corpus), que pode introduzir viés na análise por quartil; as limitações do Genderize.io para nomes do Leste Asiático (concordância regional de 57,1%), que afetam a precisão da inferência nesse segmento; e a opção por pesquisa exclusivamente em título, orientada para maximizar a precisão temática, mas que pode ter excluído estudos relevantes não explicitamente centrados em BES.

Os resultados sugerem que a relação entre género e visibilidade científica em CID é mais complexa do que frequentemente descrito na literatura, reforçando a necessidade de análises contextualizadas e comparativas, nomeadamente no espaço lusófono.

Referências

- Abrizah, A., Zainab, A. N., Kiran, K., & Raj, R. G. (2013). LIS journals scientific impact and subject categorization: A comparison between Web of Science and Scopus. *Scientometrics*, 94(2), 721–740. <https://doi.org/10.1007/s11192-012-0813-7>
- Alvim, L., & Vargues, M. M. (2024). *Estudo sobre os profissionais de informação em Portugal (2021–2023)*. Associação Portuguesa de Bibliotecários, Arquivistas e Profissionais da Informação (BAD). <https://publicacoes.bad.pt/ebooks/index.php/bad/catalog/book/24>
- Bunker Whittington, K., King, M. M., & Cingolani, I. (2024). Structure, status, and span: Gender differences in co-authorship networks across 16 region-subject pairs (2009–2013). *Scientometrics*, 129, 147–179. <https://doi.org/10.1007/s11192-023-04885-1>
- Choji, T. T., Cobo, M. J., & Moral-Munoz, J. A. (2024). Is the scientific impact of the LIS themes gender-biased? A bibliometric analysis of the evolution, scientific impact, and relative contribution by gender from 2007 to 2022. *Scientometrics*, 129, 6023–6047. <https://doi.org/10.1007/s11192-024-05005-3>
- European Commission. (2025). *She figures 2024: Gender in research and innovation. Statistics and indicators*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/592260>

- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. <https://doi.org/10.1073/pnas.1914221117>
- Maddi, A., Maisonobe, M., & Boukacem-Zeghmouri, C. (2025). Geographical and disciplinary coverage of open access journals: OpenAlex, Scopus, and WoS. *PLoS ONE*. <https://doi.org/10.48550/arXiv.2411.03325>
- McBurney, J., Hennesy, C., & Gyendina, M. (2025). From bad to worse: Gender equity in LIS journals during the COVID-19 pandemic. In *Democratizing Knowledge, Access, Opportunities: Conference Proceedings ACRL 2025*. <https://www.ala.org/sites/default/files/2025-03/FromBad-Worse.pdf>
- Monroe-Gulick, A., Weaver, M. D., & Morris, S. E. (2024). Women and men in library and information science scholarship: Authorship, collaboration and citation patterns (2003–2021). *The Journal of Academic Librarianship*, 50(5), 102–115. <https://doi.org/10.1016/j.acalib.2024.102939>
- Nazim, M., Ali, A., Gupta, A., & Ali, M. (2025). Exploring gender disparities in scholarly communication metrics within digital library research. *DESIDOC Journal of Library and Information Technology*. 45(4), pp. 339-349. <https://publicationsdrdo.in/index.php/djlit/article/download/21027/8621/89623>
- Santos, C., Monteiro, R., Lopes, M., & Ferreira, V. (2023). From late bloomer to booming: A bibliometric analysis of Women's, Gender, and Feminist Studies in Portugal. *Social Sciences*, 12(7), 396. <https://doi.org/10.3390/socsci12070396>
- Sebo, P. (2022). How accurate are gender detection tools in predicting the gender for Chinese names? *Journal of the Medical Library Association*, 110(1), 71–75. <https://doi.org/10.5195/jmla.2022.1289>
- VanHelene, A. D., Khatri, I., Beau Hilton, C., & Warner, J. L. (2024). Inferring gender from first names: Comparing the accuracy of Genderize, Gender API, and the gender R package on authors of diverse nationality. *PLOS Digital Health*. <https://doi.org/10.1371/journal.pdig.0000456>
- Wang, Y. S., Lee, C. J., West, J. D., & Erosheva, E. A. (2023). Gender-based homophily in collaborations across a heterogeneous scholarly landscape. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0283106>
- Yesmin, S., & Tasmim, S. (2025). A study of editorial boards composition: A review of Q1 journals in the field of library and information science. *Scientometrics*. 130, 6227–6251. <https://doi.org/10.1007/s11192-025-05439-3>