



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

Thamyres T. Choji^{1,2}  · Manuel J. Cobo^{3,4} · Jose A. Moral-Munoz^{1,5}

Received: 8 November 2023 / Accepted: 22 March 2024
© The Author(s) 2024

Abstract

Gender equity remains a challenge both globally and within academia, despite recent efforts to change it. Moreover, beyond the authors' productivity, studies have reported that women often achieve lower scientific impact than their peers. To shed light on this complex relationship between the scientific impact and the themes addressed, this study conducts a comprehensive analysis of Library and Information Science field from 2007 to 2022 in four consecutive slides, identifying the principal themes covered in the field, analyzing the relative gender contribution rate, employing strategic diagrams, and assessing impact metrics, such as mean normalized citation score, 1% of most cited papers, and H-Classic. We employed science mapping analysis to explore a core of 45,650 documents from the Web of Science, with gender identification in 94.25% of cases. Our findings revealed a slight increase in the percentage of women authors within the field across the time, and a decline during the COVID-19 pandemic, demonstrating consistency in the themes addressed over the years. Women were overrepresented in the classic themes of LIS, human, and health-related fields, with these themes displaying lower performance rates. In contrast, men authors were overrepresented in STEM-related fields and innovation themes, associated with higher metric values. Our findings underlined the potential association between research themes and scientific performance, and provide societal and structural explanations for these observations. This study contributes valuable insights into the relationship between research themes and the scientific impact achieved by researchers in LIS, highlighting the importance of encouraging women's participation in diverse knowledge domains and challenging prevailing stereotypes within academic and professional spheres.

Keywords Gender differences · Bibliometrics · Science mapping · Co-word analysis · Scientific impact

Introduction

The debate surrounding gender inequity is more relevant than ever, as it has reached new heights, with initiatives and research shedding light on the imbalances across various contexts. Recently, the Nobel Prize in Economics recognized the work of Goldin, who delved into the historical discoveries of the gender gap over two centuries. Her research has significantly contributed by contextualizing the presence of women in the labor market, addressing aspects like social norms, institutional structures, and assessing the impacts of significant events, such as the advent of the contraceptive pill (Goldin & Katz, 2002; Nobel Prize Outreach AB, 2023). Furthermore, the current availability of data has enhanced our capacity to detect, analyze, and comprehend disparities, enabling us to examine specific contexts and uncover their significance.

Within the broader context of gender disparities, academia is not an exception. Despite recent efforts to reshape the scientific landscape, gender imbalance persists within academia (Huang et al., 2020; Lariviere et al., 2013). This issue is not solely a question of the number of women scientists (UNESCO Institute for Statistics (UIS), 2019), since they constitute the majority of Ph.D. candidates (Organisation for Economic Co-operation and Development (OECD), 2022), but they remain underrepresented in leadership roles and higher-ranking positions (Fagan & Teasdale, 2021). Notably, women scientists tend to produce less research output (Zhang et al., 2021) and achieve lower impact with their work compared with their counterparts (Abramo et al., 2015; Chen & Seto, 2022; Joanis & Patil, 2022; Kozłowski et al., 2022; Lariviere et al., 2013; Liu et al., 2020; Mauleón & Bordons, 2006). Regarding funding and grant distribution, the results are inconclusive, with studies supporting the notion that women receive less investment and that there is no difference (Bol et al., 2022; Bornmann et al., 2007; Chaudhary et al., 2021; Marsh et al., 2009). A recent study showed that a large proportion of women do not experience a sense of belonging in academic environments and are more likely to leave the profession due to this lack (Spoon et al., 2023). Nevertheless, women who remain in the academic system often receive lower incomes (Moss-Racusin et al., 2012; Darren & Steven, 2010), and, ultimately, experience the impact of disparities in hiring decisions (Shen, 2013).

The underrepresentation of women in the scientific system is more evident in some fields of knowledge. Notably, in science, technology, engineering, and mathematics (STEM) fields, this gap is significantly higher (Holman et al., 2018), and even more pronounced in certain STEM subfields, such as computer science and math (Ceci et al., 2014; De Nicola & D'Agostino, 2021; Fagan & Teasdale, 2021). In contrast, there appears to be more gender equity in human and health-related fields (Su & Rounds, 2015). The scenario gains complexity if we consider aspects such as race and ethnicity through its intersection with gender, as it could be related to the choice of research topics (Kozłowski et al., 2022). The explanations for the unequal women's participation in different fields of knowledge remain unclear, and the debate revolves around structural problems associated with gender roles (Ceci et al., 2014; Petrongolo, 2019).

To understand and evaluate the complexity of the scientific system, we employed the Science of Science (Zeng et al., 2017), which allows us to structure science as a conceptual, social, and intellectual structure and interpret it from a multidisciplinary perspective (Moral-Munoz et al., 2020). The diversity of methodological approaches allows us to delve into specific fields, understand their evolution, identify behavioral patterns, and gain deeper insights into the dynamics of knowledge production. The significance of this approach lies in its ability to uncover meaningful and interpretable insights into

the various roles within science. This, in turn, enhances our understanding of the past, which is essential for comprehending the present, and provides the foundation for new projections aimed at creating a future characterized by greater fairness and transparency. Nevertheless, due to the complexity of the scientific system, accurately quantifying and assessing knowledge transfer remains a challenge. This complexity arises from specificities, including the distinct behaviors exhibited by different fields, making it difficult to precisely evaluate each phenomenon (Bornmann, 2017). To address this complexity, the approach should involve understanding different contexts and perspectives, determining what needs to be measured, and determining the best way to do it (Nygaard & Bahgat, 2018; Zeng et al., 2017).

Given this background, conducting studies within specific fields of knowledge is a pragmatic need to deepen our understanding. This necessity arises from the fact that behavioral patterns within each discipline, field or subfield can be unique, making it crucial to comprehend and standardize measurement within each context (Butler, 2008; Dorta-González & Dorta-González, 2013). For instance, the Library and Information Science (LIS) field has traditionally been predominantly women-dominated (Piper & Collamer, 2001). In earlier times, women in the LIS field were often related to and specialized in library education, whereas men were related to information science or high status activities (Harris et al., 1985; Varlejs & Dalrymple, 1986). Nowadays, the gender scenario in the LIS field varies according to the analysis context. For instance, in terms of authorship, in the USA, LIS is a field where there are more women than men as first authors (Thelwall & Mas-Bleda, 2020), whereas the opposite occurs in India (Parabhoi et al., 2020). Nevertheless, our understanding of this field through a gender lens remains limited. Gender-focused studies in LIS have primarily centered on select LIS print and e-journals, exploring authorship, productivity, and citation patterns. Although the presence of women authors is increasing, in general numbers, in both print journals and e-journals over the years, women are more present in e-journals than in print ones. For e-journals, the publication and citation rates of women's contributions have also increased over time. Meanwhile, when looking at print journals, the increase was seen specifically in library science journals, but in information science journals, the presence of women remained stable (Linsay, 2010; Lund & Shamsi, 2021; Vinay, 2021). In contrast, the e-journals showed the opposite; in there, men's participation has increased over the years, but the citation rate shows no differences by gender (Gul et al., 2016). In this case, the citation rate showed an increase in gender collaborative teams, resulting in more grants than men-men collaborative work. In a local context, an earlier study analyzed the gender balance in the Indian context finding similarities with the journal observations: men outnumber women authors, but in this specific case, women published more in international journals than in national ones (Parabhoi et al., 2020).

Among the factors that may affect the scientific production and impact of researchers are the themes they address. To gain insights into the relationship between these characteristics, we explored the LIS field and its evolutionary dynamics, analyzing the field from 2007 to 2022. This analysis adopted a conceptual approach, including assessing the relative contribution by gender to each theme, its scientific impact, and a science mapping analysis. This approach aims to shed light on whether the themes addressed by each gender contribute to the observed disparity in their scientific impact.

The subsequent sections are organized as follows: Sect. 2 outlines the methods employed in our analysis. Then, in Sect. 3, the results obtained are presented. Sect. 4 discusses the results, addresses limitations, and suggests future research. Finally, Sect. 5 provides the conclusions of the study.

Methodology

As shown in the introduction, our hypothesis is that the difference in the scientific output and impact of women and men could derive from a subject bias. That is, each gender might address different subfields with different citation patterns. We can use science mapping analysis based on co-word networks and bibliometric indicators to test whether this pattern occurs in the LIS field. To this end, we propose a methodology based on five phases: (i) Data acquisition, (ii) Data preparation, (iii) Data selection, (iv) Theme detection and (v) Performance analysis.

Data acquisition

For further analysis of the themes addressed and their relative gender composition in the LIS field from 2007 to 2022, we retrieved the corpus from the Web of Science (WoS) using the following criteria:

- Documents indexed on the Science Citation IndexTM Expanded (SCI) or Social Sciences Citation Index® (SSCI).
- Documents under the “Information Science & Library Science” category.
- Documents produced between 2007 and 2022.
- Documents tagged as articles or reviews.

The query was executed on January 23, 2023, and retrieved 63,843 documents. The data was downloaded using the WoS API (Velez-Estevez et al., 2023).

Data preprocessing

To ensure the accuracy of our bibliometrics analysis, the preprocessing step is fundamental. In our case, we have two units of analysis: on one hand, we employed keywords to define themes and ultimately generate the science maps. On the other hand, we used authors' names to infer gender. For this analysis, the keywords used are both author keywords and the KeywordsPlus, generated by WoS (Garfield, 1990). Keyword preprocessing is crucial for preventing misinterpretation during subsequent co-word analysis. This step was conducted using the open-source software SciMAT (Cobo et al., 2011, 2012). The first stage of preprocessing involved the unification of singular and plural words, followed by a de-duplication step to terms that describe similar concepts (e.g., ‘Hirsch index’ and ‘*h*-index,’ ‘NN’ and ‘Neural-Networks’). This step also included the identification of terms with broad meanings, such as ‘study,’ ‘analysis,’ and ‘algorithms.’ To ensure the quality of this process, the authors reviewed the whole set of preprocessed keywords.

The subsequent step consists on the gender identification of the authors. To achieve this, the given name of each author was employed to determine the gender. Three datasets containing given names and gender information were used: the United States Census (Social Security Administration, 2022), the NamesDatabase repository (Leaderboard, 2022), and the Python library GenderGuesser (Pérez, 2016). To handle cases with unisex names, we developed an algorithm for disambiguation that assigns gender based on the gender associated with the name in 90% of the instances. We selected a 90% threshold based on previous research that employed the first name of the author to assign gender (Thelwall & Mas-Bleda, 2020; Thelwall et al., 2019, 2023), and expect that this threshold would be able to

ensure that the gender assignment would almost always be correct. It is worth noting that in this study, gender was considered binary due to the fact that other gender identities require self-identification.

Despite having used different APIs and methods to determine the gender of each author, there may be names for which we have not been able to associate gender. Considering that our analysis is at document-level, it was necessary to perform an additional step to filter out documents in which the number of authors of unknown gender is higher or equal than that of male or female authors. Thus, after data filtering, the corpus had 45,650 documents (71.5% of the retrieved at data acquisition step).

As mentioned above, our aim is to determine the thematic differences between men and women and also, analyze how this difference changes and evolves over time. For this reason, we perform our analysis under a longitudinal framework, splitting the corpus into four consecutive periods: (i) From 2007 to 2010, containing 9,845 documents, (ii) From 2011 to 2014, containing 11,125 documents, (iii) From 2015 to 2018, containing 12,320 documents, and (iv) From 2019 to 2022, containing 12,360 documents.

Themes detection

After data selection, four co-word networks were built using the keywords of each document for each of the previously mentioned periods. In this type of network, nodes represent the keywords found within the corpus, and edges represent the co-occurrence relationships, indicating connections between these words (Batagelj & Cerinšek, 2013; Cobo et al., 2011).

To ensure data consistency and minimize bias, we normalized the co-occurrence frequencies of the networks obtained in the previous step using the equivalence index (Callon et al., 1991). Subsequently, the *Leiden clustering* algorithm Traag et al. (2019) was used to perform the community detection. This process results in a collection of clusters that represent the topics or themes covered within each dataset. The choice of this algorithm was influenced by its ability to identify cohesive communities. After community detection, each theme was named using the most central keyword of all associated keywords. This approach selects the word most connected in each community (Cobo et al., 2011). In our case, these central words were effective in summarizing the structure of communities.

Once we determined the themes addressed in the LIS field for each dataset, four strategic diagrams were built. These diagrams provide a two-dimensional representation based on measures of centrality and density for each theme. Centrality measures the external cohesion of each theme in relationship to other detected themes, while density reflects the internal cohesion within each theme and reveals how the keywords are interconnected among themselves (Cobo et al., 2011). Based on these variables, the themes are categorized into four groups: “Motor”, “Basic and Transversal”, “High developed and Isolated” and “Emerging or Declining”. Motor themes exhibit high centrality and density, indicating that these themes hold a central position compared to other themes, and the keywords within them display strong cohesion, suggesting that these themes are well-developed and well-structured. Similarly, Basic and Transversal themes are those showing high centrality between themes, although internal cohesion is not high, representing general and transversal themes in the field. High developed and Isolated themes are themes with low centrality and high density, implying that they are in the backbone of the research area. However, the keywords within these themes exhibit high cohesion, representing specialized or peripheral themes. Finally, Emerging or Declining themes, characterized by low centrality and

density, represent themes with weak centrality and cohesion, indicating either emerging or declining themes.

Performance analysis

In order to quantify and analyze the presence and impact of women in the detected themes, a performance analysis was conducted. Various metrics were computed, including the number of papers within the theme, sum of citations, geometric mean of citations, mean normalized citation score (MNCS) (Waltman et al., 2011), top 1% of most cited papers, percentage of documents in H-Classic (Martínez et al., 2014), percentage of women and men within each theme, and relative gender contribution rate (RGCR).

The MNCS is a metric for assessing the citation impact, facilitating comparisons across different disciplines, fields, subfields, and themes. Within the scope of our analysis, MNCS is computed by considering the citations received within each theme and the average citations within the LIS field for each year. When the MNCS value equals 1, the average number of citations received within a particular theme is equivalent to the average of citations received in the LIS field. Conversely, when the MNCS exceeds 1, the average number of citations for this theme surpasses that in the LIS field (Waltman et al., 2011). Meanwhile, calculating the metric for the 1% of most cited papers involves determining the percentage of documents that achieve this categorization. This calculation relies on two components: (i) the number of most cited documents in the dataset for a given year and (ii) the number of most cited documents within each theme for the same year. The H-Classic metric computation is based on the h-index (Hirsch, 2005). This involves the examination of documents that are part of the *h*-core, which is defined as a set of papers, each having *h* or more citations (Martínez et al., 2014).

The percentage of women and men within each theme is measured from the gender assigned to the authors, enabling us to analyze representativeness in each case. So, considering the global disparity in the number of women and men researchers (Holman et al., 2018), it is expected that the distribution of men across disciplines, fields, subfields, or themes will probably outnumber women. In order to recognize, assess, and normalize this initial imbalance, we introduce the RGCR metric. The RGCR evaluates the presence of women authors within a specific theme concerning both the overall men within the theme and the global dataset. It provides a measure indicating whether the gender contribution rate is equal, higher, or lower than the average of women and men participation in the LIS field. The calculation of RGCR is based on the percentage of women and men authors within the initial dataset and their respective percentages within each. The RGCR can be computed using the following Eq. (1):

$$RGCR = \frac{\left(\frac{X_{Women-theme}}{X_{Women-global}} \right)}{\left(\frac{Y_{Men-theme}}{Y_{Men-global}} \right)} \tag{1}$$

In this equation, $X_{Women-theme}$ represents the percentage of women authors in the specific theme, $X_{Women-global}$ is the percentage of women authors in the global dataset, $Y_{Men-theme}$ denotes the percentage of men authors in the theme, and $Y_{Men-global}$ signifies the percentage of men authors in the global dataset. When RGCR is equal to or close to 1, it indicates that this theme was more gender balanced. Values greater than 1 indicate that women authors were overrepresented in that theme, whereas values lower than 1 indicate

underrepresentation. This approach provides the advantage of assessing the presence of women authors in each theme concerning the overall presence of women in the LIS field from 2007 to 2022.

To enhance our understanding of the relationship between the RGCR and MNCS achieved by each theme during the analyzed periods, a set of graphics was generated. In these visual representations, the *y-axis* represents the RGCR, the *x-axis* represents the MNCS, and the diameter of each circle corresponds to the number of documents in each topic. This delineation, combined with the strategic diagrams, provides a deeper insight into the presence of women in the themes and their importance in the field, facilitating the understanding of their similarities and differences.

Results

After carrying out the steps of the methodology, the global dataset, between 2007 and 2022, contained 45,650 documents (representing 71.5% of the data retrieved initially) authored by 126,418 authors, of which 40.85% (51,644) of the authors were women, 53.6% (67,505) were men, and 5.55% (7,269) unknown. The global dataset was divided into four consecutive periods to uncover the thematic differences between women and men and how they evolve. Information such as the number of documents used and the representativeness of the genders in each case can be observed in Fig. 1. In what follows, we analyzed each period, uncovering the themes that women and men were focused on and computing their performance and impact measures.

From 2007 to 2010

The period from 2007 to 2010 comprised 9,845 documents, in which 42,68% of the authors were women. The strategic diagram identified 21 themes, as illustrated in Figs. 2 and 3. Among these themes, women were overrepresented in four of them, with greater presence

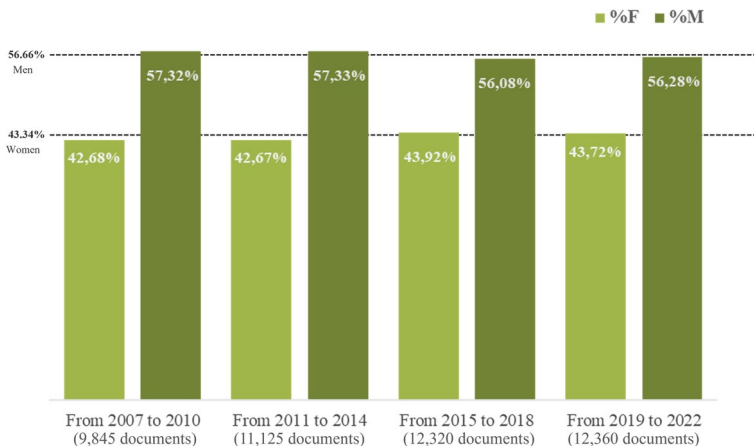


Fig. 1 Percentage gender over the analyzed periods: 2007 to 2010, 2011 to 2014, 2015 to 2018, and 2019 to 2022

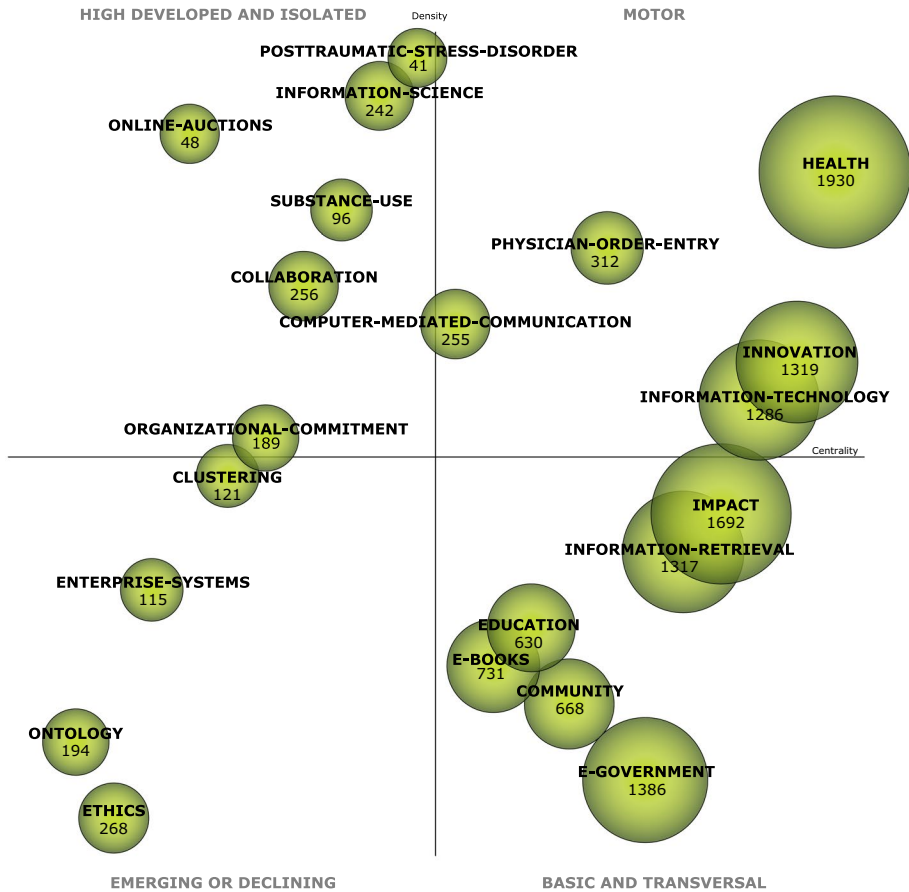


Fig. 2 Strategic diagram from 2007 to 2010, where the *x*-axis and *y*-axis represent the centrality and density, respectively, of each community. Circumference sizes indicate the number of documents in each community, while their position allows classification as 'Motor', 'Basic and Transversal', 'Emerging and Declining', and 'High developed and Isolated'

in *Substance-use* and *Posttraumatic-stress-disorder*. Meanwhile, men were overrepresented in 13 themes, with greater presence in *Clustering* and *Online-auctions*. The themes with similar ratios of each gender were *Education*, *Physician-order-entry*, *Information-retrieval* and *Computed-mediated-communication*.

Among the four themes overrepresented by women, i.e. with high RGCR values, three of them can be linked to health-related fields, such as *Health*, *Substance-use* and *Posttraumatic-stress-disorder*. The other one, *E-books*, can be related to a classic theme of LIS. *Health* had the highest productivity from 2007 to 2010 and was considered as motor theme, underscoring its importance as a theme within the LIS field. In terms of scientific impact (Table 1), *Health* presented 1.56 of MNCS, around 4% of documents in the 1% of most cited papers, and 5.5% in H-Class. Meanwhile, the productivity of *Substance-use* and *Posttraumatic-stress-disorder* was low compared to other themes, with both positioned as highly developed and isolated. Although both themes presented MNCS around 1.5, neither reached 1% of documents in the 1% of most cited papers and H-Class. In contrast,

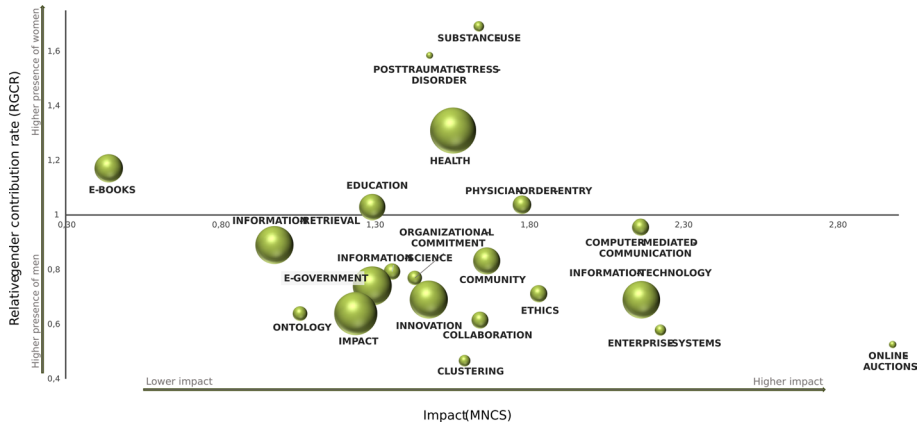


Fig. 3 Distribution of themes from 2007 to 2010 based on their impact (MNCS, *x*-axis) and relative gender contribution rate (RGCR, *y*-axis). Circumference sizes indicate the number of documents in each community. The upper quadrant displays the topics where women authors were over-represented, while the lower quadrant shows themes over-represented by men. Themes along the *x*-axis are considered more gender balanced

E-books was the theme with the lowest MNCS value and almost no documents were in the 1% of most cited papers and H-Classific.

Regarding the 13 themes overrepresented by men, i.e. with low RGCR values, most of them can be linked to STEM fields or classical themes of LIS, such as *Clustering* and *Impact*, respectively. Among these themes, *Impact*, *Information-technology*, *Innovation*, *E-government* and *Information-retrieval* were shown to have the highest productivity and in the strategic diagram were positioned as motor and/or basic and transversal. In terms of MNCS, all these themes exceed the average, with *Clustering*, *Online-auctions*, and *Enterprise-systems* presenting values higher than 2. Furthermore, all the themes overrepresented by men presented at least 0.5% of documents in the 1% of most cited papers and H-Classific.

Finally, it should be pointed out that from 2007 to 2010 the themes with the highest impact measures were those overrepresented by men. For instance, *Information-technology*, *Impact*, and *Innovation* were the most productive and exhibited a high percentage of papers in the 1% of most cited papers and H-Classific. *Online-auctions* and *Enterprise-systems* were themes that achieved the highest MNCS among 21 identified themes.

From 2011 to 2014

The period from 2011 to 2014 consisted of 11,125 documents, being women 47.67% of the authors. The percentage of women authors was slightly higher than observed from 2007 to 2010. The strategic diagram identified 20 themes (Figs. 4 and 5), six of them were overrepresented by women, and 13 themes were overrepresented by men. Only *Information-quality* presented a rate of gender similar to the global dataset (see Table 2).

The six themes overrepresented by women are more related to classical themes of the LIS field and health-related fields. Of these themes, *Librarianship*, *Academic-libraries* and *Information-literacy* did not achieve the average of MNCS, and among them, *Librarianship* was highlighted by its 3.2 RGCR value, being the theme with the highest presence of women. It showed the lowest MNCS among all identified themes

Table 1 Performance table and impact indicators of LIS field from 2007 to 2010

Theme	Docs	Cit	CGM	MNCS	%1%	%HCP	%F	%M	RGCR
Substance-use	96	5021	29.65	1.64	0.14	0.53	56.4%	43.6%	1.69
Posttraumatic-stress-disorder	41	1881	32.03	1.48	0	0.09	54.8%	45.2%	1.58
Health	1930	96665	24.07	1.56	3.92	5.5	50%	50%	1.31
E-books	731	10211	7.10	0.44	0.14	0.27	47.2%	52.8%	1.17
Physician-order-entry	312	17754	28.87	1.78	0.81	1.15	44.2%	55.8%	1.04
Education	630	25669	15.54	1.29	0.95	1.02	44%	56%	1.03
Computer-mediated-communication	255	17222	30.76	2.16	0.95	1.46	42.2%	57.8%	0.96
Ontology	194	6671	17.73	1.06	0.27	0.27	41%	59%	0.91
Information-retrieval	1317	41038	14.08	0.98	1.49	2.31	40.5%	59.5%	0.89
Community	668	36121	20.30	1.66	1.49	2.17	38.9%	61.1%	0.83
Information-science	242	10297	11.80	1.36	0.27	0.49	37.8%	62.2%	0.79
Organizational-commitment	189	8630	22.00	1.43	0.41	0.58	37.1%	62.9%	0.77
E-government	1386	57595	18.22	1.29	2.3	3.55	36.2%	63.8%	0.74
Ethics	268	16463	24.24	1.83	0.54	1.06	35.3%	64.7%	0.71
Innovation	1319	62078	22.17	1.48	2.57	3.95	34.6%	65.4%	0.69
Information-technology	1286	89746	29.53	2.17	5.68	6.43	34.6%	65.4%	0.69
Impact	1692	66232	17.98	1.24	2.17	3.59	32.8%	67.2%	0.64
Collaboration	256	13064	25.61	1.64	0.27	1.02	32%	68%	0.62
Enterprise-systems	115	8726	31.10	2.23	0.54	0.44	30.7%	69.3%	0.58
Online-auctions	48	4803	35.77	2.98	0.27	0.31	28.7%	71.3%	0.53
Clustering	121	6149	22.45	1.59	0.27	0.31	26.3%	73.7%	0.47

Docs documents produced in theme, *Cit* citations received in theme, *CGM* geometric mean of citations, *MNCS* mean normalized citation score of the theme, *%1%* percentage of documents in the 1% of most cited papers, *%HCP* percentage of documents in H-Classic, *%F* percentage of women in the theme, *%M* percentage of men in the theme, *RGCR* relative gender contribution rate of the theme

and did not present any documents among the 1% most cited papers and H-Classic. On the contrary, *Caregivers*, *Medline* and *Health* were those themes where the MNCS achieved was exceeded the average. It should be noted that *Health* was the most productive, and its documents were more present in the 1% most cited papers and H-Classic.

On the other hand, men were overrepresented in 13 themes, and most of them can be related to STEM fields, such as *Webometrics*, *Information-systems*, *Software* and *Nanotechnology*. These themes presented the lower rate of women authors among the identified themes, and all of them showed an MNCS higher than the average, with documents among the 1% most cited papers and H-Classic. Other highlighted themes in terms of presence of women were *Impact* and *Innovation*, presenting 0.69 and 0.65 of RGCR, respectively. For one hand, *Impact* presented the highest productivity in the period, was positioned as basic and transversal, and at least 4% of documents in the 1% most cited papers and H-Classic. On the other hand, *Impact* was the third in terms of production, was positioned between basic and transversal, and motor, and was the theme that presented the highest percentage of documents in the 1% most cited papers and H-Classic.

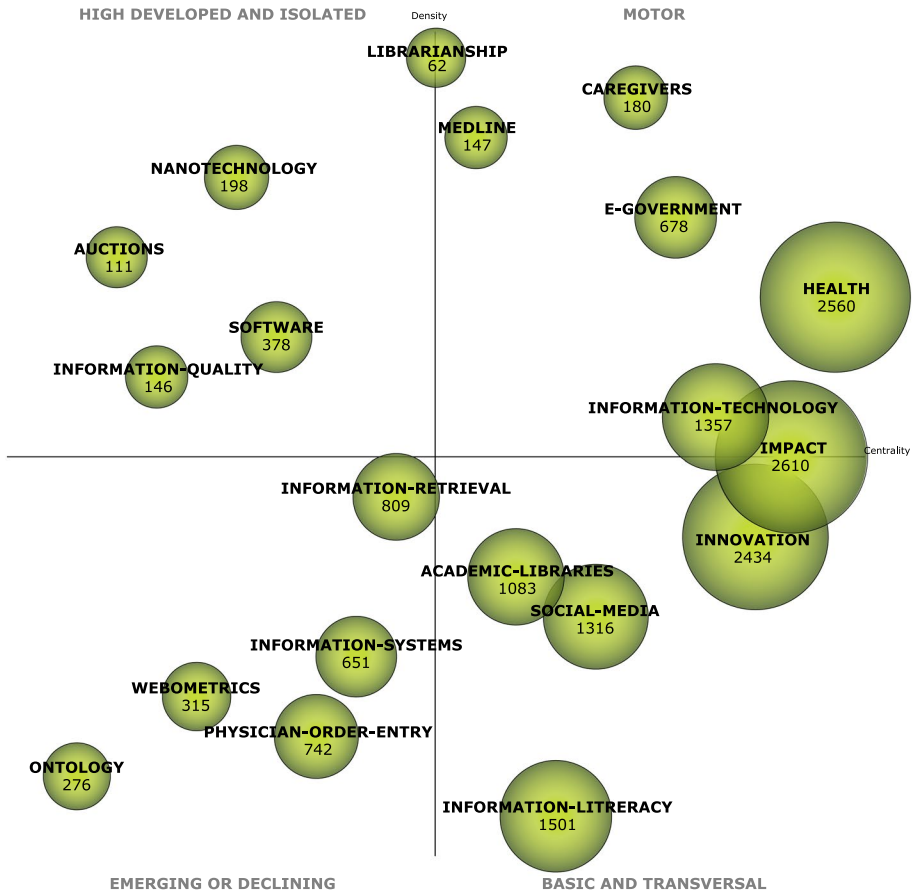


Fig. 4 Strategic diagram from 2011 to 2014, where the *x-axis* and *y-axis* represent the centrality and density, respectively, of each community. Circumference sizes indicate the number of documents in each community, while their position allows classification as 'Motor', 'Basic and Transversal', 'Emerging and Declining', and 'High developed and Isolated'

From 2015 to 2018

The period from 2015 to 2018 encompasses 12,320 documents, with 43.92% of women authors, representing an increase of 2.9% compared to the period from 2011 to 2014. Among the 21 identified themes (Figs. 6 and 7), six of them exhibited an overrepresentation of women (Table 3), 11 of men, and four presented similar rates of gender.

Once again, the six themes overrepresented by women can be related to classical themes of LIS, human, and health-related fields. Taking into account the MNCS, *Information-literacy* and *HIV* were below average, *Document-delivery*, *Physical-activities* and *Race* were on average, and just *Health* showed a 1.19 of MNCS. *Health* and *Physical-activities* were positioned as motor, *Information-literacy* as basic and transversal, and the other three themes as highly developed and isolated. Despite *Physical-activities* and *Race* presenting MNCS around 1, they did not present even 1% of documents in the 1% most cited documents and H-Class. Among the identified themes, *Health* and *Information-literacy* were

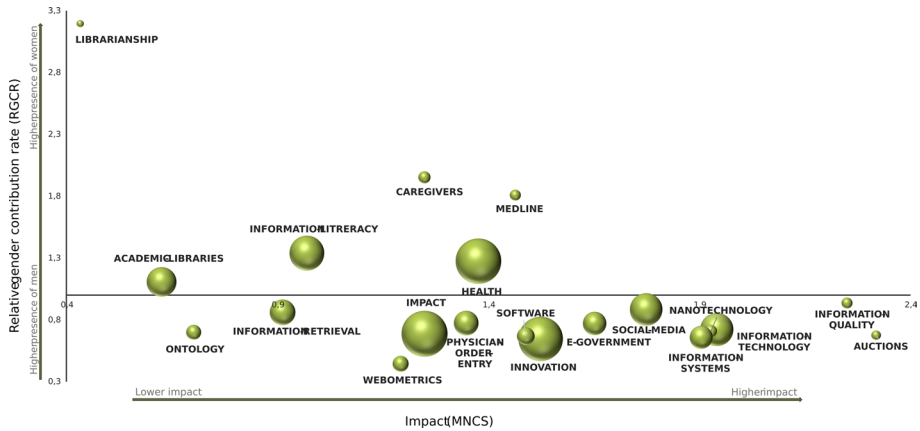


Fig. 5 Distribution of themes covered from 2011 to 2014 based on their impact (MNCS, *x-axis*) and relative gender contribution rate (RGCR, *y-axis*). Circumference sizes indicate the number of documents in each community. The upper quadrant displays the topics where women authors were over-represented, while the lower quadrant shows themes over-represented by men. Themes along the *x-axis* are considered more gender balanced

Table 2 Performance table and impact indicators of LIS field from 2011 to 2014

Theme	Docs	Cit	CGM	MNCS	%1%	%HCP	%F	%M	RGCR
Librarianship	62	670	6.00	0.43	0	0	71%	29%	3.20
Caregivers	180	5484	19.60	1.25	0.27	0.35	59.9%	40.1%	1.95
Medline	147	5309	18.29	1.46	0.41	0.31	58%	42%	1.81
Information-literacy	1501	35805	13.17	0.97	1.08	1.38	50.6%	49.4%	1.34
Health	2560	86724	18.57	1.38	4.87	4.84	49.4%	50.6%	1.27
Academic-libraries	1083	16432	7.91	0.63	0.27	0.53	45.8%	54.2%	1.11
Information-quality	146	8067	27.80	2.25	0.54	0.49	41.7%	58.3%	0.94
Social-media	1316	57099	19.27	1.77	4.87	4.44	40.3%	59.7%	0.88
Information-retrieval	809	18134	11.58	0.91	1.08	0.67	39.7%	60.3%	0.86
Physician-order-entry	742	24581	18.50	1.35	1.35	1.29	37.2%	62.8%	0.77
E-government	678	27702	20.04	1.65	1.62	1.64	37.1%	62.9%	0.77
Information-technology	1357	64811	24.16	1.94	5.28	4.84	35.6%	64.4%	0.72
Nanotechnology	198	9368	22.21	1.93	0.68	0.67	35.1%	64.9%	0.71
Ontology	276	4754	8.84	0.70	0.27	0.22	34.8%	65.2%	0.70
Impact	2610	79968	15.63	1.25	4.19	4.3	34.4%	65.6%	0.69
Auctions	111	6348	29.46	2.32	0.68	0.58	34.1%	65.9%	0.68
Software	378	13989	18.77	1.49	0.95	1.06	33.8%	66.2%	0.67
Information-systems	651	30811	21.81	1.90	2.3	2.13	33.5%	66.5%	0.66
Innovation	2434	91731	18.39	1.52	6.09	6.12	33.1%	66.9%	0.65
Webometrics	315	9043	14.82	1.19	0.41	0.62	25.3%	74.7%	0.44

Docs documents produced in theme, *Cit* citations received in theme, *CGM* geometric mean of citations, *MNCS* mean normalized citation score of the theme, *%1%* percentage of documents in the 1% of most cited papers, *%HCP* percentage of documents in H-Classic, *%F* percentage of women in the theme, *%M* percentage of men in the theme, *RGCR* relative gender contribution rate of the theme

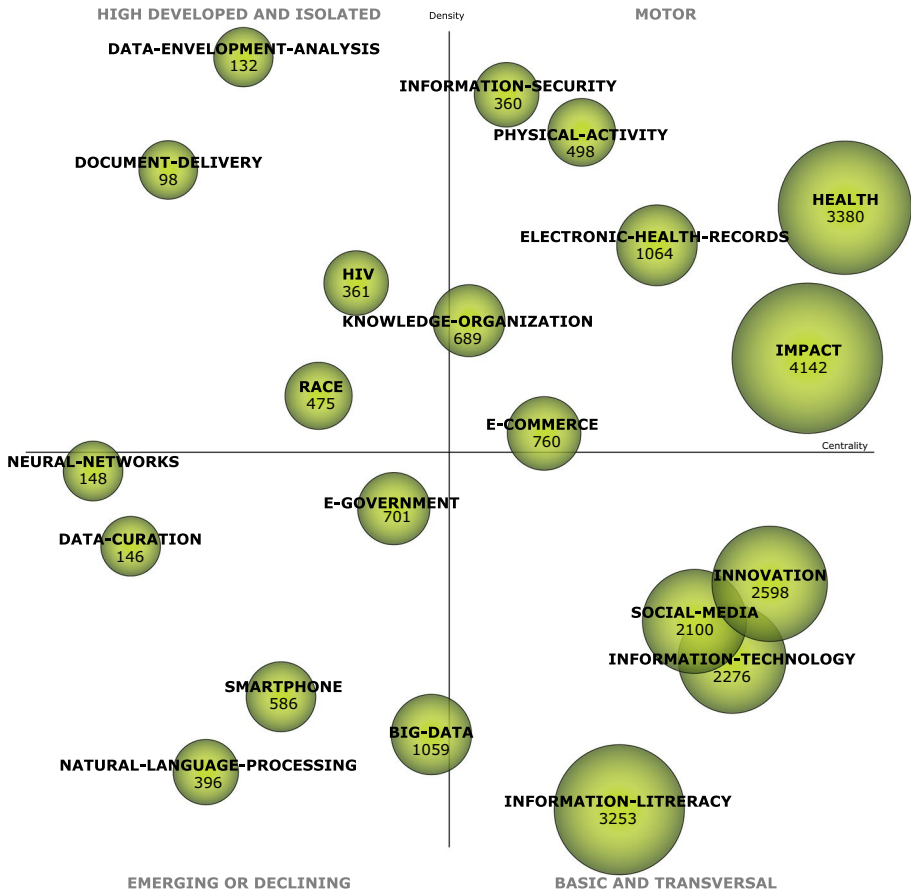


Fig. 6 Strategic diagram from 2015 to 2018, where the *x-axis* and *y-axis* represent the centrality and density, respectively, of each community. Circumference sizes indicate the number of documents in each community, while their position allows classification as 'Motor', 'Basic and Transversal', 'Emerging and Declining', and 'High developed and Isolated'

the most productive, behind only *Impact*, and both themes showed the highest percentage of documents in the 1% of most cited papers and H-Classic.

Among the 11 themes overrepresented by men, we observed that most of them can be related to classical themes of the LIS or STEM fields. In the case of STEM, some themes can be specifically related to computer science, such as *Data-curation*, *Natural-language-processing* and *Neural-networks*. Except *Knowledge organization*, all the themes presented MNCS equal or higher than the average. Among these themes, *Impact*, *Innovation* and *Information-technology* were the most productive and presented the higher percentage of documents in the 1% of most cited papers and H-Classic, with the first one positioned as motor and *Information-technology* and *Innovation* as basic and transversal.

Electronic-health-records, *Big-data*, *Smartphone* and *Social-media* were the themes in which the gender rates were similar. To them, the MNCS were between 1.35 and 1.57, and the percentage of documents in the 1% of most cited papers and H-Classic were the highest to *Big-data* and *Social-media*. The position of *Big-data* in the strategic diagram was

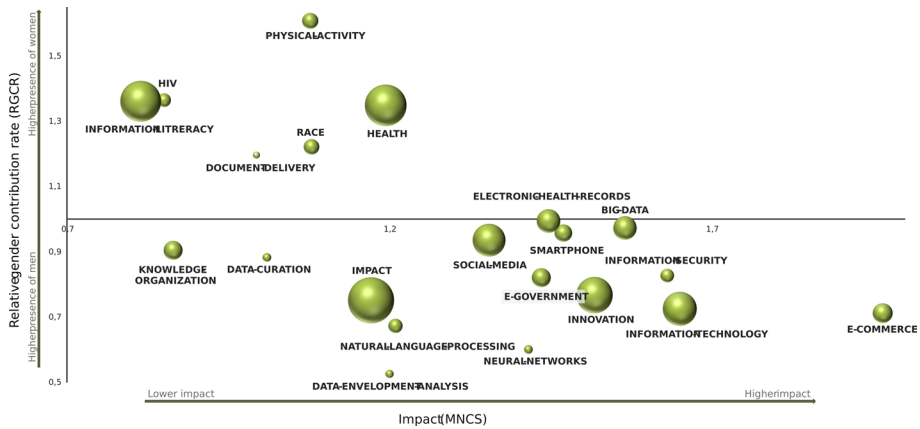


Fig. 7 Distribution of themes covered from 2015 to 2018 based on their impact (MNCS, *x-axis*) and relative gender contribution rate (RGCR, *y-axis*). Circumference sizes indicate the number of documents in each community. The upper quadrant displays the topics where women authors were over-represented, while the lower quadrant shows themes over-represented by men. Themes along the *x-axis* are considered more gender balanced

Table 3 Performance table and impact indicators of LIS field from 2015 to 2018

Theme	Docs	Cit	CGM	MNCS	%1%	%HCP	%F	%M	RGCR
Physical-activity	498	10,422	12.25	1.08	0.54	0.58	55.2%	44.8%	1.61
HIV	361	5914	10.58	0.85	0.14	0.27	51.1%	48.9%	1.36
Information-literacy	3253	51,397	8.25	0.81	3.25	1.95	51.0%	49.0%	1.36
Health	3380	78,452	12.83	1.19	5.01	4.13	50.8%	49.2%	1.35
Race	475	9881	13.50	1.08	0.41	0.4	48.3%	51.7%	1.22
Document-delivery	98	1992	3.73	0.99	0.14	0.04	47.8%	52.2%	1.20
Electronic-health-records	1064	29,852	15.33	1.45	1.76	1.06	43.2%	56.8%	0.99
Big-data	1059	32,005	14.41	1.57	3.52	2.62	42.7%	57.3%	0.97
Smartphone	586	16,571	15.51	1.47	1.35	1.15	42.3%	57.7%	0.96
Social-media	2100	54,954	14.35	1.35	4.6	3.55	41.7%	58.3%	0.94
Knowledge-organization	689	11,274	7.23	0.86	0.54	0.27	40.9%	59.1%	0.90
Data-curation	146	2840	11.25	1.01	0.27	0.18	40.3%	59.7%	0.88
Information-security	360	11,228	16.73	1.63	1.22	0.89	38.8%	61.2%	0.83
E-government	701	19,662	15.03	1.44	2.03	1.29	38.6%	61.4%	0.82
Innovation	2598	76,280	15.22	1.52	7.17	5.19	37.0%	63.0%	0.77
Impact	4142	93,903	11.65	1.17	7.04	4.79	36.5%	63.5%	0.75
Information-technology	2276	72,556	16.08	1.65	7.85	5.37	35.7%	64.3%	0.73
E-commerce	760	28,735	19.11	1.97	2.84	2.04	35.3%	64.7%	0.71
Natural-language-processing	396	9289	13.41	1.21	0.68	0.53	34.0%	66.0%	0.67
Neural-networks	148	3982	14.42	1.42	0.41	0.35	31.5%	68.5%	0.60
Data-envelopment-analysis	132	3167	13.19	1.20	0.27	0.09	28.7%	71.3%	0.53

Docs documents produced in theme, *Cit* citations received in theme, *CGM* geometric mean of citations, *MNCS* mean normalized citation score of the theme, *%1%* percentage of documents in the 1% of most cited papers, *%HCP* percentage of documents in H-Classic, *%F* percentage of women in the theme, *%M* percentage of men in the theme, *RGCR* relative gender contribution rate of the theme

between emerging and basic and transversal, which reveals the increased importance of this theme in the last years. With respect *Social-media*, its position as basic and transversal reveals its current relevance to the LIS field.

From 2019 to 2022

The period from 2019 to 2022 included 12,360 documents, with 43.72% of women authors, a decrease of 0.46% compared with the period from 2015 to 2018. As illustrated in Figs. 8 and 9, 22 themes were identified. Six of these themes were overrepresented by women, 11 by men, and five presented balanced gender ratios (Table 4).

Once again, the six themes overrepresented by women are more related to classical themes of the LIS field, human, and health-related fields, such as *Academic-libraries*, *Migrants* and *Physical-activity*. Only *Migrants* presented an MNCS higher than the

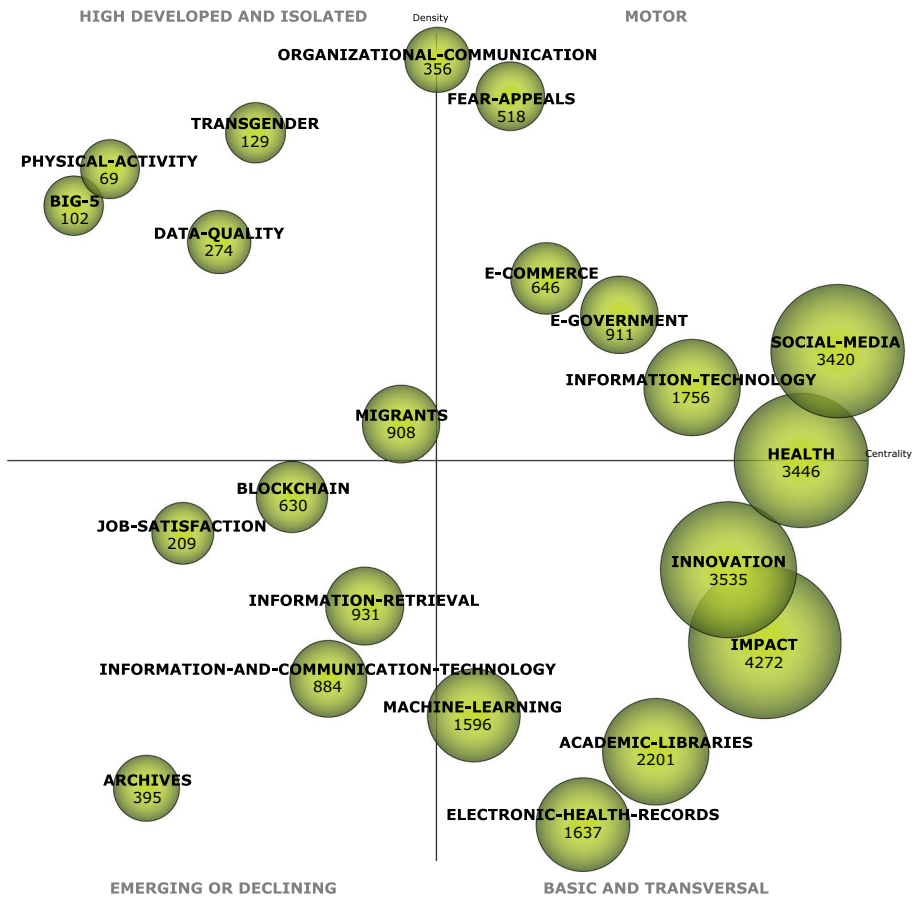


Fig. 8 Strategic diagram from 2019 to 2022, where the *x-axis* and *y-axis* represent the centrality and density, respectively, of each community. Circumference sizes indicate the number of documents in each community, while their position allows classification as 'Motor', 'Basic and Transversal', 'Emerging and Declining', and 'High developed and Isolated'

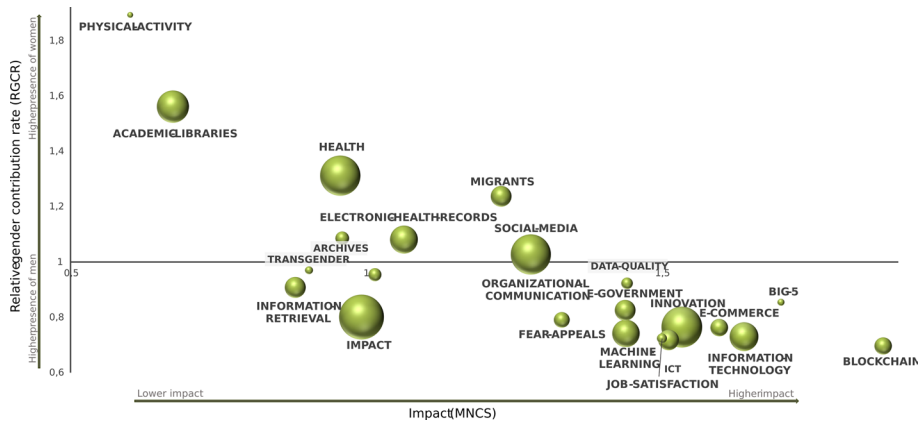


Fig. 9 Distribution of themes covered from 2019 to 2022 based on their impact (MNCS, *x-axis*) and relative gender contribution rate (RGCR, *y-axis*). Circumference sizes indicate the number of documents in each community. The upper quadrant displays the topics where women authors were over-represented, while the lower quadrant shows themes over-represented by men. Themes along the *x-axis* are considered more gender balanced

average, and its percentage of documents in the 1% most cited documents and H-Classic was around 1.5% and 0.67%, respectively. Similar to the periods analyzed previously, *Health* was the second most productive theme, left behind *Impact*. Although the presence of its documents in the 1% most cited documents and H-Classic, its MNCS was around 1, and its position changed to be between motor and basic and transversal.

Among the 11 themes overrepresented by men, we observed a diversity of themes. Some of these themes are related to STEM fields, such as *Blockchain* and *Machine-learning*, and to classical themes of the LIS field, such as *Information-and-communication-technology* and *Impact*. All themes presented MNCS equal or higher than the average, and at least 1% of documents of each theme were in the 1% most cited papers. *Blockchain* achieved 1.87 of MNCS, the highest MNCS of the themes overrepresented by men, although there were 3% of documents in the 1% most cited papers. The most productive themes were *Impact* and *Innovation*, their MNCS were 0.99 and 1.53, respectively, and, once again, both were positioned as basic and transversal.

The five themes where the gender ratio was similar were more diverse and covered themes such as *Information-retrieval*, *Data-quality*, *Organizational-communication*, *Transgender* and *Social-media*. Except for *Social-media*, their position in the strategic diagram was highly developed and isolated. Their MNCS ranged from 0.88 to 1.44, highlighting that the theme with the highest MNCS was *Data-quality*, which can be related to computer science. Regarding the presence of documents in the 1% most cited papers and H-Classic, *Social-media* presented 5.3% and 2%, exceeding the other four themes.

Discussion

Previous research highlighted disparities in the scientific impact achieved by women (Abramo et al., 2015; Chen & Seto 2022; Joanis & Patil, 2022; Kozlowski et al., (2022; Lariviere et al., 2013; Liu et al., 2022; Mauleón & Bordons, 2006). This can be attributed

Table 4 Performance table and impact indicators of LIS field from 2019 to 2022

Theme	Docs	Cit	CGM	MNCS	%1%	%HCP	%F	%M	RGCR
Physical-activity	69	352	3.73	0.60	0	0	59.1%	40.9%	1.89
Academic-libraries	2201	11,197	3.26	0.67	1.08	0.44	54.4%	45.6%	1.56
Health	3446	25,869	4.36	0.96	3.25	1.2	50.1%	49.9%	1.31
Migrants	908	8799	5.02	1.23	1.49	0.67	48.6%	51.4%	1.24
Archives	395	3298	3.58	0.96	0.27	0.18	45.4%	54.6%	1.09
Electronic-health-records	1637	13,972	4.81	1.06	1.62	0.67	45.3%	54.7%	1.08
Social-media	3420	34,393	4.98	1.28	5.95	2.48	44.0%	56.0%	1.03
Transgender	129	921	4.11	0.90	0	0	42.6%	57.4%	0.97
Organizational-communication	356	3005	4.83	1.02	0.54	0.18	42.2%	57.8%	0.95
Data-quality	274	3213	5.79	1.44	0.41	0.13	41.4%	58.6%	0.92
Information-retrieval	931	6775	3.89	0.88	0.81	0.31	41.0%	59.0%	0.91
Big-5	102	1369	6.93	1.70	0.27	0.13	39.5%	60.5%	0.85
E-government	911	9568	5.29	1.44	2.03	0.75	38.7%	61.3%	0.82
Impact	4272	32,620	4.28	0.99	5.28	2.04	38.0%	62.0%	0.80
Fear-appeals	518	5866	5.53	1.33	0.95	0.31	37.7%	62.3%	0.79
Innovation	3535	42,638	5.73	1.53	9.74	3.77	36.9%	63.1%	0.76
E-commerce	646	8005	6.61	1.60	1.49	0.58	36.8%	63.2%	0.76
Machine-learning	1596	16,702	5.32	1.44	3.65	1.46	36.2%	63.8%	0.74
Information-technology	1756	23,155	6.42	1.64	4.6	2	35.8%	64.2%	0.73
Job-satisfaction	209	2562	6.46	1.50	0.41	0.22	35.6%	64.4%	0.72
Information-and-communication-technology	884	10,903	5.43	1.51	2.44	0.89	35.4%	64.6%	0.72
Blockchain	630	9417	6.26	1.87	2.98	1.02	34.7%	65.3%	0.70

Docs documents produced in theme, *Cit* citations received in theme, *CGM* geometric mean of citations, *MNCS* mean normalized citation score of the theme, *%1%* percentage of documents in the 1% of most cited papers, *%HCP* percentage of documents in H-Classic, *%F* percentage of women in the theme, *%M* percentage of men in the theme, *RGCR* relative gender contribution rate of the theme

to several factors, including research productivity, the number of citations received, and the diversity of themes addressed. Exploring the potential factors associated with the differences in impact achieved is a first step towards understanding how and why these dynamics occur. In that sense, we conducted a bibliometric analysis to determine if women and men focused on different themes in the LIS field, which could be one of the aspects that generate this imbalanced impact. To achieve this, we analyzed the evolution over four periods: 2007 to 2010, 2011 to 2014, 2015 to 2018, and 2019 to 2022. Our objective was to examine highlighted themes within the field, assess their impact, and analyze the rate of gender in each theme.

Our findings indicate an increase in the presence of women over the analyzed periods, except for years from 2019 to 2022, and show significant differences among fields and specializations in terms of gender representation and the MNCS achieved for each theme. In general, our observations are in line with the previous results obtained in the LIS field for print and e-journals, showing an increase in the presence of women over the years (Lindsay, 2010; Lund & Shamsi, 2021); Parabhoi et al., 2020; Vinay, 2021). Despite this, these previous studies did not cover the pandemic period, which makes our analysis valuable in

examining LIS in this specific context. Regarding the identified themes and their characteristics, themes with higher RGCR, where women were overrepresented, tend to be classic themes in LIS or related to human and health-related themes. In contrast, themes with low RGCR, where men were overrepresented, are often related to information technology and computer science, which frequently achieve higher MNCS.

A drop in the number of women authors was observed between 2019 and 2022, its decline coincided with the period marked by the COVID-19 pandemic. Earlier studies have revealed that women researchers were disproportionately affected during the pandemic, often due to increased domestic and parental care responsibilities (Lamarre et al., 2020), which may have created cumulative advantages for men researchers (Liu et al., 2022; Reichelt et al., 2021). Furthermore, previous evidence reported that health and medicine-related fields were among the most negatively affected by the pandemic (Squazzoni et al., 2021), an area predominantly occupied by women (Ceci et al., 2014; Cheryan et al., 2017). Our results support this observation, as the theme *Health* was most addressed by women over the analyzed periods and exhibited decreasing MNCS values and impact indicators from 2019 to 2022, suggesting its vulnerability to the pandemic's effect.

Beyond the temporal difference, the results show a gender imbalance in the composition of themes within the LIS field. Specifically, human and health-related themes showed lower impact rates when compared to STEM fields, suggesting that these trends may negatively affect women's scientific performance. Furthermore, despite the increasing number of women in the field over the years, the themes they address remain relatively constant. Additionally, our study highlights that women are also predominant in themes such as *Race* and *Librarianship*. In the case of the first one, the MNCS achieved is lower than the average in the LIS field, and its position in the strategic diagram reflects that it is highly specialized and isolated. Similarly, *Librarianship* is the theme with the highest women contribution rate and presented the lowest MNCS among all identified themes. Both observations reveal a gender imbalance among the identified themes, suggesting that women in the LIS field addressed themes that are not generally associated with high impact rates and can be considered highly specialized.

In contrast, our findings reveal that themes related to STEM fields, such as *Blockchain*, *Neural-network* and *Machine-learning*, were overrepresented by men authors. These themes presented higher MNCS values and were more present in indicators of excellence, such as the 1% of most cited papers and H-Classic. This observation is in line with previous studies, that women authors were underrepresented in STEM fields (Thelwall & Mas-Bleda, 2020), and this disparity can be even more pronounced in specific subfields, such as mathematics, statistics, and computer science (Ceci et al., 2014; Cheryan et al., 2017). For example, *Blockchain* achieved the highest MNCS during the period from 2019 to 2022 and presented one of the lowest rates of women authors, and despite growing awareness of gender-related aspects within this subfield, it remains predominantly men-dominated (Di Vaio et al., 2023).

Furthermore, *Innovation* was another theme overrepresented by men authors from 2007 to 2022. This theme exhibited high MNCS value, the highest percentage of documents in indicators of excellence, and its position on the strategic diagram underscored its high relevance within LIS. Previous innovation-related studies identified that research conducted by diverse, gender-balanced, and larger teams tends to achieve higher citation rates (Hofstra et al., 2020; Larivière et al., 2015) and receive more research grants (Gul et al., 2016). Nevertheless, women researchers were shown to have greater asymmetries between themselves and their co-researchers (Whittington, 2018) and are less inclined to collaborate at an international level (Aksnes et al., 2019; Kwiek & Roszka, 2021). These

disparities in collaboration patterns may be related to the fact that international collaboration rates vary according to the disciplines and the researcher's career stage. For instance, it has been shown that humanities-related fields tend to have lower rates of international collaboration, and senior researchers are likelier to engage in international collaboration (Aksnes et al., 2019; Kwiek & Roszka, 2021). So, to address the gender imbalance in the innovation landscape, some researchers advocate for the adoption of a gendered innovation approach, which involves an awareness of how gender roles, stereotypes, and power dynamics can influence research, design, and innovation (Jenkins et al., 2019; Schiebinger, 2021; Schiebinger & Klinge, 2018).

Considering these results, there is a need to critically examine how power dynamics related to gender roles help to shape the distribution of men and women across different fields. It is essential to clarify that intellectual differences between women and men have no foundation in biological basis (Ceci et al., 2014; De Nicola & D'Agostino, 2021). The understanding of gender imbalances can be approached from two perspectives. One perspective focuses on individual choices and places the problem of inequity on women, attributing the imbalance to their life decisions while dismissing differences between gender roles in actual society. On the other hand, the social and structural perspective considers broader characteristics, such as gender roles and oppression, to explain the gender inequity in STEM (Miner et al., 2018). The construction of gender roles has been studied from various approaches, including the construction and perpetuation of stereotypes within the education system, especially at the primary education level, and through representations in social media. Textbooks and the media often portray white and cisgender men as scientists and white women as teachers (Corsbie-Massay & Wheatly, 2022; Kerkhoven et al., 2016; Mitchell & McKinnon, 2019), creating invisible barriers for women and individuals from diverse linguistic and cultural backgrounds. This perpetuation of stereotypes can influence women's experience within the education system and affect their sense of belonging in specific academic and professional environments (Casad et al., 2021; Rainey et al., 2018; Spoon et al., 2023; Wright et al., 2007). This structural and systematic phenomenon offers at least a partial explanation for why women and men tend to engage in different fields, subfields, and research themes.

The implications of this social phenomenon within the scientific system are vast. Varying citation patterns and the impact of different themes can influence individual performance. Beyond its effects on the scientific community's performance, unequal gender representation can also significantly impact science. Scientific knowledge produced within this imbalanced context often reflects the gendered and racialized identities of scientists (Kozlowski et al., 2022), resulting in a biased perspective on the knowledge that excludes and limits its full potential. Several studies have demonstrated that increasing diversity and collaboration among scientists can positively affect scientific impact and innovation, ultimately expanding the body of knowledge produced (Hofstra et al., 2020; Kozlowski et al., 2022; Larivière et al., 2015; Yang et al., 2022).

Although our study provides valuable insights, some limitations need to be addressed. First, our analysis relied on the Web of Science database, which has been reported to have limited coverage in human-related fields (Mongeon & Paul-Hus, 2016; Singh et al., 2021). Despite this limitation, the corpus used in our research proved sufficient for identifying research themes and revealing gender overrepresentation in various themes. Second, the method used for gender identification is more suited to Western names, as it relies on the author's first name. This approach may not effectively determine the gender of individuals from different cultural backgrounds, such as Russians, for whom gender indicators are typically found at the end of the last name. In future research, we plan to address

these limitations by employing different bibliographical databases and implementing an improved methodology to improve the accuracy of gender assignments. Moreover, given the complexity of addressing gender differences, previous studies have suggested using different metrics to avoid biases in gender analysis (Cameron et al., 2016). Self-citations, for instance, are one of these concerns, and studies have demonstrated this behavior is most observed in men authors, which could inflate their performance and potentially mask the impact of the research on the scientific community (Andersen et al., 2019; Cameron et al., 2016; King et al., 2017; Nielsen, 2016). Another limitation of our analysis was that self-citations were not excluded from the calculus of the performance indicators in each community due to a lack of data. This perspective would allow us to analyze in which communities the impact performance was inflated by self-citations and compare if there were communities mostly represented by one gender. Therefore, while our study provides valuable insights, these limitations underscore the need for continued research to refine methodologies and address biases, ultimately fostering a more comprehensive understanding of gender dynamics in academic research.

Conclusions

This research aimed to shed light on the factors that can affect the scientific impact achieved by women. As detailed throughout this study, women and men presented different citation patterns within the LIS field, and in part, bias may result from the different themes addressed by each group. In this sense, our findings underscore that women and men addressed classical themes of LIS, although men were more involved in themes that achieved higher MNCS values. Meanwhile, women were overrepresented in human and health-related themes, exhibiting low MNCS values and a low percentage of documents in impact indicators, such as the 1% of most cited papers and H-Classic. In contrast, men were overrepresented in STEM-related themes and demonstrated more linkage with innovation themes. Both STEM and innovation-related themes exhibited high MNCS values and a large number of documents within the impact indicators measured. These results highlight the significance of the themes addressed by each gender and their scientific impact, introducing a debate on societal and structural factors that may influence the selection of research fields and themes, as well as the broader implications of gender disparity in science, emphasizing the necessity for more inclusive practices. Our results also reveal a slight increase in women authors in the LIS field over the years, despite the decrease during the COVID-19 pandemic. This decline highlights that the pandemic had a disproportionate impact on women authors. To address the gender imbalance in the dataset, we introduced the RGCR metric to measure and assess gender contributions across different themes. Despite the findings, limitations related to the bibliographical database and the gender identification method were exposed. Nevertheless, this study contributes to understanding the complex relation between research themes, gender contribution within these themes, and their scientific impact achieved over time.

Acknowledgements This extend our contribution presented at the 19th ISSI Conference, 2023 (Choji et al., 2023). This work has been part of the Spanish State Research Agency through the projects PID2019-105381GA-I00/AEI/10.13039/501100011033 (iScience) and C-ING-165-UGR23, co-funded by the Regional Ministry of University, Research and Innovation and by the European Union under the Andalusia ERDF Programme 2021–2027. Also, it is part of the first author's (Choji, Thamyres T.) thesis.

Funding Funding for open access publishing: Universidad de Cádiz/CBUA. Open Access funding is provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, *9*(1), 25–38. <https://doi.org/10.1016/j.joi.2014.11.002>
- Aksnes, D. W., Piro, F. N., & Rørstad, K. (2019). Gender gaps in international research collaboration: A bibliometric approach. *Scientometrics*, *120*(2), 747–774. <https://doi.org/10.1007/s11192-019-03155-3>
- Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *eLife*, *8*, e45374. <https://doi.org/10.7554/eLife.45374>
- Batagelj, V., & Cerinšek, M. (2013). On bibliographic networks. *Scientometrics*, *96*(3), 845–864. <https://doi.org/10.1007/s11192-012-0940-1>
- Bol, T., de Vaan, M., & van de Rijt, A. (2022). Gender-equal funding rates conceal unequal evaluations. *Research Policy*, *51*(1), 104399. <https://doi.org/10.1016/j.respol.2021.104399>
- Bornmann, L. (2017). Measuring impact in research evaluations: A thorough discussion of methods for, effects of and problems with impact measurements. *Higher Education*, *73*(5), 775–787. <https://doi.org/10.1007/s10734-016-9995-x>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, *1*(3), 226–238. <https://doi.org/10.1016/j.joi.2007.03.001>
- Butler, L. (2008). ICT assessment: Moving beyond journal outputs. *Scientometrics*, *74*(1), 39–55. <https://doi.org/10.1007/s11192-008-0102-7>
- Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*(1), 155–205. <https://doi.org/10.1007/BF02019280>
- Cameron, E. Z., White, A. M., & Gray, M. E. (2016). Solving the productivity and impact puzzle: Do men outperform women, or are metrics biased? *BioScience*, *66*(3), 245–252. <https://doi.org/10.1093/biosci/biv173>
- Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., & Petzel, Z. W. (2021). Gender inequality in academia: Problems and solutions for women faculty in STEM. *Journal of Neuroscience Research*, *99*(1), 13–23. <https://doi.org/10.1002/jnr.24631>
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, *15*(3), 75–141. <https://doi.org/10.1177/1529100614541236>
- Chaudhary, A. M. D., Naveed, S., Safdar, B., Saboor, S., Zeshan, M., & Khosa, F. (2021). Gender differences in research project grants and r01 grants at the national institutes of health. *Cureus*, *13*(5), e14930. <https://doi.org/10.7759/cureus.14930>
- Chen, T.-H.K., & Seto, K. C. (2022). Gender and authorship patterns in urban land science. *Journal of Land Use Science*, *17*(1), 245–261. <https://doi.org/10.1080/1747423X.2021.2018515>
- 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>

- Choji, T. T., Cobo, M. J., Velez-Estevez, A., & Moral-Munoz, J. A. (2023). Thematic differences by gender in the field of Library and Information Science from 2017 to 2021. Paper presented at the 19th in International Conference of the International Society for Scientometrics and Informetrics (ISSI2023) (University of Bloomington, Indiana, USA 2–5 July 2023).
- Cobo, M. J., López-Herrera, A., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Cobo, M. J., López-Herrera, A., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63(8), 1609–1630. <https://doi.org/10.1002/asi.22688>
- Corsbie-Massay, C. L., & Wheatly, M. G. (2022). The role of media professionals in perpetuating and disrupting stereotypes of women in science, technology, engineering and math (STEM) fields. *Frontiers in Communication*, 7, 1027502. <https://doi.org/10.3389/fcomm.2022.1027502>
- Darren, S., & Steven, A. S. (2010). Does gender and race have an impact on earnings in the library and information science labor market in the United States of America? *Library and Information Science Research E-Journal*, 20, 1–21. <https://doi.org/10.32655/LIBRES.2010.2.1>
- De Nicola, A., & D'Agostino, G. (2021). Assessment of gender divide in scientific communities. *Scientometrics*, 126(5), 3807–3840. <https://doi.org/10.1007/s11192-021-03885-3>
- Di Vaio, A., Hassan, R., & Palladino, R. (2023). Blockchain technology and gender equality: A systematic literature review. *International Journal of Information Management*, 68, 102517. <https://doi.org/10.1016/j.ijinfomgt.2022.102517>
- Dorta-González, P., & Dorta-González, M. I. (2013). Comparing journals from different fields of science and social science through a JCR subject categories normalized impact factor. *Scientometrics*, 95(2), 645–672. <https://doi.org/10.1007/s11192-012-0929-9>
- Fagan, C., & Teasdale, N. (2021). Women professors across STEM and non-STEM disciplines: Navigating gendered spaces and playing the academic game. *Work, Employment and Society*, 35(4), 774–792. <https://doi.org/10.1177/0950017020916182>
- Garfield, E. (1990). Essays of an information scientist: Journalology, keywords plus, and other essays. *Current Contents*, 13(32), 3–7.
- Goldin, C., & Katz, L. F. (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 110(4), 730–770. <https://doi.org/10.1086/340778>
- Gul, S., Shah, T. A., Hamade, S. N., Mushtaq, R., & Koul, I. (2016). Effects of gender in library and information science research: A case study of The Electronic Library. *The Electronic Library*, 34(3), 488–503. <https://doi.org/10.1108/EL-08-2014-0126>
- Harris, R. M., Gillian Mitchel, B., & Cooley, C. (1985). The gender gap in library education. *Association for Library and Information Science Education*, 25(3), 167–176. <https://doi.org/10.3138/jelis.56.1.75>
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17), 9284–9291. <https://doi.org/10.1073/pnas.1915378117>
- Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology*, 16(4), e2004956. <https://doi.org/10.1371/journal.pbio.2004956>
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (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>
- Jenkins, F., Keane, H., & Donovan, C. (2019). Work (still) in progress: Establishing the value of gendered innovations in the social sciences. *Women's Studies International Forum*, 72, 104–108. <https://doi.org/10.1016/j.wsif.2018.04.006>
- Joanis, S. T., & Patil, V. H. (2022). First-author gender differentials in business journal publishing: top journals versus the rest. *Scientometrics*, 127(2), 733–761. <https://doi.org/10.1007/s11192-021-04235-z>
- Kerkhoven, A. H., Russo, P., Land-Zandstra, A. M., Saxena, A., & Rodenburg, F. J. (2016). Gender stereotypes in science education resources: A visual content analysis. *PLoS ONE*, 11(11), e0165037. <https://doi.org/10.1371/journal.pone.0165037>
- King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men set their own cites high: Gender and self-citation across fields and over time. *Socius Sociological Research for a Dynamic World*. <https://doi.org/10.1177/2378023117738903>

- Kozłowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences*, 119(2), e2113067119. <https://doi.org/10.1073/pnas.2113067119>
- Kwiek, M., & Roszka, W. (2021). Gender disparities in international research collaboration: A study of 25,000 university professors. *Journal of Economic Surveys*, 35(5), 1344–1380. <https://doi.org/10.1111/joes.12395>
- Lamarre, P.V., Sugimoto, C.R., Lariviere, V. (2020). The decline of women’s research production during the coronavirus pandemic.
- Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323–1332. <https://doi.org/10.1002/asi.23266>
- Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. (2013). Global gender disparities in science. *Nature*, 504, 211–213. <https://doi.org/10.1038/504211a>
- Leaderboard. (2022). Database of the popularity of UK and US names. "<https://github.com/Leaderboard/NamesDatabase>".
- Linsay, R.-E. (2010). Gender and citation in two LIS e-journals : A bibliometric analysis of LIBRES and information research. *Library and Information Science Research E-Journal*, 20(1), 1–18. <https://doi.org/10.32655/LIBRES.2010.1.2>
- Liu, J., Song, Y., & Yang, S. (2020). Gender disparities in the field of economics. *Scientometrics*, 125(2), 1477–1498. <https://doi.org/10.1007/s11192-020-03627-x>
- Liu, M., Zhang, N., Hu, X., Jaiswal, A., Xu, J., Chen, H., & Bu, Y. (2022). Further divided gender gaps in research productivity and collaboration during the COVID-19 pandemic: Evidence from coronavirus-related literature. *Journal of Informetrics*, 16(2), 101295. <https://doi.org/10.1016/j.joi.2022.101295>
- Lund, B., & Shamsi, A. (2021). Women authorship in library and information science journals from 1981 to 2020: Is equitable representation being attained? *Journal of Information Science*, 49(5), 1335–1343. <https://doi.org/10.1177/01655515211050026>
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O’Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79(3), 1290–1326. <https://doi.org/10.3102/0034654309334143>
- Martínez, M. A., Herrera, M., López-Gijón, J., & Herrera-Viedma, E. (2014). H-classes: Characterizing the concept of citation classics through h-index. *Scientometrics*, 98(3), 1971–1983. <https://doi.org/10.1007/s11192-013-1155-9>
- Mauléon, E., & Bordons, M. (2006). Productivity, impact and publication habits by gender in the area of materials science. *Scientometrics*, 66(1), 199–218. <https://doi.org/10.1007/s11192-006-0014-3>
- Miner, K. N., Walker, J. M., Bergman, M. E., Jean, V. A., Carter-Sowell, A., January, S. C., & Kaunas, C. (2018). From her problem to our problem: Using an individual lens versus a social-structural lens to understand gender inequity in stem. *Industrial and Organizational Psychology*, 11(2), 267–290. <https://doi.org/10.1017/iop.2018.7>
- Mitchell, M., & McKinnon, M. (2019). Human or objective faces of science? Gender stereotypes and the representation of scientists in the media. *Public Understanding of Science*, 28(2), 177–190. <https://doi.org/10.1177/0963662518801257>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of science and scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Moral-Munoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, 29(1), e290103. <https://doi.org/10.3145/epi.2020.ene.03>
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479. <https://doi.org/10.1073/pnas.1211286109>
- Nielsen, M. W. (2016). Gender inequality and research performance: Moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education*, 41(11), 2044–2060. <https://doi.org/10.1080/03075079.2015.1007945>
- Nobel Prize Outreach AB. (2023). Press release. NobelPrize.org. Retrieved from <https://www.nobelprize.org/prizes/economic-sciences/2023/press-release/>.
- Nygaard, L. P., & Bahgat, K. (2018). What’s in a number? How (and why) measuring research productivity in different ways changes the gender gap. *Journal of English for Academic Purposes*, 32, 67–79. <https://doi.org/10.1016/j.jeap.2018.03.009>
- Organisation for Economic Co-operation and Development (OECD). (2022). Education at a glance 2022: OECD indicators. OECD.

- Parabhoi, L., Sahu, R.R., Dewey, R.S., Parabhoi, D. (2020). Gender Disparity Among Indian Library and Information Science Professionals: a 20-year sample of publications from 1999–2018. *Library Philosophy and Practice—Electronic Journal*, 4382. Retrieved from <https://digitalcommons.unl.edu/libphilprac/4386>.
- Petrongolo, B. (2019). The gender gap in employment and wages. *Nature Human Behaviour*, 3(4), 316–318. <https://doi.org/10.1038/s41562-019-0558-x>
- Piper, P. S., & Collamer, B. E. (2001). Male librarians: Men in a feminized profession. *The Journal of Academic Librarianship*, 27(5), 406–411. [https://doi.org/10.1016/S0099-1333\(01\)00226-9](https://doi.org/10.1016/S0099-1333(01)00226-9)
- Pérez, I. S. (2016). Gender-guesser: Get the gender from first name. <https://github.com/lead-ratings/gender-guesser>
- Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how sense of belonging influences decisions to major in STEM. *International Journal of STEM Education*, 5(1), 10. <https://doi.org/10.1186/s40594-018-0115-6>
- Reichelt, M., Makovi, K., & Sargsyan, A. (2021). The impact of COVID-19 on gender inequality in the labor market and gender-role attitudes. *European Societies*, 23, S228–S245. <https://doi.org/10.1080/14616696.2020.1823010>
- Schiebinger, L. (2021). Gendered innovations: Integrating sex, gender, and intersectional analysis into science, health & medicine, engineering, and environment. *Tapuya: Latin American Science, Technology and Society*, 4(1), 1867420. <https://doi.org/10.1080/25729861.2020.1867420>
- Schiebinger, L., & Klinge, I. (2018). Gendered innovation in health and medicine. *Advances in experimental medicine and biology*, 1065, 643–654. https://doi.org/10.1007/978-3-319-77932-4_39
- Shen, H. (2013). Inequality quantified: Mind the gender gap. *Nature*, 495(7439), 22–24. <https://doi.org/10.1038/495022a>
- 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>
- Social Security Administration. (2022). Social security’s public data listing. Retrieved from “<https://www.ssa.gov/data.json>”.
- Spoon, K., LaBerge, N., Wapman, K. H., Zhang, S., Morgan, A. C., Galesic, M., Fosdick, B. K., Larremore, D. B., & Clauset, A. (2023). Gender and retention patterns among U.S. faculty. *Science Advances*, 9(42), eadi2205. <https://doi.org/10.1126/sciadv.adi2205>
- Squazzoni, F., Bravo, G., Grimaldo, F., García-Costa, D., Farjam, M., & Mehmani, B. (2021). Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals. *PLoS One*, 16(10), e0257919. <https://doi.org/10.1371/journal.pone.0257919>
- Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, 6, 189. <https://doi.org/10.3389/fpsyg.2015.00189>
- Thelwall, M., Bailey, C., Tobin, C., & Bradshaw, N.-A. (2019). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, 13(1), 149–169. <https://doi.org/10.1016/j.joi.2018.12.002>
- Thelwall, M., Kousha, K., Stuart, E., Makita, M., Abdoli, M., Wilson, P., & Levitt, J. (2023). Do bibliometrics introduce gender, institutional or interdisciplinary biases into research evaluations? *Research Policy*, 52(8), 104829. <https://doi.org/10.1016/j.respol.2023.104829>
- Thelwall, M., & Mas-Bleda, A. (2020). A gender equality paradox in academic publishing: Countries with a higher proportion of female first-authored journal articles have larger first-author gender disparities between fields. *Quantitative Science Studies*, 1(3), 1260–1282. https://doi.org/10.1162/qss_a_00050
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1), 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- UNESCO Institute for Statistics (UIS). (2019). Women in Science (Fact Sheet No. 55)
- Varlejs, J., & Dalrymple, P. (1986). Publication output of library and information science faculty. *Journal of Education for Library and Information Science*, 27(2), 71–89. <https://doi.org/10.2307/40323511>
- Velez-Estevez, A., Perez, I. J., García-Sánchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, 60(4), 103385. <https://doi.org/10.1016/j.ipm.2023.103385>
- Vinay, R. S. (2021). LIS research productivity according to gender. *Library Philosophy and Practice*, 5024, 4431.
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47. <https://doi.org/10.1016/j.joi.2010.08.001>
- Whittington, K. B. (2018). A tie is a tie? Gender and network positioning in life science inventor collaboration. *Research Policy*, 47(2), 511–526. <https://doi.org/10.1016/j.respol.2017.12.006>

- Wright, C., Thompson, S., & Channer, Y. (2007). Out of place: Black women academics in british universities. *Women's History Review*, 16(2), 145–162. <https://doi.org/10.1080/09612020601048704>
- Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F., & Uzzi, B. (2022). Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences*, 119(36), e2200841119. <https://doi.org/10.1073/pnas.2200841119>
- Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, 714, 1–73. <https://doi.org/10.1016/j.physrep.2017.10.001>
- Zhang, L., Sivertsen, G., Du, H., Huang, Y., & Glänzel, W. (2021). Gender differences in the aims and impacts of research. *Scientometrics*, 126(11), 8861–8886. <https://doi.org/10.1007/s11192-021-04171-y>

Authors and Affiliations

Thamyres T. Choji^{1,2}  · Manuel J. Cobo^{3,4} · Jose A. Moral-Munoz^{1,5}

✉ Thamyres T. Choji
choji@uca.es

Manuel J. Cobo
mjcobo@decsai.ugr.es

Jose A. Moral-Munoz
joseantonio.moral@uca.es

- ¹ Department of Nursing and Physiotherapy, Faculty of Nursing and Physiotherapy, University of Cadiz, 11009 Cádiz, Spain
- ² Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain
- ³ Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain
- ⁴ Unit for Computational Humanities and Social Sciences (U-CHASS), University of Granada, 18071 Granada, Spain
- ⁵ Biomedical Research and Innovation Institute of Cadiz (INiBICA), University of Cadiz, 11002 Cádiz, Spain