Trends in Women Studies, 2011-2020: A Computational Text Analysis

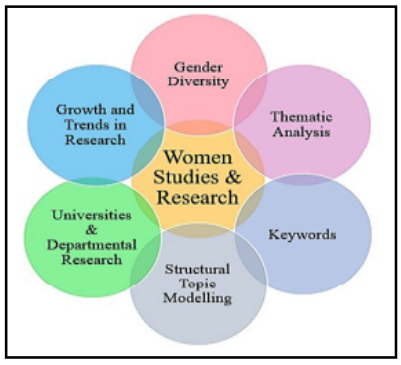
Suman Das[1], Manoj Kumar K[2], Anup Kumar Das[3] and Manorama Tripathi[4]

[1]Jayaprakash Narayan Centre of Excellence in The Humanities, IIT Indore
[2]INFLIBNET, Gandhinagar, Gujarat, India
[3]Centre for Studies in Science Policy, Jawaharlal Nehru University, New Delhi, India
[4]Dr. B. R. Ambedkar Central Library, Jawaharlal Nehru University, New Delhi, India

Abstract

The interdisciplinary field of Women's Studies is manifested in diverse forms within the educational and research landscape of Indian universities and institutions. By scrutinizing trends and patterns in research, particularly focused in PhD theses, a comprehensive understanding of this field's evolution can be gleaned. In India, the discipline of Women's Studies is in a constant state of transformation, continually adapting its dimensions.

In this study, we present our findings derived from an examination of the core theses produced over the last decade (2011-2020), exploring various facets of Women's Studies across multiple departments. Leveraging the Shodhganga digital theses repository, we meticulously selected 1389 theses as the basis of our analysis. The analysis reveals that the southern region of India has exhibited the highest publication output of these theses. To gain deeper insights into the breadth of research topics addressed by scholars, we employed computational text analysis techniques, specifically employing two prominent methods: Structural Topic Modelling (STM) and Latent Dirichlet Allocation (LDA). These methodologies enabled us to unravel the hierarchical arrangement of topics and the co-occurrence patterns within the documents.



Visual Abstract

In conclusion, our study underscores the significance and utility of Electronic Theses and Dissertations (ETD) repositories, particularly in the context of Women's Studies in India. These repositories not only facilitate comprehensive analysis but also provide a valuable resource for researchers and scholars aiming to comprehend the trajectory.

Keywords: Electronic Theses and Dissertations (ETD), ETDs on Women, Latent Dirichlet Allocation (LDA), Structural Topic Modelling (STM), Text Analysis, Women Studies

---

Corresponding Author: Dr. Manorama Tripathi, Email: manoramatripathi2@yahoo.com

1. Introduction

The inception of Women's Studies within the Indian academic sphere occurred during the 1970s. Gradually but persistently, this realm of study attained institutionalization. At present, the landscape boasts more than 150 Centres' exclusively dedicated to researching aspects related to gender, femininity, women's development, and the disparities that persist. As elucidated by Dutoya (2022), Women's Studies has witnessed a steady progression from the 1990s onwards, evolving into a discipline of substance and significance. This evolution has encompassed the introduction of numerous academic programs, including PhD-level ones (Dutoya, 2022; Wöhrer, 2016).

Researchers have diligently explored the multifaceted domain of women's studies from a myriad of perspectives. This exploration has spanned areas such as empowerment, societal development, inequality, identity, human rights violations, as well as the experiences of distinct demographic groups, including tribal and Muslim women, rural laborers, and inhabitants of urban slums.

Over the preceding decade, a discernible shift has transpired within women's studies, wherein certain pivotal themes have gained paramount importance. Women's empowerment, entrepreneurship, achieving equilibrium between work and personal life, and securing a robust social standing have emerged as focal points central to the discipline. Simultaneously, researchers have evinced a burgeoning interest in delving into previously less explored arenas, encompassing Dalit studies, issues of sexual harassment, breast cancer, as well as matters pertaining to food and nutrition.

Mushtaq et al.'s scholarly endeavor directs its focus towards gender studies and inequalities, underscoring the escalating significance of women's empowerment in propelling this field forward. Our own research augments the existing body of Indian studies by meticulously scrutinizing pivotal topics extracted from abstracts of PhD theses. In doing so, we illuminate an array of valuable insights into the diverse themes and prevailing trends within this dynamic and evolving discipline.

2. Related work

Universities worldwide engage in academic research across their departments and laboratories, with the primary goal of contributing novel knowledge. This cycle of generating, sharing, and safeguarding knowledge assumes a critical role in propelling socio-economic advancement and achieving exceptional performance benchmarks. As nations aspire to evolve into knowledge-based economies, they should accord paramount importance to creating, disseminating, and protecting knowledge. The research endeavors of a country can serve as a catalyst for progress across the three pillars of sustainable development—environmental, economic, and social (Sisa et al., 2020). In this journey, universities emerge as pivotal agents, facilitating the evolution of a global knowledge-driven economy.

During the span of 2015 to 2020, the landscape of social science research has witnessed notable enhancements, as evidenced by a remarkable surge of 140% in research publications, escalating from 10,500

in 2015 to an impressive 25,811 by October 2020 (Nanda, 2021). The accessibility of research findings can bestow significant advantages on all stakeholders, including policy makers and practitioners, empowering them to make well-informed decisions. This becomes particularly vital in tackling pressing global challenges such as health crises, climate fluctuations, and food insufficiency. Furthermore, the provision of equitable access to research outcomes, regardless of researchers' financial capabilities or geographical location, holds substantial merit.

In summation, the adoption of an open access approach for disseminating research findings bears the potential to accelerate the trajectory of scientific advancement, foster collaboration, and amplify the accessibility of information on a fair and inclusive basis.

2.1 Text Analysis

In 1949, Roberto Busa collaborated with IBM to develop an Index Variorum, automating the conversion of 10,000 handwritten texts of St. Thomas Aquinas' works (Ries, 2012). The growing accessibility and diversity of text analysis tools, enabled by text mining, have had substantial positive impacts on various fields. Consequently, Busa is rightfully recognized as a trailblazer in the realm of humanities computing.

IBM computer scientist Hans Peter Luhn introduced the concept of utilizing automated text analysis for business intelligence. He harnessed textual data, descriptive metadata, and summary information to enhance business insights (Goyal et al., 2018). The prominence of unstructured textual data has led to integrating textual analysis into data-driven business intelligence.

In the early 21st century, computational linguistics experienced remarkable growth, evolving into a multidisciplinary field encompassing various computational and cognitive approaches. These included machine learning (ML), deep learning, artificial intelligence (AI), computational cognition, and human-computer interaction (HCI). A notable recent achievement has been the development of large language models (LLMs) like ChatGPT, an automated chatbot. The field of text mining has continued to ebb and flow in relevance alongside other contemporary trends within the technology industry. Within this context, our present study is focused on women's studies literature, aiming to uncover latent patterns within the text.

2.2 Text clustering

In general, clustering is a technique to separate objects into different classes. Text documents can be clustered in various levels of granularity by cluster objects such as documents, paragraphs, and sentences. Typically, two types of clustering algorithms are used in computational techniques: supervised and unsupervised. Document clustering finds wide-ranging applications, such as enabling collection browsing, facilitating corpus summarization, and assisting in document classifications.

Clustering and classification are closely related but differ in their learning approaches. Clustering is an unsupervised learning method, while classification uses supervised learning. In classification, predefined classes exist in the dataset, and the goal is to assign a new dataset to a specific class. Text classification

employs machine learning techniques such as decision trees, rule-based classifiers, Support Vector Machine (SVM), and Bayesian classifiers (Sharma et al., 2021). It has various real-life applications such as email classification, spam filtering, news filtering, document organization etc.

LDA, STM, Embedded Topic Model (ETM) and Deep latent Dirichlet allocation (DLDA) are popular probabilistic models in the humanities and social sciences. At the same time, Top2Vec and BERTopics represent documents and their semantic relationships. The topic vector can be jointly embedded with the document, representing semantic similarity. The BERT model leverages sentence structure and class-based term frequency-inverse document frequency (TF-IDF)(Sharma et al., 2021a).

In this paper, we focus on topic models; we analyze STM (Hierarchical structure of the text) and LDA (interactive topic selection of LDAvis) to understand dense clusters allowing for interpretable topics and important words in the topic descriptions.

2.3 Topic Modelling

Topic modelling is a statistical method used to extract meaningful themes from a collection of text documents; it identifies latent topics by analyzing and identifying word co-occurrence patterns, providing valuable insights into the underlying structure and content of the text corpus. It is widely used to analyze large text collections. Computing power is harnessed to propel advancements in text modelling through topic modelling (Park & Han, 2023). It offers an answer to the fundamental area of content analysis by Lasswell, and Larner while finding a useful analytical approach for the PhD theses analysis.LDA and STM models are novel approaches that fall under Bayesian modelling (Rose et al., 2022). They operate on the assumption that in each document, the distribution of topics over the words is a blend of overarching topics present throughout the entire corpus. According to Blei (2012), LDA to model (k) topics across a corpus using algorithms to infer the structure of text and output two critical matrices, per document topic proportion (q) and topic-word distribution (b). STM incorporates metadata, that can better understand the context in which topics emerge, whereas LDA generates topics solely based on word co-occurrence patterns. We studied both the metadata and content of large corpus.

Traditional content analysis research involves predefining codes for analysis, while latent topic models require researchers to specify an initial number of topics (k). Iterative processes prioritize interpretability and validation over relying solely on statistical tests. The computational approach can produce conflicting results, but researchers must balance semantic coherence and topic exclusivity.

We employ STM and LDA models to analyze PhD theses abstract for textual analysis in the women's studies area. Although the usage of this method in literature content analysis of Ph.D. has been limited, we introduced the basic idea of STM and compared it with the findings of LDA.

The STM model provides us with a deeper understanding of the structural variables within a document and allows us to explore the hierarchical organization of topics (Chen et al., 2020). The STM model can be

effectively utilized for extracting and comprehending document metadata, including publication dates, universities, and the respective department from which it was published. STM model also provides correlated topics that can be correlated with one another. In particular, LDA shared prior Dirichlet parameter (content) parameter a (h), while STM generated document-specific covariates X(Y) (N. Hu et al., 2019). Simply put, the characteristics of STM make it a better fit for research in many applied social science disciplines than LDA.

3. Objectives

India's higher education system comprises 56 universities with central funding, 318 universities funded by the States, 185 privately funded state universities, 129 deemed-to-be universities, and 51 institutions of national importance. Additionally, there are more than 37,000 colleges affiliated with various universities in India. Notably, Shodhganga, a repository for research theses, receives submissions from 729 universities, several of which focus on women's studies.

Research in the field of women's studies holds significant importance as it aids in comprehending the socio-economic challenges faced by women. This understanding is essential for the formulation of effective policies and models across various sectors. Numerous theses are being submitted on women-centric subjects, including gender diversity, underscoring the significance of this area of research.

The objectives of the study are as follows:

- To investigate the main topics of Women's Studies subject area studied in the theses submitted by the Indian researchers during 2011-2020 and available in the Shodhganga repository.
- To highlight the gender diversity among researchers and supervisors in Women's Studies.
- To determine the Indian universities and their specific departments or centers that exhibit higher research output and the context surrounding this achievement.
- To identify hierarchical structure of research issues in the theses on women studies.

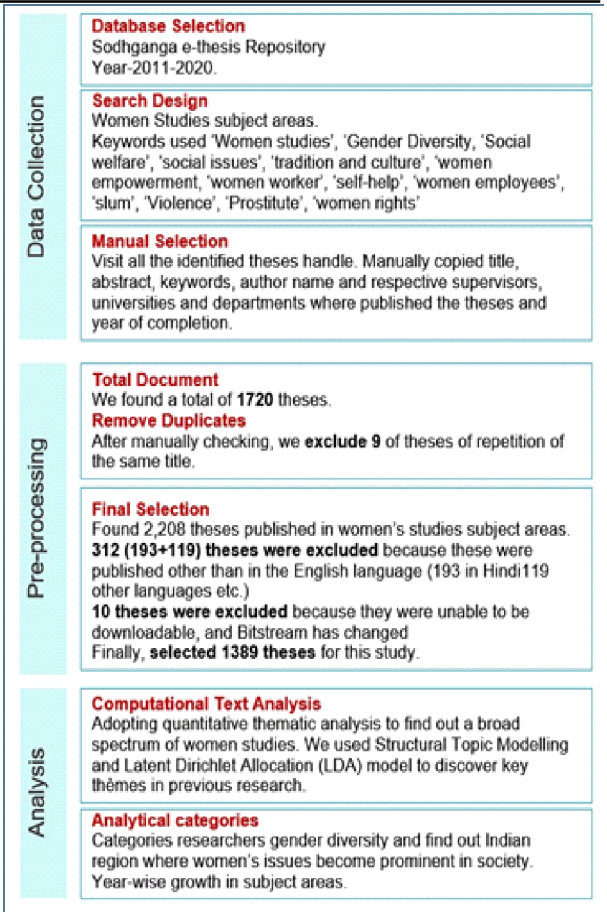
4. Data and Methods

The data used for analysis was taken from Shodhganga. Shodhganga (http://shodhganga.inflibnet.ac.in) is a comprehensive repository (Mehta et al., 2022) housing theses submitted to various Indian universities. This platform captures, indexes, stores, disseminates and preserves Electronic Theses and Dissertations (ETDs) submitted by Indian universities. As of May 2023, the repository boasts an extensive collection, featuring over 479,241 theses, 10,856+ synopses, and 69 PDFs and Fellowship reports. These contributions stem from 720+ universities scattered across the nation.

Within the specified time frame (2011-2020), there were 1720 theses specifically pertaining to women's studies across central, state, private universities, and other institutes. To narrow down our focus, we employed specific terms such as 'Women Studies', 'Gender Diversity', 'Social Welfare', 'Social Issues',

'Tradition and Culture', 'Women Empowerment', 'Women Worker', 'Self-help', 'Women Employees', 'slum', 'Violence', 'Prostitute', and 'Women Rights' to conduct our search for relevant topics.

For our analysis, we considered the abstracts (1988) of the selected theses. We also included details such as authors, supervisors, universities, and department information. To determine the gender of authors and supervisors, we initially manually categorized them as male, or female based on their first names. To validate this classification, we employed the GenderAPI software (https://gender-api.com/), which has been affirmed for its reliability by Paswan and Singh (2020). Our verification using two random samples of 100 instances each demonstrated an accuracy of 97% or higher. It is important to note that gender determination was not feasible in cases where the full name was not available.



To discern the predominant themes within women's studies over the past decade, we have employed a computational text analysis model known as Structural Topic Modelling (STM). This approach is highly effective for scrutinizing textual content (X. Han, 2020). For data preprocessing, we have utilized the Natural Language Toolkit (NLTK) package, which aids in tokenization and the transformation of text into term matrices (accessible at https://www.nltk.org/nltk_data/). Employing the RStudio environment, we have conducted the Latent Dirichlet Allocation (LDA) model (accessible at https://cran.r-project.org/web/packages/lda/index.html).

Subsequently, we have retrieved per-topic-per-word computational models from the generated LDA model using the 'tidytext' package (available at https://cran.r-project.org/web/packages/tidytext/index.html).

It is important to note that we have excluded 55 theses from our analysis due to their being written in languages other than English, such as Hindi, Gujarati, Tamil, Telugu, Bengali, and others. Additionally, we excluded 10 more theses as their PDF bitstreams for the abstracts chapters had been altered, rendering them unreadable.

5. Results

5.1 Gender Diversity and year-wise Growth

In figure 1, we found that the Indian universities awarded 1389 theses on women studies over the ten years 2011-2020. The research interest has increased significantly in this subject area since 2016. The number of theses awarded increased year by year. As per the World Bank Country Report 2021 (https://www.worldbank.org/en/country/india/overview#3), they have financially supported Education, social protection, health and nutrition, rural water supply and sanitation domains over the last five years. Nevertheless, still, there is a massive gap in investment in human capital (childhood development, education, health, social protection and sanitation, etc.), competitiveness and enabling job creation (business climate, skilling, female labour, etc.) in the country.

In the gender diversity ratio, we found that women researchers lead the research field followed by male researchers (19%) and 4% is unclassified; choose socio-economic and cultural problems in their research, followed by 28.73% of male researchers. But in supervision guidance where male faculties (54%) dominate; 4% author and 6% supervisor are unclassified due to their short names. Over 103 theses were supervised by multiple advisors. India is currently in a stage of development where female scholars are underrepresented in senior leadership roles, with males predominantly occupying leadership positions.

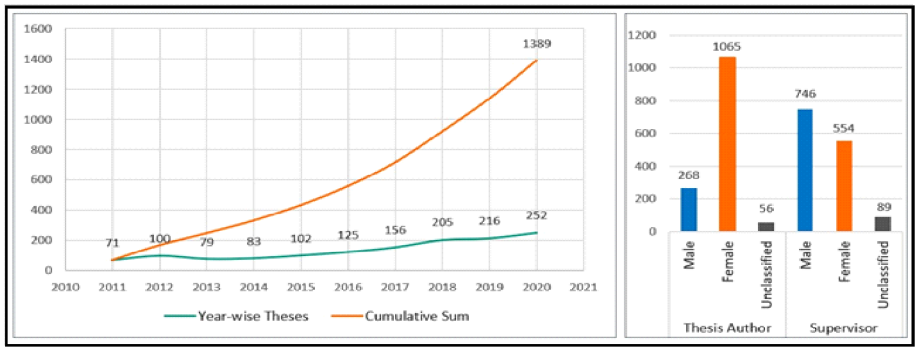


Figure 1: Gender categorization over ten years of published PhD theses on women studies subject

5.2 Leading Universities and department in Women Studies

Table 1: Contributions of theses by universities

Types of University	No of Theses Published
State University	902
Central University	265
Deemed University	143
Private University	79
Total	**1389**

India stands as the world's largest democracy, accommodating a population of over 1.4 billion people as of 2023, according to Worldometer data. Within the landscape of Indian Higher Education Institutions (HEIs), distinct gendered organizational structures persist, perpetuating an entrenched glass ceiling.

Nonetheless, strides are being made to narrow the gender gap in the corporate sphere through diverse initiatives. As observed by Karuna Chanana (2020), the realm of HEIs seems relatively untouched by these changes, particularly in terms of the participation of women in managerial and leadership positions. Women's representation in science, technology, engineering, and mathematics (STEM) fields remains largely on the periphery, as noted by Swarup & Dey (2020). It's noteworthy that this societal issue is being approached as a research subject, primarily by women.

There are four distinct types of universities offering PhD programs in Women's Studies. Among these, state universities have granted the highest number of theses, followed by central universities and deemed-to-be universities. Table 1 provides a comprehensive view of the number of theses awarded by various Indian universities, illuminating their substantial contributions to research.

Andhra University, categorized as a state university, takes the lead with 73 (5%) published theses, followed closely by Aligarh Muslim University, a central university, with 45 (3%) theses. In the case of Andhra University, departments such as Commerce & Management, Sociology, Social Work, English Literature, and Legal Studies emerge as key contributors, focusing on research pertaining to women's issues. On the other hand, Aligarh Muslim University, Mother Teresa Women's University, Jadavpur University, and University of Calcutta stand out as universities with dedicated Women's Studies departments. Additionally, institutions like Banaras Hindu University, Jamia Millia Islamia University, and Manonmaniam Sundaranar University also deserve mention as notable contributors. This compilation underscores the varying research productivity of different universities across states, underscoring their pivotal role in advancing academic knowledge and scholarship across diverse domains.

Table 2: Top 20 Universities which contributed theses during 2011-2020

Sl. No	University	Place	State	No. of Theses Published
1	Andhra University	Vishakhapatnam	Andhra Pradesh	73
2	Aligarh Muslim University	Aligarh	Uttar Pradesh	45
3	Madurai Kamaraj University	Madurai	Tamil Nadu	39
4	Mother Teresa Women's University	Kodaikanal	Tamil Nadu	38
5	Banaras Hindu University	Varanasi	Uttar Pradesh	35
6	Manonmaniam Sundaranar University	Tirunelveli	Tamil Nadu	35
7	Panjab University	Chandigarh	Punjab	31
8	University of Mysore	Mysore	Karnataka	31

9	Gauhati University	Guwahati	Assam	26
10	University of Delhi	New Delhi	Delhi	26
11	Assam University	Silchar	Assam	24
12	University of Calcutta	Kolkata	West Bengal	24
13	Jadavpur University	Kolkata	West Bengal	23
14	Jamia Milia Islamia University	Delhi	Delhi	21
15	Maharshi Dayanand University	Rohtak	Haryana	21
16	Bharathiar University	Coimbatore	Tamil Nadu	20
17	Punjabi University	Patiala	Punjab	19
18	Swami Ramanand Teerth Marathwada University	Nanded	Maharashtra	19
19	University of North Bengal	Darjeeling	West Bengal	18
20	Tata Institute of Social Sciences	Mumbai	Maharashtra	17

The allocation of awarded theses across Indian states mirrors a richly diverse academic landscape. Taking the lead, Tamil Nadu (18%) claims the highest count, closely followed by Uttar Pradesh (10%) and Andhra Pradesh (9%). States such as Karnataka, West Bengal, and Delhi also make substantial contributions, spotlighting a dynamic intellectual atmosphere. The provided data highlights a widespread zeal for knowledge pursuit and academic excellence spanning the entirety of India. This is evident in the substantial research output emanating from both densely populated regions and relatively smaller states like Maharashtra, Punjab, and Assam.

It's evident that women's issues are pervasive throughout India, with the southern and northern regions exhibiting a notable concentration in comparison to other parts of the country.

5.3 Topics prevalence through STM and LDA model

To estimate the optimal number of topics in the Short Text Modelling (STM) approach, we have initiated the topic modelling process. Determining the ideal number of topics is not a straightforward task, as it hinges on both "analytical utility" and "interpretability" aligned with the research question.

In comparison with other methods, we have observed that this particular approach stands out due to its improved speed and accuracy. This similarity in approach can also be seen in the works of Albalawi, Kim et al., and Zhang et al. Following a similar methodology, we executed the STM model by specifying twenty topics. The results depicting the relative goodness of fit for each number of topics are illustrated in Figure 2.

Figure 2 indicates that the range of 10 to 25 topics yields notably high held-out likelihood, strong semantic coherence, maximized lower bound, and minimal residuals. Taking into consideration Griffiths and Steyver's (2004) two-state algorithm, we opted to maximize the number of topics at k = 20. Consequently, we determine that the optimal number of latent topics would be around 18, based on this decision.

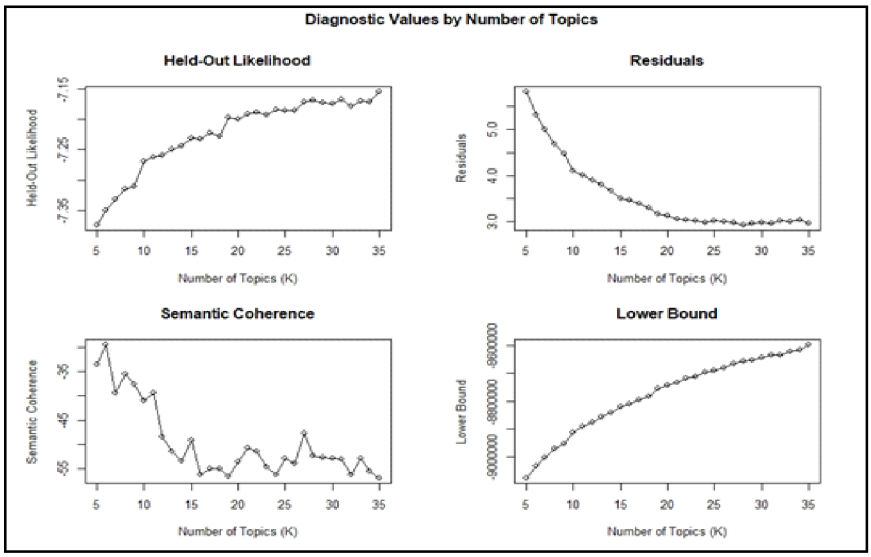


Figure 2: Determining the optimal range of latent topics through model-fit statistics

It's worth noting that Kuster and Garrido (2020) arrived at a similar conclusion regarding the optimal number of topics. To further validate our findings, we conducted multiple algorithms runs and compared the respective results.

Figure 3 presents the results of the unsupervised machine learning STM model, showcasing 20 topics. Each topic is accompanied by the top five words stems associated with it, and they are ranked based on prevalence within the corpus. Women's studies scholars will likely identify several familiar themes within these theses, as these topics align broadly with course syllabi, intellectual discussions, and scholarly publications. The model provides a comprehensive output derived from the corpus; with the significance of each topic measured by the proportion of document content it encompasses.

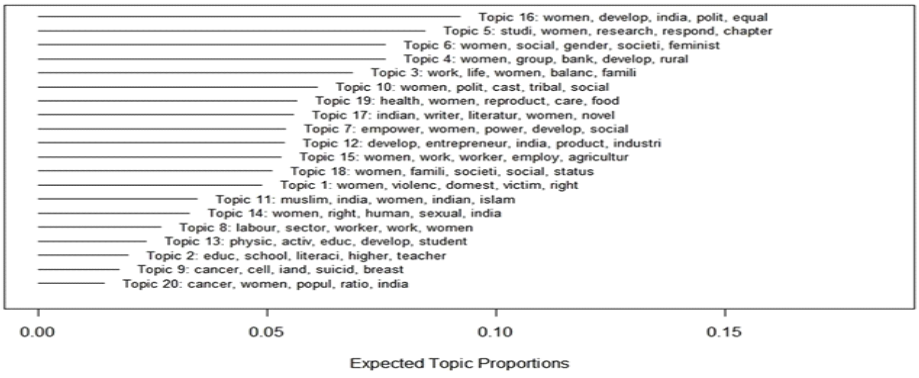


Figure 3: Level for a 20- topic (top five associated word stems) produced by structural topic model of 1389 women studies theses

The x axis represents the proportion of each topic within the overall corpus.

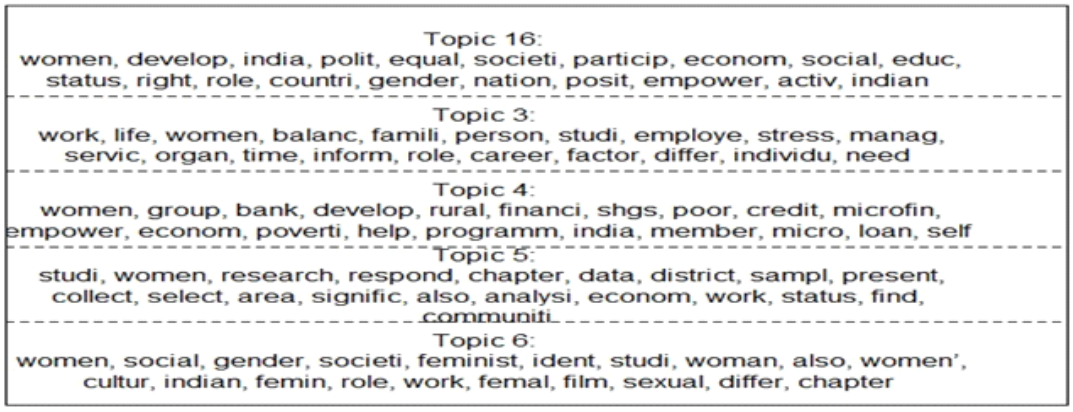


Figure 4: Top twenty FREX words associated with the first four topics

The figure represents a hierarchical structure of topic proportions using random words. In Figure 3, a specific topic is highlighted from the entire corpus. Notably, Topic Sixteen addresses women's development and political participation, shedding light on their presence and inequalities in India. This topic holds the highest prevalence within the corpus, as observed in the work of Chen et al. (2020). Meanwhile, Topics Five and Six delve into areas such as women's studies, research, societal significance, gender responsibilities, and feminist movements.

To emphasize each topic in the literature, we assigned a single label to each one. Figure 4 provides an illustrative example of the top five topics, accompanied by their top twenty FREX (Frequency and Exclusivity) words. These words exhibit a high level of exclusivity to their respective topics. Notably, topic 16 boasts the highest probability mean value of 0.092, followed by topic 5 with 0.084, topic 4 with 0.075, and topic 3 also with 0.075.

We employed the Maximum a Posteriori (MAP) algorithm, as outlined in the work of Tonidandel et al. (2022), to estimate document topic loading. Figure 5 presents a histogram depicting the proportions of topics in all the PhD theses from the years 2011 to 2020. The median estimate of document topic proportions is indicated by the dashed red line. The probabilistic statistical model reveals that there exists a "limited number of topics when the probability of a key topic is high, and the probability of non-key topics is low" (Sharma et al., 2021, p.8). Evidently, each extracted topic bears only limited relevance to several documents within the theses.

Moving on to Figure 6, this visual aids in comparing the relationships between four and five topics. Stem word sizes are scaled based on their frequencies, and their arrangement indicates their closeness to specific topics. The vertical positioning of words is randomized. On the horizontal axis, stem words are placed towards the far left or right to signify their alignment with particular topics. Notably, Figure 6 underscores that topic 4 is strongly linked with terms like women, empowerment, development, and economic and financial

conditions. Conversely, topic 6 prominently incorporates the term “Indian literature writer” throughout the entire observed period.

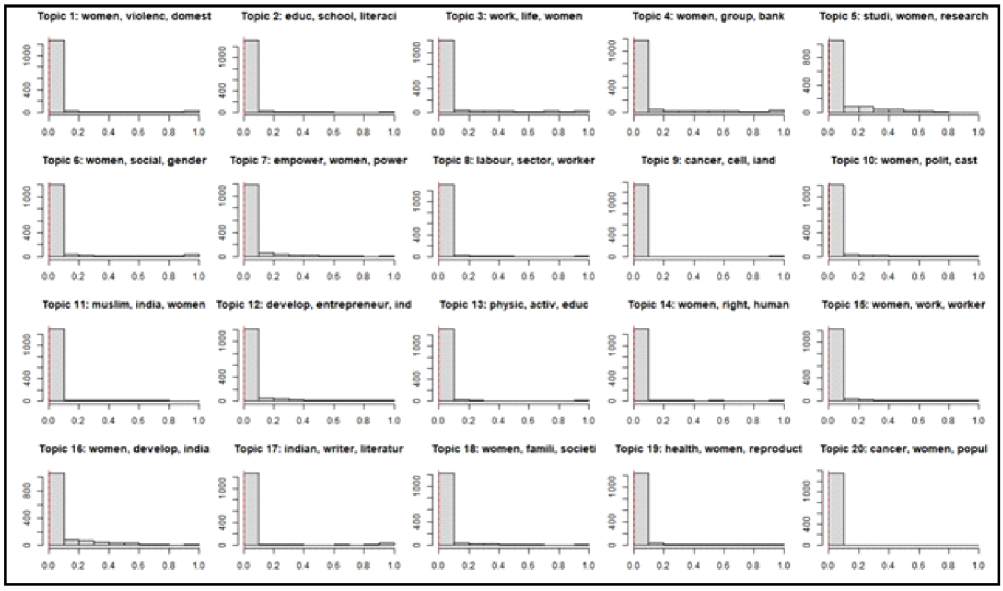


Figure 5: Estimates for selected document-topic loadings (MAP method)

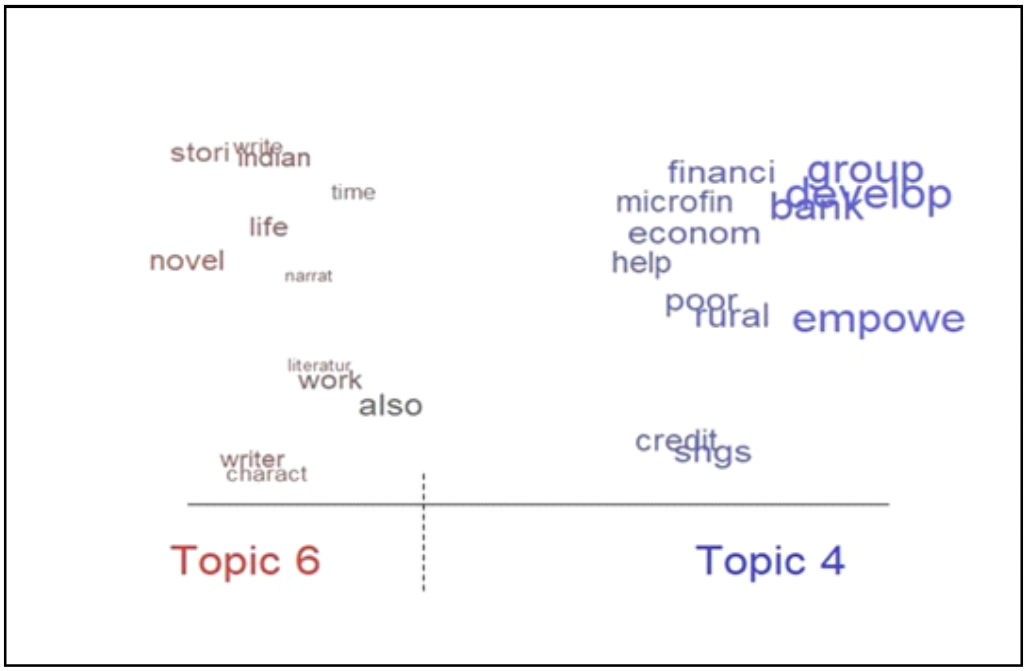


Figure 6: comparative graph between topic 6 and topic 4

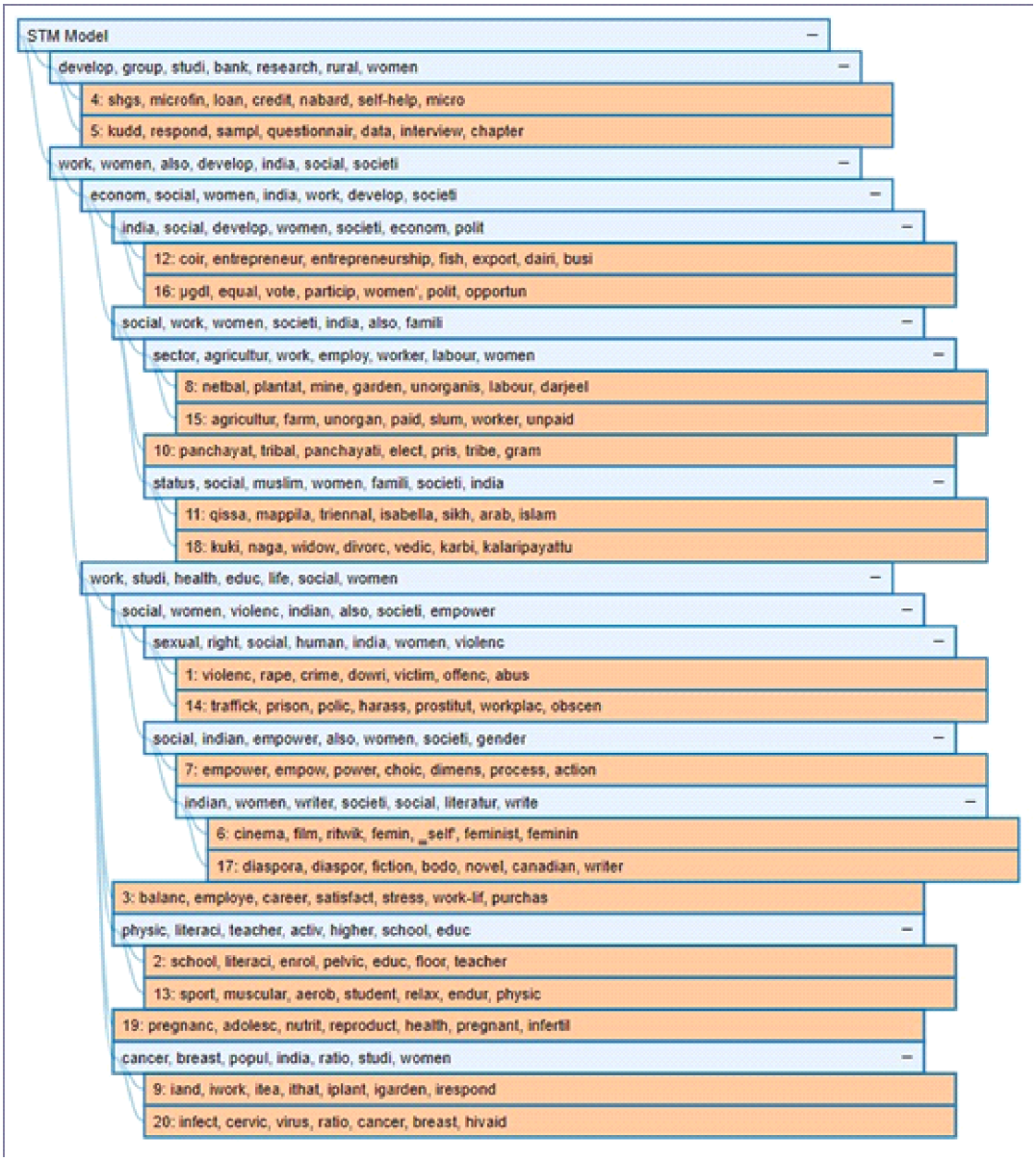


Figure 7: Hierarchical display of structural topic model (STM)

We have tried to exclude some of the stem words such as also, HIV aids, sector etc. But when we exclude the model gives over fitting results. So, we decide not exclude these stem words.

Further analysis has uncovered the hierarchical structure of the topics and their correlations within the model. These topics can be organized into a hierarchical framework, leading to the formation of eight clusters: (4, 5), (12, 16), (8, 15), (11, 18), (1, 14), (6, 17), (2, 13), and (9, 20), as depicted in Figure 7.

In a broader context, our findings shed light on the central research domains within women's studies and their interconnections through the application of structural topic modelling and visualization techniques. It has become evident that the economic and financial conditions affecting women in India are subpar, prompting researchers to focus on topics concerning women's empowerment and developmental issues. Another significant concern pertains to the inadequate societal status and position of women. Scholars have directed their attention towards topics involving violence, oppression, sexual harassment, and criminal offenses. Within the context of India, an entrenched patriarchal framework continues to exert its influence, leading to a gradual erosion of women's individual identities.

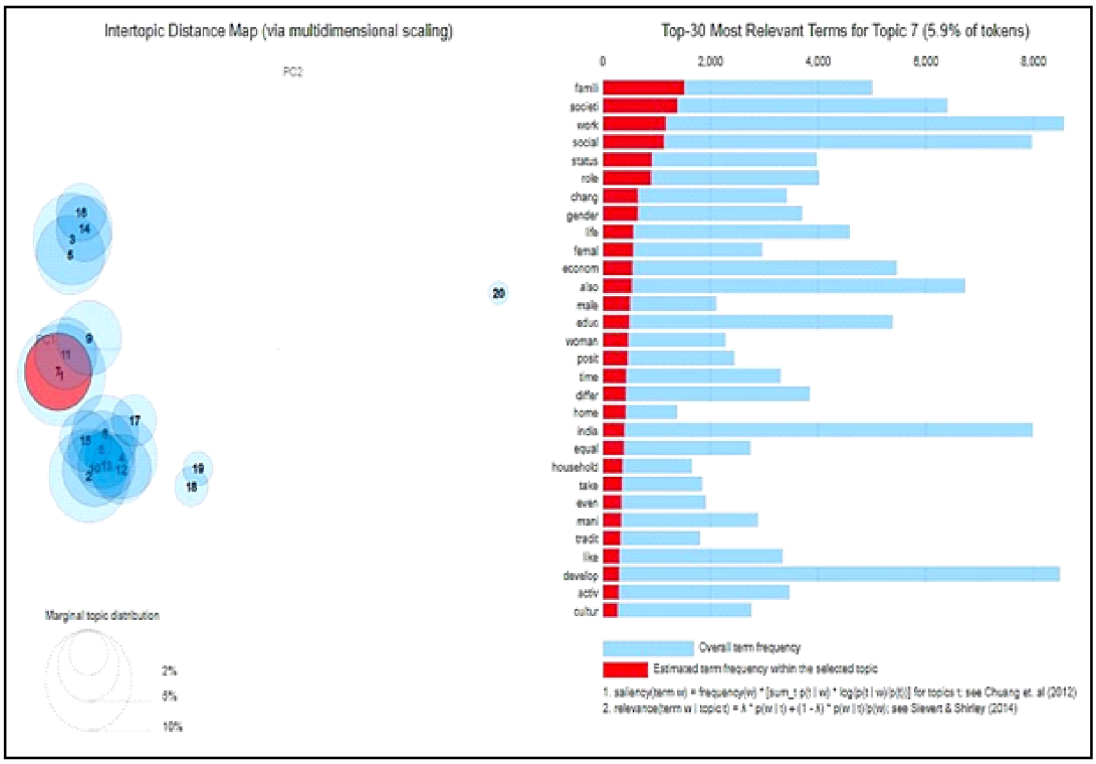


Figure 8: Topic visualization select topic 7 via LDAvis

We employed LDAvis (S. Han et al., 2020), an interactive web tool for visualization, to represent the chosen LDA model. LDAvis assists in analyzing the interconnections between topics and terms. In Figure 4, the right panel displays the top 30 relevant terms associated with the selected topic. The red bar represents the term's frequency specifically within the chosen topic, while the blue bar reflects the term's overall frequency. For interpreting the topic, we adopt a value of ë=1, which emphasizes the term's frequency within that particular topic. On the other hand, the left panel portrays the relative size of each topic, indicating its prevalence across all texts. Inter-topic distances are adjusted through PCA. Users can also define relevance matrices to explore the most significant terms for each topic.

Upon analysis, we observed that topics 10, 7, 3, and 19 do not cluster with other topics. In Figure 8, we visualize topic 7 to gain insights into its prevalence areas. Our findings suggest that researchers focus on aspects such as women's work-life balance and societal status within this topic. Additionally, topic 3 emphasizes themes related to gender, inequality, Dalit issues, and caste problems.

5.4 Keyword Analysis

In the tree plot illustrating the top 45 author keywords, noteworthy terms like "women empowerment" and "Self-help groups" emerged as the most frequent, holding the top two positions in terms of frequency, as depicted in Fig. 5. Notably, the range of words encompassed within the author's keyword tree plot is extensive. The increasing prevalence of terms such as "heterogeneity" underscores the significance of key women's issues-related keywords within the field of women's studies. This suggests their likely growth and prominence in the future.

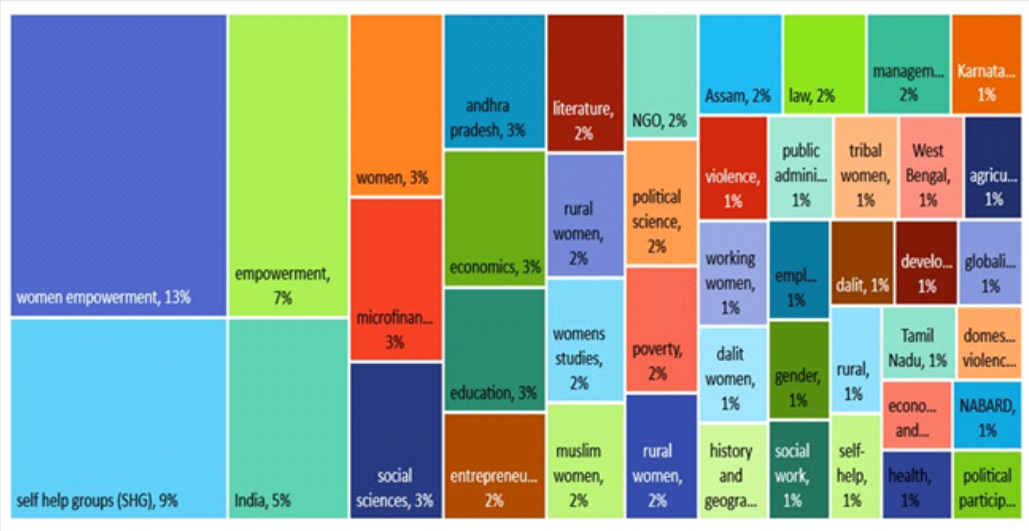


Figure 9: shows the Tree Plot of Top 45 Author Keywords

(Excluding Common Words like Women, social science, humanities

6. Findings

Our analysis unveils a diverse spectrum of vital research focuses within the realm of women's studies. These include subjects such as women's literature, women's health and nutrition, economic and labour issues, and the challenges faced by slum dwellers. These findings underscore both the breadth and depth of research activities within women's studies. Moreover, these findings might also hint at the blurring of boundaries within the field, reflecting its increasing integration and significance in society. As stated by Buckley, this field could potentially evolve into "an area of application for applied concepts from other disciplines," solidifying its position as a pivotal pillar in society.

Our findings predominantly focus on a restricted time frame and largely revolve around the Indian context. However, it's worth noting that a subset of the theses also addresses issues in other countries such as Nigeria, Bangladesh, Nepal, Sri Lanka, Egypt, Mongolia, and Kazakhstan. Notably, there is an overlap between Topic Twenty (as shown in Figure 3) and Topic Nine.

7. ETDs and Their Importance

There is a need for continued discussion on addressing sociocultural issues, as it is an essential pillar of societal transformations. Moreover, NGOs and government policymakers would benefit from more extensive discussions about societal problems and challenges (Bohr & Dunlap, 2018). Shodhganga ETD can play a significant role in investigating research trends in Indian Universities. The LDAvis model provided an interactive HTML file for an informative viewing of women studies research on twenty topics. Such kind of interactive tools could interest novice scholars who possess a strong motivation to swiftly comprehend the nuances of specific research areas quickly. Research serves as a vital medium of communication, aimed at addressing critical global challenges such as climate change, food scarcity, and the containment and eradication of diseases. Recent events, including the far-reaching pandemic that led to loss of lives and livelihoods, underscore the urgency of these endeavours. The dissemination of research findings is intended for universal access, targeting audiences from all walks of life.

In line with the UGC regulations of 2018 theses are now required to be promptly uploaded to the national repository, Shodhganga, immediately upon degree conferral. This directive reflects a broader global trend towards open access, as concerted endeavours are made worldwide to facilitate unrestricted access to research outcomes. Electronic Theses and Dissertations (ETDs) repositories play a pivotal role for multiple stakeholders.

ETDs provide valuable insights into prevailing research trends, shedding light on extensively explored subjects as well as areas warranting greater attention. Moreover, they uphold the integrity of research by fostering an environment where scholars are conscious that their work is subject to examination and assessment by peers. This awareness encourages researchers to uphold rigorous standards in their respective fields.

The significance of ETDs extends to their impact on university rankings. These repositories can enable ranking bodies to discern institutions demonstrating exemplary research performance and those that may require improvement. In essence, ETDs facilitate knowledge dissemination, contribute to enhancing the quality of research and Education, and highlight important social issues.

8. Conclusion

Women's studies research in India holds immense significance due to its focus on addressing critical issues such as gender inequality, intersectionality, social transformation, and policy formulation. Through our exploration, we have highlighted the persistent presence of gender disparities in the country. Researchers

have illuminated societal norms and discriminatory practices prevalent in a patriarchal society, which hinder the advancement and empowerment of women.

Our findings underscore that women's health awareness and social status in rural areas remain at a rudimentary stage, particularly regarding girls' Education, where preference is often given to boys. Meanwhile, urban areas present challenges, including inadequate access to clean drinking water and sanitation.Researchers emphasise women's empowerment, the Sustainable Development Goals (SDGs), societal status, economic challenges, girl child education, and women's reproductive health. Despite higher rates of women's employment, gender inequality persists in leadership roles. The insights gained from women's studies research play a pivotal role in informing policy discussions and shaping the implementation of legislation and initiatives to advance and safeguard women's rights.

Enhanced dialogues among non-governmental organizations (NGOs) and government authorities regarding the socio-economic issues would benefit immensely. In sum, women's studies research in India strives to foster a more equitable, just, and inclusive society, promoting well-being and progress for all individuals.

References

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using Topic Modelling Methods for Short-Text Data: A Comparative Analysis. Frontiers in Artificial Intelligence, 3(July), 1–14. https://doi.org/10.3389/frai.2020.00042

Bohr, J., & Dunlap, R. E. (2018). Key Topics in environmental sociology, 1990–2014: results from a computational text analysis. Environmental Sociology, 4(2), 181–195. https://doi.org/10.1080/23251042.2017.1393863

Chanana, K. (2020). Women and leadership: Strategies of gender inclusion in institutions of higher Education in India. In Strategies for Supporting Inclusion and Diversity in the Academy: Higher Education, Aspiration and Inequality. https://doi.org/10.1007/978-3-030-43593-6_8

Chen, X., Zou, D., Cheng, G., & Xie, H. (2020). Computers & Education Detecting latent topics and trends in educational technologies over four decades using structural topic modelling/ : A retrospective of all volumes of Computers & Education. Computers & Education, 151(September 2019), 103855. https://doi.org/10.1016/j.compedu.2020.103855

David M, B. (2012). Probabilistic Topic Models. Communications of the ACM, 55(4), 77–84. https://doi.org/10.1145/2133806.2133826

Dutoya, V. (2022). Journal of International Women' s Studies Women' s Studies , Gender Studies , and LGBT / Queer Studies/ : Defining and Debating the Subject of Academic Knowledge in India. 23(2).

Goyal, A., Gupta, V., & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques: A systematic review. Computer Science Review, 29, 21–43. https://doi.org/10.1016/j.cosrev.2018.06.001

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America, 101(SUPPL. 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Han, S., Ye, S., & Zhang, H. (2020). Visual exploration of Internet news via sentiment score and topic models. Computational Visual Media, 6(3), 333–347. https://doi.org/10.1007/s41095-020-0178-4

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent Dirichlet allocation topic model. Scientometrics, 125(3). https://doi.org/10.1007/s11192-020-03721-0

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about/ ? Text analysis using structural topic model. Tourism Management, 72(March 2018), 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2014). Interactive topic modelling. Machine Learning, 95(3), 423–469. https://doi.org/10.1007/s10994-013-5413-0

Kim, S., Nelson, J. G., & Williams, R. S. (1985). Mixed-basis band-structure interpolation scheme applied to the fluorite-structure compounds NiSi2, AuAl2, AuGa2, and AuIn2. In Physical Review B (Vol. 31, Issue 6). https://doi.org/10.1103/PhysRevB.31.3460

Nanda, P. K. (2021). Social science research more than doubled in five years, says education ministry. Mint. https://www.livemint.com/news/india/social-science-research-more-than-doubled-in-five-years-says-education-ministry-11616155809875.html

Park, J., & Han, A. Y. (2023). Nurse Education Today Medication safety education in nursing research/ : Text network analysis and topic modelling. Nurse Education Today, 121(December 2022), 105674. https://doi.org/10.1016/j.nedt.2022.105674

Paswan, J., & Singh, V. K. (2020). Gender and research publishing analyzed through the lenses of discipline, institution types, impact and international collaboration: a case study from India. Scientometrics, 123(1), 497–515. https://doi.org/10.1007/s11192-020-03398-5

Ries, T. (2012). Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviour, Products and Institutions. Willard McCarty (ed). In Literary and Linguistic Computing (Vol. 27, Issue 2). https://doi.org/10.1093/llc/fqs006

Rose, R. L., Puranik, T. G., Mavris, D. N., & Rao, A. H. (2022). Application of structural topic modelling to aviation safety data. Reliability Engineering and System Safety, 224(April), 108522. https://doi.org/10.1016/j.ress.2022.108522

Sharma, A., Rana, N. P., & Nunkoo, R. (2021a). Fifty years of information management research: A conceptual structure analysis using structural topic modelling. International Journal of Information Management, 58. https://doi.org/10.1016/j.ijinfomgt.2021.102316

Sharma, A., Rana, N. P., & Nunkoo, R. (2021b). Fifty years of information management research: A conceptual structure analysis using structural topic modelling. International Journal of Information Management, 58, 102316. https://doi.org/10.1016/J.IJINFOMGT.2021.102316

Swarup, A., & Dey, T. (2020). Women in science and technology: An Indian scenario. Current Science, 119(5). https://doi.org/10.18520/cs/v119/i5/744-748

Tonidandel, S., Summerville, K. M., Gentry, W. A., & Young, S. F. (2022). Using structural topic modelling to gain insight into challenges faced by leaders. Leadership Quarterly, 33(5), 101576. https://doi.org/10.1016/j.leaqua.2021.101576

Wöhrer, V. (2016). Gender studies as a multi-centred field? Centres and peripheries in academic gender research. Feminist Theory, 17(3), 323–343. https://doi.org/10.1177/1464700116652840

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: A statistical framework. International Journal of Machine Learning and Cybernetics, 1(1–4), 43–52. https://doi.org/10.1007/s13042-010-0001-0