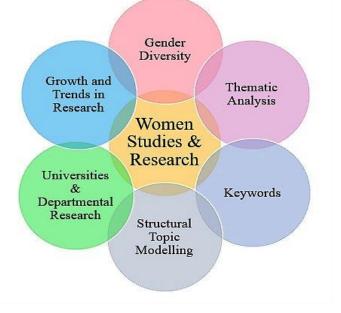
# Trends in Women Studies, 2011-2020: A Computational Text Analysis

Suman Das, IIT Indore Manoj Kumar K, INFLIBNET Centre Anup Kumar Das, JNU Manorama Tripathi, JNU



# Acknowledgement

This work has been supported by ICSSR major funding project F. No 02/199/GN/2021-22/ICSSR/RP/MJ

Public Policy Implications of Social Sciences Doctoral Research Carried Out in Indian Universities: Content Analysis Based on Awarded Doctoral Theses



## **Major Thrust Areas**



Why Women Studies? its Importance?



**Research Pattern on PhD Theses** 

**Computational Text mining** 



------

Women Issues in Indian Context



# **Objectives**

- To investigate the main topics of Women's Studies subject area during 2011-2020 and available in the Shodhganga repository;
- Highlight the gender diversity among researchers and ٠ supervisors in Women's Studies;
- Determine the Indian universities and their specific departments or centers that exhibit higher research output;
- Identify hierarchical structure of research issues in the ٠ theses on women studies.

|                 | Data 🔍   |
|-----------------|--|
| Data Collection | Database Selection<br>Sodhganga e-thesis Repository<br>Year-2011-2020.   |
|                 | Search Design<br>Women Studies subject areas.<br>Keywords used 'Women studies', 'Gender Diversity, 'Social<br>welfare', 'social issues', 'tradition and culture', 'women<br>empowerment, 'women worker', 'self-help', 'women employees',<br>'slum', 'Violence', 'Prostitute', 'women rights'   |
| Dat             | Manual Selection<br>Visit all the identified theses handle. Manually copied title,<br>abstract, keywords, author name and respective supervisors,<br>universities and departments where published the theses and<br>year of completion.  |
| Pre-processing  | Total Document<br>We found a total of 1720 theses.<br>Remove Duplicates<br>After manually checking, we exclude 9 of theses of repetition of<br>the same title.   |
|                 | Final Selection<br>Found 2,208 theses published in women's studies subject areas.<br>312 (193+119) theses were excluded because these were<br>published other than in the English language (193 in Hindi119<br>other languages etc.)<br>10 theses were excluded because they were unable to be<br>downloadable, and Bitstream has changed<br>Finally, selected 1389 theses for this study. |
| Analysis        | Computational Text Analysis<br>Adopting quantitative thematic analysis to find out a broad<br>spectrum of women studies. We used Structural Topic Modelling<br>and Latent Dirichlet Allocation (LDA) model to discover key<br>thèmes in previous research.   |
|                 | Analytical categories<br>Categories researchers gender diversity and find out Indian<br>region where women's issues become prominent in society.<br>Year-wise growth in subject areas.   |





## **Computational Text Analysis**

**1949, Roberto Busa** collaborated with **IBM** to develop an Index Variorum. 10,000 handwritten texts of St. Thomas Aquinas' works

Harnessed textual data, descriptive metadata, and summary information to enhance business insights

21st century, computational linguistics. included machine learning (ML), deep learning, artificial intelligence (AI), computational cognition, and humancomputer interaction (HCI).

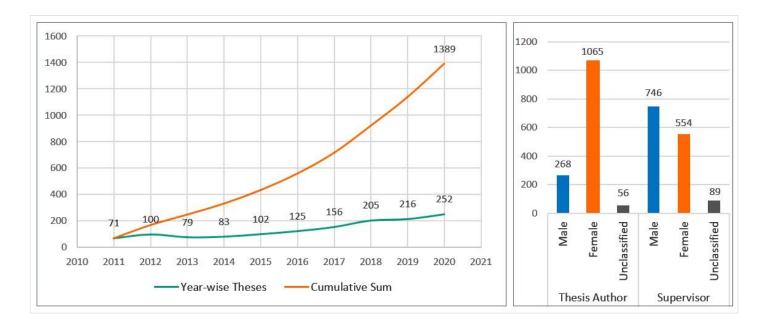
 Humanities are academic discipline that study aspect of human society and culture.

## **Text clustering and Topic Modelling**

- □ Technique to separate objects into different classes.
- Levels of granularity by cluster objects such as documents, paragraphs, and sentences
- Clustering is an unsupervised learning method, while classification uses supervised learning.
- We used STM and LDA (Bayesian Model) to understand dense clusters allowing for interpretable topics and important words in the topic descriptions
- 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
- STM and LDA models to analyse PhD theses abstract and recommendation for textual analysis
- \* 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 ( $\theta$ ) and topic-word distribution ( $\beta$ ).
- STM model provides us with a deeper understanding of the hierarchical structural of text



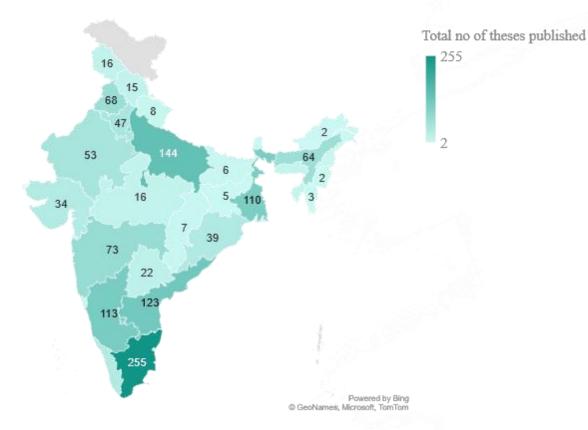
### Gender Diversity and year-wise Growth







## **Universities published theses during 2011-2020**



Sh@dh ganga

### a reservoir of Indian theses

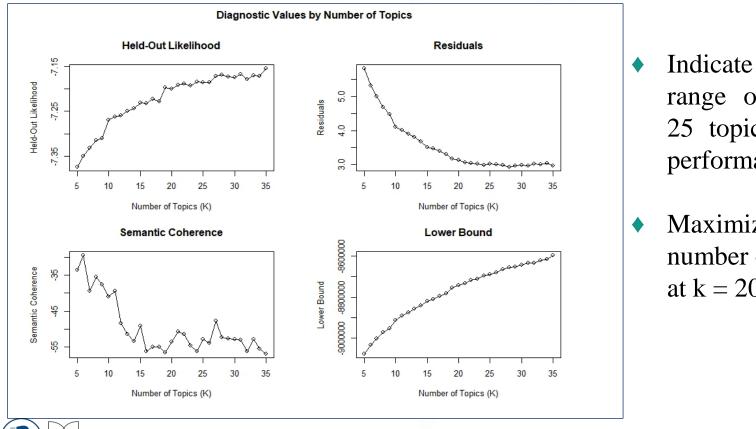
| Types of             | No of Theses |  |  |  |  |
|----------------------|--------------|--|--|--|--|
| University           | Published    |  |  |  |  |
| State University     | 902          |  |  |  |  |
| Central University   | 265          |  |  |  |  |
| Deemed<br>University | 143          |  |  |  |  |
| Private University   | 79           |  |  |  |  |
| Total                | 1389         |  |  |  |  |



## **Topics prevalence through STM and LDA model**



## **Topics prevalence through STM and LDA model**

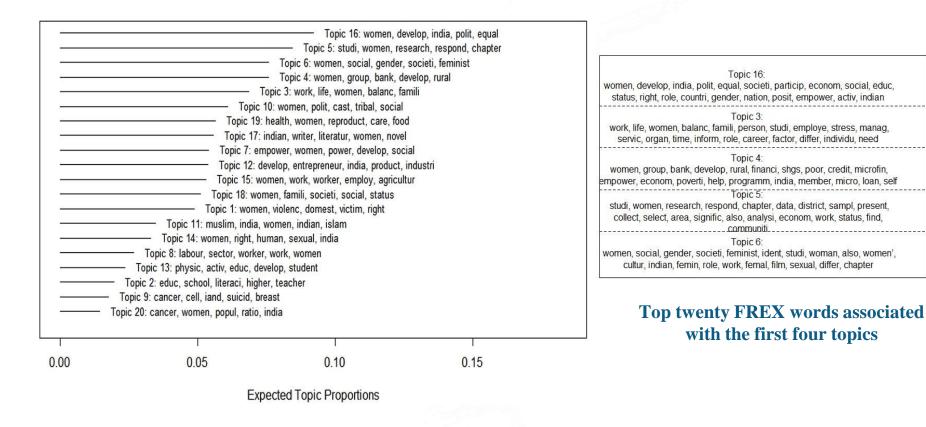


the range of 10 to 25 topics, High performance

Maximize number of topics at k = 20.



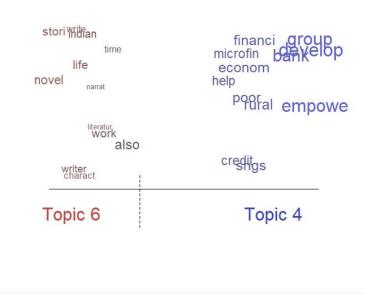
## **Topics prevalence through STM model**







### **Topics prevalence through STM**



Comparative graph between topic 6 and topic 4

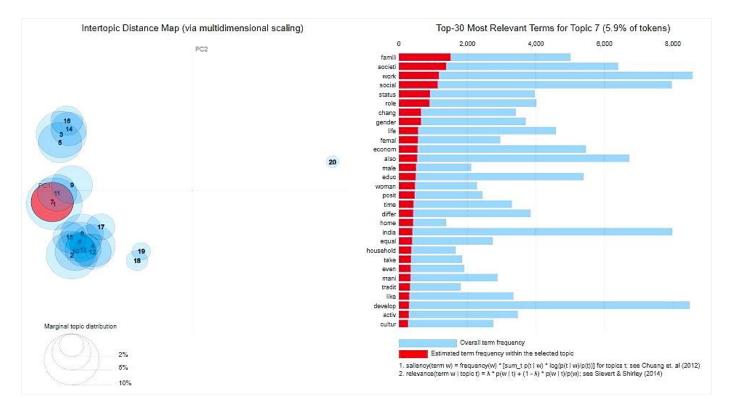
| evelop, g | roup, studi, bank, research, rural, women -                         |
|-----------|---|
| 4: shg    | is, microfin, Ioan, credit, nabard, self-help, micro                |
| 5: kud    | ld, respond, sampl, questionnair, data, interview, chapter          |
| ork, wom  | ien, also, develop, india, social, societi –                        |
| econo     | m, social, women, india, work, develop, societi -                   |
| 1 ir      | idia, social, develop, women, societi, econom, polit -              |
| 17        | 12: coir, entrepreneur, entrepreneurship, fish, export, dairi, busi |
|           | 16: µgdl, equal, vote, particip, women', polit, opportun            |
| s         | ocial, work, women, societi, india, also, famili –                  |
| Т         | sector, agricultur, work, employ, worker, labour, women -           |
|           | 8: netbal, plantat, mine, garden, unorganis, labour, darjeel        |
|           | 15: agricultur, farm, unorgan, paid, slum, worker, unpaid           |
|           | 10: panchayat, tribal, panchayati, elect, pris, tribe, gram         |
|           | status, social, muslim, women, famili, societi, india –             |
|           | 11: qissa, mappila, triennal, isabella, sikh, arab, islam           |
|           | 18: kuki, naga, widow, divorc, vedic, karbi, kalaripayattu          |
| work,     | studi, health, educ, life, social, women –                          |
| s         | ocial, women, violenc, indian, also, societi, empower —             |
|           | sexual, right, social, human, india, women, violenc -               |
|           | 1: violenc, rape, crime, dowri, victim, offenc, abus                |
|           | 14: traffick, prison, polic, harass, prostitut, workplac, obscen    |
|           | social, indian, empower, also, women, societi, gender -             |
|           | 7: empower, empow, power, choic, dimens, process, action            |
|           | indian, women, writer, societi, social, literatur, write            |
|           | 6. cinema, film, ritwik, femin, _self', feminist, feminin           |
|           | 17: diaspora, diaspor, fiction, bodo, novel, canadian, writer       |
| 3         | : balanc, employe, career, satisfact, stress, work-lif, purchas     |
| p         | hysic, literaci, teacher, activ, higher, school, educ -             |
|           | 2: school, literaci, enrol, pelvic, educ, floor, teacher            |
|           | 13: sport, muscular, aerob, student, relax, endur, physic           |
|           | 9: pregnanc, adolesc, nutrit, reproduct, health, pregnant, infertil |
| C         | ancer, breast, popul, india, ratio, studi, women –                  |



Hierarchical display of structural topic model (STM) 12

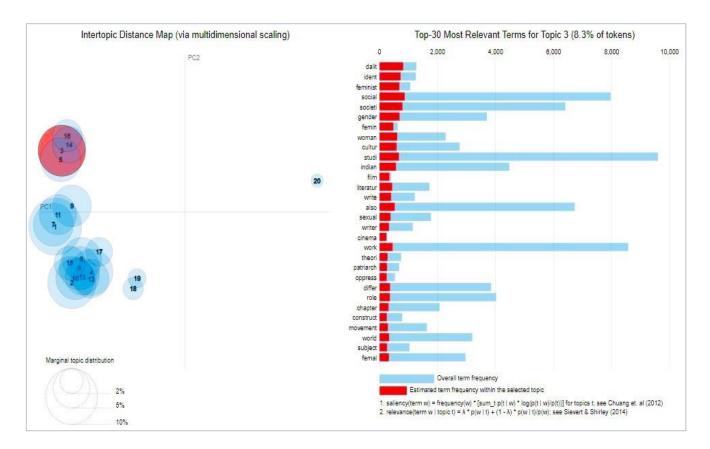


### **Topics prevalence through LDA**



**Topic visualization select topic 7 via LDAvis** 





## **Topic visualization of topic 3**

14



## **Keyword Analysis**

|                            |                    |                        | andhra            | literature,              | NGO, 2%               | Assam, 2% law,           |                        |                       | anagem<br>2%               | Karnata<br>1%               |  |
|----------------------------|--------------------|------------------------|-------------------|--------------------------|-----------------------|--------------------------|------------------------|-----------------------|----------------------------|-----------------------------|--|
|                            |                    | women, 3%              | pradesh, 3%       | 2%<br>rural              | political             | violence,<br>1%          | public<br>admini<br>1% | tribal<br>womer<br>1% |                            | agricu<br>1%                |  |
| women empowerment, 13%     | empowerment,<br>7% | microfinan<br>3%       | economics, 3%     | women,<br>2%             | science,<br>2%        | working<br>women,<br>1%  | empl<br>1%             | dalit, 19             | develo<br>6 1%             | globali<br>1%               |  |
|                            |                    |                        | education, 3%     | womens<br>studies,<br>2% | poverty,<br>2%        | dalit<br>women,<br>1%    | gender,<br>1%          | rural,<br>1%          | Tamil<br>Nadu, 1%<br>econo | domes<br>violenc            |  |
| self help groups (SHG), 9% | India, 5%          | social<br>sciences, 3% | entrepreneu<br>2% | muslim<br>women,<br>2%   | rural<br>women,<br>2% | history<br>and<br>geogra | social<br>work,<br>1%  | self-<br>help,<br>1%  | and<br>health,<br>1%       | 1%<br>political<br>particip |  |





## Conclusion

- Researchers emphasise women's empowerment, the Sustainable Development Goals (SDGs), Millennium Development Goals (MDGs), societal status, economic challenges, girl child education (Beti Bachao Beti Padhao), and women's reproductive health.
- □ Urban areas present different challenges, including inadequate access to clean drinking water and sanitation, the plight of slum dwellers, disparities across income groups, and the juggling act faced by working women to balance their responsibilities.
- Despite higher rates of women's employment, gender inequality persists in leadership roles.
- Enhanced dialogues among civil societies and government authorities regarding the mentioned socio-economic issues would be immensely beneficial. 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.



# Team Presentation



Mr. Suman Das Research Associate

JPN Centre of Excellence in the Humanities, IIT Indore



Mr. Manoj Kumar K Scientist-F (CS) Information and Library Network Centre



Dr. Anup Kumar Das Documentation Officer. Centre for Studies in Science Policy Jawaharlal Nehru University http://anupkumardas.blogspot.com Dr. Manorama Tripathi Librarian JNU Central Library Jawaharlal Nehru University utrainitie