News Media Coverage of India's National River Linking Program (2004-2021): A Case Study of the Ken-Betwa Link using Topic Modeling

Funding: PSU RIPERIA

PennState College of Earth and Mineral Sciences
Department of Geography

Authors

Harman Singh, M.Sc. + Ph.D. student in Geography
Matt Hansen, R&D Engineer at the Institute for Computational and Data Sciences
Trevor Birkenholtz, Associate Professor in Geography

Contact Email: (hxs5376@psu.edu)

Introduction

Largescale water infrastructure is on the rise globally. News media reporting of infrastructure development both reflects and influences public opinion.

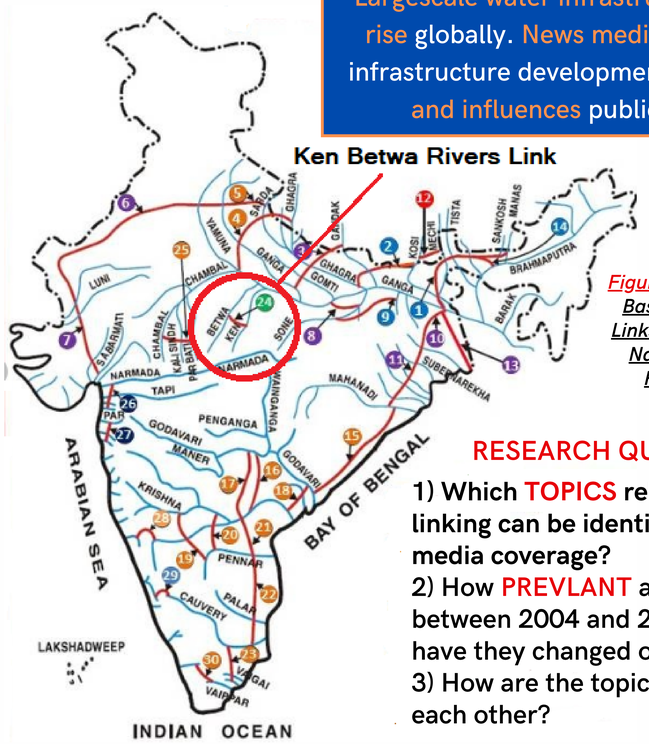


Figure 1: Proposed Inter-Basin Water Transfer Links. Map made by the National Institute of Hydrology, India

RESEARCH QUESTIONS:

1) Which TOPICS related to river linking can be identified from news media coverage?
2) How PREVLANT are these topics between 2004 and 2021 and how have they changed over time?
3) How are the topics RELATED to each other?

Results

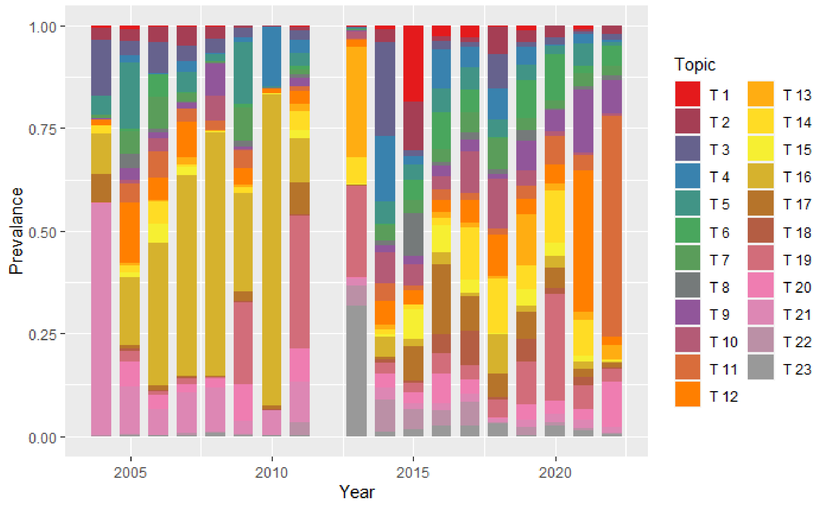
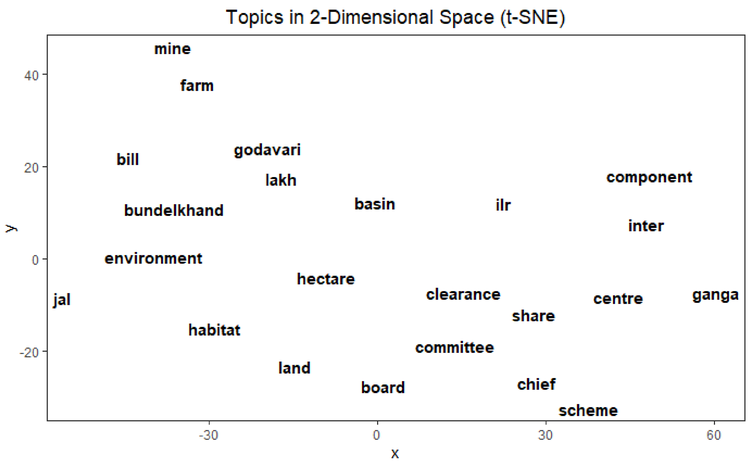


Figure 4: Illustrates temporal changes and prevlance of topics between 2004 and 2021.

The LDA model revealed 23 TOPICS from the news corpus. Topics prior to 2020 focused on environmental and government policy. In more recent years, such as 2021, topics 11 "Farm" and 12 "INRLP (dam)" dominate the new media narrative eluding to the construction of a dam that is expected to recharge and increase the groundwater levels to help the farming community with crop production and livestock output.

The WARDS METHOD was chosen to calculate the distance between the topics as it minimizes the variance between clusters. Topics with similar content were clustered together in the t-SNE plot and the closer the topic label to each other the more similar the topics (and their subset words).

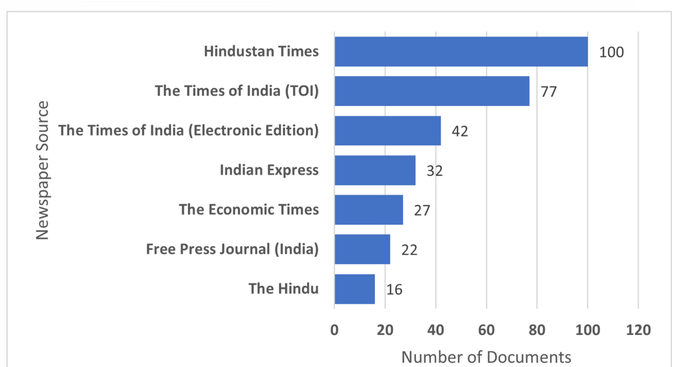
Figure 5: T-distributed stochastic neighbor embedding (t-SNE) reduces the tendency to crowd points together in the center of the plot and gives each datapoint a location that can reveal structures at many different scales.



Methodology

DATA SOURCE: LexisNexis

Figure 2: Distribution of news articles across newspaper sources (TOTAL:316)



Latent Dirichlet Allocation(LDA) TOPIC MODEL:
Topic Modeling is a form of unsupervised machine learning. LDA works as a generative PROBABILISTIC model that allows extraction of variables that are not directly observed.

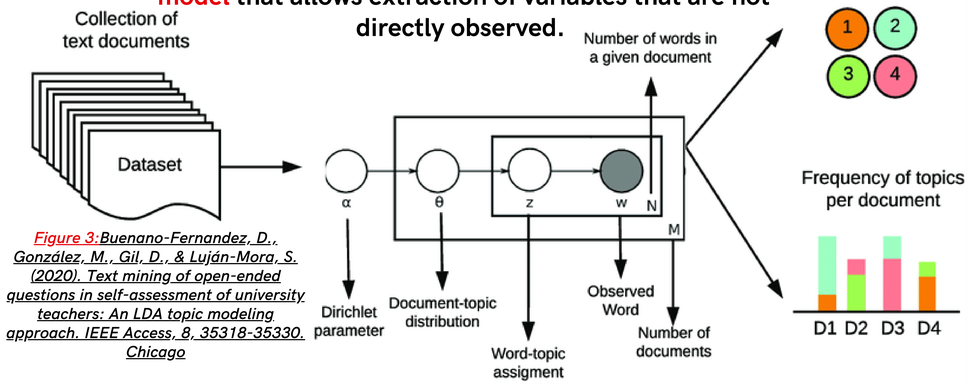


Figure 3: Buenano-Fernandez, D., González, M., Gil, D., & Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. IEEE Access, 8, 35318-35330. Chicago

Conclusion

By using computational analytical techniques on social science qualitative data we have advanced recent research on topic modeling by examining the case of newspaper media representations of water infrastructural development. We discovered 3 themes in our data: "INRLP Comparative Perspective: Ken-Betwa and other river links in India", "Environmental Conservation Perspective", and "Government Policy Perspective".

LIMITATIONS:

Choice of newspapers, size of the corpus, choice of topic model, performance metrics, and the requirement for a certain degree of human intervention and perspective.