A Topic Model Analysis of Agricultural and Applied Economics Association Publications from 2003-2024

Darsh Mandera, Abdalrahman Alalwan, Haozhou Zhang, Poojitha Balamurugan, Norbert Wilson, Leslie Collins, and Boyla Mainsah

Duke University
Contact: Norbert.Wilson@duke.edu

Selected Poster prepared for presentation at the 2025 AAEA & WAEA Joint Annual Meeting in Denver, CO: July 27-29, 2025

A Topic Model Analysis of Agricultural and Applied Economics Association Publications from 2003-2024

Darsh Mandera, Abdalrahman Alalwan, Haozhou Zhang, Poojitha Balamurugan, Norbert Wilson, Leslie Collins, Boyla Mainsah

Duke University | AMLL | Duke SANFORD WORLD FOOD POLICY CENTER | Data+

Introduction

Understanding trends in research topics, funding, and citations over time offers valuable insights for researchers. This study aims to identify trends in research topics and funding in agricultural and applied economics research projects. We analyze a corpus of research papers published in the *American Journal of Agricultural Economics* (AJAE) and *Applied Economics Perspectives and Policy* (AEPP), two journals with significant impact in the field of agricultural and applied economics. We aim to provide insights into research priorities over the last two decades.

Objectives

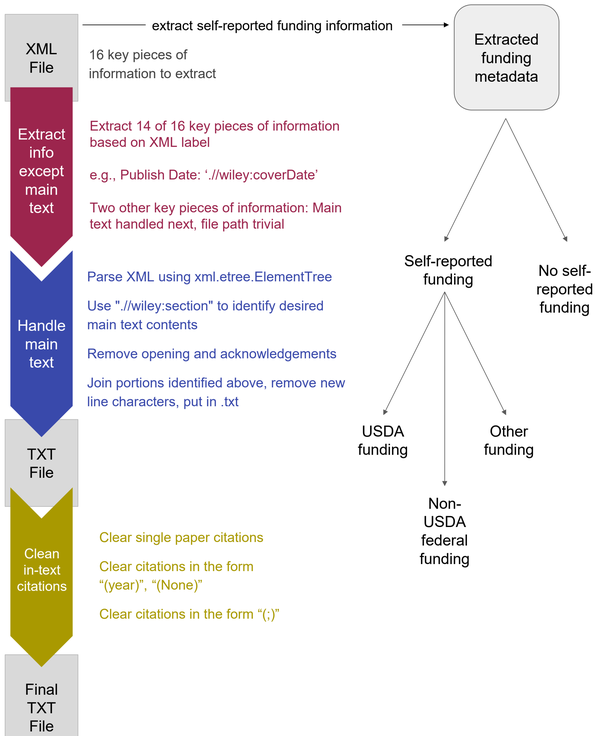
- Analyze a dataset of agricultural economics research papers to identify trends in research topics and funding patterns over time.
- Use natural language processing techniques for topic extraction.
- Perform time series analysis on the extracted topics to identify longitudinal trends in research topics and funding.

Data

Published articles

- From Wiley, we accessed 4,102 XML files representing 4,102 publications from the *AJAE* and *AEPP*, spanning the period from February 2003 to June 2024.
- We identified 3,417 records of year-by-year citation counts for each document, identified by DOI, as extracted from the Web of Science.
- After filtering non-research articles (book reviews, listings) and aligning the filtered publication corpus with citation records based on DOI, the reduced dataset consists of 3,049 XML files.

Data extraction & cleaning



Procedure

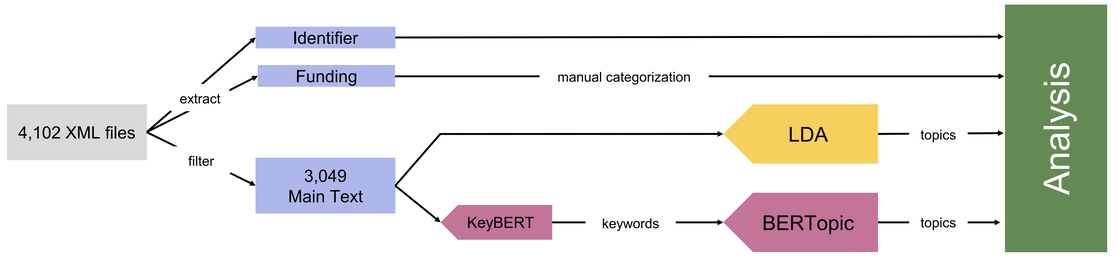
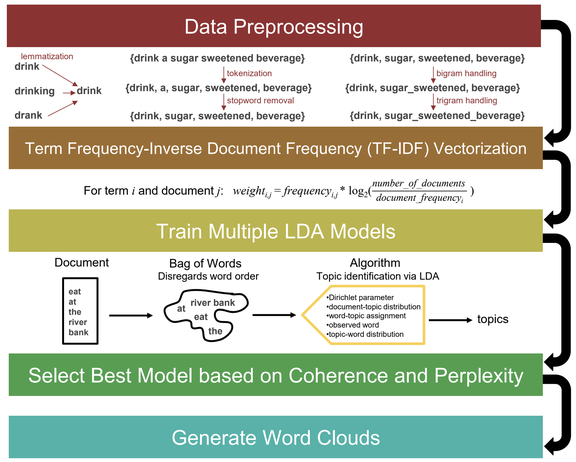


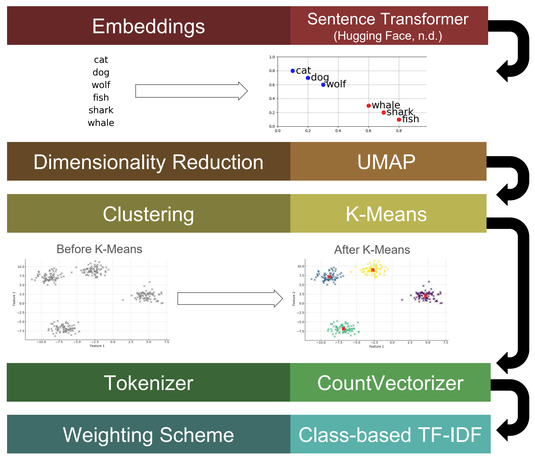
Figure 1: Flowchart of topic identification and data analysis

Methods

Latent Dirichlet Allocation (LDA)



BERTopic



Results

LDA

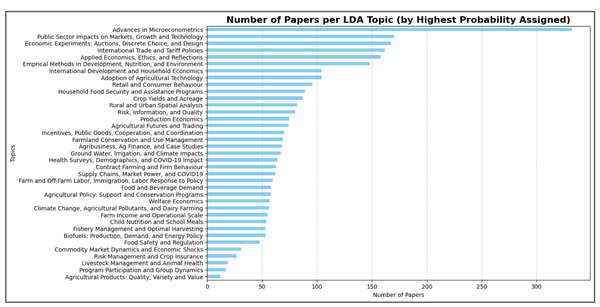


Figure 2: Number of papers attributed to LDA topics

Topic 8: Health Surveys, Demographics, and COVID-19 Impact

Topic 3: Livestock Management and Animal Health

Figure 3: LDA topic word clouds

BERTopic

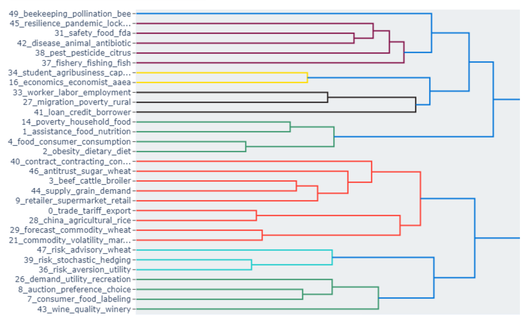
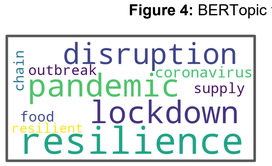


Figure 4: BERTopic topic hierarchical grouping



Topic 45: Pandemic Resilience and Supply Chains

Topic 2: Obesity and Dietary Patterns

Figure 5: BERTopic word clouds

- Extracted topics from each topic model were labeled by an internal subject matter expert.
- LDA is more effective at capturing the thematic content of each paper, providing a general overview of research themes, while BERTopic is more effective at capturing the semantic nuances, offering more coherent topics.
- A significant portion of papers in the corpus (total N = 3049) were attributed to the following topics: "Advances in Microeconometrics" (N = 332), "Public Sector Impacts on Markets, Growth and Trade" (N = 170), and "Economic Experiments: Auctions, Discrete Choice Models, and Incentives" (N = 167).

Longitudinal Trend Analysis

Research Topics

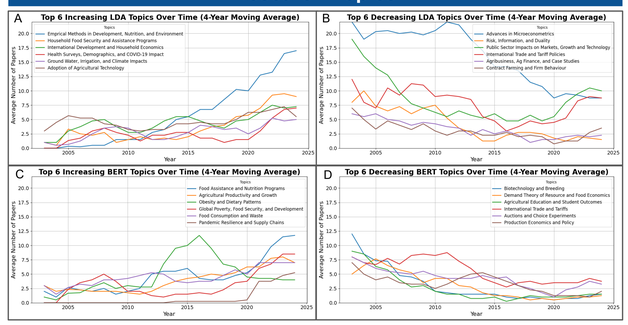


Figure 6: Longitudinal prevalence of topics grouped by increasing and decreasing trends

- We discover similarities in topic themes and trends of topics extracted with both topic models:
  - We observe an increase in papers on "Household and Development Economics" and related topics over the past two decades.
  - We also see a noticeable growth of COVID-19-related topics in 2020, which include "Health Surveys, Demographics, and COVID-19 Impact" and "Pandemic Resilience and Supply Chains."
- We identify no coherent themes in topics that decrease in frequency over time.

Funding Reporting

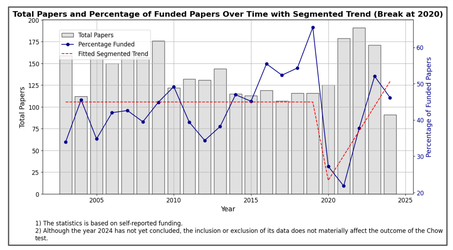


Figure 7: Trends in yearly papers published and funding proportion

- Figure 7 shows the total number of papers per year and the proportion of papers per year with reported funding.
- We hypothesized that there is a structural break in the percentage of papers reported receiving research funding around the year 2020, potentially due to the impact of the COVID-19 pandemic.
- The Chow test of structural stability reveals a structural break in the proportion of papers with reported funding ($p<.001$), with a breakpoint around 2020.

Conclusion and Limitations

- We find consistency in the themes and longitudinal trends of topics extracted by both topic models across the *AJAE* and *AEPP*.
- The impact of the 2020 COVID-19 pandemic was reflected in changes in patterns in published research topics and funding reporting in agricultural and applied economics.
- A limitation is the need for independent topic label assignment by external subject matter experts.
- Future work will analyze the relationships between topics, funding, and citations.

References

Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with Python*. O'Reilly Media, Inc.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993-1022. https://jmlr.csail.mit.edu/papers/v3/blei03a.html

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://arxiv.org/abs/1810.04805

Grootendorst, M. (2020). *KeyBERT: Minimal keyword extraction with BERT* (Version 0.3.0). Zenodo. https://doi.org/10.5281/zenodo.4461265

Hugging Face. (n.d.). *all-MiniLM-L6-v2* [Software]. Hugging Face. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

Jockers, M. (2021). Expanded stopwords list. Retrieved June 15, 2024, from https://www.matthewjockers.net/macroanalysisbook/expanded-stopwords-list

Kvamsdal, S. F., Belik, I., Hopland, A. O., & Li, Y. (2021). A machine learning analysis of the recent environmental and resource economics literature. *Environmental and Resource Economics*, *79*(1), 93-115. https://doi.org/10.1007/s10640-021-00554-0

Pattnaik, D., Kumar, S., Burton, B., & Lim, W. M. (2022). Economic Modelling at thirty-five: A retrospective bibliometric survey. *Economic Modelling*, *107*, 105712. https://doi.org/10.1016/j.econmod.2021.105712

Princeton University. (2010). *About WordNet*. WordNet. Princeton University. https://wordnet.princeton.edu/citing-wordnet

Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2).

Yin, J., & Wang, J. (2014). A Dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 233-242). ACM. https://doi.org/10.1145/2623330.2623715

Acknowledgements

WILEY | AAEA Agricultural & Applied Economics Association | DUKE UNIVERSITY LIBRARIES