Reflecting Populist Tensions over Climate Change Policy in Academic Discourse

Piotr Cichocki

Adam Mickiewicz University, Poznan, Poland

Presentation Agenda

- Algorythmic text mining of academic papers using topic modelling on climate change and climate policy
- Common project with the following paper by Mariusz Baranowski focusing on substantive results
- We decided to coordinate our presentations so that I focus more on the background of the project, data treatment and methodology, and Mariusz presents the substantive results

Project description (1)

„Determinants of populist opposition to climate policy"

Research project financed by:
Polish National Science Centre
2020/37/B/HS6/02998

- **Explaining the role of structural or ideological factors in generating opposition to climate policy**
- **Main focus: secondary analysis of cross-national survey data and qualitative research in selected countries**
- **Supplementary analysis: text mining of academic discourse both as a part of the literature review and to incorporate discourse variables into the main analysis**

Project description (2)

Literature review: the narrow path

- Specific keywords, snowballing
- Relatively few results, manageable by traditional means of literature review
- Lacks breadth and is prone to terminological shifts, e.g., the case of populism

Literature review: the broad path

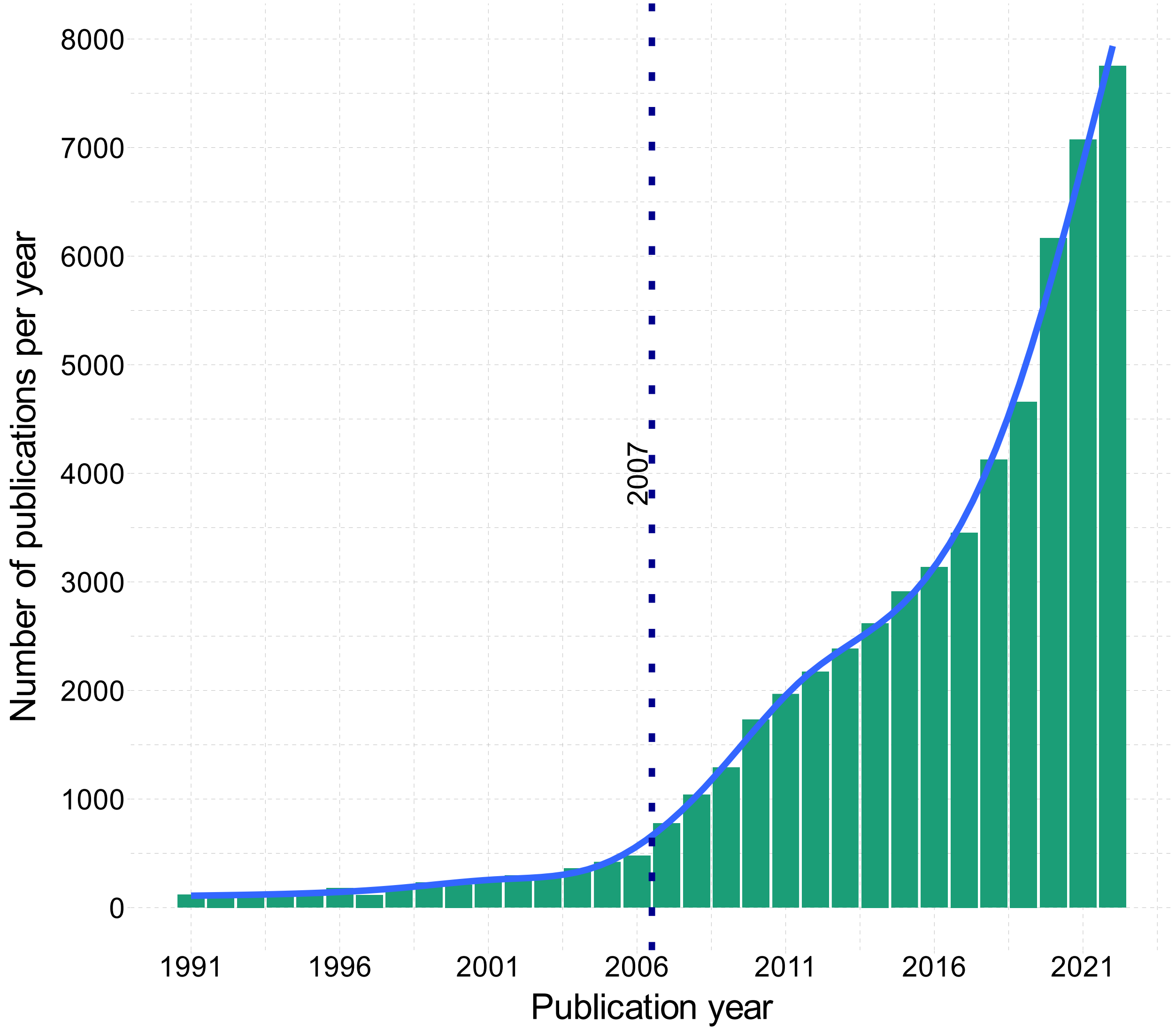
- General keywords a wide net
- A high number of records, impractical to handle by traditional means in its entirety
- Broad exploratory approach allowing the identification of unknown patterns

Text mining of SCOPUS records

- Recent advances in Natural Language Processing (NLP) and Machine Learning (ML) facilitated a rapid expansion in the possible applications of Text Mining
- Large-scale analyses of literature employ NLP for information extraction, i.e., pulling out structured information from unstructured text data
- Latent Dirichlet Allocation (LDA) as a statistical model uncovers hidden thematic structures in large collections of documents, enabling the automatic discovery of topics
- Topic modelling typically works within the traditional “bag-of-words" paradigm of NLP, which represents a document as an unordered list of tokenised words
- Topics are identified by the probability distribution of tokens within a given vocabulary, and each document has an assigned probability of association with every topic based on the topics it holds in the “bag”

Scopus search query (16/06/2023):
TITLE-ABS-KEY("climate change")
OR
TITLE-ABS-KEY("climate crisis")
OR
TITLE-ABS-KEY("global warming")
AND
PUBYEAR > 1990 AND
PUBYEAR < 2023 AND
SRCTYPE(j) AND
(LIMIT-TO(SUBJAREA , "SOCI")) AND
(LIMIT-TO(LANGUAGE , "English")) AND
(LIMIT-TO(PUBSTAGE , "final"))

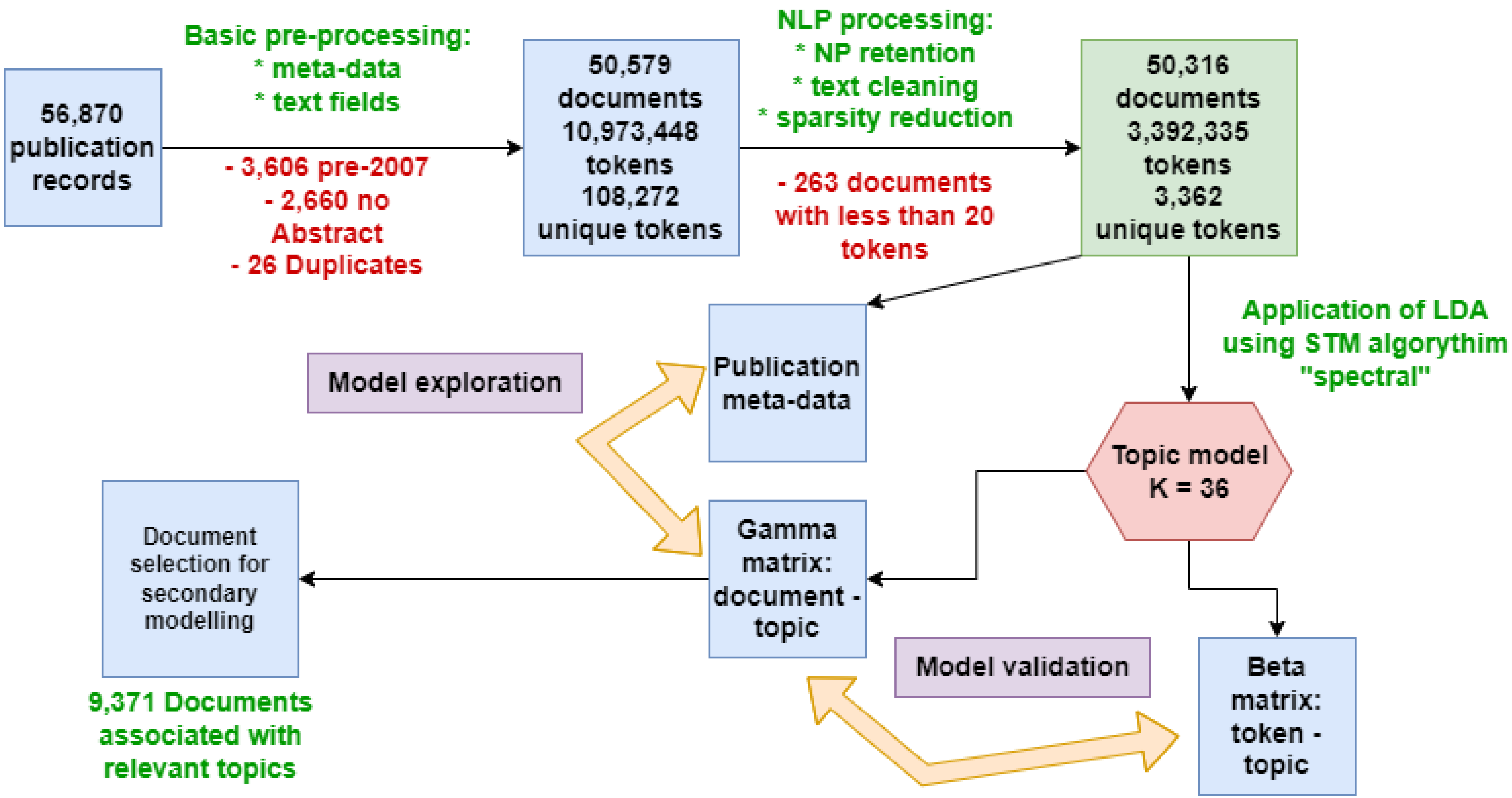
- **56,870 publication records**
- Access to content description through the Abstract and Title fields
- Keywords inconsistently applied and may lead to bias
- Access to publication metadata, e.g., publication year, author affiliations, declared funding



Temporal cut-off:

- A precipitous rise in the number of publications
- Much more than the usual inflation in academic publishing
- Cut-off applied at >= 2007, so as to focus on the last 15 years of high prominence

Corpus pre-processing for the topic modelling



Overview of the primary topic model

- **Exploratory analysis:** aiming at a reasonably exhaustive inventory of topics
- **Using K = 36 setting within STM:** following exploratory analysis
- **Primary LDA use:** selection of documents associated with discussions of climate change attitudes

DTM → **LDA**

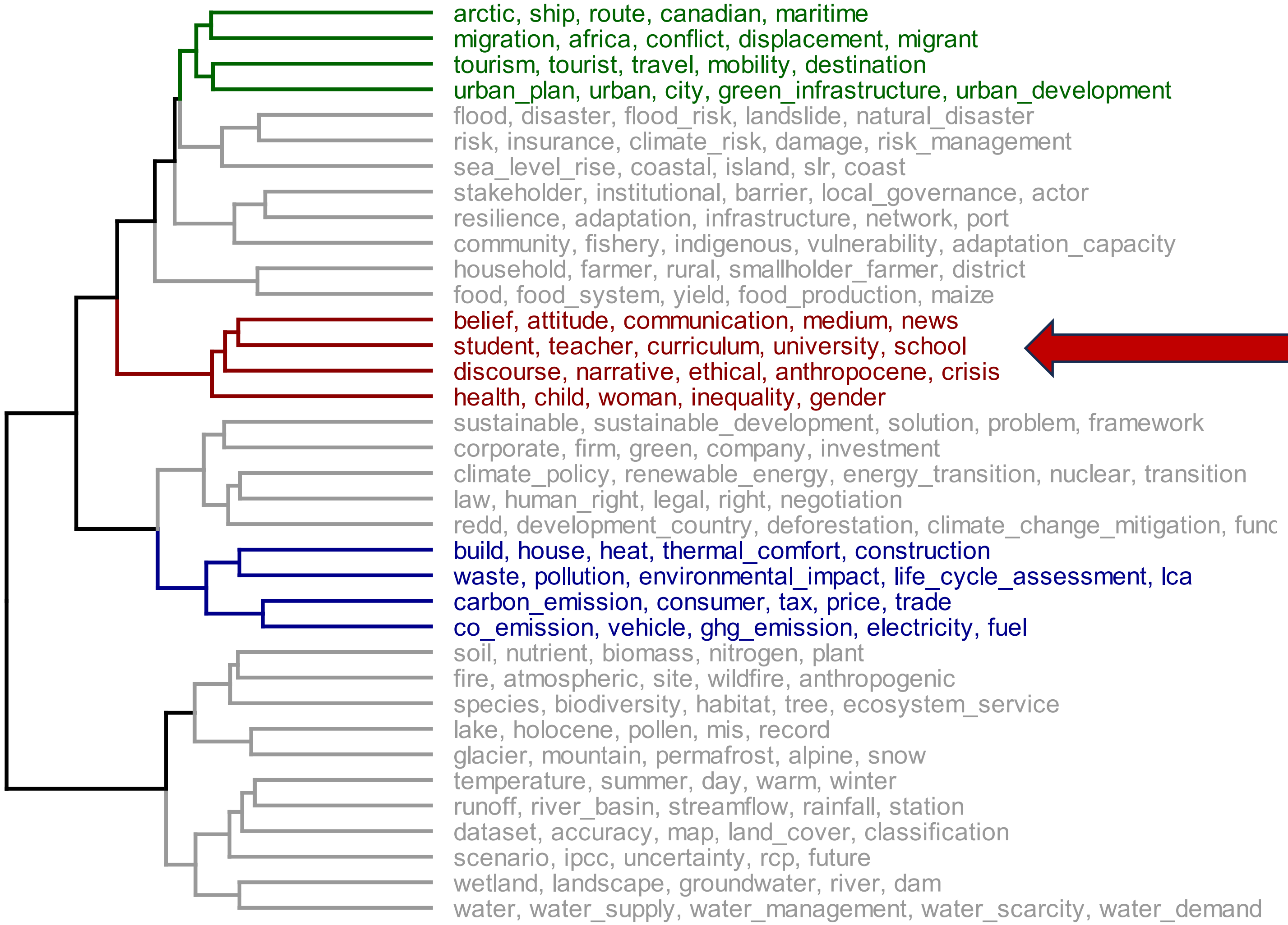
Χ – Measure the presence of a subject in a document

β – probability word affiliation to the topic

```
LDA(DTM_for_modelling, k = N,
    method = "Gibbs",
    control = list(alpha = 50/k,
                   delta = 0.01,
                   iter = 2000,
                   seed = 1981,
                   thin = 1))
```

- LDA takes DTM as input (does not see tokens, only indexes)
- The LDA algorithm is subject to control parameters that affect the output
- The key problem is the top-down determination of the number of K-themes

Dendrogram - Hierarchical clustering of 36 topics based on within-document co-occurences



Our cluster of topics for document selection

Topics included in document selection for secondary LDA

Labels
belief, attitude, communication, medium, news, public, frame, citizen, behavior, opinion, message,trust, emotion, intention, motivation
student, teacher, curriculum, university, school, science, learn, education, teach, professional, knowledge, course, skill, interdisciplinary, literacy
discourse, narrative, ethical, anthropocene, crisis, climate_crisis, conte mporary, notion, essay, idea, debate, movement, neoliberal, cultural, society
health, child, woman, inequality, gender, public_health, man, covid, pandemic, disease, worker, poverty, youth, covid_pandemic, well being

What does "document selection" mean

```
tidy(stm_out, matrix = "gamma", document_names = docnames(dtm_large)) %>%
  group_by(document) %>%
  slice_max(order_by = gamma, n = 3, with_ties = FALSE) %>%
  mutate(rank_order = rank(gamma, ties.method = "random")) %>%
  ungroup() %>%
  arrange(document,desc(rank_order)) %>%
  group_by(document) %>%
  mutate(quotient = lag(gamma, n = 1)/gamma) %>%
  filter(is.na(quotient) | quotient <= 1.25) %>%
  filter(rank_order %in% c(2, 3) | rank_order == 1 & 2 %in% rank_order) %>%
  ungroup() %>%
  filter(gamma >=0.15) %>%
  filter(topic %in% c(10, 29)) %>%
  pull(document)
```

Documents selected

- [1:9648] 1 8 9 19 34 35 43 44 59 67

Out of 9648 documents:

- 9094 „belong” to a single topic
- 277 to both topics

Topic 10 (discourse): 6190
Topic 29 (attitudes): 3458

Secondary LDA is performed on the subset to reveal its topical structure

- An alternative approach to generating many topics outright
- Useful for initial screening, especially when access to full texts in machine-readable texts is not straightforward

	1	2	3	4	5	6	7
1	0,00375	0,00017	0,00094	0,00756	0,00708	0,00078	0,00011
2	0,00874	0,00227	0,00637	0,00694	0,00616	0,00268	0,00121
3	0,00771	0,0246	0,01412	0,02141	0,0136	0,00696	0,00255
4	0,01023	0,00552	0,22863	0,00192	0,00256	0,26149	0,0083
5	0,01405	0,05194	0,00615	0,00451	0,00096	0,01411	0,3484
6	0,04759	0,00044	0,00234	0,07026	0,01438	0,00309	0,00074
7	0,03901	0,00135	0,00746	0,01296	0,1147	0,00353	0,00069
8	0,02054	0,01058	0,01713	0,00247	0,0025	0,07999	0,01318
9	0,00834	0,00115	0,00274	0,07192	0,24636	0,0629	0,00062
10	0,26488	0,03777	0,04114	0,02011	0,00229	0,05228	0,03927
11	0,07615	0,00108	0,01401	0,00205	0,06451	0,01317	0,02877
12	0,00969	0,0045	0,0079	0,0032	0,08532	0,01812	0,00722
13	0,00468	0,00038	0,00223	0,00935	0,01224	0,0006	0,00014
14	0,01478	0,00136	0,00649	0,01141	0,0474	0,01595	0,00475
15	0,00559	0,00295	0,04815	0,00355	0,00414	0,00693	0,00098
16	0,00681	0,00453	0,01443	0,00476	0,00447	0,02322	0,00407
17	0,00715	0,02829	0,18518	0,00153	0,00525	0,00531	0,00175
18	0,0197	0,00148	0,00269	0,09101	0,00971	0,00206	0,00045
19	0,00476	0,00164	0,00435	0,00178	0,05039	0,0042	0,00113
20	0,00339	0,00294	0,00725	0,00381	0,01898	0,0017	0,00031
21	0,00558	0,00089	0,00287	0,1357	0,005	0,00335	0,00112
22	0,00754	0,00014	0,00079	0,25555	0,06272	0,00096	0,00013
23	0,22593	0,06862	0,10145	0,01053	0,00571	0,0297	0,1187
24	0,00312	0,00107	0,00297	0,00497	0,0048	0,00163	0,00023
25	0,00661	0,00019	0,00126	0,08968	0,10894	0,00258	0,00036
26	0,02078	0,6982	0,04285	0,0071	0,00342	0,03827	0,03448
27	0,00704	0,00029	0,0042	0,0217	0,04436	0,00284	0,0002
28	0,00524	0,01016	0,00721	0,0405	0,00557	0,0095	0,00136
29	0,02234	0,01471	0,01	0,00565	0,00259	0,17016	0,02285
30	0,0077	0,0024	0,04727	0,00157	0,00554	0,06604	0,00324
31	0,02047	0,00111	0,01162	0,0025	0,00437	0,04156	0,01609
32	0,03969	0,00396	0,04118	0,00984	0,01256	0,00604	0,00327
33	0,01634	0,00393	0,02323	0,0439	0,01192	0,00549	0,00161
34	0,02419	0,00592	0,0637	0,00188	0,00197	0,03418	0,32967
35	0,0014	0,00227	0,00299	0,01132	0,00265	0,00385	0,00065
36	0,0085	0,00118	0,0167	0,00511	0,00485	0,0048	0,0014

Thanks for your attention