

Emerging Trends of Sustainability Reporting in the ICT Industry: Insights from Discriminative Topic Mining

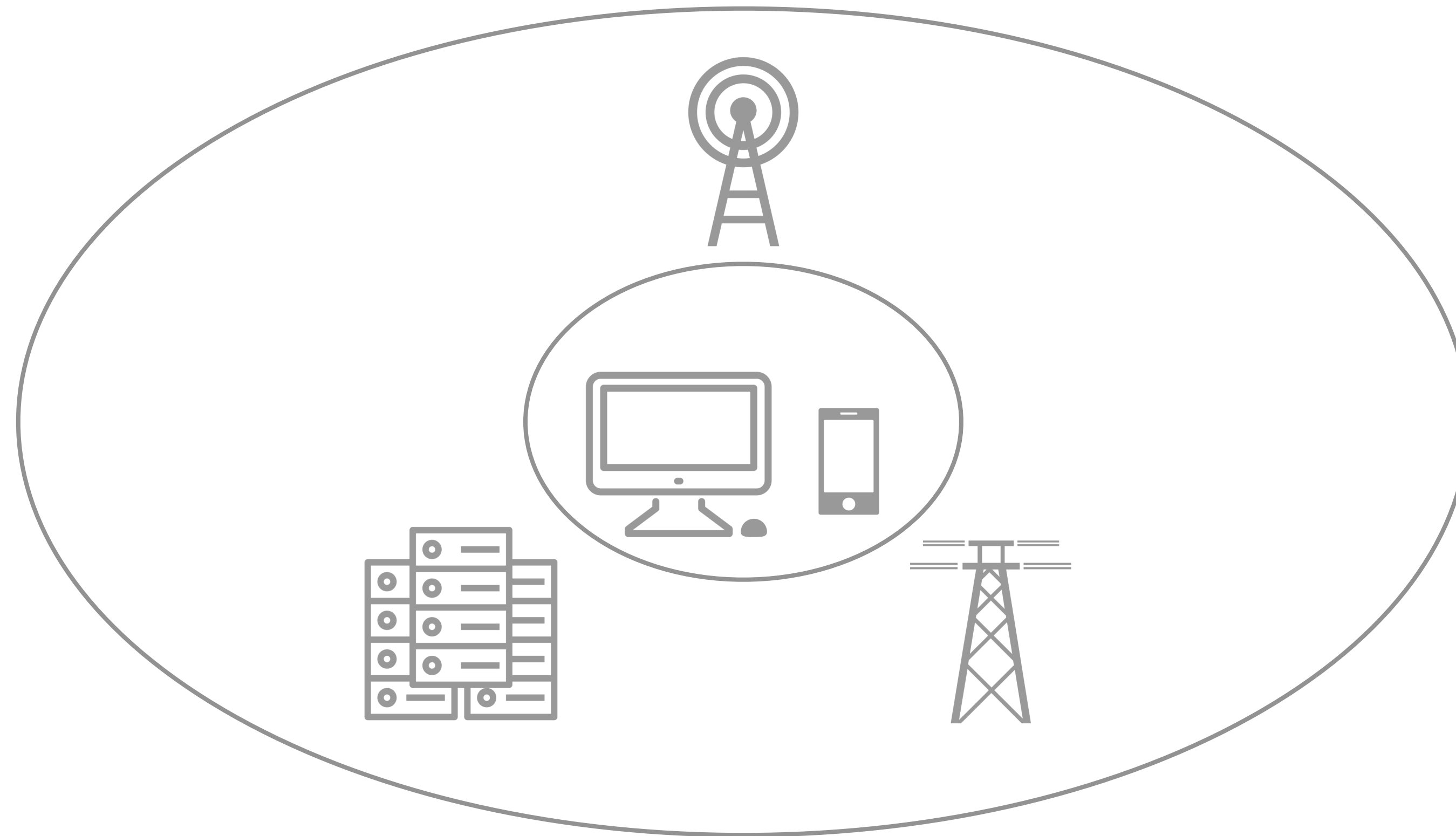
Lin Shi

linshi@stanford.edu

Nhi Truong

ntruongv@stanford.edu

ICT Industry and Climate Change



1-3%
Global GHG Emissions

How do ICT companies' sustainability reporting shift around **climate change**? Who are the important **institutional stakeholders** in their reporting?

Relevant work

- Prior work analyzes companies' sustainability reporting from the Global Reporting Initiative (GRI) between 1995 to 2005 using the Latent Dirichlet Algorithm (LDA) algorithm (Székely and Brocke, 2017)
- Pro: explore main topics covered in the reports
- Con: summarized topics are often very broad and generic

Our work

- Scraping, cleaning, adding to GRI dataset for 1999 - 2019
- Analyzing specifically the ICT industry
- Applying discriminative topic modeling to uncover specific narratives

Dataset: The Global Reporting Initiative (GRI)



Full Dataset:

22,534 environmental reports by
8,088 companies in 118 countries
between 1999 to 2019



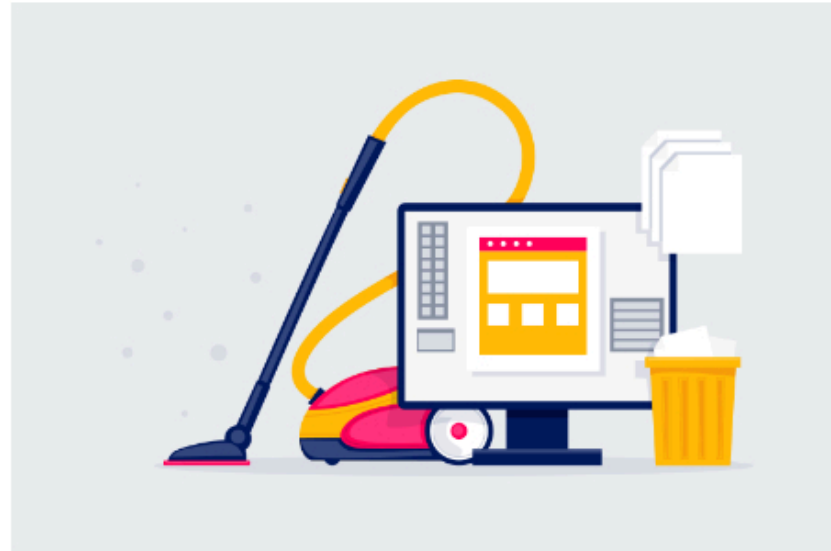
Our analysis:

ICT companies

2,187 environmental reports by
429 companies between 1999 to 2019

Method

Preprocessing



Remove:

- special characters
- stop words
- company names
- standalone numbers

Seed Phrase

- climate change
- sustainable products
- environmental health safety
- carbon footprint
- circular economy
- code of conduct
- life cycle assessment

...

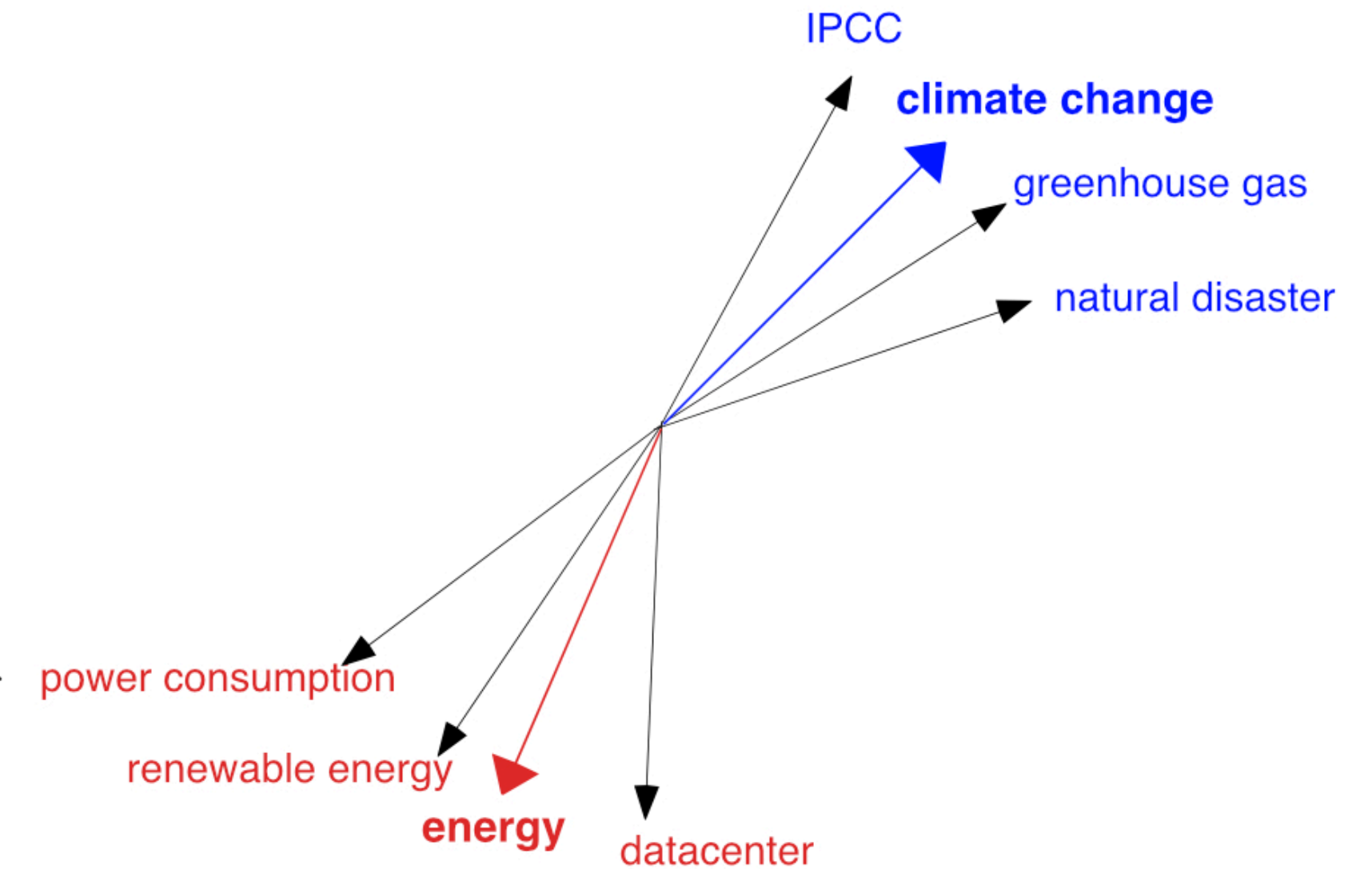
Phrasal Segmentation

AutoPhrase (Jingbo et al., 2018)

"management ensure effectiveness feasibility tackle identified risks prevention transfer risk post **disaster_recovery** **emergency_response** **early_warning** fire chemical disaster earthquake typhoon **climate_change** disruption information systems epidemic diseases corruption shortage human resources **financial_crisis** disruption supply system disruption **supply_chain**"

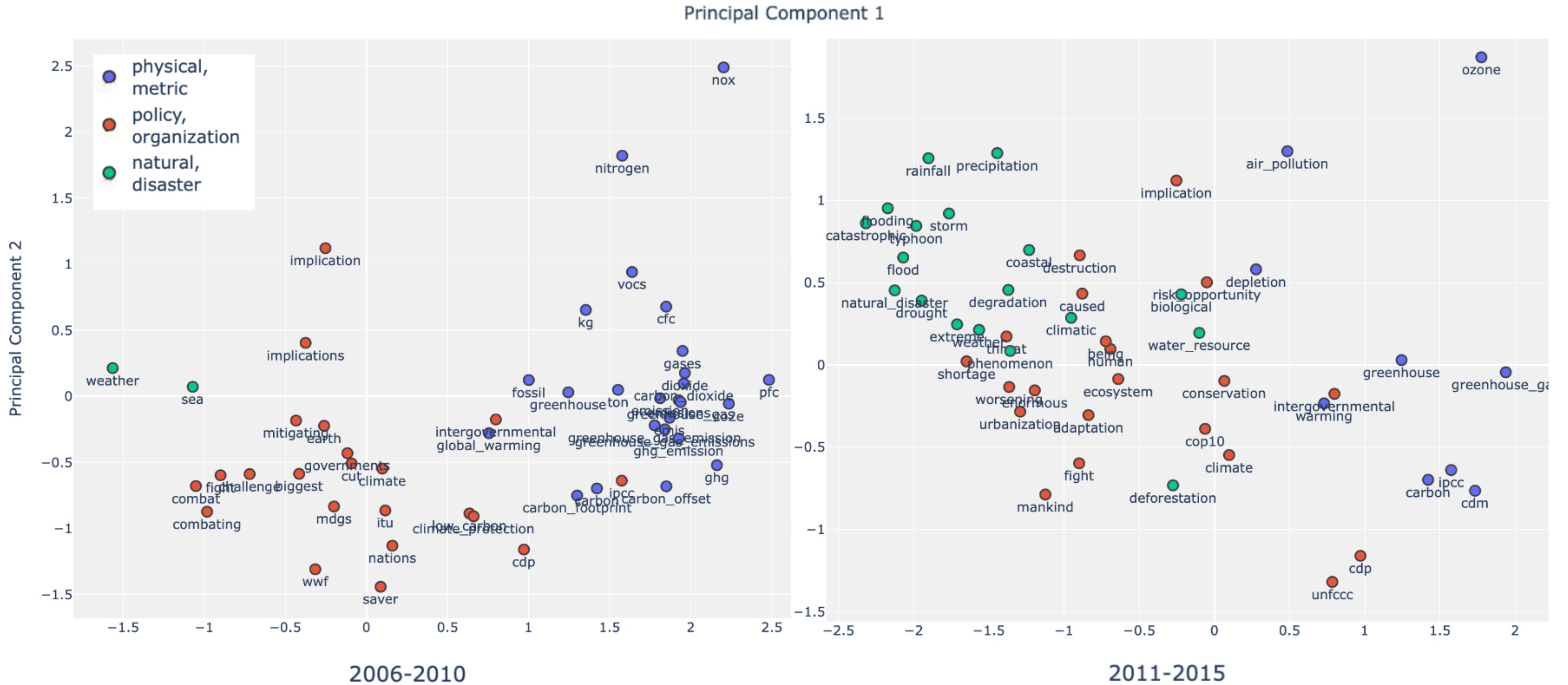
CatE

(Meng et al., 2020)

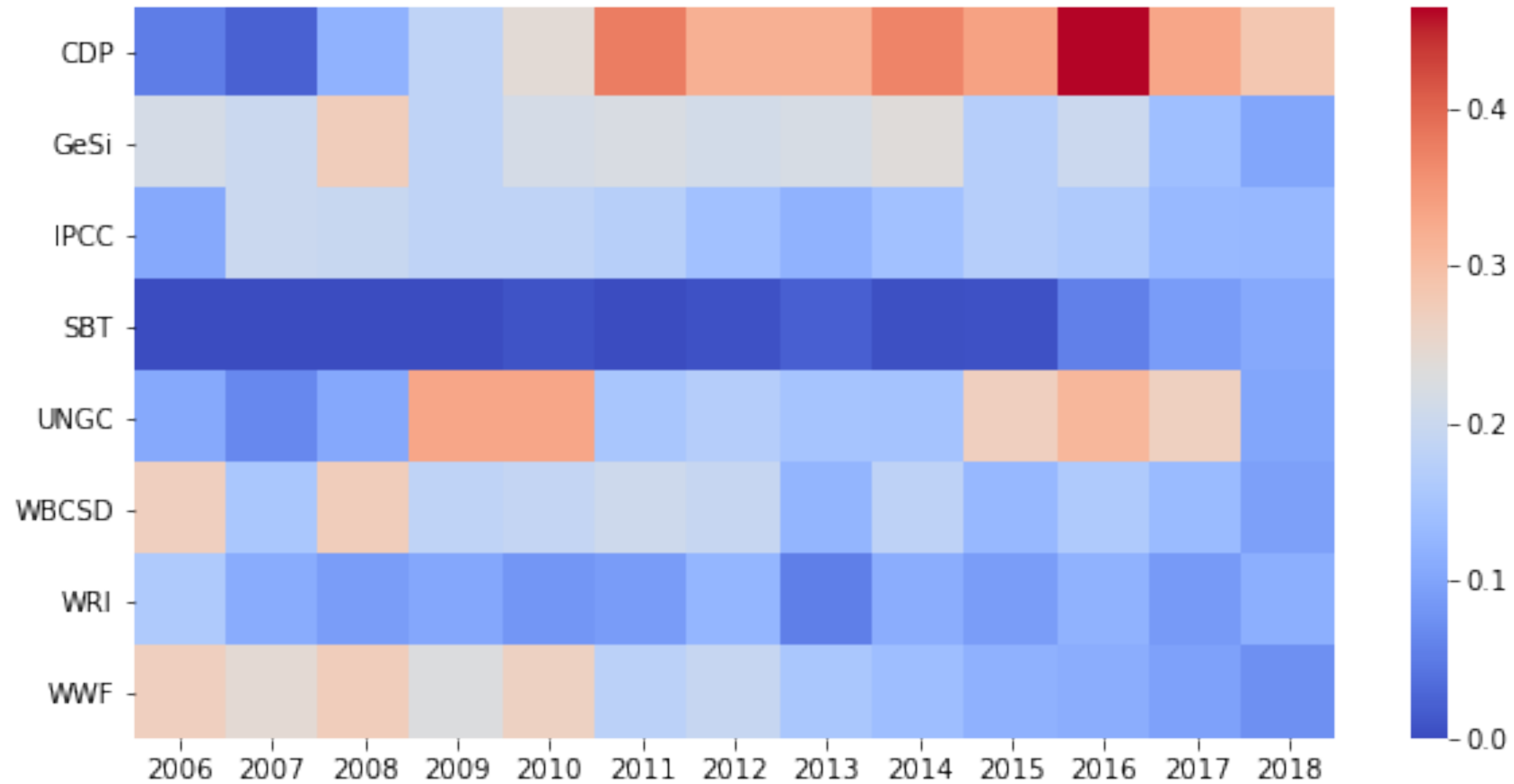


- provide topics: climate change, energy
- train 200 dimensional word embedding
- extract top 100 representative words per topic
- bootstrap 15 times
- choose top 10 runs per topic with highest coherence score
- extract top 100 common words per topic

Results: Climate change

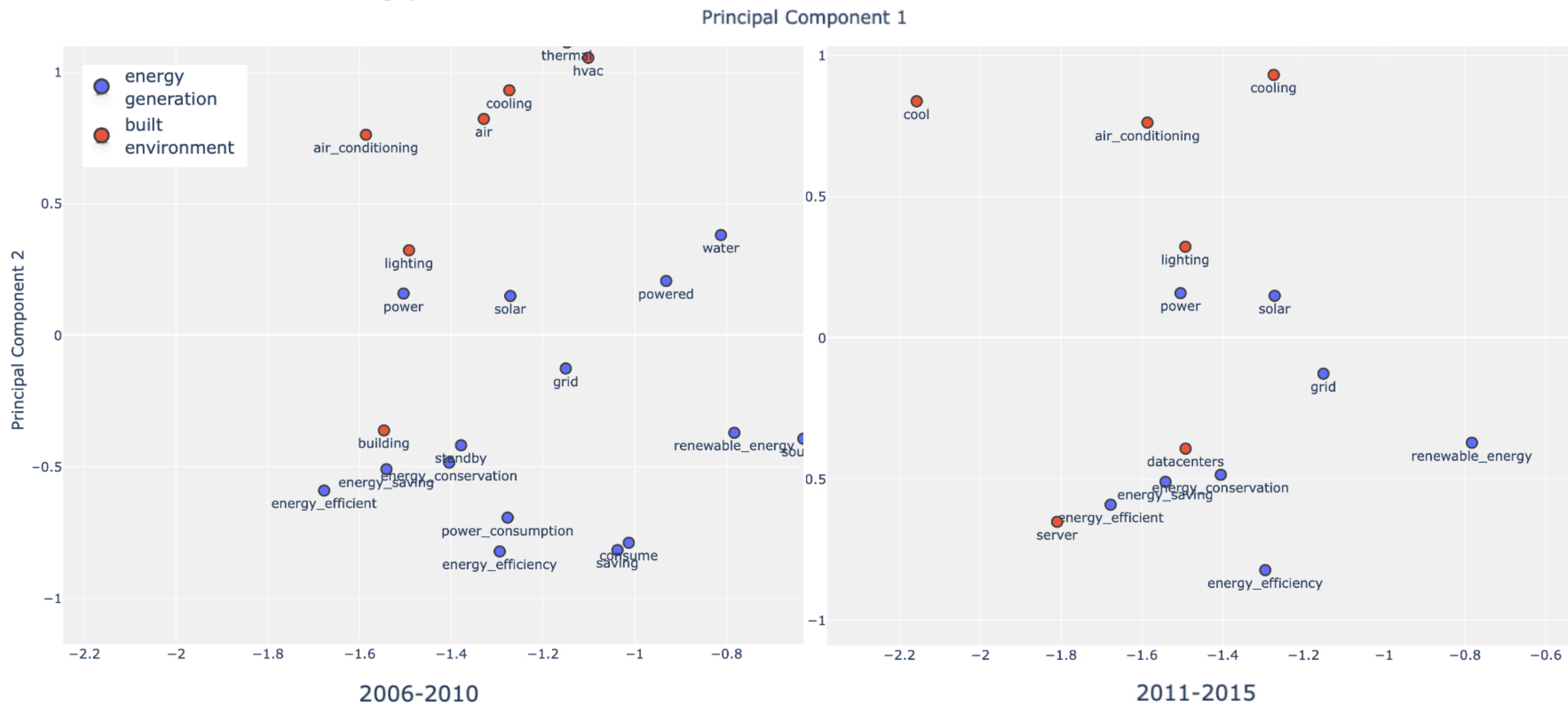


Results: Institutional Influence



percentage of documents mentioning organizations over years

Results: Energy



Future work

- Apply this framework to other ICT relevant topics (e.g. certification, audits)
- Compare the ICT sector to other sectors that have longstanding history of sustainability practices (e.g. food, textile, and forest products)
- Compare machine-based text analysis with manual content analysis on companies' specific sustainability practices

Thank you!