

Observing the evolution of the climate change debate across different public spheres

Monika Ewa Rakoczy, Robin Lamarche-Perrin,
Armin Pournaki, Nina Varchavsky



SORBONNE
UNIVERSITÉ



Media spheres

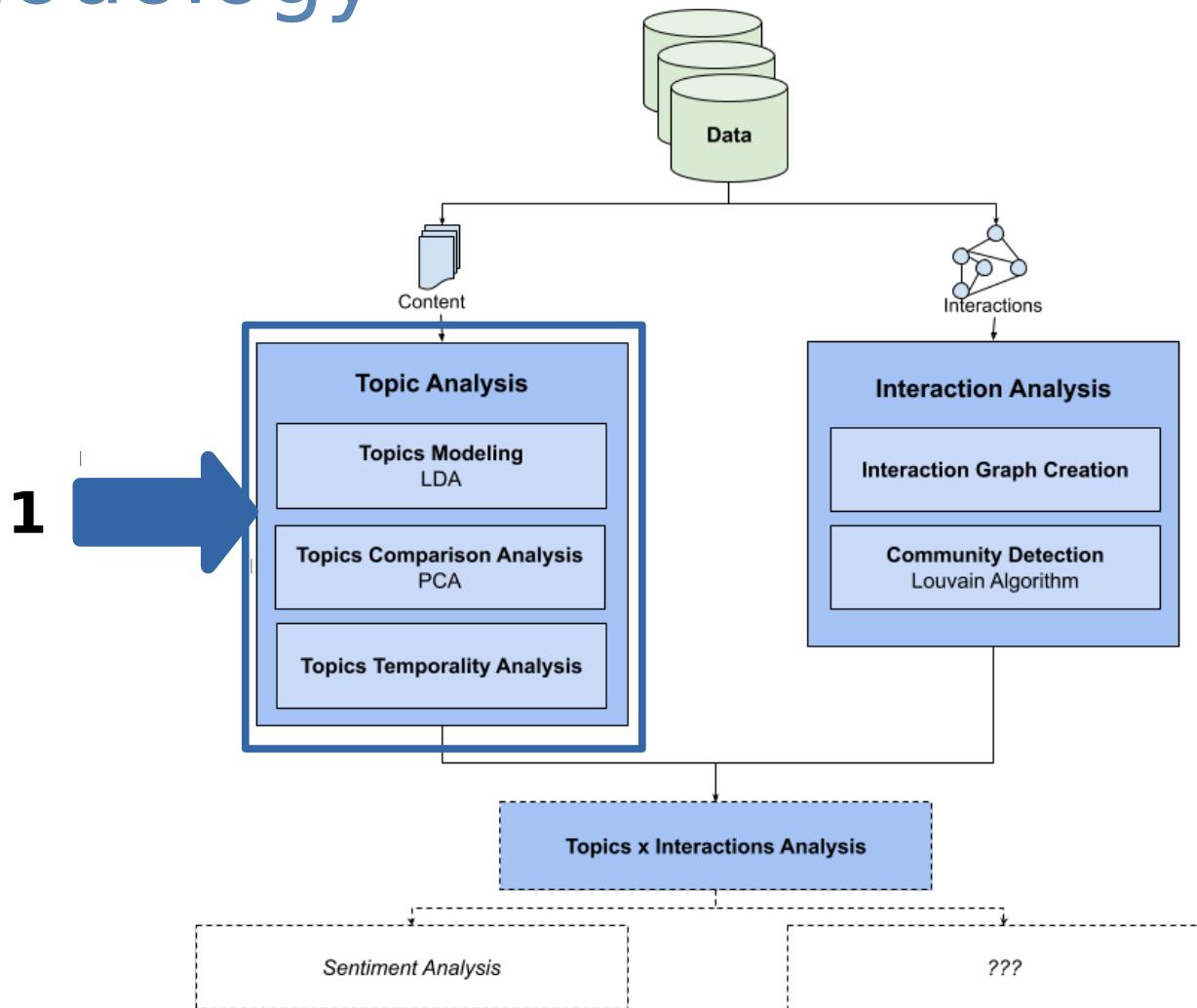
Dataset	Media sphere	Time span	Volume
Twitter	Social media	01/2016 – 04/2019	26.7M
Guardian	Mass media	01/2016 – 04/2019	79K
UK Parliament Speeches	Political	01/2016 – 12/2019	2.6K

All filtered by the use of “climate change”

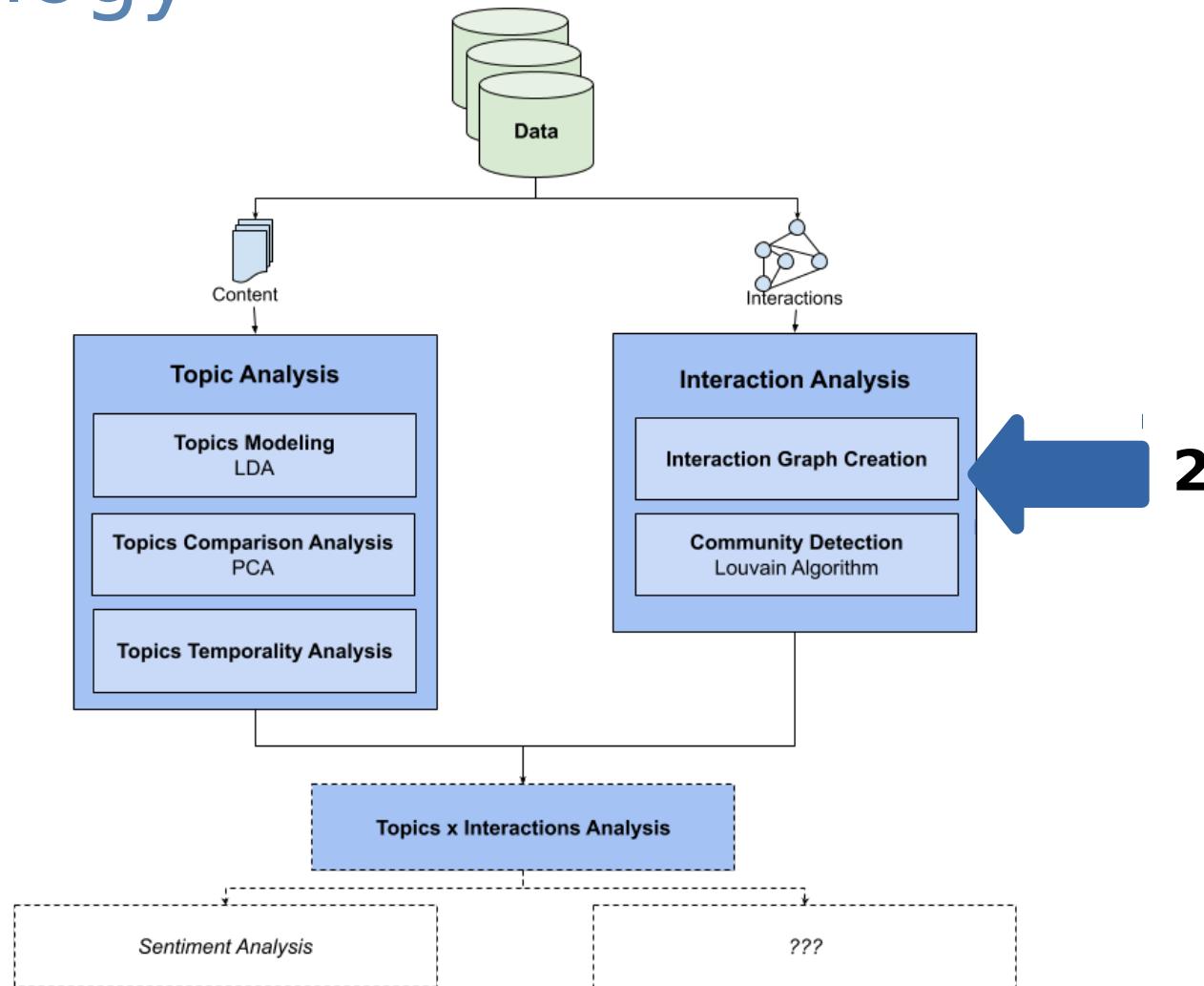
Research Questions

- **How is climate change discussed in different media spheres?**
 - What particular topics are discussed by each of them?
 - How does the used vocabulary differ in each of the spheres?
 - How are climate-related events treated by them?
 - What kind of interaction patterns do we observe in such discussions?

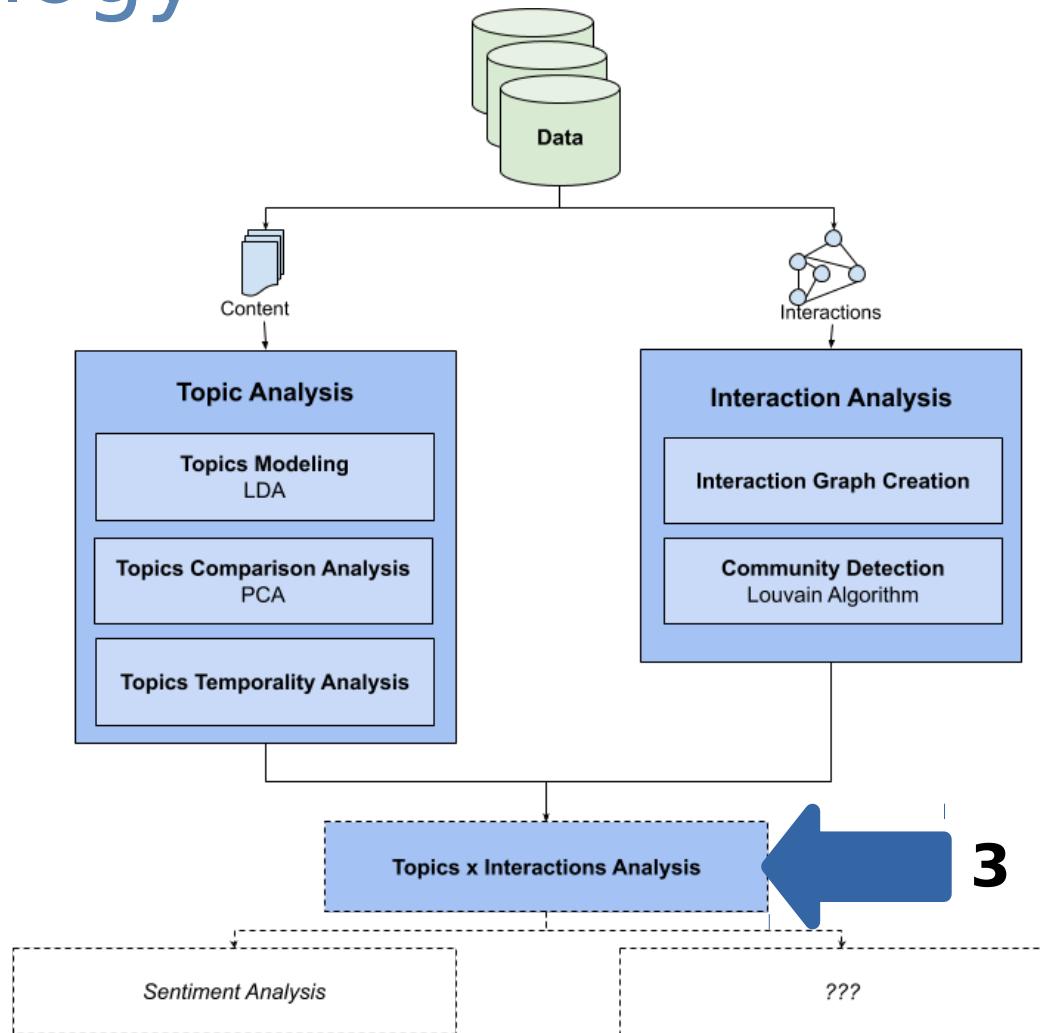
Methodology



Methodology



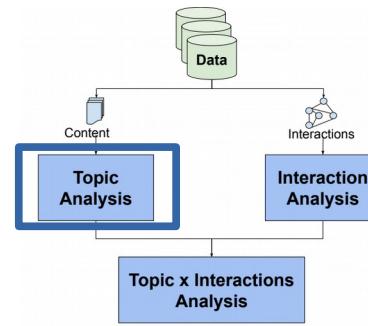
Methodology



Topic Modeling

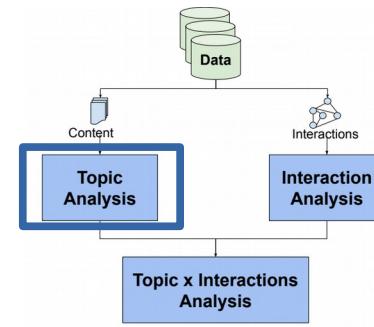
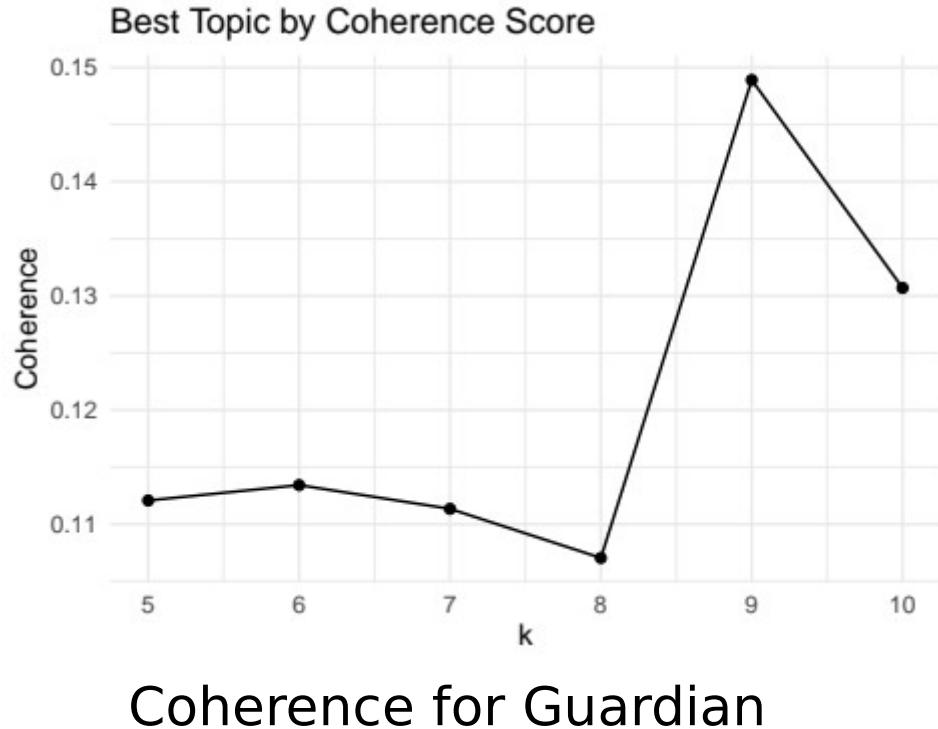
- **Latent Dirichlet Allocation (LDA)**
- Training LDA model using topic numbers from 5 to 10
- Smaller number of topics that human can intuitively understand/verify while reading a document

Dataset	Document
Twitter	Tweet
Guardian	Article
UK Parliament Speeches	Speech



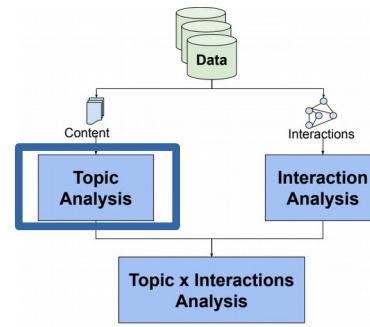
Topic Modeling

- Determining the “correct” topics number:
 - Coherence score

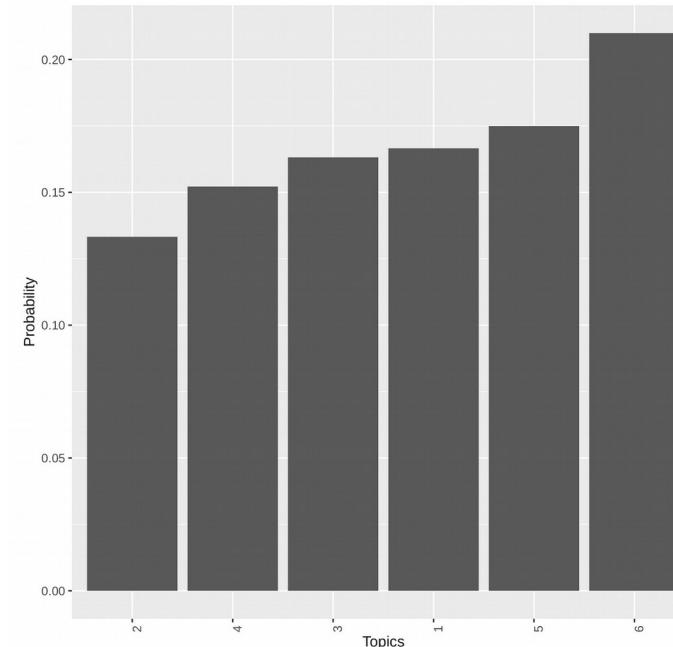


Topic Modeling

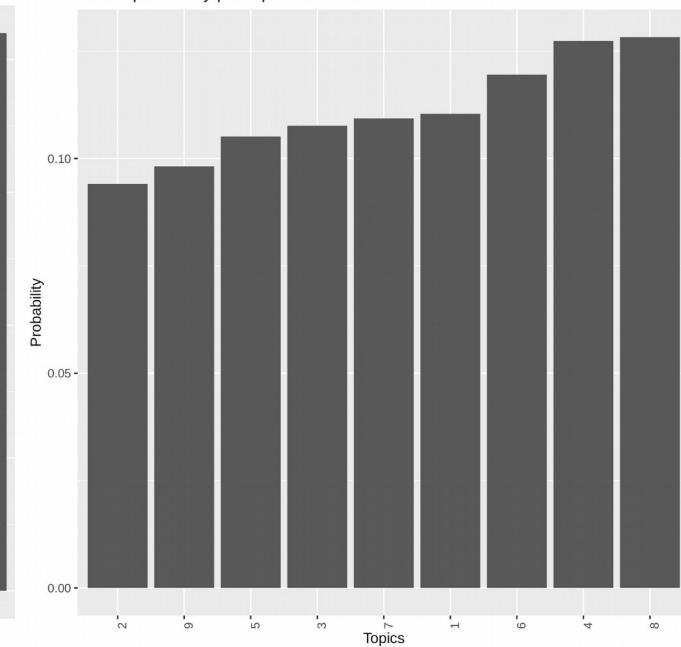
- Determining the “correct” topics number:
 - Coherence score
 - Global probability of topic (are the topics equally probable for all corpus?)



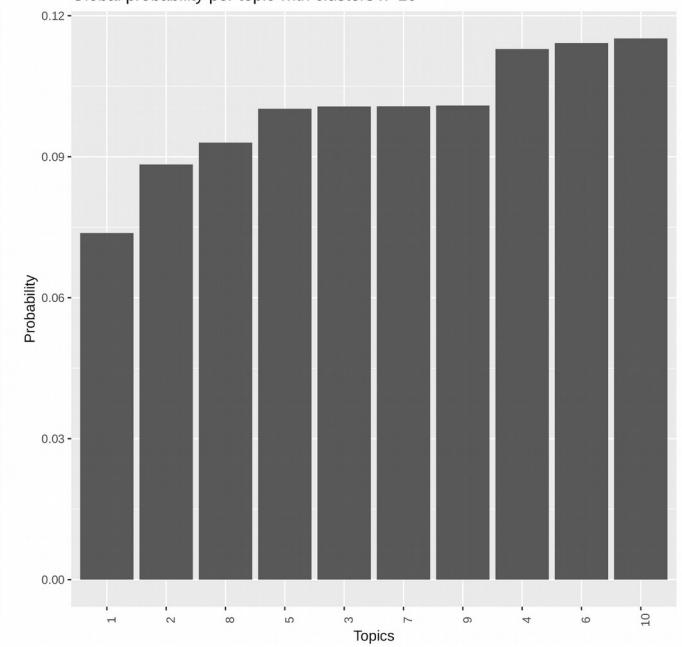
Global probability per topic with clusters k=6



Global probability per topic with clusters k=9

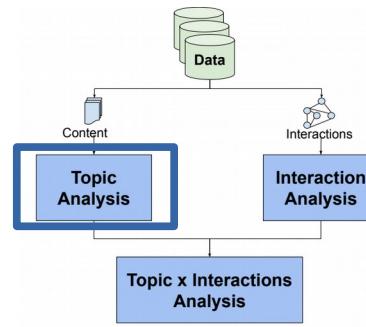
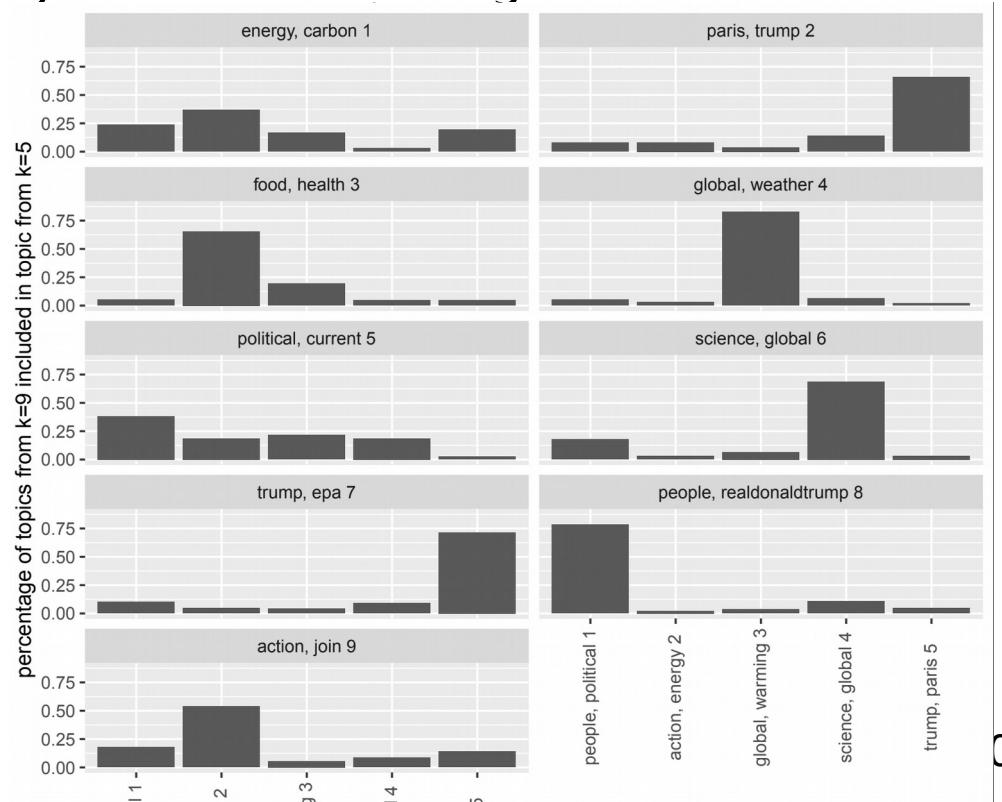
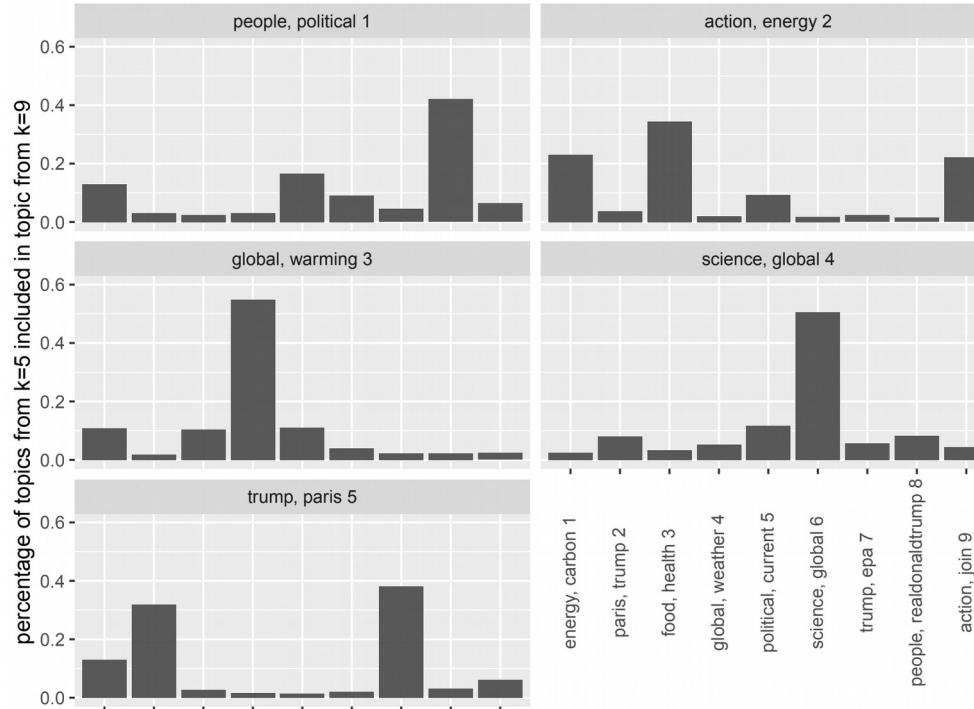


Global probability per topic with clusters k=10



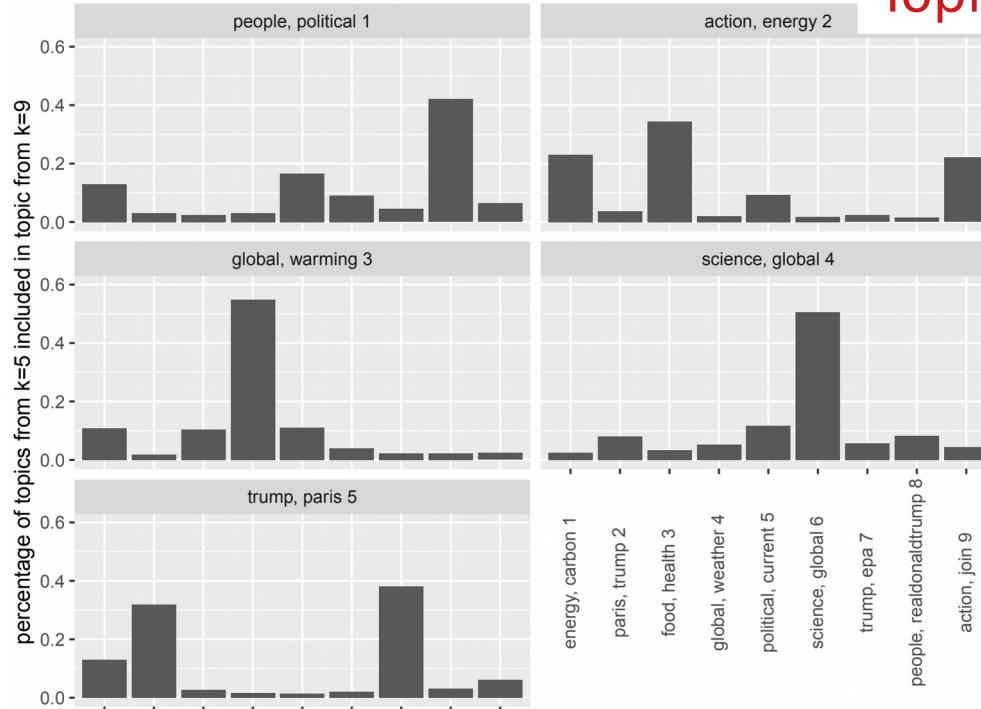
Topic Modeling

- Determining the “correct” topics number:
 - Similarity of topics for different number of clusters (are the topics split into two equally after increasing the number of clusters k?)

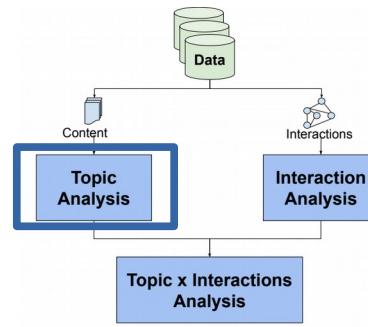
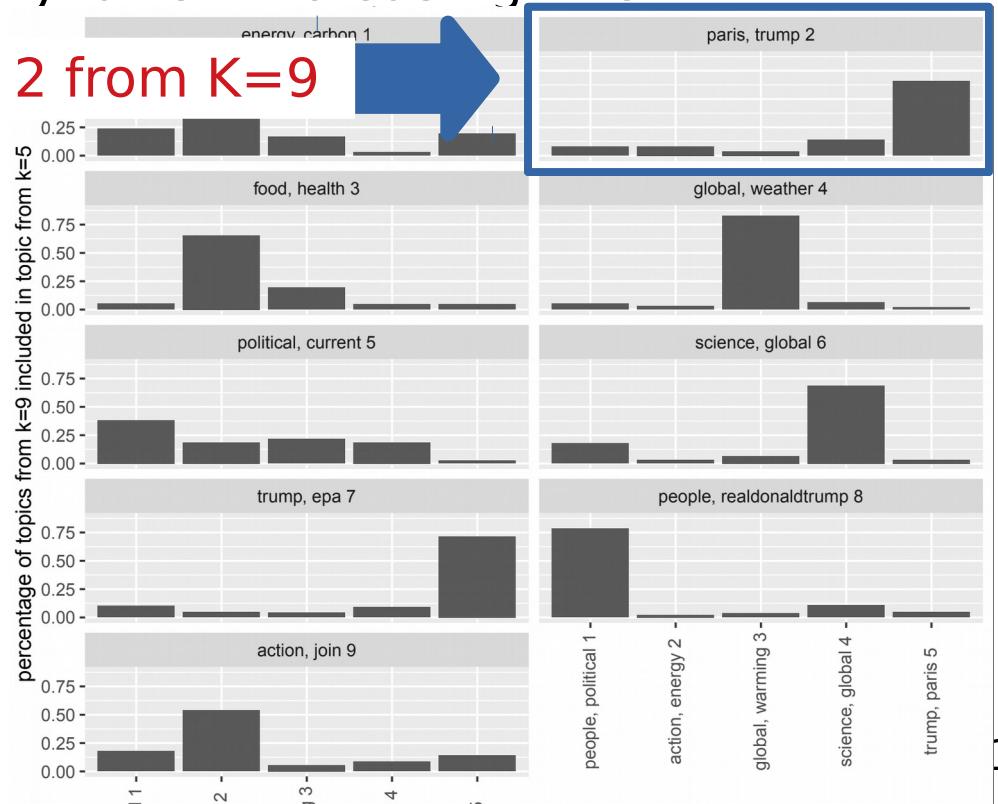


Topic Modeling

- Determining the “correct” topics number:
 - Similarity of topics for different number of clusters (are the topics split into two equally after increasing the number of clusters k?)

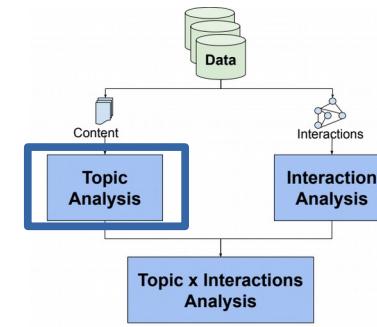
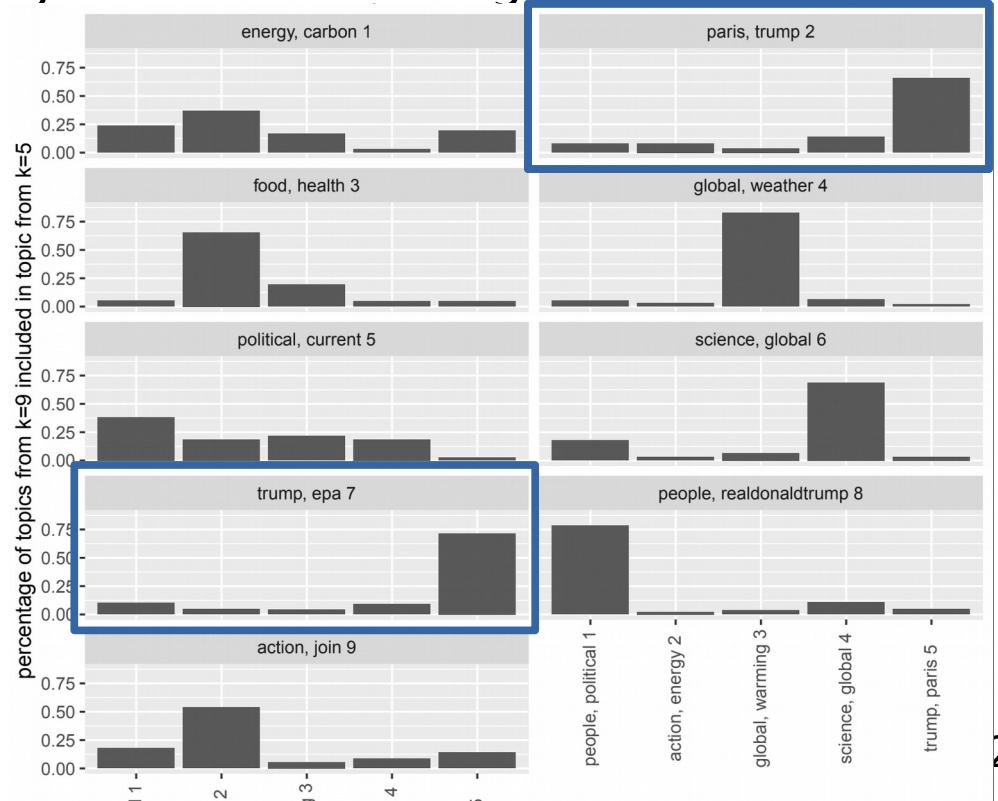
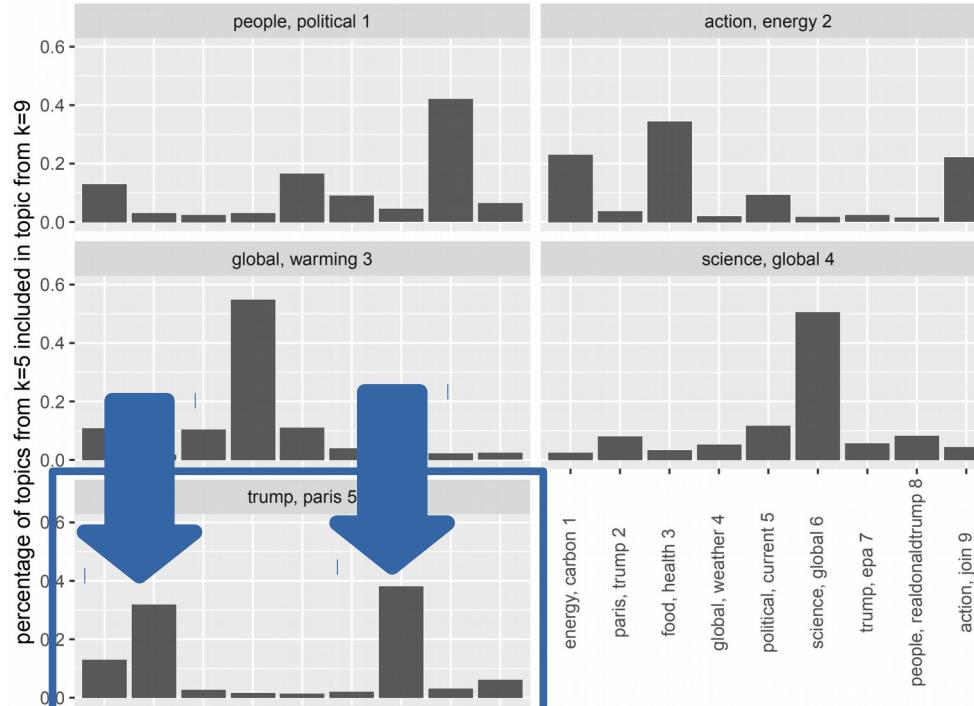


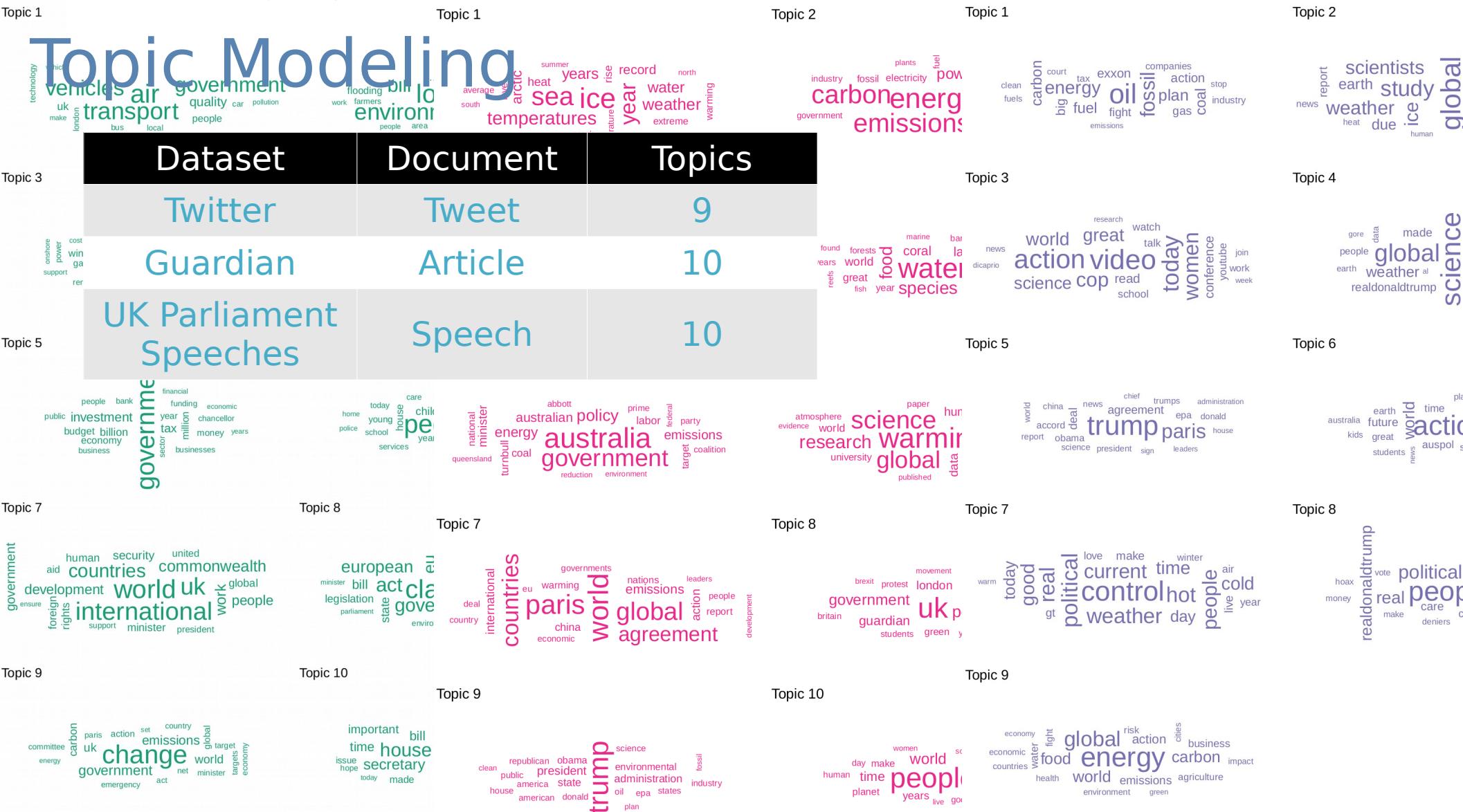
Topic 2 from K=9



Topic Modeling

- Determining the “correct” topics number:
 - Similarity of topics for different number of clusters (are the topics split into two equally after increasing the number of clusters k?)





Topic Modeling

Twitter T2

A word cloud visualization for Twitter T2. The words are colored in shades of blue and purple. The most prominent words include 'global' (large), 'warming' (large), 'arctic' (large, circled in green), 'sea' (large, circled in green), 'scientists' (medium), 'study' (medium), 'weather' (medium), 'ice' (medium), 'heat' (small), 'due' (small), 'news' (small), 'report' (small), 'earth' (small), 'world' (small), 'extreme' (small), 'record' (small), 'years' (small), 'year' (small), 'water' (small), and 'human' (small).

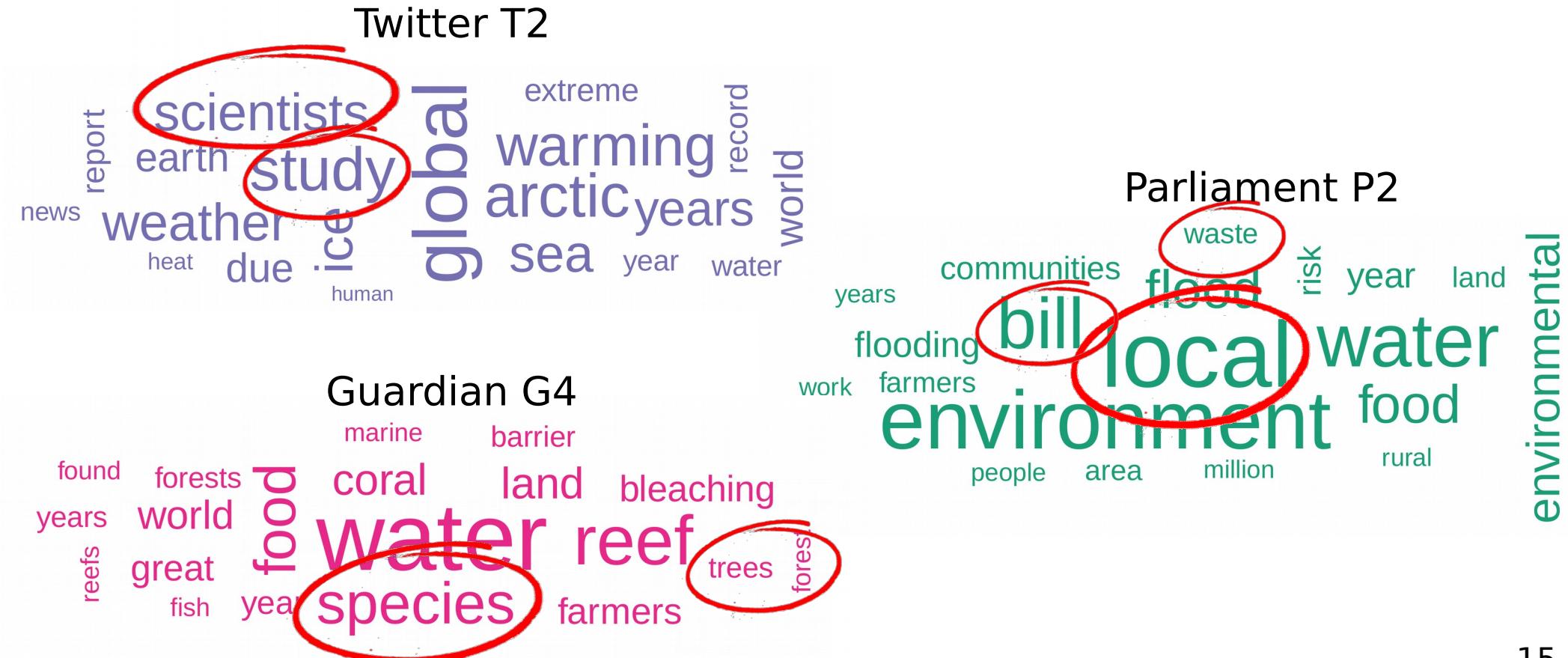
Guardian G4

A word cloud visualization for Guardian G4. The words are colored in shades of pink. The most prominent words include 'water' (large, circled in green), 'reef' (large, circled in green), 'coral' (medium), 'land' (medium), 'bleaching' (medium), 'species' (large), 'forests' (medium), 'world' (medium), 'great' (medium), 'fish' (medium), 'year' (medium), 'marine' (small), 'barrier' (small), 'found' (small), 'years' (small), 'forest' (small), and 'trees' (small).

Parliament P2

A word cloud visualization for Parliament P2. The words are colored in shades of green. The most prominent words include 'local' (large), 'environment' (large), 'bill' (medium), 'flood' (large, circled in green), 'water' (large, circled in green), 'food' (medium), 'rural' (small), 'risk' (small), 'year' (small), 'land' (small), 'waste' (small), 'communities' (small), 'people' (small), 'area' (small), 'million' (small), 'farmers' (small), 'work' (small), 'years' (small), and 'flooding' (small).

Topic Modeling



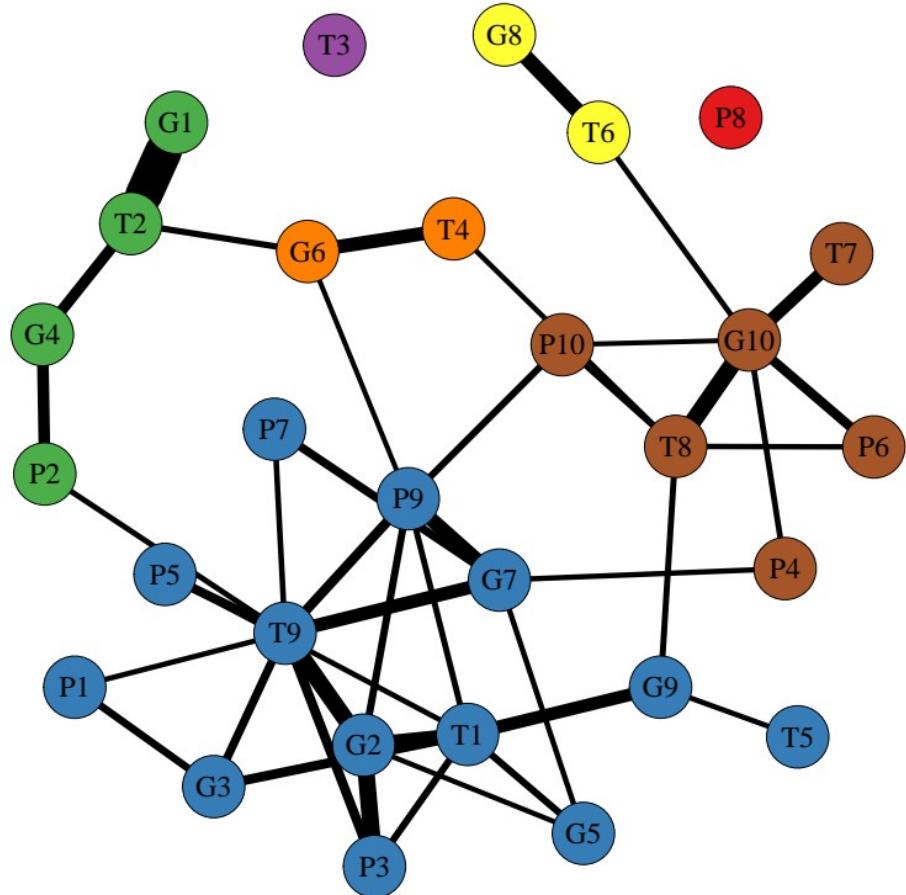
Topics Alignment

- Comparison of lexical topics' distributions
- Kullback-Leibler divergence

$$D_{KL}(t_1\|t_2) = \sum_w P(w|t_1) \log \left(\frac{P(w|t_1)}{P(w|t_2)} \right)$$

where

$P(w|t)$ probability of word w
given topic t



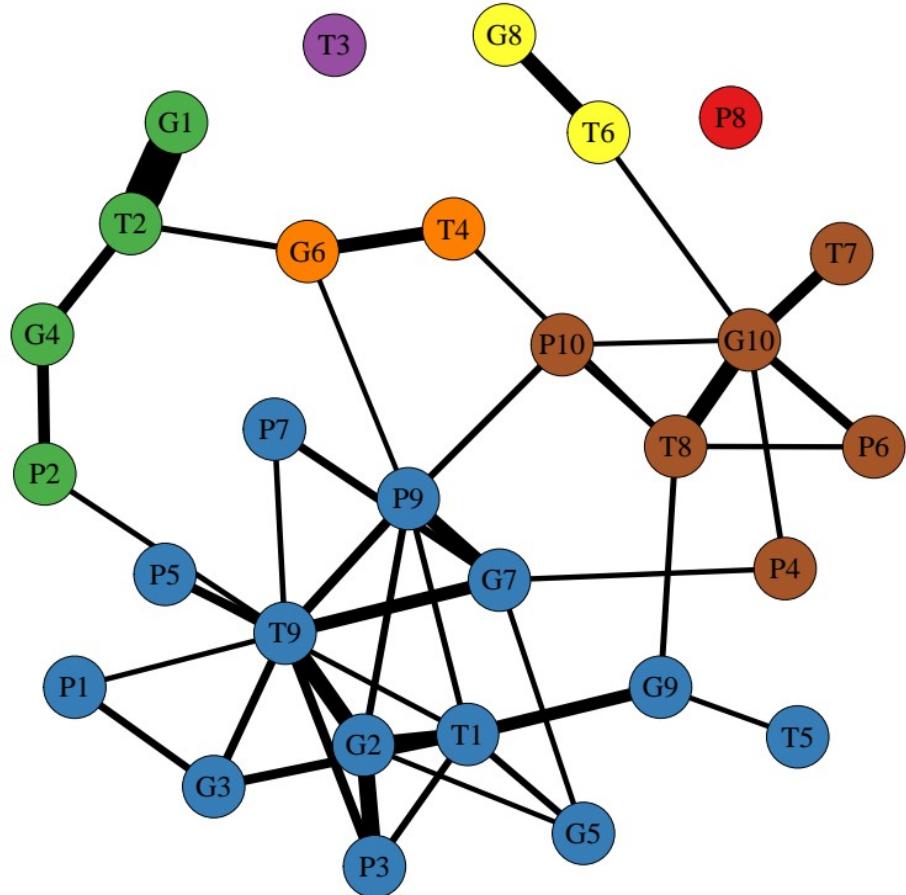
Topics Alignment

- Comparison of lexical topics' distributions
- Kullback-Leibler divergence

$$D_{KL}(t_1\|t_2) = \sum_w P(w|t_1) \log \left(\frac{P(w|t_1)}{P(w|t_2)} \right)$$

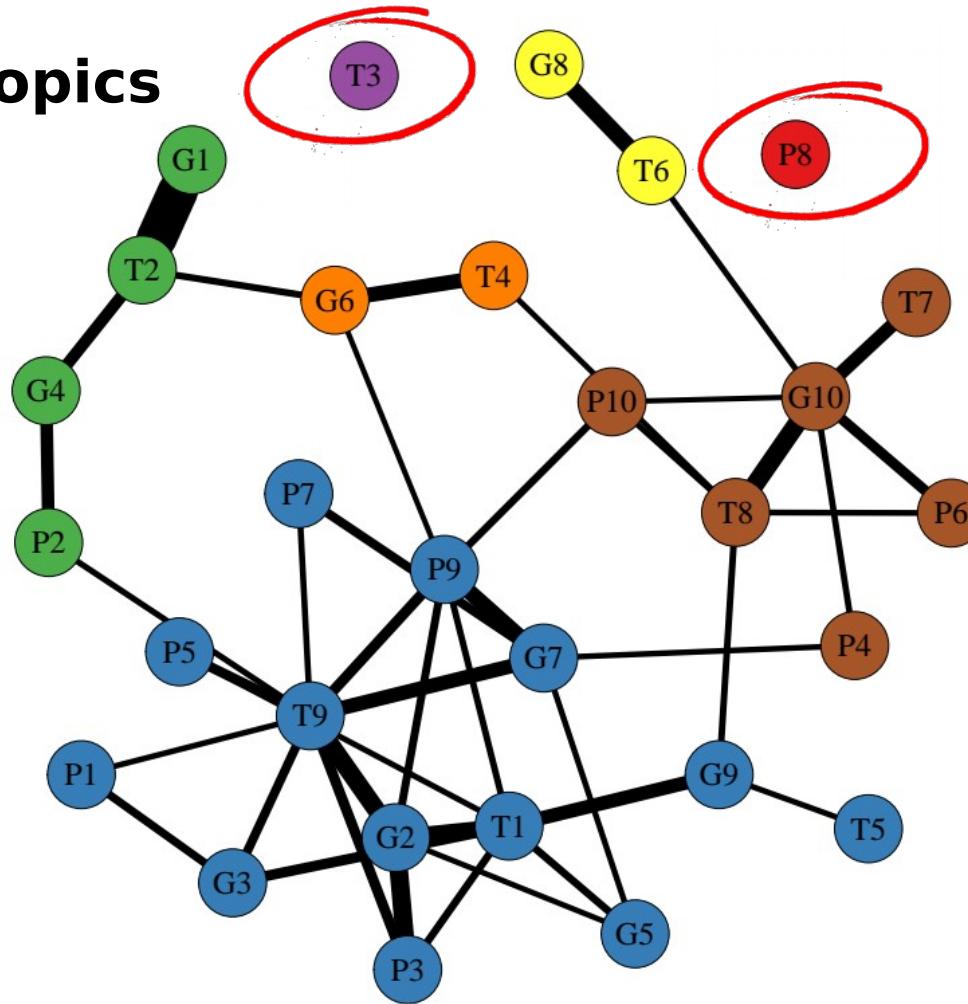
where

$P(w|t)$ probability of word w
given topic t



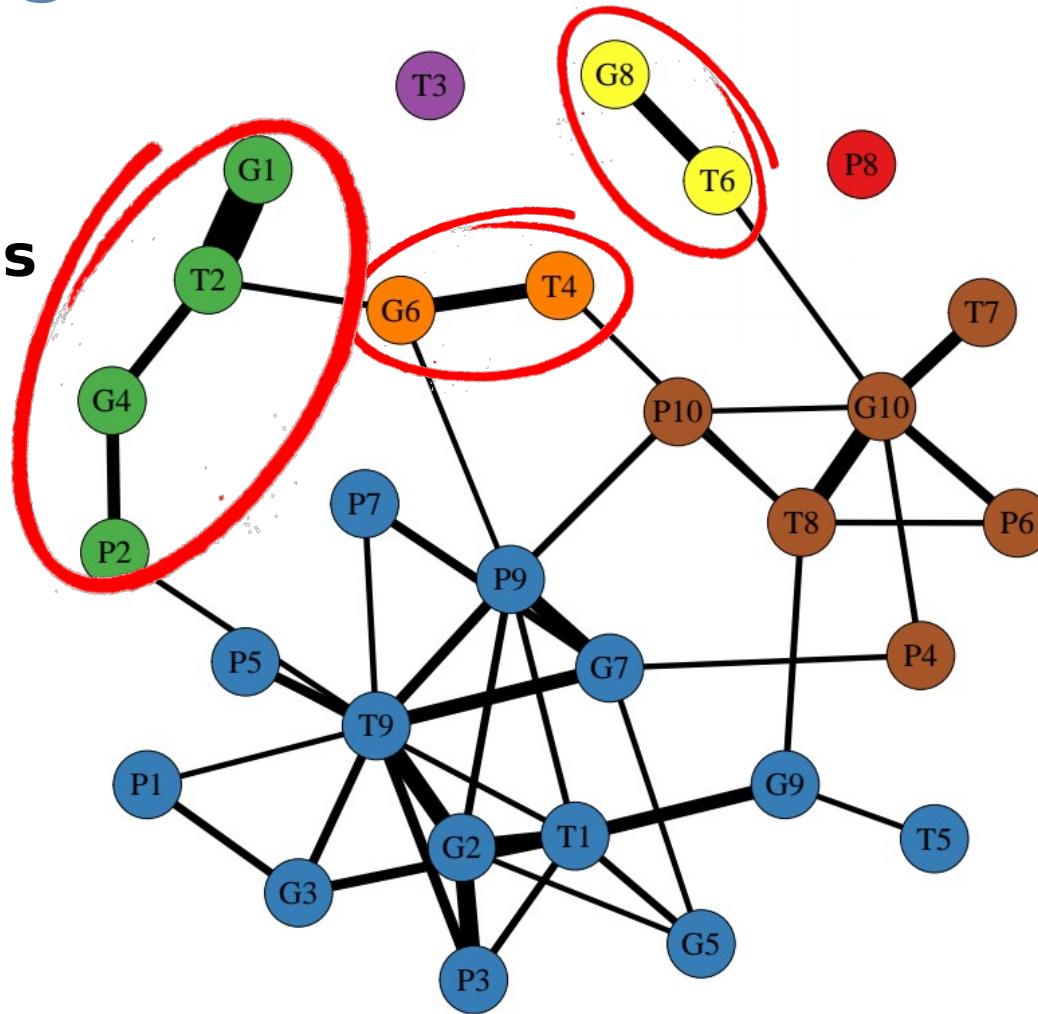
Topics Alignment

Isolated topics

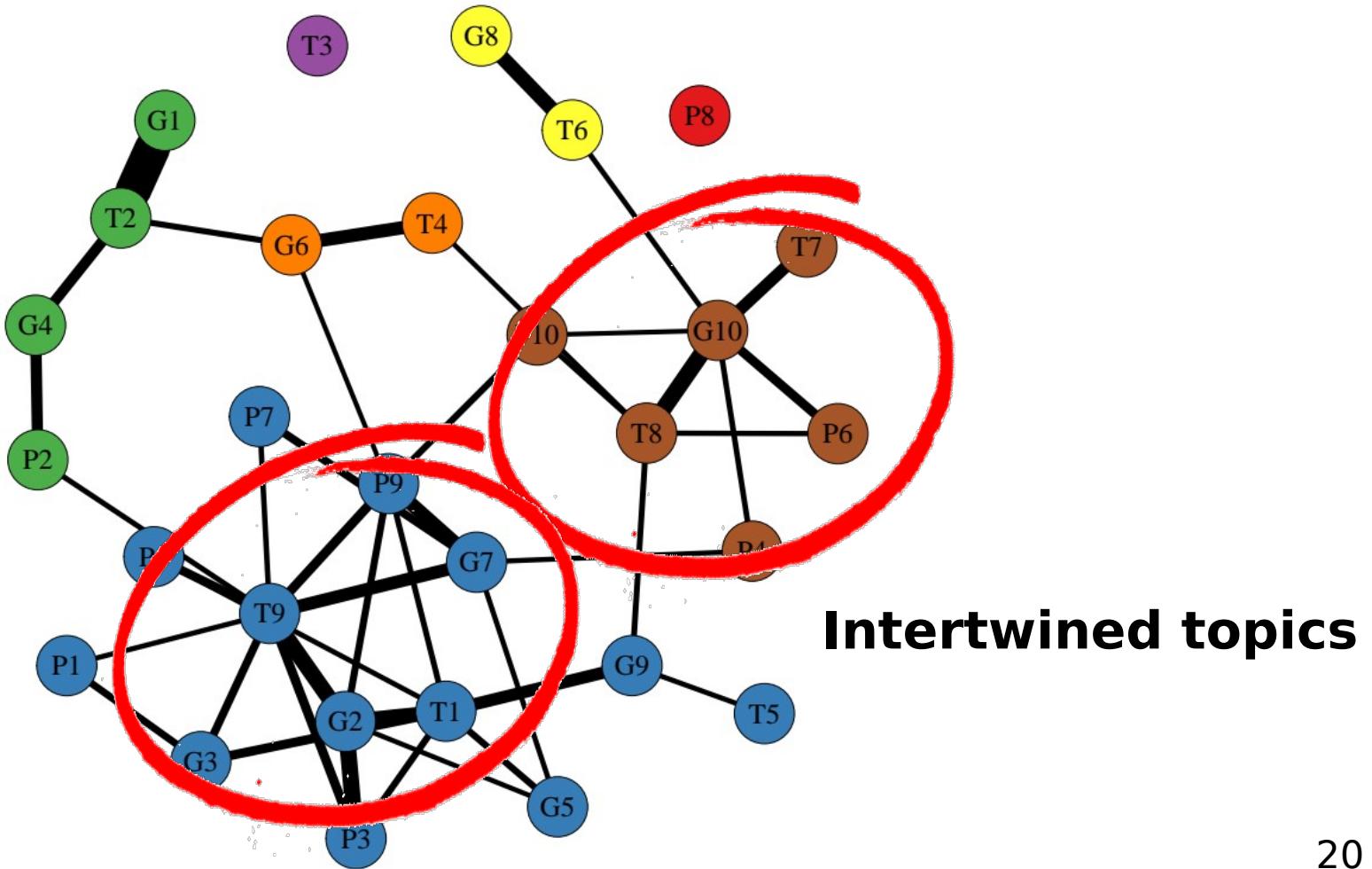


Topics Alignment

Aligned topics

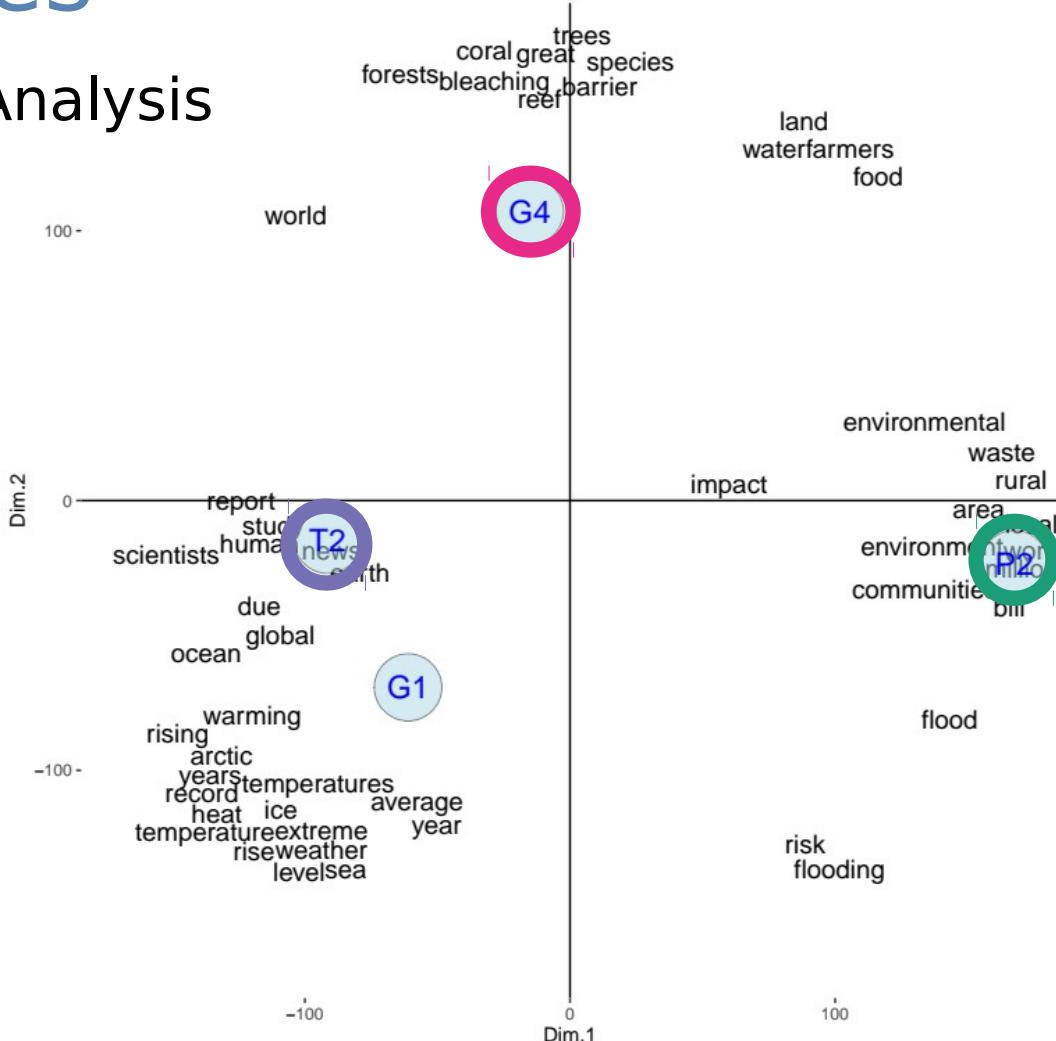
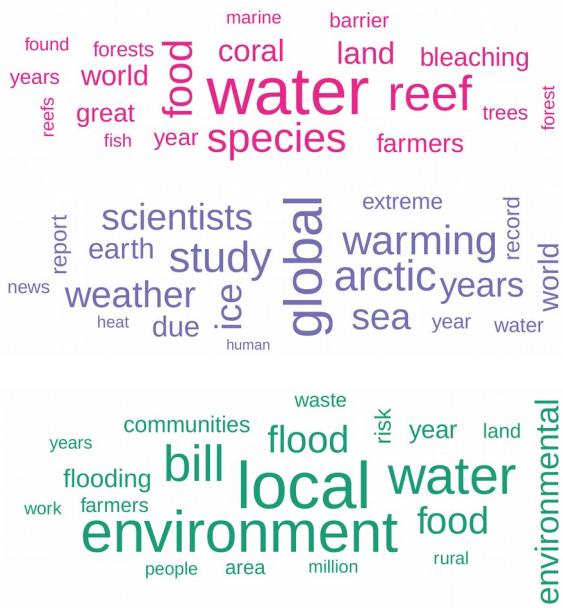


Topics Alignment



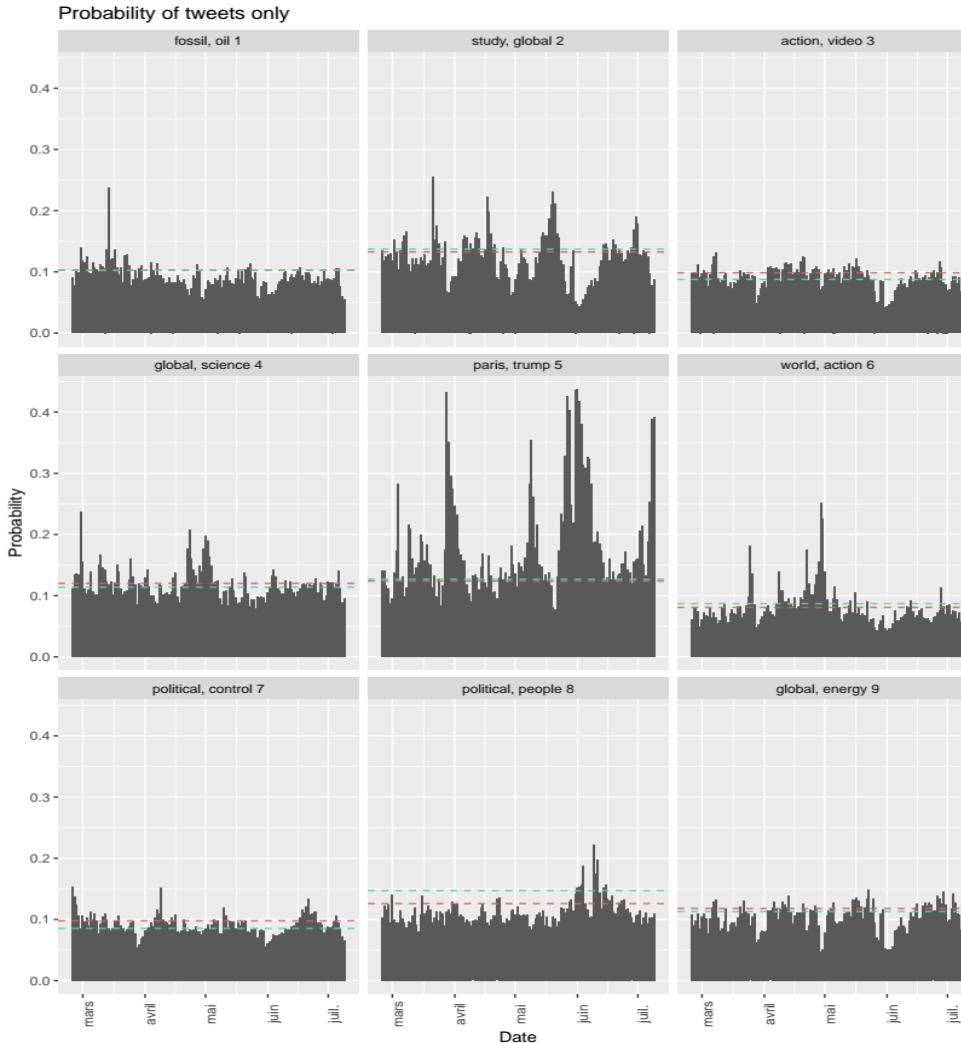
Topic specificities

- Principal Component Analysis (PCA)
- Variables = words
- Observations = topics



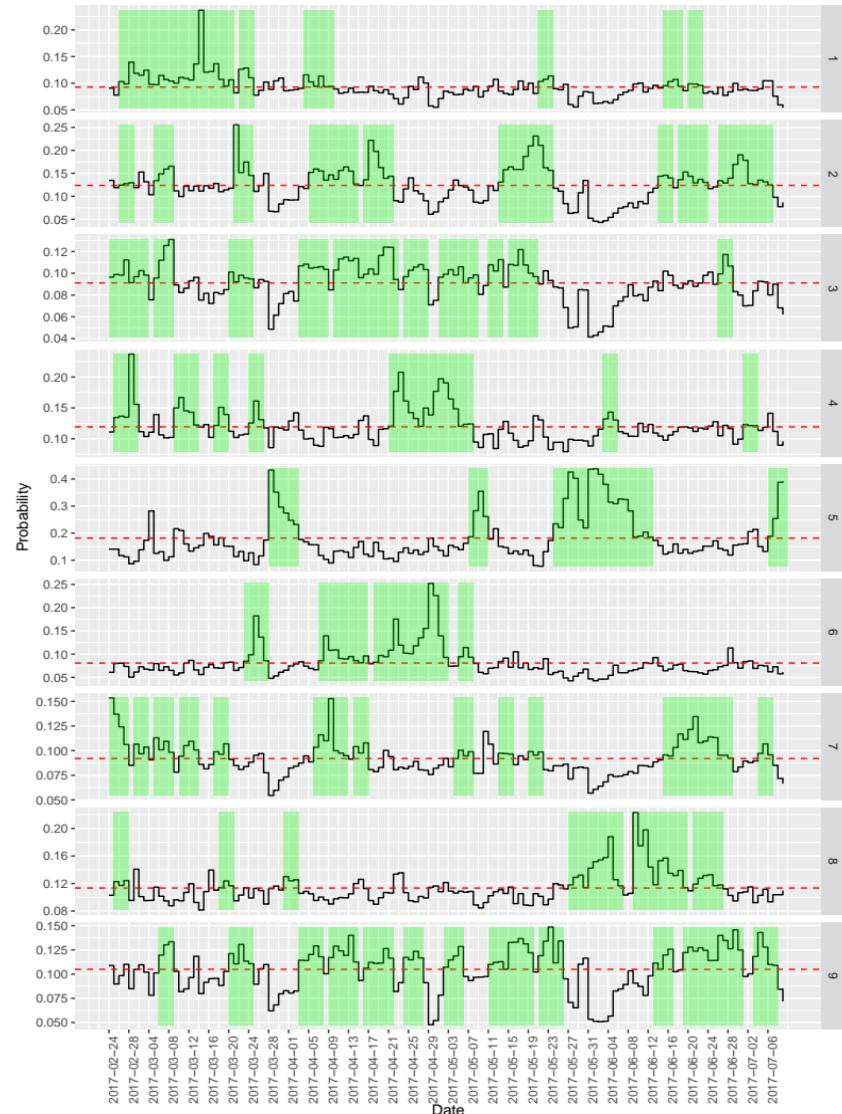
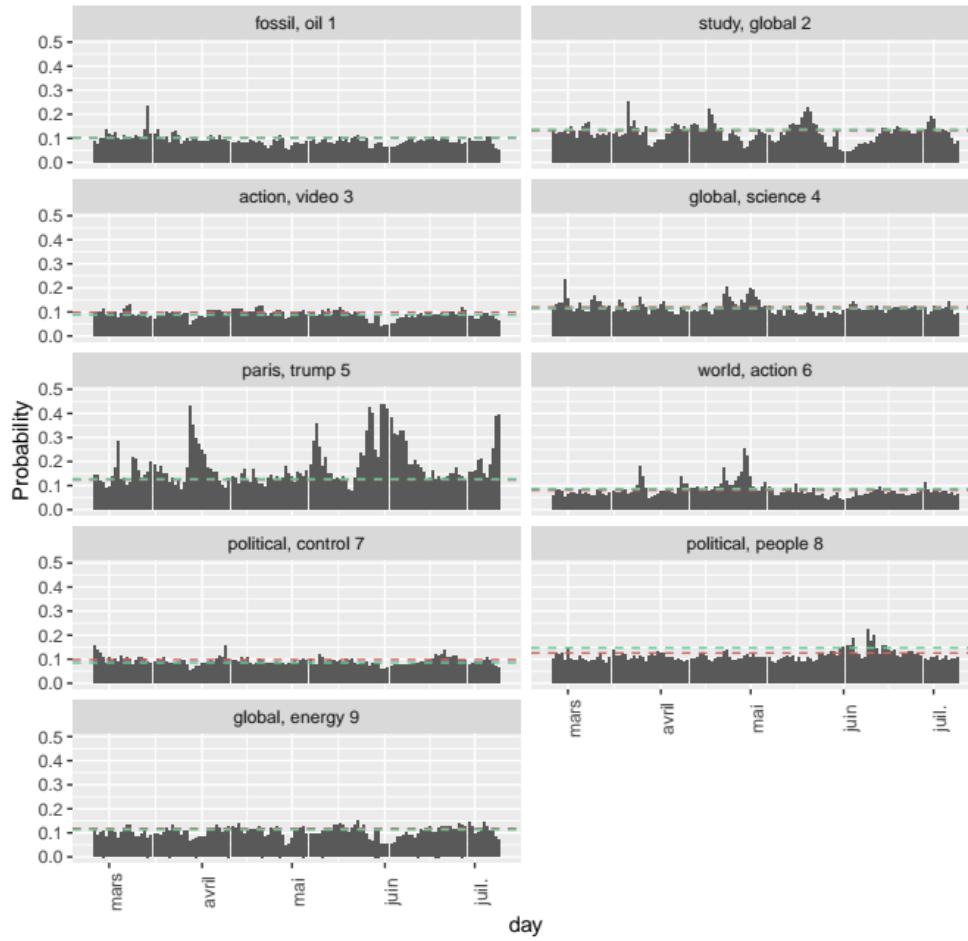
Temporal Aspect

- Focus on probability of topics change in time
- Real life events discovery based on probability peaks



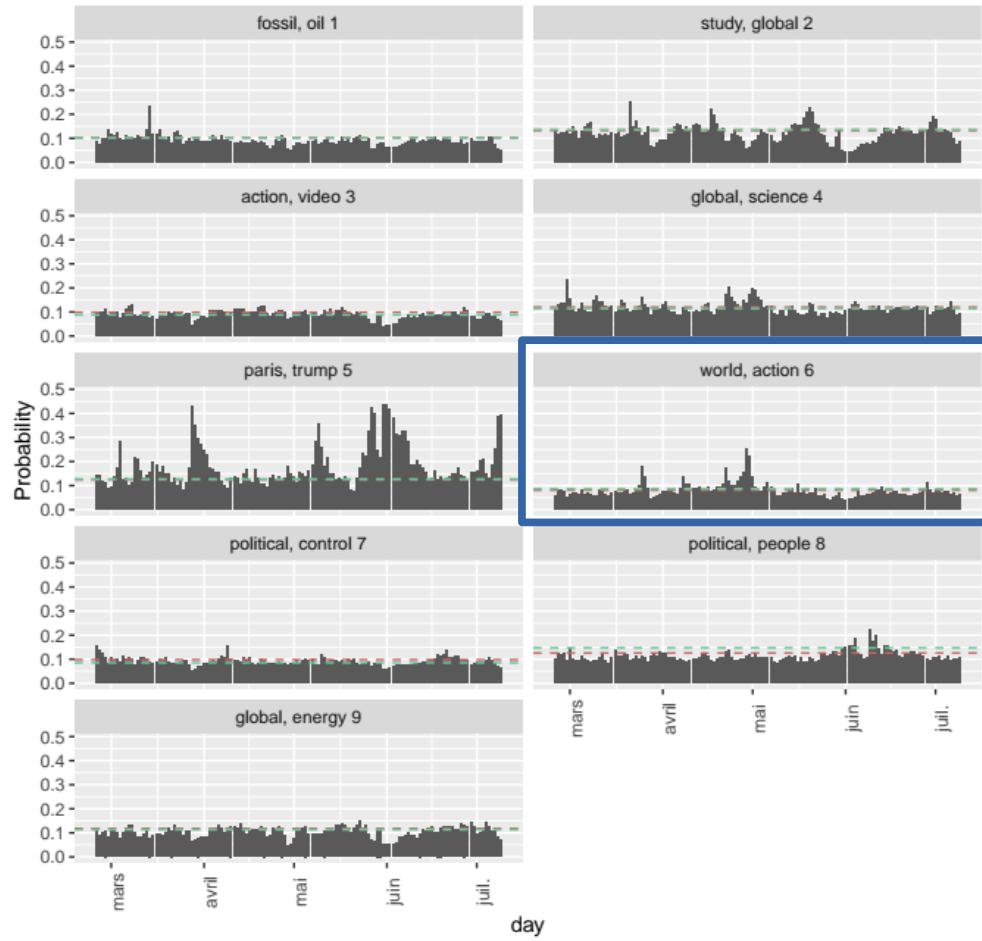
Events detection

Probability of tweets only



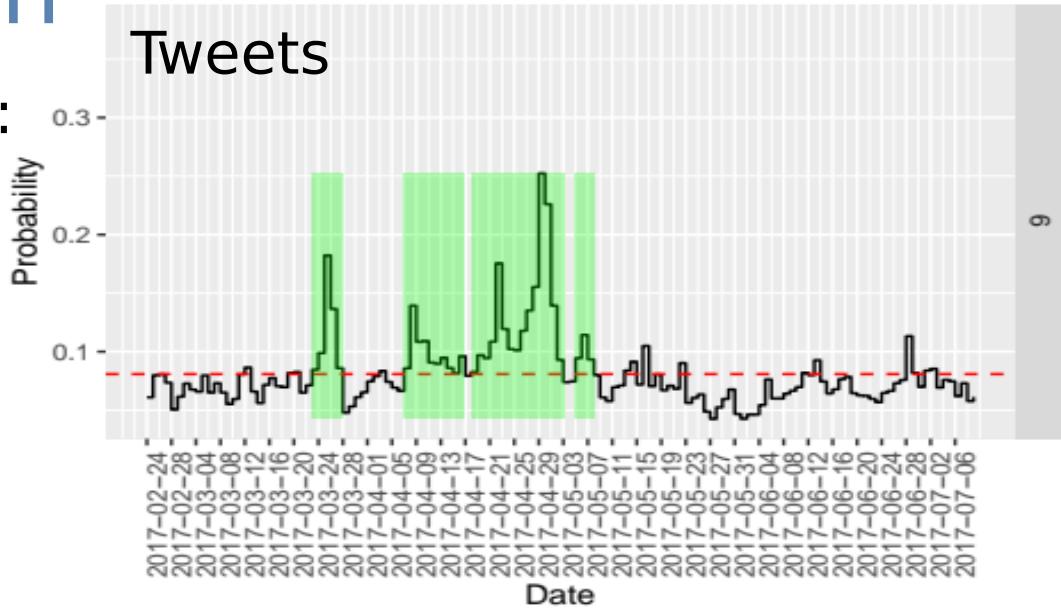
Events detection

Probability of tweets only



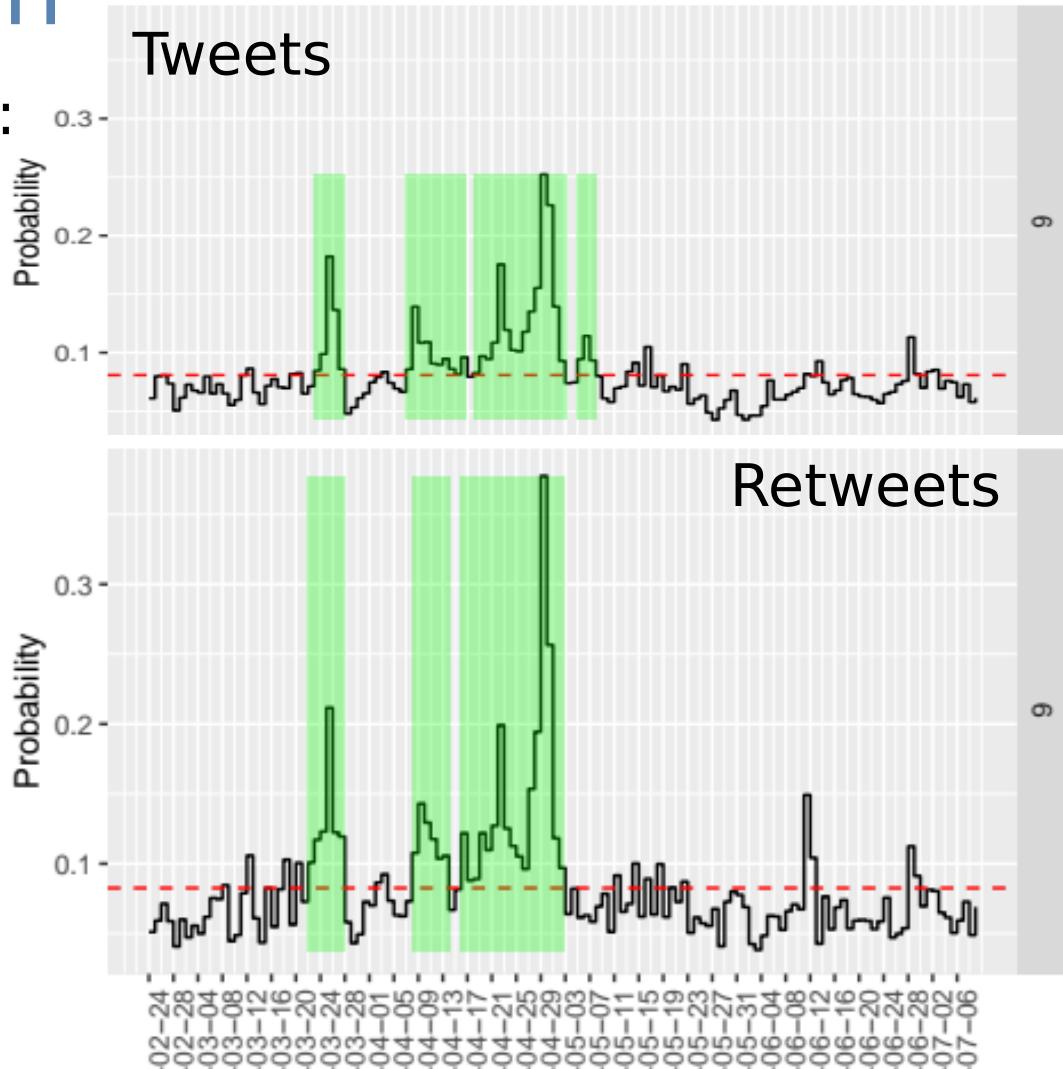
Events detection

- Two different semantics:
 - about information



Events detection

- Two different semantics:
 - about information
 - about exchange of information



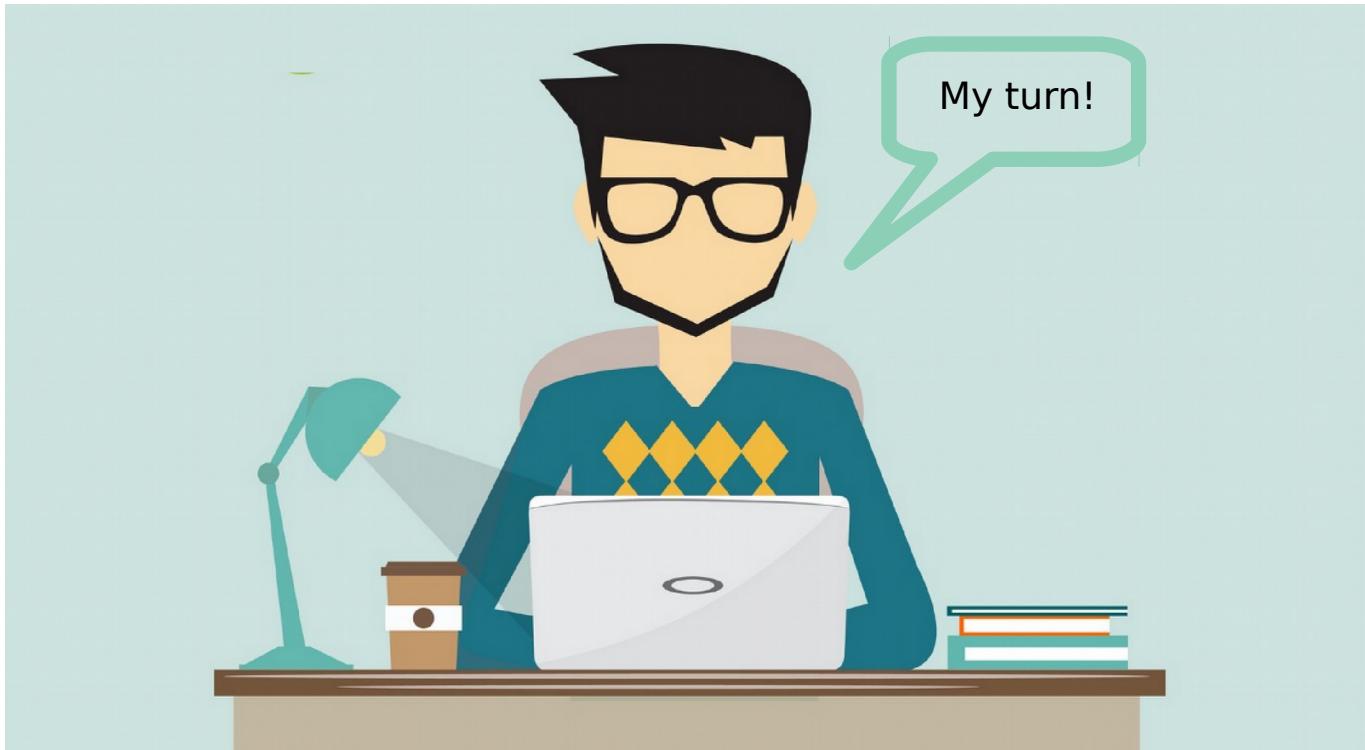
Event determination

- Top tweets (max number of retweets)

Date	Retweets	User Name	Followers	Text
2017-04-29	23715	CNN	33889033	time-lapse, bird's-eye video shows thousands of protesters marching toward white house for action on climate change
2017-04-29	17202	HillaryClinton	14184025	great to see ppl take to the streets & combat climate change, protect the next generation & fight for jobs & economic justice. #climatemarch
2017-04-29	5516	mattmfm	40582	wow: enormous crowd for the climate march today, just a week after the march for science. people are fired up.
2017-04-29	5229	LeoDiCaprio	17422580	honored to join indigenous leaders and native peoples as they fight for climate justice. join me in standing with them. #climatemarch
2017-04-29	2984	BuzzFeed News	870719	leonardo dicaprio (and thousands of other people) are at the #climatemarch in washington, dc to demand action on climate change

- Real-word event: Climate March 2017

Armin's part



PENELOPE Integration

- Use our components to exploratory observatory - general questions
- Choice of one of datasets
- Using pre-trained topics
- Using prepared network
- Dividing page into panels:
 - Topic analysis (basic info about topics)
 - Topic comparison
 - Time analysis
 - Interaction analysis

PENELOPE Integration

Integration

Select base dataset: Twitter Select number of topics: K 5 Save changes Cancel

Topics Analysis Topics Comparison Topics in Time Interactions

Wordclouds

Topic 1 Topic 2

Topic 3 Topic 4

Topic 5 Topic 6

Word frequencies per topic

Select topic: Topic 1 Select number of top words: Top 10 Top 20 Top 30 Go

Download frequencies for selected topic Download

Download frequencies for all topics Download

Download all frequencies for all topics Download all

Overall mean probability per topic

Overall median probability per topic

PENELOPE Integration

Select base dataset:

Twitter

Select number of topics:

- K 5
- K 8
- K 10

Save changes

Cancel

Topics Analysis

Topics Comparison

Topics in Time

Interactions

Select comparison dataset:

Guardian

UK Parliament

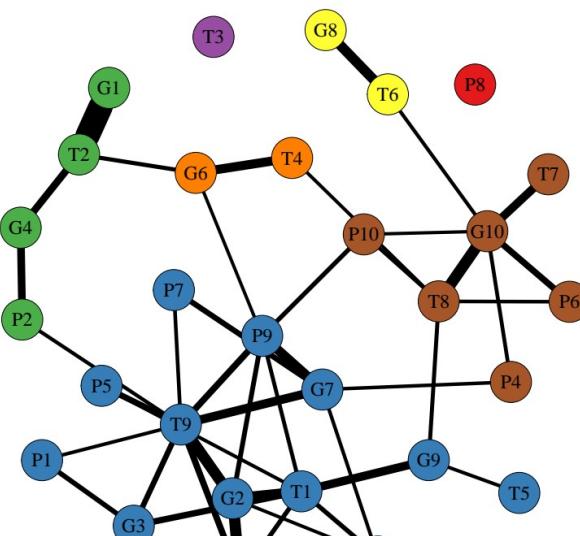
Select number of topics:

- K 5
- K 8
- K 10

Select number of topics:

- K 5
- K 8
- K 10

Show Comparison Graph

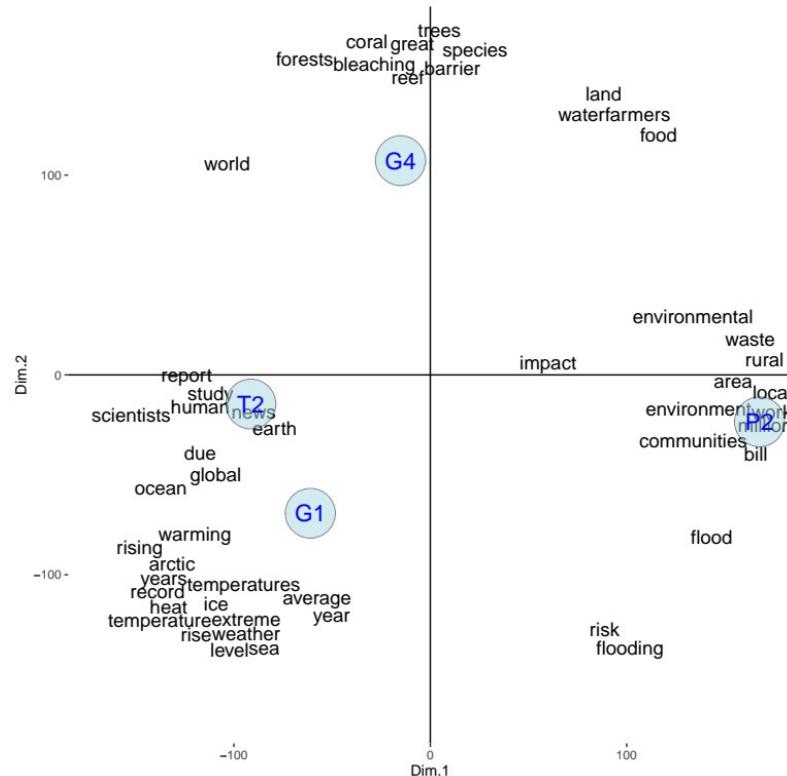


PENELOPE Integration

Show PCA

Select topics for PCA:

- Guardian G1 Twitter T1 UK Parliament P1
- Guardian G2 Twitter T2 UK Parliament P2
- Guardian G3 Twitter T3 UK Parliament P3
- Guardian G4 Twitter T4 UK Parliament P4
- Guardian G5 Twitter T5 UK Parliament P5



PENELOPE Integration

The screenshot displays the PENELOPE Integration interface. At the top, there is a header bar with three colored circles (red, orange, green) and a search bar containing a magnifying glass icon. Below the header, there are two dropdown menus: "Select base dataset" set to "Twitter text" and "Select number of topics" set to "K 8". There are also "Save changes" and "Cancel" buttons.

The main area contains a navigation bar with four tabs: "Topics Analysis", "Topics Comparison", "Topics in Time", and "Interactions". The "Topics Analysis" tab is active.

Below the tabs, there are four bar charts arranged in a 2x2 grid, each representing a different topic:

- Tweets:**
 - Topic 1:** Shows bars for four categories (purple, blue, green, blue) across four time periods.
 - Topic 2:** Shows bars for four categories (purple, blue, green, blue) across four time periods.
- Retweets:**
 - Topic 1:** Shows bars for four categories (purple, blue, green, blue) across four time periods.
 - Topic 2:** Shows bars for four categories (purple, blue, green, blue) across four time periods.

To the right of the charts, there is a legend with three checkboxes:

- Show events
- Show trends
- Show world events

PENELOPE Integration

Select topic to analyse:

- Topic 1
- Topic 2
- Topic 3

Select dates to analyse

From:

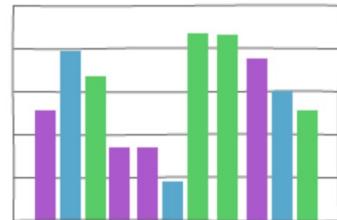
To:

Show me!

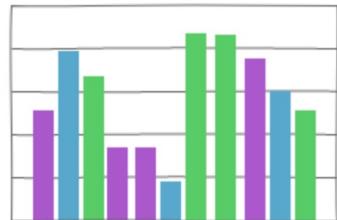
Cancel

Selected topic analysis

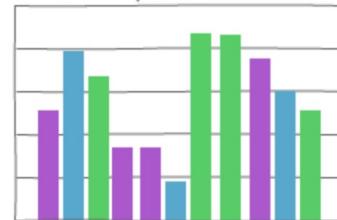
Tweets



Retweets



Ratio tweets/retweets



Show events

Show trends

Show world events

Show events

Show trends

Show world events

Show world events

Show stats

Tweets Number	<input type="text"/>
Retweets Number	<input type="text"/>
Number of users twitting	<input type="text"/>

PENELOPE Integration

Show stats

Tweets Number	
Retweets Number	
Number of users twitting	

Selected number of top tweets to show

- Top 5
- Top 10
- Top 20

Show me!

Top Retweeted Tweets

User	Followers	Text	Date	Retweets

PENELOPE Integration

- Here we proposed using components to general observatory: general questions about *data*
- Other ideas?
 - Building different pipeline integrating additional components ?
 - Specialised observatory: more specialized questions – focus on one particular event/theme (e.g. forest fires) ?

