Blending psychology and data science

A mixed-methods framework for analysing social media with an application to climate change tweets

Matthew Andreotta

School of Psychological Science, University of Western Australia

Data61, CSIRO

http://tiny.cc/xup88y

Everyone involved

Cecile Paris

Data61, CSIRO

Mark Hurlstone

Behavioural Economics Laboratory, UWA

Simon Farrell

Memory and Decision Making Laboratory, UWA

Fabio Boschetti

Oceans & Atmosphere, CSIRO

Iain Walker

University of Canberra

Robertus Nugroho

Soegijapranata Catholic University, Inodnesia

Students striking on the steps of South Australia's Parliament House. (Gabriella Marchant for ABC News, 2019)

How does the public conceptualise climate change?

Climate change is one of the greatest challenges for humanity (Schneider, 2011).

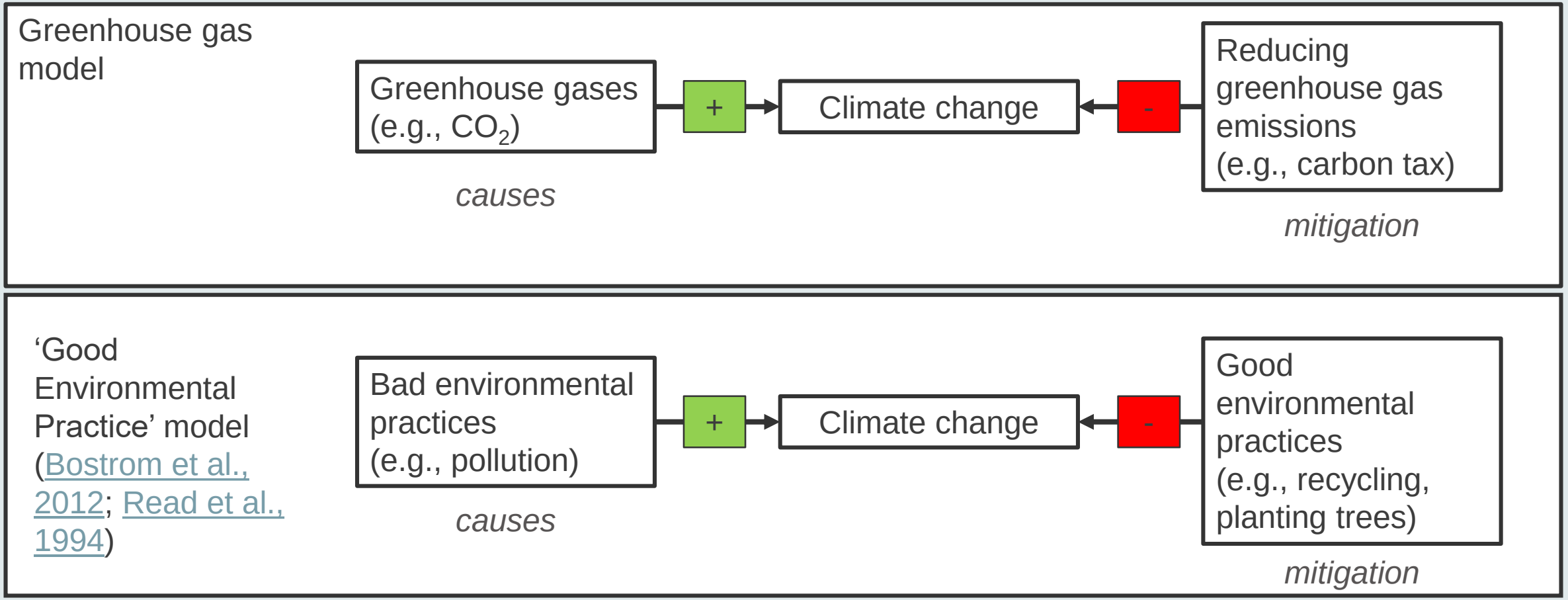
Public perceptions are critical.

How can we monitor, understand, and communicate with the public?

(Ian Webster for CRED, 2009)

A mental model perspective

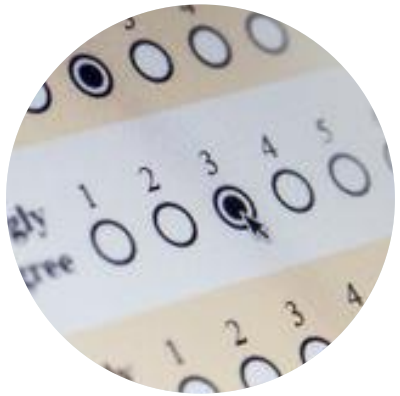
- Mental models are internalised models of a world, can be simulated and generate inferences
- Mental models attempt to *mirror* the world, but often fall short (constrained cognitive capacity, constraints from ideology)
- Differences in mental models of climate change correspond to differences in policy support.



Mental models

Different mental models of climate change generate different inferences

Collecting public perceptions



Quantitative

Surveys, experiments

- Easier to analyse
- Top-down

Qualitative

Interviews, focus groups

- Intrusive data collection
- Intensive to collect, transcribe, and analyse data

Observation/Naturalistic

Social media

- Environment not controlled by experimenter
- Diverse data
- No probes required to generate data
- Can be difficult to understand meaning

Our Study

Research Question

When Australian users tweet about climate change, what are the topics being discussed?

Analysis

Topic modelling

Topic modelling

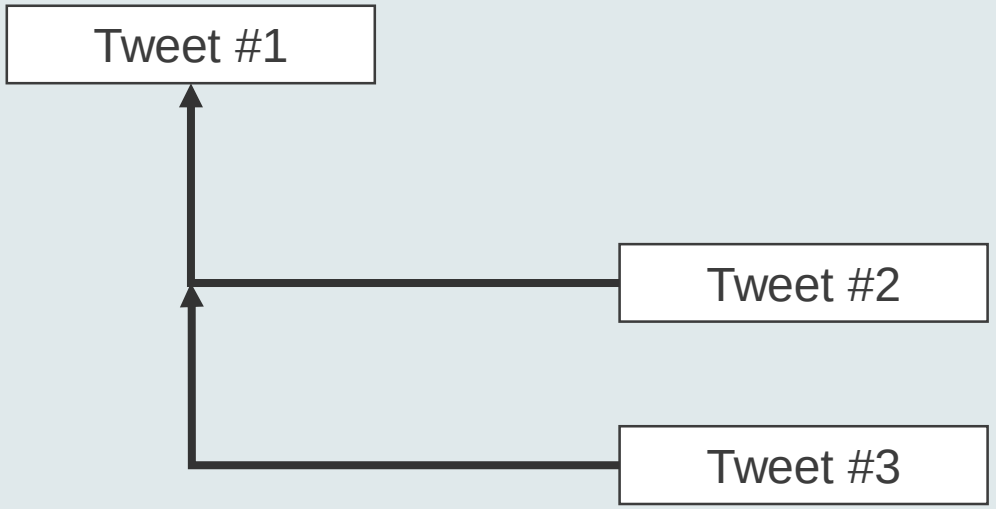
*Identifying abstract representations of semantically related words and concepts (**topics**)*

Popular approaches (e.g., *LDA*, *PLSA*)

- Relies on patterns of word co-occurrence between tweets
- Represent topics as a list of keywords
- Word co-occurrences are limited on Twitter (tweets are short)
- Popular approaches suffer (Nugroho, Zhao, Yang, Paris, & Nepal, 2017)

Non-Negative Matrix inter-joint Factorization (*NMijF*; Nugroho et al., 2017)

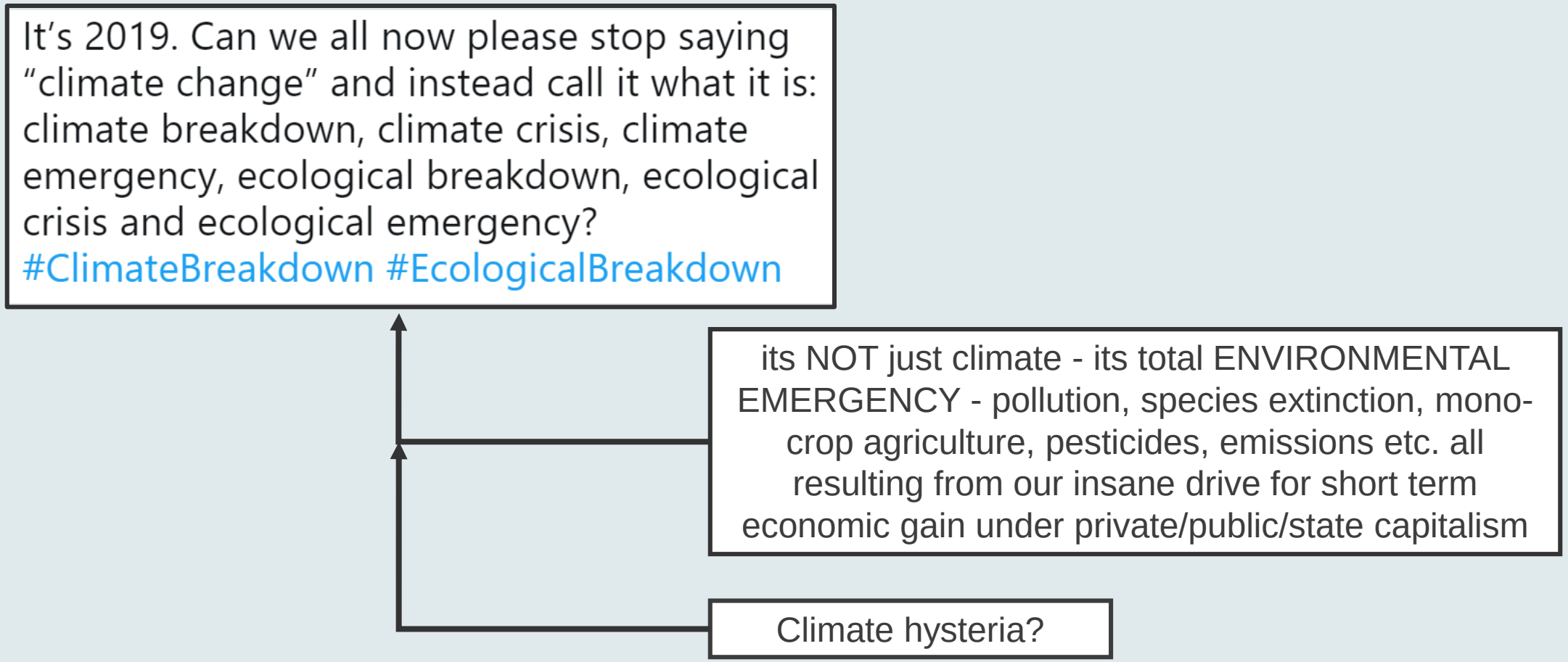
- Relies on patterns of word co-occurrence **and on patterns of social and temporal information**
- Outperforms popular approaches (Nugroho et al., 2017)



Social interactions

Tweet #1, #2, and #3 may fall within the same topic

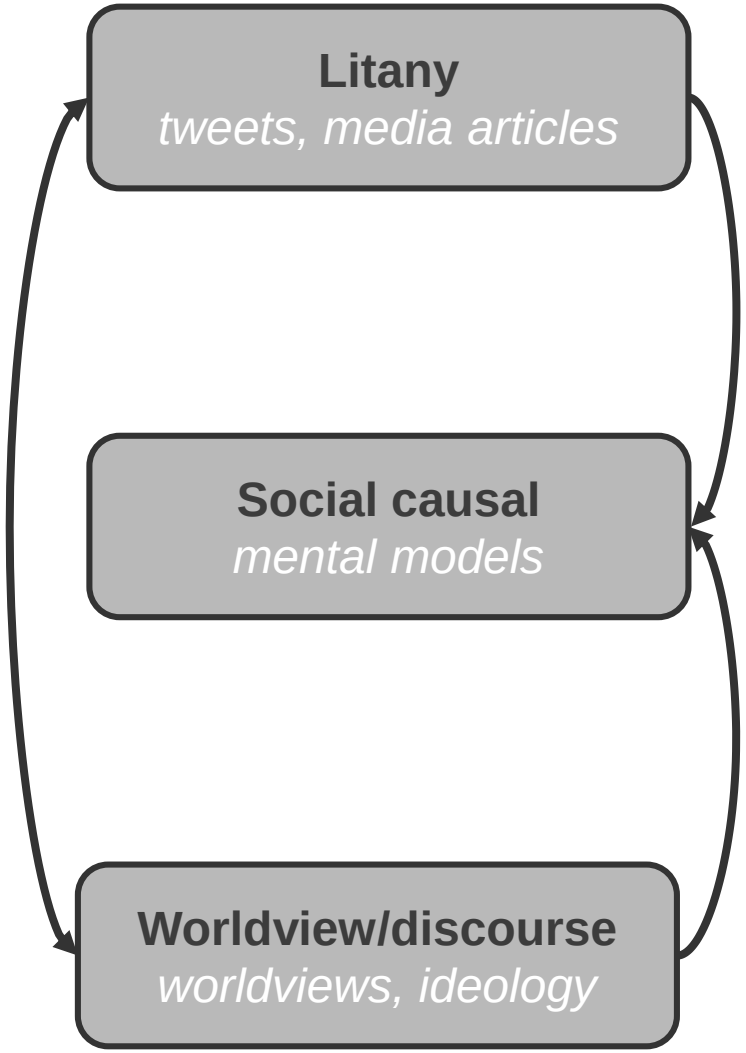
A quick detour into psychology



Social interactions

Tweet #1, #2, and #3 may fall within the same topic

(Naming climate change / Defining climate change)



Causal layered analysis (CLA)

(Inayatullah, 2003, 2004)

Litanies: issues presented as public description/social truth

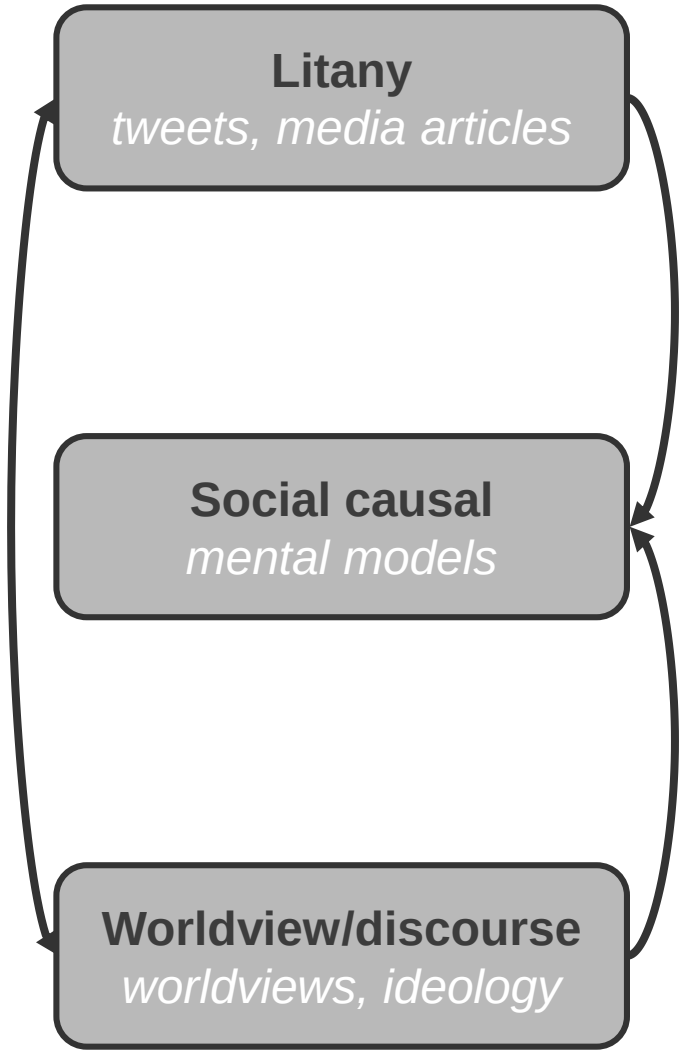
Social causal: issues presented as systematic &/or technical explanations

Worldview/discourse: issues presented as deep, complex understandings which shape the way meaning is constructed

CLA

The tweet

its NOT just climate - its total ENVIRONMENTAL EMERGENCY - pollution, species extinction, mono-crop agriculture, pesticides, emissions etc. all resulting from our insane drive for short term economic gain under private/public/state capitalism



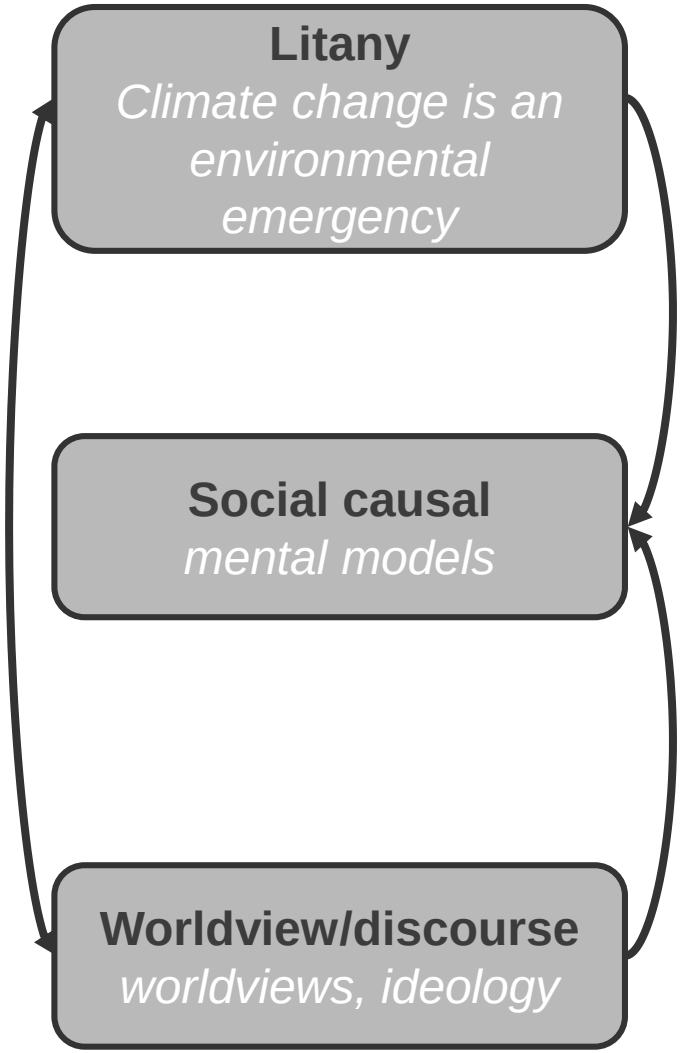
CLA

The tweet

its NOT just climate - **its total ENVIRONMENTAL EMERGENCY** - pollution, species extinction, mono-crop agriculture, pesticides, emissions etc. all resulting from our insane drive for short term economic gain under private/public/state capitalism

The litany

Climate change is an environmental emergency



CLA

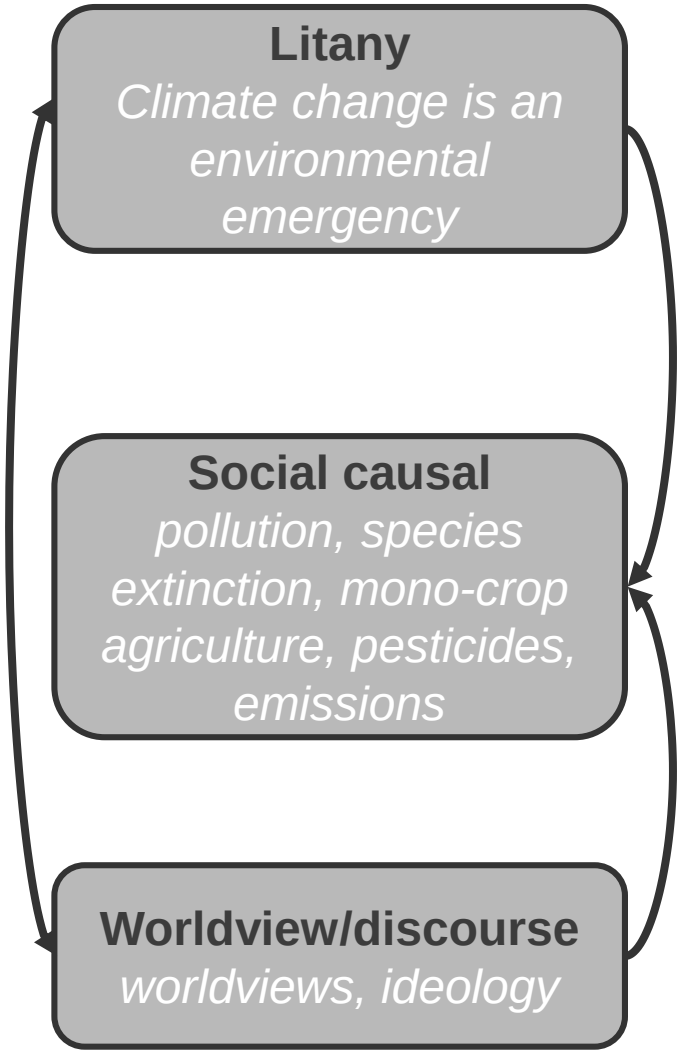
The tweet

its NOT just climate - its total ENVIRONMENTAL EMERGENCY - **pollution, species extinction, mono-crop agriculture, pesticides, emissions etc.** all resulting from our insane drive for short term economic gain under private/public/state capitalism

The social causal

Climate change will destabilise and destroy many biological and human systems

User has a mental model which considers these systems interlinked and sensitive to changes in climate



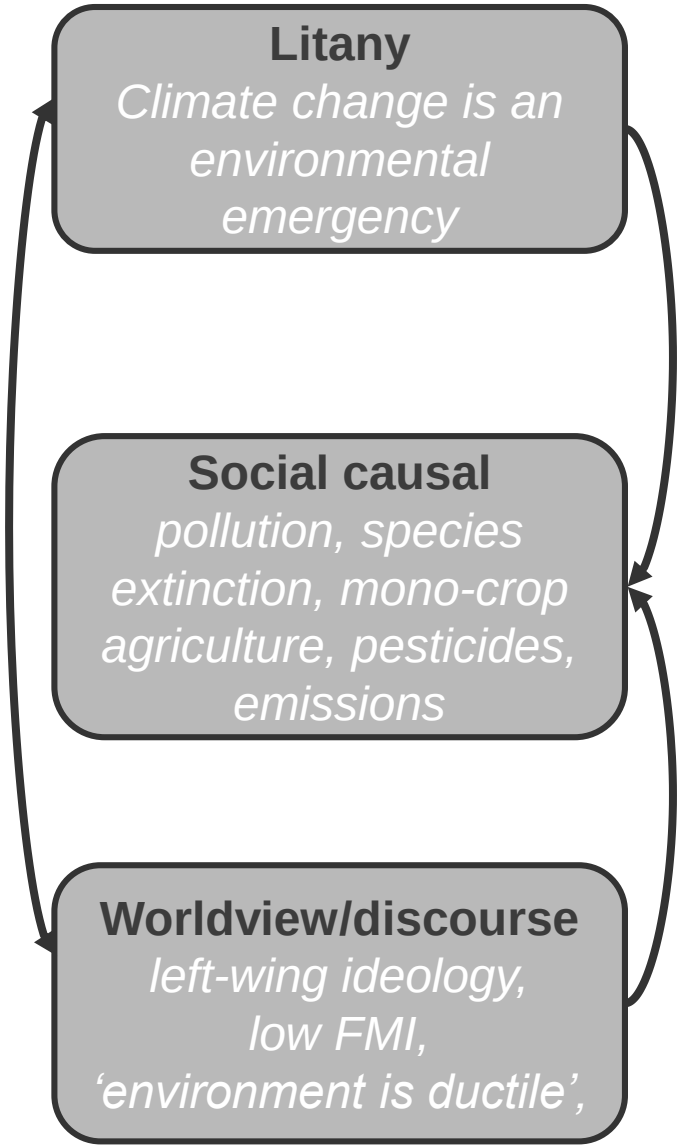
CLA

The tweet

its NOT just climate - its total ENVIRONMENTAL EMERGENCY - pollution, species extinction, mono-crop agriculture, pesticides, emissions etc. all resulting from **our insane drive for short term economic gain under private/public/state capitalism**

The worldview/discourse

- Left-wing ideology
- Low Free-Market Endorsement (Heath & Gifford, 2006)
- 'Environment is ductile' worldview (Price et al., 2014)



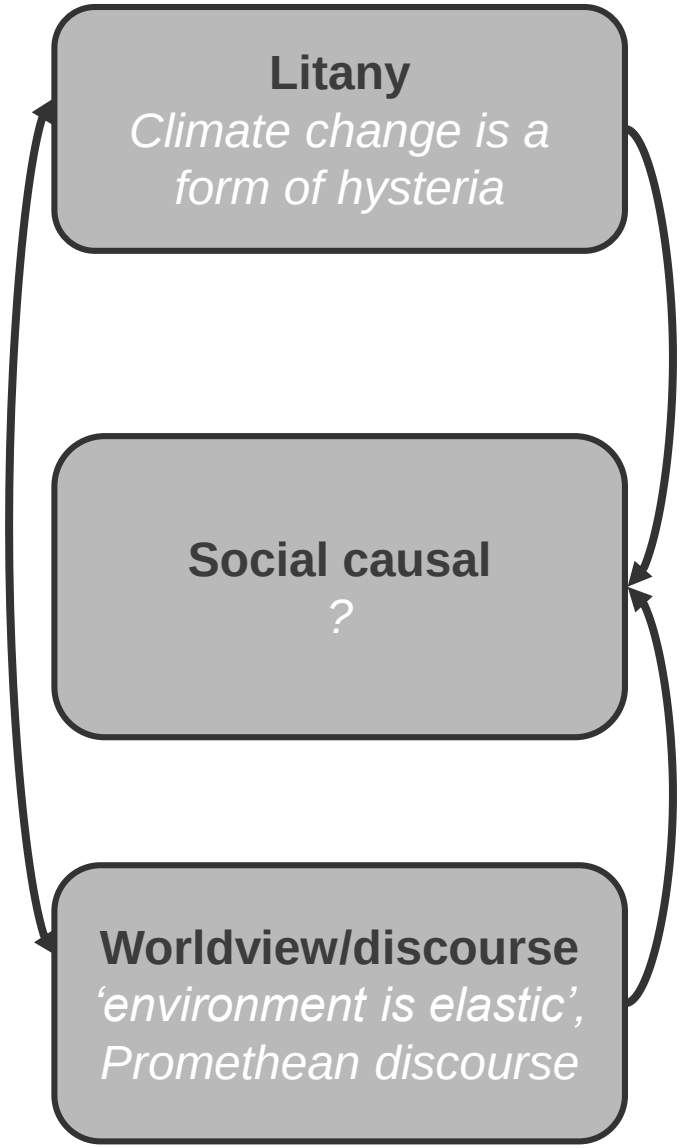
CLA

The tweet

its NOT just climate - its total ENVIRONMENTAL EMERGENCY - pollution, species extinction, mono-crop agriculture, pesticides, emissions etc. all resulting from our insane drive for short term economic gain under private/public/state capitalism

Another tweet (same topic)

Climate hysteria?



Back to the study

What to use?

Topic modelling

- Provides a semi-automated approach to analysing social media data
- Can group content on the basis of semantic relationships
- May be difficult to identify the underlying ideas, assumptions, and conceptualisations of users
- Constrained to the output of the topic models (not all topics may be meaningful)

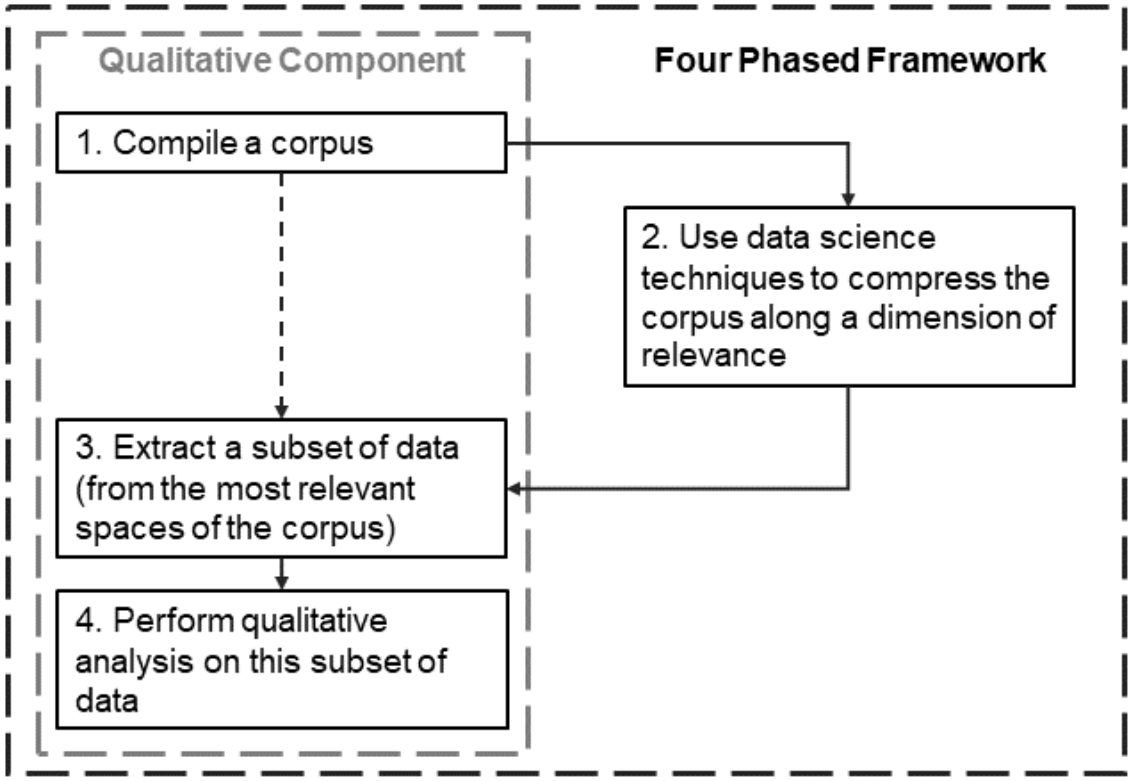
Qualitative approaches (e.g., CLA)

- Researcher can flexibly apply known theories of behaviour when analysing data (if desired)
- Numerous frameworks exist, each ensuring a particular type of insight is derived from the analysis
- Qualitative approaches are laborious (and often infeasible) to apply over Twitter

Our research question

When Australian users tweet about climate change, what are the topics being discussed?

Answered with: topic modelling ***and thematic analysis*** (Braun & Clarke, 2006, 2012)

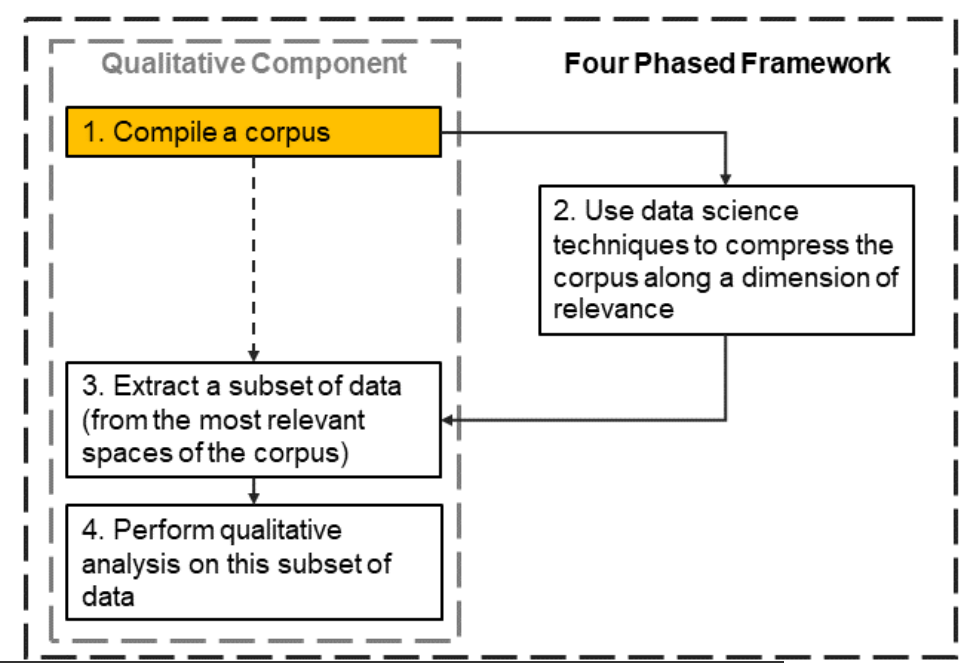


Four-Phased Framework (Andreotta et al., 2019)

We benefited from our topic solution, without being constrained to it.

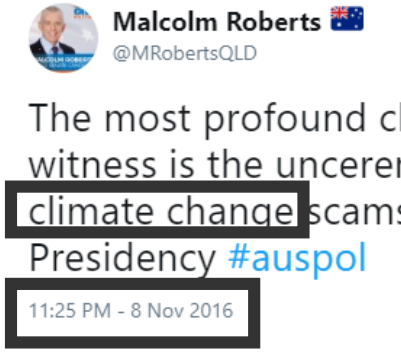
Through a qualitative analysis we can identify themes to answer our research question

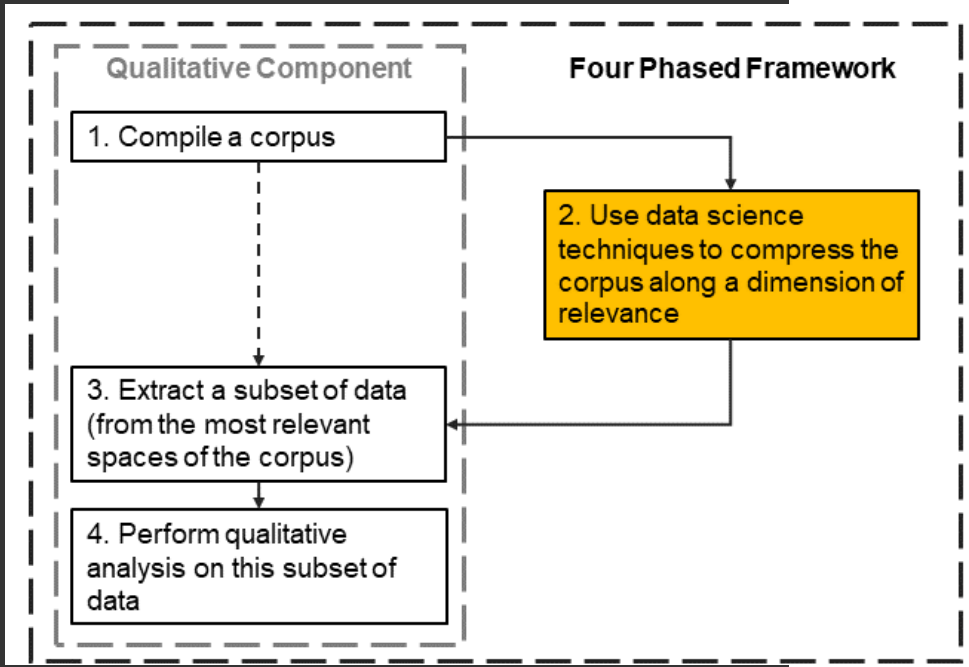
However, qualitative analyses are intensive, and we must select a subset of data to analyse.



Our data

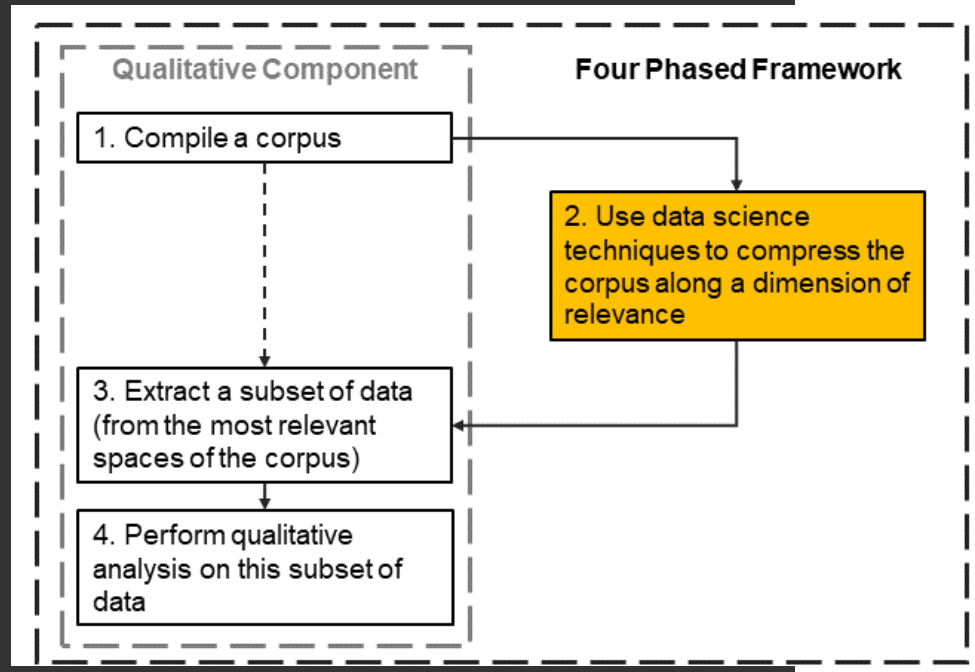
- Collected through ESA (Cameron et al., 2012; CSIRO, 2019)
- Had to satisfy three criteria:
    1. Australian location
    2. Posted in 2016
    3. References climate change (e.g., “#climatechange”)
- 201,506 tweets





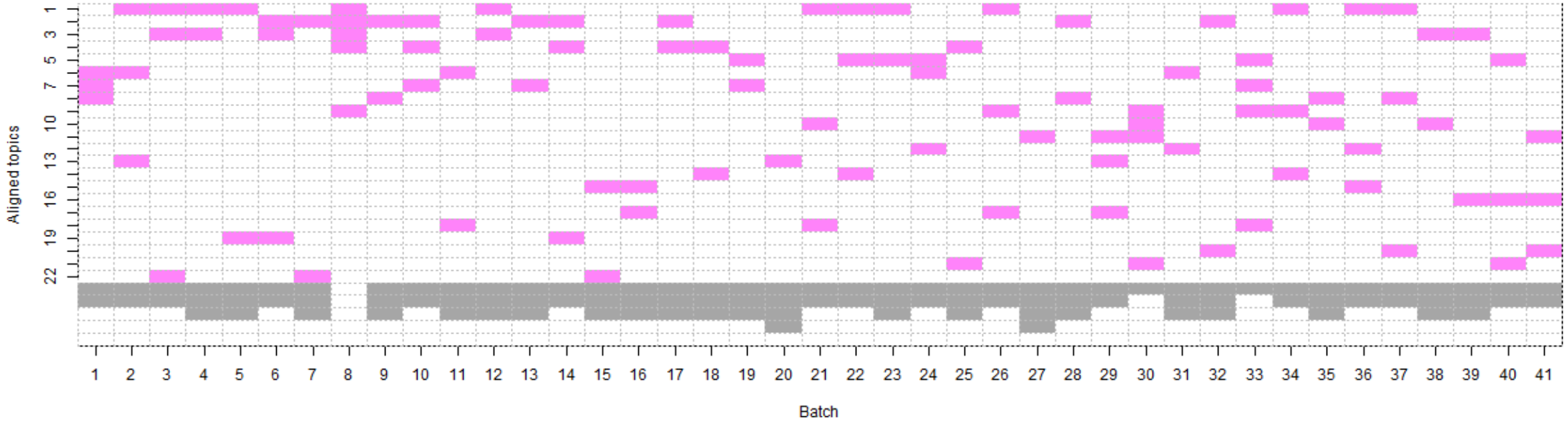
Data science

- Divided corpus into 41 batches
- Used NMijF process to derive 5 topics per batch
- 5 topic solution seemed most meaningful (also tried 10 and 20 topics per batch)



Not all topics are equally relevant

- Topics were associated with transitory environmental features, such as TV/news cycles (e.g., #insiders, #QandA) and extreme weather events Kirilenko et al., 2015; Sisco et al., 2017)
- We were ***most*** interested in topics which repeatedly emerged throughout 2016.



Identifying repeated topics

Used a topic alignment algorithm (Chuang at al., 2015) to identify topics which occurred in multiple batches

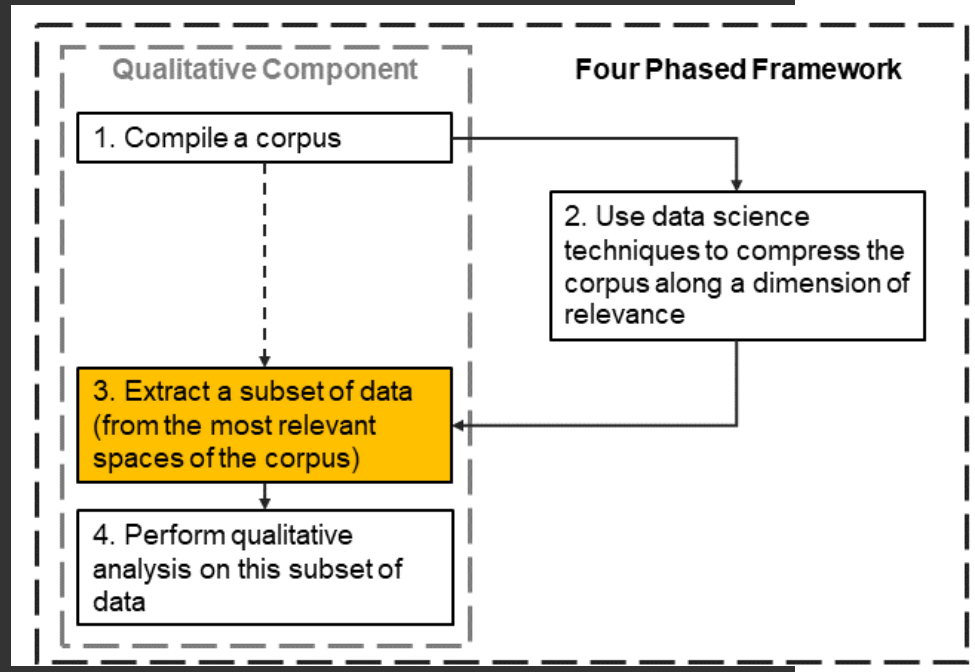
Common topics defined as topics which shared 3+ keywords

We were interested in topics which occurred throughout the year (i.e., in at least three batches)

https://github.com/AndreottaM/TopicAlignment

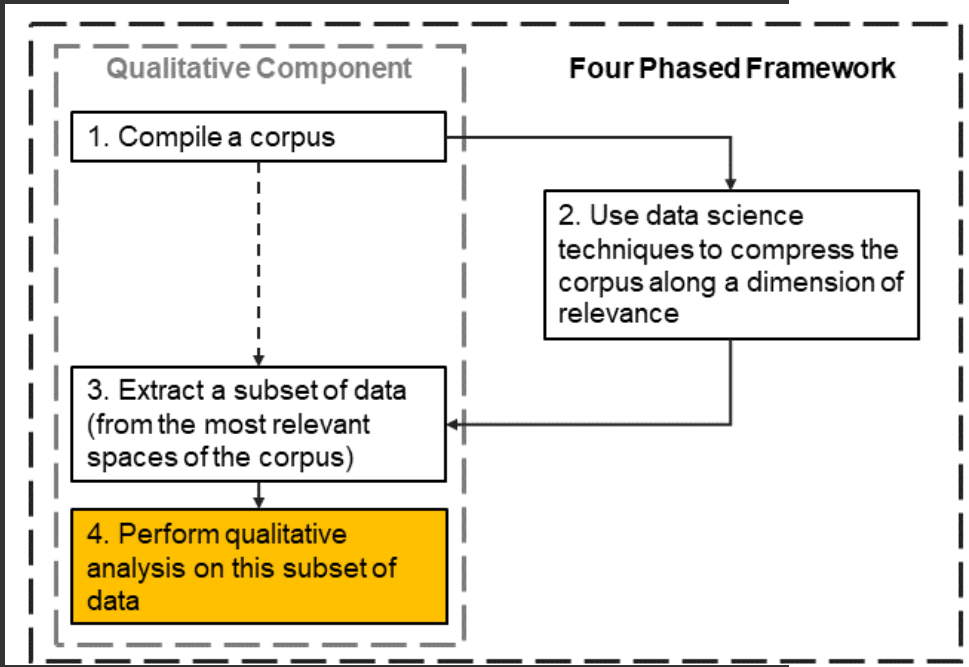
The grouped topics

Group	Total batches	Proportion of corpus (%)	Common keywords
1	13	9.83	action need now
2	10	4.92	#greatbarrierreef barrier great reef
3	7	4.53	coal new
4	6	3.00	action pai plan real
5	6	2.35	denial malcolm nation new one robe senat
6	5	3.54	hottest new record year
7	5	3.47	#qanda energi need renew
8	5	2.93	fight govt one peopl
9	5	2.45	#parisagr agreement pari ratifi time world
10	4	2.23	#qanda emiss health impact need polici risk talk
11	4	1.68	extrem link make now power renew scientist weather
12	3	1.84	debat nation one senat
13	3	1.75	action believ malcolm peopl real world
14	3	1.62	impact iss malcolm peopl
15	3	1.58	#qanda need real reef
16	3	1.54	#qldpol #scienc #wapol denier need year
17	3	1.53	flood iss now polici scientist
18	3	1.45	action impact klein naomi world
19	3	1.39	malcolm repo risk scientist warn
20	3	1.38	latest stop thank world
21	3	1.26	latest peopl planet thank year
22	3	1.21	level rise sea



Extracting a subset of data

- Randomly sampled 10 tweets from each ‘topic group’
- Trade-off between comprehensiveness and feasibility
- Allows us to sample from the most informative areas of our corpus (cf., responses to rare events)



Thematic analysis

- Reading the subset of data
- Code for themes
- Develop a codebook
- Have another rater apply codes, reach consensus on each tweet

- Use keywords and topics to inform initial generation of themes
- Use topics to form abductive hypotheses to test with thematic analysis

Themes

39.6 %

Climate change (CC) action

“Yes! Let’s start working together for real solutions on climate change #QandA”

19.6 %

Consequences of CC

“Reefs of the future could look like this if we continue to ignore #climatechange”

15.9 %

Conversations on CC

“not so gripping from No Principles Malcolm. Not one mention of climate change in his pitch.”

12.7 %

CC deniers

“According to Senator elect Malcolm Roberts, NASA fiddles the figures on Climate Change.”

9.1 %

The legitimacy of CC and climate science

“Do we have an international convention on ‘Cloud Seeding’? Or it comes under United Nation’s climate change agreement?”

Themes vs. Topics (e.g., sea level rise)

Consequences of CC

“Want to see more flooding in Qld and Paris (& everywhere else)? That s what #climate change is bringing.”

Conversations on CC

“Trouble is climate change deniers use weather info to muddy debate. Careful???????????????”

Mixed-methods framework

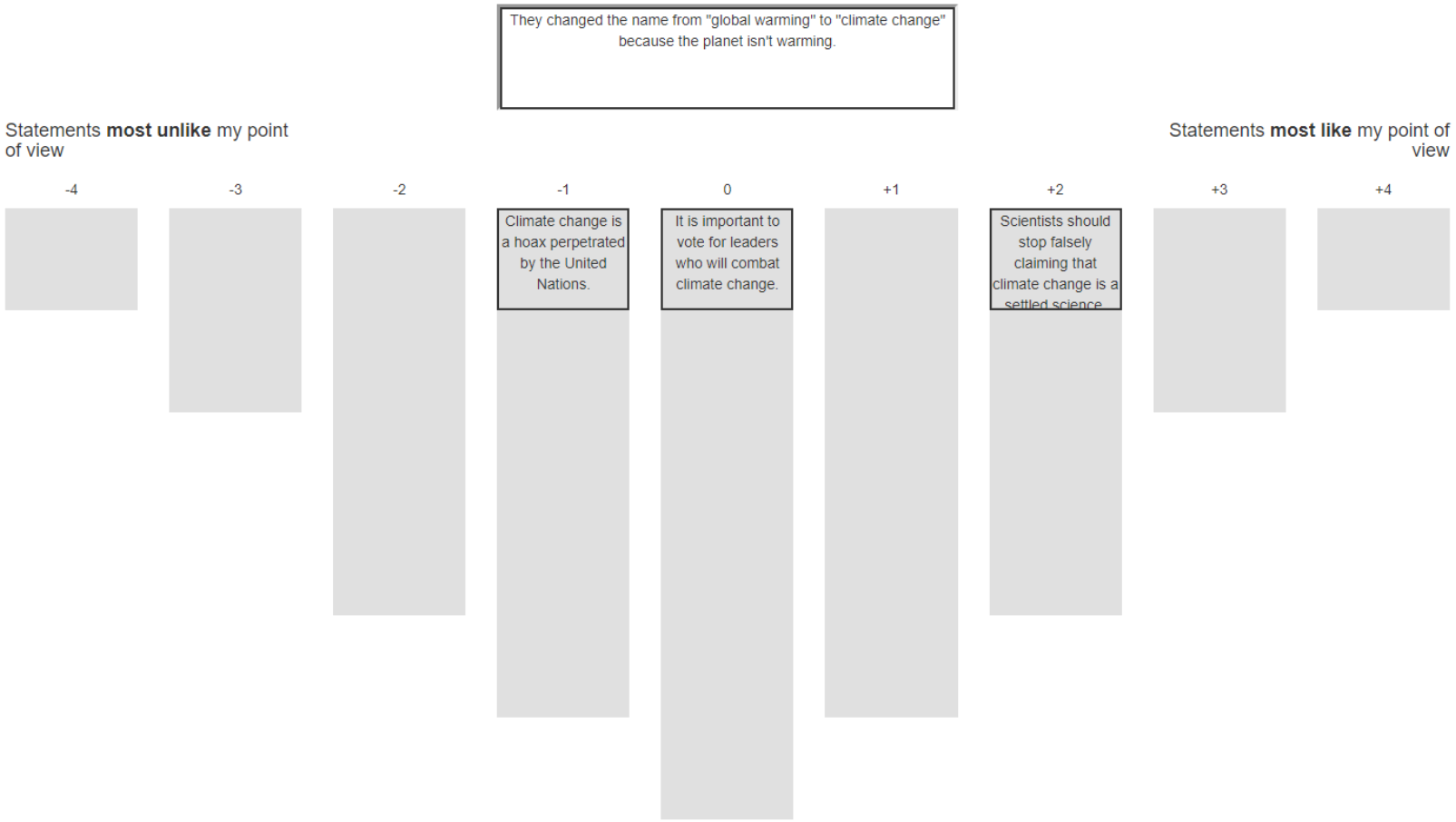
How did Qual approaches improve Quant?

- Could conduct analyses with an understanding of socio-political events, context, theories, etc.
- Could decide the ‘conceptual granularity’ of a topic/theme
- Can control where topics/themes fall on a semantic-latent dimension

How did Quant. approaches improve Qual?

- Identify the most relevant space of the corpus for answering our research question
- Forms the foundation for qualitative explorations of the data set

Andreotta, M., Nugroho, R., Hurlstone, M. J., Boschetti, F., Farrell, S., Walker, I., & Paris, C. (2019). Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. *Behavior research methods*, 1-16. https://doi.org/10.3758/s13428-019-01202-8.

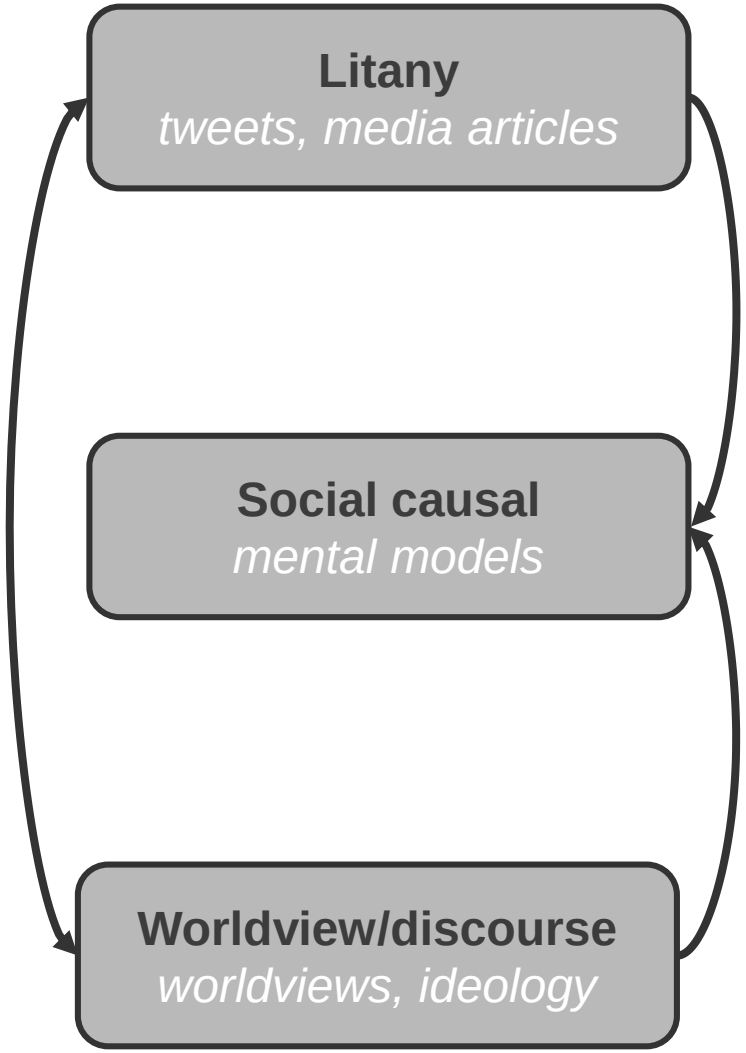


Next study: Segmenting the public

For each theme, we extract 6 litanies (broadly representing the diverse views of Twitter users)

Participants complete a card sorting task.

We will identify unique profiles of climate change interpretations



Predicting segment membership

Understand the predictors of interpretative profile membership.

What is a better predictor of climate change perceptions? The mental models people have of the climate, or their underlying political ideology, worldviews, etc.?

Different strategies and challenges for updating beliefs rooted in ideology vs. rooted in explanations.

Recommended Reading

Causal Layered Analysis (CLA)

- Inayatullah, S. (2004). The causal layered analysis (CLA) reader: Theory and Case Studies of an Integrative and Transformative Methodology.
- Bishop, B. J., & Dzidic, P. L. (2014). Dealing with Wicked Problems: Conducting a Causal Layered Analysis of Complex Social Psychological Issues. *American Journal of Community Psychology*, 53(*1-2*).

Thematic Analysis (TA)

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(*2*), 77–101.

Thank you

http://tiny.cc/xup88y

This research was supported by an Australian Government Research Training Program (RTP) Scholarship from the University of Western Australia and a scholarship from the CSIRO Research Office to me, as well as a grant from the Climate Adaptation Flagship of the CSIRO awarded to Dr. Mark Hurlstone and Dr. Iain Walker.