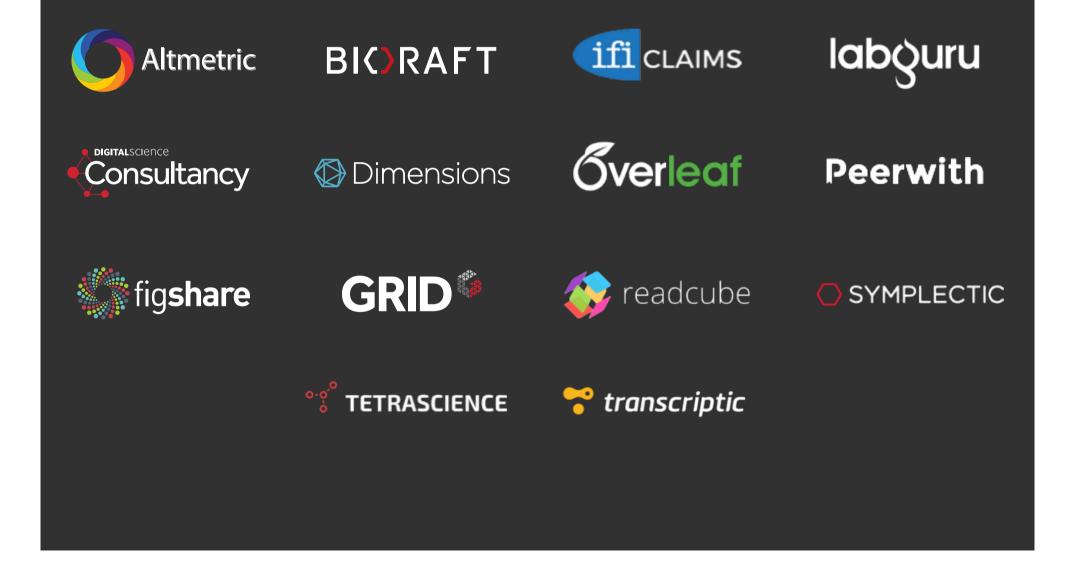
Visualising topic modelling.. Simplifying or complexifying?

Hélène Draux





Part of the **Digital Science** family



The Consultancy Team



Prof Jonathan Adams Chief Scientist Former Director of Research Evaluation at Thomson Reuters Advised on research metrics to governments around the world



Simon Porter VP Academic Relationships and Knowledge Architecture Formerly of U of Melbourne Worked in Research administration, library, IT



Dr Martin Szomszor Consultant Data Scientist One of Top 50 UK Information Age data leaders 2015



Dr Phill Jones Director of Publishing Innovation Former academic scientist and scholarly publisher Works with SSP, the STM association, ALPSP



Aaron Sorenson Bibliometrics Engagement Leader

Formerly of Temple U and GE Healthcare

Scientometrics editor of Journal of Alzheimer Disease



Dr Hélène Draux Data Scientist – Research Data management and visualization expert

Geodata specialty



Mike Taylor Head of Research Metrics

Former Senior Product Manager, Informetrics at Elsevier Works with FORCE11, RDA,

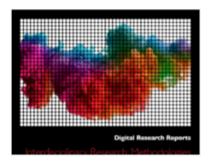
and NISO



Our Customers and Collaborators



Digital Research Reports



Interdisciplinary Research: Methodologies for Identification and Assessment



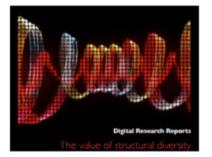
Examining Implications of Brexit for the UK Research Base



The Societal and Economic Impacts of Academic Research



The Implications of International Research Collaboration for UK Universities



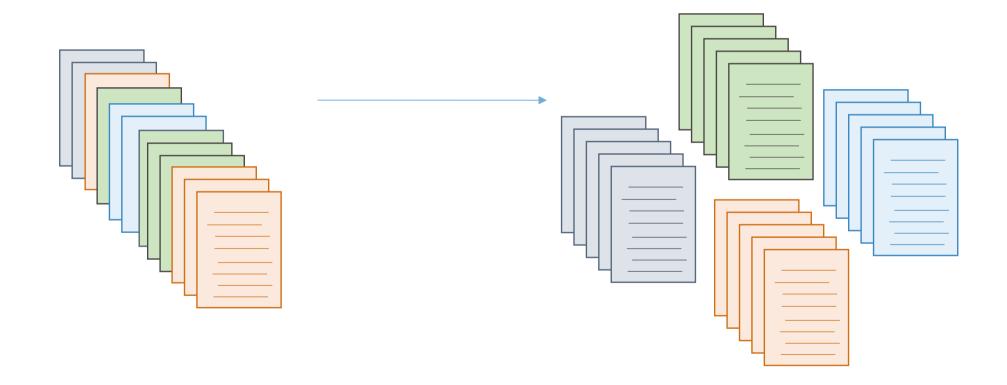
The Value of Structural Diversity



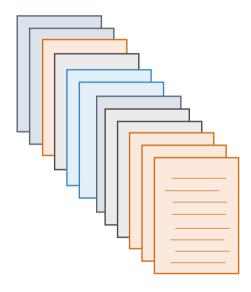
The Diversity of UK Research and Knowledge



Aggregating similar documents together (= unsupervised classification)

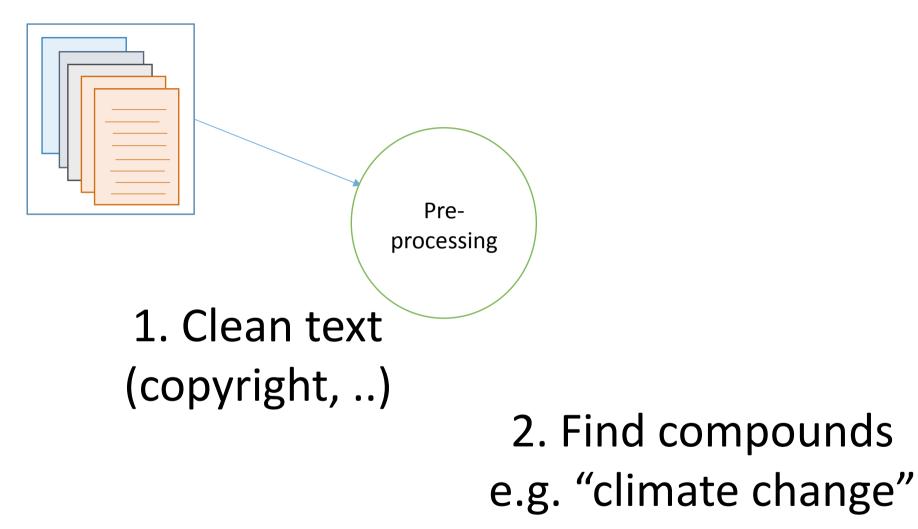




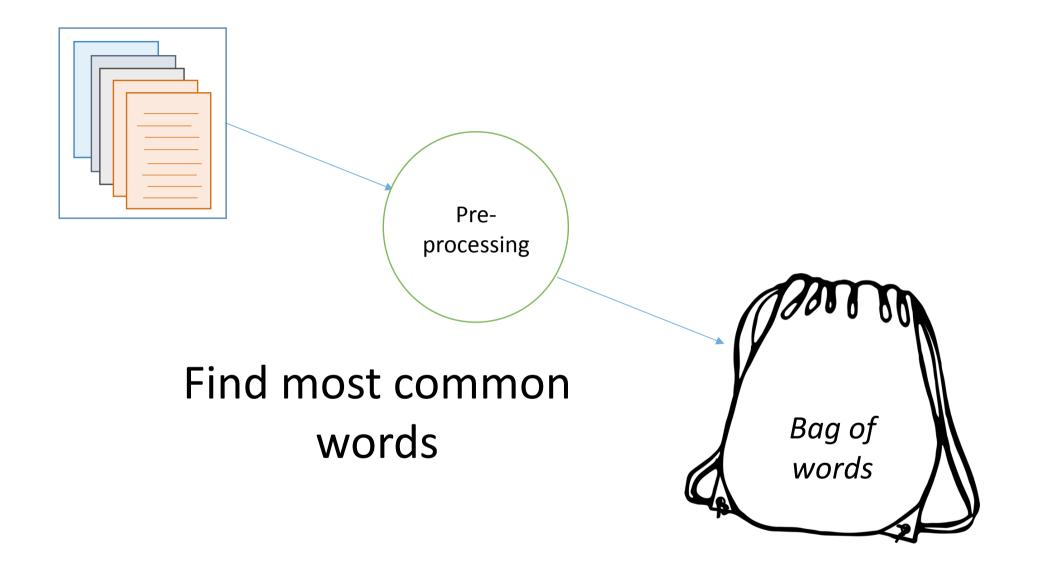


Corpus of documents

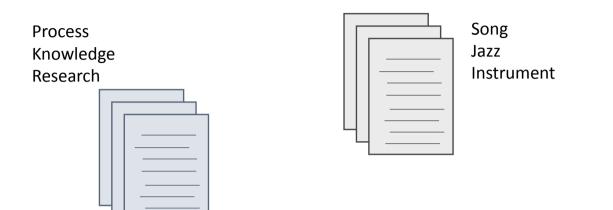








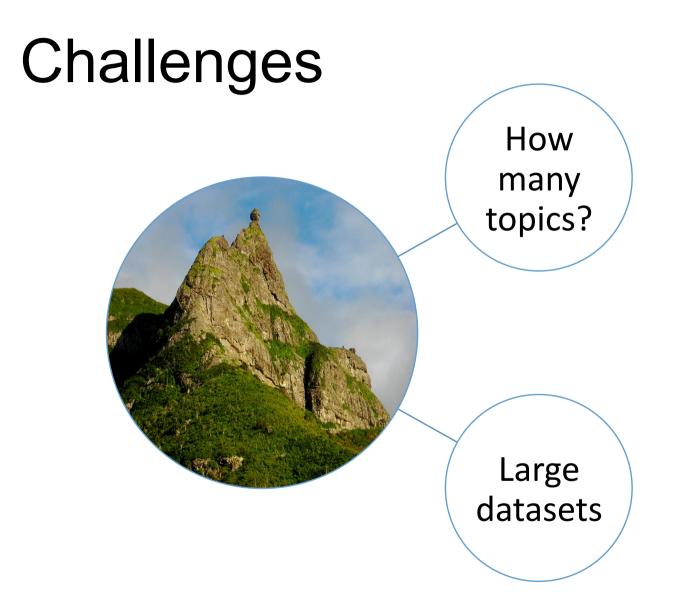




History Heritage Gallery

Group by term frequency







Number of topics



Depends on the purpose



Depends on the homogeneity of the data



Explore transition – Alluvial diagram

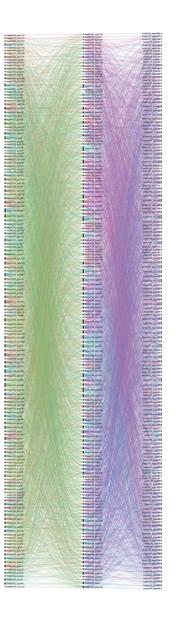
	model26_topic6	model27_topic21
model25_topic22	model26_topic22	model27_topic22
model25_topic6	model26_topic2	model27_topic6
model25_topic2	model26_topic18	model27_topic18
model25_topic24	model26_topic21	model27_topic2
model25_topic10	model26_topic25	model27_topic10
model25_topic18	model26_topic7	model27_topic1
model25_topic7	model26_topic10	model27_topic7
model25_topic9	model26_topic24	model27_topic26
model2.9_upped	model26_topic9	model27_topic24
model25_topic3	model26_topic15	hodel21_topic24
model25_topic4	model26_topic4	model27_topic4
model25_topic15	model26_topic1	model27_topic9
model2.3_upic rd		model27_topic16
model25_topic1	model26_topic23	model27_topic15
model25_topic23	model26_topic14	model27_topic14
model25_topic14		model27_topic23
mode/25_topic8	model26_topic8	model27_topic8
model25_topic16	model26_topic3	model27_topic25
model25_topic21	model26_topic5	model27_topic3
model25_topic20	model26_topic20	model27_topic20
	model/26_topic11	
model25_topic11		model27_topic11
model25_topic5	model26_topic0	model27_topic5
model25_topic13	model26_topic13	model27_topic13
model25_topic17	model/26_topic17	model27_topic17
model25_topic12	model26_topic12	model27_topic12
model25_topic19	model26_topic16	model27_topic0
model25_topic0	model26_topic19	model27_topic19

- Follow the terms
- Terms in model with:
 - 25 topics
 - 26 topics
 - 27 topics



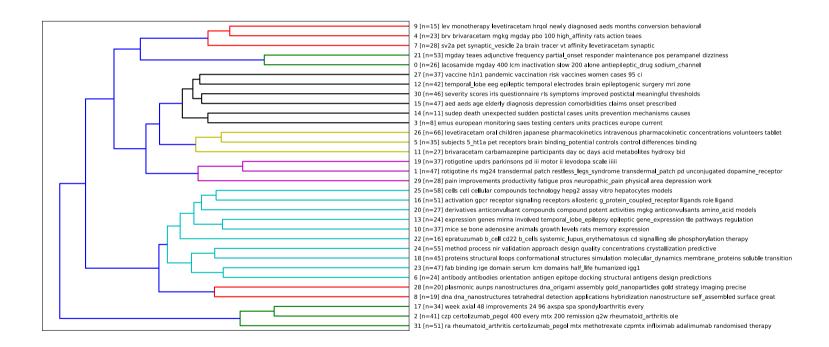
Hard to scale up

• 145 to 146 to 147



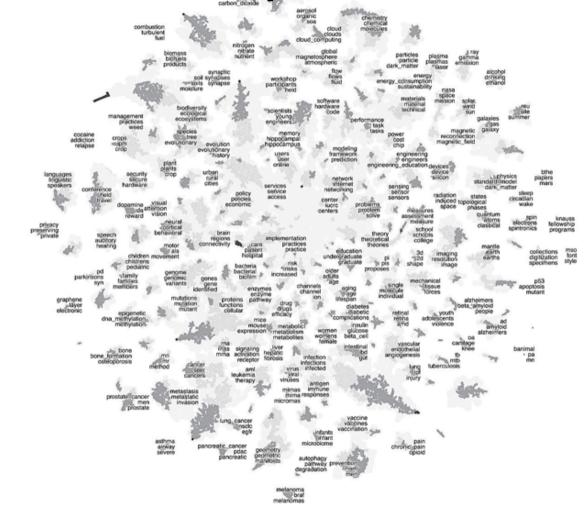


Clusters of topics - *Dendrogram*





Overview - Landscape





Visualisation of topic models

Simplifies the models

.. but some visualisations make them more complicated

