Visualising topic modelling..
Simplifying or complexifying?

Hélène Draux
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

DIGITALscience Consultancy
Make better decisions faster
Our Customers and Collaborators
Aggregating similar documents together (= unsupervised classification)
Corpus of documents
1. Clean text (copyright, ..)
2. Find compounds e.g. “climate change”
Find most common words
Group by term frequency
Challenges

- How many topics?
- Large datasets
Number of topics

- Depends on the purpose

- Depends on the homogeneity of the data
Explore transition – *Alluvial diagram*

- 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