A Topography of Climate Change Research

Max Callaghan

MCC

November 28, 2017

MCC

- To contribute evidence-based policy-making on climate change, the IPCC aims to *comprehensively* assess
- These assessments should be aim to balance legitimacy, credibility and relevance (Cash and Clark, 2001)
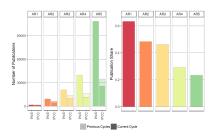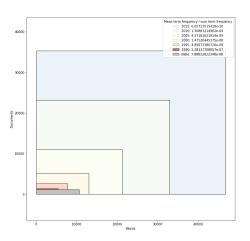
Figure: Source: Minx et al. (2017)

- Comprehensive, credible and relevant assessments become more challenging as the literature grows

To understand, and to aid, scientific assessments of climate change, we need to machine read the literature

Approach - Words, words, words

- Topic modelling is a way of reducing the dimensionality of a corpus of documents
- A large matrix of documents x words is factorised by a matrix of topics x words and a matrix of topics x documents (Lee and Seung, 1999)
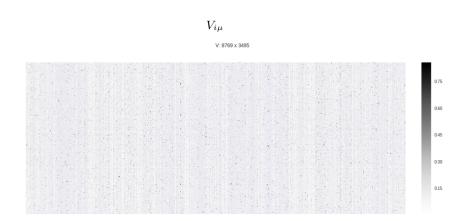- Topics describe the latent structure of the document corpus

$$V_{i\mu}$$

V: 8769 x 3495

Figure: A topic model of 3495 documents on climate change from the year 2000

MCC

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia}H_{a\mu}$$
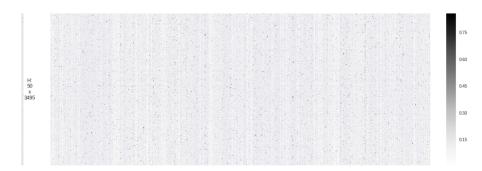
V: 8769 x 3495

H:
50
x
3495

0.75

0.60

0.45

0.30

0.15

V ≈ WH

W: 8769 x 50

Figure: A topic model of 3495 documents on climate change from the year 2000

- What topics are discovered endogenously from a corpus of $> 200{,}000$ abstracts on climate change?
- How do these topics describe the structure of the literature?
- Which topics have grown recently?
- Which topics are better and less well represented in IPCC assessment reports?

Preliminary results - explanation

Biomass carbon density of hunan typical forest types

Chen, J.; Li, X.; Wang, F.; Zeng, Y.; Zeng, Z.; 2016

The forest carbon reserve is very important to forest ecosystems. The amount of carbon of forest plays an important role in improvement of the global warming. Both field surveys and laboratory analysis were employed to investigate biomass and biomass carbon density in six typical forest types (Cupressus funebris forest, Eucalyptus forest, Pinus massoniana forest, Cunninghamia lanceolata forest, Quercus fabri forest and Populus tremula forest) of the Hunan Province. Results show that the biomass, biomass carbon and carbon density of the selected six forest types increase with the increasing ages. The carbon density per unit for young forests, middle forests and premature-mature-overmature forests of each forest type were as follows: 30.1, 73.4 and 121 t/hm2 in Cupressus funebris forest, 25.6, 39.7 and 97.1 t/hm2 in Eucalyptus forest, 17.7, 48.4 and 80.9 t/hm2 in Pinus massoniana forest, 22.5, 43.9 and 99.5 t/hm2 in Cunninghamia lanceolata forest, 16.6, 19.6 and 59.1 t/hm2 in Quercus fabri forest, and 16.6, 26.7 and 53.7 t/hm2 in Populus tremula forest. Because the forest types in Hunan Province are mainly in the young and middle-aged forest stands, the biomass carbon density is regarded to increase. This study provides important information for forest management and evaluation of carbon sequestration. © 2016, World Food Ltd. and WFL Publishers. All Rights Reserved.

related topics

{forest, tropic, deforest}

{biomass, aboveground, root}

{carbon, sequestr, organ}

{food, agricultur, secur}

Adjust topic threshold: 0.002

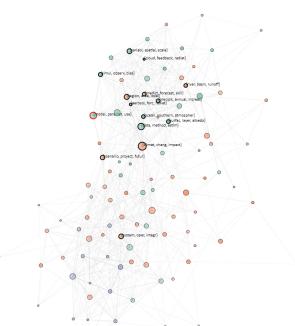- Documents are mixtures of topics, based on the words which occur in them

MCC

Primary Working Group
- 1
- 2
- 3

- A network of comprehensible topics is generated with 100 topics
- Topics can be matched to the IPCC working group from which the majority of the topic documents are referenced in
- Topics from the same working group are **significantly** more likely to be correlated with each other than those which are not
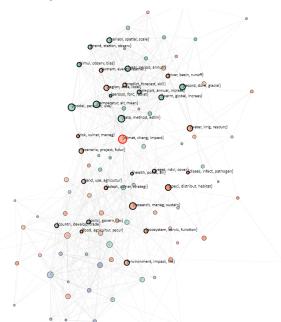
Preliminary results - structure - WGI

Primary Working Group
- 1
- 2
- 3

- The largest topic in WGI is on models.

Preliminary results - structure - WGII
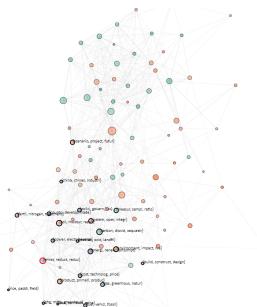

Primary Working Group
- 1
- 2
- 3

- The largest primarily WGII topic is on climate change impacts
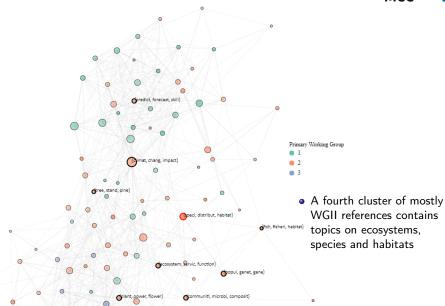
Preliminary results - structure - WGIII

- The largest WGIII topic is on emissions reductions.

MCC

Primary Working Group
- 1
- 2
- 3

{predict, forecast, skill}

{climat, chang, impact}

{tree, stand, pine}

{speci, distribut, habitat}

{fish, fisheri, habitat}

{ecosystem, servic, function}

{popul, genet, gene}

{plant, power, flower}

{communiti, microbi, composit}

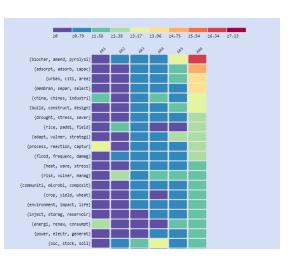- A fourth cluster of mostly WGII references contains topics on ecosystems, species and habitats

MCC

- In later assessment periods, a large meta-topic on research priorities and sustainability emerges
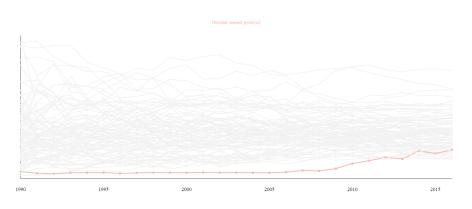
Preliminary results - growth

- Negative emissions related topics have shown strong growth since the end of AR5
- As have topics on cities and extreme weather events

- Biochar has emerged as an entirely new topic in the last 10 years

Preliminary results - gaps in coverage

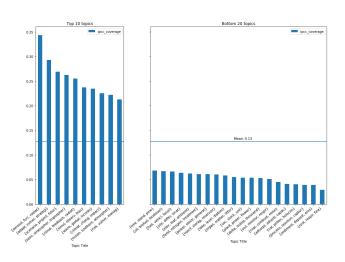How can we get a sense of which topics are better covered in IPCC reports?

- Each document d either matches or does not match an IPCC reference
- For each topic h, we can sum the scores for each category of document
- The "IPCC proportion" of each topic is the proportion of the sum of the document score accounted for by documents which match IPCC references.

Preliminary results - gaps in coverage

- The physical science aspects of climate change, as well topics on impacts, adaptation and scenarios are well covered by the IPCC

- Topics on specific technological solutions (particularly NETs), as well as soils, are less well covered
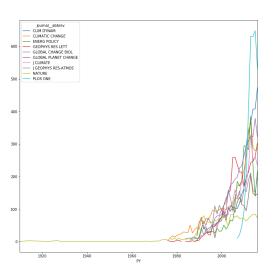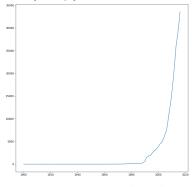
- Endogenously discovered topics make substantive sense of an unmanageable dataset of climate-relevant literature
- Topic modelling discovers over-arching topics such as that on sustainability and research priorities, as well as individual, fast growing topics such as biochar or CCS
- Topics from the same IPCC working group are more likely to be connected to each other.
- Quantitative evidence is found to support policy makers' dissatisfaction with a lack of 'solution orientation' in IPCC reports (Kowarsch et al., 2017)
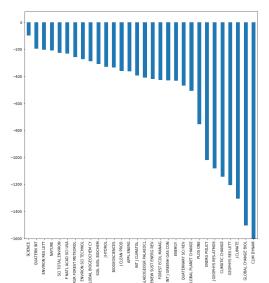
Extra results - Journals

- Nature has been publishing about climate change for a long time
- Plos One has very recently overtaken all other journals
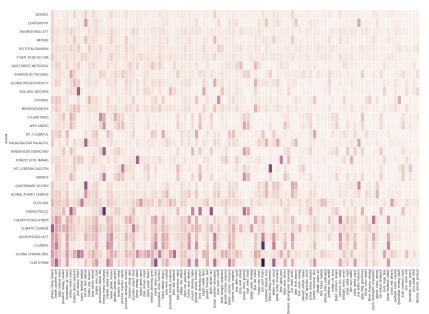- The number of journals has risen very steeply

Extra results - Journals

- Journal Entropy describes the distribution of topics in a journal (Hall et al., 2008)
- High values mean the journal deals with a wider range of topics

$$H(z|c,y) = -\sum_{i=1}^{K} \hat{p}(z_i|c,y) \; log \; \hat{p}(z_i|c,y)$$

Cash, D. W. and Clark, W. C. (2001). From science to policy : assessing the assessment process. *Social Science Research Network*, (November):1–45.

Hall, D., Jurafsky, D., and Manning, C. D. (2008). Studying the history of ideas using topic models. *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (October):363–371.

Kowarsch, M., Jabbour, J., Flachsland, C., Kok, M. T. J., Watson, R., Haas, P. M., Minx, J. C., Alcamo, J., Garard, J., Riousset, P., Pintér, L., Langford, C., Yamineva, Y., von Stechow, C., O'Reilly, J., and Edenhofer, O. (2017). A road map for global environmental assessments. *Nature Climate Change*, 7(6):379–382.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91.

Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J., and Edenhofer, O. (2017). Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy*.