

Growth and Diversification of Research Topics

30 Years of Climate Change Research

Carlo Debernardi*, Marco Seeber, Mattia Cattaneo

September 8, ESHS 2022 - Brussels



UNIVERSITÀ
DEGLI STUDI
DI MILANO



UiA Universitetet
i Agder



B E H A V E

Aim

Characterize, from a quantitative point of view, the process of growth and dissemination of the research topics

- Our data
- Why a focus on climate change
- What is a research topic
- Operationalisation: state of the art
- Our proposal
- A practical example: dynamic research topics

Data collection

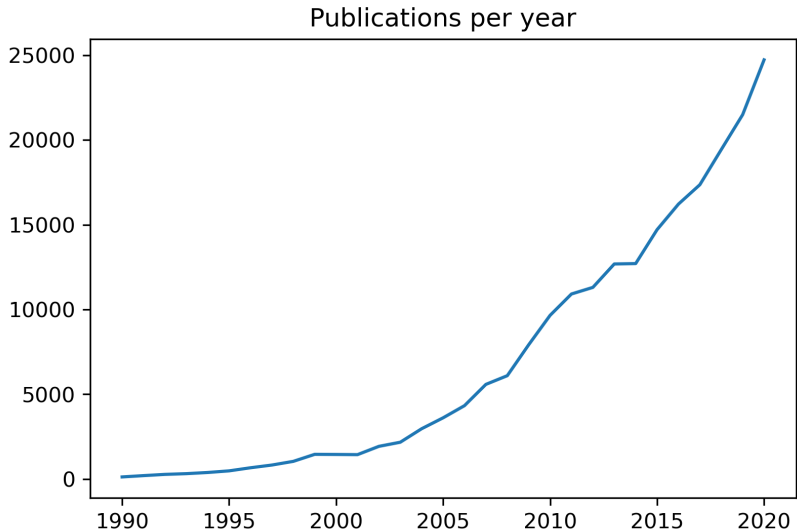
Scopus search over abstracts, titles and keywords for “*climate change*” and “*global warming*”

We then kept only articles, conference papers, reviews, book chapters and books (e.g. editorials, notes and letters are excluded), and removed all entries without an abstract or with insufficient textual information

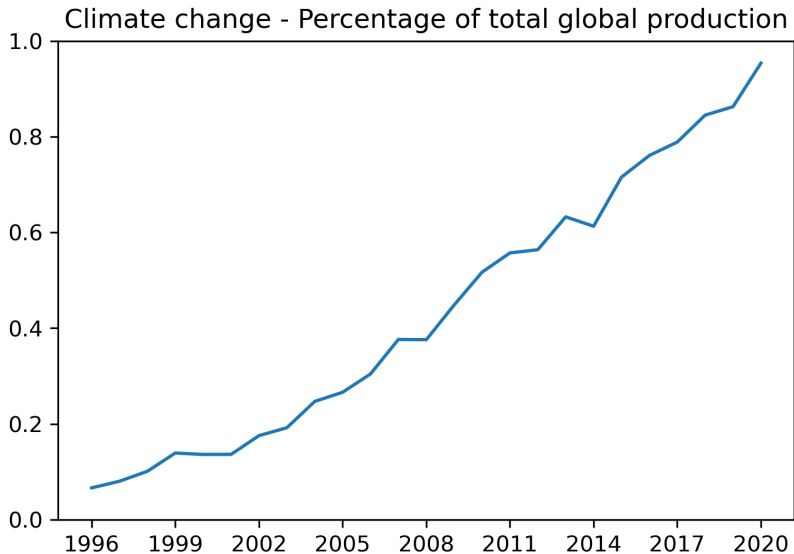
We retrieved metadata about 214,503 publications (after cleaning) from more than 400k authors

Time span: 1990 - 2020

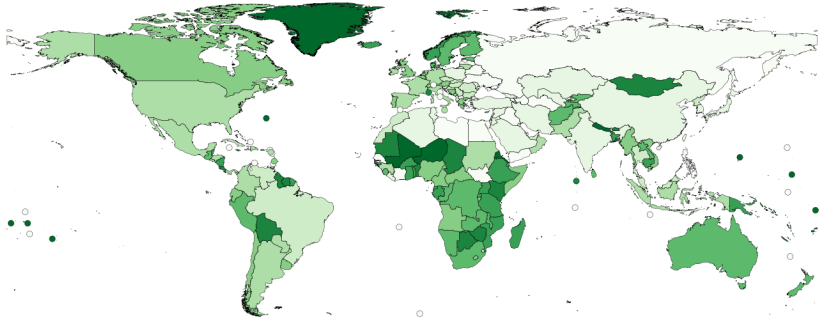
Why Climate Change



Why Climate Change



Why Climate Change



2016-2020 % of climate change research over country total publications

(darker color = higher value, max around 10%)

What is a research topic

Two main dimensions: **social** and **cognitive**

- “communities of one hundred members, sometimes considerably less”
- *Kuhn (1970)*
- “invisible college” of approximately 100 “core” scientists - *Price (1986)*
- “the easiest way to define a research specialty is through its social embodiment [...] self-organized network of researchers who tend to study the same research topics, attend the same conferences, read and cite each other’s research papers and publish in the same research journals” - *Morris and Veer Martens (2009)*

On the cognitive side: the level of abstraction is problematic and the terminology not well defined (eg. discipline, field, research area, topic etc.)

How to operationalise a research topic

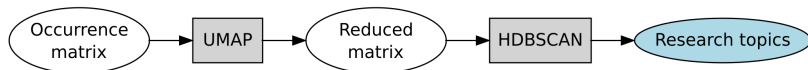
difficult conceptual definition + no "ground truth" = operational definition

Scientometrics		Topic Modeling (LDA)
Data	Citational relations	Textual data
Topic as	Cluster of documents	Mixture of words

Main limits:

- assumptions on properties of clusters or topics
- **number of clusters / topics as a parameter**
- forced attribution (no noise/uncertainty)
- interpretability

A quantitative approach to identify RTs



The document \times word occurrence matrix is reduced through UMAP (McInnes & Healy 2018)

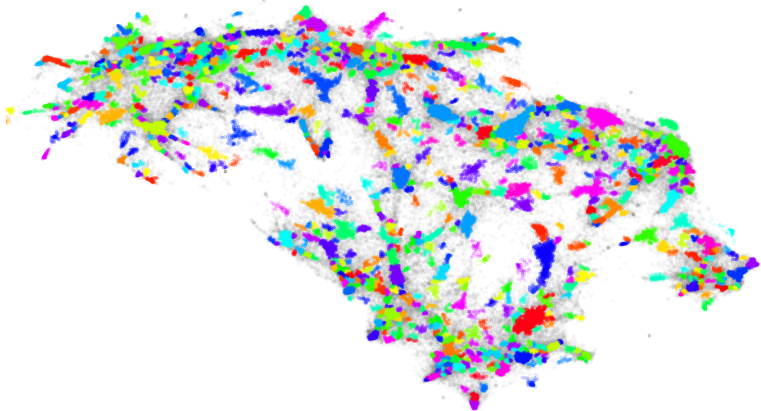
Nonlinear dimensionality reduction that preserves nearest neighbors

The resulting matrix is then used to define clusters with HDBSCAN (Campello et al. 2013)

Density-based clustering, noise-aware (around 70% of the points are classified)

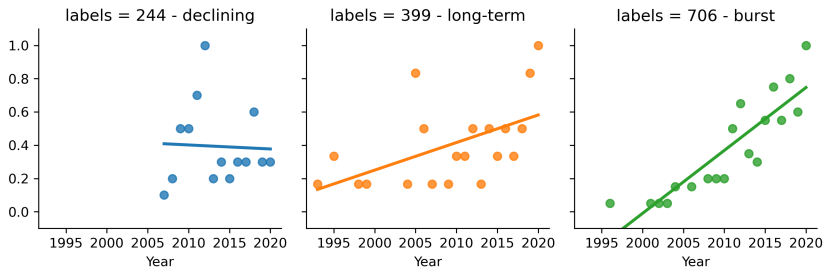
A little visual help...?

Embedding



2d visualization - 871 research topics - how to make sense of this?

Growth dynamics of RTs



244: optimization of production (eg. steel, “clean coal”) and water distribution processes

399: ocean warming, surface temperature and ozone hole

706: risk management and natural disasters (cyclones, flooding etc)

Collaboration patterns

Do research topics diffuse over collaboration networks?

- Germination vs contagion (dependent on the level of abstraction)

How can we measure the cohesion of a research community?

- Relative size of the giant component in the co-authorship network

Notes:

- Difference in giant components measured according to publications and authors (cohesion external to the topic)
- High levels of growth in publications seem to be associated with low cohesion

A typology of RTs growth

We can try to build a typology of research topics growth from different perspectives:

growth / cohesion	variety / diversity
normalized growth rate (publications)	interdisciplinarity
normalized growth rate (authors)	internationality
relative size of the giant component	institutional variety

Just as an example, let's take the RTs above the average in all of those categories. . .

The most dynamic RTs

- 633 : effect on fauna of water acidification, temperature and salinity
- 132 : climate change and public health (esp. respiratory diseases and allergies)
- 497 : farming and climate change (esp. new pests and development of resilient crops)
- 022 : farming and climate change (esp. mycotoxins affecting maize)
- 706 : risk management and natural disasters (cyclones, flooding etc)
- 009 : biomass and biofuel, sustainability and perspectives
- 042 : glacial lakes outburst floods, moraine dams, landsliding
- 674 : climate change, governance and future perspectives
- 821 : economics and policies (esp. energy supply, carbon and fuel tax)
- 784 : climate justice, anthropocene, archaeology of climate change

Conclusions

- Lack of discussion on the conceptual definition of RTs (level of abstraction)
- Quantitative methods are not necessarily bounded to a coarse-grained definition of RTs
- Admitting more (smaller) RTs is theoretically justified and empirically promising
- A fine-grained analysis of this kind might be a useful tool for practitioners and scholars interested in the dynamics of certain research areas

Next step: develop a fully-fledged typology of the possible growth patterns

Thanks for your attention

carlo.debernardi.w@gmail.com