Bibliometric analysis of TreesLab scientific production

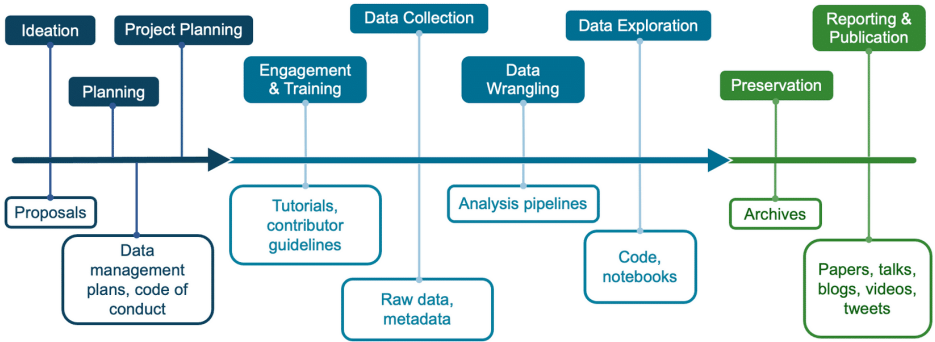
Alber Sánchez alber.ipia@inpe.br
Guilherme Mataveli guilherme.mataveli@inpe.br

Research assistant - TreesLab
National Institute for Space Research - INPE
Brazil

January 5, 2024

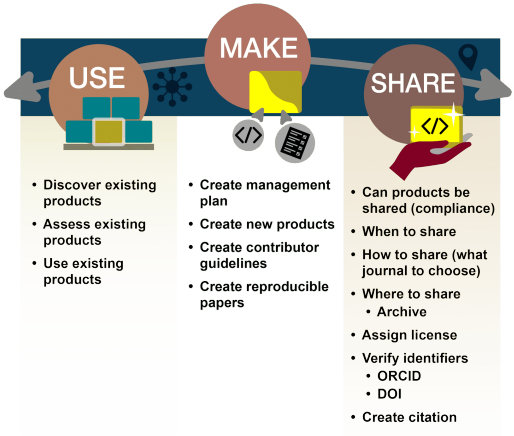
Overview

The research cycle



Source: OpenScience101.org

Use, make, share open results



Source: OpenScience101.org

What is bibliometric analysis?

- Bibliometrics is the measurement of physical units of publications, bibliographic citations, and surrogates for them [3].
- The bibliometric methodology encapsules the application of quantitative techniques (i.e., bibliometric analysis — e.g., citation analysis) on bibliometric data (e.g., units of publication and citation) [5].

Bibliometrix package

- R package for bibliometric analysis [1].
- It allows quantitative research in bibliometrics and scientometrics.
- Statistical analysis of publications.
- Useful for performance evaluation and policymaking.
- It includes a Web Application (biblioshiny) for non-programmers!

Bibliographic databases

- Scopus.
- Web of science.

ELSEVIER
Scopus

Clarivate
Web of Science™

Scopus and Web of Science at INPE

- To query them, use a computer at INPE and your Café login.

TreesLab publications

What constitutes a TreesLab's publication?

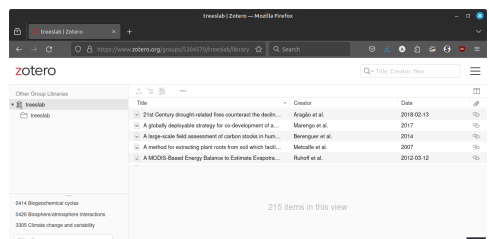
- Any publication whose authors agree to add it to the TreesLab's publication list.

Who are the members of TreesLab?

- Researchers, posdocs, Phd & masters students who consent to be part of the TreesLab.

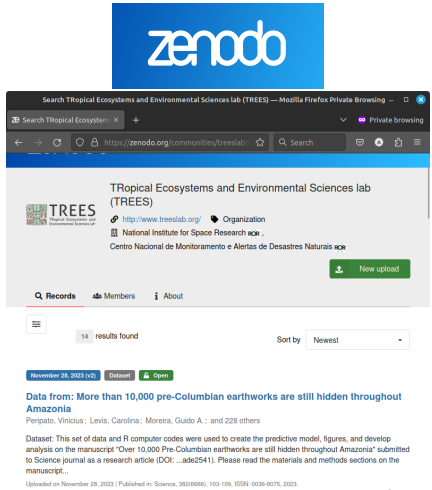
TreesLab publication list

- The publication list is available using Zotero, both online (click here) and as a desktop application.
- To add a publication, send its DOI to the TreesLab' mailing list.
- Publications wihtout DOI (e.g. GeoInfo, SBSR) should be added to online document *treeslab_sem_doi* (click here).
- TreesLab's Zotero Group is called *treeslab*.
- The former list (Mendeley) is deprecated.



TreesLab Publications' data

- TreesLab has a Zenodo Community (click here).
- To become a member, go to Zenodo, click on communities, search for *treeslab*, and follow the instructions.
- Upload your paper's data (no membership required).
- Make a request to the TreesLab community's administrators to include your paper's data.



Documents and references

- *Document (or citing document)*: Scientific document (article, review, conference proceeding, etc.) included in a bibliographic collection.
- *Reference (or cited reference)*: Scientific document included in at least one of the reference lists (bibliography) of the document set. Then *"a reference is cited by one or more documents"* [1].
- *Cited document*: Scientific document included in a bibliographic collection and, at the same time, it is cited in at least one other document in the collection. Cited documents are a subset of the reference set.

Global and local citations

Global citations.

- Measures the number of citations a document has received from documents contained in the entire database (e.g. WoS or Scopus).
- Measures the impact of a document in the whole bibliographic database.
- For many documents, a large part of global citations could come from other disciplines!

Local citations.

- Measures the number of citations a document has received from documents included in the analyzed collection.
- Is calculated analyzing the whole reference set.
- Measures the impact of a document in the analyzed collection.

Assumptions

- We assumed that TreesLab's publications are a study subject.

Method

1. Get TreesLab publications' DOIs.
2. Query Scopus and Web of Science.
3. Run analysis using *R*.

Overview

Description	Results
Timespan	2003:2023
Sources (Journals, Books, etc)	75
Documents	199
Annual Growth Rate %	13.68
Document Average Age	8.21
Average citations per doc	96.55
References	10842
Author's Keywords (DE)	500
Authors	1231
Authors of single-authored docs	1
Co-Authors per Doc	15.4
International co-authorships %	6.533

Documents by type

Description	Results
article	166
letter	9
review	8
conference paper	3
editorial	3
erratum	2
short survey	2
article; proceedings paper	1
biographical-item	1
correction	1
data paper	1
editorial material	1
note	1

Authors' productivity

Authors	Articles
ARAGÃO L	117
ANDERSON L	92
MALHI Y	61
SHIMABUKURO Y	33
PHILLIPS O	30
METCALFE D	21
ARAGAO L	20
ARAI E	19
BAKER T	19
MEIR P	19
QUESADA C	19

Authors	Articles Fractionalized
ARAGÃO L	12.65
ANDERSON L	12.25
SHIMABUKURO Y	5.34
MALHI Y	4.96
ARAI E	2.43
DE O G	1.82
MATAVELI G	1.79
WAGNER F	1.77
DALAGNOL R	1.66
ARAGAO L	1.64

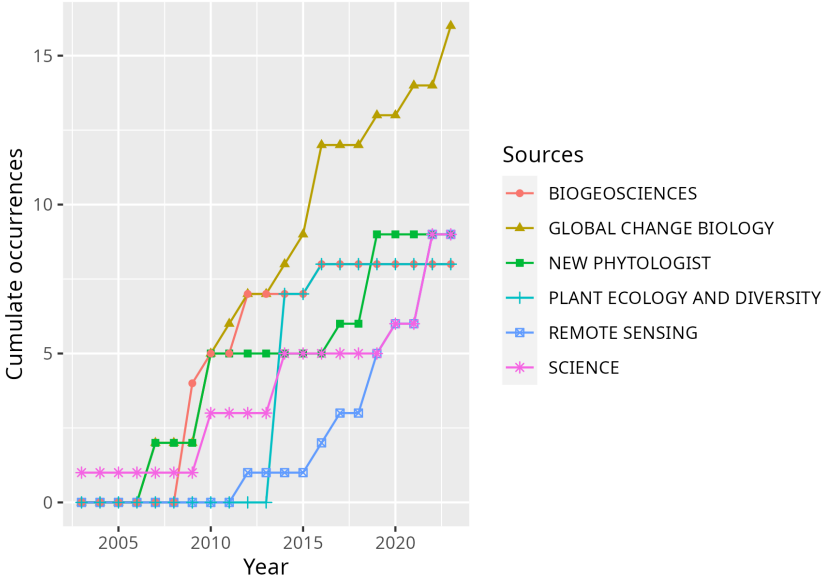
Most cited papers

Paper	TC	TCperYear	NTC
NEMANI RR, 2003, SCIENCE	2797	127.1	1.00
LUYSSAERT S, 2007, GLOBAL CHANGE BIOL	784	43.6	2.30
MORTON DC, 2006, PROC NATL ACAD SCI U S A	746	39.3	2.50
LUYSSAERT S, 2007, GLOBAL CHANGE BIOL1	743	41.3	2.18
BRIENEN RJW, 2015, NATURE	708	70.8	4.25
BARLOW J, 2016, NATURE	655	72.8	7.82
MALHI Y, 2009, PROC NATL ACAD SCI U S A	629	39.3	3.35
ARAGÃO LEOC, 2018, NAT COMMUN	445	63.6	5.29
PHILLIPS OL, 2010, NEW PHYTOL	421	28.1	2.88
ARAGÃO LEOC, 2007, GEOPHYS RES LETT	396	22.0	1.16

Most relevant sources

Sources	Articles
GLOBAL CHANGE BIOLOGY	16
NEW PHYTOLOGIST	9
REMOTE SENSING	9
SCIENCE	9
BIOGEOSCIENCES	8
PLANT ECOLOGY AND DIVERSITY	8
NATURE	6
PLOS ONE	6
SCIENTIFIC REPORTS	6
ENVIRONMENTAL RESEARCH LETTERS	5
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B: BIOLOG...	5

Production over time - sources

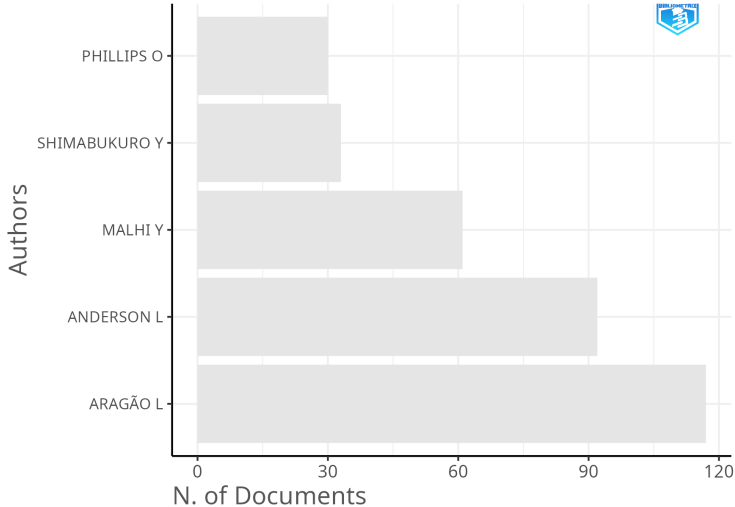


Most relevant keywords

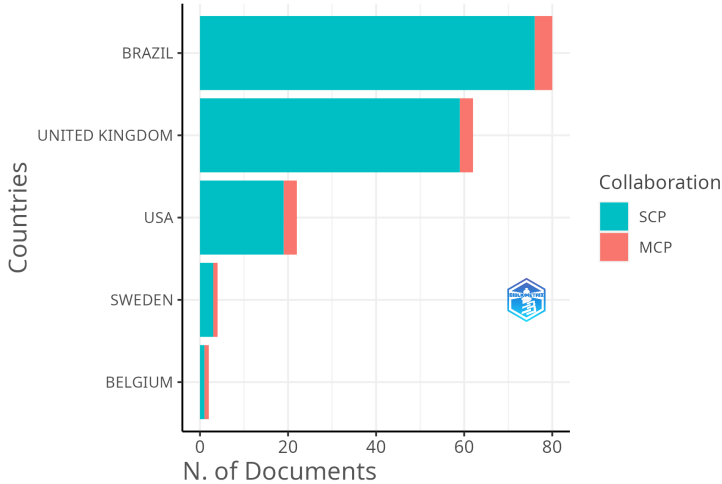
Author Keywords (DE)	Articles
AMAZON	25
DEFORESTATION	19
REMOTE SENSING	18
DROUGHT	17
FIRE	12
TROPICAL FOREST	12
TROPICAL FORESTS	12
CLIMATE CHANGE	11
MODIS	10
PHENOLOGY	10

Keywords-Plus (ID)	Articles
BRAZIL	100
DEFORESTATION	63
AMAZONIA	55
CARBON	54
CLIMATE CHANGE	52
DROUGHT	47
REMOTE SENSING	44
TROPICAL FOREST	41
BIOMASS	39
FOREST	37

Most Productive Authors

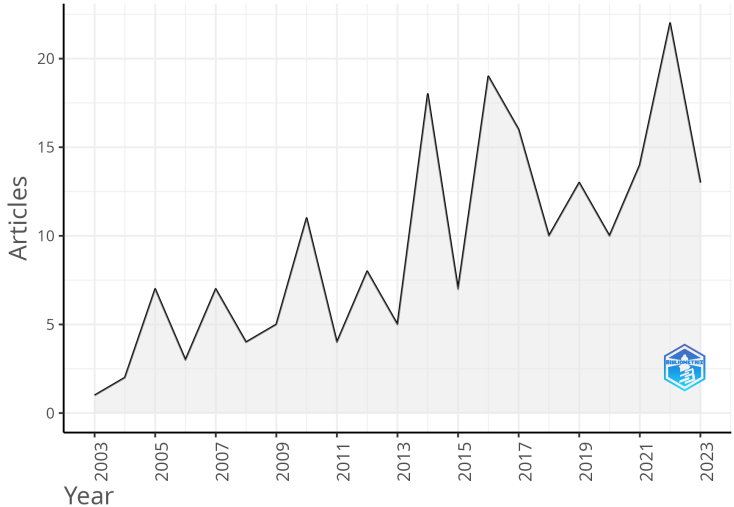


Most Productive Countries

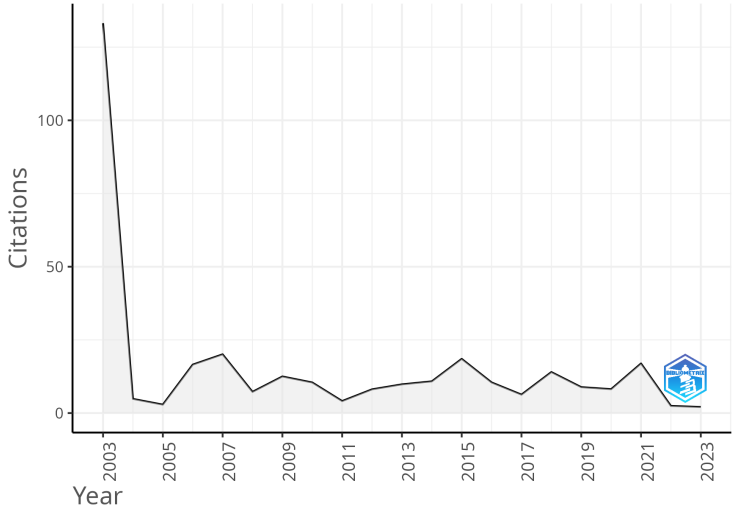


SCP: Single Country Publications, MCP: Multiple Country Publications

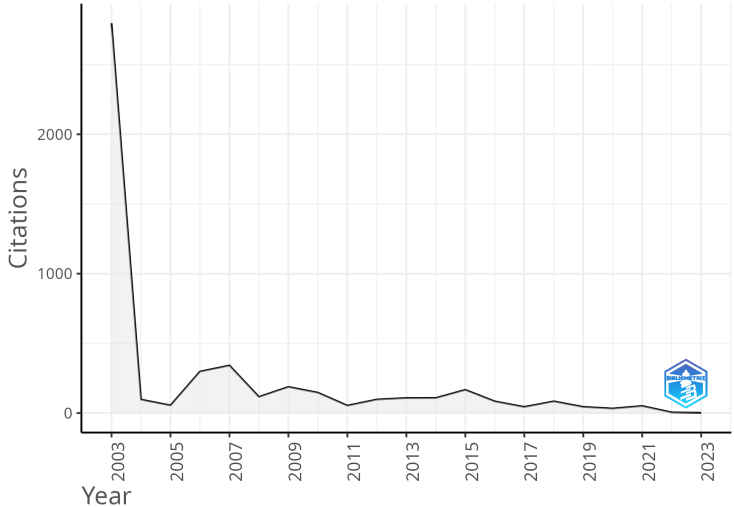
Annual Scientific Production



Average Article Citation per Year



Average Total Citation per Year



Structures of knowledge

- Science mapping aims at displaying the structural and dynamic aspects of scientific research [2].
- *Science mapping* allows investigating scientific knowledge from a statistical point of view:
  - *Conceptual*: What science talks about; themes and trends.
  - *Intellectual*: How the work of an author influences a given scientific community.
  - *Social*: How authors, institutions, and countries interact with each other.

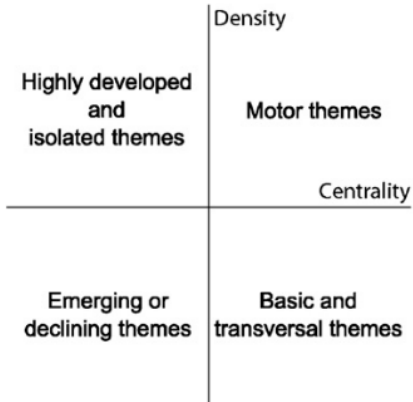
Conceptual structure

Represent relations among concepts or words in a set of publications.

- Words whih appear together in a document would be related in a network (co-words network).
- Factorial analysis helps to identify subfields by means of data reduction techniques.
- Mixed approach.

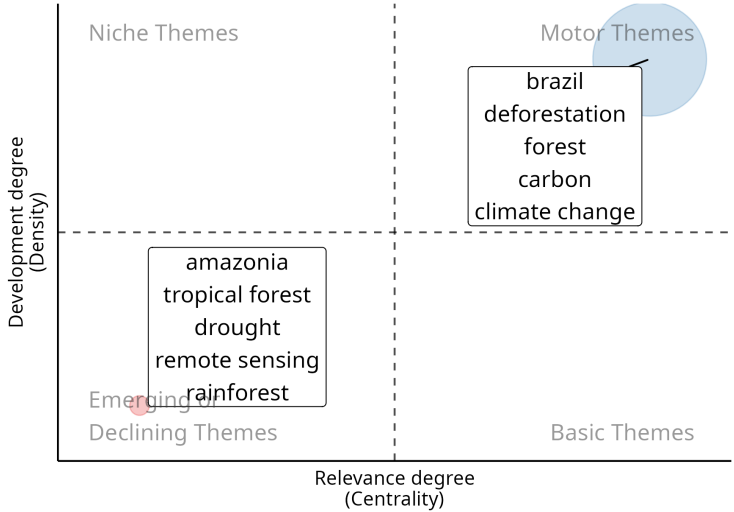
The strategic diagram

- Upper-right: Themes are related externally to concepts applicable to other themes that are conceptually closely related.
- Upper-left: Well-developed internal ties but unimportant external ties; marginal importance for the field.
- Lower-left: Mainly represents emerging or disappearing themes.
- Lower-right: Important for a research field but are not developed.

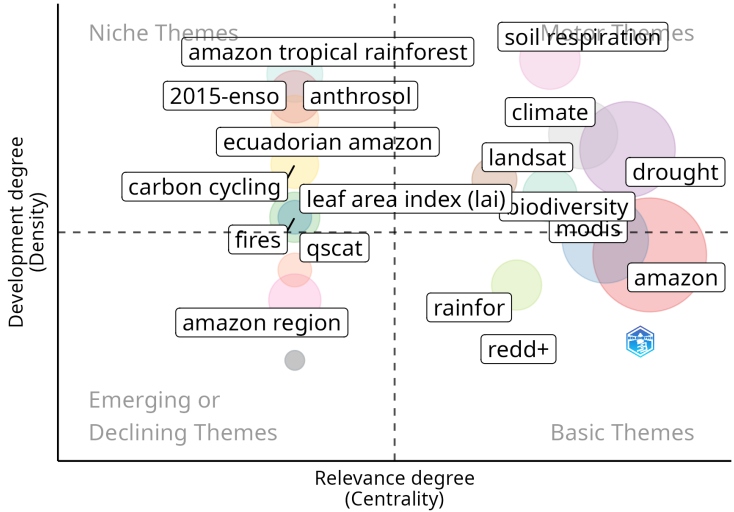


The strategic diagram. Source [4]

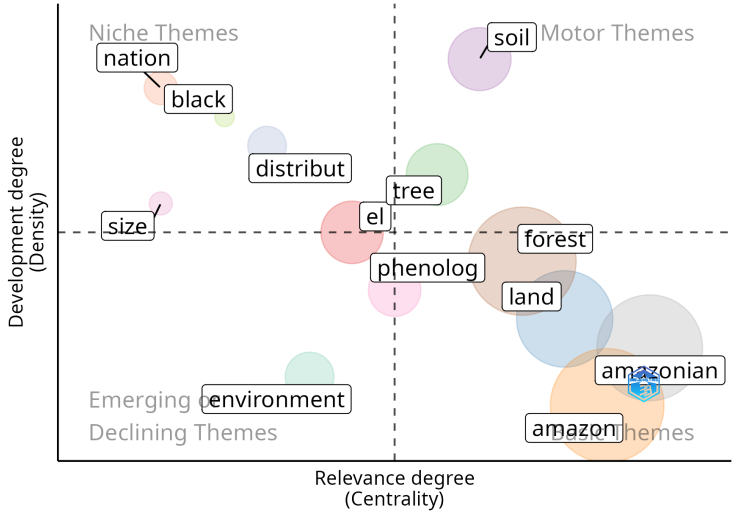
Thematic map (keyword plus)



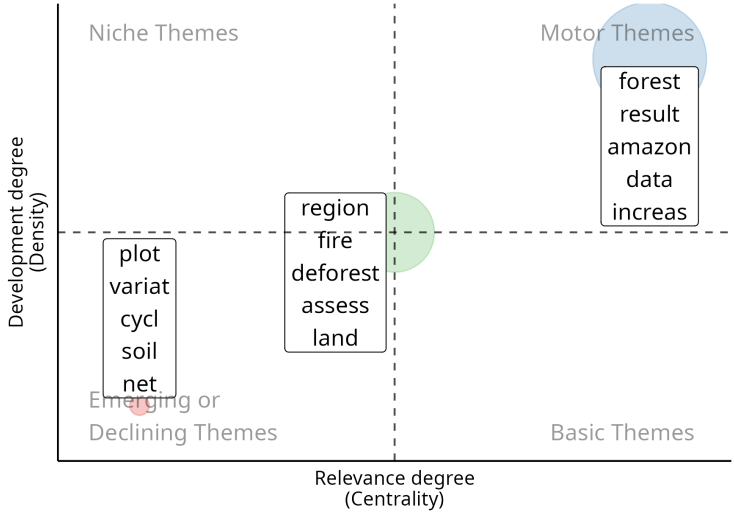
Thematic map (authors' keywords)



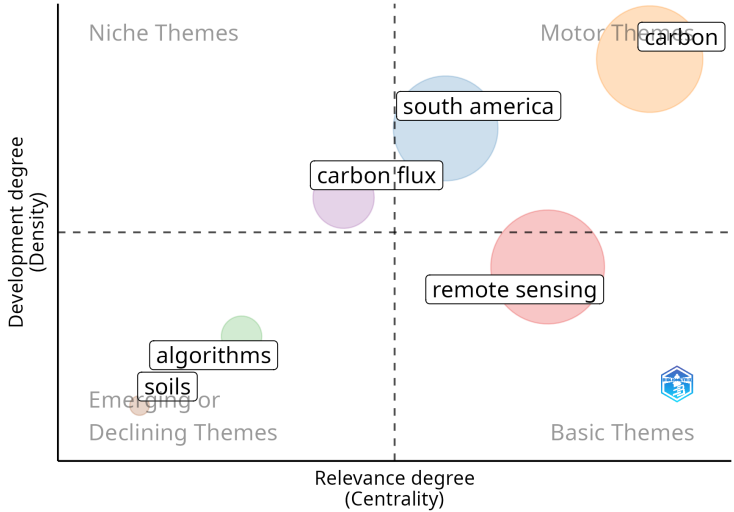
Thematic map (titles)



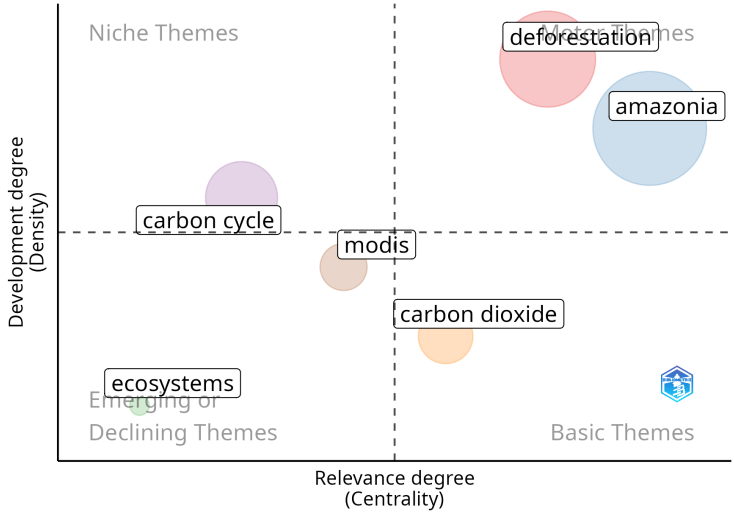
Thematic map (abstracts)



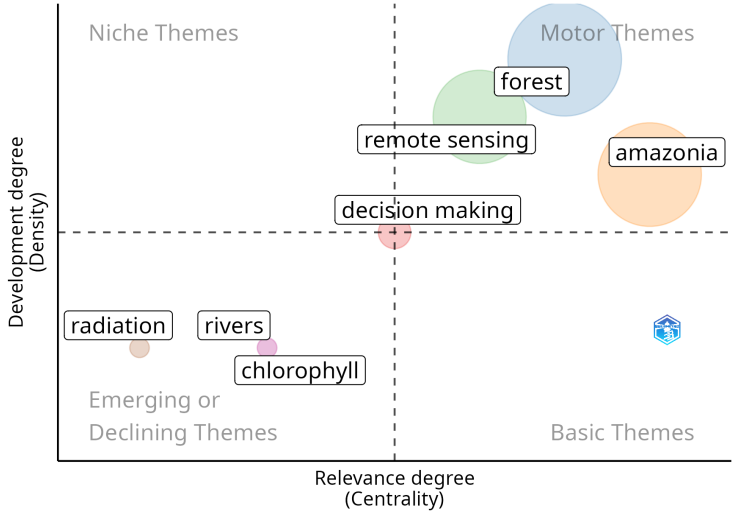
Thematic evolution - 1



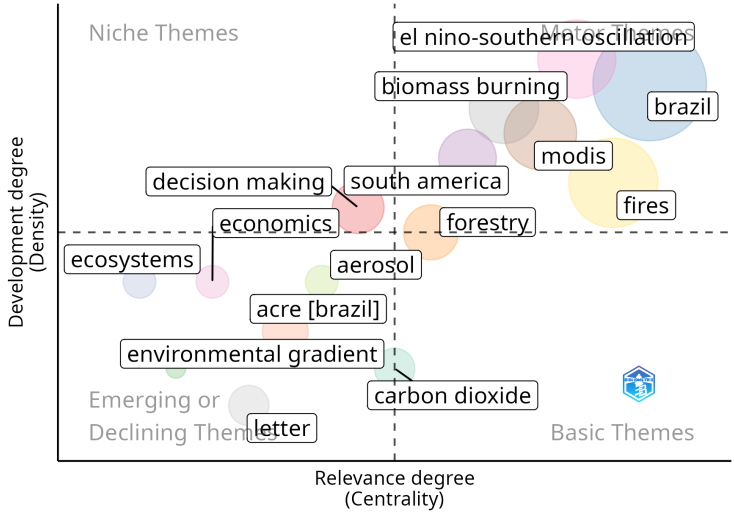
Thematic evolution - 2



Thematic evolution - 3



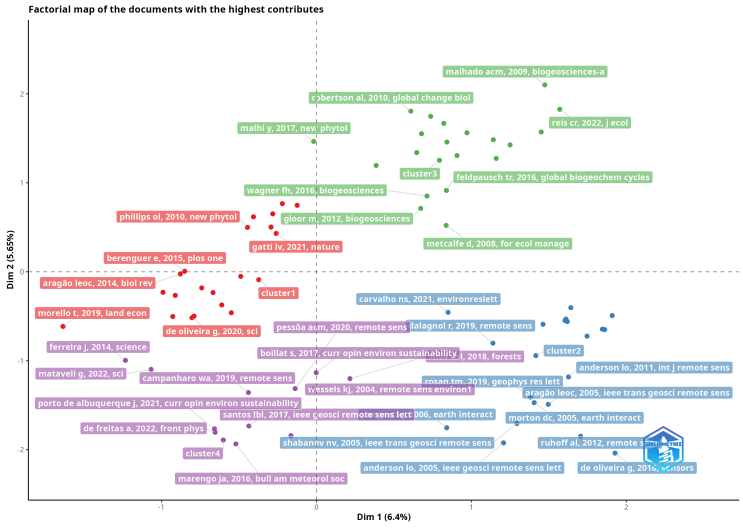
Thematic evolution - 4



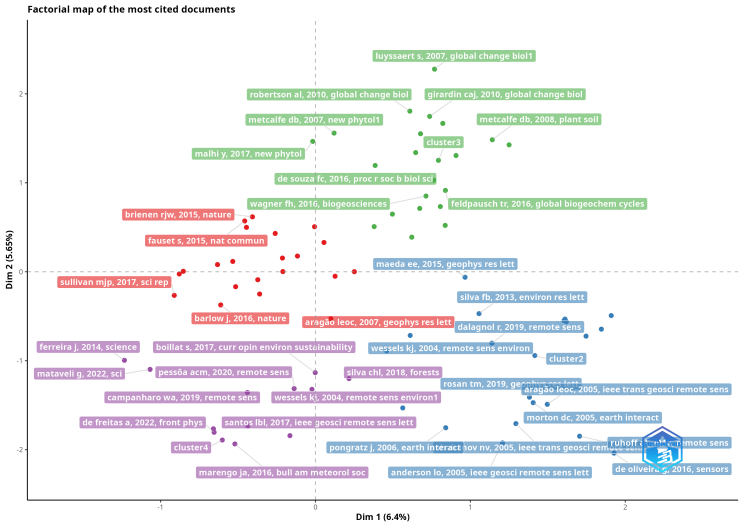
Factorial Map - Most contributing documents

- Identify the link between topic and documents.
- Plot the document associated to the highest absolute contribution.
- Absolute contributions measure the weight of each document in the information summarized by the two axes.
- The colors represent the clusters.

Factorial Map - Most contributing documents



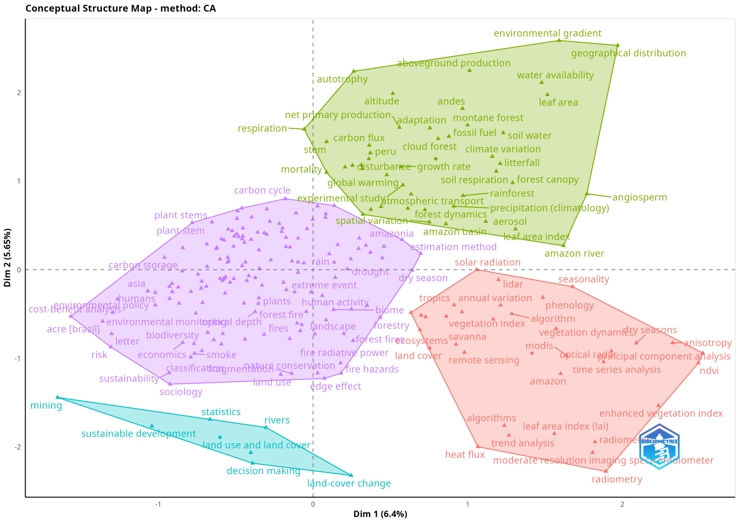
Factorial Map - Most cited documents



Map of words

- Clusters are identified by hierarchical clustering.
- Each color corresponds to a topic.

Map of words

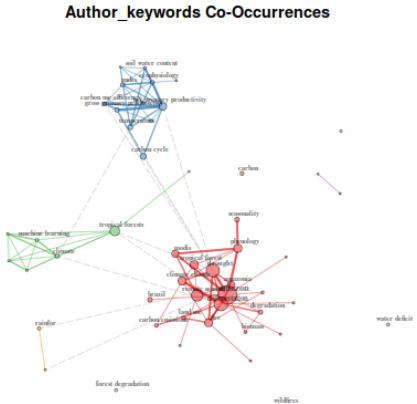


Conceptual Structure Map - method: CA
Dim 1 (6.4%)
Dim 2 (5.65%)
environmental gradient
geographical distribution
aboveground production
water availability
autotrophy
altitude
andes
leaf area
net primary production
adaptation
montane forest
respiration
fossil fuel
soil water
carbon flux
cloud forest
climate variation
stem
peru
disturbance
growth rate
litterfall
mortality
soil respiration
forest canopy
global warming
rainforest
experimental study
atmospheric transport
angiosperm
precipitation (climatology)
forest dynamics
aerosol
spatial variation
amazon basin
leaf area index
amazon river
carbon cycle
plant stems
plant stem
amazonia
estimation method
carbon storage
rain
drought
dry season
solar radiation
seasonality
asia
extreme event
lidar
humans
tropics
annual variation
environmental policy
cost-benefit analysis
plants
human activity
biome
phenology
forest fire
algorithm
environmental monitoring
optical depth
fires
landscape
forestry
vegetation index
acre [brazil]
biodiversity
forest fires
vegetation dynamics
dry seasons
letter
ecosystems
savanna
anisotropy
modis
risk
economics
smoke
fire radiative power
land cover
remote sensing
optical radar
principal component analysis
classification
fragmentation
nature conservation
fire hazards
time series analysis
ndvi
sustainability
land use
edge effect
amazon
sociology
mining
statistics
rivers
enhanced vegetation index
algorithms
leaf area index (lai)
sustainable development
radiometer
land use and land cover
trend analysis
heat flux
moderate resolution imaging spectroradiometer
decision making
land-cover change
radiometry

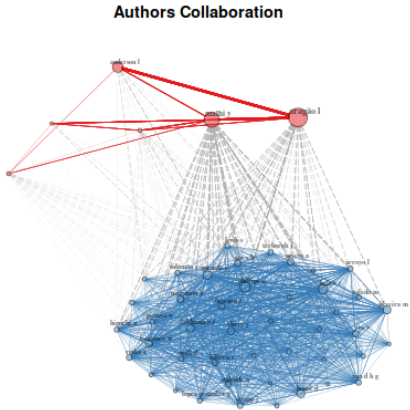
Network analysis

- A network is a representation of the co-occurrence matrix.
- Diagonal elements are the occurrences of each item in the collection.
- Non-diagonal elements are the co-occurrence of two item in a collection.

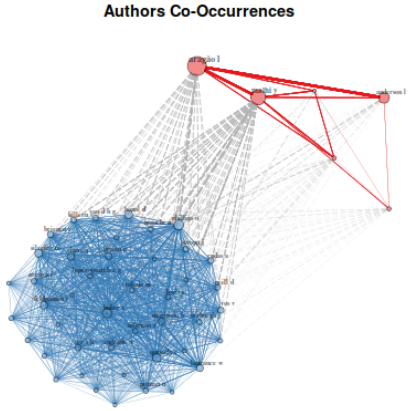
Network - Co-occurrences Authors-Keywords



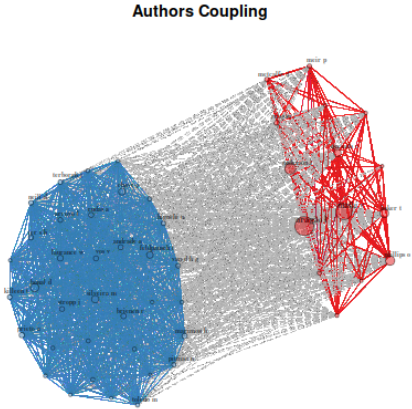
Network - Authors collaboration



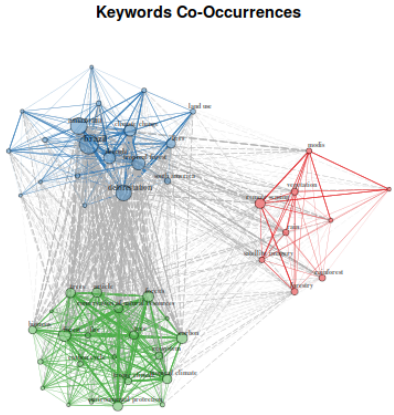
Network - Authors co-occurrences



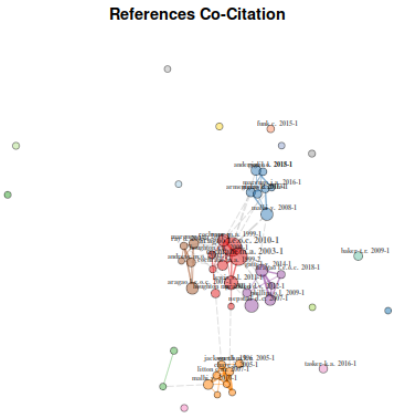
Network - Authors coupling



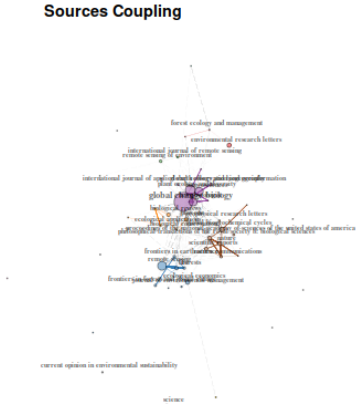
Network - Keyword co-occurrences



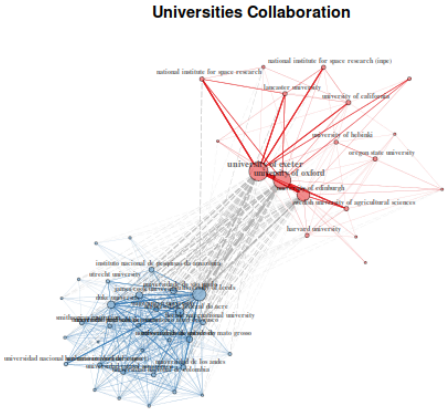
Network - References co-citation



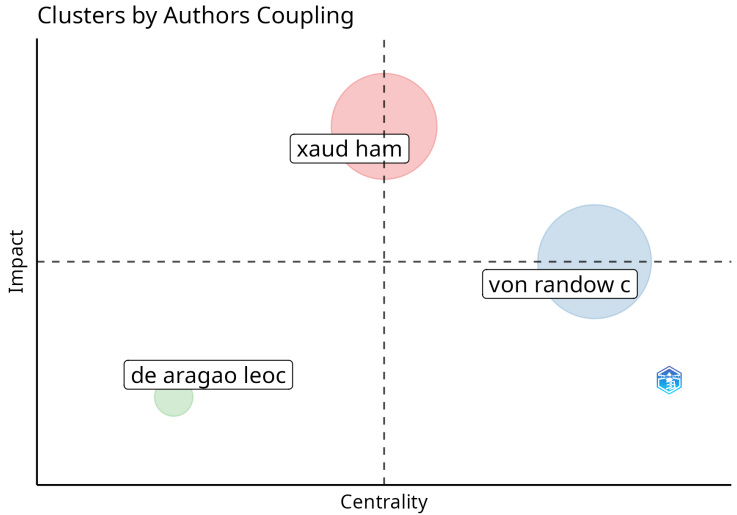
Network - Sources coupling



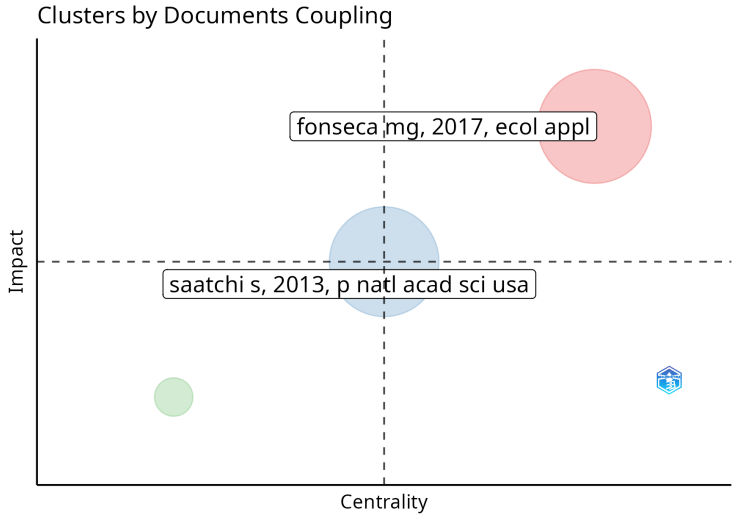
Network - Universities collaboration



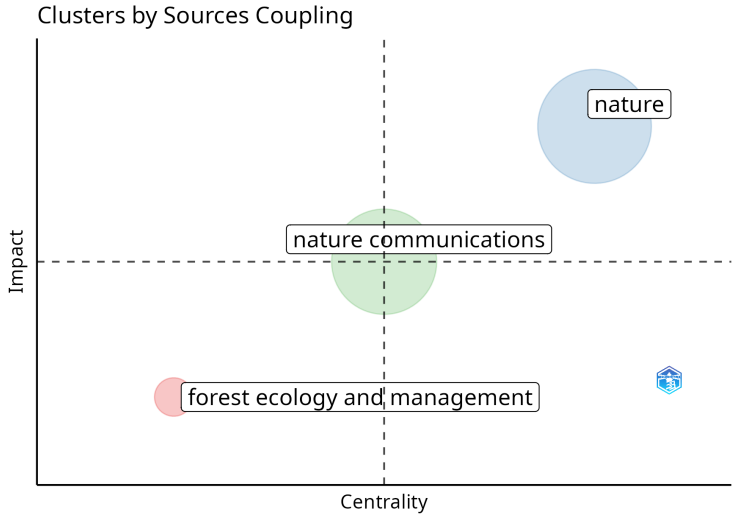
Coupling network analysis (authors)



Coupling network analysis (documents)



Coupling network analysis (sources)



Take home message

- TODO.

References I

[1] Massimo Aria and Corrado Cuccurullo. "Bibliometrix : An R-tool for Comprehensive Science Mapping Analysis". In: *Journal of Informetrics* 11.4 (Nov. 2017), pp. 959–975. ISSN: 17511577. DOI: 10.1016/j.joi.2017.08.007. (Visited on 11/14/2023).

[2] Katy Börner, Chaomei Chen, and Kevin W. Boyack. "Visualizing Knowledge Domains". In: *Annual Review of Information Science and Technology* 37.1 (Jan. 2003), pp. 179–255. ISSN: 0066-4200, 1550-8382. DOI: 10.1002/aris.1440370106. (Visited on 12/21/2023).

[3] R. N. Broadus. "Toward a Definition of "Bibliometrics"". In: *Scientometrics* 12.5-6 (Nov. 1987), pp. 373–379. ISSN: 0138-9130, 1588-2861. DOI: 10.1007/BF02016680. (Visited on 12/19/2023).

References II

[4] M.J. Cobo et al. "An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field". In: *Journal of Informetrics* 5.1 (Jan. 2011), pp. 146–166. ISSN: 17511577. DOI: 10.1016/j.joi.2010.10.002. (Visited on 12/27/2023).

[5] Naveen Donthu et al. "How to Conduct a Bibliometric Analysis: An Overview and Guidelines". In: *Journal of Business Research* 133 (Sept. 2021), pp. 285–296. ISSN: 01482963. DOI: 10.1016/j.jbusres.2021.04.070. (Visited on 12/11/2023).