Bibliometric analysis of TreesLab scientific production

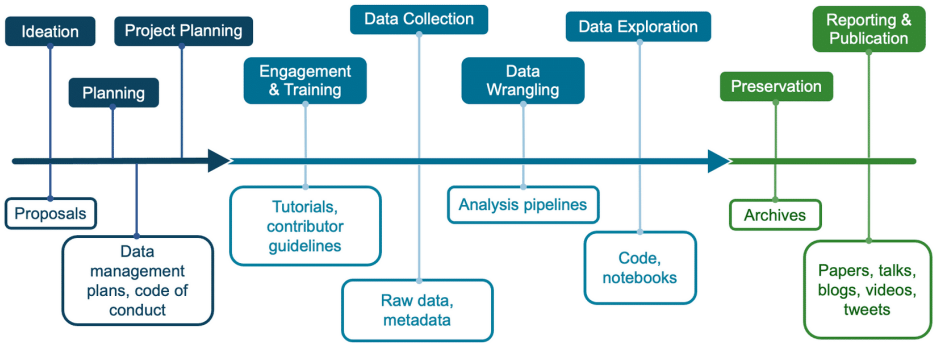
Alber Sánchez alber.ipia@inpe.br
Guilherme Mataveli
Debora Dutra

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

July 8, 2024

Introduction.

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].

Bibliographic databases

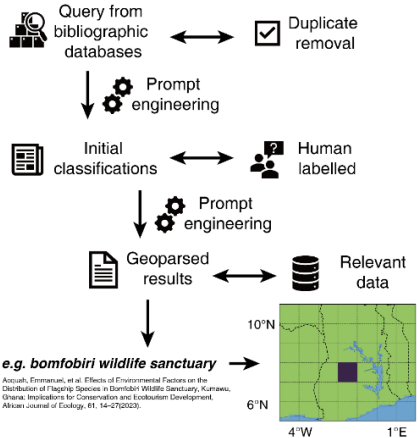
- Scopus.
- Web of science.

ELSEVIER
Scopus

Clarivate
Web of Science™

Bibliometrics and LLMs

- *Large language models reveal big disparities in current wildfire research* [6].
- This is a potential new trend in bibliometric analysis.
- Feed database data into LLM (ChatGPT).
- Extract, besides bibliometric information, localization of AOI.



Source: [6].

Method.

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 databases.
3. Run analysis.

ORCID
Connecting research and researchers

ELSEVIER
Scopus

Clarivate
Web of Science™

TreesLab publications

What constitutes a TreesLab's publication?

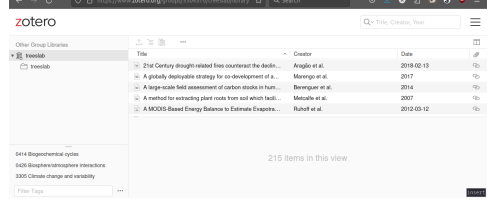
- Any publication whose authors agree to add it to the TreesLab's publication list.

Who are the members of TreesLab?

- Researchers, posdocs, Phd & masters students who consent on being part of the TreesLab.

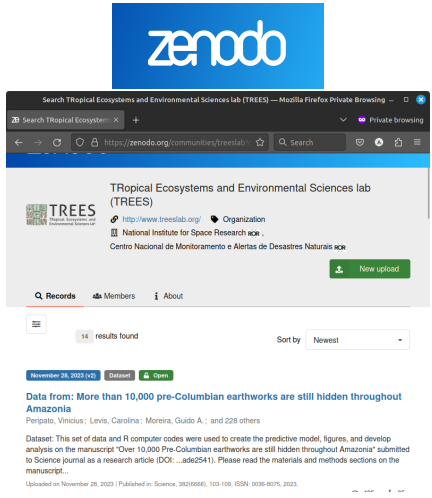
TreesLab publication list

- The publication list is available using Zotero, both online (click here) and as a desktop application.
- Currently, it has 306 items (2024-04-17).
- 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).
- Currently, it has 16 open and 2 restricted datasets (2024-04-017).
- 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.



From publication list to DOIs

1. Use Zotero to export publication list to BibTex format.
2. Use Bash script to extract DOIs to a text file.
3. Use a text editor to build queries for both Scopus and Web Of Science databases.

Scopus and Web of Science at INPE

- To run the queries, use a computer at INPE and your Café login.
- The database interface offers advanced options, where you can type complex queries.
- Store the resutla as either CSV or Plain Text for further processing.

Bibliometrix

- 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!

Bibliometrix field tags

- Some of the column names used by Bibliometrix.
- Find the complete list clicking here.

Field Tag	Description
AU	Authors' Names
TI	Document Title
SO	Journal Name (or Source)
JI	ISO Source Abbreviation
DT	Document Type
DE	Authors' Keywords
ID	Keywords associated by SCOPUS or WoS database
AB	Abstract
C1	Authors' Affiliations
RP	Corresponding Author's Affiliation
CR	Cited References
TC	Times Cited
PY	Publication Year
SC	Subject Category
UT	Unique Article Identifier
DB	Bibliographic Database

Source: Bibliometrix - Data Importing and Converting.

Results.

Overview

Description	Results
Timespan	2003:2023
Sources (Journals, Books, etc)	84
Documents	210
Annual Growth Rate %	15.55
Document Average Age	7.71
Average citations per doc	85.76
References	1
Author's Keywords (DE)	560
Authors	1329
Authors of single-authored docs	1
Co-Authors per Doc	15.6
International co-authorships %	86.67

Documents by type

Description	Results
article	175
letter	10
review	9
editorial material	6
article; proceedings paper	2
biographical-item	1
correction	1

Authors' productivity

Authors	Articles
ARAGAO L	140
ANDERSON L	92
MALHI Y	59
SHIMABUKURO Y	41
PHILLIPS O	31
SILVA C	31
ARAI E	21
BAKER T	20
BARLOW J	19
DALAGNOL R	18
MARIMON B	18
MATAVELI G	18
METCALFE D	18
QUESADA C	18

Authors	Articles Fractionalized
ARAGAO L	16.37
ANDERSON L	12.27
SHIMABUKURO Y	6.87
MALHI Y	4.92
MATAVELI G	2.97
ARAI E	2.79
SILVA C	2.70
DE O G	2.32
DALAGNOL R	1.96
WAGNER F	1.83

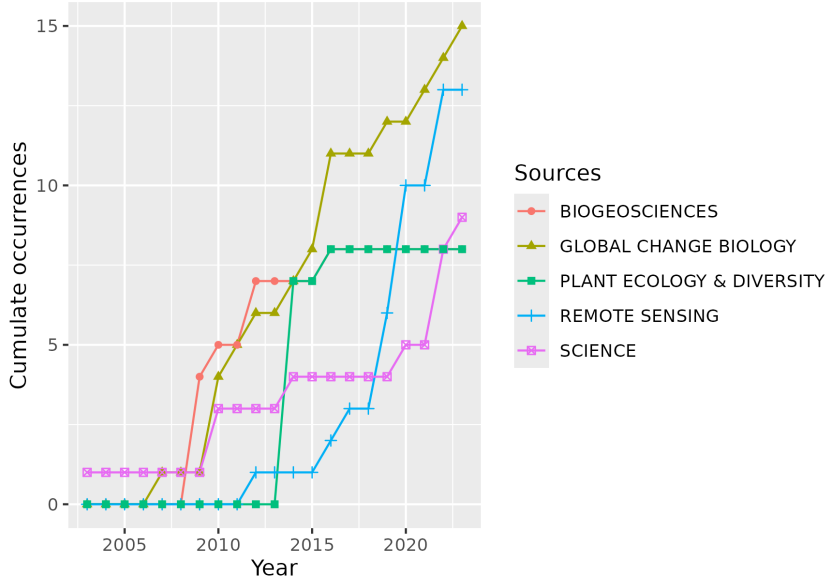
Most cited papers

Paper	TC	TCperYear	NTC
NEMANI R, 2003, SCIENCE	2683	122.0	1.00
LUYSSAERT S, 2007, GLOBAL CHANGE BIOL	766	42.6	3.03
BRIENEN R, 2015, NATURE	750	75.0	4.31
MORTON D, 2006, P NATL ACAD SCI USA	674	35.5	4.09
BARLOW J, 2016, NATURE	646	71.8	7.39
MALHI Y, 2009, P NATL ACAD SCI USA	596	37.2	4.02
ARAGAO L, 2018, NAT COMMUN	462	66.0	4.77
PHILLIPS O, 2010, NEW PHYTOL	433	28.9	3.00
ARAGAO L, 2007, GEOPHYS RES LETT	375	20.8	1.48
GATTI L, 2021, NATURE	369	92.2	8.90

Most relevant sources

Sources	Articles
GLOBAL CHANGE BIOLOGY	15
REMOTE SENSING	13
SCIENCE	9
BIOGEOSCIENCES	8
PLANT ECOLOGY & DIVERSITY	8
NEW PHYTOLOGIST	7
NATURE	6
INTERNATIONAL JOURNAL OF REMOTE SENSING	5
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGI...	5
PLOS ONE	5
SCIENTIFIC REPORTS	5

Production over time - sources

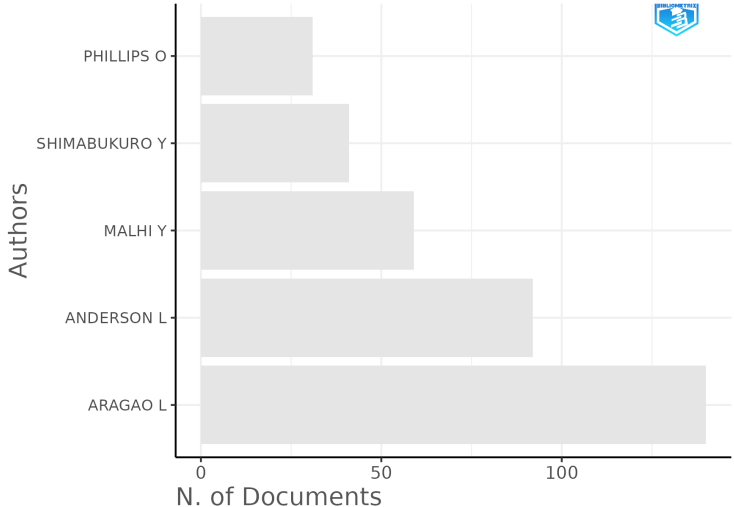


Most relevant keywords

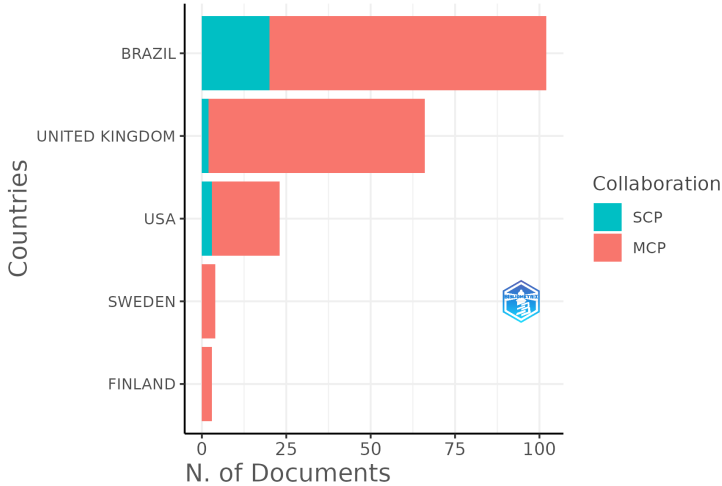
Author Keywords (DE)	Articles
AMAZON	29
REMOTE SENSING	21
DEFORESTATION	19
DROUGHT	17
FIRE	13
TROPICAL FOREST	13
MODIS	12
CLIMATE CHANGE	11
PHENOLOGY	11
TROPICAL FORESTS	11

Keywords-Plus (ID)	Articles
DEFORESTATION	49
CLIMATE-CHANGE	33
CARBON	25
RAIN-FOREST	25
TROPICAL FORESTS	25
BIOMASS	19
DYNAMICS	19
FOREST	19
PATTERNS	19
LAND-USE	18

Most Productive Authors

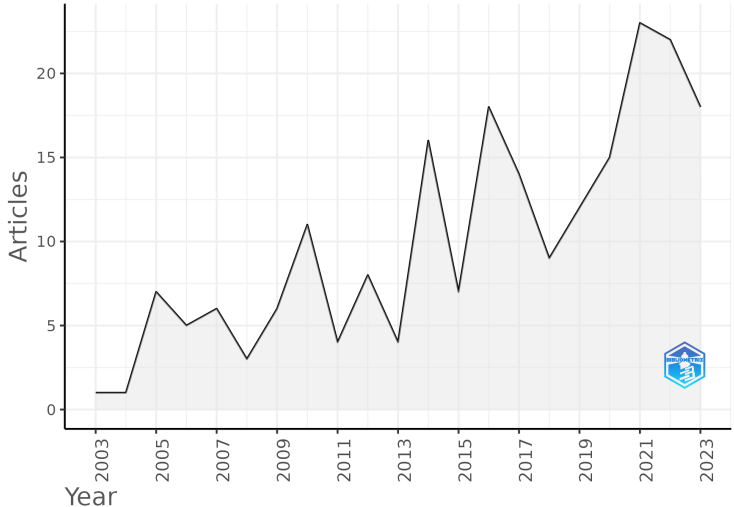


Most Productive Countries

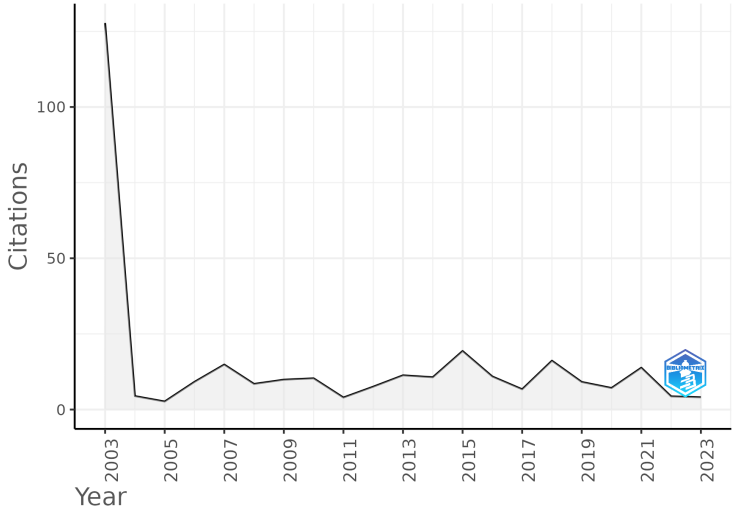


SCP: Single Country Publications, MCP: Multiple Country Publications

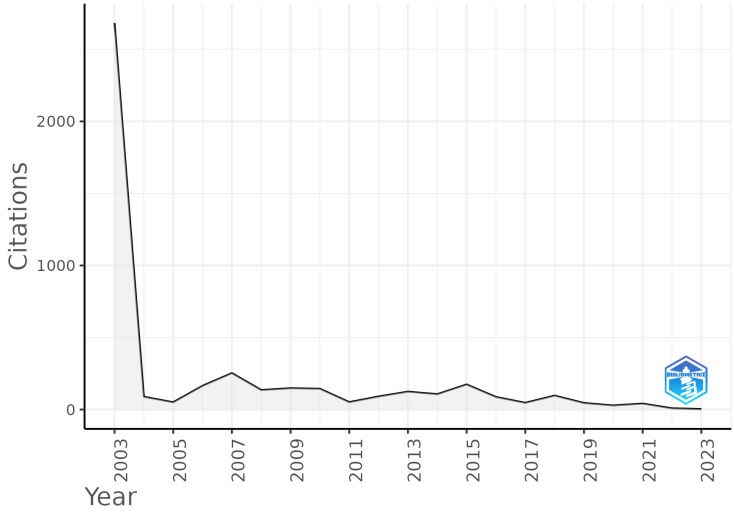
Annual Scientific Production



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.

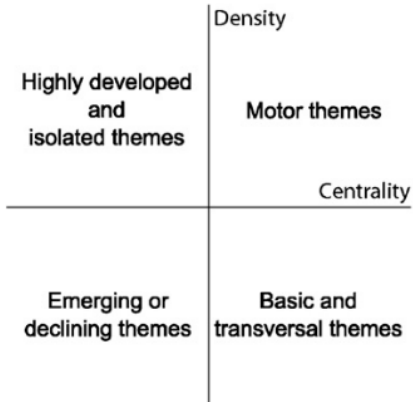
Conceptual structure

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.

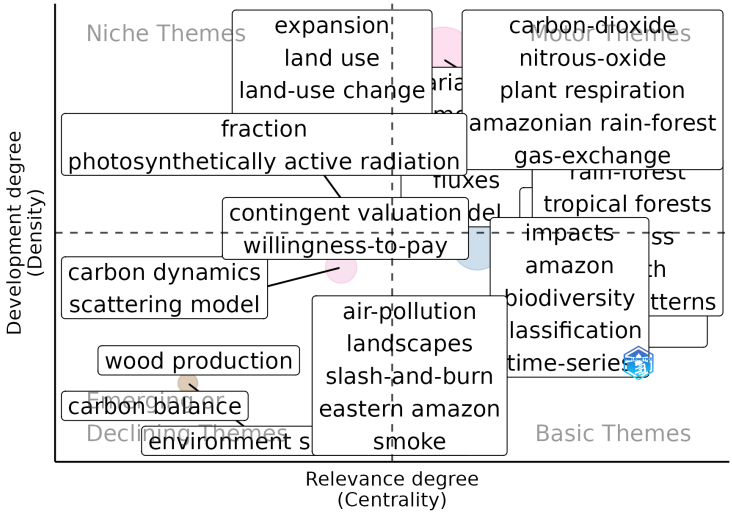
The strategic diagram

- 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.

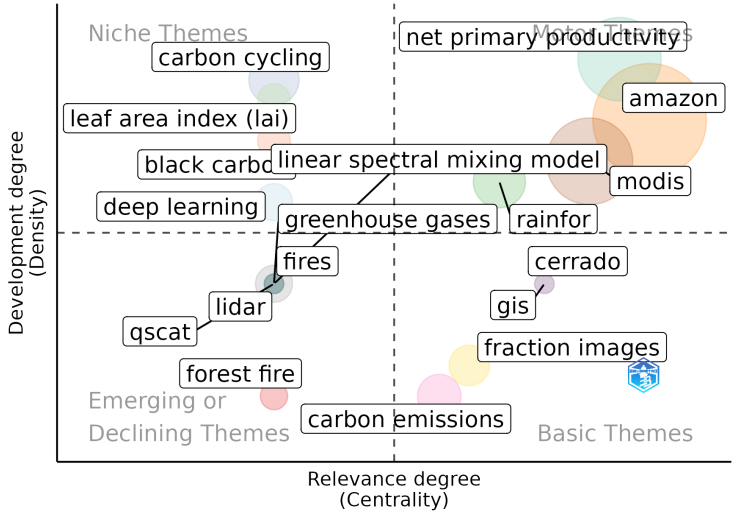


The strategic diagram. Source [4]

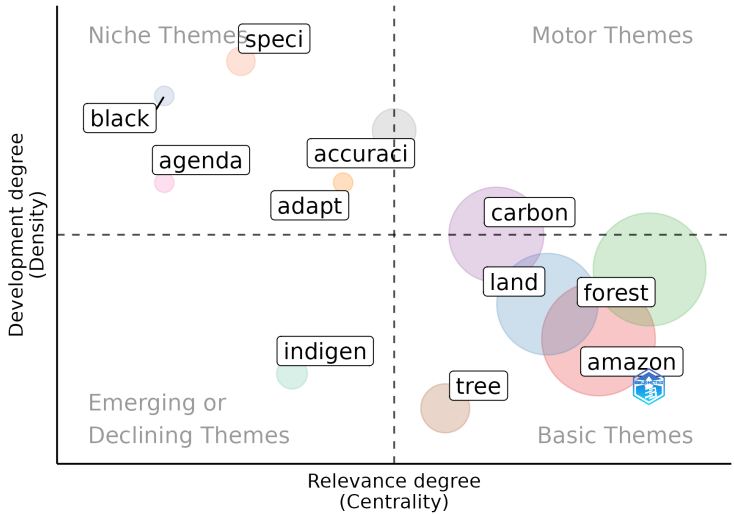
Thematic map (keyword plus)



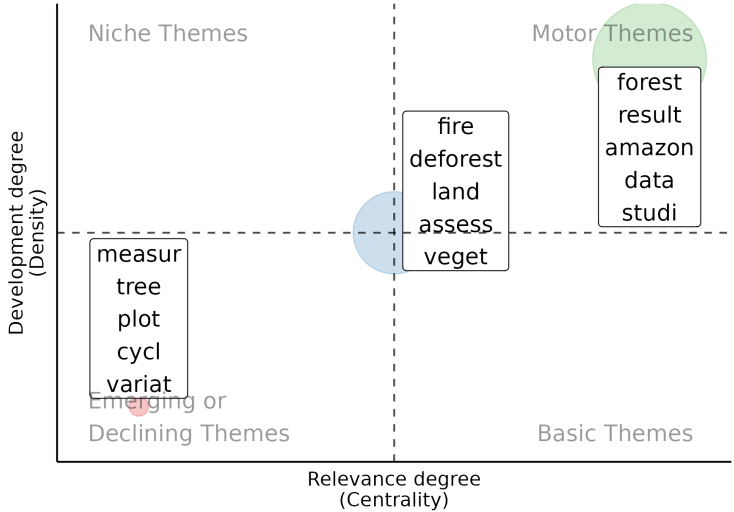
Thematic map (authors' keywords)



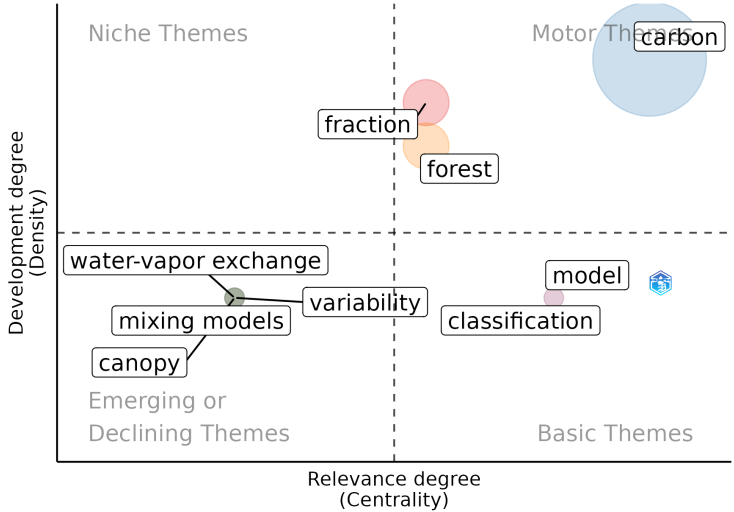
Thematic map (titles)



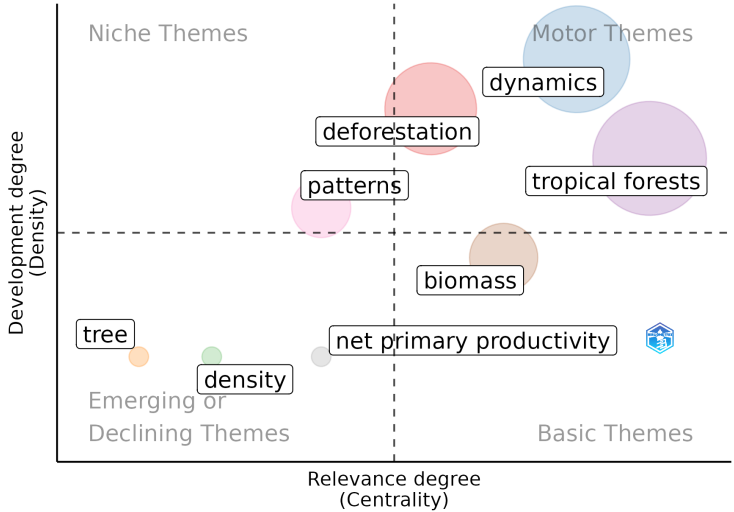
Thematic map (abstracts)



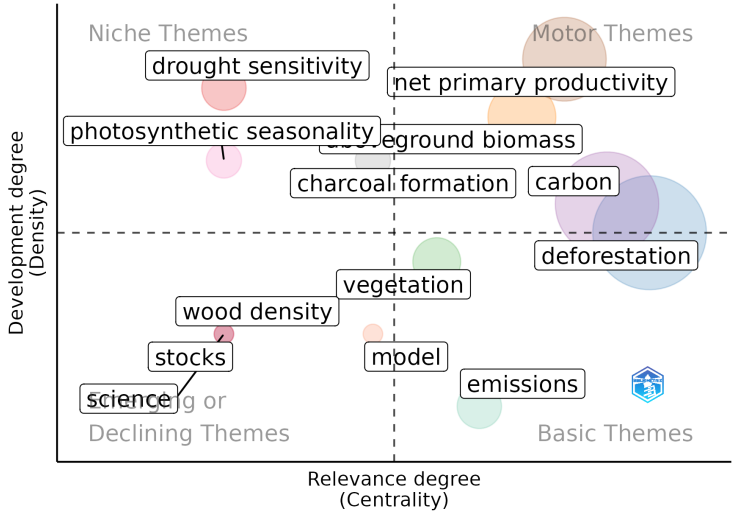
Thematic evolution - 2003:2007



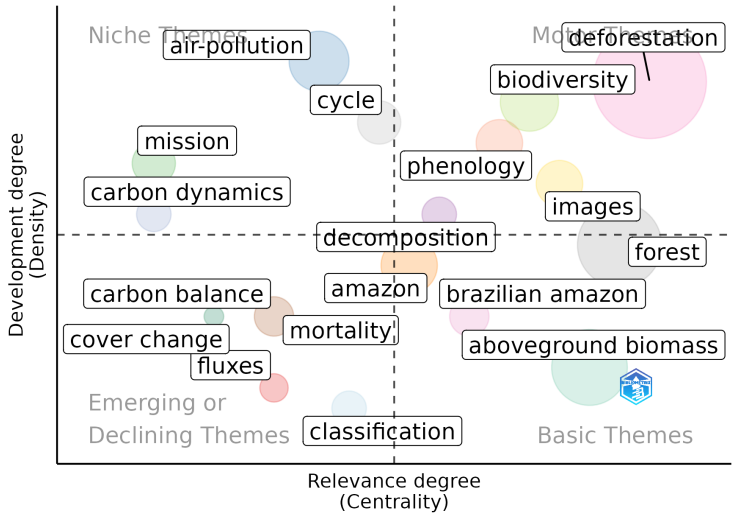
Thematic evolution - 2008:2012



Thematic evolution - 2013:2017



Thematic evolution - 2018:2022



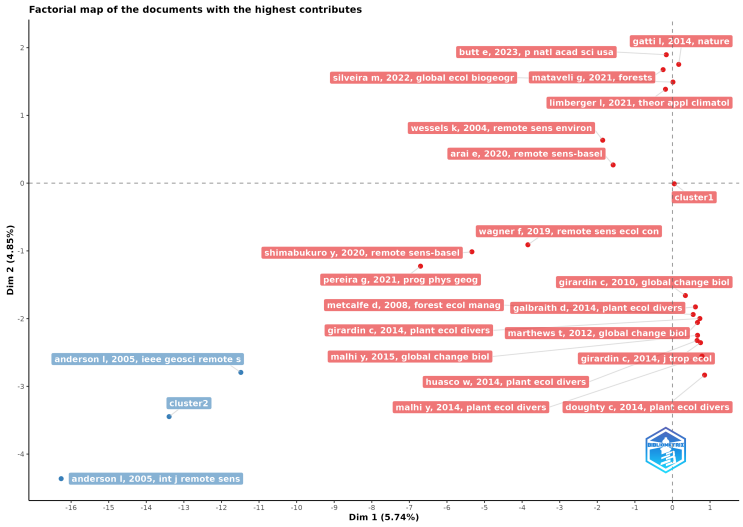
Factorial Map - Most contributing documents

- 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.

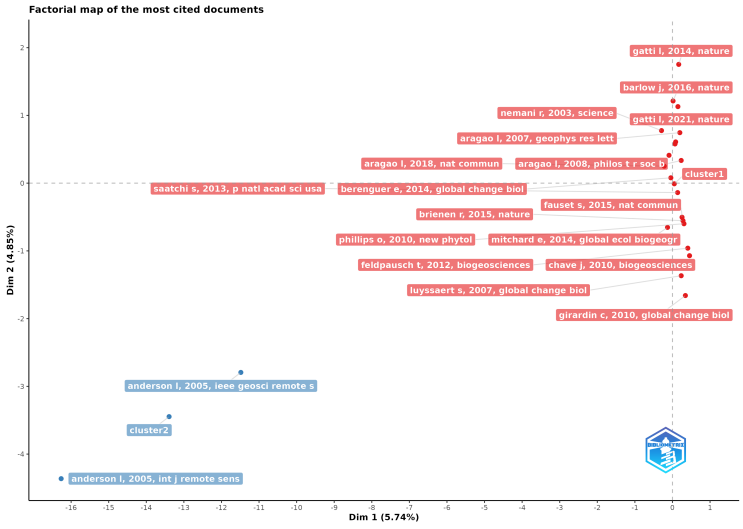
Factorial Map - Most contributing documents

- Identify the link between topic and most cited documents.
- Plot documents associated to the highest global citations.
- The colors represent the clusters.

Factorial Map - Most contributing documents



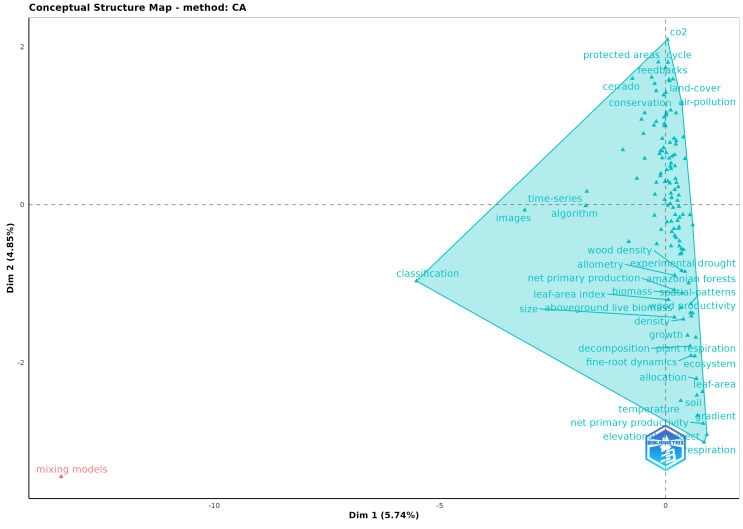
Factorial Map - Most cited documents



Map of words

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

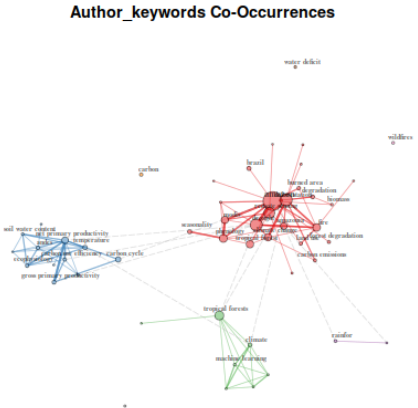
Map of words


Conceptual Structure Map - method: CA
co2
protected areas
cycle
feedbacks
cerrado
land-cover
conservation
air-pollution
time-series
images
algorithm
wood density
allometry
experimental drought
classification
net primary production
amazonian forests
biomass
spatial patterns
leaf-area index
wood
productivity
size
aboveground live biomass
density
growth
decomposition
plant respiration
fine-root dynamics
ecosystem
allocation
leaf-area
soil
temperature
gradient
net primary productivity
elevation
respiration
mixing models
Dim 1 (5.74%)
Dim 2 (4.85%)
2
0
-2
-10
-5
0

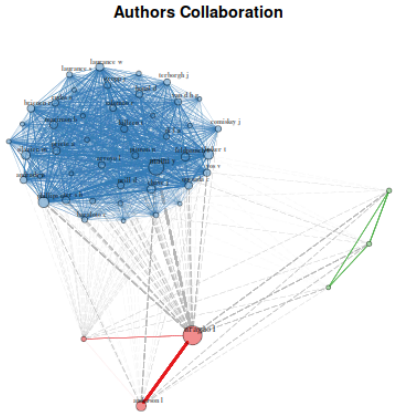
Network analysis

- 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.

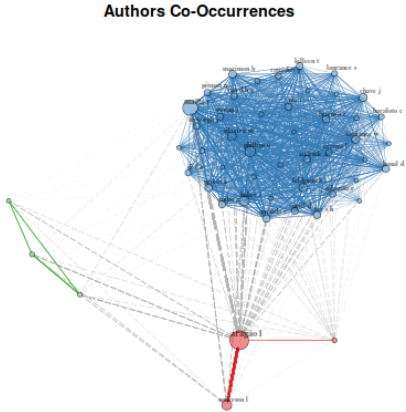
Network - Co-occurrences Authors-Keywords



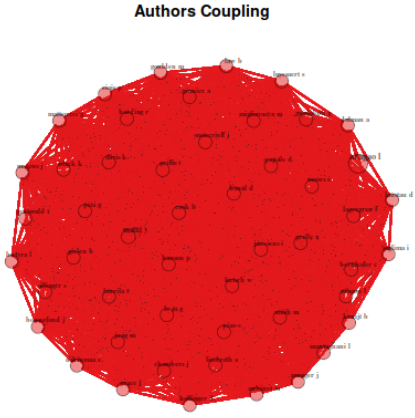
Network - Authors collaboration



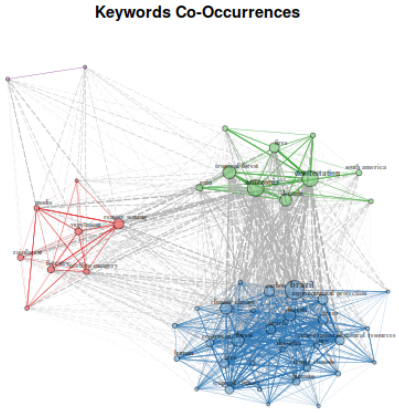
Network - Authors co-occurrences



Network - Authors coupling

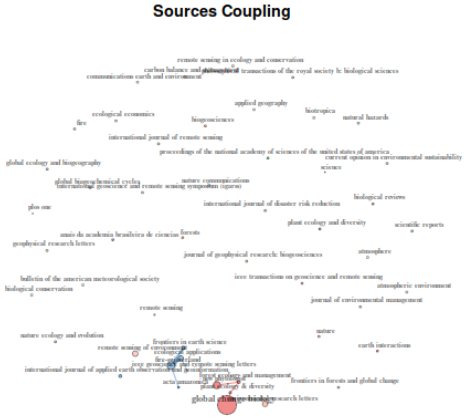


Network - Keyword co-occurrences

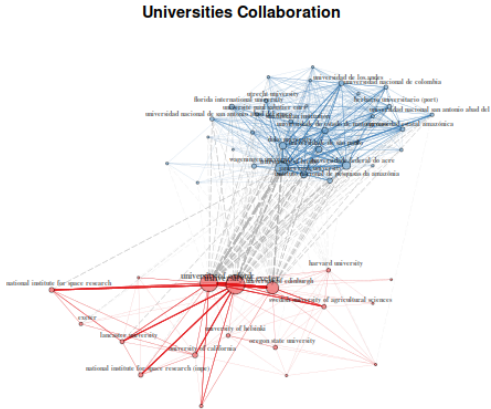


Network - References co-citation

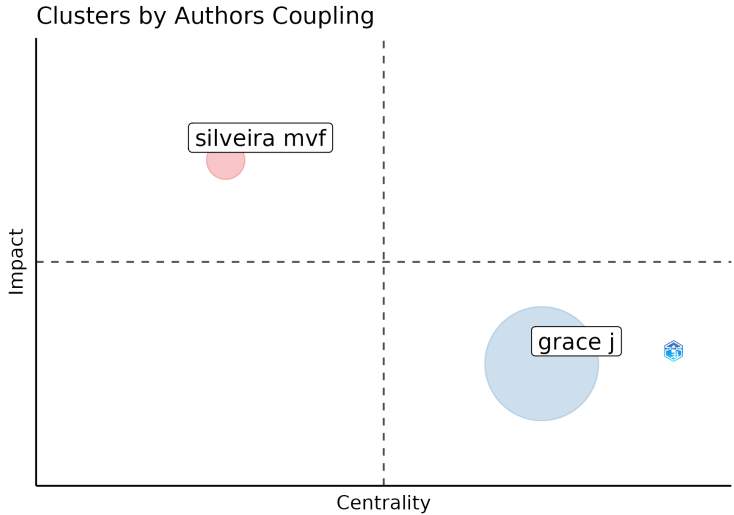
Network - Sources coupling



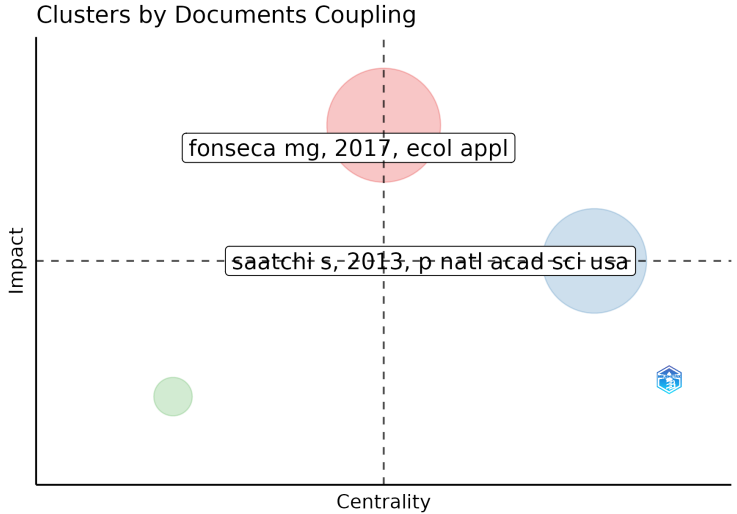
Network - Universities collaboration



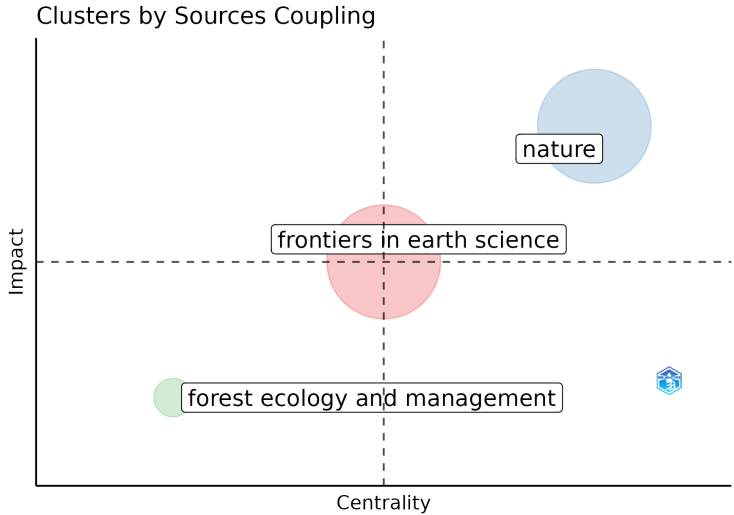
Coupling network analysis (authors)



Coupling network analysis (documents)



Coupling network analysis (sources)



Take home message

- It's healthy for TreesLab to look at itself once in a while.
- An introspection excercise reveals trends and patterns along 20 years of research, spotting ways to move forward.
- The tools and services presented here, can be chained for regular reporting and monitoring.
- Besides, these tools allow exploring new or converging knowledge areas, easing the TreesLab path into potential opportunities in interdisciplinar research.

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).

[6] Zhengyang Lin et al. "Large Language Models Reveal Big Disparities in Current Wildfire Research". In: *Communications Earth & Environment* 5.1 (Apr. 2024), p. 168. ISSN: 2662-4435. DOI: 10.1038/s43247-024-01341-7. (Visited on 04/10/2024).