

# 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

November 29, 2024

# Overview I

## Introduction

The research cycle

Bibliometrics

## Method

Definitions

Assumptions

Get TREES Lab publications' DOIs

Query Scopus and Web of Science

Run analysis using *R*

## Results

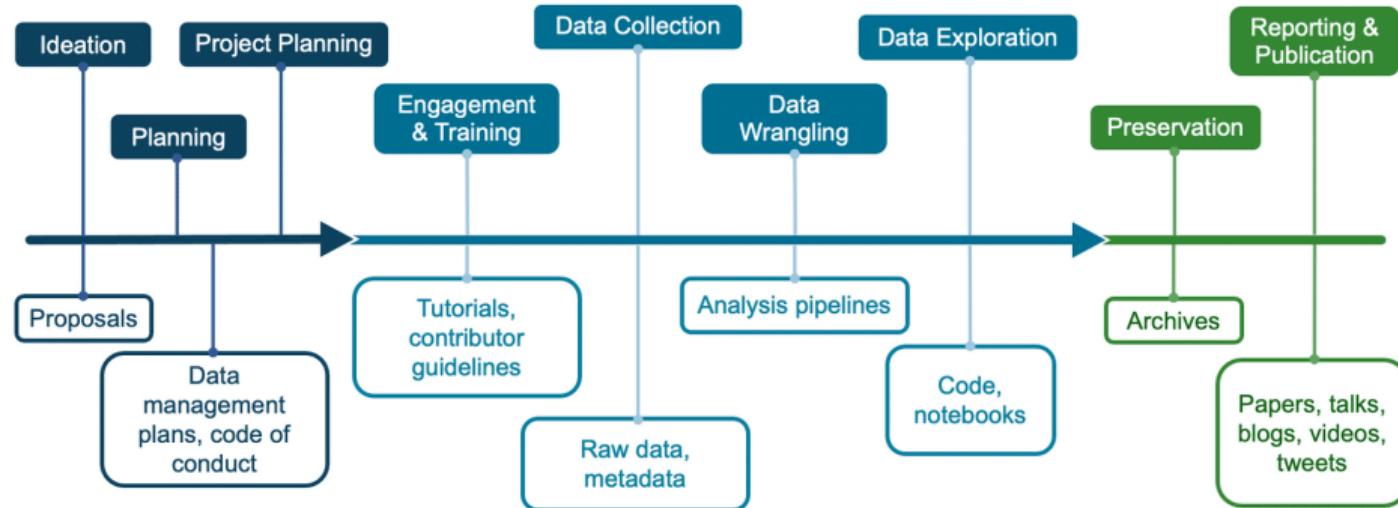
TREES Lab publication list (complete)

TREES Lab publications with DOI

## Summary

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 [**broadus1987**].
- ▶ The bibliometric methodology encapsulates the application of quantitative techniques (i.e., bibliometric analysis — e.g., citation analysis) on bibliometric data (e.g., units of publication and citation) [**donthu2021**].

# Bibliographic databases

- ▶ Scopus.
- ▶ Web of science.

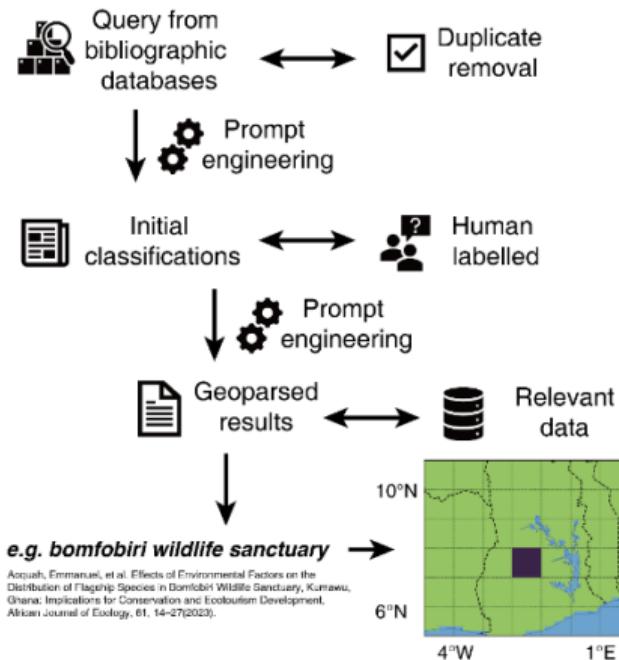


**ELSEVIER**  
Scopus

 Clarivate  
**Web of Science™**

# Bibliometrics and LLMs

- ▶ Large language models reveal big disparities in current wildfire research [lin2024].
- ▶ This is a potential new trend in bibliometric analysis.
- ▶ Feed database data into LLM (ChatGPT).
- ▶ Extract, besides bibliometric information, localization of AOI.



Source: [lin2024].

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*" [aria2017].
- ▶ *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.



# 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.



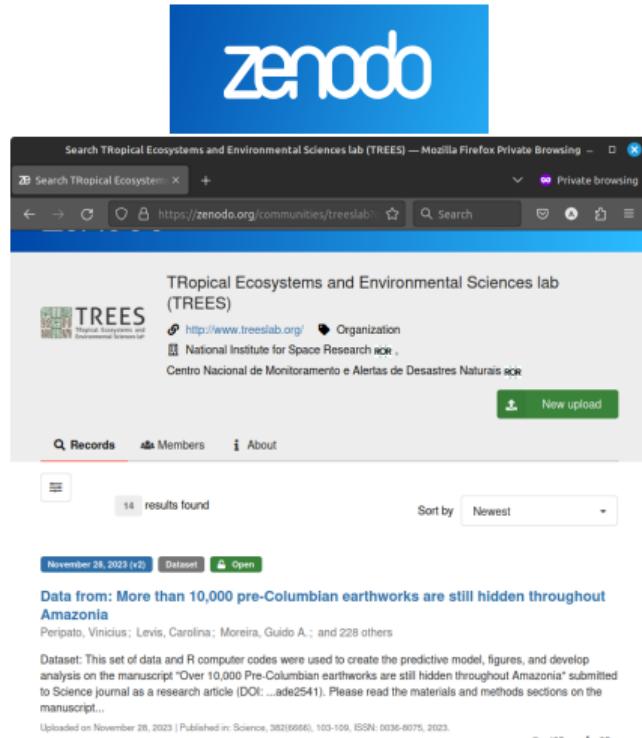
# TreesLab publication list

- ▶ 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 without 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.

Title	Creator	Date
21st Century drought-related fires counteract the decline in tree mortality risk in the western United States	Aneglio et al.	2018-02-19
A globally deployable strategy for co-development of a... A large-scale field assessment of carbon stocks in humic... A method for extracting plant roots from soil which facil... A MODIS-Based Energy Balance to Estimate Evapotra...	Marengi et al. Berenguer et al. Metcalfe et al. Ruhoff et al.	2017 2014 2007 2012-03-12
...		

# 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 results as either CSV or Plain Text for further processing.



Você está acessando esse portal por: INPE  
Acesso CAFÉ ▾

Home > Acervo > Lista de bases e coleções

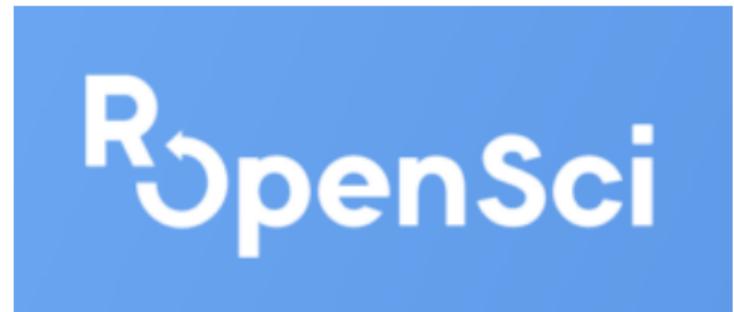
**Lista de bases e coleções**

Órgãos do

Sobre Acervo

## bib2df

- ▶ *bib2df*: Parse a BibTex file to a tibble [**ottolinger2024**].
- ▶ Most R analysis is done using data frames or tibbles.
- ▶ BibTex is *de facto* standard for managing bibliographic references.
- ▶ *bib2df* is part of rOpenSci.
- ▶ rOpenSci is a initiative that foster *R* software tools that lower barriers to working with local and remote scientific datasets.



# Bibliometrix

- ▶ R package for bibliometric analysis [**aria2017**].
- ▶ 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.

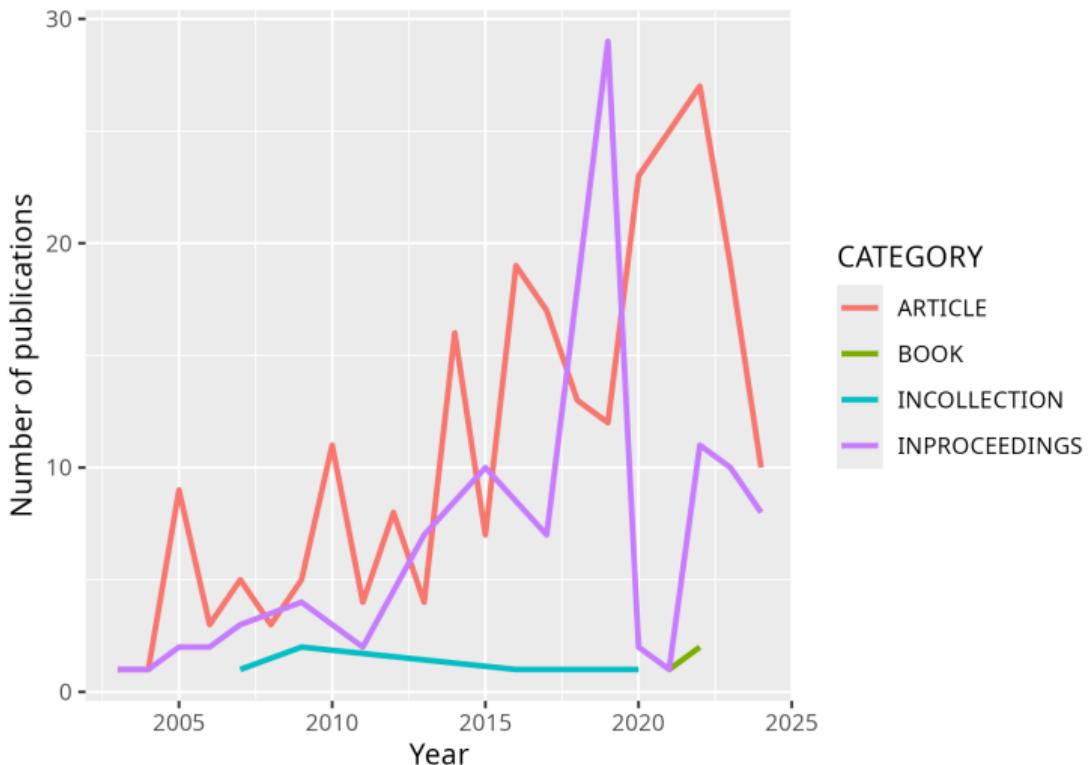
Results.

TREES Lab publication list (complete).

## Documents by type (complete)

CATEGORY	n
ARTICLE	242
INPROCEEDINGS	100
TECHREPORT	28
INCOLLECTION	5
BOOK	3
MISC	1

# Annual Scientific Production (complete)



TREES Lab publications with DOI.

## Overview

Description	Results
Timespan	2003:2024
Sources (Journals, Books, etc)	86
Documents	221
Annual Growth Rate %	12.1
Document Average Age	7.33
Average citations per doc	84.86
References	1
Author's Keywords (DE)	571
Authors	1411
Authors of single-authored docs	2
Co-Authors per Doc	15.6
International co-authorships %	85.97

## Documents by type

Description	Results
article	182
letter	13
review	10
editorial material	6
article; proceedings paper	2
biographical-item	1
correction	1

## Authors' productivity

Authors	Articles
ARAGAO L	144
ANDERSON L	98
MALHI Y	59
SHIMABUKURO Y	42
PHILLIPS O	31
MATAVELI G	24
SILVA C	24
ARAI E	21
BAKER T	20
DALAGNOL R	20

Authors	Articles	Fractionalized
ARAGAO L	144	16.76
ANDERSON L	98	12.73
SHIMABUKURO Y	42	6.98
MALHI Y	31	4.92
MATAVELI G	24	3.59
ARAI E	21	2.79
SILVA C	20	2.54
DALAGNOL R	20	2.08
WAGNER F	20	2.04
		1.83

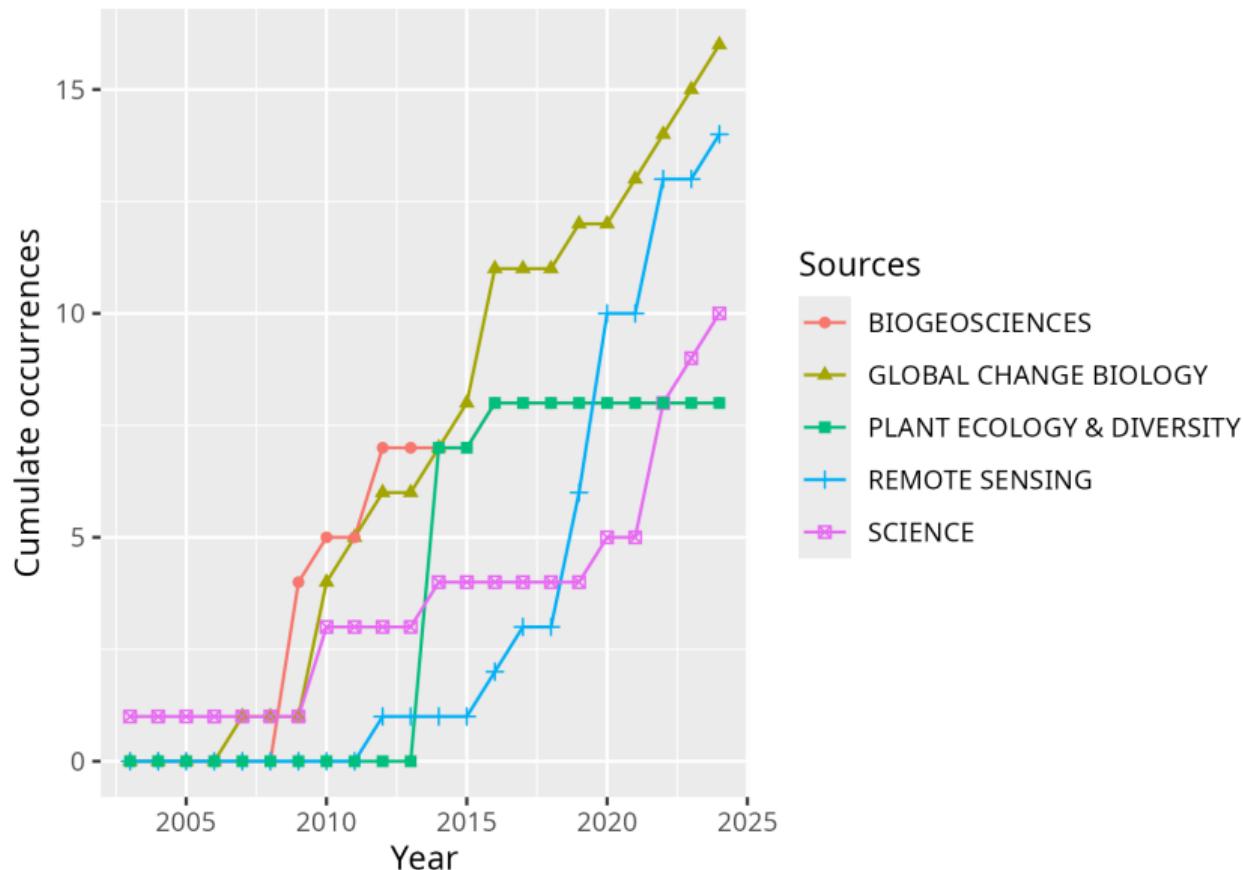
## Most cited papers

Paper	TC	TCperYear	NTC
NEMANI R, 2003, SCIENCE	2741	124.6	1.00
BRIENEN R, 2015, NATURE	783	78.3	4.31
LUYSSAERT S, 2007, GLOBAL CHANGE BIOL	776	43.1	3.06
MORTON D, 2006, P NATL ACAD SCI USA	681	35.8	4.09
BARLOW J, 2016, NATURE	673	74.8	7.45
MALHI Y, 2009, P NATL ACAD SCI USA	604	37.8	4.05
ARAGAO L, 2018, NAT COMMUN	496	70.9	4.80
PHILLIPS O, 2010, NEW PHYTOL	446	29.7	3.01
GATTI L, 2021, NATURE	435	108.8	9.06
GATTI L, 2014, NATURE	373	33.9	3.43

## Most relevant sources

Sources	Articles
GLOBAL CHANGE BIOLOGY	16
REMOTE SENSING	14
SCIENCE	10
BIOGEOSCIENCES	8
PLANT ECOLOGY & DIVERSITY	8
NATURE	7
NEW PHYTOLOGIST	7
ENVIRONMENTAL RESEARCH LETTERS	5
FIRE-SWITZERLAND	5
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

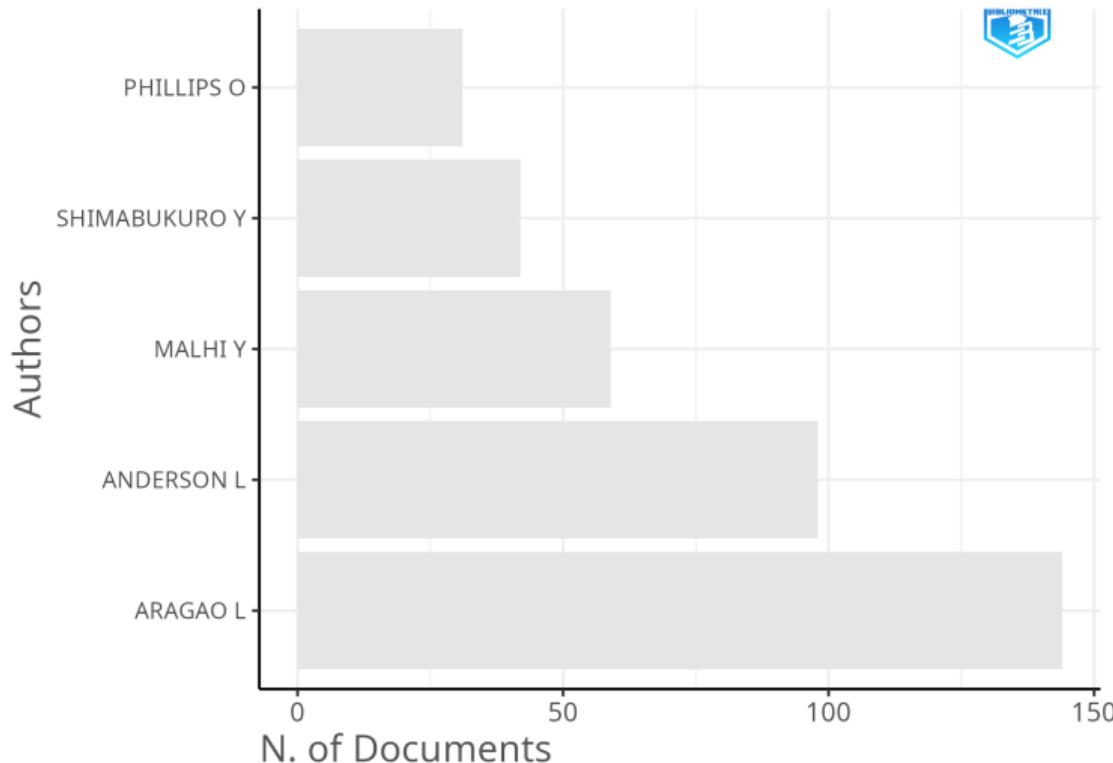


# Most relevant keywords

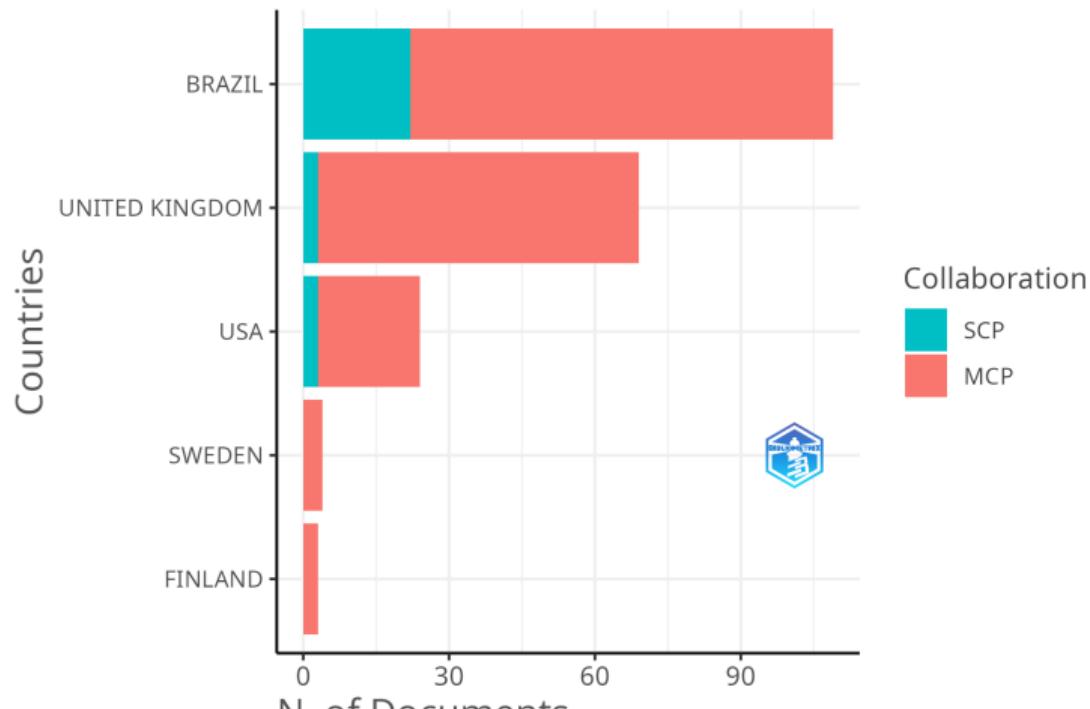
Author Keywords (DE)	Articles
AMAZON	31
DEFORESTATION	21
REMOTE SENSING	21
DROUGHT	17
FIRE	13
TROPICAL FOREST	13
MODIS	12
CLIMATE CHANGE	11
PHENOLOGY	11
TROPICAL FORESTS	11

Keywords-Plus (ID)	Articles
DEFORESTATION	52
CLIMATE-CHANGE	36
CARBON	26
RAIN-FOREST	26
TROPICAL FORESTS	25
FOREST	23
BIOMASS	20
LAND-USE	20
PATTERNS	20
DYNAMICS	19

# Most Productive Authors

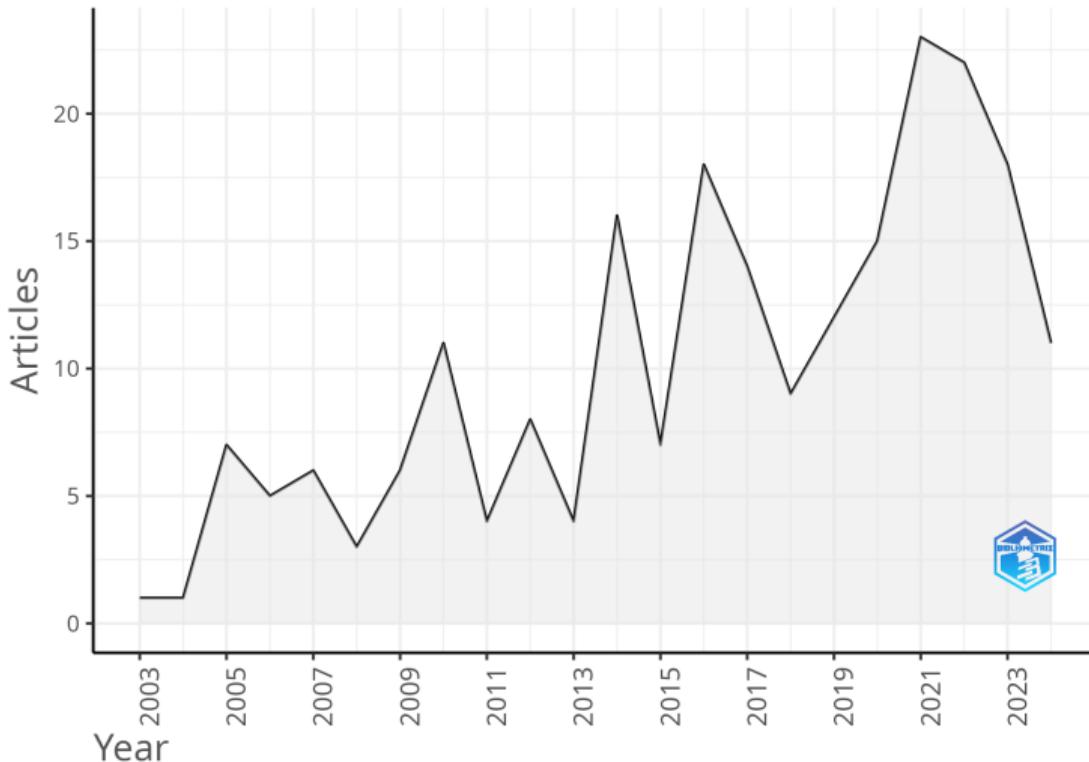


# Most Productive Countries

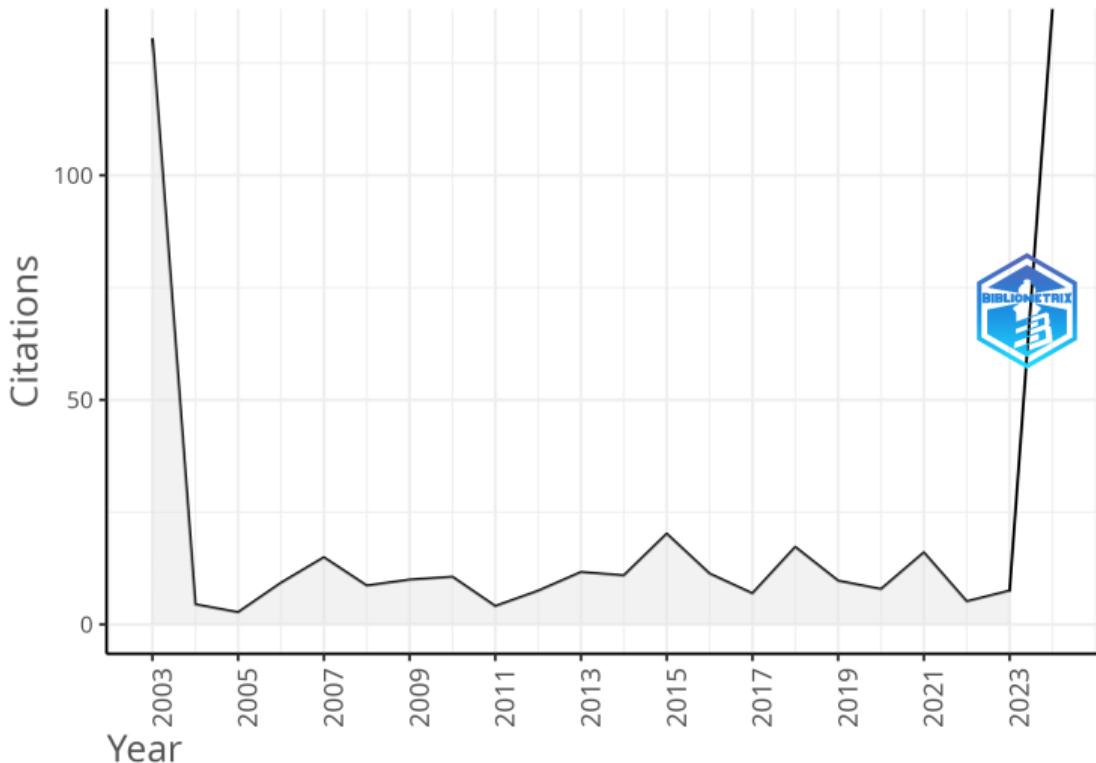


SCP: Single Country Publications, MCP: Multiple Country Publications

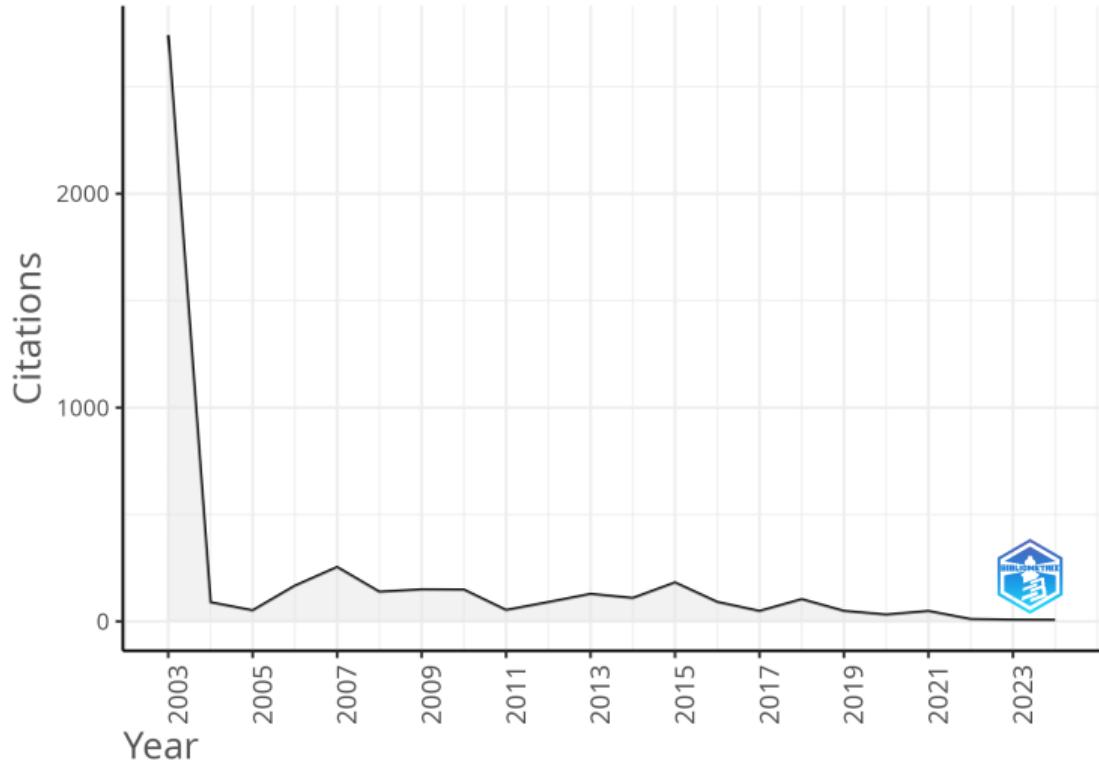
# Annual Scientific Production



# 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 [**borner2003**].
- ▶ *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.

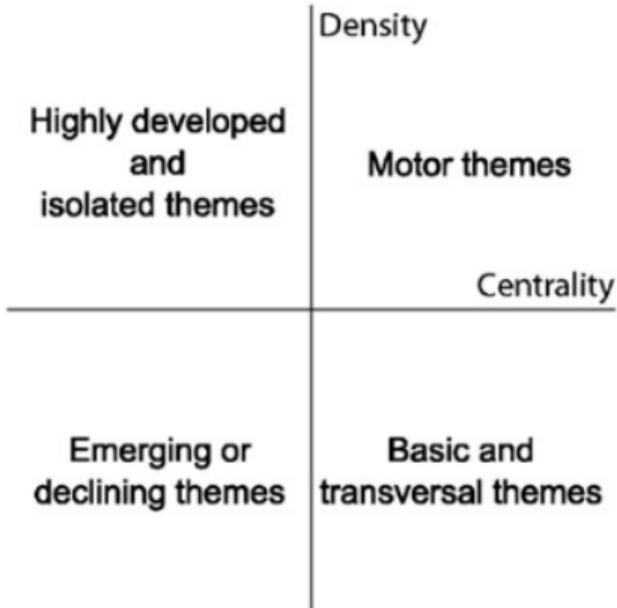
## Conceptual structure

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

- ▶ Words which 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.

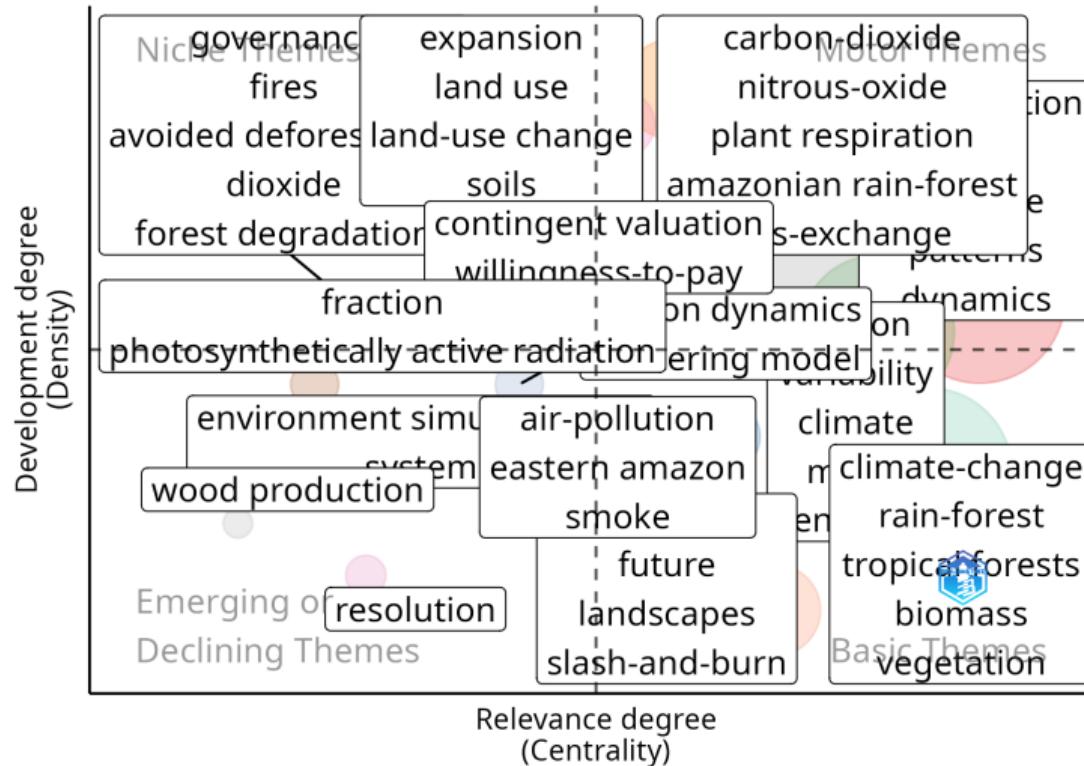
## The strategic diagram

- ▶ 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.

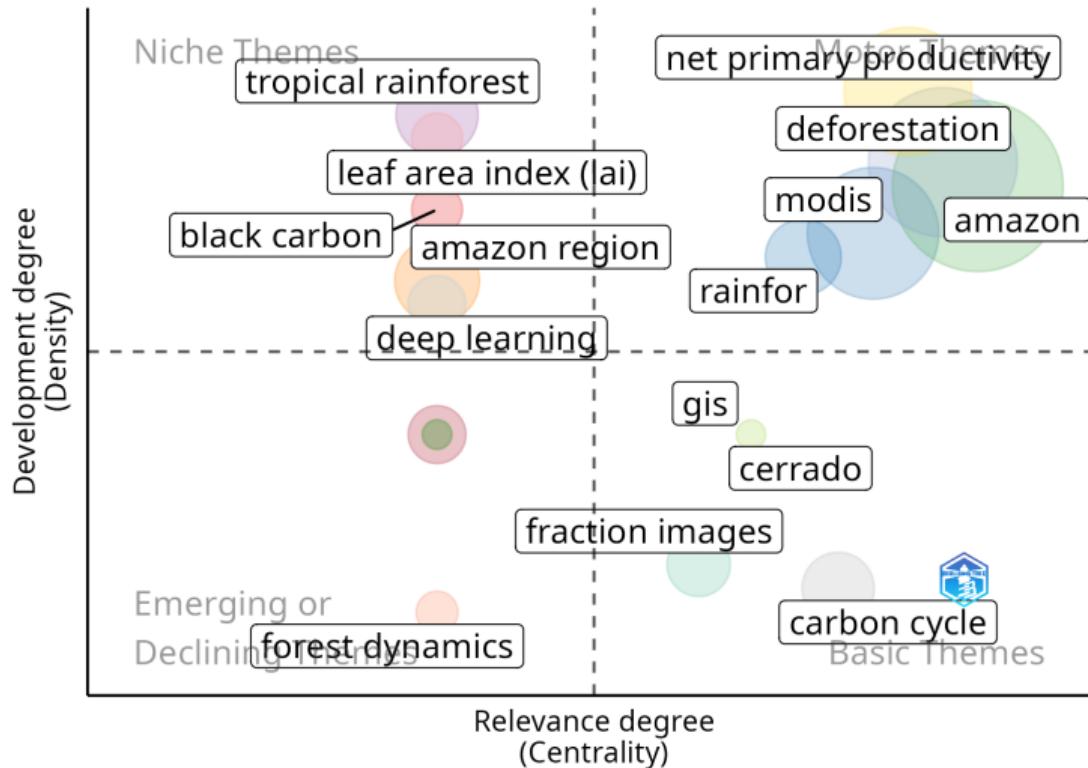


The strategic diagram. Source [cobo2011]

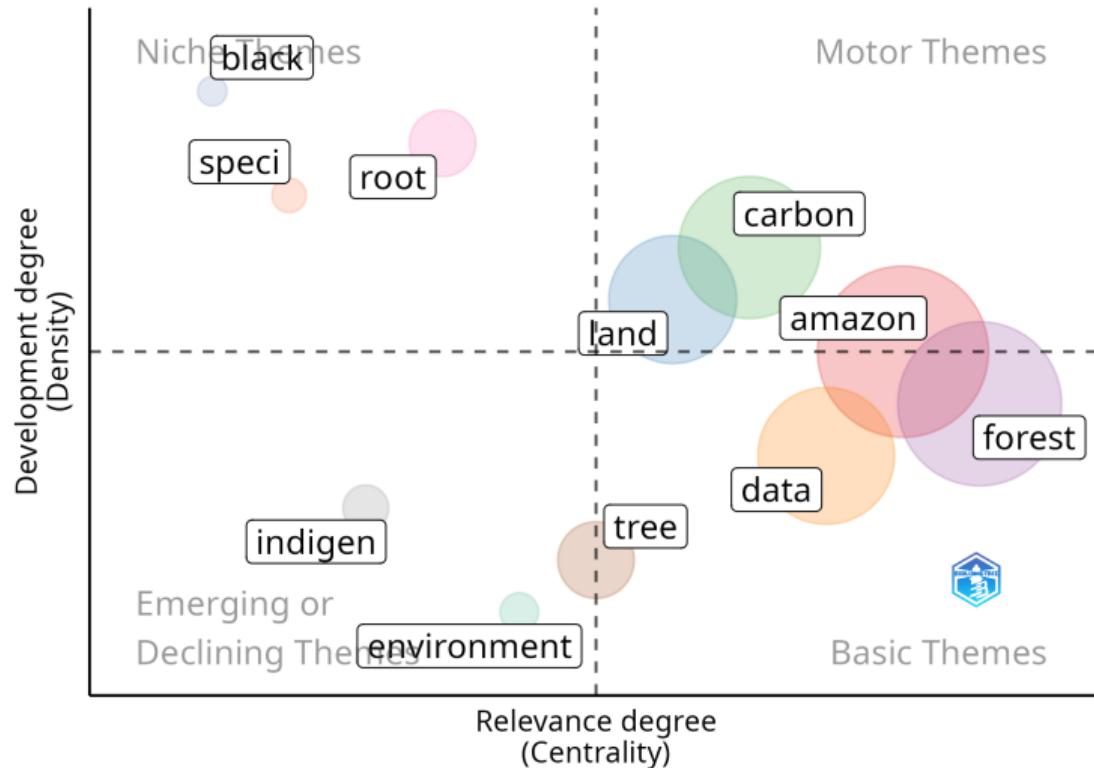
# Thematic map (keyword plus)



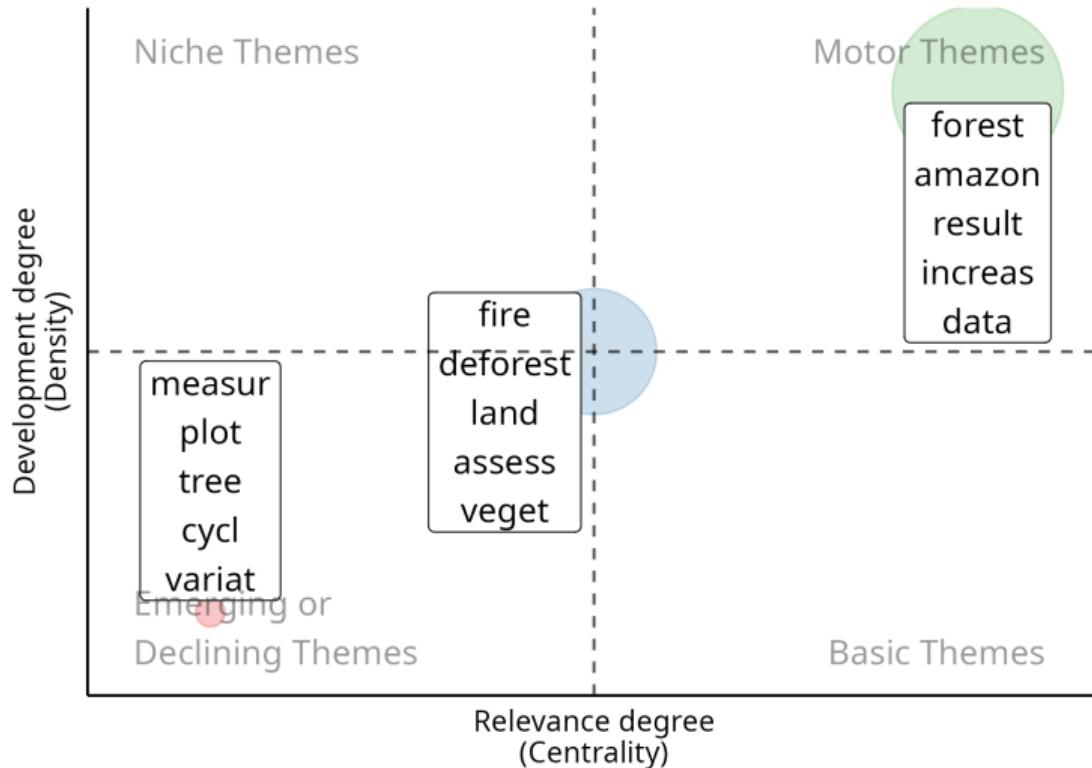
# Thematic map (authors' keywords)



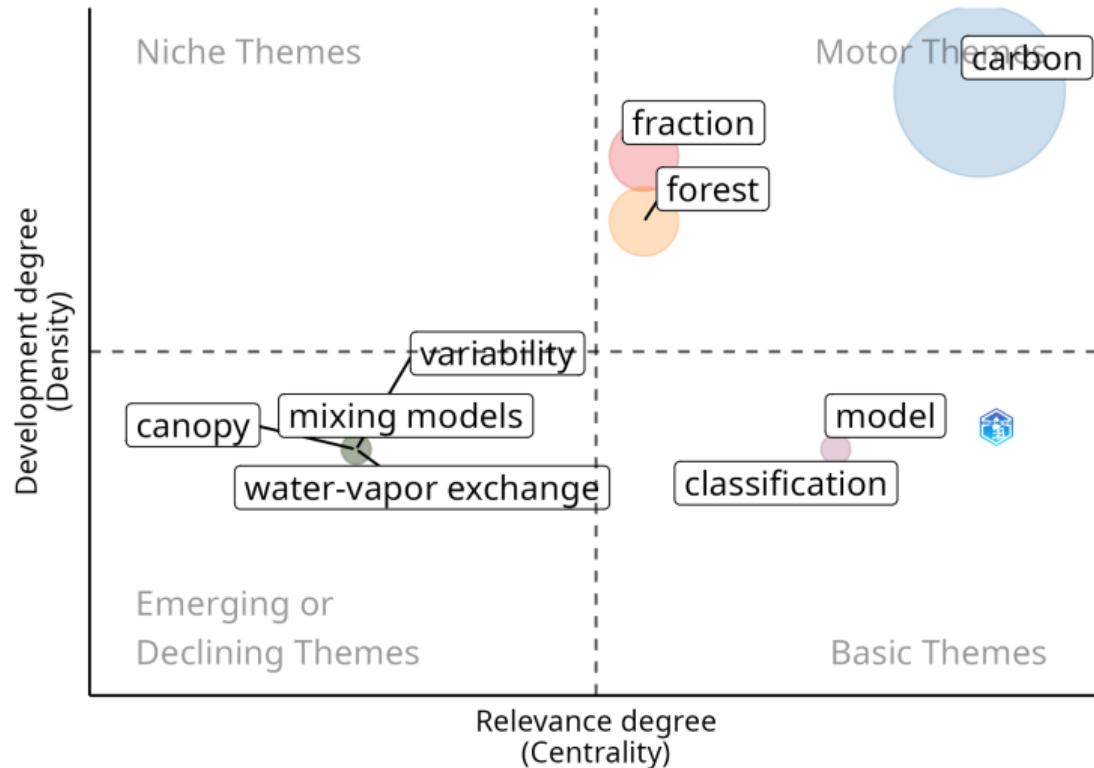
# Thematic map (titles)



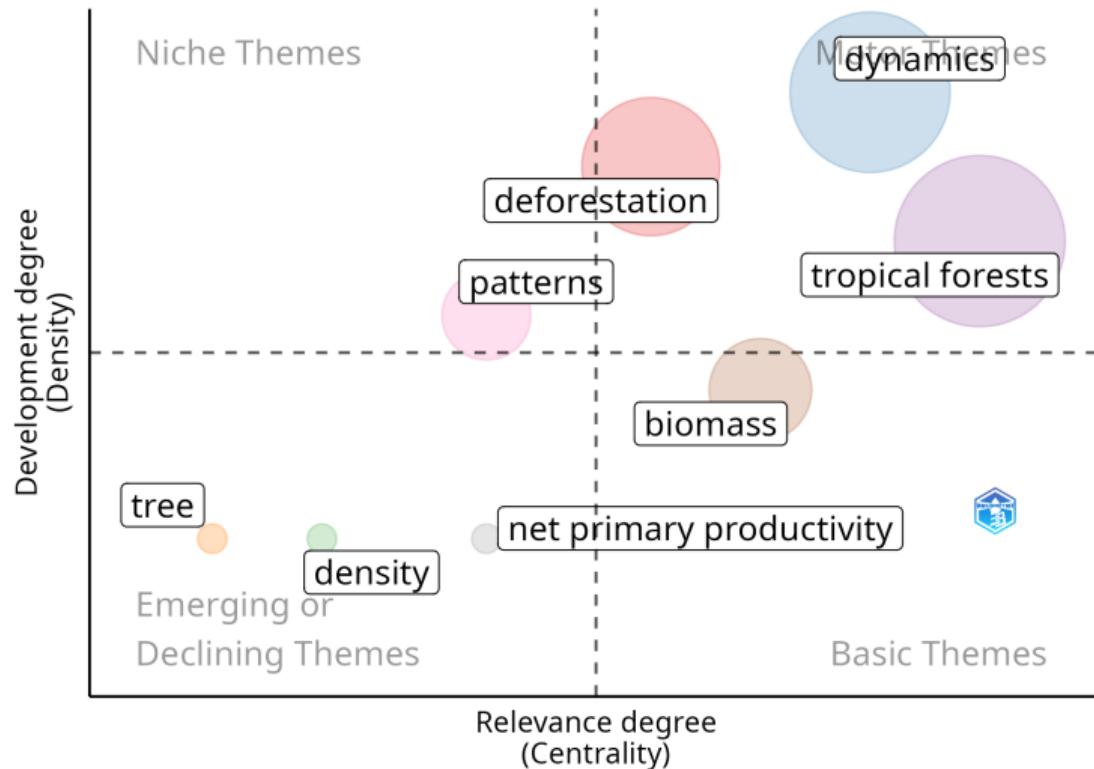
## Thematic map (abstracts)



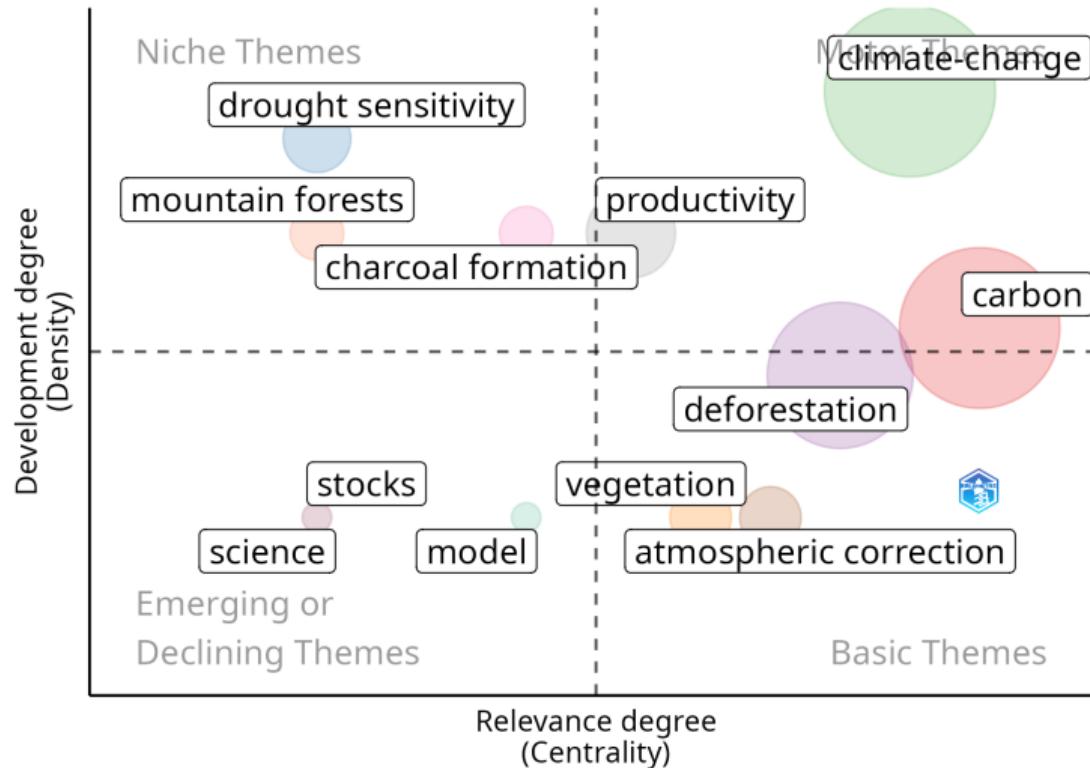
# Thematic evolution - 2003:2007



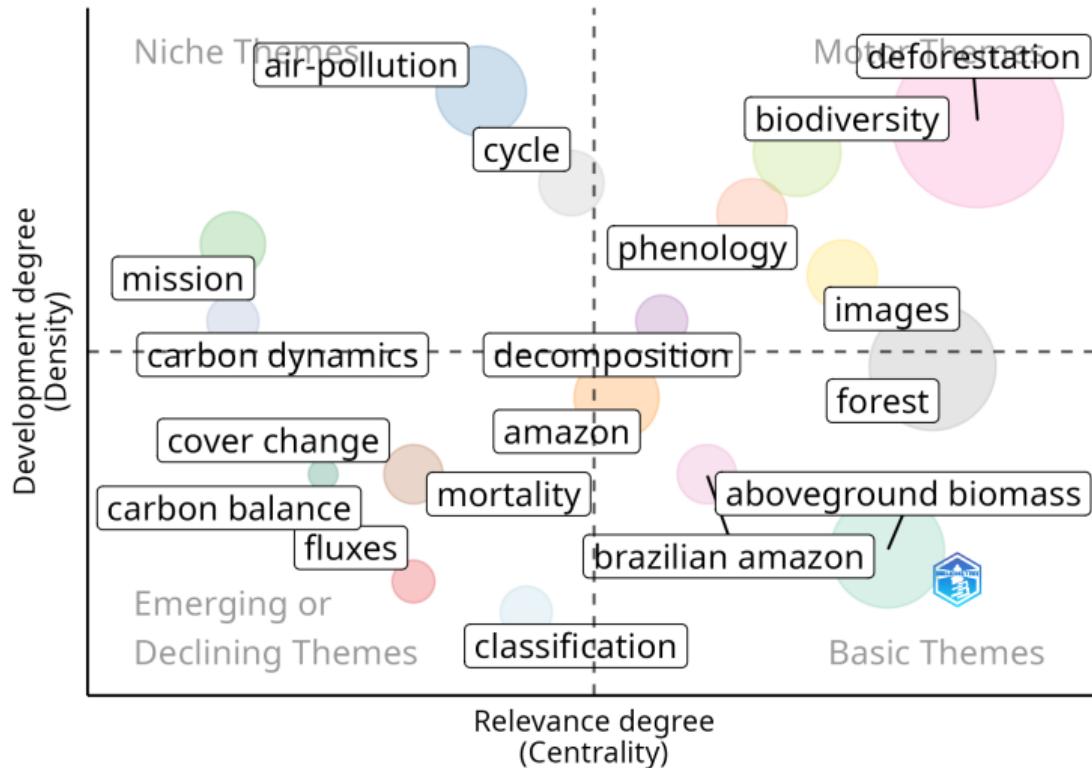
## Thematic evolution - 2008:2012



# Thematic evolution - 2013:2017



# Thematic evolution - 2018:2022



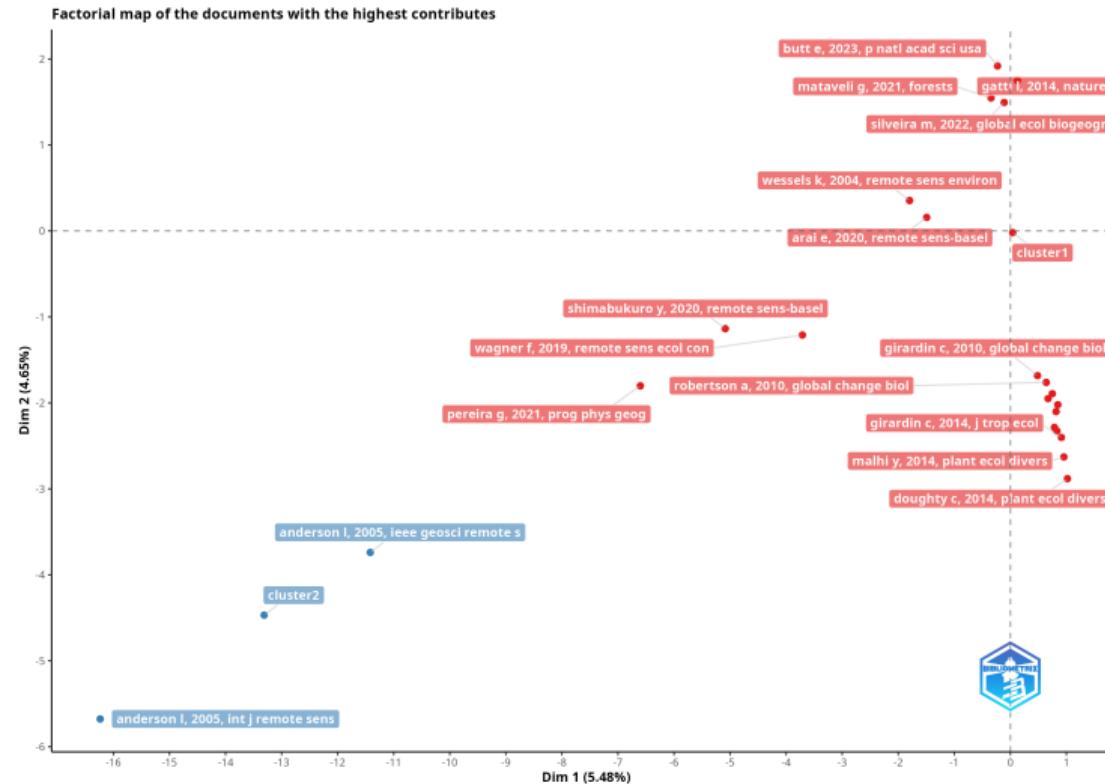
## Factorial Map - Most contributing documents

- ▶ 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.

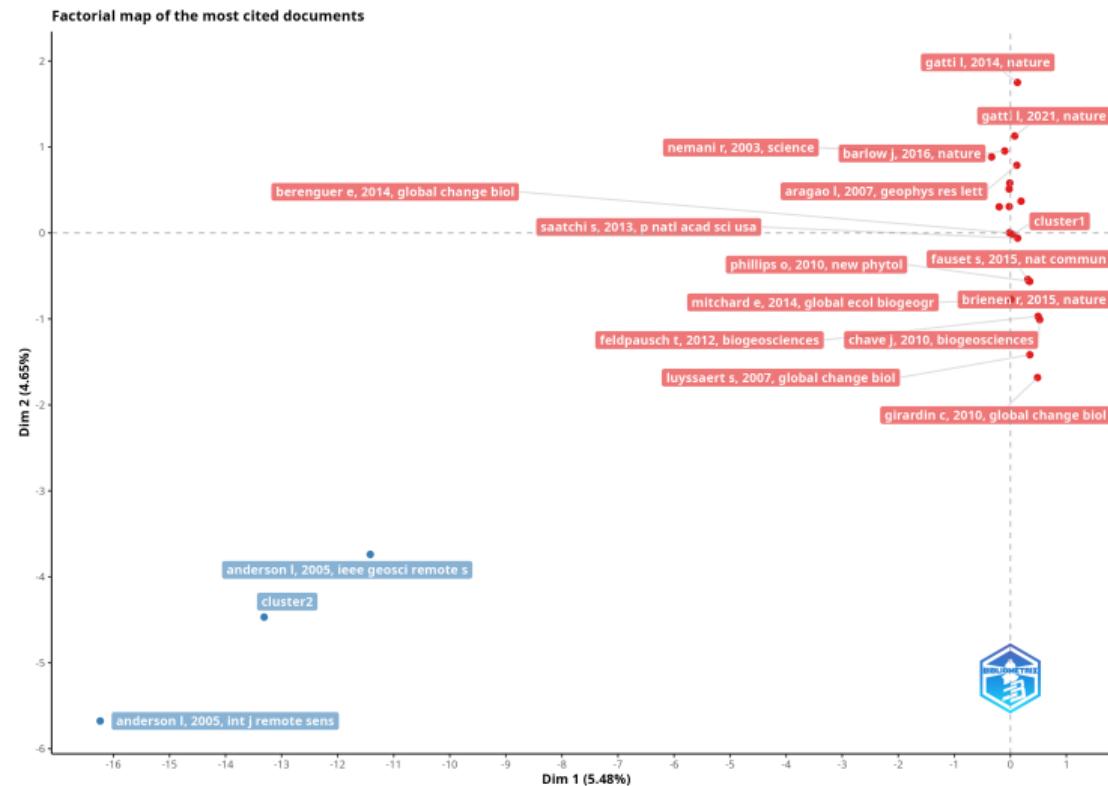
## Factorial Map - Most contributing documents

- ▶ 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



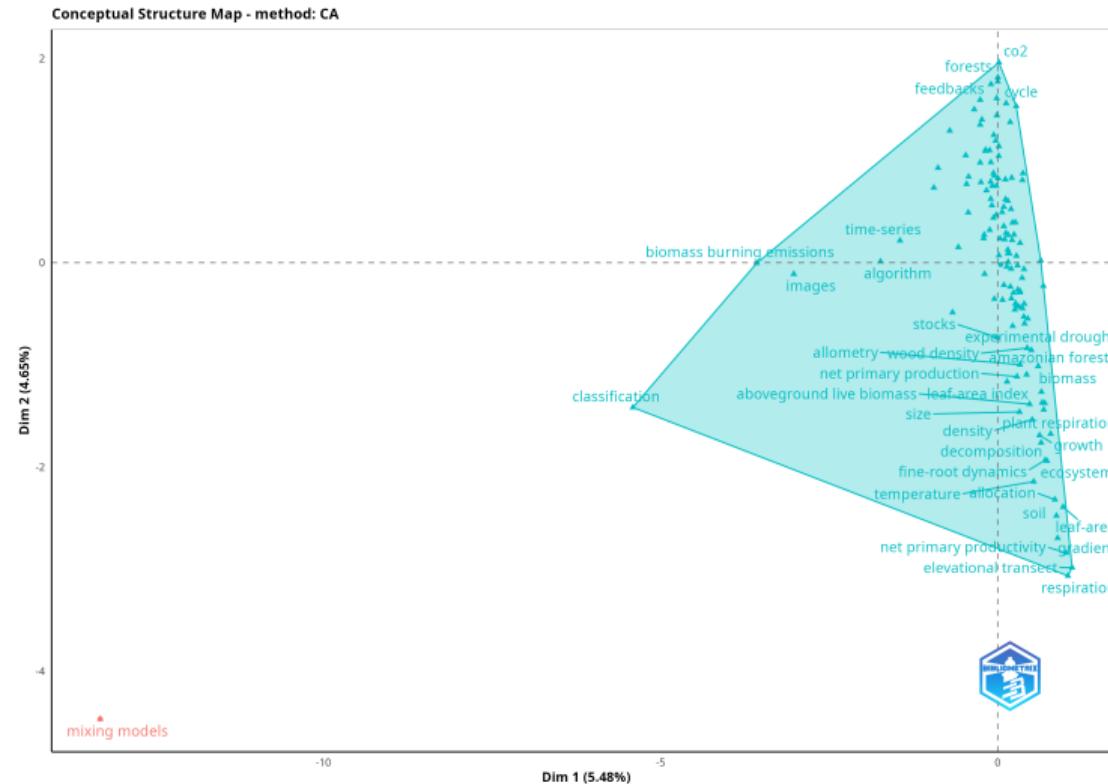
## Factorial Map - Most cited documents



## Map of words

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

# Map of words



## 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.

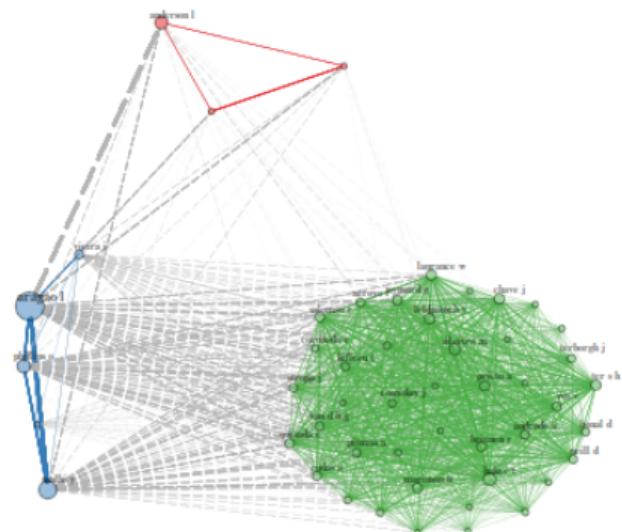
## Network - Co-occurrences Authors-Keywords

### **Author\_keywords Co-Occurrences**

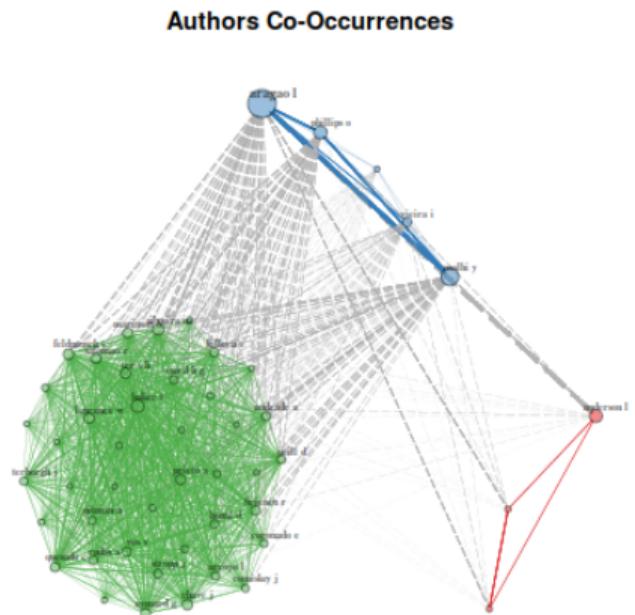


## Network - Authors collaboration

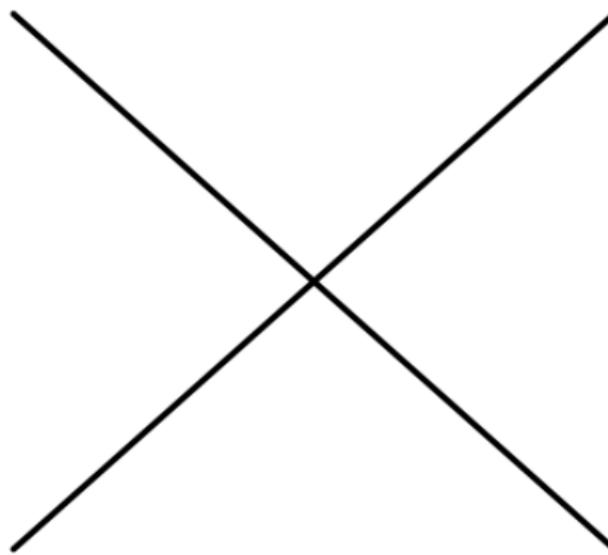
## Authors Collaboration



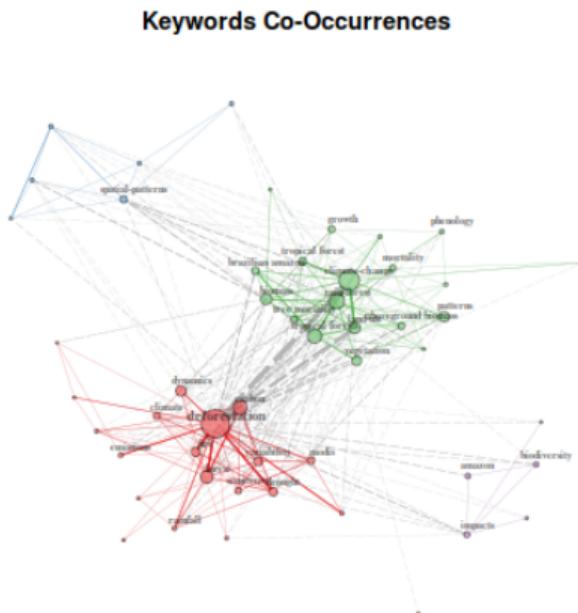
# Network - Authors co-occurrences



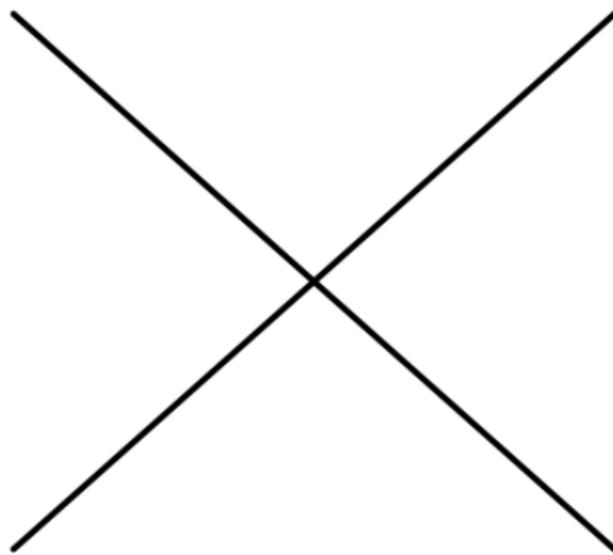
## Network - Authors coupling



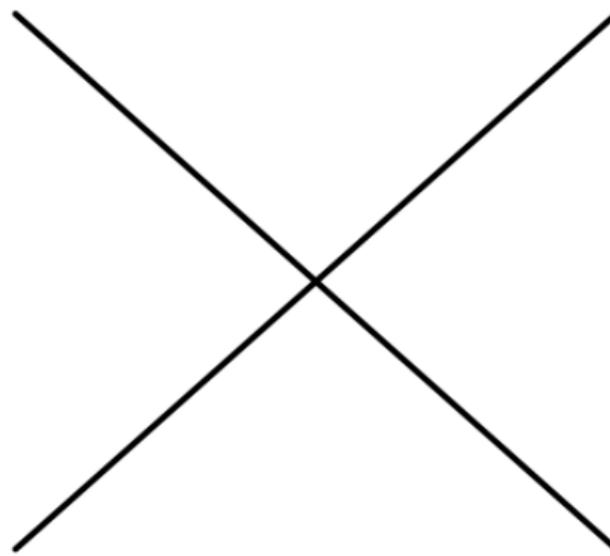
## Network - Keyword co-occurrences



## Network - References co-citation

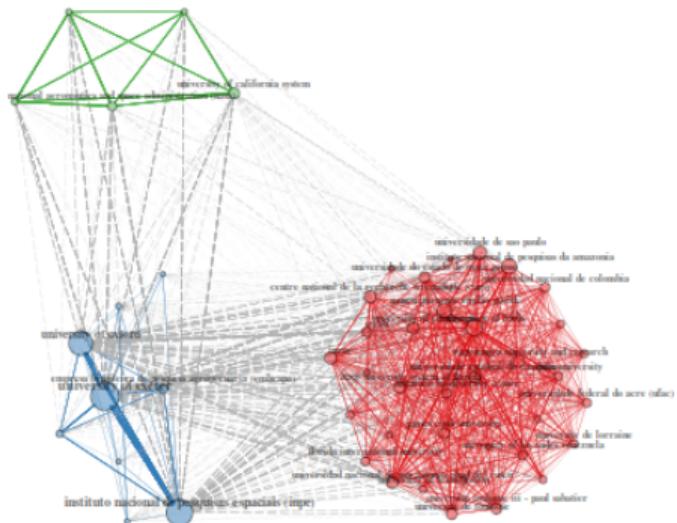


## Network - Sources coupling

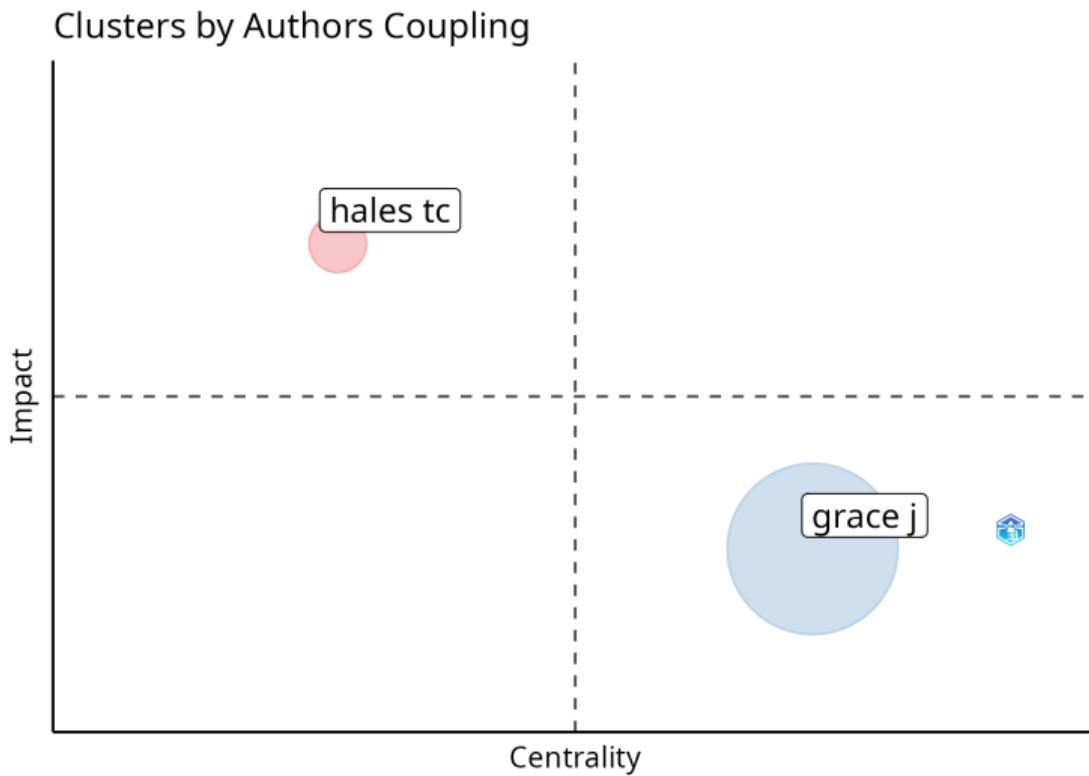


## Network - Universities collaboration

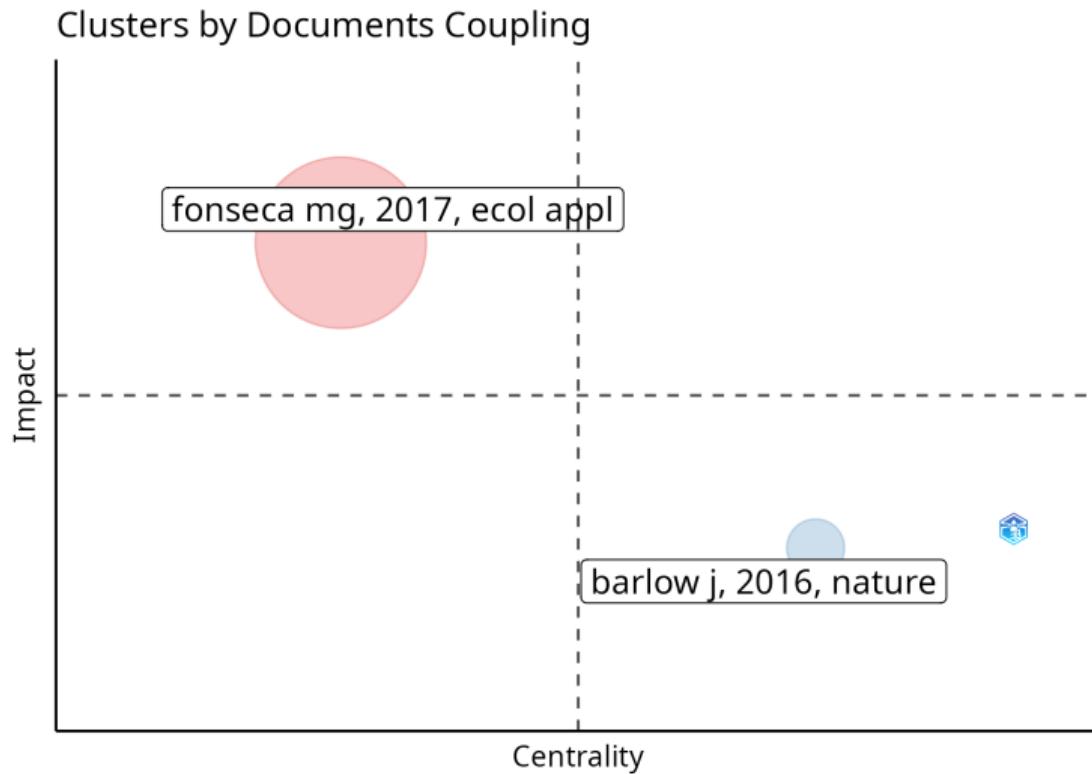
## **Universities Collaboration**



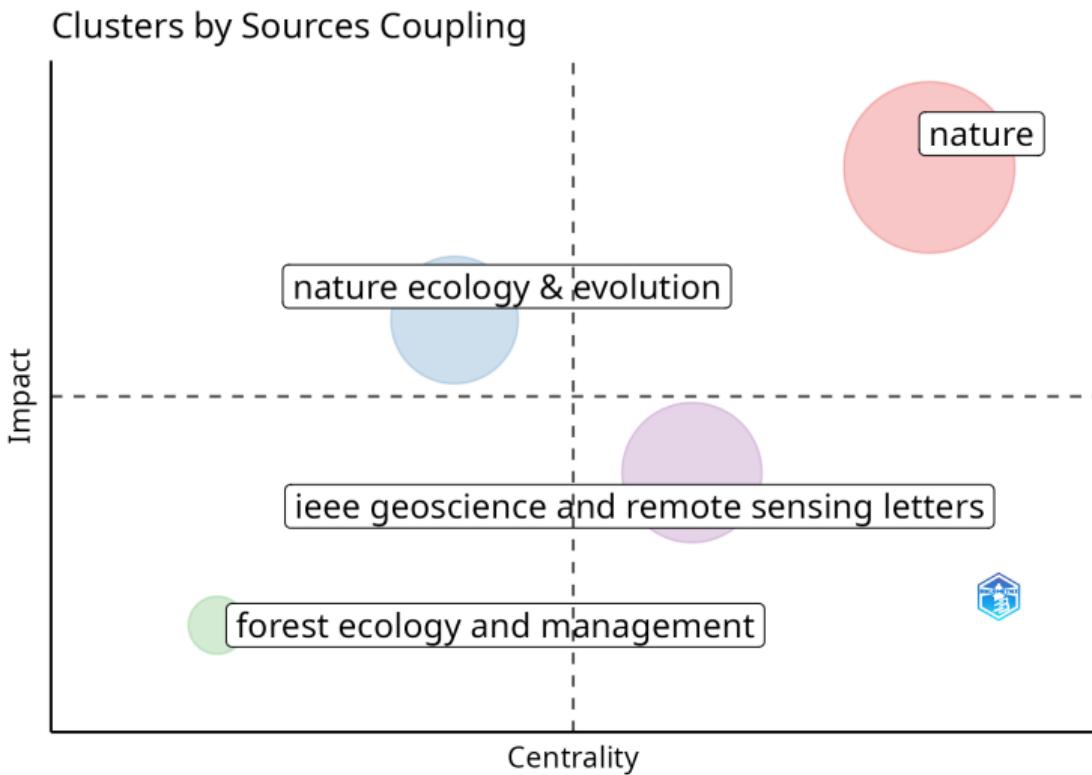
## Coupling network analysis (authors)



## 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 exercise 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 interdisciplinary research.

# References I