Systematic Reviews using R

Abhay Singh

11/09/2020

What is R?

According to the official webpage:

- R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.http://www.r-project.org/about.html

According to Wikipedia

- R is a free software programming language and a software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls and surveys of data miners are showing R's popularity has increased substantially in recent years.

To Summarise

- R (R Core Team, 2020), is a highly capable statistical programming language which has gained immense popularity among researchers, data analysts and in analytical professions around the globe.
- R provides a flexible and extensible free environment to conduct research and analysis and also contribute in the form of open source routines and packages which promotes reproducibility.
- R is the most amazing free statistical software ever!
- This recent video by Revolution Analytics does a great job in summarizing R

Systematic Reviews using R

Brief Introduction

- R's package ecosystem is one of its major advantages, there are packages available for most widely used statistical and data analysis & visualisation techniques used several packages added almost daily on new and upcoming methods published by academic researchers or industry practitioners.
- R provide packages for various areas of interest (see https://cran.r-project.org/web/views/ for a list of task views grouping packages according to their functionality) including systematic literature review or the related field of meta analysis.
- Bibliometrix (Aria and Cuccurullo (2017)), Revtools (Westgate (2018)) and Litsearchr (E. Grames et al. (2019),E. M. Grames et al. (2019)) of the Metaverse (https://rmetaverse.github.io/) project, Adjutant (Crisan, Munzner, and Gardy (2018)), Metagear (Lajeunesse (2016))) are a few providing various functionality.

- Bibliometrix is by far the most popular with several publications using the package
- The package webpage (http://www.bibliometrix.org/Papers.html) provides a list of publications utilising the package. (for example see, Lajeunesse (2016); Addor and Melsen (2019)) and hence we will use the package to demonstrate some of its functionality.

- We use the bibliometrix package for this example.

Upcoming Workshop on 15 and 24 September 2020

Registration Link

Demonstration using R

Files and Data available here https://github.com/singh1985/systematic_review_pres

Main Results

Descriptive Analysis

- Using the inbuild data in the package (for reproducibility)
  - Manuscripts including the term “bibliometrics” in the title.
  - Period: 1975 - 2017
  - Database: SCOPUS
  - Format: bibtex

Productive Authors

Table 1: Most Productive Authors

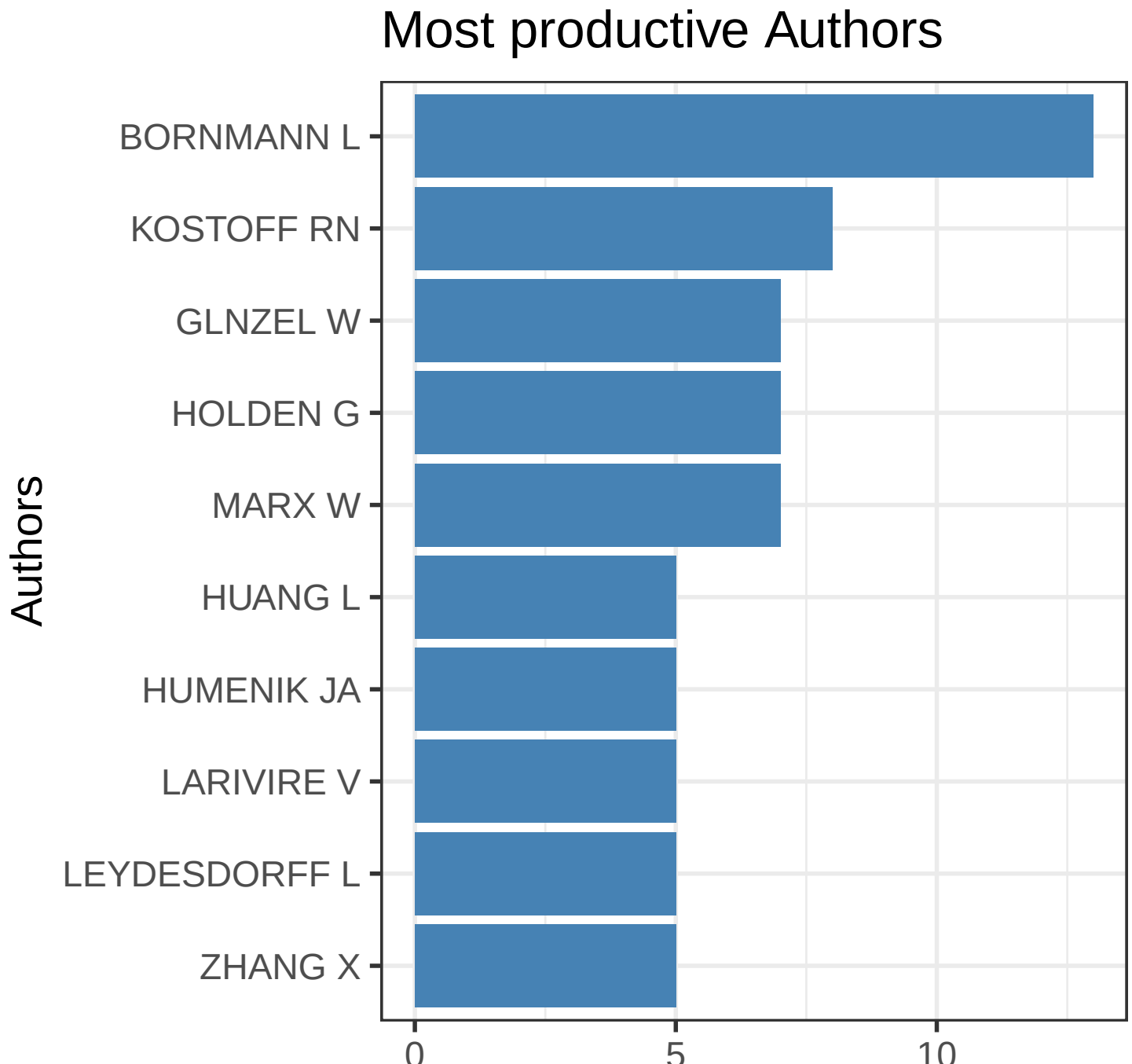
Authors	Articles	Authors	Articles Fractionalized
BORNMANN L	13	BORNMANN L	6.75
KOSTOFF RN	8	HOLDEN G	4.25
GLNZEL W	7	WHITE HD	4.00
HOLDEN G	7	MARX W	3.42
MARX W	7	ATKINSON R	3.00
HUANG L	5	NA	3.00
HUMENIK JA	5	GLNZEL W	2.67
LARIVIRE V	5	KIRBY A	2.50
LEYDESDORFF	5	PERITZ	2.50

Most cited papers

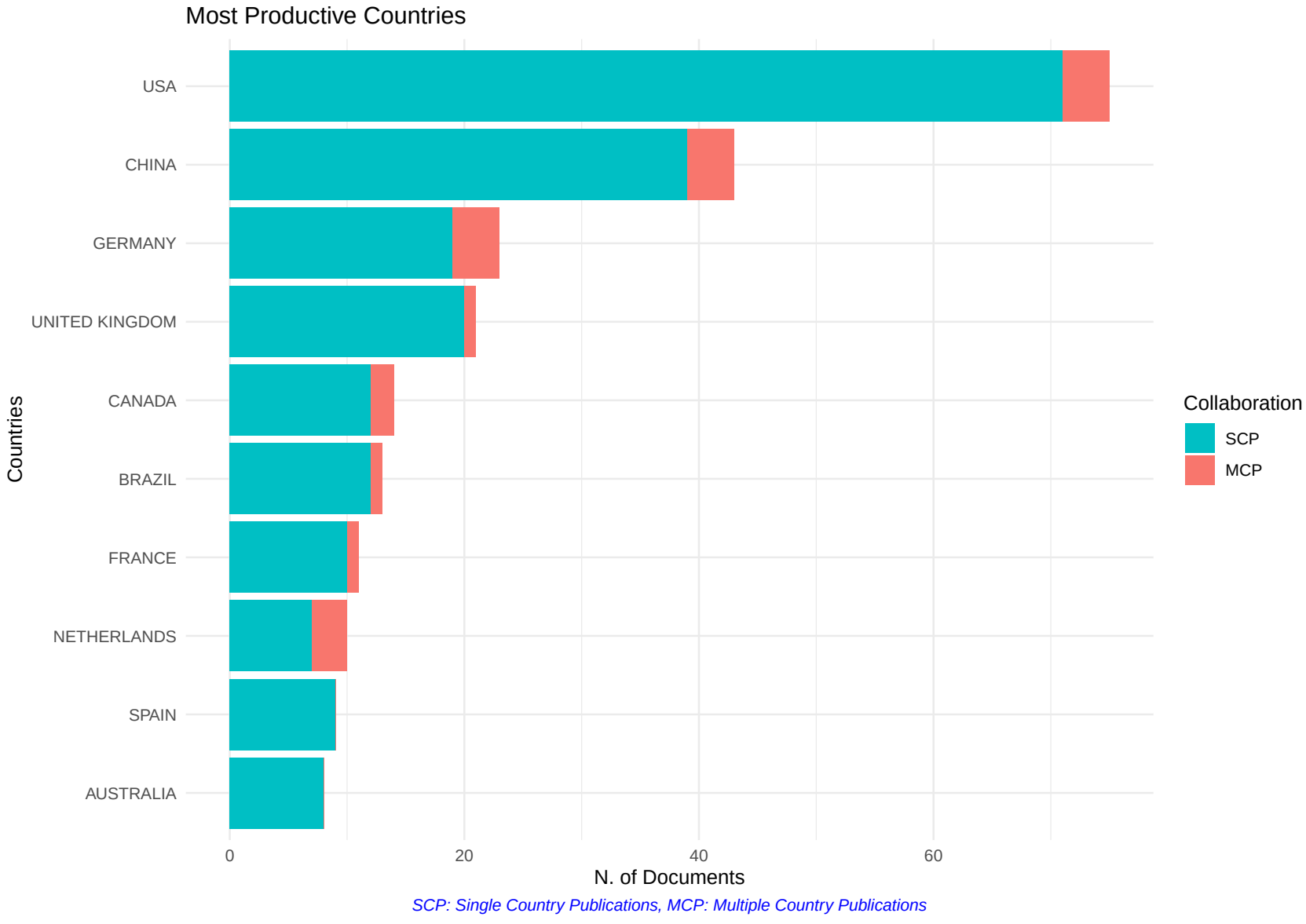
Table 2: Most Cited Papers

Paper	TC	TCperYear
DAIM TU , 2006, TECHNOL FORECAST SOC CHANGE	331	22.07
BORGMAN CL , 2002, ANNU REV INF SCI TECHNOL	312	16.42
WEINGART P, 2005, SCIENTOMETRICS	208	13.00
NARIN F, 1994, SCIENTOMETRICS	169	6.26
CRONIN B, 2001, J INF SCI	160	8.00
HOOD WW , 2001, SCIENTOMETRICS	144	7.20
HICKS D , 2015, NATURE	130	21.67
CHEN Y-C , 2011,	129	12.90

Summary Plot-1 (Most Porductive Authors)

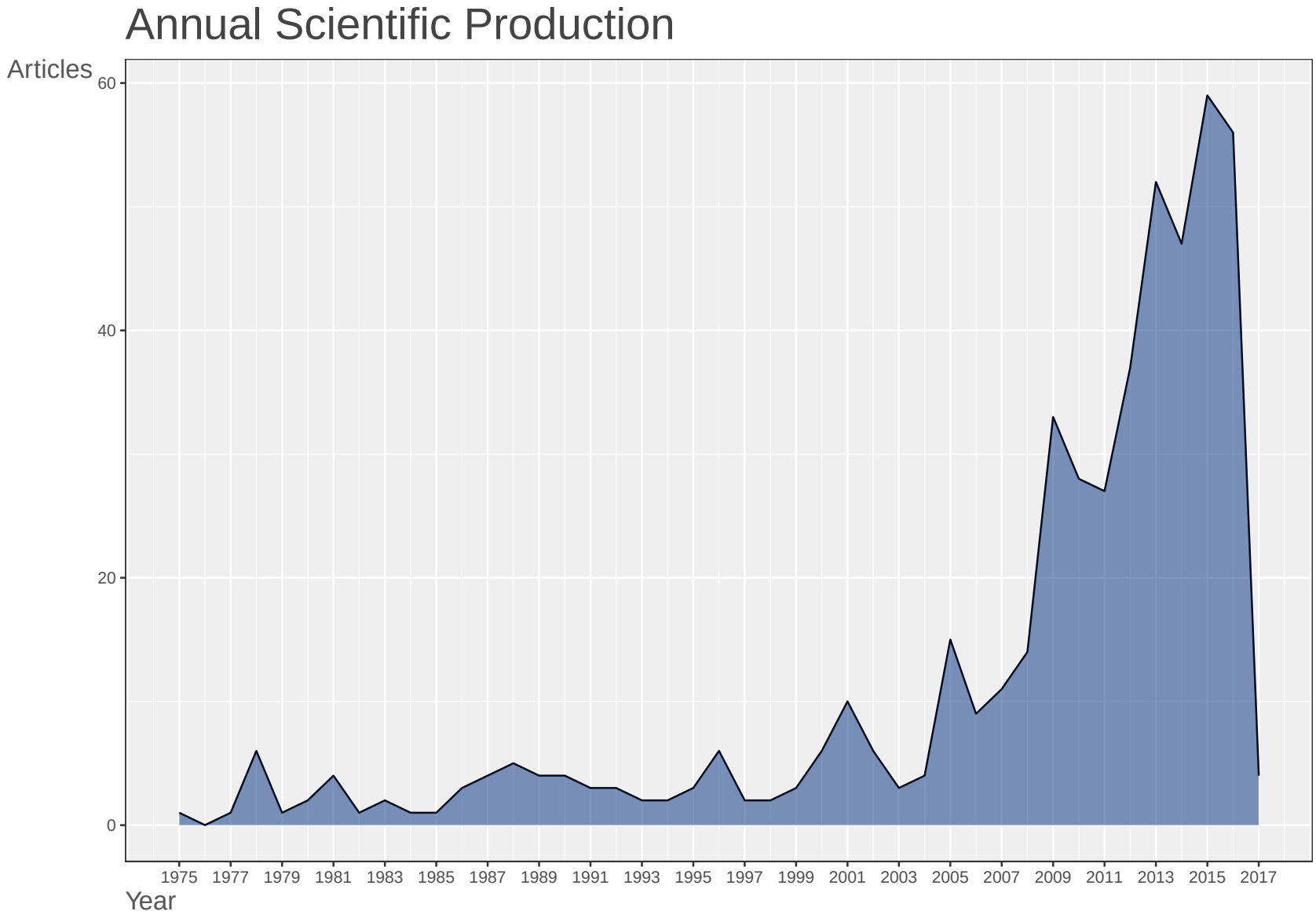


Summary Plot-2 (Most Productive Countries)

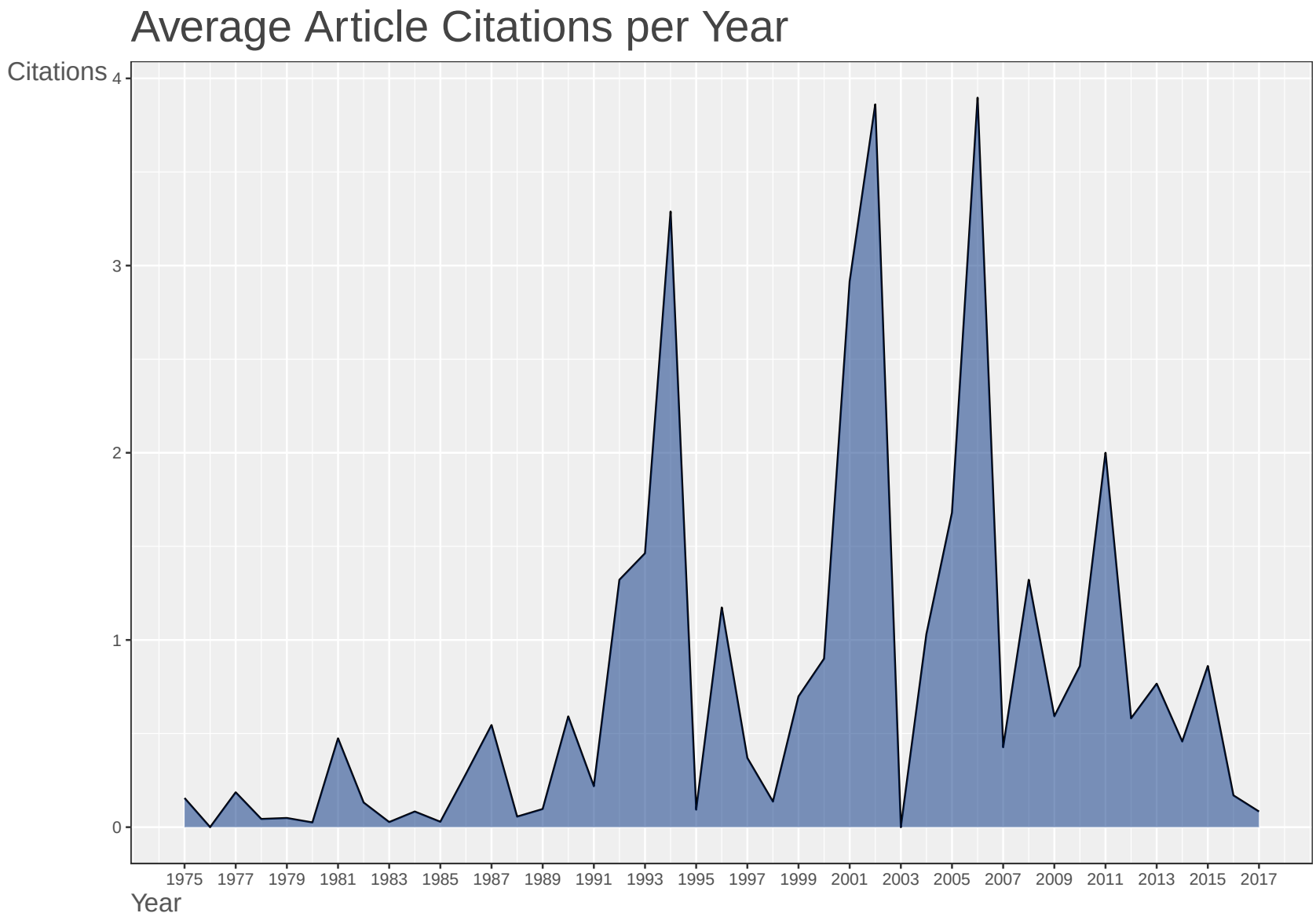


SCP: Single Country Publications, MCP: Multiple Country Publications

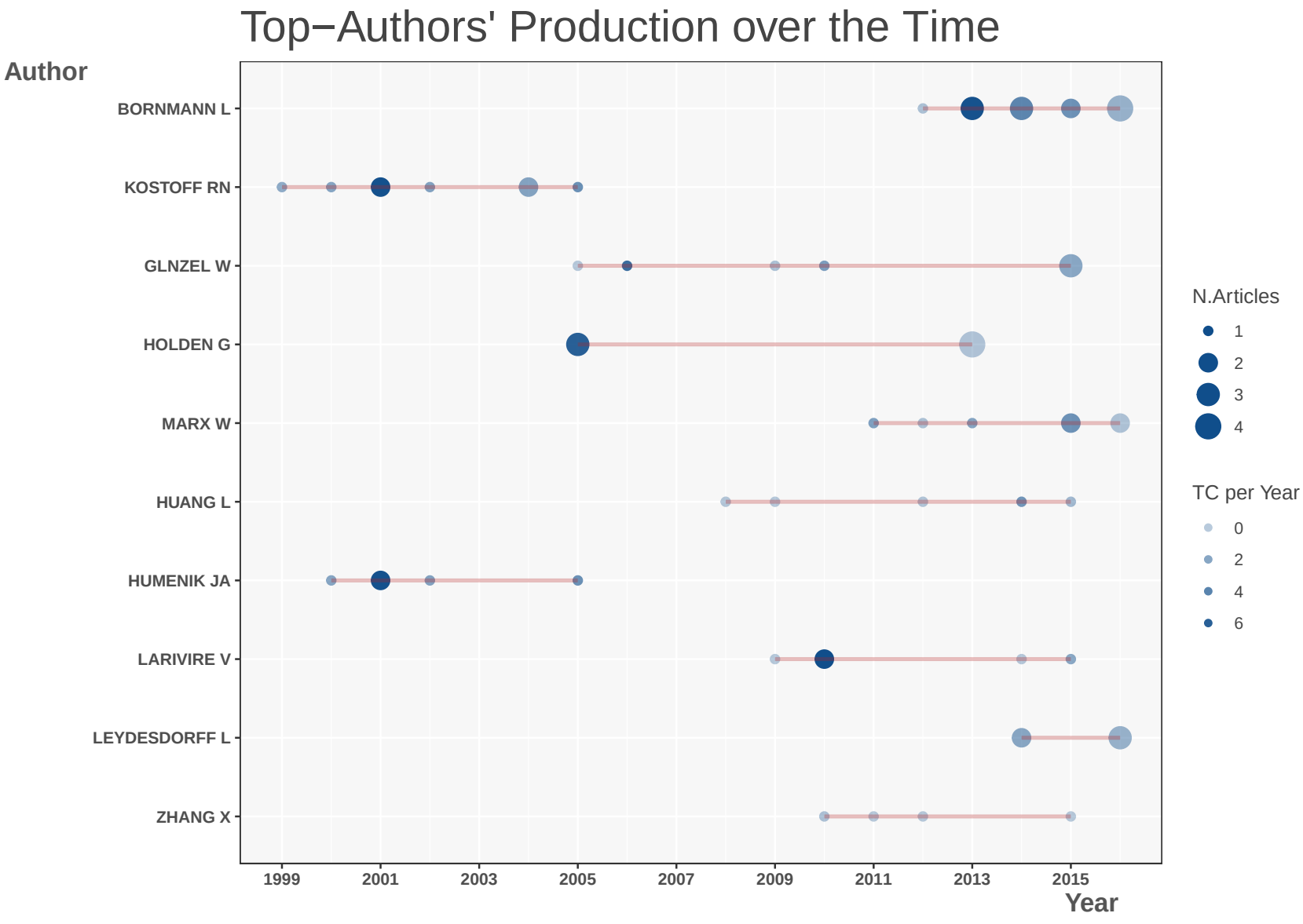
Summary Plot-3 (Annual Scientific Production)



Summary Plot-4 (Average Article Citation)

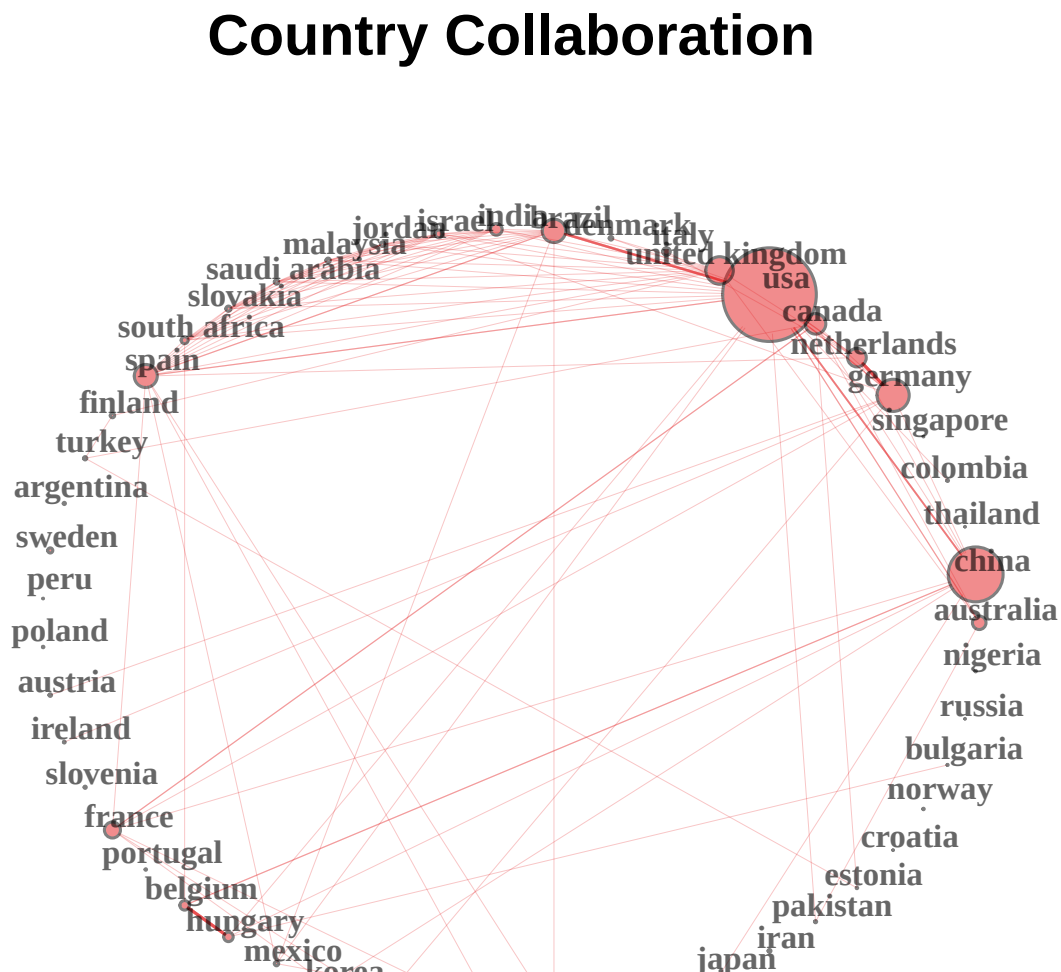


Top Authors Over time



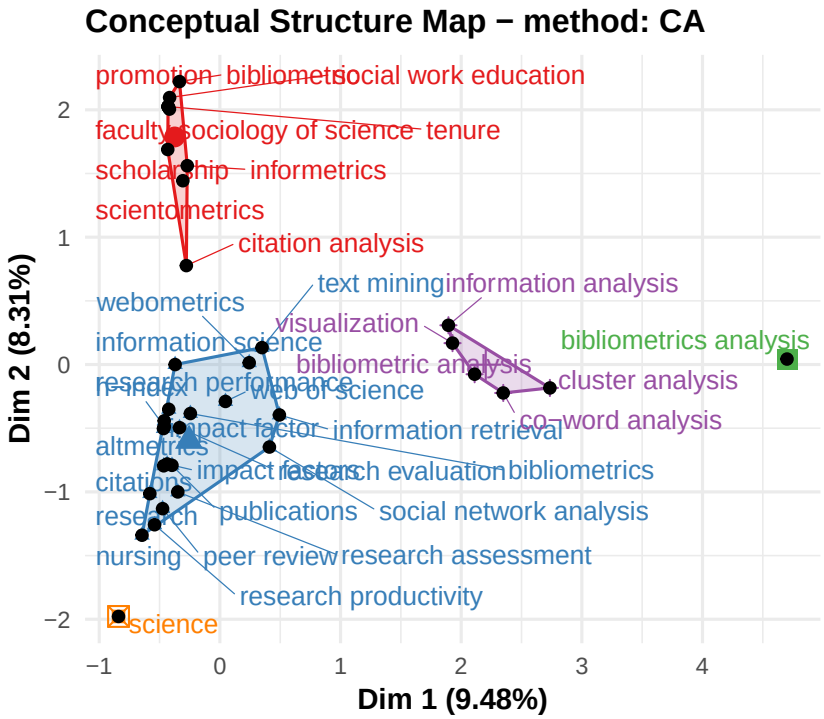
Network Plots

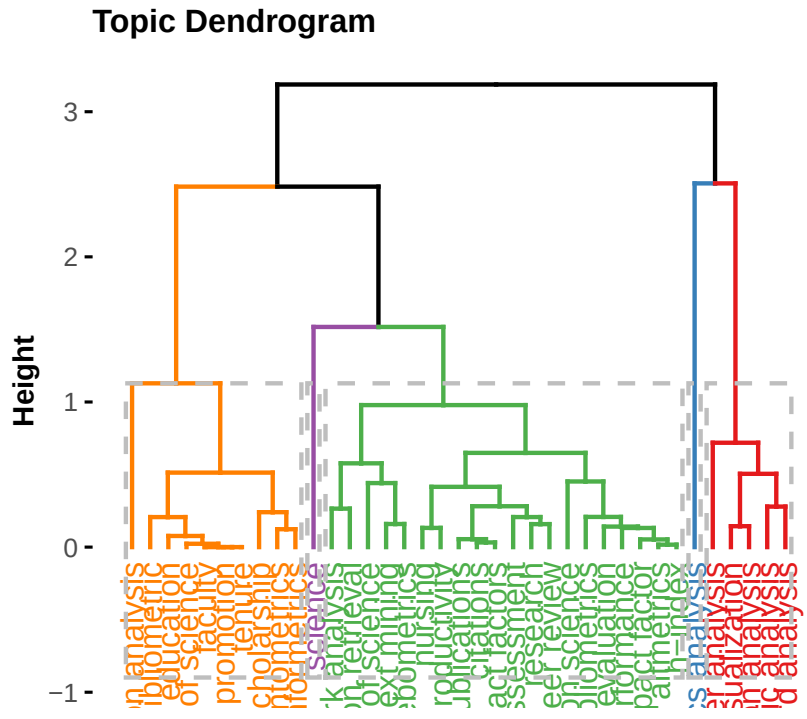
- The package also facilitates various network analysis like, co-citation analysis, coupling analysis, collaboration analysis or co-occurrence analysis. Figure-2 shows a key word co-occurrence plot

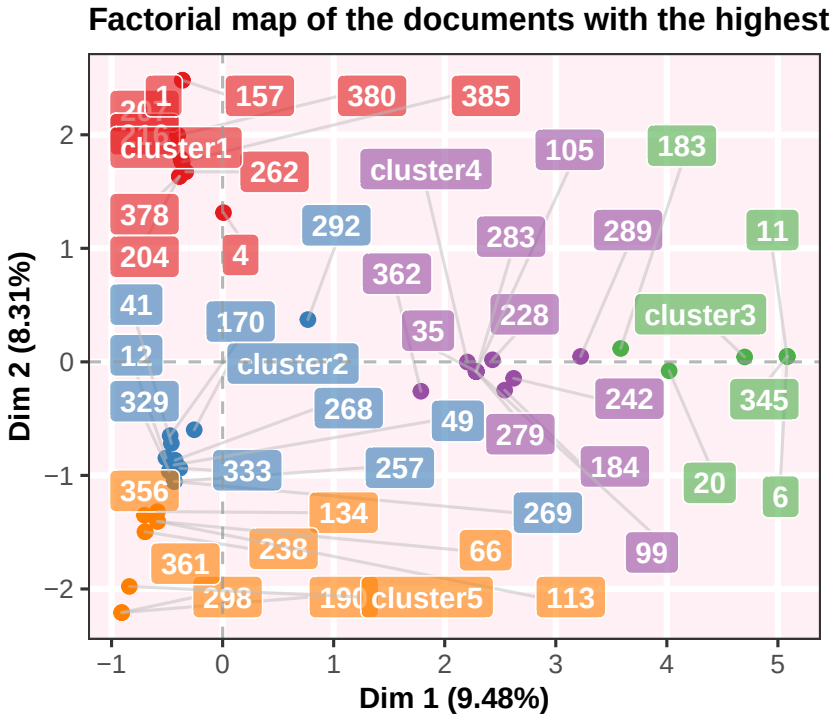


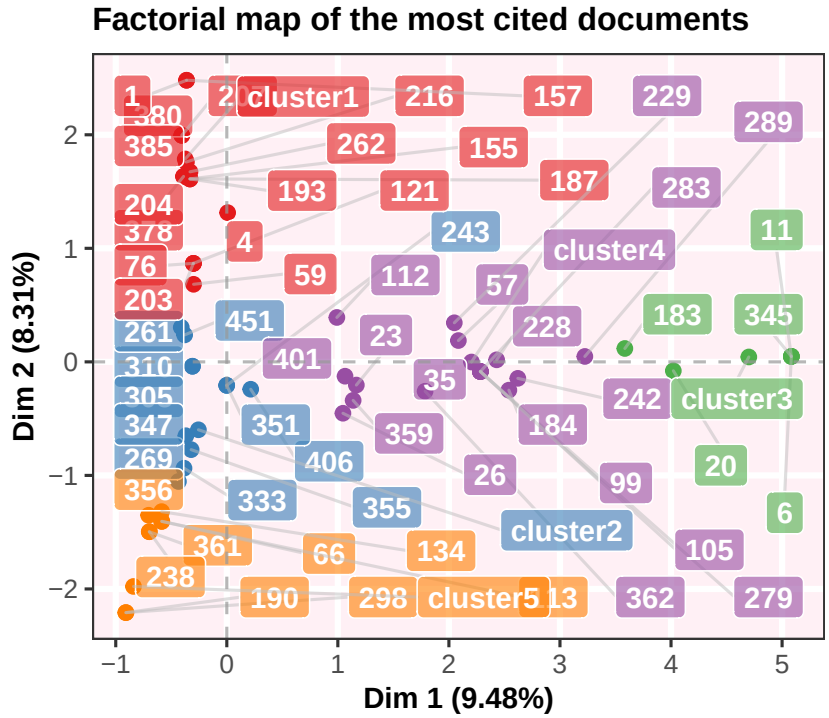
Co-word Analysis

- Analysis of the conceptual structure among the articles analysed.
- Bibliomentrix can conduct a co-word analysis to map the conceptual structure of a framework using the word co-occurrences in a bibliographic database.
- The analysis in the figure is conducted using the Correspondence Analysis and K-Means clustering using Author's keywords. This analysis includes Natural Language Processing and is conducted without stemming.


Conceptual Structure Map − method: CA
promotion
bibliometrics
social work education
faculty
sociology of science
tenure
scholarship
informetrics
scientometrics
citation analysis
text mining
information analysis
webometrics
visualization
information science
bibliometric analysis
bibliometrics analysis
cluster analysis
research performance
h−index
web of science
co−word analysis
impact factor
information retrieval
altmetrics
impact measure
research evaluation
bibliometrics
citations
research
publications
social network analysis
nursing
peer review
research assessment
research productivity
science
Dim 2 (8.31%)
Dim 1 (9.48%)


Topic Dendrogram
Height


Factorial map of the documents with the highest
cluster1
cluster4
cluster2
cluster3
cluster5
Dim 2 (8.31%)
Dim 1 (9.48%)


Factorial map of the most cited documents
cluster1
cluster4
cluster2
cluster3
cluster5
Dim 2 (8.31%)
Dim 1 (9.48%)

Author collaboration network

There is a gui too!

```
biblioshiny()
```

References

Addor, N, and LA Melsen. 2019. "Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models." *Water Resources Research* 55 (1): 378–90.

Aria, Massimo, and Corrado Cuccurullo. 2017. "Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis." *Journal of Informetrics* 11 (4): 959–75. https://doi.org/10.1016/j.joi.2017.08.007.

Crisan, Anamaria, Tamara Munzner, and Jennifer L Gardy. 2018. "Adjutant: An R-Based Tool to Support Topic Discovery for Systematic and Literature Reviews." *Bioinformatics* 35 (6): 1070–2.

Grames, Eliza M., Andrew N. Stillman, Morgan W. Tingley, and Chris S. Elphick. 2019. "An Automated Approach to Identifying Search Terms for Systematic Reviews Using Keyword Co-Occurrence Networks." *Methods in Ecology and Evolution* 0 (ja). https://doi.org/10.1111/2041-210X.13268.

Grames, Eliza, Andrew Stillman, Morgan Tingley, and Chris Elphick. 2019. *Litsearchr: Automated Search Term Selection and Search*