# Systematic Literature Review

Abhay Singh

## Bibliometrix Analysis using R

- Bibliometrix (https://www.bibliometrix.org/) allows R users to import a bibliography database generated using SCOPUS and Web of Science stored either as a Bibtex (.bib) or Plain Text (.txt) file.

- The package has simple functions which allows for descriptive analyses as shown in table-1 to table-3.

- The analysis can also be easily visualised as shown in figure-1.

```r
library(bibliometrix)   #load the package
library(pander)   #other required packages
library(knitr)
library(kableExtra)
library(ggplot2)
# use scopuscollection data from the package Manuscripts
# including the term 'bibliometrics' in the title.  Period:
# 1975 - 2017 Database: SCOPUS Format: bibtex
data("scopusCollection")
file1 = data("scopusCollection")
# M=convert2df(file='insert
# filename',format='bibtex',dbsource = 'scopus')#convert the
# data to data frame
```

## Descriptive Analysis

```r
# Descriptive analysis
M = scopusCollection   #just to reuse the other code
res1 = biblioAnalysis(M, sep = ";")
s1 = summary(res1, k = 10, pause = FALSE, verbose = FALSE)

d1 = s1$MainInformationDF   #main information
d2 = s1$MostProdAuthors   #Most productive Authors
d3 = s1$MostCitedPapers   #most cited papers
pander(d1, caption = "Summary Information")
```

Table 1: Summary Information

| Description | Results |
|---|---|
| MAIN INFORMATION ABOUT DATA | |
| Timespan | 1975:2017 |
| Sources (Journals, Books, etc) | 280 |
| Documents | 487 |
| Average years from publication | 11.6 |
| Average citations per documents | 10.36 |

| Description | Results |
|---|---|
| Average citations per year per doc | 0.7799 |
| References | 12245 |
| DOCUMENT TYPES | |
| article | 417 |
| book | 12 |
| conference | 58 |
| DOCUMENT CONTENTS | |
| Keywords Plus (ID) | 1436 |
| Author's Keywords (DE) | 722 |
| AUTHORS | |
| Authors | 949 |
| Author Appearances | 1187 |
| Authors of single-authored documents | 162 |
| Authors of multi-authored documents | 787 |
| AUTHORS COLLABORATION | |
| Single-authored documents | 184 |
| Documents per Author | 0.513 |
| Authors per Document | 1.95 |
| Co-Authors per Documents | 2.44 |
| Collaboration Index | 2.6 |

## Productive Authors

```
s1$MostProdAuthors
```

```
##      Authors     Articles Authors      Articles Fractionalized
## 1    BORNMANN L      13    BORNMANN L                     6.75
## 2    KOSTOFF RN       8    HOLDEN G                       4.25
## 3    GLNZEL W         7    WHITE HD                       4.00
## 4    HOLDEN G         7    MARX W                         3.42
## 5    MARX W           7    ATKINSON R                     3.00
## 6    HUANG L          5    NA                             3.00
## 7    HUMENIK JA       5    GLNZEL W                       2.67
## 8    LARIVIRE V       5    KIRBY A                        2.50
## 9    LEYDESDORFF L    5    PERITZ BC                      2.50
## 10   ZHANG X          5    SMITH DR                       2.50
```

```
pander(d2, caption = "Most Productive Authors", table.split = Inf)
```

Table 2: 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 L | 5 | PERITZ BC | 2.50 |
| ZHANG X | 5 | SMITH DR | 2.50 |

## Most cited papers

```
pander(d3, caption = "Most Cited Papers")
```

Table 3: 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, SCIENTOMETRICS | 129 | 12.90 |
| D'ANGELO CA , 2011, J AM SOC INF SCI TECHNOL | 81 | 8.10 |
| GLNZEL W , 2006, SCIENTOMETRICS | 78 | 5.20 |

## Information Plots

```
p1 = plot(res1, pause = FALSE)
```

## Summary Plot-1 (Most Porductive Authors)
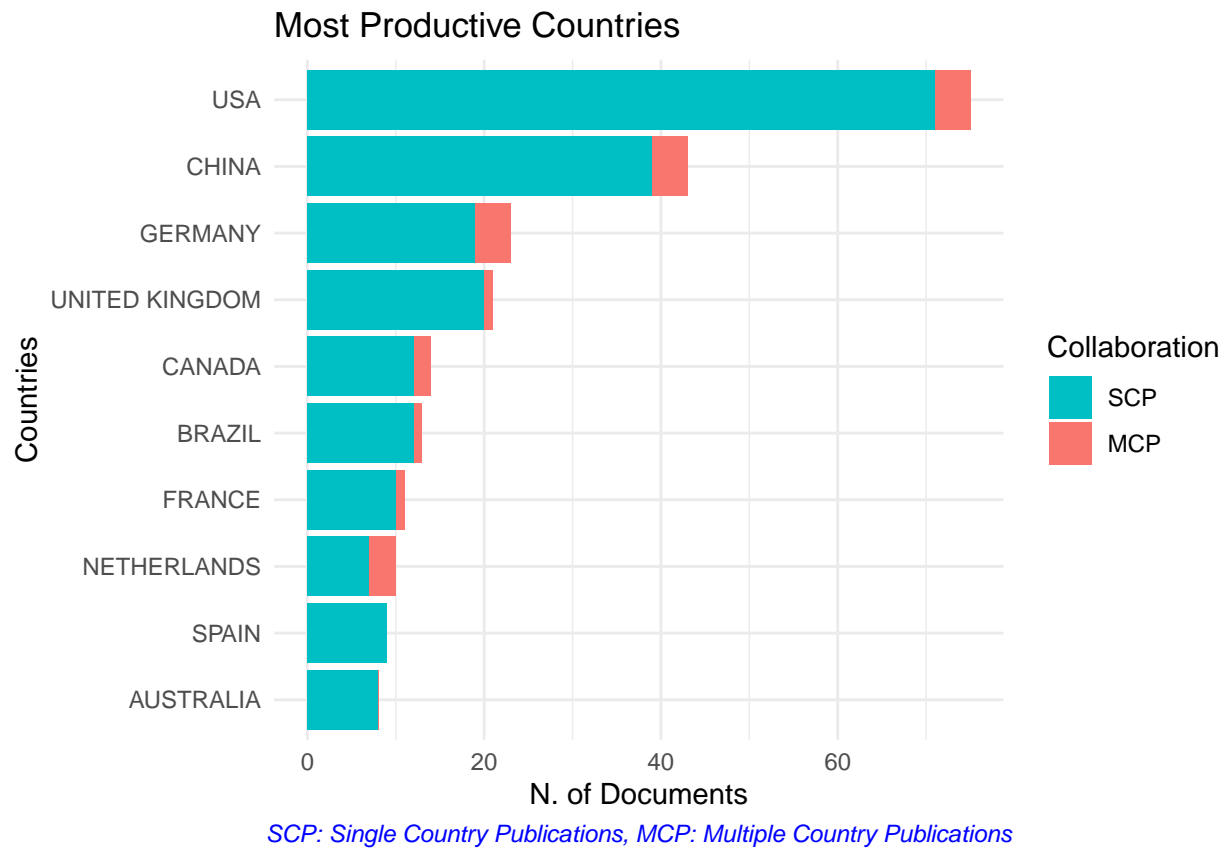
```
library(ggplot2)
theme_set(theme_bw())


p1[[1]] + theme_bw() + scale_x_discrete(limits = rev(levels(as.factor(p1[[1]]$data$AU))))
```

## Most productive Authors
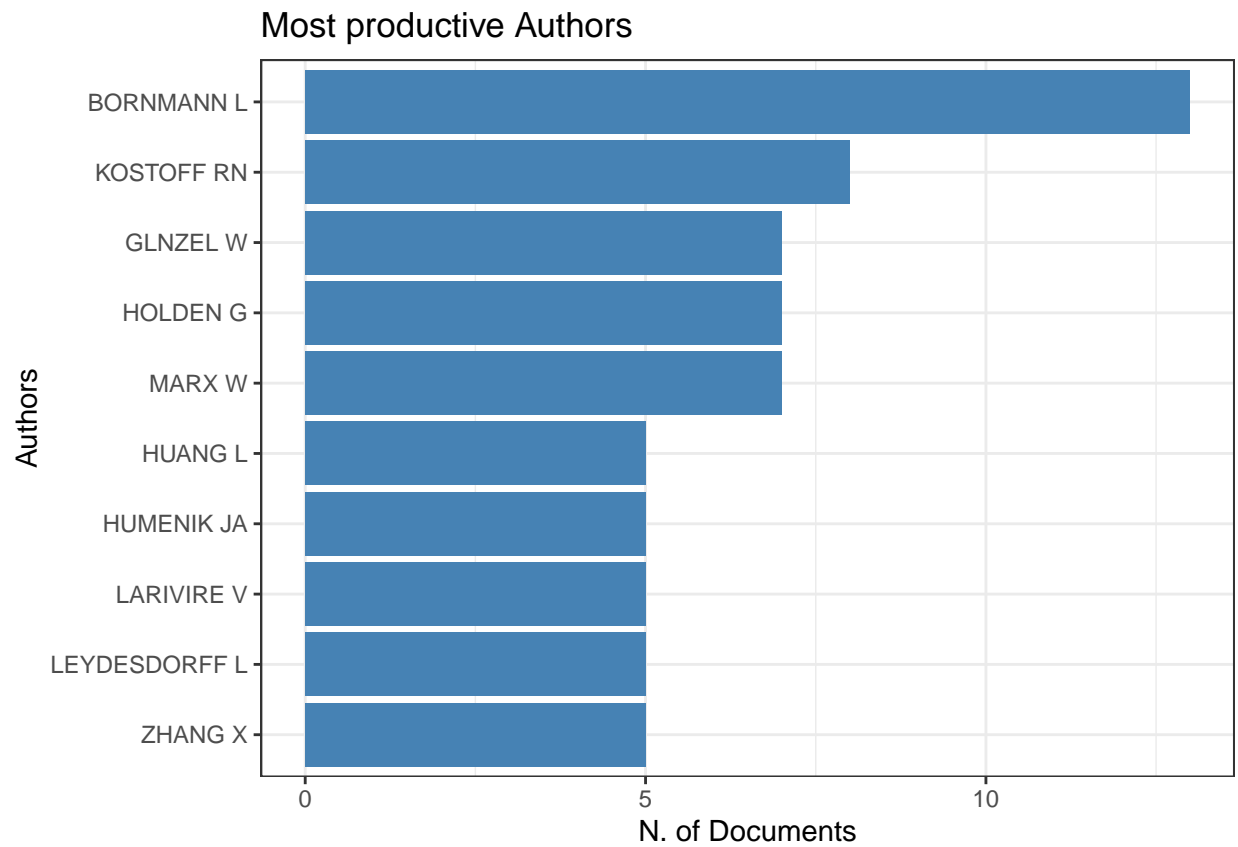


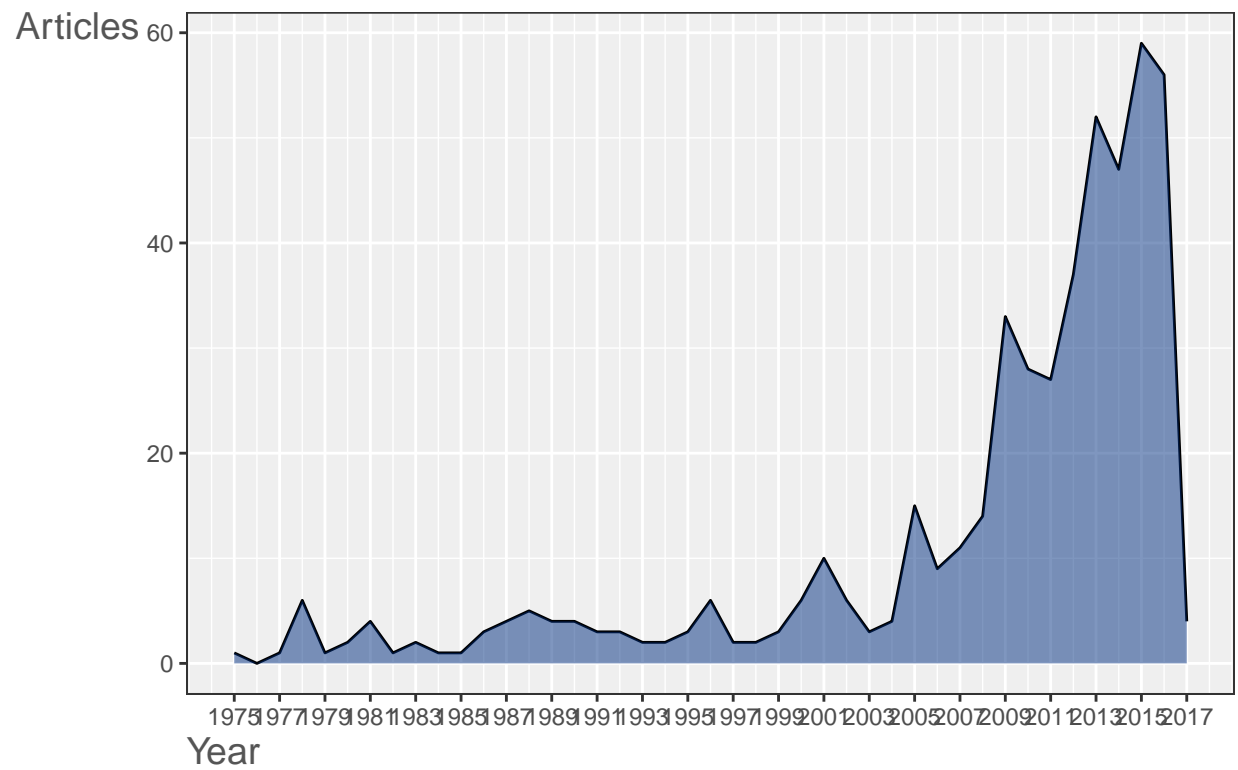## Summary Plot-2 (Most Productive Countries)

```
p1[[2]]
```

## Most Productive Countries



*SCP: Single Country Publications, MCP: Multiple Country Publications*

## Summary Plot-3 (Annual Scientific Production)

```
p1[[3]]
```

# Annual Scientific Production

Articles

60

40

20

0

1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017

Year

## Summary Plot-4 (Average Article Citation)
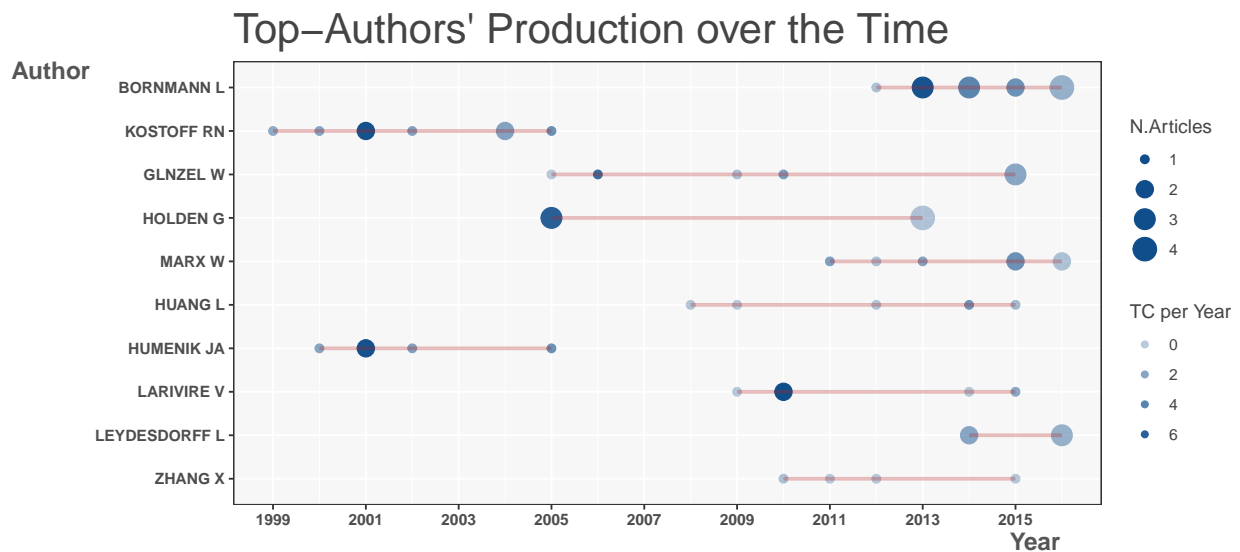
```
p1[[4]]
```

# Average Article Citations per Year

Citations

Year
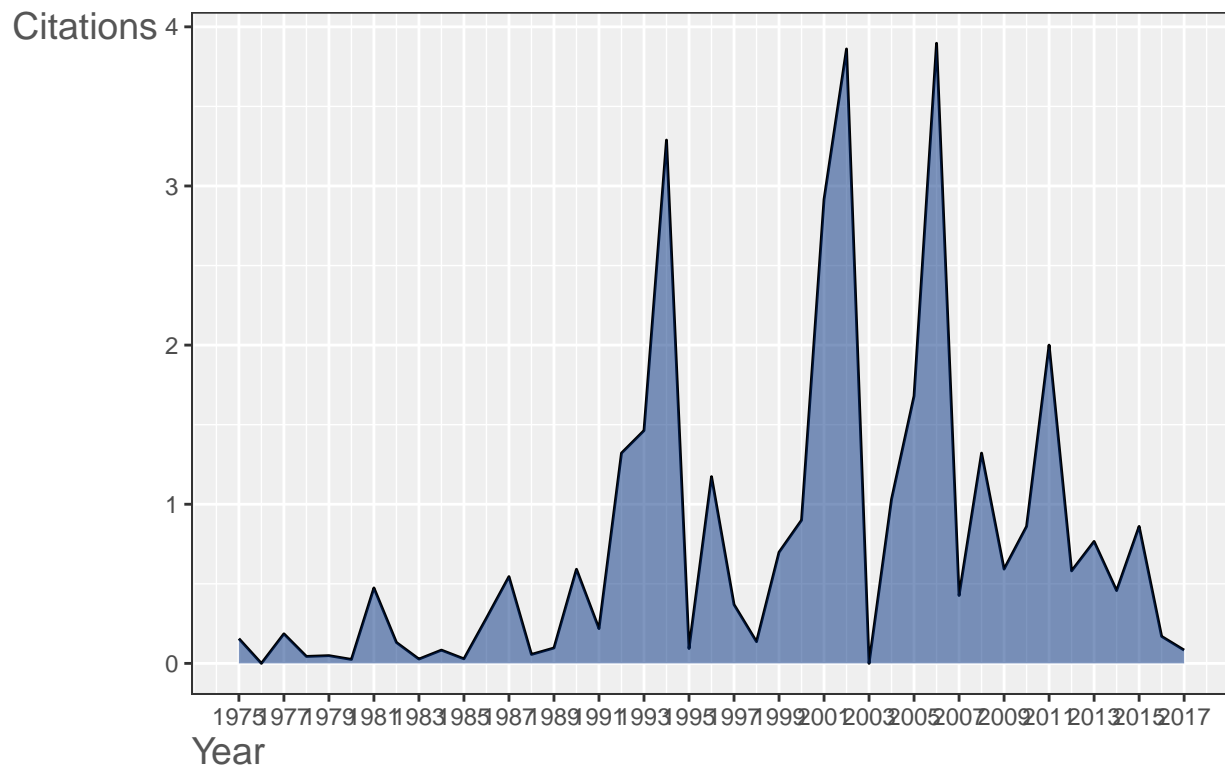
* A graph for author statistics over time can also be produced.

- Figure-1 shows a graph of top 10 authors over time. The information from these plots can be easily extracted to summarise them in a table.

```
topAU = authorProdOverTime(M, k = 10, graph = TRUE)
```

# Top−Authors' Production over the Time

Author

Year

- 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

```
M <- metaTagExtraction(M, Field = "AU_CO", sep = ";")
NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "countries",
    sep = ";")
# Plot the network
net = networkPlot(NetMatrix, n = dim(NetMatrix)[1], Title = "Country Collaboration",
    type = "circle", size = TRUE, remove.multiple = FALSE, labelsize = 0.7,
    cluster = "none")
```
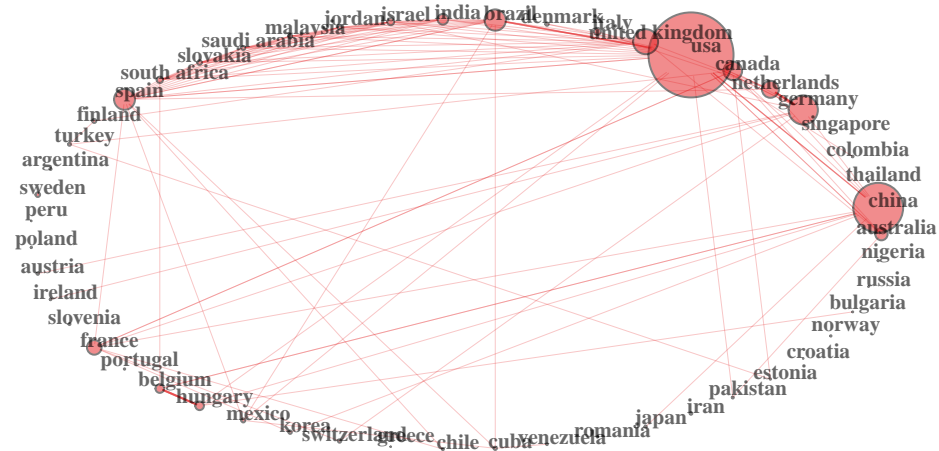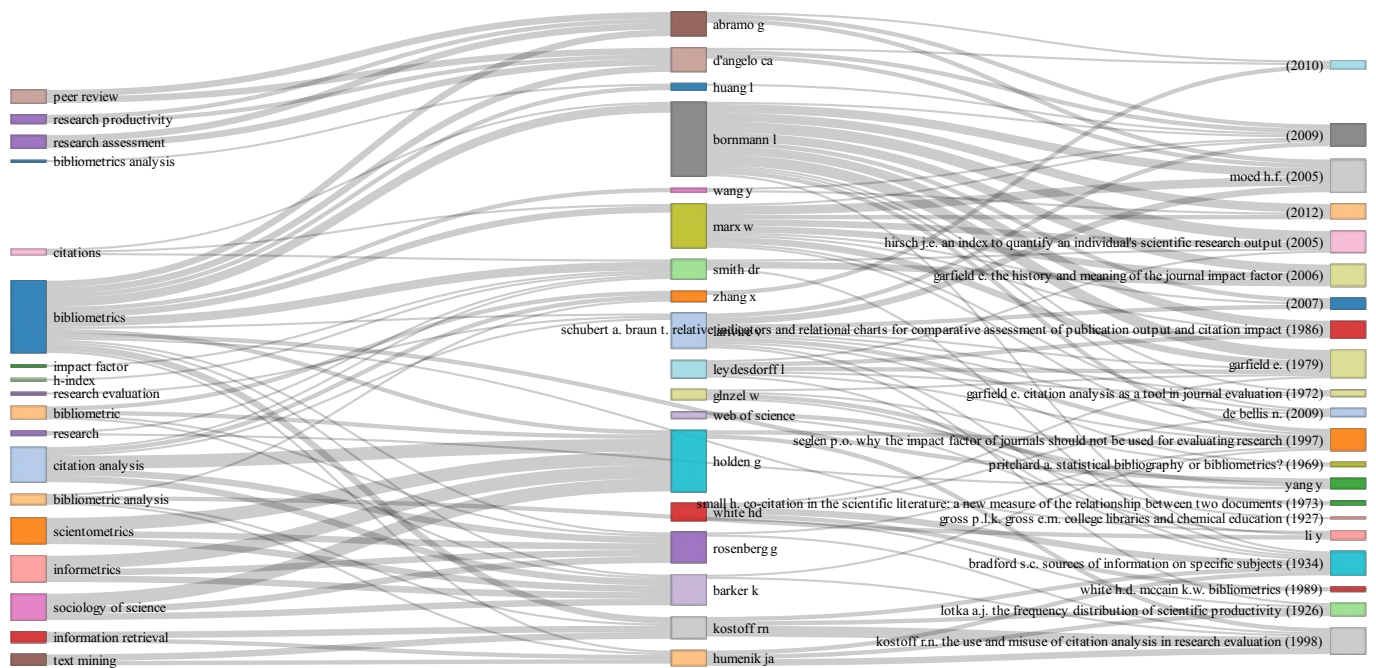
## Country Collaboration



Figure 1: Country Collaboration

- Bibliometrix provides another useful function to plot a Sankey diagram to visualise multiple attributes at the same time. For example, figure-9 provides a three fields plot for Author, Author Keywords and Cited References.

```
threeFieldsPlot(M, fields = c("DE", "AU", "CR"))
```

## 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 Figure-2 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.

```
library(gridExtra)
CS = conceptualStructure(M, field = "DE", method = "CA", minDegree = 4,
    clust = 5, stemming = FALSE, labelsize = 10, documents = 10,
    graph = FALSE)

grid.arrange(CS[[4]], CS[[5]], CS[[6]], CS[[7]], ncol = 2, nrow = 2)
```

## Author collaboration network

```
NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "authors",
    sep = ";")
net = networkPlot(NetMatrix, n = 20, Title = "Author collaboration",
    type = "auto", size = 10, size.cex = T, edgesize = 3, labelsize = 0.6)
```
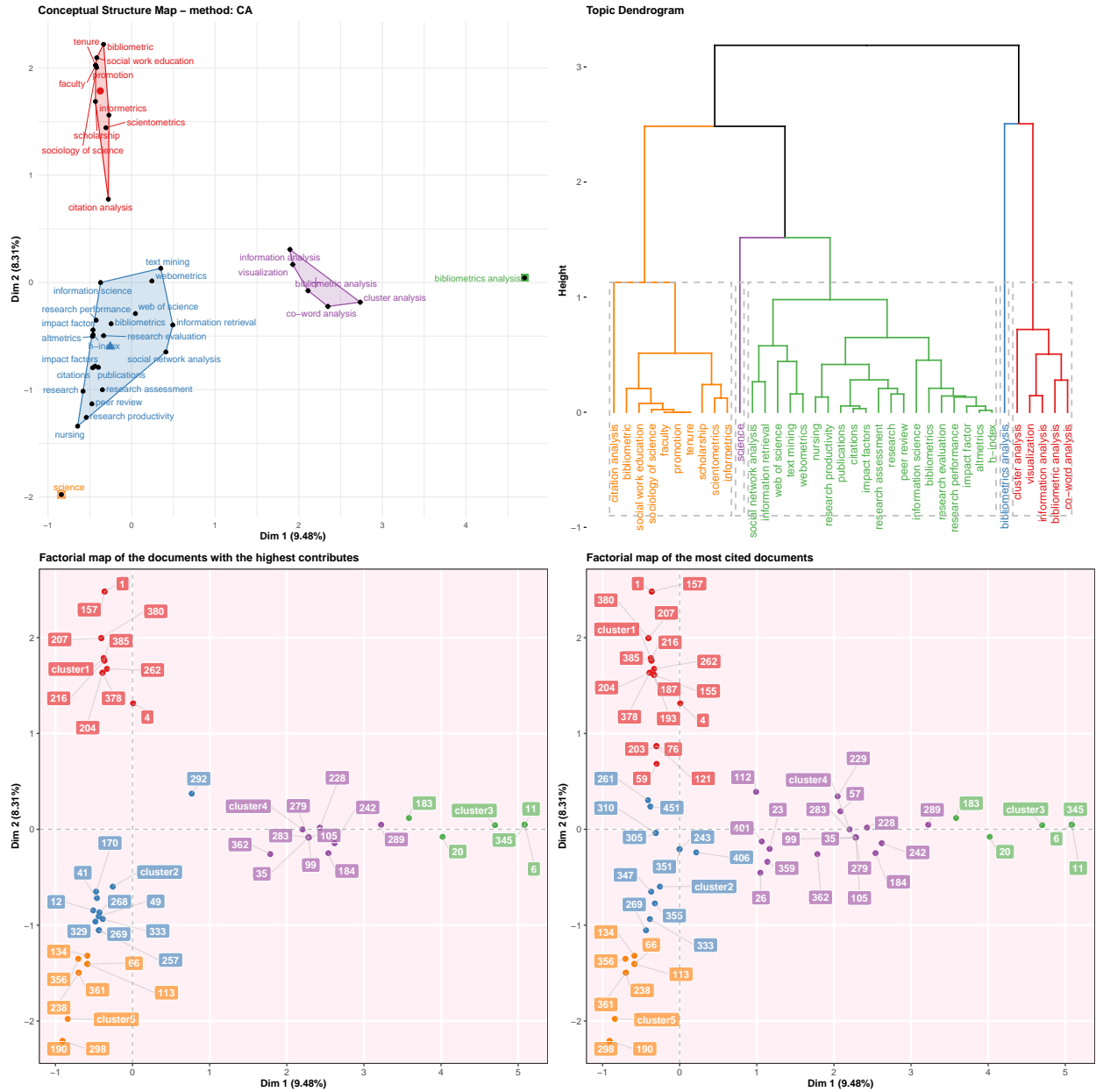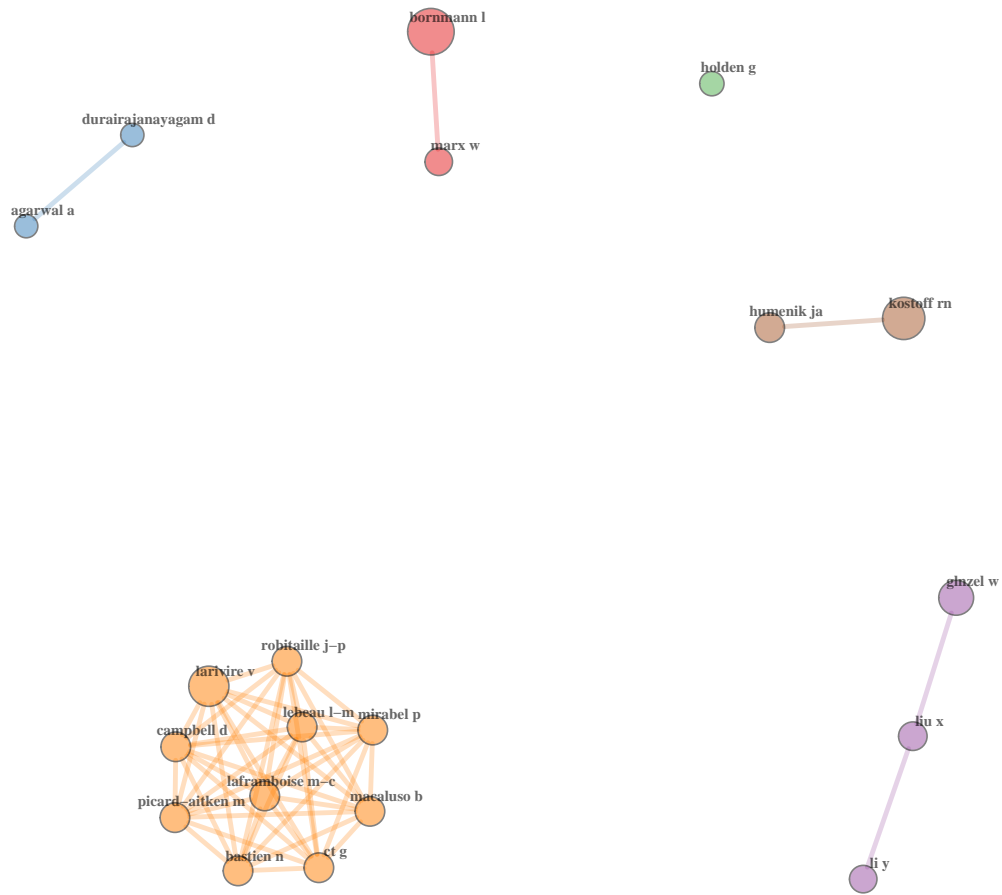
Figure 2: Conceptual Structure
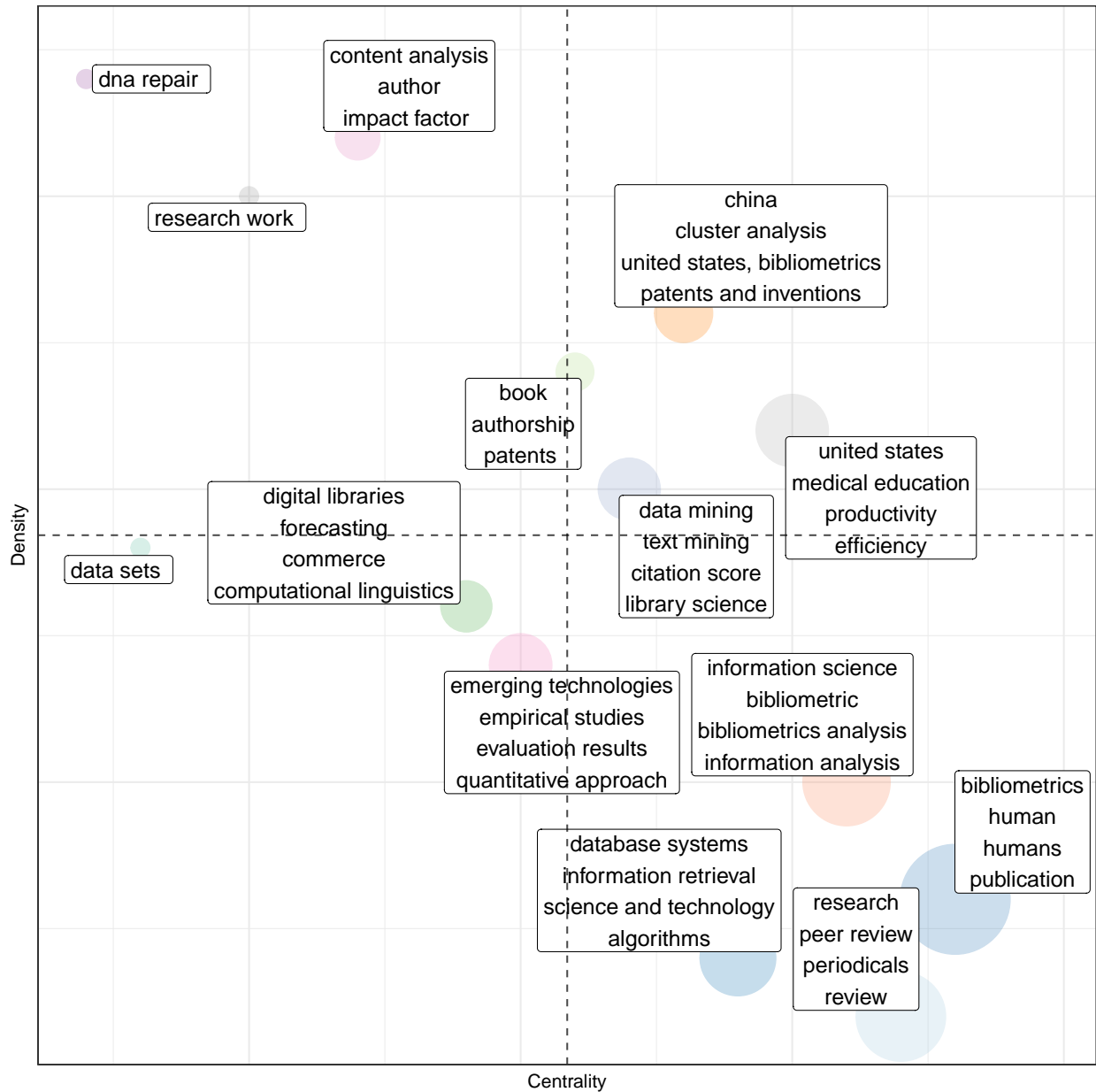
**Author collaboration**



# Thematic Map

Co-word analysis draws clusters of keywords. They are considered as themes, whose density and centrality can be used in classifying themes and mapping in a two-dimensional diagram.

Thematic map is a very intuitive plot and we can analyze themes according to the quadrant in which they are placed: (1) upper-right quadrant: motor-themes; (2) lower-right quadrant: basic themes; (3) lower-left quadrant: emerging or disappearing themes; (4) upper-left quadrant: very specialized/niche themes.

```r
# Map2=thematicEvolution(M3,field='ID',n=1000,stemming=FALSE,repel=TRUE,years=2000)
Map = thematicMap(M, field = "ID", n = 1000, minfreq = 5, stemming = FALSE,
    size = 0.5, n.labels = 4, repel = TRUE)
plot(Map$map)
```

The plot is a two-dimensional thematic map with "Centrality" on the x-axis and "Density" on the y-axis, showing clusters of terms:

- dna repair
- content analysis, author, impact factor
- research work
- china, cluster analysis, united states, bibliometrics, patents and inventions
- book, authorship, patents
- digital libraries, forecasting, commerce, computational linguistics
- data sets
- data mining, text mining, citation score, library science
- united states, medical education, productivity, efficiency
- emerging technologies, empirical studies, evaluation results, quantitative approach
- information science, bibliometric, bibliometrics analysis, information analysis
- bibliometrics, human, humans, publication
- database systems, information retrieval, science and technology, algorithms
- research, peer review, periodicals, review

## There is a gui too!

```
biblioshiny()
```

This concludes the example. There are various online sources to take this further