

# Learning bibliometrix

M.Lagisz

04/06/2019

## Before the lesson:

Please make sure you got the latest RStudio and latest R version installed.

## Lesson objectives:

- \* learn to perform a search in academic literature database
- \* download search results and import them into R
- \* summarise bibliometric data
- \* make a few types of simple bibliometric networks
- \* plot bibliometric networks

## Lesson outline:

- \* About this lesson
  - \* Getting bibliometric data
  - \* Summarising bibliometric data
  - \* Creating bibliometric networks
  - \* More resources
- 

## About this lesson

This lesson is prepared for those who are already familiar with R coding language, R markdown and RStudio. By the end of this tutorial you should be able to create a simple html document containing markdown-formatted text, images and R code, all in R Studio.

---

## Getting bibliometric data

You can do analyses of literature on any topic. In this lesson we will have a look at the academic literature related to the concept of Terminal Investment. Terminal Investment hypothesis predicts increased investment of resources into reproduction as the chances of survival decrease. This can be observed as increased reproductive effort in older animals or in animals challenged with factors signalling threat to their survival (e.g., predation, pathogens, parasites).

Terminal investment in animals is usually studied in three main ways:

1. via observational studies of correlations of age and reproductive effort,
2. in experimental studies where animals are subject to immune challenges and their subsequent reproductive effort is compared to unchallenged animals of the same age,
3. in experimental studies where reproductive response to immune challenge is compared between animals of older ages versus younger ages.

You can read more on this Wiki page: [https://en.wikipedia.org/wiki/Terminal\\_investment\\_hypothesis](https://en.wikipedia.org/wiki/Terminal_investment_hypothesis)  
[\(https://en.wikipedia.org/wiki/Terminal\\_investment\\_hypothesis\)](https://en.wikipedia.org/wiki/Terminal_investment_hypothesis)

We hope the topic is quite appealing and quite easy to understand. There are several published reviews on terminal investment hypothesis and we can expect many publications related to this topic,

as well as many researchers working on it. Is this so?

Thus, we will try to run bibliometric analyses on the relevant sample of literature. Note that some R packages (and many other online/software tools) are available (and more are being developed) that can perform some of the tasks which we will practice during this exercise, and often much more. For your own project you may want to try to use some of them, but there is no single “perfect” tool that fits all possible analyses and that is easy and usable for all disciplines and types of research questions. Note that the main purpose of this exercise is to familiarize you with the basic principles/issues of bibliometric analyses. You can always learn more in your own time if you are interested.

---

### The search

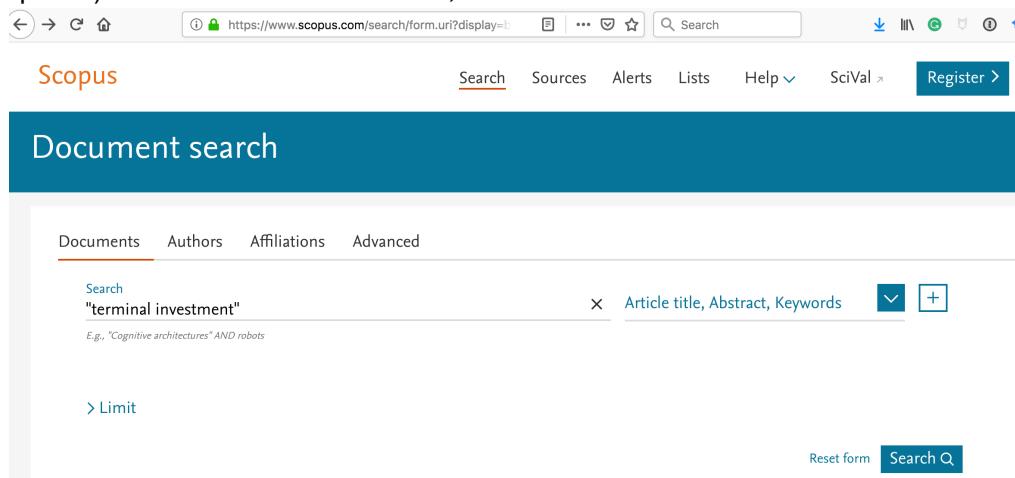
First, we need to find a representative sample of academic publications on our topic of choice. For this, we will use cross-disciplinary database of academic literature, Scopus. This database has the largest coverage of the published literature and should give us the most complete picture.

Note that we have free access to this database on campus, but you will not be able to access it from outside the campus unless (you use UNSW or other university proxy servers). An alternative database, commonly used for broad academic literature searches and analyses, is Web of Science: <https://www.webofknowledge.com/> (<https://www.webofknowledge.com/>).

---

## TASK 1

Go to Scopus search page: <https://www.scopus.com/search/form.uri?display=basic> (<https://www.scopus.com/search/form.uri?display=basic>) and enter “terminal investment” (with the quotes) in the basic search window, as follows:



The screenshot shows the Scopus search interface. At the top, there is a navigation bar with icons for back, forward, search, and user profile. Below the navigation bar is the Scopus logo and a search bar containing the query "terminal investment". The search bar also includes a dropdown menu, a plus sign for adding terms, and a search button. Below the search bar, there is a blue header bar with the text "Document search". Underneath the header, there are tabs for "Documents", "Authors", "Affiliations", and "Advanced". The "Documents" tab is selected. The search query "terminal investment" is entered in the search bar. Below the search bar, there is a note: "E.g., "Cognitive architectures" AND robots". At the bottom of the search form, there are buttons for "Reset form" and "Search".

---

Press “Search” button. You shoud see something like this:

---

## 160 document results

View secondary documents View 71 patent results View 22 Mendeley Data

TITLE-ABS-KEY ("terminal investment")

[Edit](#) [Save](#) [Set alert](#) [Set feed](#)

Search within results...

Analyze search results Show all abstracts Sort on: Date (newest)

Document title	Authors	Year	Source	Cited by
<a href="#">1 Males can evolve lower resistance to sexually transmitted infections to infect their mates and thereby increase their own fitness</a>	Johns, S., Henshaw, J.M., Jennings, M.D., Head, M.L.	2019	Evolutionary Ecology 33(2), pp. 149-172	1
<a href="#">2 Investment strategy of Chinese terminal operators along the "21st-Century Maritime Silk Road"</a>	Wang, L., Zheng, Y., Ducruet, C., Zhang, F.	2019	Sustainability (Switzerland) 11(7),2066	0

View abstract View at Publisher Related documents

View abstract View at Publisher Related documents

Hey, this does not look good... - very few documents were found and some of them are completely unrelated (building shipping terminals).

Why is that?

This is because our search is too simple. It allows us only to find the papers that explicitly mention "terminal investment" phrase in their title, abstract or keywords. To find a better set of papers for the analyses, we need a more sophisticated search string. Additionally, we will focus our topic a little bit more and aim to find papers that use immune challenge approach in wild or semi-wild animal species (so, we try to exclude established lab model species such as mice and rats, domesticated animals such as dogs and pigs, and humans). Finding the best search string is a bit of an art, so we just provide you with this one to save time:

```
(TITLE-ABS-KEY ( ( "terminal investment" OR "reproductive effort" OR "fecundity compensation" OR "reproductive compensation" OR "reproductive fitness" OR "reproductive investment" OR "reproductive success" OR "Life History Trade-Off*" OR "Phenotypic Plasticity" ) AND ( "immune challenge*" OR "immunochallenge*" OR "infect*" OR lipopolysaccharide OR lps OR phytohemagglutinin OR pha OR "sheep red blood cells" OR srbc OR implant OR vaccine* ) ) AND NOT TITLE-ABS-KEY ( load OR human OR people OR men OR women OR infant* OR rat OR rats OR mouse OR mice OR pig* OR pork OR beef OR cattle OR sheep OR lamb* OR chicken* OR calf* OR *horse* ))
```

## TASK 2

You need to copy and paste the above search string into the Advanced Search tab of the Scopus Search page:

## Advanced search

Documents Authors Affiliations Advanced

Search tips 

Enter query string

TITLE-ABS-KEY (( "terminal investment" OR "reproductive effort" OR "fecundity compensation" OR "reproductive compensation" OR "reproductive fitness" OR "reproductive investment" OR "reproductive success" OR "Life History Trade-Off\*" OR "Phenotypic Plasticity") AND ( "immune challeng\*" OR "immunochalleng\*" OR "infect\*" OR lipopolysaccharide OR lps OR phytohemagglutinin OR pha OR "sheep red blood cells" OR srbc OR implant OR vaccin\*)) AND NOT (load OR human OR people OR men OR women OR infant\* OR rat OR rats OR mouse OR mice OR pig\* OR pork OR beef OR cattle OR sheep OR lamb\* OR chicken\* OR calf\* OR \*horse\*))

Outline query Add Author name / Affiliation Clear form Search 

Press “Search”" button. You shoud see somethink like this:

Scopus  Sources Alerts Lists Help SciVal  Register > Login 

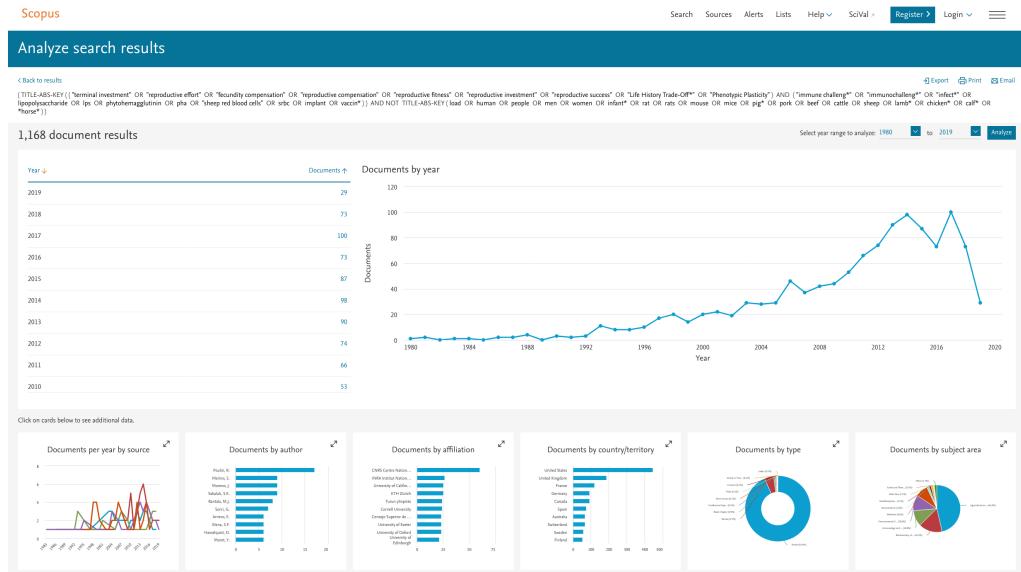
1,167 document results  View 8 patent results

(TITLE-ABS-KEY (( "terminal investment" OR "reproductive effort" OR "fecundity compensation" OR "reproductive compensation" OR "reproductive fitness" OR "reproductive investment" OR "reproductive success" OR "Life History Trade-Off\*" OR "Phenotypic Plasticity") AND ( "immune challeng\*" OR "immunochalleng\*" OR "infect\*" OR lipopolysaccharide OR lps OR phytohemagglutinin OR pha OR "sheep red blood cells" OR srbc OR implant OR vaccin\*)) AND NOT TITLE-ABS-KEY (load OR human OR people OR men OR women OR infant\* OR rat OR rats OR mouse OR mice OR pig\* OR pork OR beef OR cattle OR sheep OR lamb\* OR chicken\* OR calf\* OR \*horse\*))

 Edit  Save  Set alert 

Analyze search results					
Show all abstracts Sort on: Date (newest)					
All	Export	Download	View citation overview	View cited by	Add to List
 1	Mechanisms of network evolution: a focus on socioecological factors, intermediary mechanisms, and selection pressures	Sueur, C., Romano, V., Sosa, S., Puga-Gonzalez, I.	2019	Primates 60(3), pp. 167-181	1
 2	Males can evolve lower resistance to sexually transmitted infections to infect their mates and thereby increase their own fitness	Johns, S., Henshaw, J.M., Jennings, M.D., Head, M.L.	2019	Evolutionary Ecology 33(2), pp. 149-172	1

There are over 1,000 records retrieved from the Scopus database (some look relevant and many are not, but that is always the case). On the left of the results window you can see simple filters: year, most common author names, subject areas, etc. You can explore the whole set roughly by using “Analyze search results” link above the table of the hits:



## **TASK 3**

Next, we will export the bibliometric records for more detailed bibliometric analyses in R. To do so, close the Scopus analyses window to go back to the list of records found. First, select all records by clicking box “All” in the left top of the list of references. Then click the “Export” link to the right.

A pop-up window with the export options will appear.

First, select the format of the export: we will uses .bib file (BibTex format of references, one of the standard ones).

Second, select which fields have to be exported by clicking the boxes on top of each column (or as needed).

For bibliometric analyses on the citations among papers, it is essential to tick the box next to “Include references” (i.e. data on the cited documents).

Note that, unfortunately, Scopus limits number of exported records to 2000. For longer lists of records, you will need to split them in smaller chunks for the export and then merge into a single larger dataset (not covered in this tutorial; WoS export limits are 500 records).

Export document settings [?](#)

You have chosen to export 1167 documents

Select your method of export

 Mendeley   RefWorks  SciVal [?](#)  RIS Format  
 EndNote,  
 Excel  BibTeX  Plain Text  
 Reference Manager [?](#)  ASCII in HTML

---

What information do you want to export?

<input checked="" type="checkbox"/> Citation information	<input checked="" type="checkbox"/> Bibliographical information	<input checked="" type="checkbox"/> Abstract & keywords	<input checked="" type="checkbox"/> Funding details	<input type="checkbox"/> Other information
<input checked="" type="checkbox"/> Author(s)	<input checked="" type="checkbox"/> Affiliations	<input checked="" type="checkbox"/> Abstract	<input checked="" type="checkbox"/> Number	<input type="checkbox"/> Tradenames & manufacturers
<input checked="" type="checkbox"/> Document title	<input checked="" type="checkbox"/> Serial identifiers (e.g. ISSN)	<input checked="" type="checkbox"/> Author keywords	<input checked="" type="checkbox"/> Acronym	<input type="checkbox"/> Accession numbers & chemicals
<input checked="" type="checkbox"/> Year	<input checked="" type="checkbox"/> PubMed ID	<input checked="" type="checkbox"/> Index keywords	<input checked="" type="checkbox"/> Sponsor	<input type="checkbox"/> Conference information
<input checked="" type="checkbox"/> Source title	<input checked="" type="checkbox"/> Publisher	<input checked="" type="checkbox"/> Funding text	<input checked="" type="checkbox"/> Include references	
<input checked="" type="checkbox"/> volume, issue, pages	<input checked="" type="checkbox"/> Editor(s)			
<input checked="" type="checkbox"/> Citation count	<input checked="" type="checkbox"/> Language of original document			
<input checked="" type="checkbox"/> Source & document type	<input checked="" type="checkbox"/> Correspondence address			
<input checked="" type="checkbox"/> DOI	<input checked="" type="checkbox"/> Abbreviated source title			

Click “Export” button. A file named “Scopus” (with extension matching your export type file, e.g., bib) will be saved to your downloads folder.

Note that when you export references with their reference lists included in the records, the resulting files are quite large (in our case around 16Mb).

In case you did not succeed expoerting the files (or wish to work with exactly the same ones we used, or you cannot acces Scopus), the files downloaded on 27/05/2019 are provided (the standard way os to store them in a “/data” subdirectory).

---

## TASK 4

Create a new Rmarkdown file to save your code (you can do this within new RStudio project). Install and upload *bibliometrix* R package:

```
install.packages("bibliometrix", dependencies=TRUE) ### installs bibliometrix package and dependencies
library(bibliometrix) #uploads the package
# Note: output not displayed for this chunk
```

Upload the file exported from Scopus (you can use one provided) into RStudio (note that the file path you may need to use on your computer may be different, e.g., “H:/Users/z1234567/Downloads /scopus.bib”).

Then, convert the data from that file into internal *bibliometrix* format.

```

tmp <- readFiles("data/scopus.bib")
bib <- convert2df(tmp, dbsource = "scopus", format = "bibtex") # Convert to a
bibliometric data frame
#>
#> Converting your scopus collection into a bibliographic dataframe
#>
#> Articles extracted 100
#> Articles extracted 200
#> Articles extracted 300
#> Articles extracted 400
#> Articles extracted 500
#> Articles extracted 600
#> Articles extracted 700
#> Articles extracted 800
#> Articles extracted 900
#> Articles extracted 1000
#> Articles extracted 1100
#> Articles extracted 1167
#> Done!
#>
#>
#> Generating affiliation field tag AU_UN from C1: Done!
names(bib)
#> [1] "AU"        "TI"        "SO"        "JI"        "AB"        "DE"
#> [7] "ID"        "LA"        "DT"        "DT2"       "TC"        "CR"
#> [13] "C1"       "DI"        "AR"        "RP"        "BE"        "FU"
#> [19] "BN"       "SN"        "PN"        "PP"        "PU"        "PM"
#> [25] "DB"       "VL"        "PY"        "AU_UN"     "AU1_UN"    "AU_UN_NR"
#> [31] "SR_FULL"  "SR"
#write.csv(bib, "data/bib_as_df.csv", row.names = FALSE) #if you want to save
this data frame as a csv file

```

After some processing, an object called “bib” is created. It contains a data frame with each row corresponding to one exported publication from Scopus and with each column corresponding to a field exported from Scopus online database. (Note, if you tried to achieve this by exporting a csv file directly from Scopus, you would likely get a messy data frame, due to missing field values shifting the cells between columns).

## TASK 5

What are the contents of the columns of our “bib” data frame? Columns are labelled with a two-letter tags: AU, TI, SO, JI, AB, DE, ID, LA, DT, DT2, TC, CR, C1, DI, AR, RP, BE, FU, BN, SN, PN, PP, PU, PM, DB, VL, PY, AU\_UN, AU1\_UN, AU\_UN\_NR, SR\_FULL, SR.

For a complete list and descriptions of field tags used in *bibliometrix* you can have a look at this file: [http://www.bibliometrix.org/documents/Field\\_Tags\\_bibliometrix.pdf](http://www.bibliometrix.org/documents/Field_Tags_bibliometrix.pdf) ([http://www.bibliometrix.org/documents/Field\\_Tags\\_bibliometrix.pdf](http://www.bibliometrix.org/documents/Field_Tags_bibliometrix.pdf))

Our data frame contains just a subset of these codes. Which ones?

Note that column **bib\$AU** contains authors of each paper (as surnames and initials) separated by semicolon (;). We can easily split these strings and can extract a list of all author names to a vector:

```

# head(bib$AU) #have a look at the few few records on your screen
authors <- bib$AU
authors <- unlist(strsplit(authors, ";")) #split the records into individual authors
authors <- authors[order(authors)] #order alphabetically
head(authors) #have a look again
#> [1] "ABBOTT J"          "ABE A"           "ABEDON ST"        "ABO SHEHADA M"
#> [5] "ABOUL SOUD MAM"   "ABRANTES N"
# View(unique(authors)) #use to see all the values
# write.csv(authors, "data/author_list_uncleaned.csv", row.names = FALSE) #if you want to save this data frame as a csv file

```

## TASK 6

Cited references for each inculded paper are in the CR column of the “bib” data frame. They are in a single string, also seperated by semicolon (;). We can have a look at them and check whether familiar names were cited, e.g.:

```

dim(bib) #dimensions of the data frame
#> [1] 1167 32
names(bib) #names of the columns of the data frame
#> [1] "AU"      "TI"      "SO"      "JI"      "AB"      "DE"
#> [7] "ID"      "LA"      "DT"      "DT2"     "TC"      "CR"
#> [13] "C1"     "DI"      "AR"      "RP"      "BE"      "FU"
#> [19] "BN"     "SN"      "PN"      "PP"      "PU"      "PM"
#> [25] "DB"     "VL"      "PY"      "AU_UN"   "AU1_UN"  "AU_UN_NR"
#> [31] "SR_FULL" "SR"
#bib$CR[1] #display a list of cited references for the first paper in the data frame
#(we are not displaying it in this doucment as it is a very long string! - examine it on your screen instead)
#look whether some of these names are cited:
grep("NAKAGAWA, S.", bib$CR)
#> [1] 2 6 7 20 33 36 37 56 72 75 102 109 121 145 152 166 207
#> [18] 222 249 285 293 312 330 361 362 368 370 401 440 455 471 475 489 501
#> [35] 512 560 562 573 590 620 655 690 713 730 770
grep("CORNWELL, W.", bib$CR)
#> [1] 15
bib[grep("CORNWELL, W.", bib$CR), c(1:3)] #who is citing?
#>
AU
#> MULETZ-WOLZ CR, 2019, J EVOL BIOL MULETZ WOLZ CR;BARNETT SE;DIRENZO GV;ZAMUDIO KR;TOLEDO LF;JAMES TY;LIPS KR
#>
TI
#> MULETZ-WOLZ CR, 2019, J EVOL BIOL DIVERSE GENOTYPES OF THE AMPHIBIAN-KILLING FUNGUS PRODUCE DISTINCT PHENOTYPES THROUGH PLASTIC RESPONSES TO TEMPERATURE
#> SO
#> MULETZ-WOLZ CR, 2019, J EVOL BIOL JOURNAL OF EVOLUTIONARY BIOLOGY

```

# Summarising bibliometric data

## TASK 7

Luckily, *bibliometrix* package has a handy function that summarises the information contained in the “bib” data frame, so we can get some quick facts about our set of papers.

**Note:** this and the following tasks require quite a bit of computational power, they may be slow or even halt on your computer.

In such case, for this exercise, make your data frame smaller by subsetting it, e.g.:

“`bib <- bib[1:500, ] #taking first 500 records`”. However, the results and plots you will produce with a subsetted data frame will differ from the ones presented in this document.

```

# Preliminary descriptive analyses
results <- biblioAnalysis(bib, sep = ";")
summary(object = results, k = 10, pause = TRUE)
#>
#>
#> Main Information about data
#>
#> Documents 1167
#> Sources (Journals, Books, etc.) 380
#> Keywords Plus (ID) 6388
#> Author's Keywords (DE) 3332
#> Period 1980 - 2019
#> Average citations per documents 27
#>
#> Authors 3918
#> Author Appearances 4728
#> Authors of single-authored documents 80
#> Authors of multi-authored documents 3838
#> Single-authored documents 84
#>
#> Documents per Author 0.298
#> Authors per Document 3.36
#> Co-Authors per Documents 4.05
#> Collaboration Index 3.54
#>
#> Document types
#> ARTICLE 1090
#> ARTICLE IN PRESS 1
#> BOOK CHAPTER 10
#> CONFERENCE PAPER 6
#> ERRATUM 1
#> LETTER 1
#> NOTE 2
#> REVIEW 53
#> SHORT SURVEY 3
#>
#> Hit <Return> to see next table:
#>
#> Annual Scientific Production
#>
#> Year Articles
#> 1980 1
#> 1981 2
#> 1983 1
#> 1984 1
#> 1986 2
#> 1987 2
#> 1988 4
#> 1990 3
#> 1991 2
#> 1992 3
#> 1993 11

```

```

#>    1994      8
#>    1995      8
#>    1996     10
#>    1997     17
#>    1998     20
#>    1999     14
#>    2000     20
#>    2001     22
#>    2002     19
#>    2003     29
#>    2004     28
#>    2005     29
#>    2006     46
#>    2007     37
#>    2008     42
#>    2009     44
#>    2010     53
#>    2011     66
#>    2012     74
#>    2013     90
#>    2014     98
#>    2015     87
#>    2016     73
#>    2017    100
#>    2018     73
#>    2019     28
#>
#> Annual Percentage Growth Rate 9.698031
#>
#> Hit <Return> to see next table:
#>
#> Most Productive Authors
#>
#> Authors          Articles Authors          Articles Fractionalized
#> 1   POULIN R           17   POULIN R            8.25
#> 2   MERINO S            9   ELENA SF            3.03
#> 3   MORENO J            9   HURD H             2.92
#> 4   SAKALUK SK           9   BENESH DP            2.83
#> 5   RANTALA MJ            8   MORET Y             2.62
#> 6   JOKELA J              7   ROY BA             2.58
#> 7   SORCI G              7   TSENG M             2.50
#> 8   ARRIERO E             6   WEBSTER JP            2.50
#> 9   ELENA SF              6   KOELLA JC            2.33
#> 10  HASSELQUIST D            6   TURNER PE            2.28
#>
#> Hit <Return> to see next table:
#>
#> Top manuscripts per citations
#>
#> Paper
TC TCperYear
#> 1 FOLSTAD I, 1992, AMERICAN NATURALIST

```

```

1827      67.7
#> 2  SCHULZ B, 2005, MYCOL RES
719      51.4
#> 3  SCHRECK CB, 2001, AQUACULTURE
356      19.8
#> 4  BONNEAUD C, 2003, AM NAT
345      21.6
#> 5  NORDLING D, 1998, PROC R SOC B BIOL SCI
306      14.6
#> 6  GUSTAFSSON L, 1994, PHILOSOPHICAL TRANSACTIONS - ROYAL SOCIETY OF LONDON
N, B  300      12.0
#> 7  OTS I, 1998, FUNCT ECOL
286      13.6
#> 8  GARCIA DE LEANIZ C, 2007, BIOL REV
252      21.0
#> 9  SPRENT JI, 2007, NEW PHYTOL
244      20.3
#> 10 LOVE OP, 2005, AM NAT
225      16.1
#>
#> Hit <Return> to see next table:
#>
#> Corresponding Author's Countries
#>
#>          Country Articles   Freq SCP MCP MCP_Ratio
#> 1  USA           275 0.2935 216  59    0.215
#> 2  UNITED KINGDOM 101 0.1078 64   37    0.366
#> 3  FRANCE        73 0.0779 52   21    0.288
#> 4  CANADA         52 0.0555 38   14    0.269
#> 5  GERMANY        52 0.0555 28   24    0.462
#> 6  SPAIN          50 0.0534 28   22    0.440
#> 7  FINLAND        35 0.0374 19   16    0.457
#> 8  SWEDEN          29 0.0309 17   12    0.414
#> 9  SWITZERLAND    28 0.0299 16   12    0.429
#> 10 AUSTRALIA      20 0.0213 16   4     0.200
#>
#>
#> SCP: Single Country Publications
#>
#> MCP: Multiple Country Publications
#>
#> Hit <Return> to see next table:
#>
#> Total Citations per Country
#>
#>          Country       Total Citations Average Article Citations
#> 1  USA             8121                  29.53
#> 2  UNITED KINGDOM 4566                  45.21
#> 3  FRANCE          2296                  31.45
#> 4  NORWAY          2180                  155.71
#> 5  GERMANY          1831                  35.21
#> 6  SWEDEN          1653                  57.00

```

```

#> 7 CANADA 1402 26.96
#> 8 SPAIN 1123 22.46
#> 9 FINLAND 1078 30.80
#> 10 SWITZERLAND 1031 36.82
#>
#> Hit <Return> to see next table:
#>
#> Most Relevant Sources
#>
#> Sources Articles
#> 1 JOURNAL OF EVOLUTIONARY BIOLOGY 41
#> 2 PROCEEDINGS OF THE ROYAL SOCIETY B: BIOLOGICAL SCIENCES 40
#> 3 EVOLUTION 35
#> 4 PARASITOLOGY 35
#> 5 OECOLOGIA 32
#> 6 PLOS ONE 30
#> 7 AMERICAN NATURALIST 25
#> 8 FUNCTIONAL ECOLOGY 24
#> 9 BMC EVOLUTIONARY BIOLOGY 22
#> 10 BEHAVIORAL ECOLOGY AND SOCIOBIOLOGY 20
#>
#> Hit <Return> to see next table:
#>
#> Most Relevant Keywords
#>
#> Author Keywords (DE) Articles Keywords-Plus (ID) Articles
#> 1 PHENOTYPIC PLASTICITY 68 FEMALE 469
#> 2 REPRODUCTION 45 ANIMALS 461
#> 3 FITNESS 41 ANIMAL 445
#> 4 REPRODUCTIVE SUCCESS 39 ARTICLE 445
#> 5 LIFE HISTORY 36 MALE 407
#> 6 IMMUNITY 31 REPRODUCTION 365
#> 7 TRADE OFF 31 PHYSIOLOGY 316
#> 8 PARASITE 28 NONHUMAN 285
#> 9 LIFE HISTORY TRADE OFFS 27 REPRODUCTIVE SUCCESS 259
#> 10 VIRULENCE 27 HOST PARASITE INTERACTION 239

```

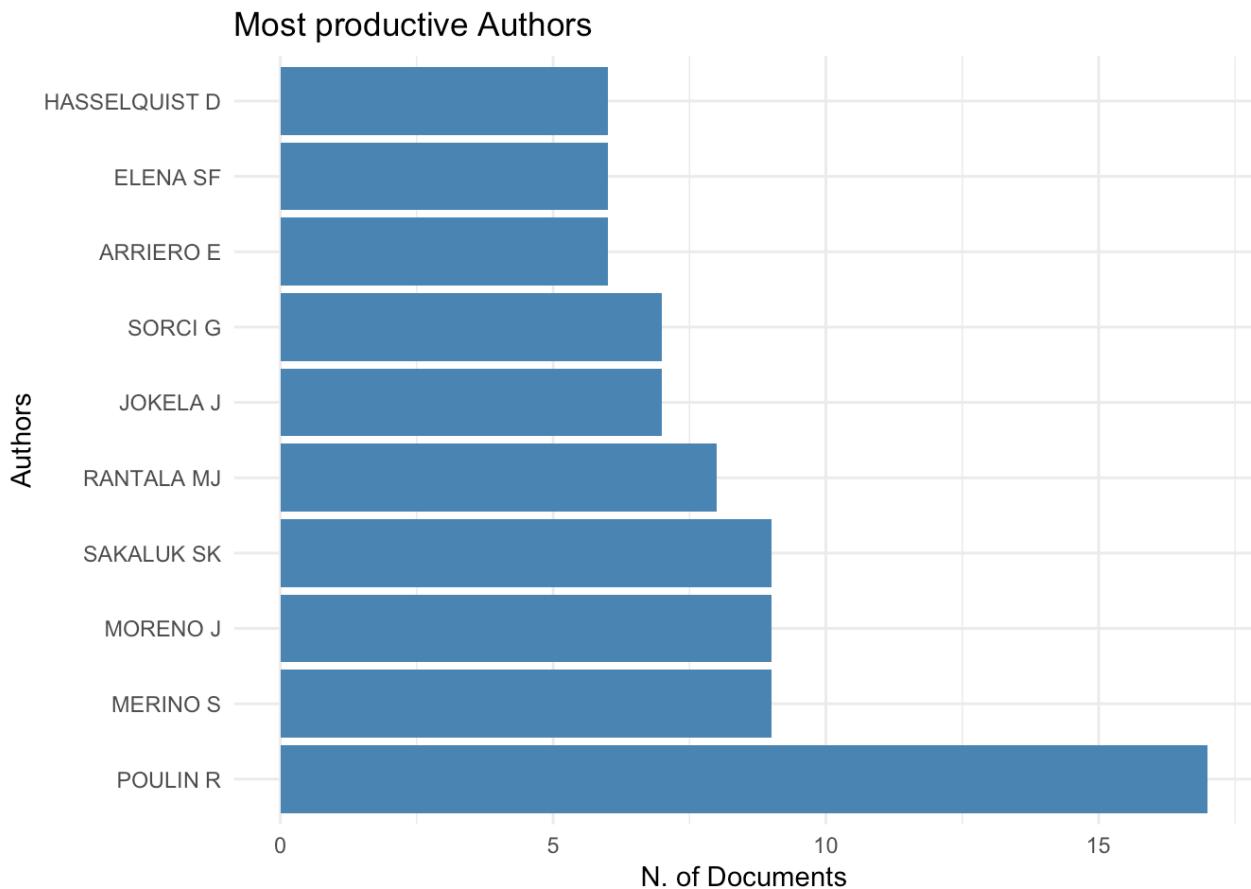
Using summary function on *bibliometrix* results, we can get several screens with various tables summarising bibliometric data from our data frame - how many documents, journals, keywords, authors, publications timespan, collaboration index, annual publication growth rate, most prolific authors, publications per country, per journal, per keywords, etc.

You can automatically plot some of these tables (hit “return”" to display next graph, and later you can use arrows in the top left of the plots pane to move back and forth between consecutive graphs saved in the RStudio memory):

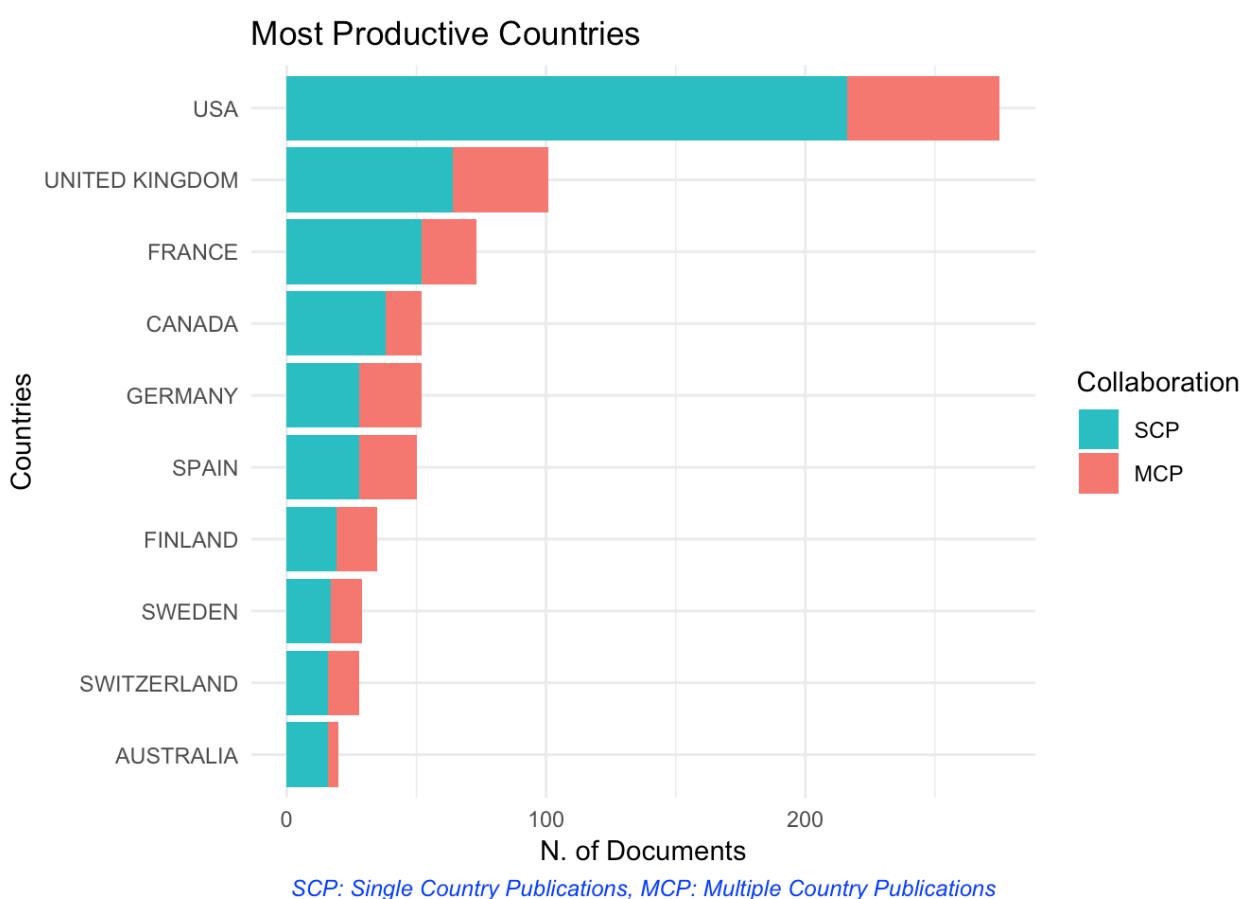
```

plot(results, k = 10, pause=TRUE) #this takes top 10 values from each plottable table
#> Hit <Return> to see next plot:

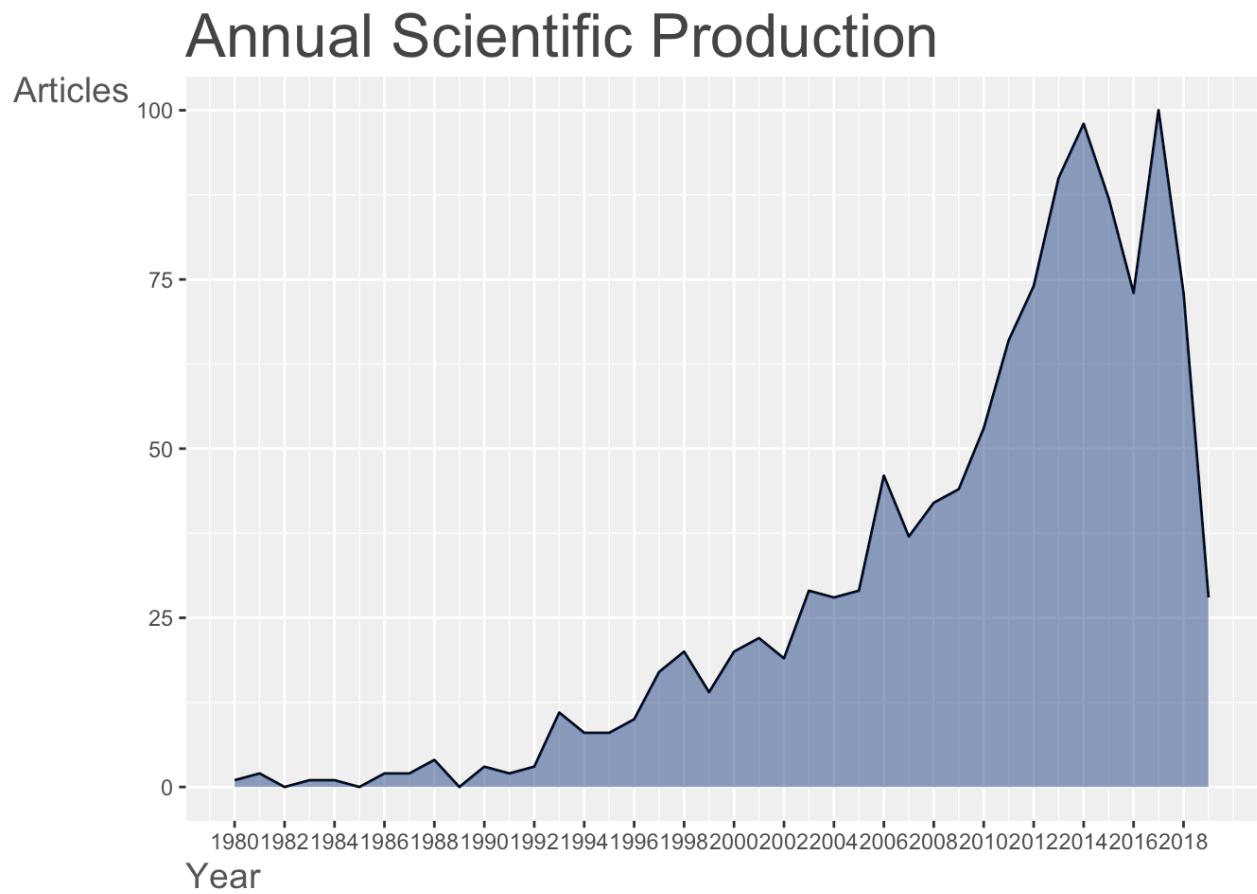
```



```
#> Hit <Return> to see next plot:
```

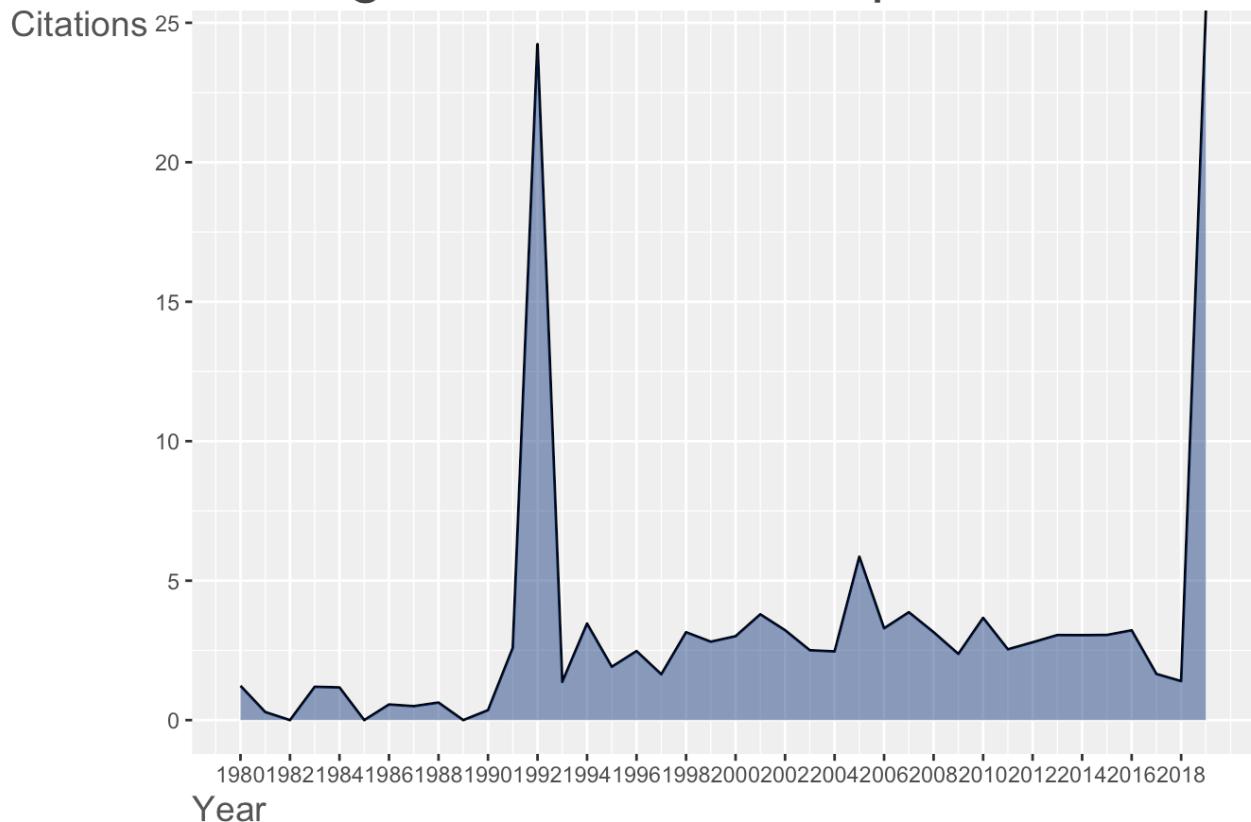


```
#> Hit <Return> to see next plot:
```



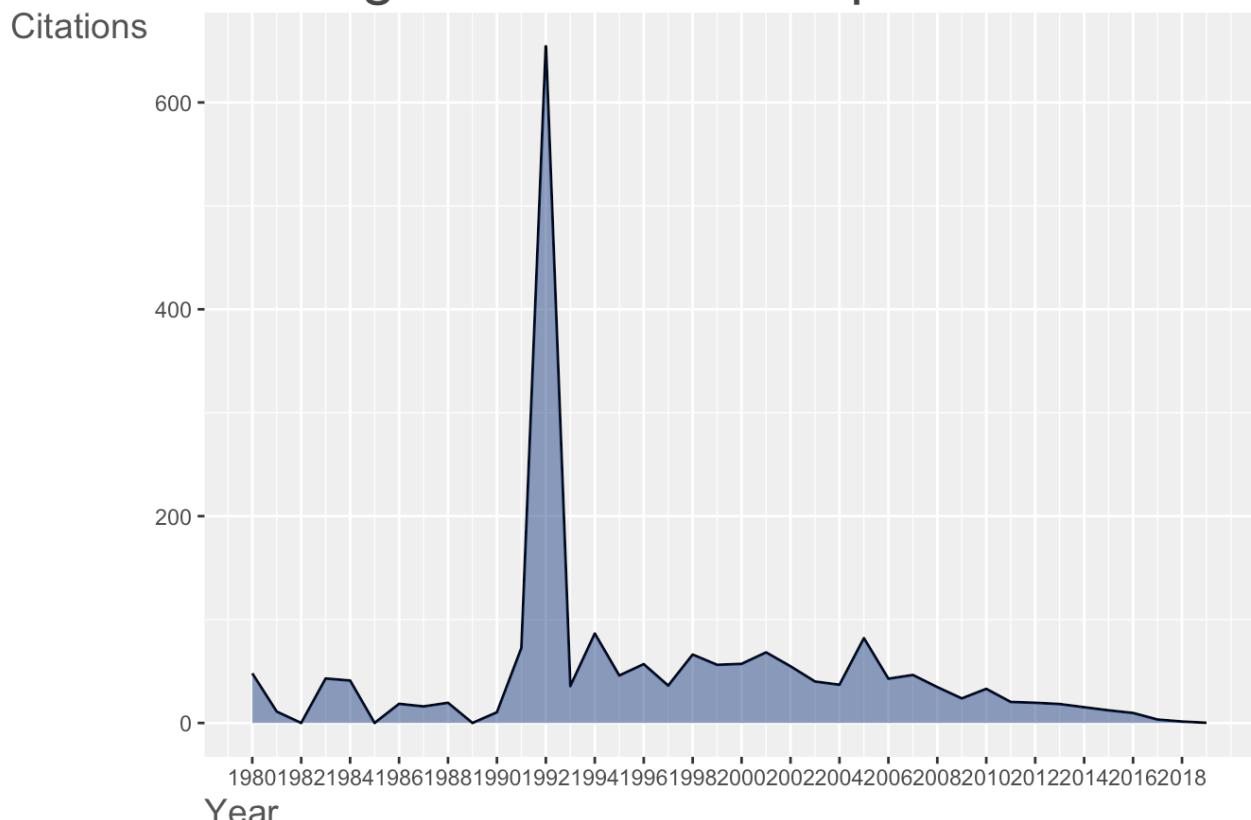
```
#> Hit <Return> to see next plot:
```

## Average Article Citations per Year



```
#> Hit <Return> to see next plot:
```

## Average Total Citations per Year



```
#the code below is for saving these plots into a pdf:
# pdf(file = "plots/bib_descriptive_plots.pdf", height = 8, width = 8, pointsize=10) #
# plot(results, k = 20, pause=FALSE) #this takes top 20 values from each plottable table
# dev.off()
```

## TASK 8

The cited papers from the CR field of the data frame can be analysed using function *citations*. Function *citations* makes it easy to generate the frequency tables of the most cited papers or the most cited first authors from the reference lists of our papers downloaded from Scopus.

Ten most cited papers:

```
mostcitedP <- citations(bib, field = "article", sep = ";")
cbind(mostcitedP$Cited[1:10])
#>
[,1]
#> LOCHMILLER, R.L., DEERENBERG, C., TRADE-OFFS IN EVOLUTIONARY IMMUNOLOGY: JUST WHAT IS THE COST OF IMMUNITY? (2000) OIKOS, 88, PP. 87-98
45
#> HAMILTON, W.D., ZUK, M., HERITABLE TRUE FITNESS AND BRIGHT BIRDS: A ROLE FOR PARASITES? (1982) SCIENCE, 218, PP. 384-387
31
#> MORET, Y., SCHMID-HEMPEL, P., SURVIVAL FOR IMMUNITY: THE PRICE OF IMMUNE STEM ACTIVATION FOR BUMBLEBEE WORKERS (2000) SCIENCE, 290, PP. 1166-1168
30
#> FORBES, M.R.L., PARASITISM AND HOST REPRODUCTIVE EFFORT (1993) OIKOS, 67, P. 444-450
29
#> STEARNS, S.C., (1992) THE EVOLUTION OF LIFE HISTORIES, , OXFORD UNIVERSITY PRESS, OXFORD
25
#> MINCHELLA, D.J., HOST LIFE-HISTORY VARIATION IN RESPONSE TO PARASITISM (1985) PARASITOLOGY, 90, PP. 205-216
24
#> SHELDON, B.C., VERHULST, S., ECOLOGICAL IMMUNOLOGY: COSTLY PARASITE DEFENCE AND TRADE-OFFS IN EVOLUTIONARY ECOLOGY (1996) TRENDS ECOL. EVOL., 11, PP. 317-321 19
#> ROLFF, J., SIVA-JOTHY, M.T., INVERTEBRATE ECOLOGICAL IMMUNOLOGY (2003) SCIENCE, 301, PP. 472-475
18
#> ANDERSON, R.M., MAY, R.M., COEVOLUTION OF HOSTS AND PARASITES (1982) PARASITOLOGY, 85, PP. 411-426
17
#> FRANK, S.A., MODELS OF PARASITE VIRULENCE (1996) Q. REV. BIOL., 71, PP. 37-78
17
```

Ten most cited authors:

```
mostcitedA <- citations(bib, field = "author", sep = ";")
cbind(mostcitedA$Cited[1:10])
#> [,1]
#> WINGFIELD J C      425
#> POULIN R            410
#> MLLER A P           394
#> SCHMID HEMPEL P     329
#> HASSELQUIST D       326
#> READ A F             306
#> ZUK M                281
#> EBERT D               275
#> SHELDON B C          263
#> BENNSCH S             229
```

The function *localCitations* generates the frequency table of the **locally** most cited authors. Locally, means that only citations are counted only within the given data set - i.e. how many times an author/paper that is in this data set has been cited by other authors/papers also in the data set.

Ten most frequent local cited authors and papers:

```
mostcitedLA <- localCitations(bib, results, sep = ";")
#> Articles analysed    100
#> Articles analysed    200
#> Articles analysed    300
#> Articles analysed    400
#> Articles analysed    500
#> Articles analysed    600
#> Articles analysed    700
#> Articles analysed    800
#> Error in grep(y, M$CR[M$PY >= Year]): invalid regular expression '\(2015\)'
#> CHANGES IN PHYTOHAEMAGGLUTININ SKIN-SWELLING RESPONSES DURING THE BREEDING SEASON IN A MULTI-BROODED SPECIES, THE EURASIAN TREE SPARROW: DO MALES WITH HIGHER TESTOSTERONE LEVELS SHOW STRONGER IMMUNE RESPONSES? [UNTERSCHIEDLICHE IMMUNANTWORTEN ANHAND PHYTOHAEMAGGLUTININ-HAUTSCHWELLUNG BEI FELDSPERLINGEN WENDRER BRUTZEIT: ZEIGEN MÄNNCHEN MIT HÖHEREN TESTOSTERONWERTEN STRKERE IMMUNANTWORTE N?], reason 'Invalid character range'
mostcitedLA[1:10]
#> Error in eval(expr, envir, enclos): object 'mostcitedLA' not found
```

## Creating bibliometric networks

So far, we looked only at the numbers - who or what gets cited most, either from the main papers list or from the lists of the references within these papers. Now it is time to look at the actual networks of citations and also other types of networks that can be created using our data set.

To do so we will be creating various rectangular matrices which reflect connections of different attributes of Papers/Authors. These matrices than can be plotted as bipartite networks and analysesd.

Co-citation or coupling networks are a special type of networks resulting from scientific papers containing references to other scientific papers.

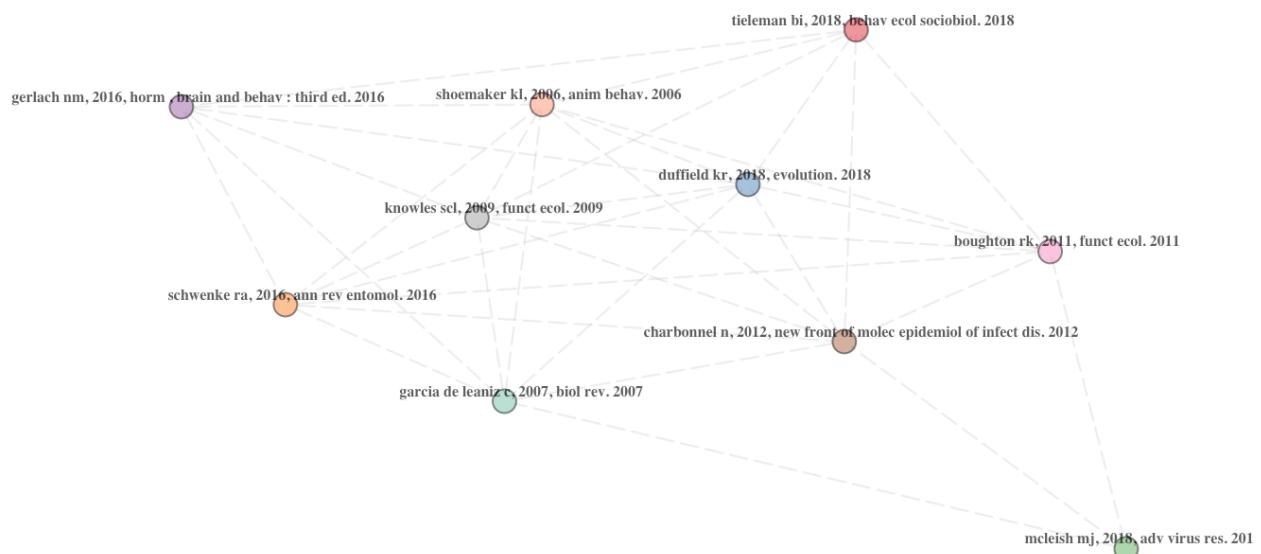
Package *bibliometrix* contains function *biblioNetwork* which makes creating bibliographic networks easy. This function can create the most frequently used coupling networks: Authors, Sources, and Countries.

## TASK 9

**Bibliographic coupling** - two articles are bibliographically coupled if they share at least one reference from their reference lists (i.e. at least one cited source appears in the reference lists/bibliographies of both papers (Kessler, 1963).

```
NetMatrix <- biblioNetwork(bib, analysis = "coupling", network = "references",
sep = ";")
net = networkPlot(NetMatrix, weighted = NULL, n = 10, Title = "Papers' bibliographic coupling",
type = "fruchterman", size = 5, remove.multiple = TRUE, labelsize = 0.5)
```

### Papers' bibliographic coupling



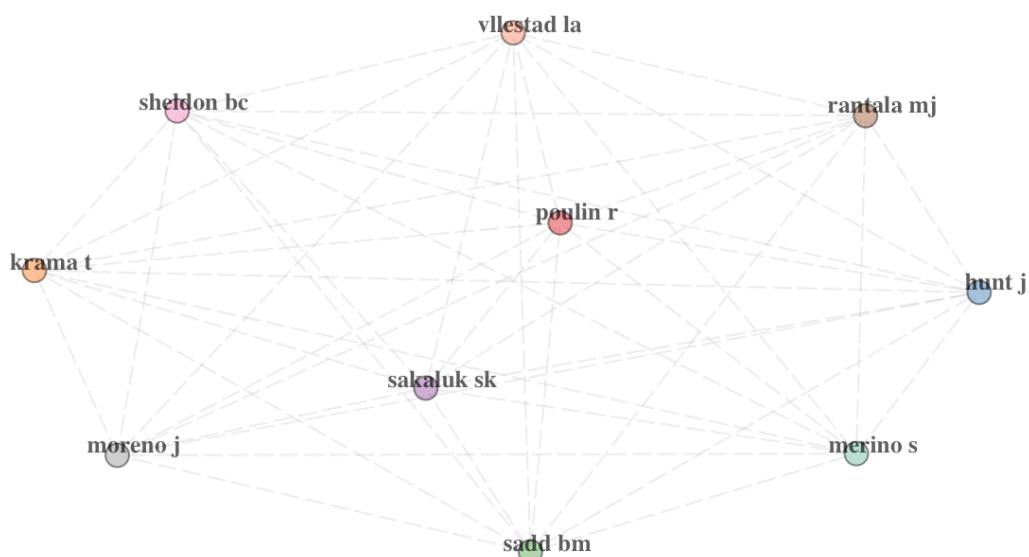
Above, we plotted only the top 10 most coupled papers (n=10), try increasing this number to 100 (would not recommend further increasing the number of displayed nodes - it gets slow and messy). What happens and why?

## TASK 10

**Author's bibliographic coupling** - two authors are bibliographically coupled if they share at least one reference from their reference lists.

```
NetMatrix <- biblioNetwork(bib, analysis = "coupling", network = "authors", sep = ";")
net = networkPlot(NetMatrix, weighted = NULL, n = 10, Title = "Authors' bibliographic coupling", type = "fruchterman", size = 5, remove.multiple = TRUE, labelsizes = 0.8)
```

### Authors' bibliographic coupling



Above, we plotted only the top 10 most coupled authors ( $n=10$ ), try increasing this number to 100 (would not recommend further increasing the number of displayed nodes to  $>50$  - it gets slow and messy).

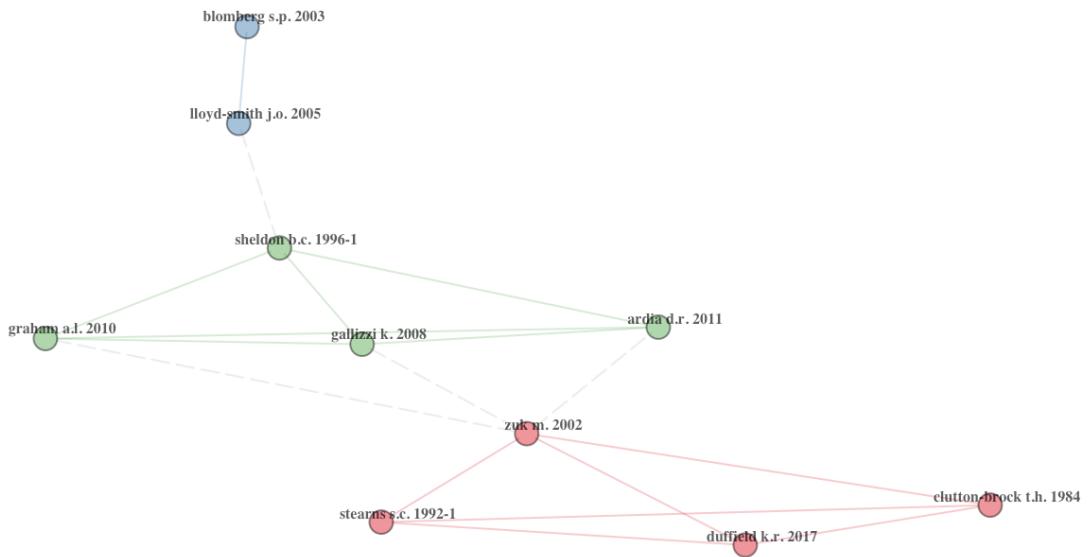
What happens and why?

## TASK 11

**Bibliographic co-citation** is kind of opposite to bibliographic coupling, in so that two papers are linked by co-citation when both are cited in a third paper.

```
NetMatrix <- biblioNetwork(bib[1:50,], analysis = "co-citation", network = "references", sep = ";")
net = networkPlot(NetMatrix, weighted=NULL, n = 10, Title = "Papers' co-citations", type = "fruchterman", size = 5, remove.multiple = TRUE, labelsize = 0.5)
```

## Papers' co-citations



Note that for creating this matrix we only used first 50 papers from our data set - this is because the resulting matrix is a matrix of ALL cited papers and it gets HUGE). Also, we plotted only the top 10 most coupled papers ( $n=10$ ), try increasing this number to 20 (would not recommend further increasing the number of displayed nodes to  $>50$  - it gets slow and messy).

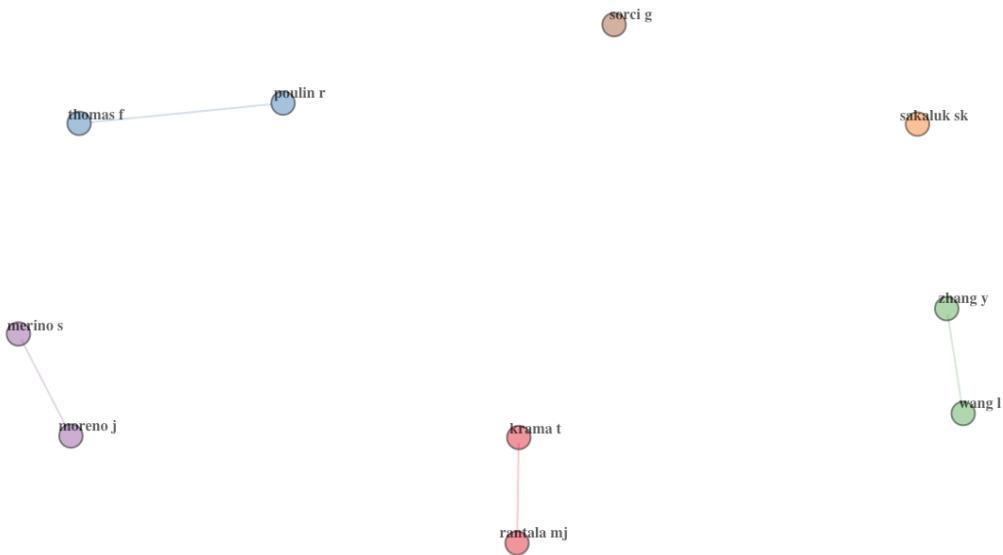
What happens and why?

## TASK 12

**Bibliographic collaboration** is a network where nodes are authors and links are co-authorships on the papers.

```
NetMatrix <- biblioNetwork(bib, analysis = "collaboration", network = "author
s", sep = ";")
net = networkPlot(NetMatrix, weighted = NULL, n = 10, Title = "Authors' collab
orations", type = "fruchterman", size = 5, remove.multiple = TRUE, labelsizes =
0.5)
```

## Authors' collaborations



Above, we plotted only the top 10 most collaborating authors (n=10), try increasing this number to 100 (would not recommend further increasing the number of displayed nodes - it gets slow and messy).

What happens and why?

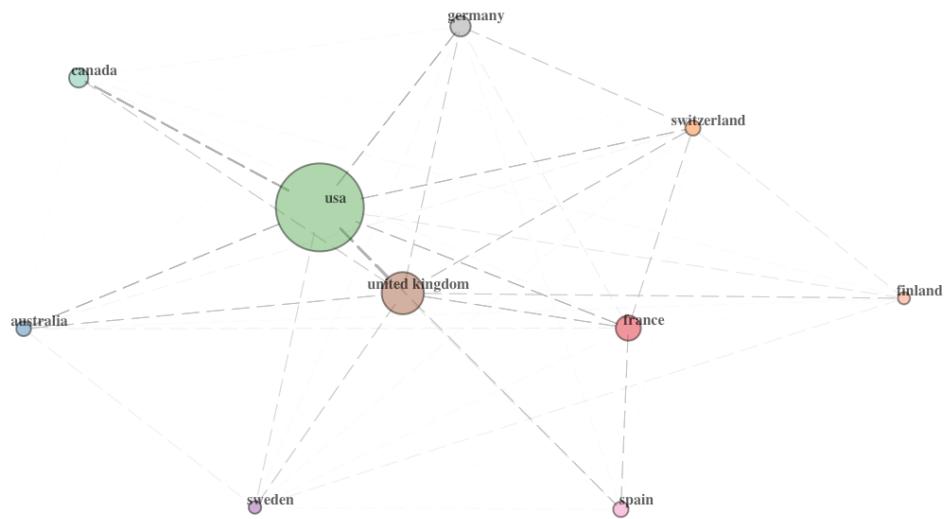
---

## TASK 13

**Country Scientific Collaboration** - we can visualise authors from which countries publish papers together most frequently.

```
bib <- metaTagExtraction(bib, Field = "AU_CO", sep = ";") #we need to extract  
countries from the affiliations first  
NetMatrix <- biblioNetwork(bib, analysis = "collaboration", network = "countries", sep = ";")  
net = networkPlot(NetMatrix, n = 10, Title = "Country Collaboration", type = "auto", size = TRUE, remove.multiple = FALSE, labelszie = 0.5)
```

## Country Collaboration



Above, we plotted only the top 10 most collaborating countries ( $n=10$ ), try increasing this number to 50 (would not recommend further increasing the number of displayed nodes to  $>100$  - it gets slow and messy).

What happens and why?

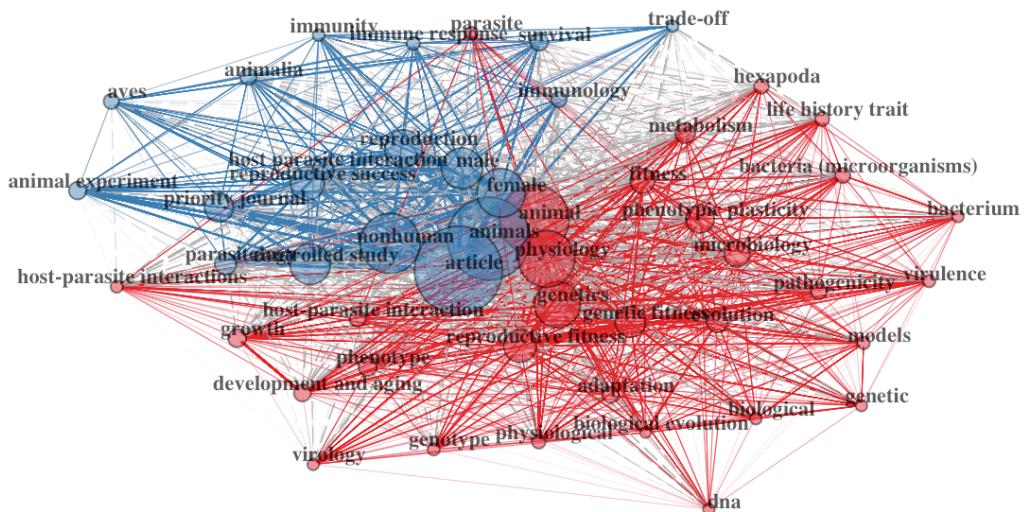
---

## TASK 14

**Keyword co-occurrences** - we can also visualise which papers share most keywords (from Scopus database).

```
NetMatrix <- biblioNetwork(bib, analysis = "co-occurrences", network = "keywords", sep = ";")
net = networkPlot(NetMatrix, n = 50, Title = "Keyword co-occurrence", type = "fruchterman", size = T, remove.multiple = FALSE, labelsize = 0.7, edgesize = 5)
```

## Keyword co-occurrence



Try replacing network = "keywords" with network = "author\_keywords" and see what happens. You can also try to display fewer/more keywords in the plot.

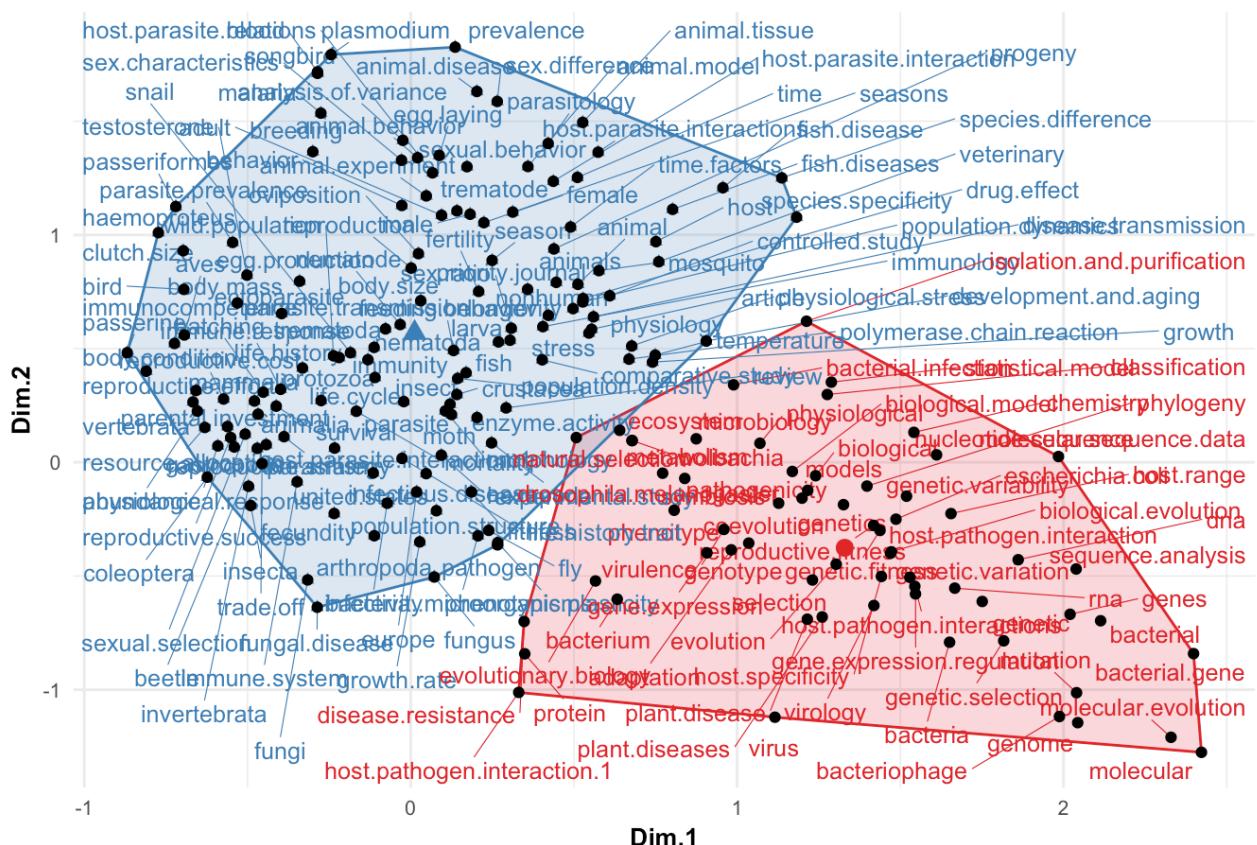
## TASK 15

**Note:** you may want to skip this step on a big data set or a slow computer.

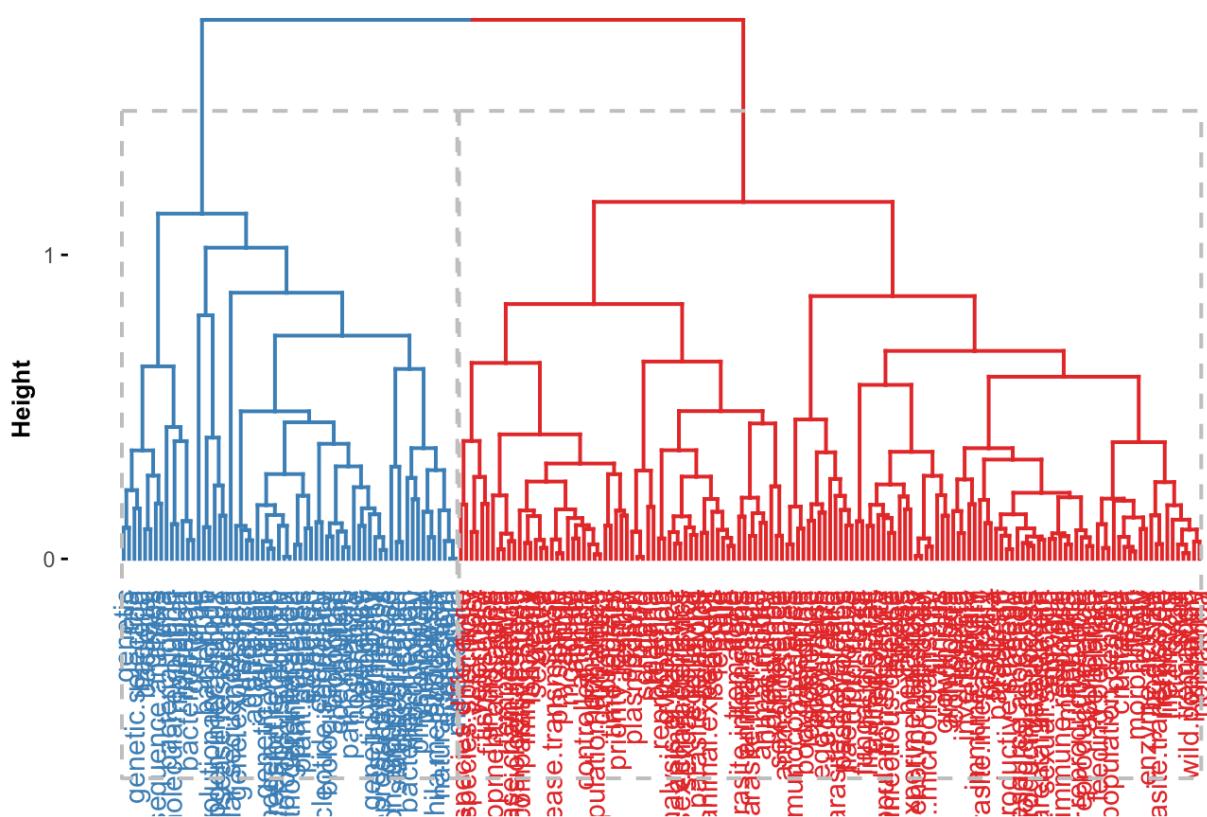
**Co-Word Analysis** - uses the word co-occurrences in a bibliographic collection to map the conceptual structure of research. It works via a separate function *conceptualStructure* that creates a conceptual structure map of a scientific field performing Correspondence Analysis (CA), Multiple Correspondence Analysis (MCA) or Metric Multidimensional Scaling (MDS) and Clustering of a bipartite network of terms extracted from keyword, title or abstract fields of the data frame.

```
CS <- conceptualStructure(bib, field = "ID", minDegree = 20, k.max = 5, stemming = FALSE, labelsize = 10)
```

## **Conceptual Structure Map - method: MCA**



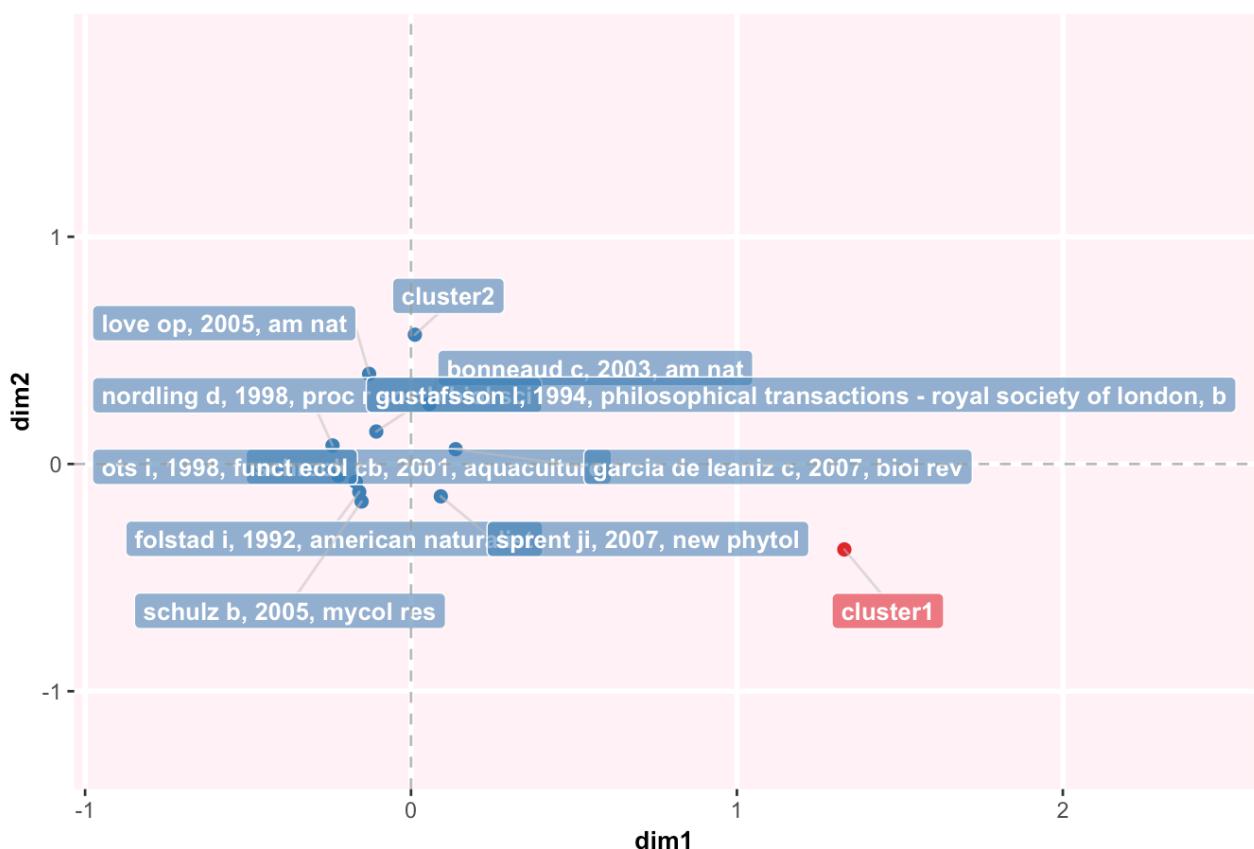
## Topic Dendrogram



### Factorial map of the documents with the highest contributes



### Factorial map of the most cited documents



The code above uses field ID, which stands for “conceptualStructure”. You could try using authors keywords, “DE” field, instead.

Is the map different?

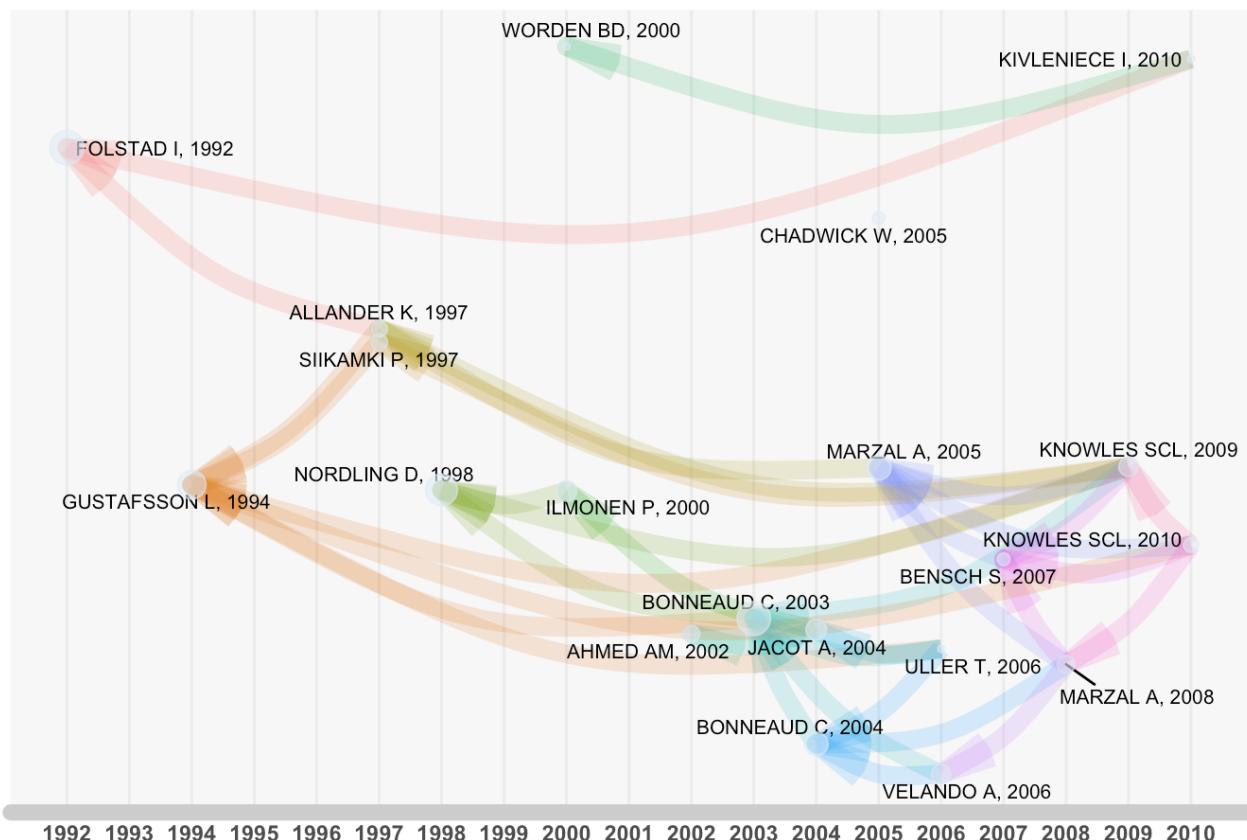
## TASK 15

**Note:** you may want to skip this step on a big data set or a slow computer.

**Historical Direct Citation Network** - represents a chronological network map of most relevant direct citations in a bibliographic collection, i.e who is citing whom and in what order. *histNetwork* function calculates a chronological direct citation network matrix which then is plotted using *histPlot*:

```
#options(width = 130)
histResults <- histNetwork(bib, min.citations = 10, sep = ";")
#> Articles analysed 100
#> Articles analysed 200
#> Articles analysed 300
#> Articles analysed 400
#> Articles analysed 500
#> Articles analysed 600
#> Articles analysed 656
net = histPlot(histResults, labelsize = 2, arrowsize = 0.5)
```

Historical Direct Citation Network



```

#>
#> Legend
#>
#>
Paper
#> 1992 - 15 FOLSTAD I, 1992, AMERICAN
NATURALIST
#> 1994 - 26 GUSTAFSSON L, 1994, PHILOSOPHICAL TRANSACTIONS - ROYAL SOCIETY O
F LONDON, B
#> 1997 - 62 SIIKAMKI P, 1997,
FUNCT ECOL
#> 1997 - 63 ALLANDER K, 1997,
FUNCT ECOL
#> 1998 - 72 NORDLING D, 1998, PROC R SOC
B BIOL SCI
#> 2000 - 100 ILMONEN P, 2000, PROC R SOC
B BIOL SCI
#> 2000 - 110 WORDEN BD, 2000,
ANIM BEHAV
#> 2002 - 143 AHMED AM,
2002, OIKOS
#> 2003 - 167 BONNEAUD C,
2003, AM NAT
#> 2004 - 194 JACOT A,
2004, EVOLUTION
#> 2004 - 199 BONNEAUD C,
2004, EVOLUTION
#> 2005 - 217 CHADWICK W, 2005, PROC R SOC
B BIOL SCI
#> 2005 - 219 MARZAL A,
2005, OECOLOGIA
#> 2006 - 236 ULLER T, 2006,
FUNCT ECOL
#> 2006 - 246 VELANDO A, 2006, PROC R SOC
B BIOL SCI
#> 2007 - 294 BENNSCH S, 2007,
J ANIM ECOL
#> 2008 - 308 MARZAL A, 2008,
J EVOL BIOL
#> 2009 - 346 KNOWLES SCL, 2009,
FUNCT ECOL
#> 2010 - 358 KIVLENIECE I, 2010,
ANIM BEHAV
#> 2010 - 387 KNOWLES SCL, 2010,
J EVOL BIOL
#> DOI Year LCS GCS
#> 1992 - 15 10.1086/285346 1992 35 1827
#> 1994 - 26 10.1098/RSTB.1994.0149 1994 26 300
#> 1997 - 62 10.1046/J.1365-2435.1997.00075.X 1997 14 47
#> 1997 - 63 10.1046/J.1365-2435.1997.00095.X 1997 14 66
#> 1998 - 72 10.1098/RSPB.1998.0432 1998 31 306
#> 2000 - 100 10.1098/RSPB.2000.1053 2000 17 203

```

```

#> 2000 - 110          10.1006/ANBE.1999.1368 2000 11  53
#> 2002 - 143 10.1034/J.1600-0706.2002.970307.X 2002 14 109
#> 2003 - 167          10.1086/346134 2003 34 345
#> 2004 - 194 10.1111/J.0014-3820.2004.TB01603.X 2004 17 105
#> 2004 - 199 10.1111/J.0014-3820.2004.TB01633.X 2004 20 119
#> 2005 - 217          10.1098/RSPB.2004.2959 2005 12  63
#> 2005 - 219          10.1007/S00442-004-1757-2 2005 21 215
#> 2006 - 236 10.1111/J.1365-2435.2006.01163.X 2006 11  63
#> 2006 - 246          10.1098/RSPB.2006.3480 2006 16 164
#> 2007 - 294 10.1111/J.1365-2656.2006.01176.X 2007 12 151
#> 2008 - 308 10.1111/J.1420-9101.2008.01545.X 2008 15 137
#> 2009 - 346 10.1111/J.1365-2435.2008.01507.X 2009 16 123
#> 2010 - 358 10.1016/J.ANBEHAV.2010.09.004 2010 11  39
#> 2010 - 387 10.1111/J.1420-9101.2009.01920.X 2010 13 136

```

Only articles with minimum of 10 citations are included in teh above analysis, if you change this number to a higher value, the analyses will be quicker and the plot less dense.

**MORE TO DO** You can use different types of network plots - just tweak “type” parameter in the *networkPlot* function (check the vignette for the available options). Type indicates the network map layout: circle, kamada-kawai, mds, etc.

You can use non-R tools to visualise bibliographic networks, e.g. VOSviewer software by Nees Jan van Eck and Ludo Waltman (<http://www.vosviewer.com> (<http://www.vosviewer.com>)). When in R function you usetype=“vosviewer”, the function will export the network a standard “pajek” network file (named “vosnetwork.net”), which can be used in other network-plotting software, including VOSviewer.

## Resources

- Bibliometrix webpage: <http://www.bibliometrix.org/> (<http://www.bibliometrix.org/>) and documentation: <http://www.bibliometrix.org/documents.html> (<http://www.bibliometrix.org/documents.html>)
- A more detailed *bibliometrix* package tutorial: <https://cran.r-project.org/web/packages/bibliometrix/vignettes/bibliometrix-vignette.html> (<https://cran.r-project.org/web/packages/bibliometrix/vignettes/bibliometrix-vignette.html>)
- Another useful R package - Revtools: <http://revtools.net/> (<http://revtools.net/>)
- Stand-alone GUI software - VOSviewer: <http://www.vosviewer.com/> (<http://www.vosviewer.com/>)