

Análise Bibliométrica com R

Neste trabalho irei apresentar o workflow básico para a realização de uma análise bibliométrica utilizando o R e o pacote Bibliometrix a partir dos dados de publicações científicas obtidas na base de dados Scopus.

A preparação do ambiente (instalação do R, RStudio e Bibliometrix) e a obtenção dos dados seguirão conforme a aula apresentada anteriormente. Mais informações sobre estas etapas podem ser obtidas aqui.

Este *markdown* foi desenvolvido baseado no material oficial do Bibliometrix.

Inicialmente, se faz necessário carregar o pacote *Bibliometrix* no R, que irá ser responsável por todo o processamento e análise de indicadores de nosso trabalho.

```
library(bibliometrix)

##
## bibliometrix
## A R tool for comprehensive bibliometric analysis of scientific literature
##
## by Massimo Aria & Corrado Cuccurullo
##
## http:\\www.bibliometrix.org
```

Com o ambiente preparado, iremos carregar os dados bibliométricos obtidos junto a base de dados Scopus.

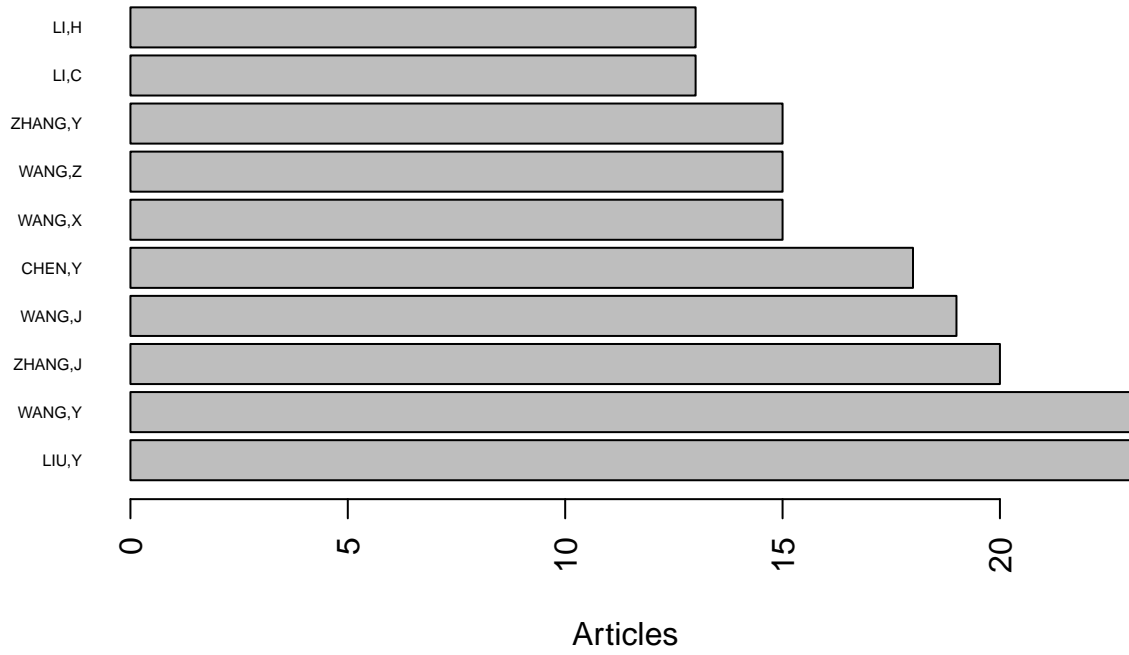
```
D <- readLines("scopus.bib")
M <- convert2df(D, dbsource = "scopus", format = "bibtex")

## 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    874
```

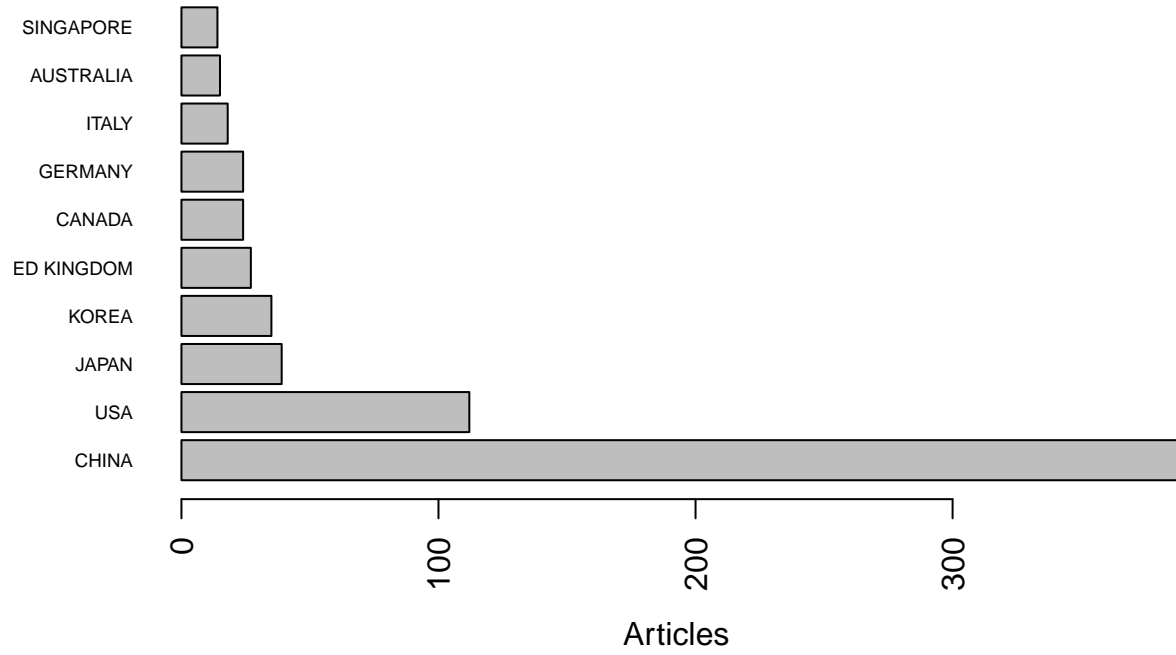
Para dar início a análise, é necessário executar o pré-processamento dos dados, preparando-os para posterior análise de indicadores. Além disso, algumas métricas interessantes podem ser extraídas.

```
### (grafico) Processar resultados bibliometricos
results <- biblioAnalysis(M, sep = ";")
plot(x = results, k = 10, pause = FALSE)
```

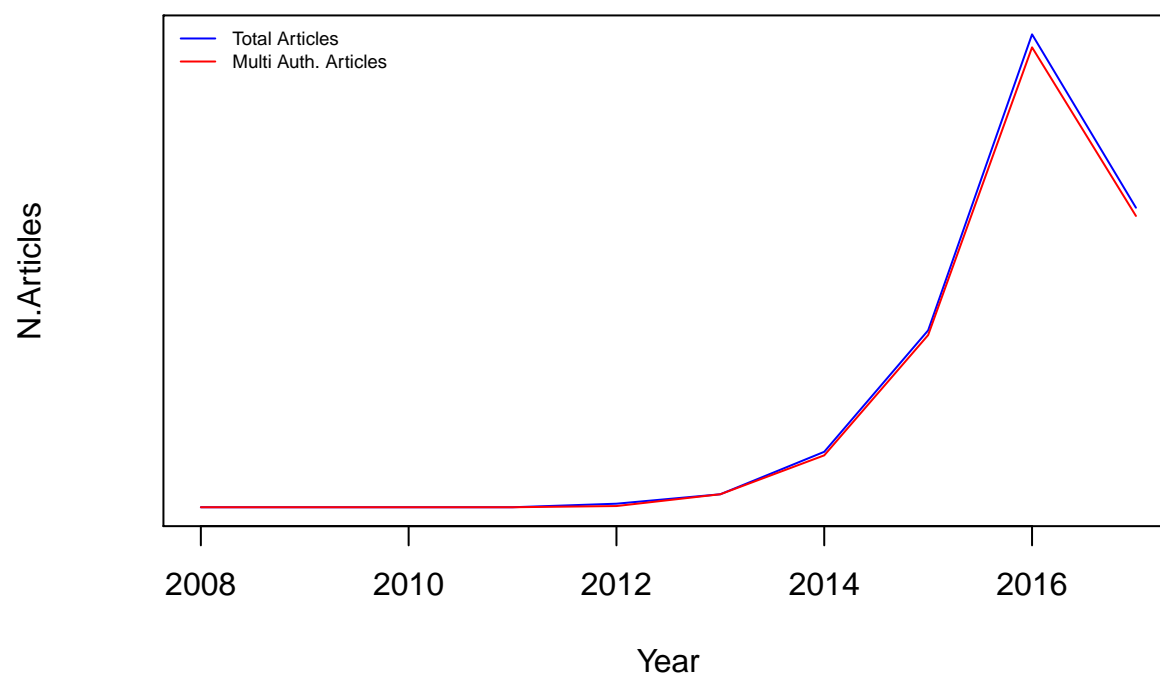
Most Productive Authors



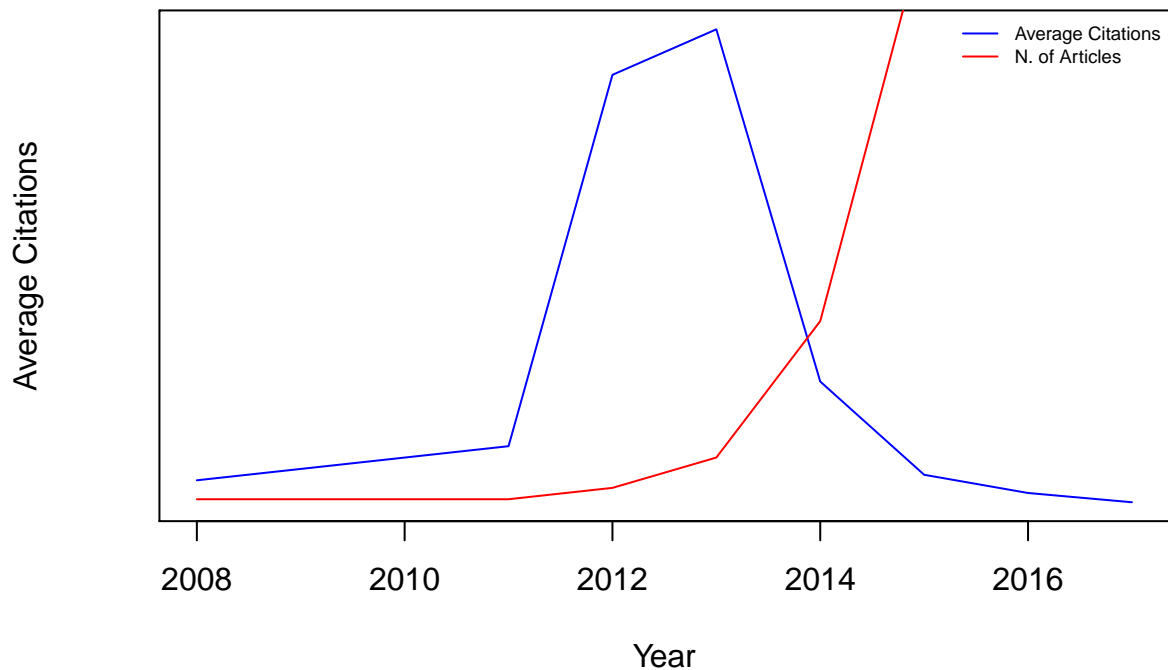
Most Productive Countries



Articles per Year



Average Citations per Article



*** Com os dados preparados, o Bibliometrix oferece um resumo completo do conjunto de publicações analisados, incluindo:

- Resumo quantitativo
 - Artigos
 - Periódicos
 - Palavras-chave (Autorais e Indexadas)
 - Citações média por artigo
 - Período contemplado na pesquisa
- Evolução histórica (p/ ano)
- Autores mais produtivos
- Artigos mais citados
- Produção científica por país
- etc.

```
### (console) Resumo dos resultados
S <- summary(object = results, k = 10, pause = FALSE)
```

```
##
##
## Main Information about data
##
## Articles 874
## Sources (Journals, Books, etc.) 370
## Keywords Plus (ID) 5023
## Author's Keywords (DE) 2051
## Period 2008 - 2017
## Average citations per article 6.585
```

```

##
## Authors 2339
## Author Appearances 3676
## Authors of single authored articles 20
## Authors of multi authored articles 2319
##
## Articles per Author 0.374
## Authors per Article 2.68
## Co-Authors per Articles 4.21
## Collaboration Index 2.74
##
##
## Annual Scientific Production
##
## Year Articles
## 2008 1
## 2011 1
## 2012 4
## 2013 12
## 2014 48
## 2015 151
## 2016 402
## 2017 255
##
## Annual Percentage Growth Rate 120.6945
##
##
## Most Productive Authors
##
## Authors Articles Authors Articles Fractionalized
## 1 LIU,Y 23 WANG,Y 5.26
## 2 WANG,Y 23 LIU,Y 5.21
## 3 ZHANG,J 20 ZHANG,Y 5.04
## 4 WANG,J 19 WANG,J 4.64
## 5 CHEN,Y 18 ZHANG,J 4.13
## 6 WANG,X 15 ZHANG,S 4.03
## 7 WANG,Z 15 CHEN,Y 3.88
## 8 ZHANG,Y 15 WANG,X 3.82
## 9 LI,C 13 ZHANG,L 3.38
## 10 LI,H 13 LI,Y 3.35
##
##
## Top manuscripts per citations
##
##
## Paper
## 1 SRIVASTAVA N;HINTON G;KRIZHEVSKY A;SUTSKEVER I;SALAKHUTDINOV R,(2014),J. MACH. LEARN. RES.
## 2 BENGIO Y;COURVILLE A;VINCENT P,(2013),IEEE TRANS PATTERN ANAL MACH INTELL
## 3 BERGSTRA J;BENGIO Y,(2012),J. MACH. LEARN. RES.
## 4 JI S;XU W;YANG M;YU K,(2013),IEEE TRANS PATTERN ANAL MACH INTELL
## 5 CHEN Y;LIN Z;ZHAO X;WANG G;GU Y,(2014),IEEE J. SEL. TOP. APPL. EARTH OBS. REMOTE SENS.
## 6 DONG C;LOY C;HE K;TANG X,(2016),IEEE TRANS PATTERN ANAL MACH INTELL
## 7 O'CONNOR P;NEIL D;LIU S;DELBRUCK T;PFEIFFER M,(2013),FRONT. NEUROSCI.
## 8 SRIVASTAVA N;SALAKHUTDINOV R,(2014),J. MACH. LEARN. RES.
## 9 SARIKAYA R;HINTON G;DEORAS A,(2014),IEEE TRANS. AUDIO SPEECH LANG. PROCESS.

```

10 CHAN T;JIA K;GAO S;LU J;ZENG Z;MA Y,(2015),IEEE TRANS IMAGE PROCESS

TC TCperYear

1 910 303.3

2 900 225.0

3 396 79.2

4 362 90.5

5 108 36.0

6 86 86.0

7 67 16.8

8 57 19.0

9 57 19.0

10 56 28.0

##

##

Most Productive Countries

##

Country Articles Freq

1 CHINA 389 0.4571

2 USA 112 0.1316

3 JAPAN 39 0.0458

4 KOREA 35 0.0411

5 UNITED KINGDOM 27 0.0317

6 CANADA 24 0.0282

7 GERMANY 24 0.0282

8 ITALY 18 0.0212

9 AUSTRALIA 15 0.0176

10 SINGAPORE 14 0.0165

##

##

Total Citations per Country

##

Country Total Citations Average Article Citations

1 CANADA 2425 101.042

2 CHINA 1051 2.702

3 USA 946 8.446

4 HONG KONG 156 12.000

5 UNITED KINGDOM 135 5.000

6 NETHERLANDS 111 8.538

7 SINGAPORE 105 7.500

8 SWITZERLAND 103 8.583

9 AUSTRALIA 92 6.133

10 JAPAN 91 2.333

##

##

Most Relevant Sources

##

##

1 NEUROCOMPUTING

2 IEEE TRANSACTIONS ON MEDICAL IMAGING

3 IEEE TRANSACTIONS ON IMAGE PROCESSING

4 PATTERN RECOGNITION

5 LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND

6 IEEE TRANSACTIONS ON MULTIMEDIA

7 NEURAL NETWORKS

```

## 8 JOURNAL OF ELECTRONIC IMAGING
## 9 PLOS ONE
## 10 SENSORS (SWITZERLAND)
## Articles
## 1 50
## 2 17
## 3 16
## 4 16
## 5 13
## 6 12
## 7 12
## 8 11
## 9 11
## 10 11
##
##
## Most Relevant Keywords
##
## Author Keywords (DE) Articles Keywords-Plus (ID)
## 1 DEEP LEARNING 505 DEEP LEARNING
## 2 CONVOLUTIONAL NEURAL NETWORK 99 NEURAL NETWORKS
## 3 CONVOLUTIONAL NEURAL NETWORKS 91 CONVOLUTION
## 4 MACHINE LEARNING 54 CONVOLUTIONAL NEURAL NETWORK
## 5 NEURAL NETWORKS 43 DEEP NEURAL NETWORKS
## 6 NEURAL NETWORK 36 LEARNING SYSTEMS
## 7 DEEP NEURAL NETWORKS 32 ARTIFICIAL INTELLIGENCE
## 8 CLASSIFICATION 23 ARTIFICIAL NEURAL NETWORK
## 9 FEATURE EXTRACTION 22 CLASSIFICATION (OF INFORMATION)
## 10 UNSUPERVISED LEARNING 20 COMPLEX NETWORKS
## Articles
## 1 421
## 2 336
## 3 255
## 4 248
## 5 174
## 6 169
## 7 131
## 8 100
## 9 98
## 10 78

```

É possível também realizar uma análise cruzada entre os trabalhos do conjunto de dados onde, por exemplo, extrair quais os papers mais citados no conjunto de dados analisado.

```

#### (console/viewer) Referências (Papers) mais citadas, no conjunto de dados
CR_Papers = citations(M, field = "article", sep = ";")
CR_Papers <- as.data.frame(CR_Papers$Cited[1:10])
names(CR_Papers) <- c('Paper', 'Citations')
CR_Papers

```

```

##
## 1 HINTON, GE, SALAKHUTDINOV, RR, REDUCING THE DIMEN
## 2 LECUN
## 3 BENGIO, Y, LEARNING DEEP ARCHITECT

```



```
## 4 HINTON, GE, SALAKHUTDINOV, RR, REDUCING TI
## 5 LECUN, Y, BOTTOU, L, BENGIO, Y, HAFFNER, P, GRADIENT-BASED LEARNING
## 6 LECUN, Y, BOTTOU, L, BENGIO, Y, HAFFNER, P, GRADIENT-BASED LE
## 7 KRIZHEVSKY, A, SUTSKEVER, I, HINTON, GE, IMAGENET CLASSIFICATION WITH DEEP CONVOLUTIONAL NEURAL NI
## 8 HINTON, GE, OSINDER, S, TEH, Y-W, A FAST LEARN
## 9 BENGIO, Y, LEARN
## 10 HINTON, GE, OSINDER, S, TEH, YW, A FAST LEARNING A
## Citations
## 1 119
## 2 48
## 3 44
## 4 40
## 5 40
## 6 36
## 7 35
## 8 34
## 9 33
## 10 27
#View(CR_Papers)
```

Ou ainda, é possível obter os autores mais referenciados dentro do conjunto de dados.

```
CR_Authors = citations(M, field = "author", sep = ";")
CR_Authors <- as.data.frame(CR_Authors$Cited[1:10])
names(CR_Authors) <- c('Author', 'Citations')
CR_Authors
```

```
## Author Citations
## 1 BENGIO, Y 1429
## 2 HINTON, GE 1274
## 3 LECUN, Y 779
## 4 HINTON, G 626
## 5 KRIZHEVSKY, A 593
## 6 SUTSKEVER, I 577
## 7 VINCENT, P 377
## 8 LAROCHELLE, H 324
## 9 SCHMIDHUBER, J 318
## 10 BOTTOU, L 312
```

```
#View(CR_Authors)
```

O Fator de Dominância é um importante indicador que mensura a produtividade de autores, calculando a razão entre o número de trabalhos em que o autor aparece como **Primeiro Autor** e o número total de trabalhos multi-autorais em que este aparece.

```
# k = 10 registros
DF <- dominance(results, k = 10)
DF
```

```
## Dominance Factor Multi Authored First Authored Rank by Articles
## LI,H 0.3846154 13 5 9
## WANG,Y 0.3478261 23 8 2
## LI,Y 0.3076923 13 4 10
```

## CHEN,Y	0.2777778	18	5	5
## ZHANG,Y	0.2666667	15	4	7
## LIU,Y	0.2173913	23	5	1
## WANG,J	0.2105263	19	4	4
## WANG,X	0.2000000	15	3	6
## LI,C	0.1538462	13	2	8
## ZHANG,J	0.1000000	20	2	3
##	Rank by DF			
## LI,H	1			
## WANG,Y	2			
## LI,Y	3			
## CHEN,Y	4			
## ZHANG,Y	5			
## LIU,Y	6			
## WANG,J	7			
## WANG,X	8			
## LI,C	9			
## ZHANG,J	10			

O índice-H (H-index) é outra métrica importante para avaliar a relevância acadêmica de um autor. Ela leva em consideração o número de artigos com citações maiores ou iguais a esse número.

Por exemplo: um autor com índice-H = 15 indica que este autor possui 15 artigos com pelo menos 15 citações.

```
authors = gsub(",", " ", names(results$Authors)[1:10])
indices <- Hindex(M, authors, sep = ";", years=10)
HI <- indices$H
HI
```

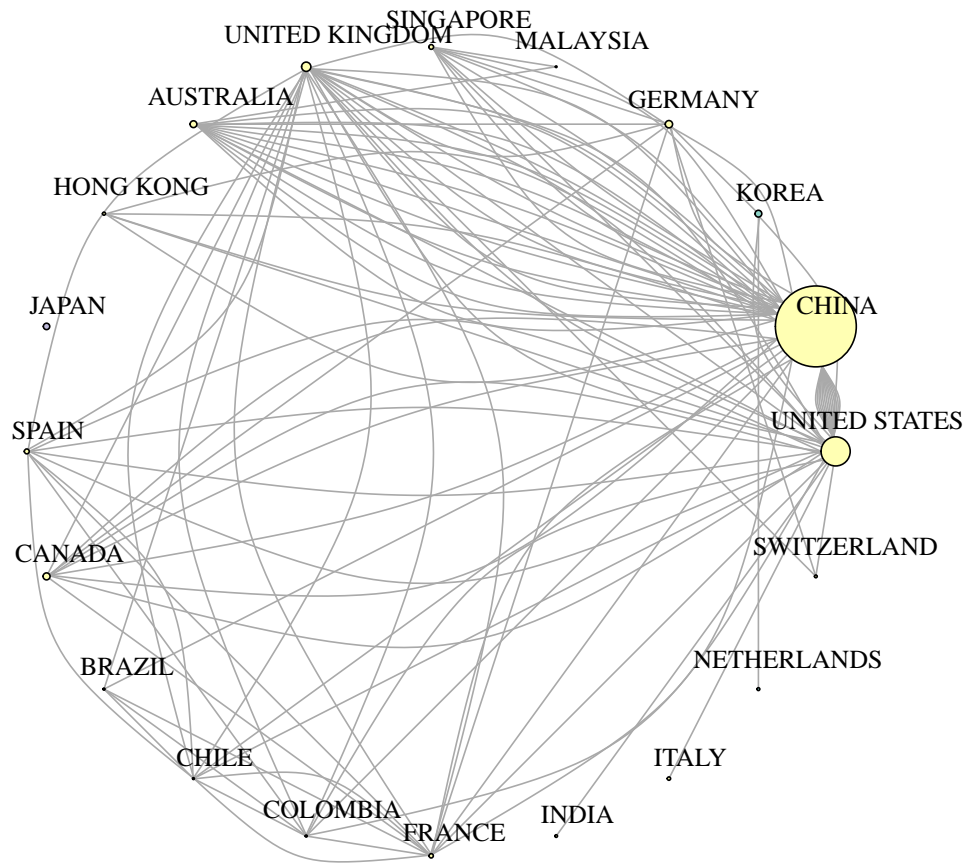
##	Author	h_index	g_index	m_index	TC	NP
## 1	LIU Y	5	8	1.2500000	81	23
## 2	WANG Y	3	6	0.7500000	47	21
## 3	ZHANG J	4	6	1.0000000	52	19
## 4	WANG J	3	7	1.0000000	54	19
## 5	CHEN Y	6	14	1.5000000	206	18
## 6	WANG X	5	7	1.0000000	64	15
## 7	WANG Z	3	5	1.0000000	32	14
## 8	ZHANG Y	3	5	0.7500000	25	15
## 9	LI C	2	4	0.6666667	20	13
## 10	LI H	3	6	0.7500000	42	14

Dentro do conjunto de dados, os atributos dos papers estão conectados entre si através do próprio paper. Assim, por exemplo, Autor(es) estão interconectados com Periódicos, Palavras-chave com Data de publicação, etc. Essas conexões de diferentes atributos geram redes bipartidas que podem ser representadas como matrizes retangulares (Papers x ATRIBUTO_QUALQUER).

Logo, por exemplo, é possível gerar uma rede que apresenta o comportamento colaborativo entre países dentro da área de pesquisa em questão, ou seja, representa o modo como ocorre a transferência de conhecimento acadêmico.

```
M <- metaTagExtraction(M, Field = "AU_CO", sep = ";")
NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "countries", sep = ";")
networkPlot(NetMatrix, n = 20, Title = "Country Collaboration", type = "circle", size=TRUE, remove.mult.
```

Country Collaboration

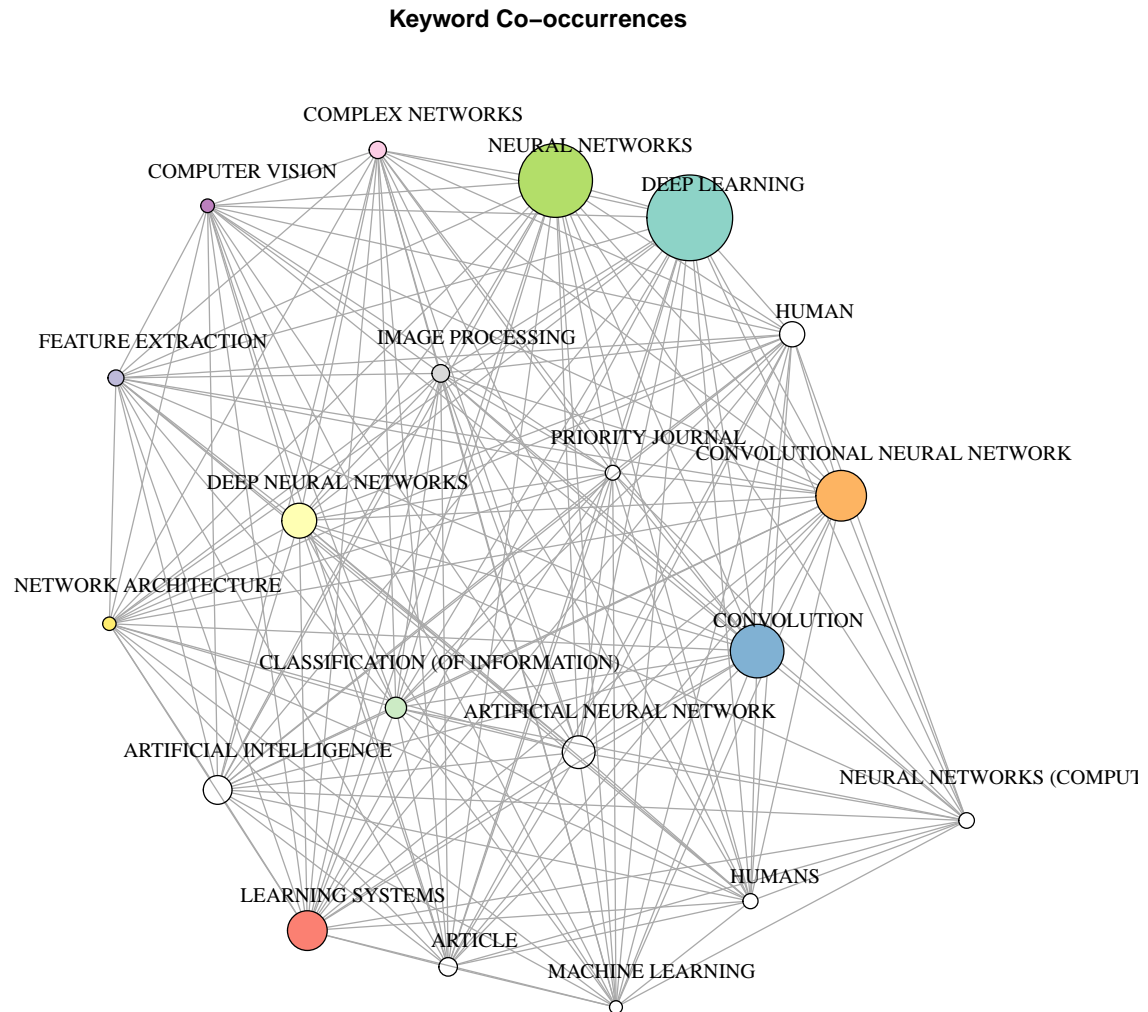


```
## IGRAPH UN-- 20 110 --
## + attr: name (v/c), id (v/c), size (v/n), color (v/c)
## + edges (vertex names):
## [1] UNITED STATES--CHINA      UNITED STATES--CHINA
## [3] UNITED STATES--CHINA      UNITED STATES--CHINA
## [5] UNITED STATES--CHINA      UNITED STATES--CHINA
## [7] UNITED STATES--CHINA      UNITED STATES--CHINA
## [9] UNITED STATES--CHINA      UNITED STATES--CHINA
## [11] UNITED STATES--CHINA      UNITED STATES--CHINA
## [13] UNITED STATES--CHINA      UNITED STATES--CHINA
## [15] UNITED STATES--KOREA      UNITED STATES--KOREA
## + ... omitted several edges
```

Ou também, pode ser gerado uma rede de coocorrências de palavras-chave, permitindo analisar subáreas da área de conhecimento estudada.

```
# (grafico) Rede de co-ocorrencias de palavras-chave (20 keywords)
```

```
NetMatrix <- biblioNetwork(M, analysis = "co-occurrences", network = "keywords", sep = ";")
networkPlot(NetMatrix, n = 20, Title = "Keyword Co-occurrences", type = "kamada", size=T)
```



```
## IGRAPH UN-- 20 184 --
## + attr: name (v/c), id (v/c), size (v/n), color (v/c)
## + edges (vertex names):
## [1] DEEP LEARNING--DEEP NEURAL NETWORKS
## [2] DEEP LEARNING--FEATURE EXTRACTION
## [3] DEEP LEARNING--LEARNING SYSTEMS
## [4] DEEP LEARNING--CONVOLUTION
## [5] DEEP LEARNING--CONVOLUTIONAL NEURAL NETWORK
```

```
## [6] DEEP LEARNING--NEURAL NETWORKS
## [7] DEEP LEARNING--COMPLEX NETWORKS
## [8] DEEP LEARNING--IMAGE PROCESSING
## + ... omitted several edges
```

Conclusão

Portanto, a linguagem R juntamente com o pacote Bibliometrix fornecem mecanismos eficientes para realizar a análise bibliométrica em conjuntos de publicações acadêmicas obtidos na Scopus, permitindo extrair conhecimento importante a partir de indicadores relevantes, os quais podem agregar qualidade na produção científica em quaisquer áreas do conhecimento.

Referências

- <http://r-project.org/>
- <http://www.bibliometrix.org/>
- <https://github.com/diegocavalca/data-science>
- <http://www.collnet.de/Berlin-2008/KumarWIS2008cir.pdf>
- <http://htmlpreview.github.io/?https://github.com/massimoaria/bibliometrix/master/vignettes/bibliometrix-vignette.html>