Short Chain Fatty Acids (SCFA)

Sarah de Oliveira

17 de maio de 2019

## Introdução

Sempre que se é iniciado um trabalho, necessita-se de uma bibliografia básica que supra a demanda de informações primordiais ao embasamento científico. Deste modo, faz-se necessária a seleção da melhor fonte de informação para o esboço de um trabalho científico, e com o auxílio de ferramentas de trabalho como o R e seus pacotes, tal seleção se torna mais fidedigna. Como no caso apresentado, traz-se um levantamento de autores e publicações relacionados com os Ácidos graxos de cadeia curta, do inglês, short chain fatty acids (SCFA). Os SCFAs são moléculas simples compostas por átomos de carbonos (1-5) e um grupo carboxila, produzidos pelo metabolismo fermentativo das bactérias que compõem a microbiota comensal intestinal. Eles são relevantes pois exercem funções biológicas a nível local no tratogastrointestinal e a nível sistêmico, sendo importantes para a modulação do sistema imune.

## Requerimentos

Para que o trabalho pudesse ser executado, fez-se necessária a utilização de alguns programas, ferramentas e pacotes do R disponíveis para o uso. Dentre eles, destacam-se: Programa R e RStudio; Git (um sistema de controle de versão); MikTex (programa de conversão do Script do Rmarkdown para PDF); Pacote “Rmarkdown” (conversão de scripts para diversos formatos); Pacote “Bibliometrix” (conjunto de ferramentas para análises na área de cientometria e bibliometria); Um arquivo Scopus.bib (arquivo baixado da base de dados Scopus com todas as informações de artigos relacionados ao tema “SCFA”).

## Desenvolvimento

#### Carregamento e conversão de dados

Após realizar o download de todos os programas necessários, os quais foram citados anteriormente, e criação do seu diretório no Git, é preciso agora fazer o download dos pacotes necessários para a criação do trabalho. Primeiramente instala-se o Rmarkdown, realiza-se sua configuração e em seguida a instalação do pacote Bibliometrix.

install.packages(“rmarkdown”) install.packages(“bibliometrix”)

OBSERVAÇÃO: Não é possível gerar o arquivo de Word com as funções de instalação dos pacotes em forma de comandos.

No caso de dúvidas basta digitar no console help(nome\_da\_função) ou ?nome\_da\_função, e no quadrante inferior direito irá aparecer um manual sobre como utilizar a função requerida.

Para realizar o carregamento do pacote utilizado na análise, digita-se a seguinte linha de comando:

library(bibliometrix)

## To cite bibliometrix in publications, please use:  
##   
## Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11(4), pp 959-975, Elsevier.  
##   
##   
## http:\\www.bibliometrix.org  
##   
##   
## To start with the shiny web-interface, please digit:  
## biblioshiny()

Depois de estabelecer seu arquivo de dados na pasta de trabalho, é necessário que ele seja carregado no seu workspace para a análise. Para isso, utiliza-se a função “readFiles” sobre o arquivo “scopus.bib”“, tal função será atribuída ao vetor de caracteres”D“.

D <- readFiles("scopus2.bib")

Desta forma agora O vetor “D” contém o nome dos arquivos baixados do site SCOPUS, pois a função “readFiles”" combina todos os arquivos de texto em um único vetor grande de caracteres. Em seguida o objeto D deve ser convertido em um data frame usando a função “convert2df” como na linha de comando abaixo, e atribuído a “M”:

M <- convert2df(D, dbsource = "scopus", format = "bibtex")

##   
## 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 921   
## Done!  
##   
##   
## Generating affiliation field tag AU\_UN from C1: Done!

#### Análise Bibliométrica

Agora o primeiro passo para realizar uma análise descritiva do quadro de dados bibliográficos é utilizar a função “biblioAnalysis”, a qual calcula as principais medidas bibliométricas e retorna um objeto da classe “bibliometrix”.

results <- biblioAnalysis(M, sep = ";")

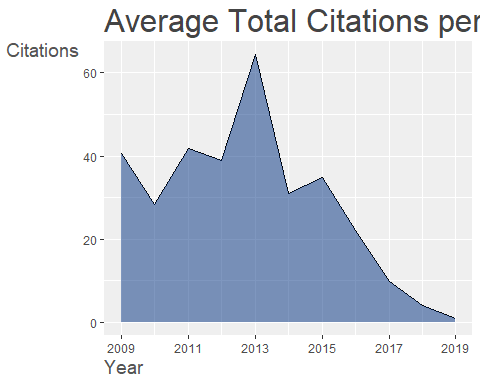
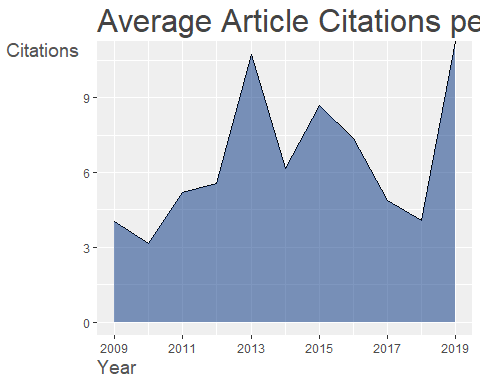
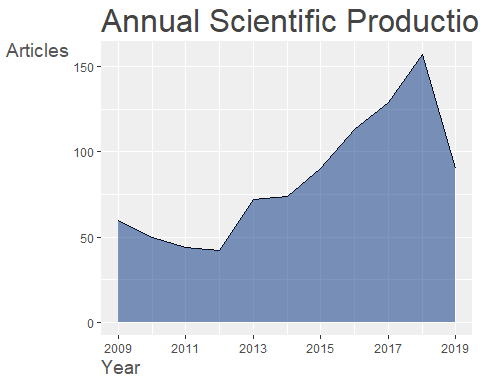
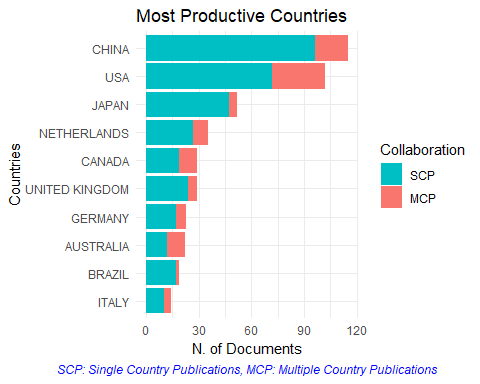
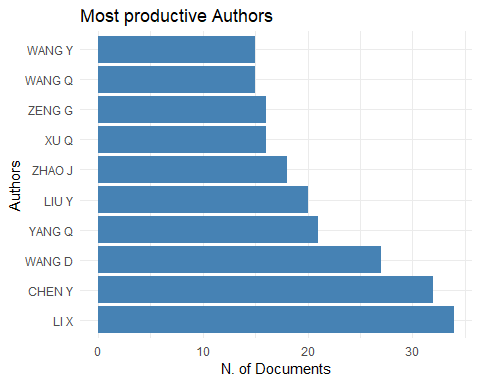
Como uma forma de resumir os principais resultados da análise bibliométrica, facilitando assim a leitura de todas as informações contidas ali, utiliza-se a função “summary”, a qual apresenta as principais informações sobre o quadro de dados bibliográficos e várias tabelas, como produção científica anual, principais artigos por número de citações, autores mais produtivos, países mais produtivos, citação total por país, fontes mais relevantes (periódicos) e palavras-chave mais relevantes.

summary(results, k = 10, pause = FALSE)

##   
##   
## Main Information about data  
##   
## Documents 921   
## Sources (Journals, Books, etc.) 479   
## Keywords Plus (ID) 7227   
## Author's Keywords (DE) 1973   
## Period 2009 - 2019   
## Average citations per documents 23.72   
##   
## Authors 4053   
## Author Appearances 5530   
## Authors of single-authored documents 25   
## Authors of multi-authored documents 4028   
## Single-authored documents 27   
##   
## Documents per Author 0.227   
## Authors per Document 4.4   
## Co-Authors per Documents 6   
## Collaboration Index 4.51   
##   
## Document types   
## ARTICLE 798   
## ARTICLE IN PRESS 6   
## BOOK CHAPTER 15   
## CONFERENCE PAPER 11   
## EDITORIAL 4   
## ERRATUM 5   
## LETTER 7   
## NOTE 4   
## REVIEW 65   
## SHORT SURVEY 6   
##   
##   
## Annual Scientific Production  
##   
## Year Articles  
## 2009 60  
## 2010 50  
## 2011 44  
## 2012 42  
## 2013 72  
## 2014 74  
## 2015 90  
## 2016 113  
## 2017 129  
## 2018 157  
## 2019 90  
##   
## Annual Percentage Growth Rate 4.137974   
##   
##   
## Most Productive Authors  
##   
## Authors Articles Authors Articles Fractionalized  
## 1 LI X 34 CHEN Y 6.74  
## 2 CHEN Y 32 LI X 4.70  
## 3 WANG D 27 WANG D 3.22  
## 4 YANG Q 21 PLUZNICK JL 2.89  
## 5 LIU Y 20 PENNER GB 2.58  
## 6 ZHAO J 18 LIU Y 2.55  
## 7 XU Q 16 FENG L 2.38  
## 8 ZENG G 16 KIM CH 2.29  
## 9 WANG Q 15 ZHAO J 2.28  
## 10 WANG Y 15 YANG Q 2.28  
##   
##   
## Top manuscripts per citations  
##   
## Paper TC TCperYear  
## 1 SMITH PM, 2013, SCIENCE 1305 217.5  
## 2 DEN BESTEN G, 2013, J LIPID RES 737 122.8  
## 3 TOLHURST G, 2012, DIABETES 631 90.1  
## 4 KOH A, 2016, CELL 396 132.0  
## 5 KIMURA I, 2013, NAT COMMUN 395 65.8  
## 6 VINOLO MAR, 2011, NUTR 370 46.2  
## 7 CANFORA EE, 2015, NAT REV ENDOCRIONOL 316 79.0  
## 8 ROS-COVIN D, 2016, FRONT MICROBIOL 303 101.0  
## 9 FROST G, 2014, NAT COMMUN 288 57.6  
## 10 TAN J, 2014, ADV IMMUNOL 288 57.6  
##   
##   
## Corresponding Author's Countries  
##   
## Country Articles Freq SCP MCP MCP\_Ratio  
## 1 CHINA 115 0.1913 96 19 0.1652  
## 2 USA 102 0.1697 72 30 0.2941  
## 3 JAPAN 52 0.0865 47 5 0.0962  
## 4 NETHERLANDS 35 0.0582 27 8 0.2286  
## 5 CANADA 29 0.0483 19 10 0.3448  
## 6 UNITED KINGDOM 29 0.0483 24 5 0.1724  
## 7 GERMANY 23 0.0383 17 6 0.2609  
## 8 AUSTRALIA 22 0.0366 12 10 0.4545  
## 9 BRAZIL 19 0.0316 17 2 0.1053  
## 10 ITALY 14 0.0233 10 4 0.2857  
##   
##   
## SCP: Single Country Publications  
##   
## MCP: Multiple Country Publications  
##   
##   
## Total Citations per Country  
##   
## Country Total Citations Average Article Citations  
## 1 USA 4346 42.608  
## 2 CHINA 2056 17.878  
## 3 NETHERLANDS 2048 58.514  
## 4 UNITED KINGDOM 1986 68.483  
## 5 JAPAN 1414 27.192  
## 6 AUSTRALIA 842 38.273  
## 7 CANADA 658 22.690  
## 8 DENMARK 488 40.667  
## 9 GERMANY 432 18.783  
## 10 FRANCE 393 39.300  
##   
##   
## Most Relevant Sources  
##   
## Sources Articles  
## 1 BIORESOURCE TECHNOLOGY 26  
## 2 PLOS ONE 26  
## 3 JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY 20  
## 4 SCIENTIFIC REPORTS 20  
## 5 NUTRIENTS 15  
## 6 JOURNAL OF ANIMAL SCIENCE 14  
## 7 JOURNAL OF DAIRY SCIENCE 14  
## 8 WATER RESEARCH 11  
## 9 POULTRY SCIENCE 10  
## 10 JOURNAL OF NUTRITION 9  
##   
##   
## Most Relevant Keywords  
##   
## Author Keywords (DE) Articles Keywords-Plus (ID) Articles  
## 1 SHORT CHAIN FATTY ACIDS 278 FATTY ACIDS 722  
## 2 SHORT CHAIN FATTY ACID 70 ARTICLE 494  
## 3 GUT MICROBIOTA 50 SHORT CHAIN FATTY ACID 445  
## 4 BUTYRATE 45 VOLATILE 431  
## 5 INFLAMMATION 39 FERMENTATION 413  
## 6 SCFA 39 MALE 411  
## 7 MICROBIOTA 35 HUMAN 376  
## 8 WASTE ACTIVATED SLUDGE 33 VOLATILE FATTY ACID 350  
## 9 ANAEROBIC FERMENTATION 27 METABOLISM 346  
## 10 OBESITY 23 CONTROLLED STUDY 337

Essa função aceita dois argumentos, “k” é um valor que indica o número de linhas de cada tabela e “pause” é um valor lógico (VERDADEIRO ou FALSO) usado para permitir pausar na rolagem da tela ou não. A partir das informações básicas dadas pela função “summary”, podemos construir gráficos que nos permitem visualizar as variáveis que estarão sendo analisadas em cima de nossos dados, como os autores mais produtivos, os países mais produtivos no quesito científico, a produção anual científica relativa ao tema procurado, etc. Para isso basta aplicar a função “plot” como descrito abaixo.

plot(x=results, k=10, pause=FALSE)



#### Análise de referências citadas

Além de tudo o que já foi mostrado, também é possível montar uma tabela de frequências das referências mais citadas ou dos os primeiros autores mais citados, tudo isso utilizando a função “citations” como mostrado nos exemplos abaixo:

* Artigos mais citados:

CR <- citations(M, field = "article", sep = ";")  
cbind(CR$Cited[1:10])

## [,1]  
## CUMMINGS, J.H., POMARE, E.W., BRANCH, W.J., NAYLOR, C.P., MACFARLANE, G.T., SHORT CHAIN FATTY ACIDS IN HUMAN LARGE INTESTINE, PORTAL, HEPATIC AND VENOUS BLOOD (1987) GUT, 28, PP. 1221-1227 59  
## TOPPING, D.L., CLIFTON, P.M., SHORT-CHAIN FATTY ACIDS AND HUMAN COLONIC FUNCTION: ROLES OF RESISTANT STARCH AND NONSTARCH POLYSACCHARIDES (2001) PHYSIOL REV, 81, PP. 1031-1064 36  
## BERGMAN, E.N., ENERGY CONTRIBUTIONS OF VOLATILE FATTY ACIDS FROM THE GASTROINTESTINAL TRACT IN VARIOUS SPECIES (1990) PHYSIOL REV, 70, PP. 567-590 25  
## CUMMINGS, J.H., SHORT CHAIN FATTY ACIDS IN THE HUMAN COLON (1981) GUT, 22, PP. 763-779 22  
## TOPPING, D.L., CLIFTON, P.M., SHORT-CHAIN FATTY ACIDS AND HUMAN COLONIC FUNCTION: ROLES OF RESISTANT STARCH AND NONSTARCH POLYSACCHARIDES (2001) PHYSIOL. REV., 81, PP. 1031-1064 22  
## VINOLO, M.A., RODRIGUES, H.G., NACHBAR, R.T., CURI, R., REGULATION OF INFLAMMATION BY SHORT CHAIN FATTY ACIDS (2011) NUTRIENTS, 3, PP. 858-876 19  
## GAO, Z., YIN, J., ZHANG, J., WARD, R.E., MARTIN, R.J., LEFEVRE, M., CEFALU, W.T., YE, J., BUTYRATE IMPROVES INSULIN SENSITIVITY AND INCREASES ENERGY EXPENDITURE IN MICE (2009) DIABETES, 58, PP. 1509-1517 18  
## LEY, R.E., TURNBAUGH, P.J., KLEIN, S., GORDON, J.I., MICROBIAL ECOLOGY: HUMAN GUT MICROBES ASSOCIATED WITH OBESITY (2006) NATURE, 444, PP. 1022-1023 18  
## MASLOWSKI, K.M., VIEIRA, A.T., NG, A., KRANICH, J., SIERRO, F., YU, D., SCHILTER, H.C., MACKAY, C.R., REGULATION OF INFLAMMATORY RESPONSES BY GUT MICROBIOTA AND CHEMOATTRACTANT RECEPTOR GPR43 (2009) NATURE, 461, PP. 1282-1286 18  
## TURNBAUGH, P.J., LEY, R.E., MAHOWALD, M.A., MAGRINI, V., MARDIS, E.R., GORDON, J.I., AN OBESITY-ASSOCIATED GUT MICROBIOME WITH INCREASED CAPACITY FOR ENERGY HARVEST (2006) NATURE, 444, PP. 1027-1031 18

Outra função que podemos usar é a “localCitations”, a qual contrói uma tabela dos autores mais citados dentro dos nossos próprios dados, ou seja, mede quantas vezes tal autor foi citado dentro da base de dados por outros autores que também estão presente no mesmo local.

CR <- localCitations(M, sep = ";")

## Articles analysed 100   
## Articles analysed 200   
## Articles analysed 300   
## Articles analysed 400   
## Articles analysed 500   
## Articles analysed 600   
## Articles analysed 700   
## Articles analysed 771

CR$Authors[1:10,]

## Author LocalCitations  
## 1699 LI X 197  
## 3140 WANG D 178  
## 438 CHEN Y 171  
## 3098 VINOLO MAR 147  
## 3331 YANG Q 132  
## 532 CURI R 126  
## 3411 ZENG G 118  
## 2498 RODRIGUES HG 110  
## 3076 VENEMA K 101  
## 133 BAKKER BM 99

CR$Papers[1:10,]

## Paper DOI Year LCS GCS  
## 207 DEN BESTEN G, 2013, J LIPID RES 10.1194/JLR.R036012 2013 72 737  
## 178 TOLHURST G, 2012, DIABETES 10.2337/DB11-1019 2012 55 631  
## 141 VINOLO MAR, 2011, NUTR 10.3390/NU3100858 2011 50 370  
## 243 KIM MH, 2013, GASTROENTEROLOGY 10.1053/J.GASTRO.2013.04.056 2013 44 207  
## 220 KIMURA I, 2013, NAT COMMUN 10.1038/NCOMMS2852 2013 38 395  
## 460 KOH A, 2016, CELL 10.1016/J.CELL.2016.05.041 2016 37 396  
## 83 ZAIBI MS, 2010, FEBS LETT 10.1016/J.FEBSLET.2010.04.027 2010 34 154  
## 119 VINOLO MAR, 2011, J NUTR BIOCHEM 10.1016/J.JNUTBIO.2010.07.009 2011 33 155  
## 319 TAN J, 2014, ADV IMMUNOL 10.1016/B978-0-12-800100-4.00003-9 2014 33 288  
## 486 ROS-COVIN D, 2016, FRONT MICROBIOL 10.3389/FMICB.2016.00185 2016 32 303

#### Ranking de dominância de autores

É possível determinar a fração artigos com múltiplos autores em que um deles é citado como primeiro autor, essa fração é chamada de fator de dominância. Para esse cálculo utilizamos a função “dominance”, gerando assim um ranking de dominância dos autores, onde “results” resgata o objeto da classe bibliometrix e “k” o número de autores considerados na análise.

DF <- dominance(results, k = 10)  
DF

## Author Dominance Factor Tot Articles Single-Authored Multi-Authored First-Authored Rank by Articles Rank by DF  
## 1 LUO J 0.58333333 12 0 12 7 1 1  
## 2 ZHAO J 0.50000000 18 0 18 9 6 2  
## 3 FENG L 0.33333333 12 0 12 4 1 3  
## 4 CHEN Y 0.18750000 32 0 32 6 9 4  
## 5 XU Q 0.18750000 16 0 16 3 5 4  
## 6 LIU X 0.15384615 13 0 13 2 3 6  
## 7 WANG D 0.11111111 27 0 27 3 8 7  
## 8 LI X 0.08823529 34 0 34 3 10 8  
## 9 WANG Y 0.06666667 15 0 15 1 4 9  
## 10 LIU Y 0.05000000 20 0 20 1 7 10

#### O índice-h dos autores

O índice-h é uma métrica na qual se mensura o impacto da produtividade e da citação das publicações de um cientista. Tal índice beseia-se em um conjunto de artigos mais citados do referente pesquisador e no número de citações que esses artigos receberam em outras publicações. Para calcular o índice-h dos autores usamos a função “Hindex”, ela também pode calcular o índice-h de fontes e suas variantes (índice-g e índice-m) em determinados dados bibliográficos. Neste caso, serão selecionados 10 autores.

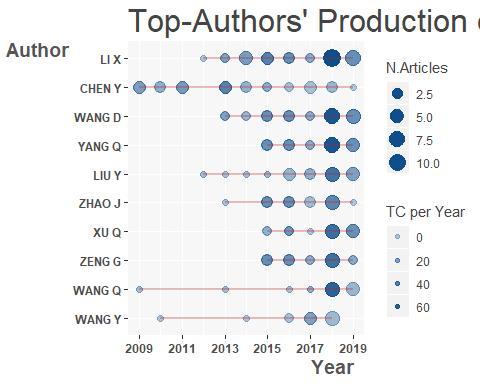
authors=gsub(","," ",names(results$Authors)[1:10])  
indices <- Hindex(M, field = "author", elements=authors, sep = ";", years = 50)  
indices$H

## Author h\_index g\_index m\_index TC NP PY\_start  
## 1 LI X 18 29 2.2500000 869 36 2012  
## 2 CHEN Y 18 30 1.6363636 1112 30 2009  
## 3 WANG D 16 26 2.2857143 727 27 2013  
## 4 YANG Q 13 21 2.6000000 512 21 2015  
## 5 LIU Y 12 19 1.5000000 372 22 2012  
## 6 ZHAO J 10 17 1.4285714 422 17 2013  
## 7 XU Q 10 16 2.0000000 319 16 2015  
## 8 ZENG G 12 16 2.4000000 463 16 2015  
## 9 WANG Q 9 15 0.8181818 307 15 2009  
## 10 WANG Y 6 10 0.6000000 118 14 2010

#### Produtividade dos principais autores ao longo do tempo

Para saber a produção dos autores (no caso, os 10 primeiros autores), em termos de número de publicações e total de citações por ano ao longo do tempo basta usar a função “AuthorProdOverTime”.

topAU <- authorProdOverTime(M, k = 10, graph = TRUE)



#### Estimativa do coeficiente de Lei de Lotka

A Lei de Lotka descreve a frequência de publicação pelos autores em qualquer área dada como uma lei do inverso do quadrado, onde o número de autores publicando um certo número de artigos é uma razão fixa para o número de autores publicando um único artigo. Deste modo, esta suposição implica que o coeficiente beta teórico da Lei de Lotka é igual a 2. A função “lotka” estima os coeficientes da Lei de Lotka para a produtividade científica. Usando a função “lotka” é possível estimar o coeficiente Beta dos nossos dados bibliográficos, avaliando através de um teste estatístico, a similaridade desta distribuição empírica com a teórica.

L <- lotka(results)

* Distribuição empírica da produtividade dos autores

L$AuthorProd

## N.Articles N.Authors Freq  
## 1 1 3331 0.8218603504  
## 2 2 461 0.1137429065  
## 3 3 129 0.0318282754  
## 4 4 58 0.0143103874  
## 5 5 20 0.0049346163  
## 6 6 16 0.0039476931  
## 7 7 8 0.0019738465  
## 8 8 5 0.0012336541  
## 9 9 8 0.0019738465  
## 10 10 1 0.0002467308  
## 11 11 3 0.0007401925  
## 12 12 2 0.0004934616  
## 13 13 1 0.0002467308  
## 14 15 2 0.0004934616  
## 15 16 2 0.0004934616  
## 16 18 1 0.0002467308  
## 17 20 1 0.0002467308  
## 18 21 1 0.0002467308  
## 19 27 1 0.0002467308  
## 20 32 1 0.0002467308  
## 21 34 1 0.0002467308

* Estimativa do coeficiente Beta

L$Beta

## [1] 2.392466

* Constante

L$C

## [1] 0.3262053

* Qualidade de ajuste

L$R2

## [1] 0.9074507

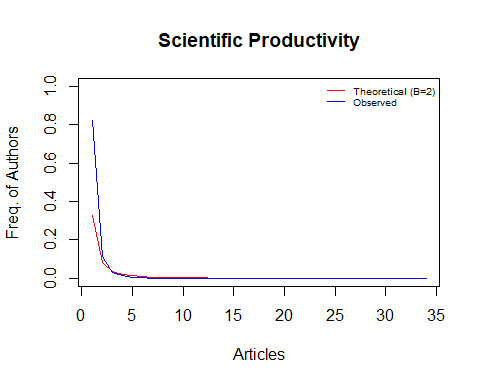
* p-Valor dos dois testes K-S da amostra

L$p.value

## [1] 0.09493348

A tabela “L$ AuthorProd” de modo geral nos fornece a distribuição da produtividade científica baseada em nossos dados. O coeficiente Beta estimado foi calculado em 2.40 com um ajuste de 0.90. O teste de duas amostras de Kolmogorov-Smirnoff fornece um valor de p em 0.094, o que significa que não há uma diferença significativa entre as distribuições de Lotka observadas e as teóricas. É possível comparar as duas distribuições através da utilização da função “plot”.

Observed=L$AuthorProd[,3]  
Theoretical=10^(log10(L$C)-2\*log10(L$AuthorProd[,1]))  
plot(L$AuthorProd[,1],Theoretical,type="l",col="red",ylim=c(0, 1), xlab="Articles",ylab="Freq. of Authors",main="Scientific Productivity")  
lines(L$AuthorProd[,1],Observed,col="blue")  
legend(x="topright",c("Theoretical (B=2)","Observed"),col=c("red","blue"),lty = c(1,1,1),cex=0.6,bty="n")



#### Matrizes de rede bibliográfica

Os elementos do artigo de certa forma estão conectados entre si através de ligações que são estabelecidas pelo próprio manuscrito como os autores, palavras-chave, etc. Essas ligações entre os elementos podem gerar redes bipartidas, as quais são representadas como matrizes retangulares. Além disso, tais publicações científicas geralmente contêm referências a outros trabalhos científicos, gerando desta forma uma rede adicional chamada de rede de co-citação ou acoplamento.

##### - Redes bipartidas

Para calcular uma rede bipartida selecionando um dos elementos presentes em nossos dados, podemos usar a função “cocMatrix”, onde A é uma matriz binária retangular, representando uma rede bipartida onde linhas e colunas são manuscritos e fontes respectivamente.

A <- cocMatrix(M, Field = "SO", sep = ";")

Caso haja necessidade há a possibilidade de classificar, em ordem decrescente, as somas da coluna de A, desse jeito você pode ver as fontes de publicação mais relevantes.

sort(Matrix::colSums(A), decreasing = TRUE)[1:5]

## BIORESOURCE TECHNOLOGY PLOS ONE   
## 26 26   
## SCIENTIFIC REPORTS JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY   
## 20 20   
## NUTRIENTS   
## 15

Seguindo este tipo de prerrogativa, é possível criar vários tipos de redes bipartidas, por exemplo, uma rede de autores ou de citações, veja os exemplos abaixo:

A <- cocMatrix(M, Field = "CR", sep = ". ")  
  
A <- cocMatrix(M, Field = "AU", sep = ";")

Entretanto se a intenção for criar uma rede do país, você precisa extrair essas informações do atributo de afiliação usando a função “metaTagExtraction”, uma vez que os países dos autores não é um elemento padrão do quadro de dados bibliográficos.

M <- metaTagExtraction(M, Field = "AU\_CO", sep = ";")  
A <- cocMatrix(M, Field = "AU\_CO", sep = ";")

Para criar uma rede de palavras-chave do autor ou uma rede de palavras-chaves Plus, é necessário utilizar as funções abaixo:

A <- cocMatrix(M, Field = "DE", sep = ";")  
  
A <- cocMatrix(M, Field = "ID", sep = ";")

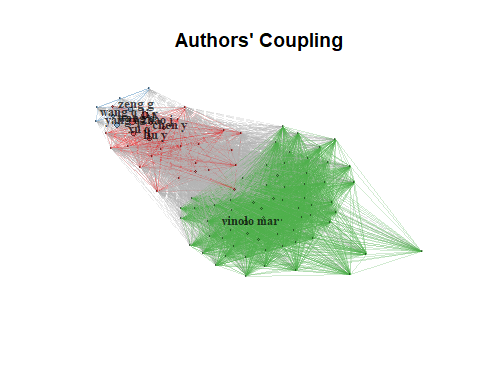
##### - Acoplamento bibliográfico

A força de acoplamento de dois artigos é definida pelo número de referências que os artigos têm em comum, isto é, dois artigos poderiam ser bibliograficamente acoplados se pelo menos uma fonte citada aparecesse nas referências bibliográficas de ambos os artigos. A função “biblioNetwork” consegue calcular a partir de um quadro de dados bibliográficos, as redes de acoplamento mais utilizadas, como Autores, Fontes e Países, para isso usa-se o código a seguir:

NetMatrix <- biblioNetwork(M, analysis = "coupling", network = "references", sep = ". ")

Com base no exposto os artigos com apenas algumas referências tenderiam a ser acoplados bibliograficamente de uma forma mais fraca que os outros se a força de acoplamento fosse medida com base na quantidade de referências que esses artigos têm em comum. Indicando assim que poderia ser mais fácil alterar para uma medida relativa de acoplamento bibliográfico. Para isso temos a função “normalizeSimilarity”, que calcula a força de associação, inclusão, similaridade de Jaccard ou Salton entre os vértices de uma rede. “normalizeSimilaridade” pode ser recuperada diretamente da função networkPlot () usando o argumento normalize.

NetMatrix <- biblioNetwork(M, analysis = "coupling", network = "authors", sep = ";")  
  
net=networkPlot(NetMatrix, normalize = "salton", weighted=NULL, n = 100, Title = "Authors' Coupling", type = "fruchterman", size=5,size.cex=T,remove.multiple=TRUE,labelsize=0.8,label.n=10,label.cex=F)



##### - Co-citação bibliográfica

Uma rede de co-citação também pode ser criada a partir da função “biblioNetwork”. A co-citação pode ser considerada como uma contrapartida do acoplamento bibliográfico uma vez que ela relaciona a co-citação de dois artigos quando ambos são citados em uma terceira publicação.

NetMatrix <- biblioNetwork(M, analysis = "co-citation", network = "references", sep = ". ")

##### - Colaboração bibliográfica

A criação de uma rede de colaboração científica também depende da função “biblio1network”, porém esta é uma rede na qual os nós são autores e os links são co-autoriais, pois a última é uma das formas mais bem documentadas de colaboração científica. Calcula-se este tipo de rede a partir do seguinte código:

NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "authors", sep = ";")

Existe a possibilidade de criar também uma rede de colaboração entre países com o seguinte código:

NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "countries", sep = ";")

#### Análise descritiva das características do gráfico de rede

Para calcular as várias estatísticas de resumo podemos utilizar a função “networkStat”, isto é, partindo de uma matriz bibliográfica podemos computar dois grupo de medidas descritivas que serão mais exploradas adiante.

NetMatrix <- biblioNetwork(M, analysis = "co-occurrences", network = "keywords", sep = ";")  
netstat <- networkStat(NetMatrix)

##### - As estatísticas resumidas da rede

Uma das medidas descritivas analisadas são as estatísticas resumidas de rede, as quais permitem descrever as seguintes propriedades estruturais de uma rede: Tamanho (número de vértices que compõe a rede); Densidade (proporção das bordas atuais de todas as bordas possíveis na rede); Transitividade (proporção de triângulos para triplos conectados); Diâmetro (distância do caminho mais curto entre dois nós na rede); Distribuição de graus (distribuição cumulativa de graus de vértices); Centralização de graus (grau normalizado da rede global); Centralização de proximidade (inverso normalizado do comprimento do caminho mais curto entre dois nós médio do vértice para outros na rede); Centralização do vetor próprio (primeiro autovetor da matriz gráfica); Centralização de proximidade (número normalizado da distância de nós que passam pelo vértice) e Comprimento médio do caminho (média da menor distância entre cada par de vértices na rede).

names(netstat$network)

## [1] "networkSize" "networkDensity" "networkTransitivity" "networkDiameter"   
## [5] "networkDegreeDist" "networkCentrDegree" "networkCentrCloseness" "networkCentrEigen"   
## [9] "networkCentrbetweenness" "NetworkAverPathLeng"

##### - Principais índices de centralidade e prestígio do vértices

Outro tipo de medida descritiva são os principais índices de centralidade e prestígio dos vértices, na qual é possível identificar os vértices mais importantes em uma rede e a propensão de dois vértices que estão conectados estarem conectados a um terceiro vértice. As estatísticas, no nível do vértice, retornadas pelo networkStat são: Centralidade do Grau; Centralidade de proximidade (quantos passos são necessários para acessar todos os outros vértices); Centralidade do autovetor (medida de estar bem conectado ao bem conectado); Centralidade da intermediação (mede a corretagem ou o potencial de gatekeeping, o número de caminhos mais curtos entre os vértices que passam por um determinado vértice); Pontuação de PageRank (aproxima a probabilidade de que qualquer mensagem chegue a um determinado vértice); Hub Score (estima o valor dos links que saem do vértice); Pontuação de Autoridade (esclarece que um vértice tem alta autoridade quando está vinculado por muitos outros vértices que estão conectando muitos outros vértices) e Classificação de Vértices (classificação obtida como uma combinação ponderada linear das medidas de vértice de centralidade e prestígio).

names(netstat$vertex)

## NULL

Se for o caso de uma análise mais resumida dos principais resultados da função “networkStat”, usa-se o resumo da função genérica (“summary”), desta forma será exibido as principais informações sobre a rede e a descrição do vértice através de várias tabelas. Além do mais é possível selecionar o número de linhas a serem exibidos durante a leitura das tabelas geradas apenas colocando o argumento “k” no código.

summary(netstat, k=10)

##   
##   
## Main statistics about the network  
##   
## Size 7301   
## Density 0.016   
## Transitivity 0.133   
## Diameter 4   
## Degree Centralization 0.845   
## Average path length 2.022   
##

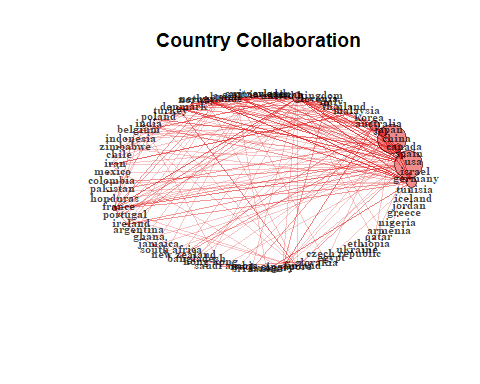
#### Visualizando redes bibliográficas

As redes bibliográficas podem ser visualizadas ou modeladas graficamente baseadas na função escolhida e logicamente com o que quer ser analisado, determinando deste modo o tipo de rede bibliográfica que será criada.

##### - Colaboração Científica entre Países

Para criar e plotar uma rede de colaboração científica entre diferentes países utiliza-se o seguinte código:

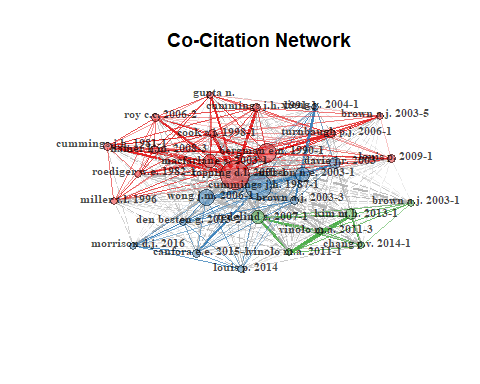
M <- metaTagExtraction(M, Field = "AU\_CO", sep = ";")  
NetMatrix <- biblioNetwork(M, analysis = "collaboration", network = "countries", sep = ";")  
  
net=networkPlot(NetMatrix, n = dim(NetMatrix)[1], Title = "Country Collaboration", type = "circle", size=TRUE, remove.multiple=FALSE,labelsize=0.7,cluster="none")



##### - Rede de co-citação

No caso de criar e plotar uma rede de co-citação, precisamos do seguinte código:

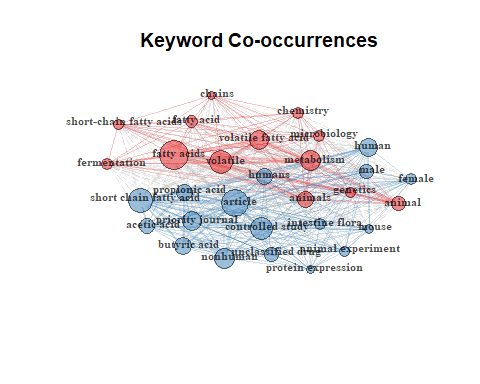
NetMatrix <- biblioNetwork(M, analysis = "co-citation", network = "references", sep = ";")  
  
net=networkPlot(NetMatrix, n = 30, Title = "Co-Citation Network", type = "fruchterman", size=T, remove.multiple=FALSE, labelsize=0.7,edgesize = 5)



##### - Co-ocorrências de palavras-chave

Por fim, para criar e plotar uma rede de co-ocorrências de palavras-chave usa-se o seguinte código:

NetMatrix <- biblioNetwork(M, analysis = "co-occurrences", network = "keywords", sep = ";")  
  
net=networkPlot(NetMatrix, normalize="association", weighted=T, n = 30, Title = "Keyword Co-occurrences", type = "fruchterman", size=T,edgesize = 5,labelsize=0.7)

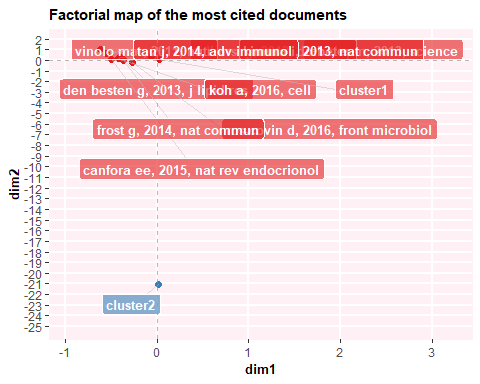
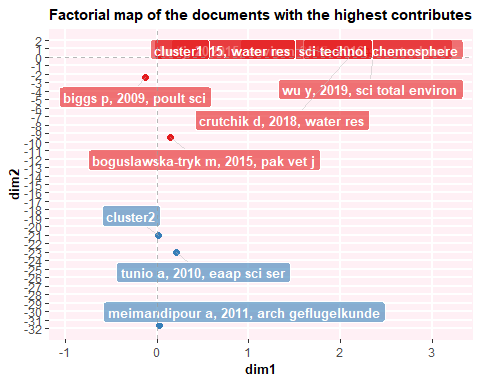
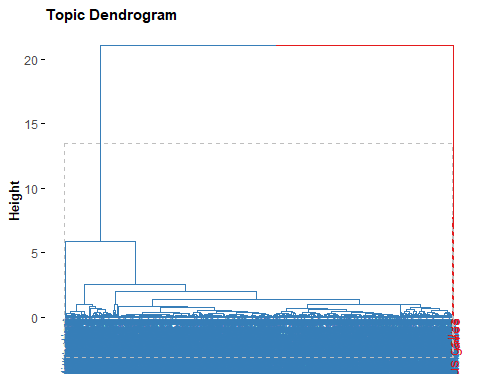
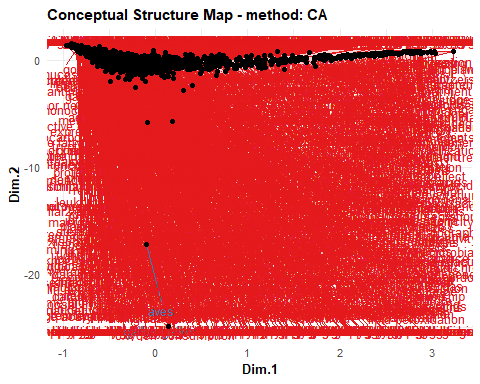


#### Análise Co-Word: A estrutura conceitual de um campo

A análise conjunta tem o propósito de mapear a estrutura conceitual de um framework usando a palavra co-ocorrências em uma base de dados bibliográficos. Ela pode ser realizada através de técnicas de redução de dimensionalidade, como Escala Multidimensional (MDS), Análise de Correspondência (AC) ou Análise de Correspondência Múltipla (ACM). Utilizando a função “conceptualStructure” podemos executar uma CA ou MCA para criar uma estrutura conceitual do campo e o K-means clustering para identificar clusters de documentos que expressam conceitos comuns, no final os resultados serão plotados em um mapa bidimensional como o exemplo a seguir.

CS <- conceptualStructure(M,field="ID", method="CA", minDegree=4, k.max=8, stemming=FALSE, labelsize=10, documents=10)

## Warning in data.frame(xmin = unlist(xleft), ymin = unlist(ybottom), xmax = unlist(xright), : row names were found from a  
## short variable and have been discarded



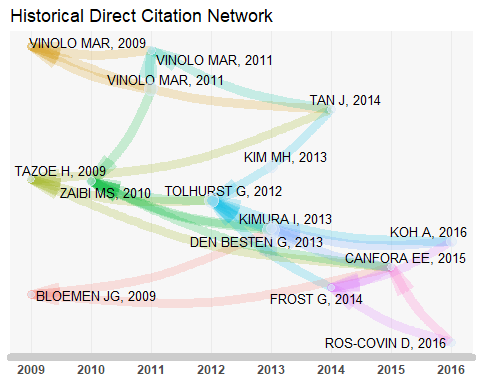
#### Rede Histórica de Citação Direta

Quando queremos representar um mapa de rede cronológica das citações diretas mais relevantes resultantes de uma base de dados bibliográficos, devemos criar um mapa historiográfico. Para isso basta usar a função “hitsNetwork” seguida de um código que use a função “histPlot” para a plotagem dos dados.

options(width=130)  
histResults <- histNetwork(M, min.citations = 10, sep = ";")

## Articles analysed 100   
## Articles analysed 200   
## Articles analysed 300   
## Articles analysed 396

net <- histPlot(histResults, n=15, size = 20, labelsize=10, size.cex=TRUE, arrowsize = 0.5, color = TRUE)



##   
## Legend  
##   
## Paper DOI Year LCS GCS  
## 2009 - 2 BLOEMEN JG, 2009, CLIN NUTR 10.1016/J.CLNU.2009.05.011 2009 29 89  
## 2009 - 11 VINOLO MAR, 2009, CLIN SCI 10.1042/CS20080642 2009 26 69  
## 2009 - 13 TAZOE H, 2009, BIOMED RES (JAPAN) 10.2220/BIOMEDRES.30.149 2009 30 157  
## 2010 - 57 ZAIBI MS, 2010, FEBS LETT 10.1016/J.FEBSLET.2010.04.027 2010 34 154  
## 2011 - 81 VINOLO MAR, 2011, J NUTR BIOCHEM 10.1016/J.JNUTBIO.2010.07.009 2011 33 155  
## 2011 - 99 VINOLO MAR, 2011, NUTR 10.3390/NU3100858 2011 50 370  
## 2012 - 125 TOLHURST G, 2012, DIABETES 10.2337/DB11-1019 2012 55 631  
## 2013 - 144 DEN BESTEN G, 2013, J LIPID RES 10.1194/JLR.R036012 2013 72 737  
## 2013 - 154 KIMURA I, 2013, NAT COMMUN 10.1038/NCOMMS2852 2013 38 395  
## 2013 - 171 KIM MH, 2013, GASTROENTEROLOGY 10.1053/J.GASTRO.2013.04.056 2013 44 207  
## 2014 - 185 FROST G, 2014, NAT COMMUN 10.1038/NCOMMS4611 2014 28 288  
## 2014 - 223 TAN J, 2014, ADV IMMUNOL 10.1016/B978-0-12-800100-4.00003-9 2014 33 288  
## 2015 - 264 CANFORA EE, 2015, NAT REV ENDOCRIONOL 10.1038/NRENDO.2015.128 2015 31 316  
## 2016 - 308 KOH A, 2016, CELL 10.1016/J.CELL.2016.05.041 2016 37 396  
## 2016 - 326 ROS-COVIN D, 2016, FRONT MICROBIOL 10.3389/FMICB.2016.00185 2016 32 303

## Respostas encontradas

Ao final do trabalho concluí-se que o pacote “bibliometrix” formnece ferramentas práticas e eficazes para uma análise bibliográfica, com as funções existentes em tal pacote é possível criar e plotar dados em formas de tabela, gráficos e redes de conexões, onde podem ser analisados os mais diversos parâmetros que estão envolvidos no banco de dados bibliográficos. Com o desenvolvimento do script que abrange os códigos baseados no pacote “bibliometrix” desnvolveu-se uma análise de dados embasada no tema Short Chain Fatty Acids, no qual encontraram-se informações relevantes e essenciais para o referenciamento bibliográfico de uma produção científica, por exemplo, descobriu-se que no ano de 2018 houve o maior índice de publicação (157) de artigos científicos. Apesar da China ser o país com o mais alto nível de publicação, grande parte dos artigos mais citados são norte-americanos. Li foi o autor mais produtivo encontrado e o mais citado dentro do nosso banco de dados, entretanto Cummings foi o mais citado de forma geral. Dentre os resultados obtidos, nota-se que as revistas com maior número de publicações relacionadas ao tema eram a Biosource Technology e Plos One. Além dos resultados fornecidos foi possível criar redes de conexões entre autores, países de colaboração e palavras-chaves. Em suma, as possibilidades de trabalho são extremamente diversificadas, basta compreender a própria necessidade e executá-la.

## Dificuldades encontradas

Dentre as dificuldades encontradas no decorrer do desenvolvimento do trabalho, vale destacar o problema em gerar um arquivo no formato de doc quando a função de instalação de pacotes (install.packages) era utilizada, por isso, no script os códigos foram retirados. Outro ponto foi a configuração do Github e aprender a lidar com o funcionamento de tal ferramenta, a maior dificuldade foi enviar os commits para armazenamento no repositório, o que exigiu mais atenção e uma pesquisa em outras fontes sobre como funciona o Controle de Versão; no entanto não foi possível unir os branches criados através da função merge, o que exigiu a criação de outro repositório durante o desenvolvimento do trabalho. Além disso, entender as funções e os argumentos pertinentes a cada uma foi um desafio, modificá-las de modo que os gráficos e redes se tornassem legíveis e esteticamente melhores foi o maior obstáculo, entretanto nem sempre o objetivo foi alcançado.

## Bibliografia

Aria, M. & Cuccurullo, C. (2017) bibliometrix: Uma ferramenta R para análise abrangente de mapeamento científico, Journal of Informetrics, 11 (4), pp. 959-975, Elsevier.

Git: Guia Prático - Roger Dudler (<https://rogerdudler.github.io/git-guide/index.pt_BR.html>)

Git push - Página do manual (<file:///C:/Program%20Files/Git/mingw64/share/doc/git-doc/git-push.html>)

Uma breve introdução à bibliometrix - Massimo Aria e Corrado Cuccurullo (<https://cran.r-project.org/web/packages/bibliometrix/vignettes/bibliometrix-vignette.html>)

SCOPUS (<http://www.scopus.com>)