

Bibliometria

Elisa Ribeiro Miranda Antunes

26 de abril de 2019

Introdução

Para o trabalho final da disciplina de R para iniciantes, foi selecionada a proposta 5, sobre dados bibliográficos. Para a bibliometria, foi utilizada a palavra chave “plant metabolomics”.

Esta palavra chave se refere aos estudos de metabolômica de plantas, ou seja, os estudos dos metabólicos, primários e secundários, presentes no Reino Vegetal. A Metabolômica de plantas é uma área em ascensão, com ampla aplicação e permite: o melhor entendimento das reações da planta às condições ambientais, a detecção de compostos de interesse medicinal ou comercial, etc.

Recursos

O pacote utilizado para este projeto foi o Bibliometrix [1]. O pacote foi inicialmente instalado e em seguida carregado através dos seguintes códigos:

```
install.packages("bibliometrix", dependencies=TRUE, repos = "http://cran.us.r-project.org")  
library(bibliometrix)
```

Desenvolvimento

O projeto foi desenvolvido em várias etapas, das quais: Importação de Dados, Cálculo dos Dados Bibliométricos, Elaboração dos Gráficos e Elaboração da Rede de Co-citação.

Importação dos Dados

Os dados foram importados da plataforma Web of Science, foram encontrados 72 resultados para a palavra chave “plant metabolomics”

Os dados exportados foram salvos no Diretório do R associado ao Git e o seguinte código foi utilizado para a leitura dos dados:

```
WoS.plant.metab <- readFiles("C:/Users/elisa/Documents/R Directory/R_iniciantes-projeto/savedrecs.bib")
```

Os dados foram então convertidos ao formato data frame pelo seguinte código:

Cálculo dos Dados Bibliométricos

Após a importação, os dados bibliométricos principais foram calculados através do seguinte código:

```
bibl.plant.metab <- biblioAnalysis(plant.metab.df, sep = ";")
```

Após o cálculo dos dados bibliométricos, foram plotados os gráficos

Elaboração dos Gráficos

Os gráficos foram plotados com os 10 autores mais produtivos, os 10 países mais produtivos, produção científica anual, média de artigos citados por ano e média de citações por ano.

Para realização dos gráficos, inicialmente resumida toda a informação bibliométrica através da função “summary”. Os códigos utilizados para esta etapa foram:

```
options(width=100)  
sum.plant.metab <- summary(object = bibl.plant.metab, k = 10, pause = FALSE)
```

```

##
##
## Main Information about data
##
## Documents 72
## Sources (Journals, Books, etc.) 45
## Keywords Plus (ID) 335
## Author's Keywords (DE) 153
## Period 2003 - 2018
## Average citations per documents 60.26
##
## Authors 250
## Author Appearances 348
## Authors of single-authored documents 5
## Authors of multi-authored documents 245
## Single-authored documents 7
##
## Documents per Author 0.288
## Authors per Document 3.47
## Co-Authors per Documents 4.83
## Collaboration Index 3.77
##
## Document types
## ARTICLE 32
## CORRECTION 1
## EDITORIAL MATERIAL 4
## LETTER 1
## MEETING ABSTRACT 9
## REVIEW 24
## REVIEW, BOOK CHAPTER 1
##
##
## Annual Scientific Production
##
## Year Articles
## 2003 2
## 2004 1
## 2005 1
## 2006 6
## 2007 2
## 2008 7
## 2009 3
## 2010 6
## 2011 2
## 2012 7
## 2013 5
## 2014 7
## 2015 5
## 2016 9
## 2017 3
## 2018 6
##
## Annual Percentage Growth Rate 7.598962
##
##
## Most Productive Authors

```

##	Authors	Articles	Authors	Articles Fractionalized
## 1	SAITO K	7	SAITO K	3.537
## 2	HALL RD	6	HALL RD	2.375
## 3	FIEHN O	5	HEGEMAN AD	2.033
## 4	KOPKA J	5	BEALE MH	1.333
## 5	HEGEMAN AD	4	WARD JL	1.333
## 6	SUMNER LW	4	SCHRIPSEMA J	1.200
## 7	WOLFENDER JL	4	FERNIE AR	1.083
## 8	BEALE MH	3	KIM HK	1.083
## 9	DE VOS RCH	3	TRETHEWEY R	1.000
## 10	FERNIE AR	3	VERPOORTE R	0.833

##

##

Top manuscripts per citations

##

##	Paper	TC	TCperYear
## 1	SUMNER LW, 2003, PHYTOCHEMISTRY	614	38.4
## 2	DE VOS RCH, 2007, NAT PROTOC	395	32.9
## 3	FIEHN O, 2008, PLANT J	285	25.9
## 4	HALL RD, 2006, NEW PHYTOL	259	19.9
## 5	SCHAUER N, 2006, TRENDS PLANT SCI	215	16.5
## 6	TOLSTIKOV VV, 2003, ANAL CHEM	197	12.3
## 7	JENKINS H, 2004, NAT BIOTECHNOL	177	11.8
## 8	KIM HK, 2011, TRENDS BIOTECHNOL	174	21.8
## 9	SANCHEZ DH, 2008, PHYSIOL PLANT	160	14.5
## 10	KIM HK, 2010, PHYTOCHEM ANAL	125	13.9

##

##

Corresponding Author's Countries

##

##	Country	Articles	Freq	SCP	MCP	MCP_Ratio
## 1	GERMANY	13	0.2000	10	3	0.231
## 2	USA	9	0.1385	7	2	0.222
## 3	NETHERLANDS	8	0.1231	7	1	0.125
## 4	CHINA	7	0.1077	6	1	0.143
## 5	JAPAN	5	0.0769	5	0	0.000
## 6	UNITED KINGDOM	5	0.0769	2	3	0.600
## 7	BELGIUM	3	0.0462	3	0	0.000
## 8	SWITZERLAND	3	0.0462	2	1	0.333
## 9	BRAZIL	2	0.0308	2	0	0.000
## 10	PORTUGAL	2	0.0308	1	1	0.500

##

##

SCP: Single Country Publications

##

MCP: Multiple Country Publications

##

##

Total Citations per Country

##

##	Country	Total Citations	Average Article Citations
## 1	USA	1247	138.56
## 2	NETHERLANDS	1100	137.50
## 3	GERMANY	760	58.46
## 4	UNITED KINGDOM	422	84.40

## 5	SWITZERLAND	230	76.67
## 6	BRAZIL	158	79.00
## 7	BELGIUM	101	33.67
## 8	JAPAN	85	17.00
## 9	PORTUGAL	68	34.00
## 10	CHINA	59	8.43

##

##

Most Relevant Sources

##

	Sources	Articles
## 1	JOURNAL OF CHROMATOGRAPHY B-ANALYTICAL TECHNOLOGIES IN THE BIOMEDICAL AND LIFE SCIENCES	4
## 2	METABOLOMICS	4
## 3	PHARMACEUTICAL BIOLOGY	4
## 4	PHYTOCHEMICAL ANALYSIS	4
## 5	ANALYTICAL CHEMISTRY	3
## 6	PHYSIOLOGIA PLANTARUM	3
## 7	PLANT PHYSIOLOGY	3
## 8	PLANTA MEDICA	3
## 9	INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES	2
## 10	JOURNAL OF CHROMATOGRAPHY A	2

##

##

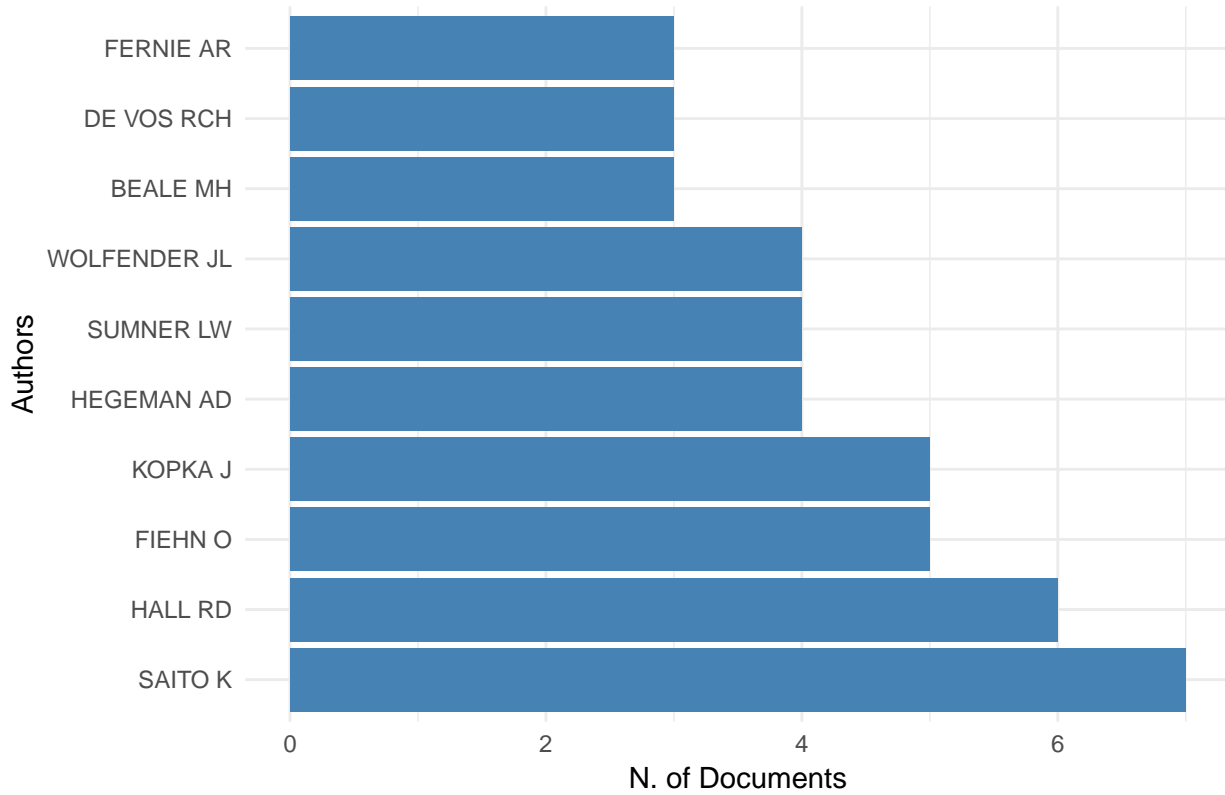
Most Relevant Keywords

##

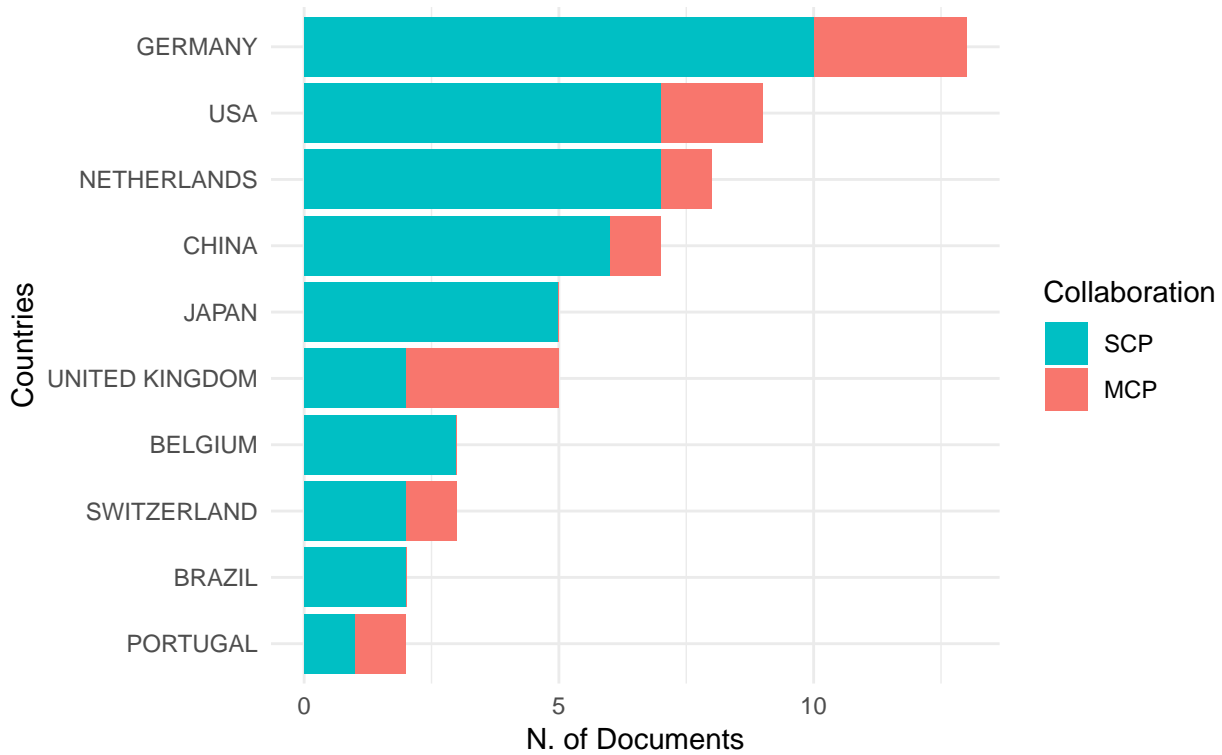
	Author Keywords (DE)	Articles	Keywords-Plus (ID)	Articles
## 1	METABOLOMICS	17	ARABIDOPSIS THALIANA	19
## 2	MASS SPECTROMETRY	14	FUNCTIONAL GENOMICS	16
## 3	PLANT METABOLOMICS	12	IDENTIFICATION	16
## 4	GC MS	6	MASS SPECTROMETRY	15
## 5	METABOLITE PROFILING	5	LIQUID CHROMATOGRAPHY	13
## 6	LC MS	4	METABOLITES	9
## 7	EXTRACTION	3	PERFORMANCE LIQUID CHROMATOGRAPHY	8
## 8	METABOLOME	3	SYSTEMS BIOLOGY	8
## 9	PLANTS	3	ARABIDOPSIS	7
## 10	ABIOTIC STRESS	2	GAS CHROMATOGRAPHY	6

```
plot(x = bibl.plant.metab, k = 10, pause = FALSE) #plot
```

Most productive Authors

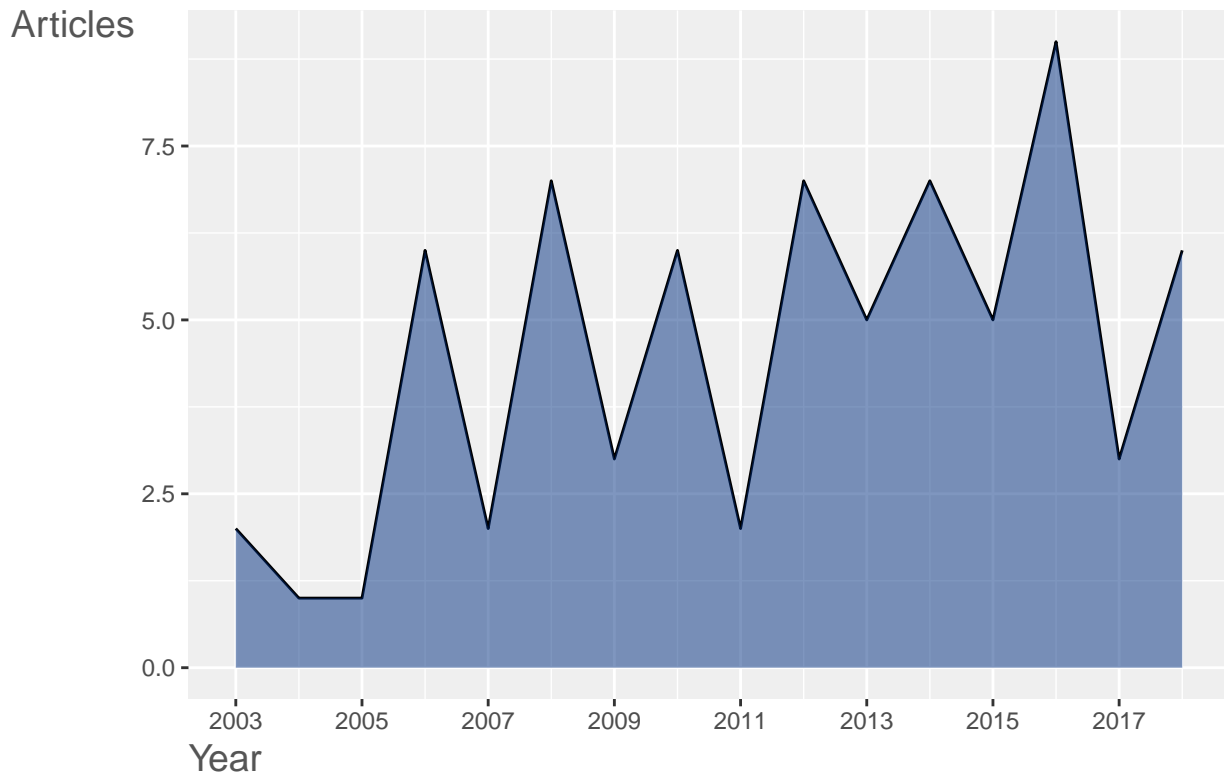


Most Productive Countries

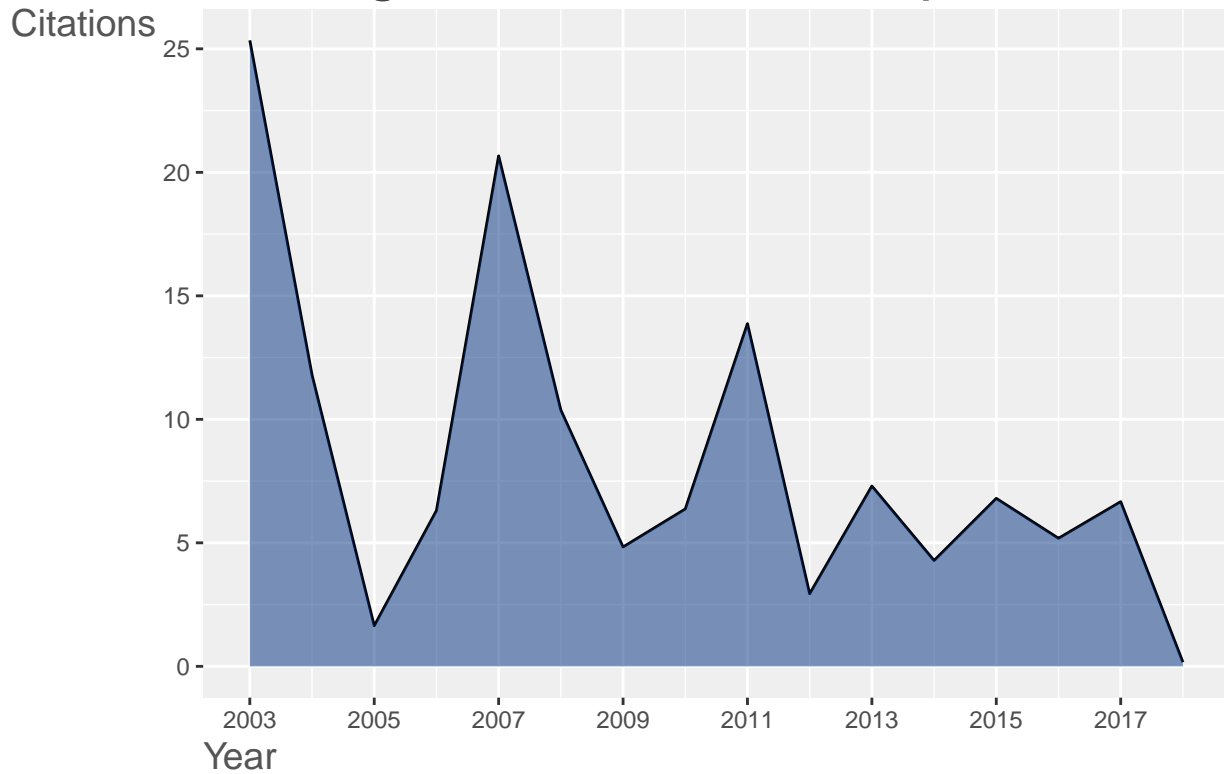


SCP: Single Country Publications, MCP: Multiple Country Publications

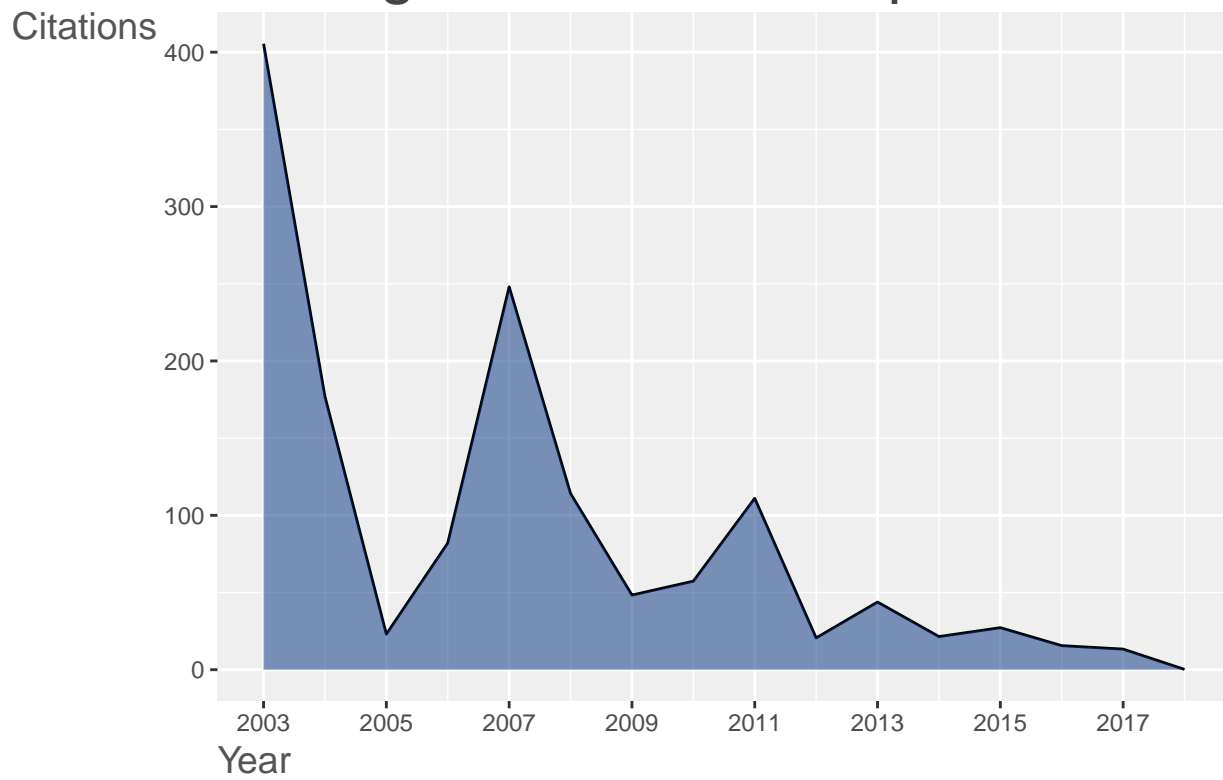
Annual Scientific Production



Average Article Citations per Year



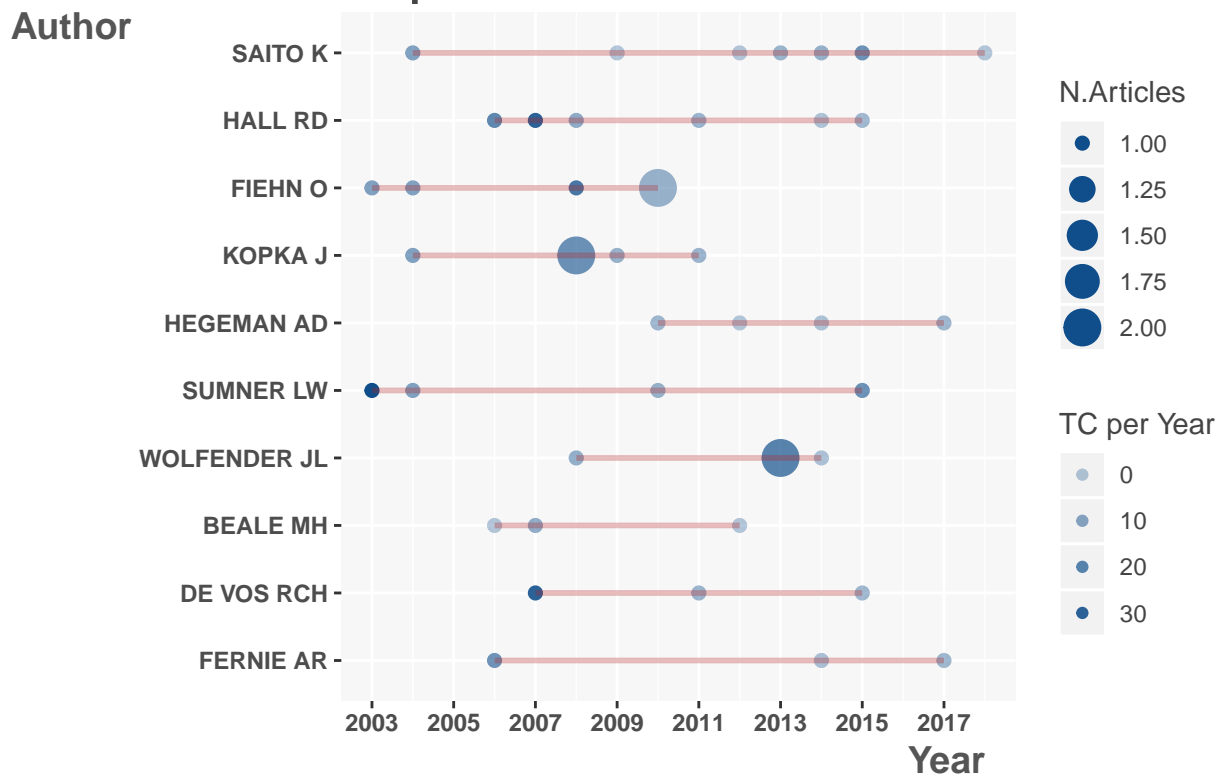
Average Total Citations per Year



Foi também plotado gráfico ilustrando a produção dos autores ao longo do tempo, através do seguinte código:

```
topAU <- authorProdOverTime(plant.metab.df, k = 10, graph = TRUE)
```

Top–Authors' Production over th



Elaboração da Rede de Co-citação

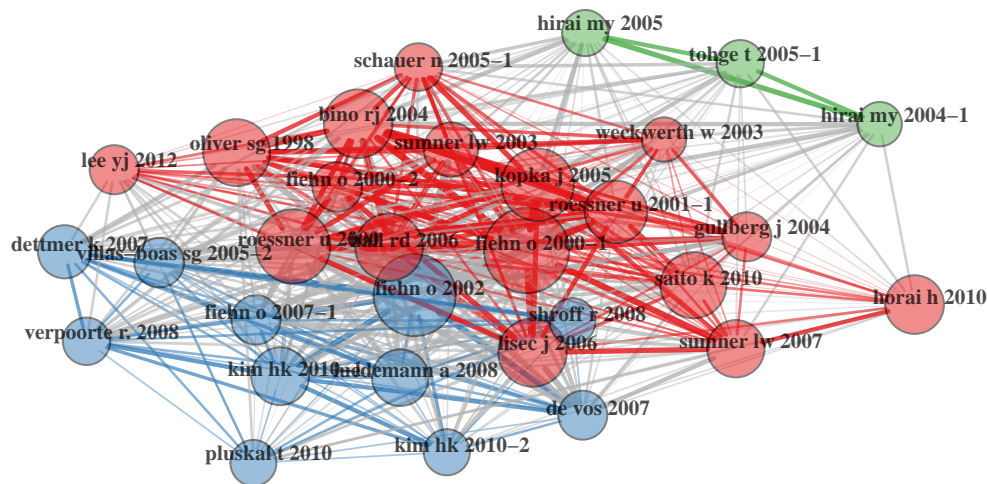
Para realização da rede de co-citação foi, inicialmente, formada a matrix através do seguinte código:

```
NetMatrix <- biblioNetwork(plant.metab.df, analysis = "co-citation",
                           network = "references", sep = ";")
```

A seguir, a rede foi plotada, utilizando o seguinte código:

```
net=networkPlot(NetMatrix, n = 30, Title = "Co-Citation Network", type = "fruchterman",
               size=T, remove.multiple=FALSE, labels=0.7, edgesize = 5)
```


Co-Citation Network



Respostas Encontradas

Através das análises realizadas foi observado que o autor mais produtivo na área é Saito K, entretanto, sua produção não aumentou de 2003 para 2017, de acordo com o gráfico de “Top Author’s production over the years”

Com relação a produção por país, a Alemanha se posicionou em primeiro lugar no ranking de produção na área de “plant metabolomics” e aproximadamente 1/5 desta produção possui colaboração múltipla entre países. Em contrapartida, Estados Unidos, em 6º lugar no ranking possui mais da metade de sua produção derivada de colaboração entre países. O Brasil, por sua vez, está em 9º lugar na área de “plant metabolomics”, com nenhuma produção colaborativa, da mesma forma que a Bélgica, de acordo com o gráfico.

De acordo com o gráfico de produção científica anual a área de “plant metabolomics” cresceu de 2003 para 2017. Apesar da presença de altos e baixos na produção a tendência foi de aumento, com pico produtivo em 2016. Porém a citação média de artigos por ano diminuiu, com pico de citação em 2003 e em constante queda até 2017, da mesma forma que a média de citação total por ano.

Dificuldades encontradas

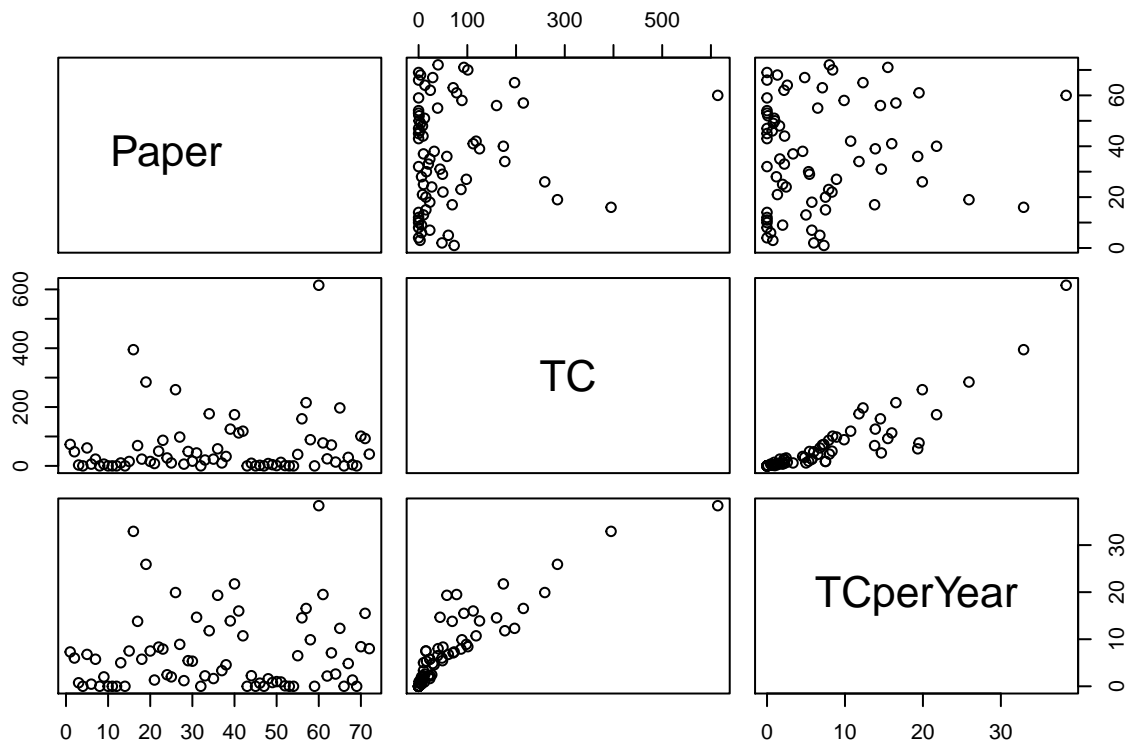
Inicialmente foram encontradas dificuldades com relação à geração do PDF através do RMarkdown mas após procurar resoluções em sites relacionados ao R o problema foi resolvido. A dificuldade estava relacionada a geração do PDF com códigos do pacote bibliometrix. Era necessário adicionar, na linha de código que instalava o pacote, o local de onde a instalação ocorreria (repos = “<http://cran.us.r-project.org>”)

Foram também encontradas dificuldades com relação às margens do documento PDF do RMarkdown. O problema foi contornado quebrando as linhas e códigos e foi solucionado após consulta em páginas de ajuda do R e adicionando o parâmetro “geometry” no topo do documento foi possível personalizar as margens do arquivo.

Além disso, houveram dificuldades, não solucionadas, em plotar gráficos com os dados dos artigos mais citados (MostCitedPapers), no modelo dos gráficos de “Most Productive Authors” e “Most Productive Countries”. A ideia era plotar um gráfico com os 10 papers mais citados e a quantidade de citações para cada, mostrando o ranking. Este gráfico conteria o nome do autor, ano e revista e quantas citações tal artigo obteve.

Alguns dos códigos que foram utilizados na tentativa de realizar tal gráfico foram:

```
plot(bibl.plant.metab[["MostCitedPapers"]]) #que retorna vários gráficos de pontos
```



```
plot(sum.plant.metab[["MostCitedPapers"]]) #que retorna um grafico vazio e mensagens de erro
```

```
plot(y = sum.plant.metab[["MostCitedPapers$Paper"]],  
     x = sum.plant.metab[["MostCitedPapers$TC"]]) #que também retorna apenas mensagem de erro
```

Bibliografia

[1] Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11(4), pp 959-975, Elsevier.