Bibliométrie

Eric Marcon

8 janvier 2021

Résumé

Utilisation de Google Scholar et de Scopus avec R pour analyser les publications d'une structure ou d'un auteur.

Table des matières


```
##
## The downloaded binary packages are in
##  /var/folders/24/8k48jl6d249_n_qfxwsl6xvm0000gn/T//RtmphmaILY/downloaded_packages
```

1 Google Scholar

Le package *scholar* permet d'accéder à l'API de Google Scholar. L'objectif est d'analyser la production d'un auteur (ou d'une structure) disposant d'un identifiant, donc d'une page, Google Scholar.

Le paramètre de base est l'identifiant de l'auteur :

```
AuthorID <- "4iLBmbUAAAAJ" # Eric Marcon
# AuthorID <- "8XqZyDUAAAAJ" # UMR EcoFoG
```

La vignette du package fournit la majorité du code utile.

```
vignette(topic = "scholar", package = "scholar")
```

1.1 Information sur l'auteur

La fonction `get_profile` retourne une liste avec les informations sur l'auteur.

```
library("scholar")
get_profile(AuthorID)
```

```
## $id
## [1] "4iLBmbUAAAAJ"
##
## $name
## [1] "Eric Marcon"
##
## $affiliation
## [1] "UMR Amap, AgroParisTech"
##
## $total_cites
## [1] 1798
##
## $h_index
## [1] 19
##
## $i10_index
## [1] 25
##
## $fields
## [1] "verified email at agroparistech.fr - homepage"
##
## $homepage
## [1] "https://ericmarcon.github.io/"
##
## $coauthors
##  [1] "Puech Florence"
##  [2] "Bruno Herault"
##  [3] "Gabriel Lang"
##  [4] "Chris Baraloto"
##  [5] "Sabrina Coste"
##  [6] "Heidy Schimann"
```

```
##  [7] "Céline Leroy"
##  [8] "Sandrine Pavoine"
##  [9] "Jerome Chave"
## [10] "Lilian Blanc"
## [11] "Jingjing Liang ()"
## [12] "Zhiyi Zhang"
## [13] "Vivien Rossi"
## [14] "Ivan Scotti"
## [15] "Céline Born"
## [16] "Carlo Ricotta"
## [17] "Cecile Richard-Hansen"
## [18] "Guitet"
## [19] "Michael Grabchak"
## [20] "François Morneau"
```

1.2 Liste des publications

La fonction `get_publications` retourne un dataframe contenant toutes les publications. Les colonnes contiennent le titre, la liste des auteurs (séparés par des virgules), le nom du journal, la pagination (sous la forme *Volume (numéro), pages*), le nombre de citations et les années correspondantes (sous la forme de vecteurs), et deux identifiants internes de la publication (`cid` et `pubid`).

```
Publications <- get_publications(AuthorID)
colnames(Publications)
```

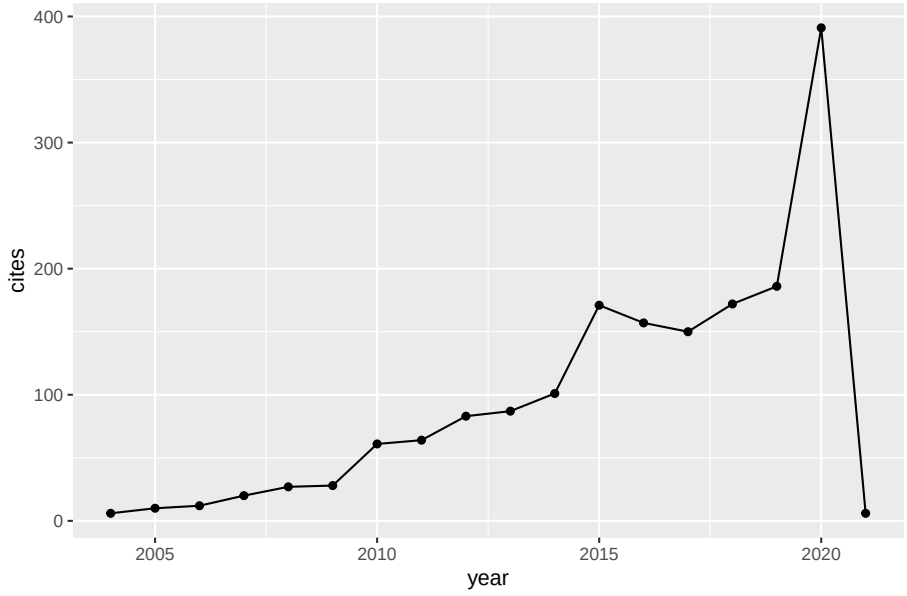
```
## [1] "title"   "author"  "journal" "number"
## [5] "cites"   "year"    "cid"     "pubid"
```

1.3 Citations par année

Evolution du nombre de citations d'un auteur :

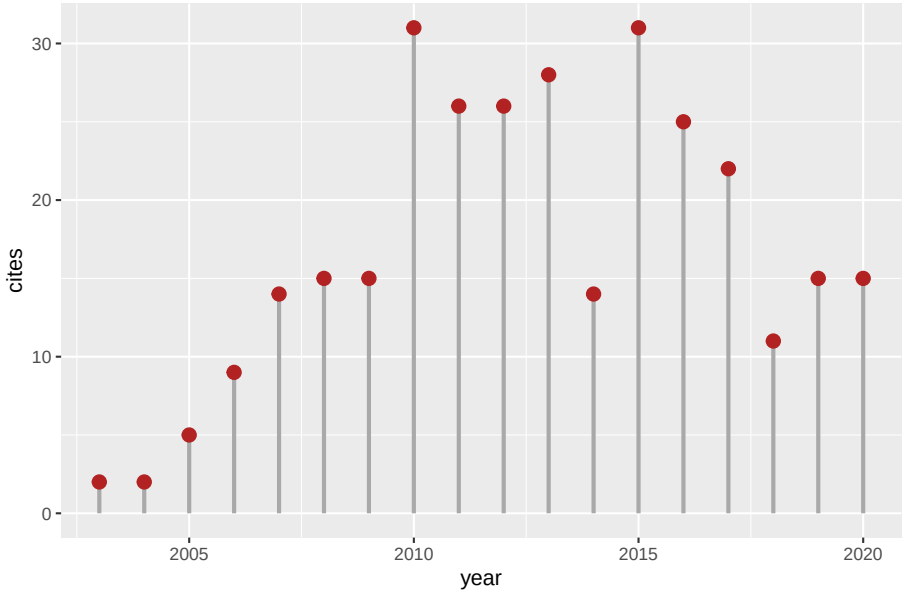
```
library("ggplot2")

get_citation_history(AuthorID) %>%
  ggplot(aes(x = year, y = cites)) +
    geom_line() +
    geom_point() +
    labs(caption= format(Sys.time(), "%Y-%m-%d %H:%M (GMT %Z)"))
```



Suivi d'un article en particulier (le plus cité : les articles sont classés par ordre décroissant du nombre de citations) :

```
NumArticle <- 1
Reference <- with(Publications[NumArticle, ],
  paste(author, " (", year, ") ", journal, ". ", number, sep=""))
get_article_cite_history(AuthorID, Publications$pubid[NumArticle]) %>%
  ggplot(aes(year, cites)) +
    geom_segment(aes(xend = year, yend = 0), size=1, color='darkgrey') +
    geom_point(size=3, color='firebrick') +
    labs(caption = Reference)
```

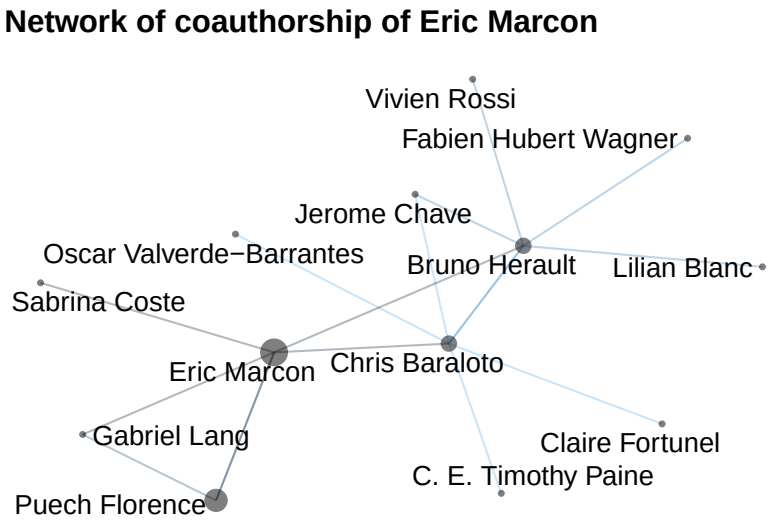


E Marcon, F Puech (2003) Journal of Economic Geography. 3 (4), 409−428

1.4 Réseau d'auteurs

`get_coauthors` retourne un dataframe contenant les coauteurs déclarés par l'auteur sur sa page et leurs coauteurs. La profondeur `n_deep` du graphe permet d'augmenter le nombre de niveaux de coauteurs mais ne peut pas être mise à 0 pour obtenir seulement les coauteurs directs. Les valeurs par défaut sont 5 coauteurs et une profondeur de 1.

```
get_coauthors(AuthorID, n_coauthors = 5, n_deep=1) %>%
  # Bug in get_coauthors
  filter(substr(coauthors, start = 1, stop = 8) != "Sort By ") %>%
  plot_coauthors
```

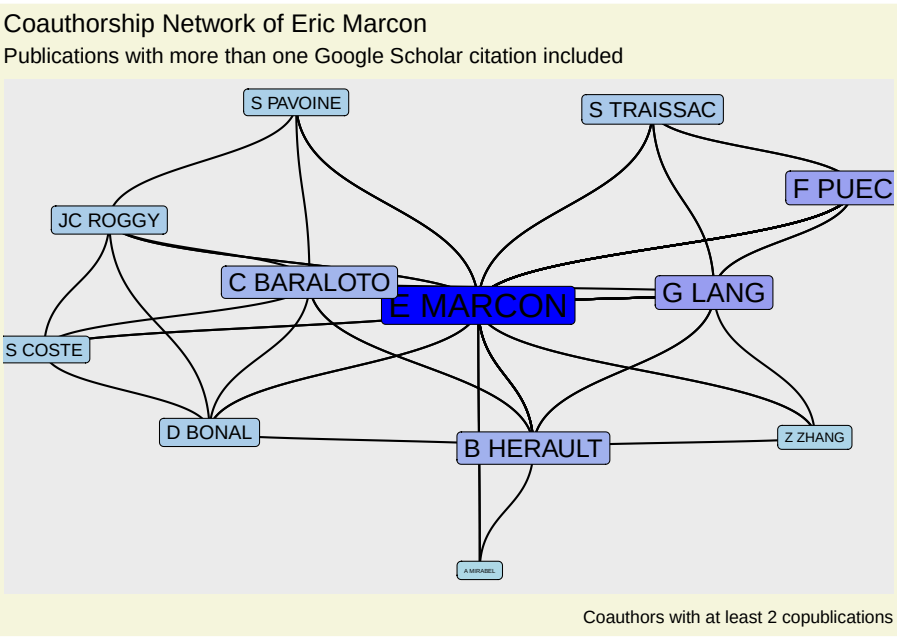


Les coauteurs réels, définis par le nombre de publications écrites en commun, sont à rechercher dans le tableau des publications.

```
# Paramètres
MinCopublications <- 2
MaxCoauteurs <- 100

library("magrittr")
# Vecteur des coauteurs de publications, sans accents
get_publications(AuthorID) %>%
  mutate(AuthorsASCII=iconv(author, from="UTF-8", to="ASCII//TRANSLIT")) %$%
  AuthorsASCII %>%
  # Suppression des accents transformés en ' sur MacOS
  str_replace("'", "") ->
  AuthorsASCII
# Auteurs uniques
AuthorsASCII %>%
  paste(collapse=", ") %>%
  str_split(pattern=", ") %>%
  unlist %>%
  # Uniformisation de la casse
  str_to_upper() %>%
  unique ->
  UniqueAuthors
# Elimination de ... (= et al.)
UniqueAuthors <- UniqueAuthors[UniqueAuthors != "..."]
# Matrice d'autorat: une ligne par article, auteurs en colonnes, valeurs logiques
PaperAuthoredBy <- sapply(UniqueAuthors, function(Author) str_detect(str_to_upper(AuthorsASCII), Author))
# Filtrage des auteurs
tibble(Author=UniqueAuthors, NbPapers=colSums(PaperAuthoredBy)) %>%
  filter(NbPapers >= MinCopublications) %>%
  arrange(desc(NbPapers)) %>%
  slice(1:MaxCoauteurs) ->
  NbPapersPerAuthor
# Recalcul de la matrice d'autorat réduite
PaperAuthoredBy <- sapply(NbPapersPerAuthor$Author,
                          function(Author) str_detect(str_to_upper(AuthorsASCII), Author))
```

```r
# Matrice d'adjacence
adjacencyMatrix <- t(PaperAuthoredBy) %*% PaperAuthoredBy
# Graphe d'adjacence
# (https://paulvanderlaken.com/2017/10/31/network-visualization-with-igraph-and-ggraph/)
library("igraph")
g <- graph.adjacency(adjacencyMatrix, mode = "undirected", diag = FALSE)
V(g)$Degree <- degree(g, mode = 'in') # Nombre de liens
V(g)$Name <- NbPapersPerAuthor$Author # Etiquettes des noeuds
# Figure
library("ggraph")
ggraph(g, layout = "auto") +
  geom_edge_diagonal(alpha = 1, label_colour = "blue") +
  geom_node_label(aes(label = Name, size = log(Degree), fill = Degree)) +
  scale_fill_gradient(high = "blue", low = "lightblue") +
  theme(
    plot.background = element_rect(fill = "beige"),
    panel.border = element_blank(),
    panel.grid = element_blank(),
    legend.position = "none",
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank()) +
  labs(title = paste("Coauthorship Network of", get_profile(AuthorID)$name),
       subtitle = "Publications with more than one Google Scholar citation included",
       caption = paste("Coauthors with at least", MinCopublications, "copublications"))
```



Nombres de publications :

```r
knitr::kable(NbPapersPerAuthor, caption="Nombre de documents par auteur",
             longtable = FALSE, booktabs = TRUE) %>%
  kableExtra::kable_styling(bootstrap_options = "striped")
```

TABLE 1 : Nombre de documents par auteur

Author	NbPapers
E MARCON	47
F PUECH	13
G LANG	10
B HERAULT	6
C BARALOTO	5
S TRAISSAC	3
S PAVOINE	3
S COSTE	2
D BONAL	2
JC ROGGY	2
Z ZHANG	2
A MIRABEL	2

2 Scopus et Web of Science

Le package *bibliometrix* permet d'exploiter les données des bases de données commerciales majeures.

La vignette du package décrit l'ensemble de ses possibilités.

```
vignette(topic = "bibliometrix-vignette", package = "bibliometrix")
```

2.1 Lecture des données

Voir la première partie de la vignette. Sur le site de Scopus (utilisé en exemple), sélectionner les références utiles et les exporter dans un fichier Bibtex. L'export doit contenir tous les champs, y compris le résumé et les documents cités.

Le fichier est ensuite lu et converti :

```
library(bibliometrix)
# Fichier de données au format bibtex, exporté de Scopus
M <- convert2df("scopus.bib", dbsource="scopus", format="bibtex")
```

```
##
## Converting your scopus collection into a bibliographic dataframe
##
## Done!
##
##
## Generating affiliation field tag AU_UN from C1:  Done!
```

2.2 Analyses basiques

Les analyses de base sont retournées par la fonction `biblioAnalysis`. Le résultat est un objet de type `bibliometrix`. Les méthodes `summary` et `plot` renvoient tous les résultats à l'écran.

```
k <- 5 # Nombre d'auteurs à afficher
BA <- biblioAnalysis(M)
summary(BA, k)
```

```
##
##
## MAIN INFORMATION ABOUT DATA
##
##  Timespan                             2001 : 2020
##  Sources (Journals, Books, etc)       299
##  Documents                            859
##  Average years from publication       8.12
##  Average citations per documents      32.73
##  Average citations per year per doc   3.632
##  References                           42751
```

```
##
## DOCUMENT TYPES
##  article               793
##  book chapter          3
##  conference paper      19
##  data paper            2
##  editorial             1
##  erratum               4
##  letter                4
##  note                  4
##  review                28
##  short survey          1
##
## DOCUMENT CONTENTS
##  Keywords Plus (ID)                    5239
##  Author's Keywords (DE)                2629
##
## AUTHORS
##  Authors                               5279
##  Author Appearances                    11438
##  Authors of single-authored documents  7
##  Authors of multi-authored documents   5272
##
## AUTHORS COLLABORATION
##  Single-authored documents             7
##  Documents per Author                  0.163
##  Authors per Document                  6.15
##  Co-Authors per Documents              13.3
##  Collaboration Index                   6.19
##
##
## Annual Scientific Production
##
##  Year    Articles
##     2001        1
##     2002        4
##     2003       27
##     2004       18
##     2005       16
##     2006       21
##     2007       31
##     2008       26
##     2009       50
##     2010       76
##     2011       67
##     2012       69
```

```
##     2013       51
##     2014       50
##     2015       70
##     2016       61
##     2017       53
##     2018       53
##     2019       46
##     2020       69
##
## Annual Percentage Growth Rate 24.96303
##
##
## Most Productive Authors
##
##    Authors        Articles Authors        Articles Fractionalized
## 1      DEJEAN A        145     DEJEAN A                     26.75
## 2      BARALOTO C      106     BARALOTO C                   15.72
## 3      ORIVEL J         93     ORIVEL J                     15.04
## 4      HRAULT B         87     HRAULT B                     14.41
## 5      LEROY C          74     LEROY C                      11.73
## 6      BONAL D          73     CORBARA B                    11.03
## 7      CORBARA B        71     BONAL D                       8.51
## 8      CRGHINO R        56     CLAIR B                       8.37
## 9      CHAVE J          53     CRGHINO R                     8.34
## 10     STAHL C          45     ALMRAS T                      7.50
##
##
## Top manuscripts per citations
##
##                            Paper                                     DOI   TC TCperYear
## 1  PHILLIPS OL, 2009, SCIENCE            10.1126/science.1164033          1032      79.4
## 2  DAZ S, 2016, NATURE                   10.1038/nature16489               775     129.2
## 3  LUYSSAERT S, 2007, GLOBAL CHANGE BIOL 10.1111/j.1365-2486.2007.01439.x  635      42.3
## 4  TER STEEGE H, 2013, SCIENCE           10.1126/science.1243092           569      63.2
## 5  LIANG J, 2016, SCI                    10.1126/science.aaf8957           412      68.7
## 6  BRIENEN RJW, 2015, NATURE             10.1038/nature14283               410      58.6
## 7  MOUILLOT D, 2013, PLOS BIOL           10.1371/journal.pbio.1001569      399      44.3
## 8  SIEFERT A, 2015, ECOL LETT            10.1111/ele.12508                 354      50.6
## 9  KUNSTLER G, 2016, NATURE              10.1038/nature16476               323      53.8
## 10 PHILLIPS OL, 2010, NEW PHYTOL         10.1111/j.1469-8137.2010.03359.x  308      25.7
##
##
## Corresponding Author's Countries
##
##           Country Articles    Freq SCP MCP MCP_Ratio
## 1  FRANCE              392 0.71014 222 170     0.434
```

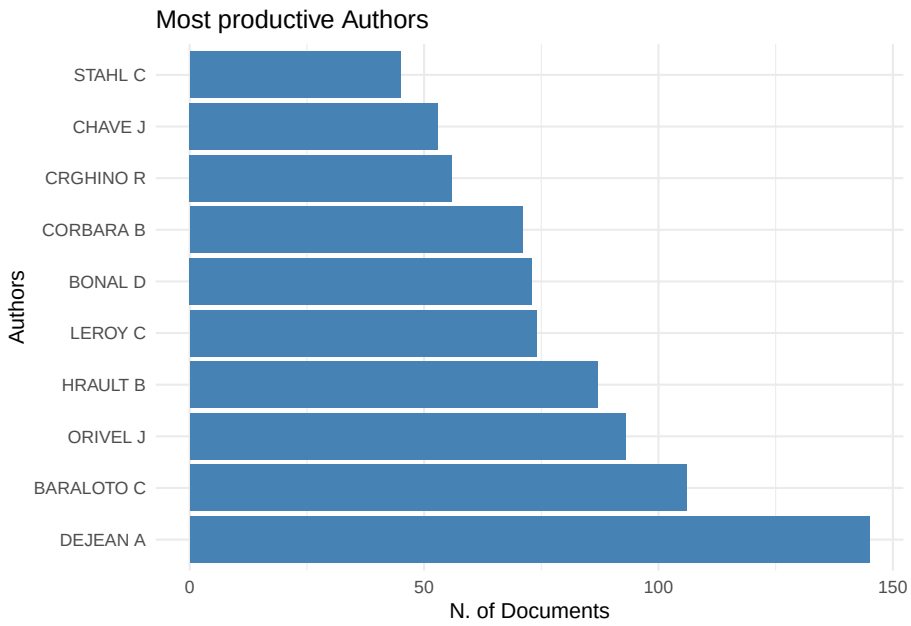
```
## 2  USA                  22 0.03986   3  19     0.864
## 3  UNITED KINGDOM       21 0.03804   0  21     1.000
## 4  BRAZIL               19 0.03442   0  19     1.000
## 5  GERMANY              14 0.02536   0  14     1.000
## 6  JAPAN                13 0.02355   4   9     0.692
## 7  BELGIUM              11 0.01993   0  11     1.000
## 8  CANADA                9 0.01630   1   8     0.889
## 9  NETHERLANDS           5 0.00906   0   5     1.000
## 10 SWITZERLAND           5 0.00906   0   5     1.000
##
##
## SCP: Single Country Publications
##
## MCP: Multiple Country Publications
##
##
## Total Citations per Country
##
##      Country      Total Citations Average Article Citations
## 1  FRANCE                  10633                       27.1
## 2  UNITED KINGDOM           2578                      122.8
## 3  BELGIUM                  1114                      101.3
## 4  USA                      1070                       48.6
## 5  ARGENTINA                 775                      775.0
## 6  NETHERLANDS               708                      141.6
## 7  BRAZIL                    492                       25.9
## 8  GERMANY                   416                       29.7
## 9  ITALY                     395                       98.8
## 10 JAPAN                     332                       25.5
##
##
## Most Relevant Sources
##
##                   Sources        Articles
## 1  PLOS ONE                            37
## 2  ANNALS OF FOREST SCIENCE            36
## 3  BIOTROPICA                          20
## 4  COMPTES RENDUS - BIOLOGIES          20
## 5  NEW PHYTOLOGIST                     15
## 6  SCIENTIFIC REPORTS                  15
## 7  ECOLOGY                             14
## 8  FOREST ECOLOGY AND MANAGEMENT       14
## 9  FUNCTIONAL ECOLOGY                  14
## 10 GLOBAL CHANGE BIOLOGY               14
##
##
```

```
## Most Relevant Keywords
##
##    Author Keywords (DE)      Articles Keywords-Plus (ID)     Articles
## 1        FRENCH GUIANA             93        FRENCH GUIANA         330
## 2        TROPICAL FOREST           31        ARTICLE               220
## 3        TROPICAL RAINFOREST       23        ANT                   174
## 4        FUNCTIONAL DIVERSITY      18        BIODIVERSITY          156
## 5        FUNCTIONAL TRAITS         17        ANIMALS               147
## 6        TENSION WOOD              17        TREE                  136
## 7        AMAZONIA                  16        ECOSYSTEM             135
## 8        AMAZON                    15        TROPICAL FOREST       135
## 9        BIODIVERSITY              15        ANIMAL                126
## 10       TROPICAL TREES            15        RAINFOREST            124
```

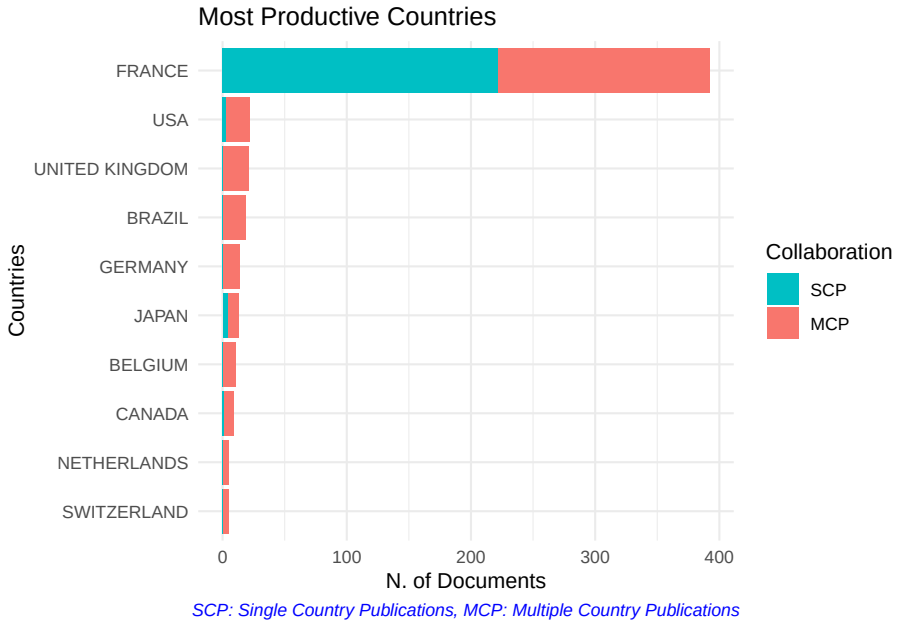
Pour les afficher séparément, il faut stocker le résultat dans une variable (qui est une liste) et appeler ensuite chacun de ses membres.

```
# plot(BA) renvoie tous les graphiques à la suite. Stocker.
BAP <- plot(BA)
```

```
# Graphiques disponibles
BAP$MostProdAuthors
```

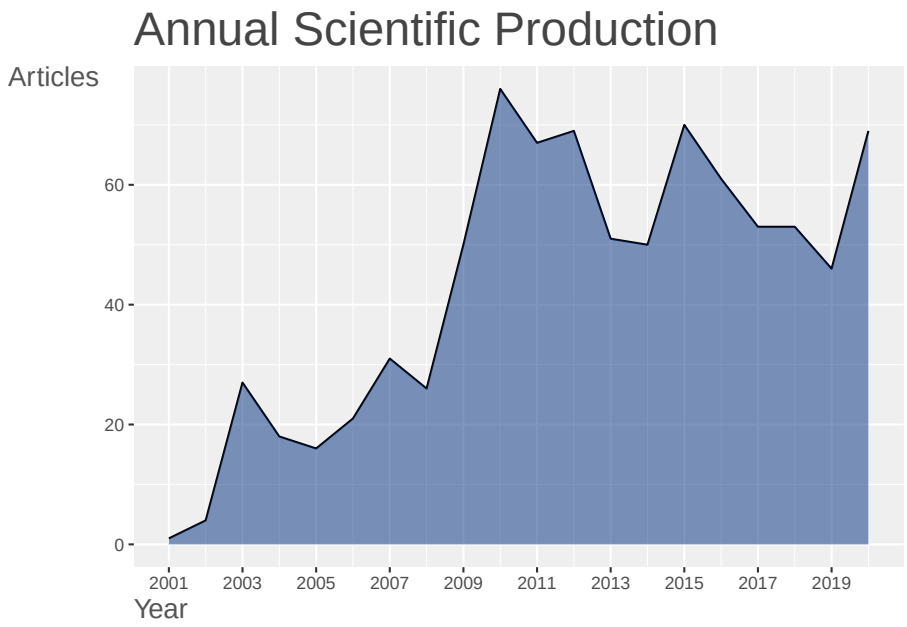



```
BAP$MostProdCountries
```

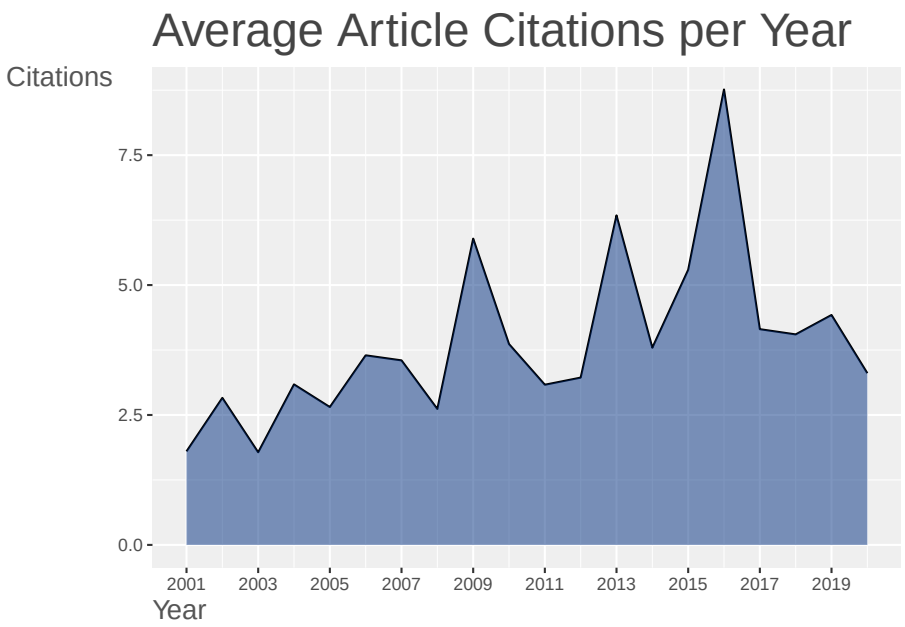


SCP: Single Country Publications, MCP: Multiple Country Publications

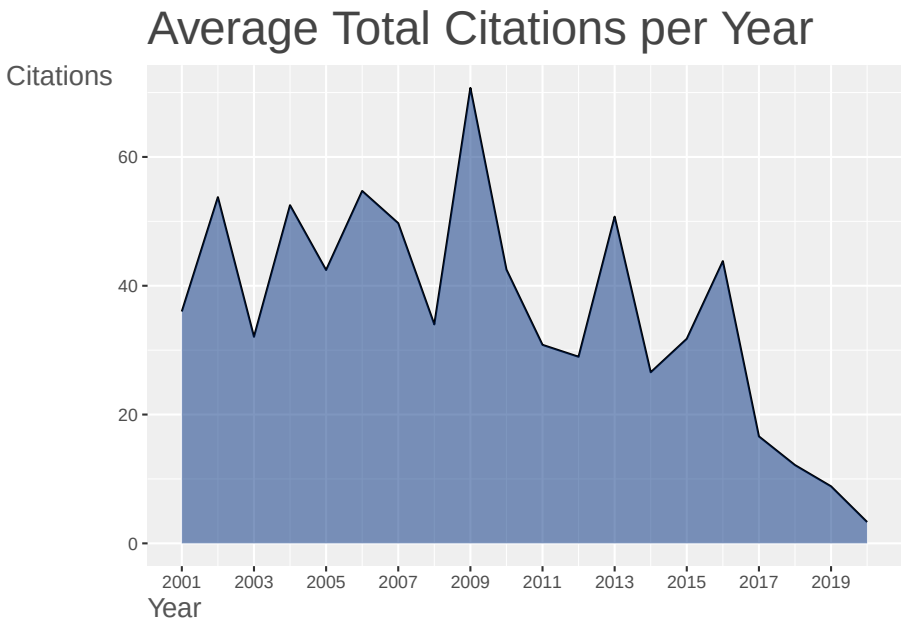
```
BAP$AnnualScientProd
```



```
BAP$AverArtCitperYear
```




```
BAP$AverTotCitperYear
```



2.3 h index

L'indice h peut être calculé par auteur ou source, et depuis un nombre d'années choisi.

Pour tous les auteurs :

```
Hindex(M, elements = dominance(BA)$Author, years=50)$H %>%
  arrange(desc(h_index))
```

```
##         Author h_index g_index  m_index   TC  NP PY_start
## 1   BARALOTO C      46      89 2.421053 7996 106     2003
## 2      BONAL D      36      73 1.714286 5829  73     2001
## 3      CHAVE J      32      53 1.684211 5813  53     2003
## 4     HRAULT B      31      58 2.066667 3578  87     2007
## 5     DEJEAN A      25      39 1.315789 2347 144     2003
## 6     ORIVEL J      23      35 1.533333 1632  93     2007
## 7      LEROY C      20      28 1.052632 1030  74     2003
## 8    CORBARA B      20      32 1.052632 1194  71     2003
## 9    CRGHINO R      18      25 1.285714  789  56     2008
## 10     STAHL C      16      30 1.333333  951  45     2010
```

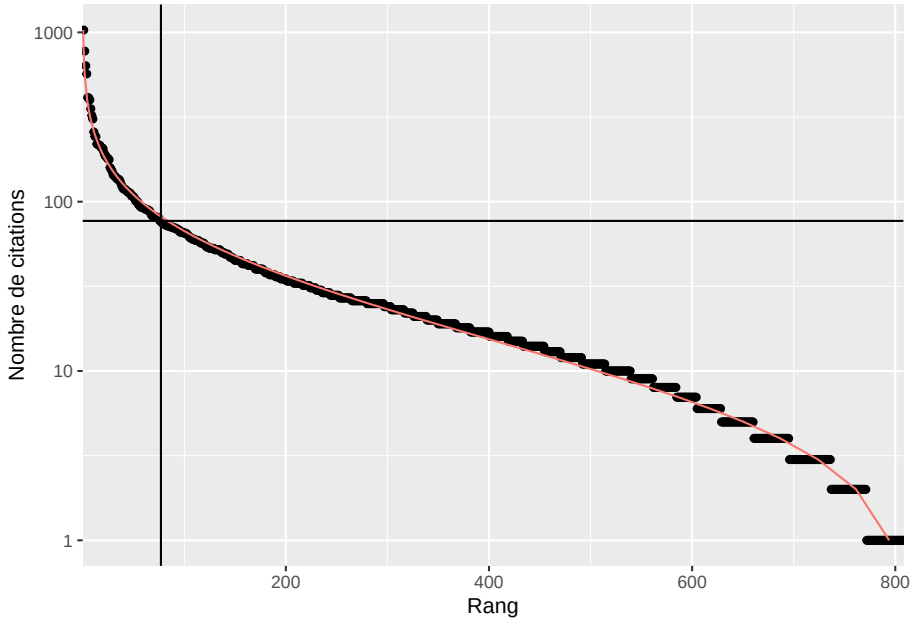
Pour l'indice de toute la base bibliographique :

```
(h <- Hindex(M, elements="*", years=50)$H)
```

```
##   Author h_index g_index  m_index    TC  NP PY_start
## 1      *      77     132 3.666667 28116 859     2001
```

Le graphique rang-citations peut être tracé par le package entropart.

```
library("entropart")
# Courbe rang-abondance, ajustée à une distribution log-normale
autoplot(as.AbdVector(M$TC), ylab = "Nombre de citations", xlab = "Rang", Distribution = "lnorm") +
# Ajout de l'indice h
  geom_hline(yintercept = h$h_index) +
  geom_vline(xintercept = h$h_index)
```



2.4 Documents et auteurs cités

Les documents les plus cités par la base bibliographique sont retournés par la commande `citations`, par article ou par auteur.

```
CAR <- citations(M, field = "article")
CAR$Cited[1:5] %>%
  as_tibble %>%
  rename(Article = CR, Citations=n) %>%
  knitr::kable(caption =
      "Citations les plus fréquentes par les documents de la base de données bibliographique",
      longtable = TRUE, booktabs = TRUE) %>%
  kableExtra::kable_styling(full_width=TRUE, bootstrap_options = "striped")
```

TABLE 2 : Citations les plus fréquentes par les documents de la base de données bibliographique

Article	Citations
KRAFT, N.J.B., VALENCIA, R., ACKERLY, D.D., FUNCTIONAL TRAITS AND NICHE-BASED TREE COMMUNITY ASSEMBLY IN AN AMAZONIAN FOREST (2008) SCIENCE, 322, PP. 580-582	18

CHAVE, J., COOMES, D., JANSEN, S., LEWIS, S.L., SWENSON, N.G., ZANNE, A.E., TOWARDS A WORLDWIDE WOOD ECONOMICS SPECTRUM (2009) ECOLOGY LETTERS, 12, PP. 351-366	16
CRGHINO, R., LEROY, C., DEJEAN, A., CORBARA, B., ANTS MEDIATE THE STRUCTURE OF PHYTOTELM COMMUNITIES IN AN ANT-GARDEN BROMELIAD (2010) ECOLOGY, 91, PP. 1549-1556	13
FINE, P.V.A., MESONES, I., COLEY, P.D., HERBIVORES PROMOTE HABITAT SPECIALIZATION BY TREES IN AMAZONIAN FORESTS (2004) SCIENCE, 305, PP. 663-665	13
NEPSTAD, D.C., TOHVER, I.M., RAY, D., MOUTINHO, P., CARDINOT, G., MORTALITY OF LARGE TREES AND LIANAS FOLLOWING EXPERIMENTAL DROUGHT IN AN AMAZON FOREST (2007) ECOLOGY, 88, PP. 2259-2269	13

Les auteurs les plus cités :

```
CAU <- citations(M, field = "author")
CAU$Cited[1:5] %>%
  as_tibble %>%
  rename(Auteur=CR, Citations=n) %>%
  knitr::kable(
    caption="Auteurs les plus cités par les documents de la base de données bibliographique",
    longtable = TRUE, booktabs = TRUE) %>%
  kableExtra::kable_styling(bootstrap_options = "striped")
```

Table 3 : Auteurs les plus cités par les documents de la base de données bibliographique

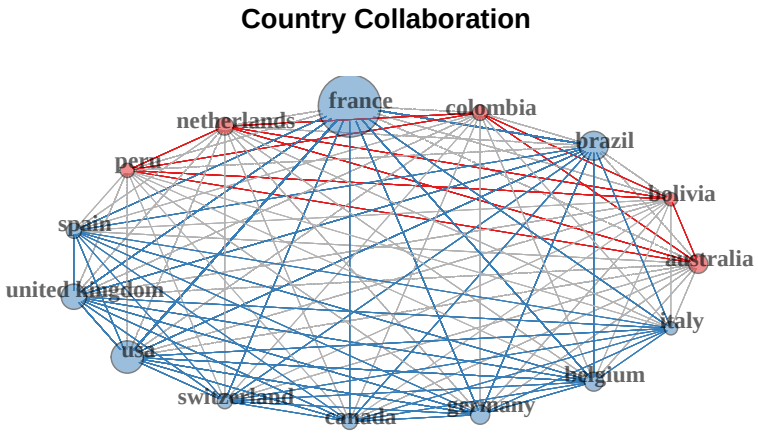
Auteur	Citations

DEJEAN A	927
BARALOTO C	522
ORIVEL J	495
BONAL D	492
CHAVE J	442

2.5 Collaborations

Un réseau de collaboration entre les pays des auteurs est retourné par la fonction `biblioNetwork`.

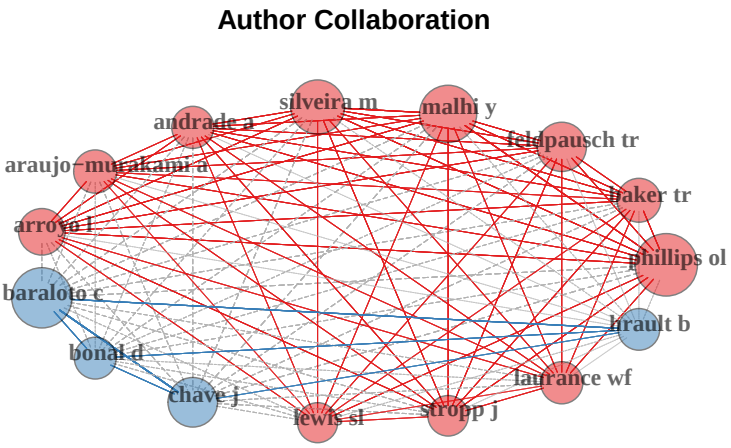
```
NbCountries <- 15
# Create a country collaboration network
mAU_CO <- metaTagExtraction(M, Field = "AU_CO", sep = ";")
NetMatrix <- biblioNetwork(mAU_CO, analysis = "collaboration",
    network = "countries", sep = ";")
# Plot the network
netC <- networkPlot(NetMatrix, n = NbCountries, Title = "Country Collaboration",
    type = "circle", size = TRUE, remove.multiple = FALSE)
```



Le réseau des auteurs est obtenu de la même façon.

```
NbAuthors <- 15
# Réseau d'auteurs
AuthorNet <- biblioNetwork(M, analysis = "collaboration",
```

```
    network = "authors", sep = ";")
netA <- networkPlot(AuthorNet, n = NbAuthors, Title = "Author Collaboration",
    type = "circle", size = TRUE, remove.multiple = FALSE)
```



3 Analyse des résumés

Les résumés des publications se trouvent dans la colonne `AB` de la base importée par *bibliometrix*. Ils sont en Anglais.

3.1 Corpus

Le package `tm` permet de constituer un corpus.

```
library("tm")
M$AB %>%
  VectorSource %>%
  VCorpus %>%
  tm_map(PlainTextDocument) %>%
  tm_map(content_transformer(tolower)) ->
  MonCorpus
```

La fonction `tm_map` permet d'appliquer une fonction quelconque à chaque élément du corpus, c'est-à-dire à chaque résumé. Les fonctions standard, n'appartenant pas au package `tm`, doivent être appliquées par l'intermédiaire de la fonction `content_transformer` pour ne pas dégrader la structure du corpus : dans le code précédent, la fonction `tolower` est appliquée à chaque résumé pour le passer en minuscules, alors que la création de corpus est en majuscules.

3.2 Nettoyage du corpus

Des mots sémantiquement identiques ont plusieurs formes. Le traitement le plus rigoureux consiste à les réduire à leur radical mais le résultat n'est pas très lisible. La fonction `stemDocument` permet de le faire : il suffit de l'utiliser à la place de `PlainTextDocument` dans le code ci-dessus. Un bon compromis consiste à supprimer les formes plurielles, par une fonction ad-hoc : ce sera fait plus tard.

Les déterminants, conjonctions, etc. sont les mots les plus fréquents mais n'ont pas d'intérêt pour l'analyse. La fonction `removeWords` permet de retirer une liste de mots. `stopwords` fournit la liste de ces mots dans une langue au choix. `removeNumbers` retire les nombres comme *one*, *two*, etc. et la fonction `removePunctuation` retire la ponctuation.

```
MonCorpus %<>% tm_map(removePunctuation) %>%
  tm_map(removeNumbers) %>%
  tm_map(removeWords, stopwords("english"))
```

Une liste de mots complémentaire est nécessaire pour supprimer des mots inutiles mais fréquents. Elle peut être complétée de façon itérative pour retirer des mots parasites du résultat final.

```
ExtraWords <- c("use", "used", "using", "results",
    "may", "across", "high", "higher", "low", "show",
    "showed", "study", "studies", "studied", "however",
    "can", "our", "based", "including", "within", "total",
    "among", "found", "due", "also", "well", "strong",
    "large", "important", "first", "known", "one",
    "two", "three")
MonCorpus %<>% tm_map(removeWords, ExtraWords)
```

3.3 Mots du corpus

L'objectif est de transformer le corpus en un vecteur d'abondance des mots utilisés. `TermDocumentMatrix` crée un objet spécifique au package *tm* qui pose des problèmes de traitement. Cet objet est transformé en un vecteur d'abondances.

```
TDM <- TermDocumentMatrix(MonCorpus, control = list(minWordLength = 3))
AbdMots <- sort(rowSums(as.matrix(TDM)), decreasing = TRUE)
```

Le vecteur de mots contient des formes singulières et plurielles. Elles peuvent être regroupées selon un modèle simple : si un mot existe avec et sans *s* ou *es* final, la forme singulière est sans *s* ou *es*. Des pluriels particuliers peuvent être ajoutés selon les besoins.

```
# Adapté de https://github.com/mkfs/misc-text-mining/blob/master/R/wordcloud.R
aggregate_plurals <- function (v) {
    aggr_fn <- function(v, singular, plural) {
        if (! is.na(v[plural])) {
            v[singular] <- v[singular] + v[plural]
            v <- v[-which(names(v) == plural)]
        }
        return(v)
```

```
    }
    for (n in names(v)) {
        n_pl <- paste(n, 's', sep='')
        v <- aggr_fn(v, n, n_pl)
        n_pl <- paste(n, 'es', sep='')
        v <- aggr_fn(v, n, n_pl)
        # cas particuliers
        if (endsWith(n, "y")) {
          n <-  substr(n, 1, nchar(n)-1)
          n_pl <- paste(n, 'ies', sep='')
          }
        if (n == "genus") {
          n_pl <- "genera"
          v <- aggr_fn(v, n, n_pl)
        }
    }
    return(v)
}

AbdMots %<>% aggregate_plurals
```

3.4 Nuage de mots

Le résultat final est un nuage de mots.

```
library("wordcloud")
df <- data.frame(word = names(AbdMots), freq = AbdMots)
wordcloud(df$word, df$freq, max.words = 100, random.order = FALSE,
    rot.per = 0.35, use.r.layout = FALSE, colors = brewer.pal(8,
        "Dark2"))
```