



Primeros pasos y técnicas avanzadas con BiblioShiny App

Materiales didácticos

Daniel Torres-Salinas

Versión 0.8 - work in progress

Contenido

Úsese el siguiente texto como un conjunto de anotaciones y comentarios que sirven de guía al curso introductorio del software Bibliometrix dónde trabajamos directamente con esta herramienta. Este curso incluye dos sesiones:

- BIBLIOMETRIX (I): Introducción y primeros pasos. en este curso hacemos una introducción al software bibliometrix desarrollado por Massimo Aria y Corrado Cuccurullo. En el mismo no centraremos en su interfaz biblioshiny explicando como instalarlo en tu pc. a continuación haremos una descripción de las funcionalidades y veremos las opciones de carga de datos.
- BIBLIOMETRIX (II): En esta segunda parte comenzaremos a conocer más profundamente el software bibliometrix ilustrando los diferentes tipos de análisis y gráficos que realiza a nivel de autor, fuentes y documentos. Asimismo veremos cómo se llevan a cabo los análisis estructurales en sus tres niveles: conceptual, intelectual y social.

0) ¿Por qué Bibliometrix?

Muchos son los programas informáticas y aplicaciones que existen para realizar análisis bibliométricos y calcular indicadores, sin embargo Bibliometrix reúne una serie de características que nos animan a recomendarlo, podemos señalar las siguientes:

- 1) Está programado en r lo que garantiza futuros desarrollos y la suma de los aportes de la comunidad científica y programadora de r
- 2) Se puede utilizar con una interfaz de Biblioshiny lo que es estupendo para todos aquellos que no quieran aprender y nos permite trabajar inmediatamente
- 3) Trabaja con colecciones de diferentes bases de datos y también con diferentes formatos en los que se descargan los registros
- 4) Incorpora un compendio los análisis más sencillos pero también las técnicas más complejas (co-citation networks, factorial análisis)
- 5) Ha incorporado las principales funcionalidades de otros conocidos programas bibliométricos como Scimat y HistCite.
- 6) Puede ser empleado por profesionales para la realización de informes bibliométricos, por investigadores de todas las áreas para realizar análisis sistemáticos y por investigadores bibliométricos para sus análisis

1) Introducción y primeros pasos

1.1. Características generales

Los creadores

Debemos comenzar esta exposición mencionando y reconociendo la labor de sus creadores que tuvieron desarrollar Bibliometrix en r y distribuirlo de manera gratuita, nos referimos a

[Massimo Aria](#) (derecha) & [Corrado Cuccurullo](#) (izquierda)



Formatos Bibliográficos.

El software con el que trabajaremos lo podéis encontrar en la página web de Bibliometrix: <https://www.bibliometrix.org/index.html>

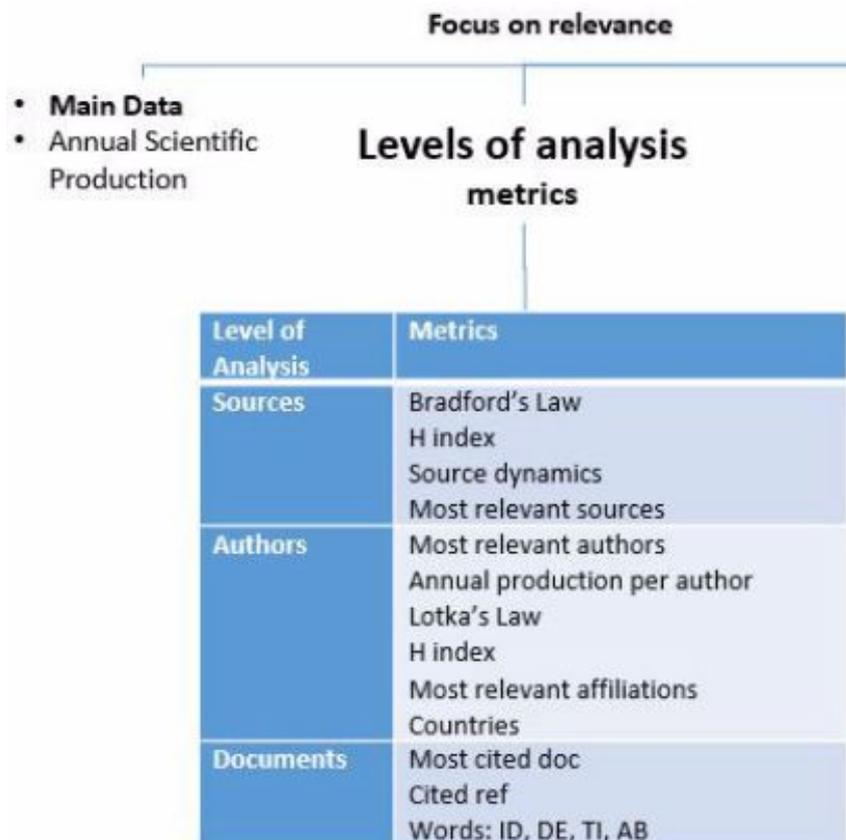
Parte de los materiales presentados en estos materiales provienen de esta página web y de los tutoriales y explicaciones de los autores que podéis encontrar en la misma.

Es importante conocer que existen dos versiones de Biblioshiny, una versión en R que requiere conocimientos de este lenguaje y una versión que utiliza biblioshiny que nos permite trabajar con un interfaz y de forma más sencilla

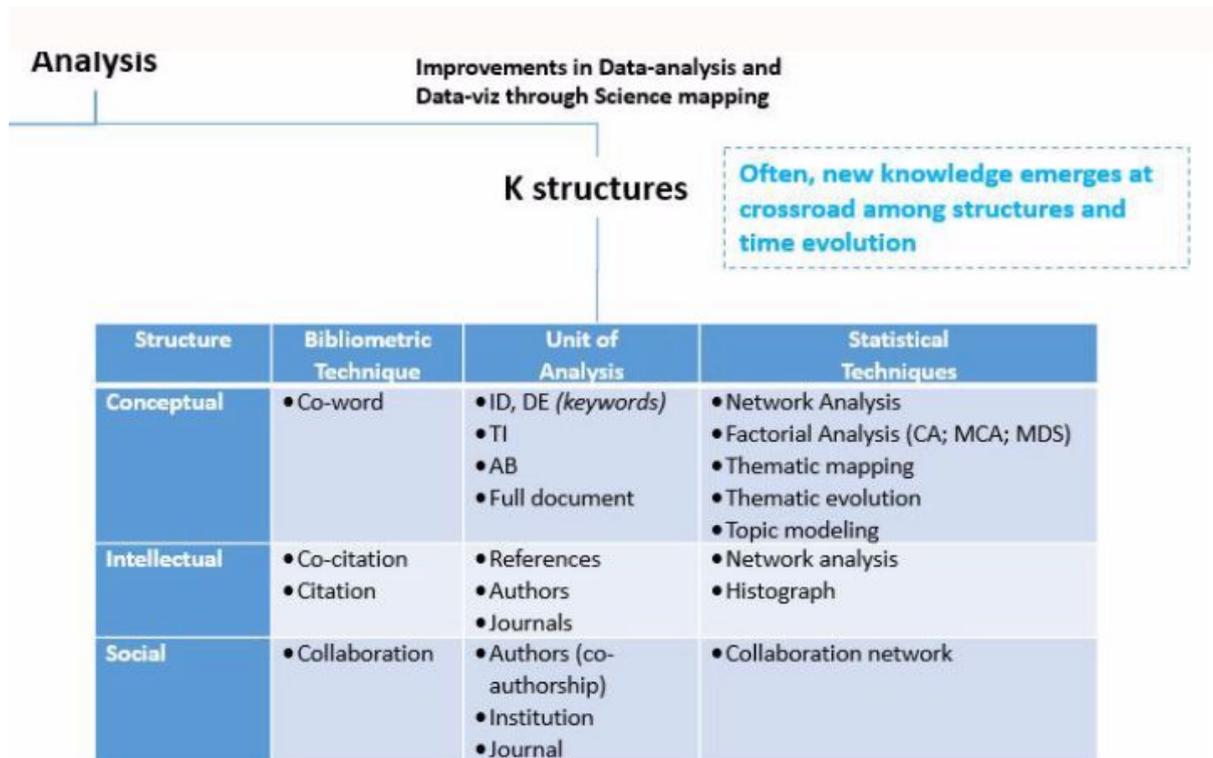
Tipos de análisis que hace bibliometrix

| | |
|---|---|
| Niveles de análisis: - Fuentes - Autores - Documentos | Análisis estructurales - Estructura Conceptual - Estructura Intelectual - Estructura Social |
|---|---|

Niveles de análisis



Niveles estructurales



Recursos recomendados

Diapositivas

[Diapositivas originales con Biblioshiny](#)

Vídeos

- [Bibliometrics \(17\): The Biblioshiny App For Non-Coders | Bibliometrix R Package](#)
- [Introducción básica a Bibliometrix](#)
- [Bibliometría Now: Primeros pasos con Biblioshiny - Bibliometrix](#)

1.2. Instalación de biblioShiny App

Resumen del proceso de instalación

1. Instalamos R
2. Instalamos R-Studio
3. Ejecutamos R-Studio
4. Instalamos el paquete de r de Bibliometrix
5. Ejecutamos BiblioShiny App

Veamos a continuación en detalle cada uno de estos pasos:

Instalamos R

Nos [Vamos a la dirección del cran r project](#)

Seleccionamos la opción que nos interesa (Linux, Mac, Windows)

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Seleccionamos Install R for the first time

R for Windows

Subdirectories:

| | |
|-----------------------------|---|
| base | Binaries for base distribution. This is what you want to install R for the first time . |
| contrib | Binaries of contributed CRAN packages (for R \geq 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables. |
| old contrib | Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 2.13.x; managed by Uwe Ligges). |
| Rtools | Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself. |

Una vez en la pantalla seleccionamos Download R

R-4.0.3 for Windows (32/64 bit)

[Download R 4.0.3 for Windows](#) (85 megabytes, 32/64 bit)

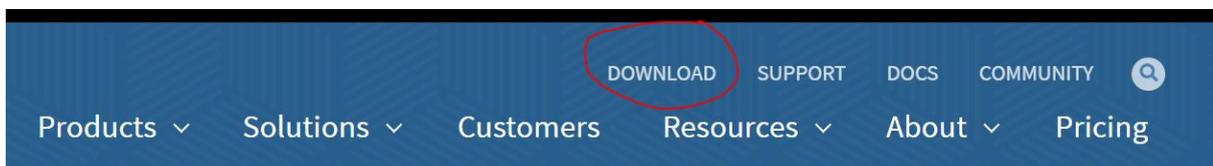
[Installation and other instructions](#)

[New features in this version](#)

Procedemos a instalar R (le decimos a todo los que nos diga que sí)

Instalamos R-Studio

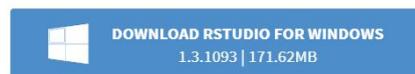
Descargamos el software desde la [web de r-studio](#)



Descargamos la versión “Free” que nos interese e instalamos (decimos a todo que sí)

RStudio Desktop 1.3.1093 - [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires Windows 10/8/7 (64-bit)

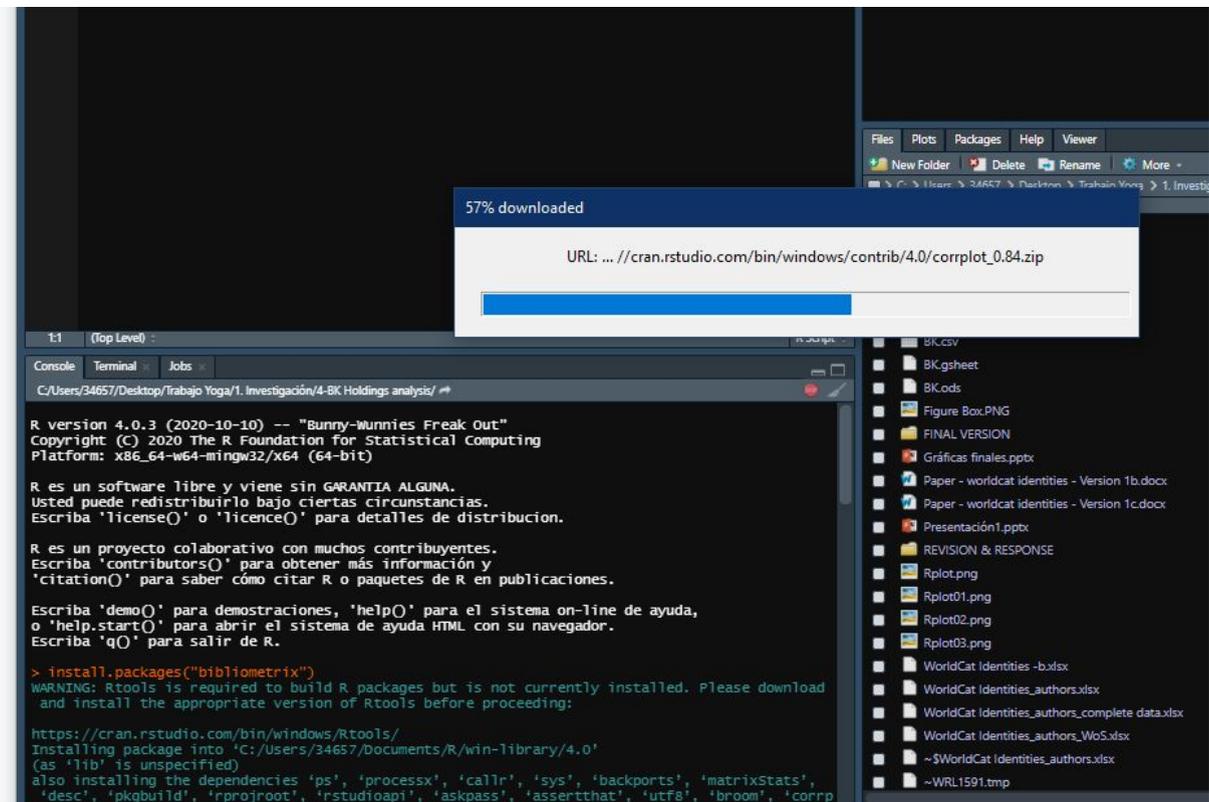


Instalamos el paquete de R de Bibliometrix. No te asustes, ahora hay que trabajar un poco con r-studio, pero solo tendremos que escribir unas cuantas líneas y pronto lo abandonaremos. Una vez instalados ejecuta R-Studio y verás que se abre una pantalla dividida en cuatro partes. Para instalar el packages en la consola, zona inferior izquierda escribimos el siguiente código

`install.packages("bibliometrix")`

y pulsamos intro, a continuación se instalará el paquete de R.

Puede tardar un rato y verás cómo se va instalando el paquete



Ejecutamos BiblioShiny App

Una vez que hemos instalado el package de r lleva el momento de ejecutar y poner en marcha BiblioShiny App. Para ello en la consola escribe estas dos líneas de código:

```
library(bibliometrix)
biblioshiny()
```

```
> library(bibliometrix)
> biblioshiny()

Listening on http://127.0.0.1:5756
```

Ahora ya podemos empezar a trabajar

1.3. Descarga y carga de los datos desde diferentes fuentes

Formatos Bibliográficos

Bibliometrix puede trabajar con cinco bases de datos diferentes, y dentro de ellas con diferentes formatos de exportación de registros.

Bases de datos, formatos y extensiones

| Base de datos | Formatos | Extensión |
|----------------|---------------------------------------|-------------------|
| Web of Science | BibTeX plaintext Ednote Desktop | Bib txt ciw |
| Scopus | BibTeX CSV export | bib txt |
| Dimensions | Bibliometrics mapping Excel | csv xlsx |
| Pubmed | Pubmed export file | txt |
| Cochrane | Plaintex | txt |

Ejemplo de formato Plain Text de Web of Science - [Ver archivo](#)

```
PT J
AU Garcovich, D
   Wu, AZ
   Sucar, AMS
   Martin, MA
AF Garcovich, Daniele
   Zhou Wu, Angel
   Sanchez Sucar, Ana-Matilde
   Adobes Martin, Milagros
TI The online attention to orthodontic research: an Altmetric analysis of
   the orthodontic journals indexed in the journal citation reports from
   2014 to 2018
SO PROGRESS IN ORTHODONTICS
LA English
DT Article
DE Altmetrics; Bibliometric; Citations count; Mendeley; Social media
ID SCORE
```

Ejemplo de formato CSV en Scopus - [Ver archivo](#)

| Authors | Author(s) ID | Title | Year | Source title | Volume | Issue | Art. No. |
|-----------------------|-------------------|------------------------|------|-----------------------|--------|----------------------|----------|
| Antonakaki D., Frago | 42161138100;66024 | A survey of Twitter re | 2021 | Expert Systems with | 164 | | 114006 |
| Li D., Li X., Gao B. | 57198514031;57206 | The Development ar | 2021 | Smart Innovation, Sy | 182 | | |
| Garcovich D., Zhou | 14123240100;57216 | The online attention | 2020 | Progress in Orthodon | 21 | 1 | 31 |
| Mettai K., Boumaraf | 57202110534;24605 | Measuring the value | 2020 | Grey Journal | 16 | Special Issue Winter | |
| Schöpfel J., Prost H. | 14619562900;15065 | The scope of open s | 2020 | Grey Journal | 16 | Special Issue Winter | |
| Korkmaz G., Kelling | 56278329300;57204 | Modeling the impac | 2020 | Social Network Analy | 10 | 1 | 7 |
| Obemeister N. | 57192652762; | Tapping into scienc | 2020 | Palgrave Communic | 6 | 1 | 74 |
| Gruetzemacher R., F | 57191160886;66035 | Forecasting extreme | 2020 | Technological Forec | 161 | | 120323 |
| Sinha N., Singh P., | 25628729100;57216 | Robotics at workplac | 2020 | International Journal | 55 | | 102210 |
| Mora L., Wu X., Pan | 57169999100;57216 | Mind the gap: Devel | 2020 | Journal of Cleaner P | 275 | | 124087 |
| Wang Z., Chen Y., G | 57215366921;56765 | Preprints as accelera | 2020 | Journal of Infometric | 14 | 4 | 101097 |
| Ortega J.L. | 14060977700; | Proposal of compose | 2020 | Journal of Infometric | 14 | 4 | 101071 |
| Haldule S., Davalbh | 57218996516;57216 | Post-publication pro | 2020 | Rheumatology Intern | 40 | 11 | |
| Kong L., Wang D. | 57219307031;55713 | Comparison of citati | 2020 | Journal of Infometric | 14 | 4 | 101095 |

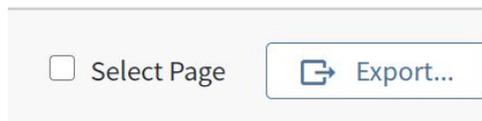
Otros formatos habituales de exportación:

- Ejemplo de formato Bibtex - [Ver página web](#)

Descarga y carga de los datos desde Web of Science

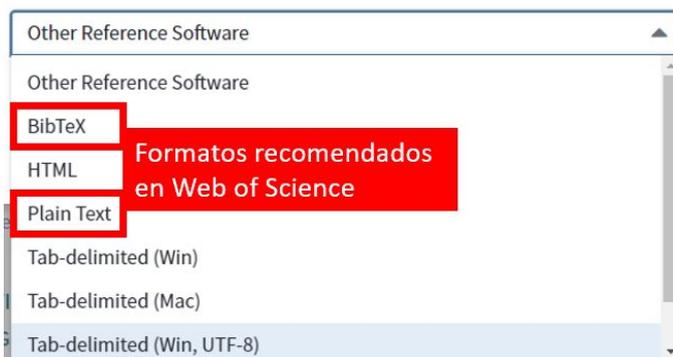
Es importante seleccionar siempre en base Web of Science Core Collection

En Web of science seleccionar



y dentro del mismo “other format files” dentro del mismo podemos seleccionar, el número de registros a descargar, el contenido “Record Content” seleccionar “Full Record”. A continuación seleccionamos el tipo de formato (File Format) con el que vamos a trabajar BibTex o Plain Text.

File Format



Pasos para descargar desde Web of Science:

- Probemos ahora con un archivo concreto de Web of Science, vamos a buscar todos los trabajos relacionados con el frente de investigación de las altmetrics, vamos a buscar en todas las bases de datos de core collection sin límite temporal y en el campo topic. Los descargamos en “Full Record and Cited References”
- El total de registros es de 799
- Unificación de archivos en Web of Science, recordar que solo podemos descargar de 500 en 500 registros. Podemos realizar una unificación manual
- [Ver archivo](#)

Pasos para cargar un archivo en BiblioShiny App

- Ir al menú data
- Seleccionar el tipo de fichero “Raw files” o “Bibliometrix files”, siempre seleccionar el formato Raw
- Seleccionar a continuación la base de datos de procedencia “Web of Science” del fichero y el formato “Plain Text”
- Pulsamos start
- Si el fichero se carga correctamente los podemos ver en pantalla



Un pequeña tip

- En el menú de carga cuando hemos cargado un fichero podemos utilizar la opción Export a bibliometrix file para convertir los registros a excel o formato r
- Poner un ejemplo

Descarga y carga de los datos desde Scopus

Pasos para descargar desde Scopus:

- Realizamos la misma búsqueda “altmetrics” seleccionando la opción documents y que busque en todos los campos
- En la pantalla de resultados seleccionar “todos” y pinchar en export nos aparecerá un menú de formación donde escogemos el formato que nos interesan así como los bloques de información concretos de información que queremos exportar
- La limitación de registros en scopus es 2000 en una única descarga, por tanto tendremos que unificar ficheros. Importante seleccionar siempre la opción de solo los 2000 registros “CSV - Only the first 2,000 documents”. Bibliometrix no trabaja con la opción de más de 2000 registros y solamente información de citas “CSV - Export all documents to a CSV file, citation information only.”

Select your method of export

MENDELEY
 ExLibris RefWorks
 SciVal *i*
 RIS Format
EndNote, Reference Manager
 CSV
Excel
 BibTeX

Plain Text
ASCII in HTML

What information do you want to export?

Citation information
 Bibliographical information
 Abstract & keywords
 Funding details

Pasos para cargar un archivo

- Ir al menú data
- Seleccionar el tipo de fichero “Raw files” o “Bibliometrix files”, siempre seleccionar el formato Raw
- Seleccionar a continuación la base de datos de procedencia “Scopus” del fichero y el formato “CSV”
- Pulsamos start
- Si el fichero se carga correctamente lo podemos ver en pantalla

Descarga de datos desde la API

Solo disponible para dos bases de datos Pubmed y Dimensions
Poco útil ya que en ambos la búsqueda a la API se reduce a 2000
Poner un ejemplo de búsqueda de Altmetrics

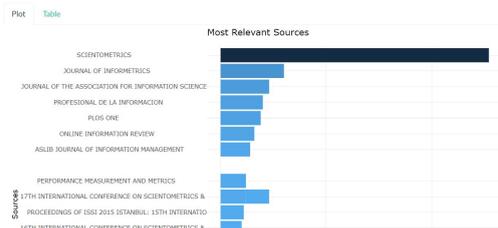
Filtrado de los documentos cargados

Esta opción de filtrado es interesante para reducir el dataset

2.2. Sources: analizando las fuentes

Most relevant Sources

- Generan un conteo por fuente de información

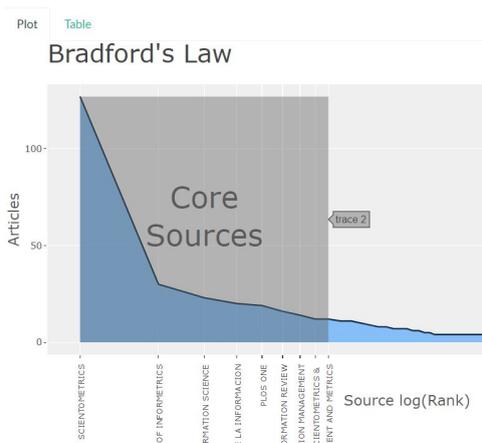


Most Local Cited Sources

- Toma las referencias de los artículos (**Pendiente**)

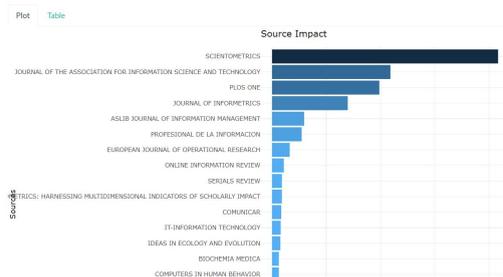
Bradford Laws

- Aplica la [ley de Bradford](#)



Sources local impact

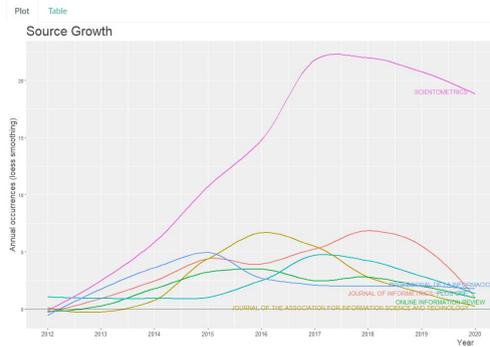
- Realización de varigos para algunos indicadore H, G, M index y citas



Sources local impact

- Nos permite obtener el gráfico de evolución
- Las opciones son datos anuales y acumulados

- Se puede incluir el intervalo de confianza y el número de revistas



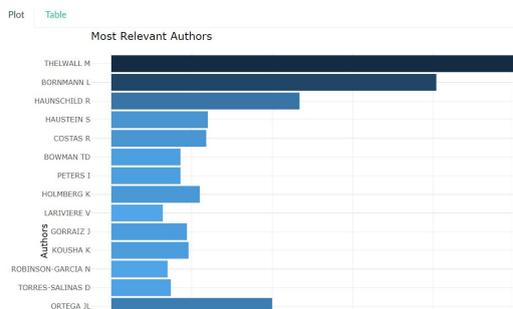
2.3. Authors: Analizando los autores

Tres tipos de nivel análisis

- Autores
- Afiliaciones
- Países

Authors: Most relevant authors

- Nos permite un gráfico de los autores más productivos
- Los más interesante es que nos permite varios recuentos de los autores:
 - N. of Documents. Número de documentos totales
 - Percentage: Igual al anterior pero en porcentaje
 - Fractionalized Frequency: Hace un recuento fracciones (considerando el número de los autores, por ejemplo si tenemos un trabajo con 4 autores este cuenta 0.25 puntos)



Authors: Most Local Cited Authors

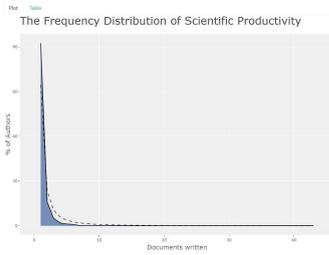
- Nos permiten comprobar los autores más influyentes en el campo

Authors: Author's Production Over Time

- **Not Working**

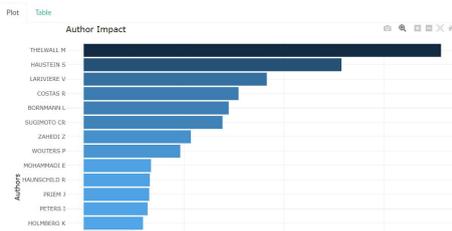
Authors: Lotka's Law

- Aplica la [ley de Lotka](#)



Authors: Author Local Impact

- Calcula varios indicadores, ojo con los indicadores H,G y M ya que quizás no son significativos

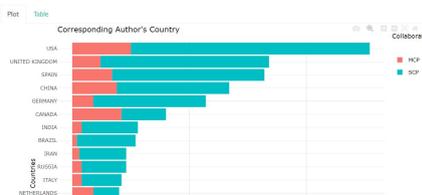


Affiliations: Most Relevant Affiliations

- Obtiene los centros más productivos, es interesante que utiliza un algoritmo de desambiguación para afinar los conteos

Countries: Corresponding Author Countries

- Obtiene
- Ojo con este conteo es posible que algunos de los documentos no cuente con esta información
- Mirar que significa MCP y SCP



Countries: Country Scientific Production y Most Cited Countries



2.3. Documents: Análisis a nivel de documento

Tres tipos de análisis

- Documentos
- Referencias Citadas
- Palabras

Documents: Most Relevant Cited Documents & Most Local Cited Documents

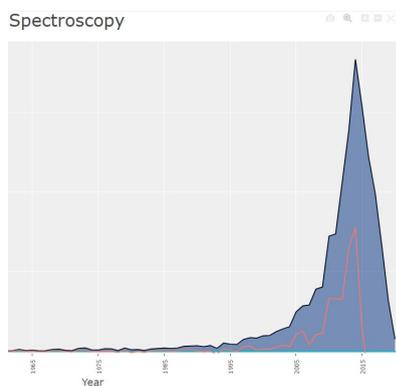
- Obtiene los centros más productivos, es i

Cited References: Most Local Cited References

- **Not Working**

Cited References: Reference Spectroscopy

- De utilidad para realizar estudios relacionados con el envejecimiento de la literatura científica



Words: Most Frequent Words

- Simples conteos de palabras basado en algunos de los campos Keywords Plus, Authors Keywords, Títulos o Abstracts. Una pregunta interesante cuál es mejor para utilizar.

Words: Word Clouds

- Se pueden cambiar diversos parámetros como Word occurrence measure (es interesante Frecuency, no le veo demasiada utilidad porque apenas a square y Log) y El resto de parámetros son de carácter estético como Shape, Font Type, Text colors, etc...

3) Análisis estructurales

Vamos a entrar en análisis complejos, algunos de ellos podrían llevar horas aprenderlos, aquí queremos hacer una exposición básico

Estructura conceptual = Trabajamos con términos

Estructura Intellectual = Trabajamos con personas

Estructura Social = Trabajamos con relaciones sociales (colaboración científica)

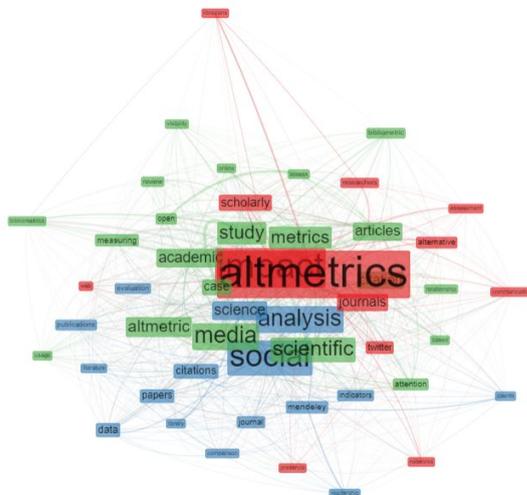
3.1. Conceptual Structure: Estructura conceptual

Dos tipos de análisis

- **Network approach**
- **Factorial approach**

Network: Co-occurrence Network

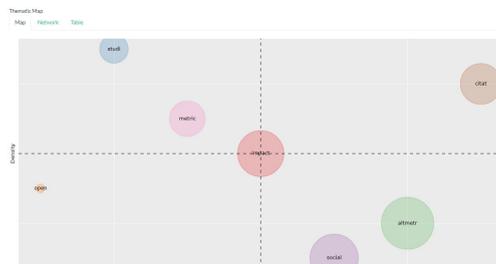
- Es un análisis donde se construye una [matriz de coocurrencia](#) de palabras, es una las técnicas más clásicas
- Tienen varios bloques de configuración relacionados con los “parámetros de la red”. Este análisis como los siguientes nos requiere más formación y conocimientos por ejemplo el layout utilizando diferentes algoritmos (kamada Kawai, Fruchterman, ...).
- También se puede normalizar los valores de la matriz con los valores de Jaccard, Dice, equivalencia, etc...
- Aplica los algoritmos de clustering
- Todos los parámetros anteriores se orienta para una correcta visualización de los fenómenos que queremos analizar
- El resto de opciones son más de carácter estético: colores, tamaños, fuentes, etiquetas, etc...
- Exportación a Pajek



●

Network: Thematic Map

- Este análisis traslada los análisis del software [SCIMAT](#)
- Hay muchos artículos sobre scimat [pero or recomiendo éste](#), en la página web de su [creador tenéis más artículos](#)
- No incorpora todas las opciones del software original pero si es muy completo
- A pesar de ser una técnica compleja han sabido incorporar la “esencia” del software y simplificar considerablemente los parámetros de configuración
 - Field
 - Word Stemming (para cortar en la raíz de las palabras y evitar duplicidades por género, por plurales, ...)
 - Se puede definir el número de palabras, número de etiquetas y la frecuencia mínima que tienes que tener las palabras para aparecer (por tanto muy útil para la poda)
 - Podemos ver el mapa de temas y la red general

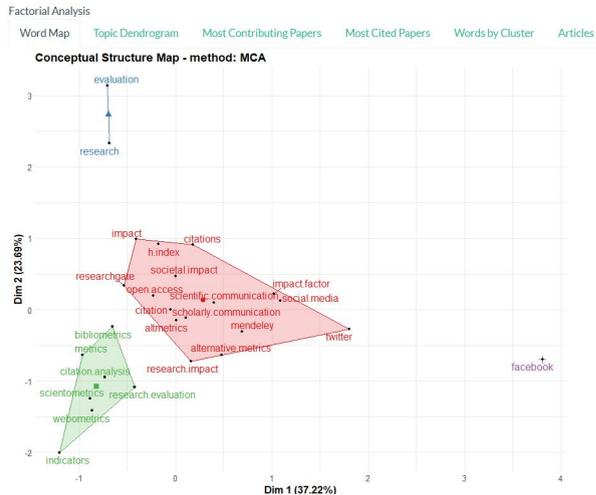


Network: Thematic Evolution

- Es un análisis muy complejo también incorporar en SCIMAT, para funcionar correctamente necesita un gran número de documentos y sobre todo un número significativo de años. [Aquí tenéis también algún ejemplo de como funciona](#)

Factorial: Factorial Analysis

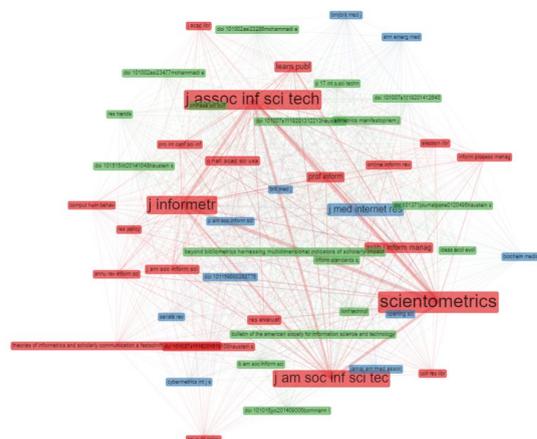
- Realiza un análisis, aquí tenéis [un ejemplo de aplicación en el ámbito de la sociología](#)
- El análisis factorial incluye tres tipos de análisis de reducción de datos: análisis de correspondencias, escalamiento multidimensional, análisis múltiple de correspondencias. La cuestión aquí radica en el análisis escoger. En este trabajo podéis tener una [referencia de cómo funciona](#). La selección de número de palabras y del número de clusters es fundamental
- Presenta la reducción bidimensional que se acompaña de pestañas como del dendograma (que matiza sustancialmente el gráfico), papers más relevantes y la clasificación de las palabras y los artículos



3.2. Intellectual Structure: Estructura intelectual

Co-Citation Network

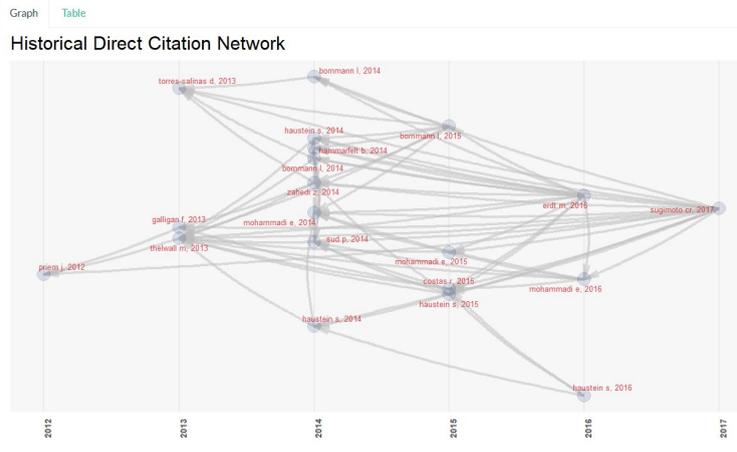
- Otra de las grandes técnicas tradicionales de las que tenemos múltiples ejemplos en la literatura científica. [Ejemplos](#). Se toma como base dos documentos citados conjuntamente, es el vínculo entre nodos.
- En Bibliometrix podemos realizar tres tipos de mapas de autores, fuentes y trabajos
- Tenemos opciones de configuración y clustering y edición
- Estar atentos al tema de la separación de campos, esto es importante por ejemplo si vamos a trabajar con documentos “Papers” podemos utilizar cualquiera (recomendamos “,”) pero si lo hacemos con un fichero WoS queremos ver autores “Autores” tenemos que utilizar “;”. En cualquier parece que se generan errores



-
- Una vez más recomendamos descargar el archivo pajel (poner ejemplo)

Historiograph

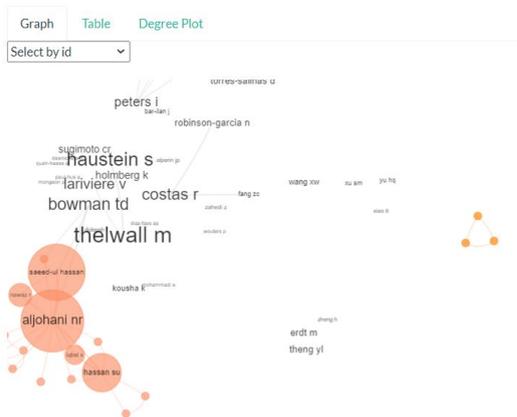
- Aquí al igual que ocurría con SCIMAT nos encontramos con una adaptación de un software clásico de la bibliometría, HISTCITE. Hay varios artículos que analizan el software ([Ejemplo](#)). Útil sobre todo para trabajos que analicen campos
- Fácil de obtener, solo un parámetro de para el número de nodos y tres para parámetros gráficos y estéticos (etiqueta y tamaño del nodo)



3.3. Social Structure: Social intelectual

Collaboration Network

- Basada en la firma conjunta de publicaciones, otras de las técnicas habituales. Se puede realizar a tres niveles: autores, instituciones y países
- Como en anteriores ocasiones los parámetros de configuración de la red son similares al igual que los de edición
- La tabla incorpora indicadores de centralidad y pagerank de gran utilidad para identificar actores de la red



Collaboration Map

- Un simple mapa

4) Ejercicios prácticos

4.1. Instalación y puesta en marcha de BiblioShiny en nuestro PC

Nuestro primer ejercicio evidentemente, si aún no lo hemos realizado, es instalar R y Rstudio y poner en marcha BiblioShiny App

4.1. Carga de datos de diferentes bases de datos

| Búsqueda y base de datos | Ejercicios |
|---|---|
| <p>Realiza la siguiente búsqueda en Web of Science Core Collection:</p> <p>Busca “Bibliometrics” en Topics Limita a la categoría Information and Library Science Limita a los años 2020 y 2019</p> | <p>1) Descarga los registros en formato BibTex y cárgalos en BiblioShiny</p> <p>2) Descarga los registros en formato Ednote Desktop y cárgalos en BiblioShiny</p> <p>2) Descarga los registros en formato Plain Text y cárgalos en BiblioShiny</p> |
| <p>Realiza la siguiente búsqueda en Scopus</p> <p>Busca “Scientometrics” en Source Title Limita la búsqueda al año 2020</p> | <p>1) Descarga los registros en formato BibTex y cárgalos en BiblioShiny</p> <p>2) Descarga los registros en formato CSV y cárgalos en BiblioShiny</p> |
| <p>Realiza la siguiente búsqueda en Dimensions</p> <p>Lutz Bornmann</p> | <p>Ojo para descargar los registros en Dimensions hay que estar logueado</p> <p>1) Descarga los registros en formato Excel-XLSX y cárgalos en BiblioShiny (topics)</p> <p>2) Descarga los registros en formato Export for Bibliometrix Mapping y cárgalos en BiblioShiny</p> |

4.2. Análisis de la producción científica de un campo

Vamos a trabajar en un ejercicio exploratorio y un poco libre. El objetivo es analizar un campo científico, queremos hacer un análisis de la producción científica española en el campo de las métricas.

Vamos a utilizar como base de datos de referencia **Web of Science** y los parámetros de búsqueda (no demasiado refinados) han sido los siguientes:

- He utilizado solo el Social Science Citation Index
- Los términos han sido TOPIC: (bibliometric* or scientometric* or altmetric* or informetric*)
- He limitado en ADDRESS: (spain)
- He utilizado todos los años
- El resultado es el siguiente



Results: 1,253
(from Web of Science Core Collection)

You searched for: TOPIC:
(bibliometric* or scientometric* or
altmetric* or informetric*) AND
ADDRESS: (spain)

Timespan: All years.
Indexes: SSCI.
...Less

Lo primero que debes hacer es descargar los registros y crear un único fichero para cargarlo en BiblioShiny.

A continuación realizada un informe que incluya al menos 8 gráficos y/o tablas diferentes en el que analices el campo científico. Asimismo obligatoriamente debes realizar al menos un análisis conceptual y un análisis estructural. Acompaña cada gráfico con su nombre o tipo de análisis y un breve comentario de los resultados.