Chapter 3 / Capítulo 3

Applied bibliometrics. From data to publication (Spanish Edition)
ISBN: 978-9915-9680-6-3
DOI: 10.62486/978-9915-9680-6-3.ch03

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

TOOLS FOR BIBLIOMETRIC ANALYSIS/ HERRAMIENTAS PARA EL ANÁLISIS BIBLIOMÉTRICO

3.1. Introducción a las herramientas bibliométricas

El software bibliométrico representa la columna vertebral metodológica que transforma datos crudos en conocimiento accionable. Sin estas herramientas especializadas, los investigadores enfrentarían la imposibilidad práctica de analizar manualmente los volúmenes masivos de publicaciones científicas que caracterizan la era contemporánea. La automatización permite procesar miles de referencias en minutos, identificando patrones de colaboración, tendencias temáticas y redes de citación que permanecerían ocultas en el ruido informativo. Esta capacidad computacional no solo optimiza recursos temporales, sino que fundamentalmente expande las fronteras de lo analizable, permitiendo abordar preguntas de investigación que antes resultaban inabarcables por limitaciones técnicas y operativas.

La elección del software adecuado determina directamente la validez y profundidad de los hallazgos bibliométricos. Cada herramienta incorpora supuestos metodológicos particulares que condicionan los tipos de análisis posibles y las interpretaciones derivadas. Plataformas como VOSviewer priorizan la visualización de redes, mientras Bibliometrix enfatiza los indicadores estadísticos y Python ofrece flexibilidad algorítmica. Esta especialización crea un ecosistema complementario donde la combinación estratégica de herramientas supera las limitaciones individuales. La competencia en el manejo de múltiples softwares se convierte así en una habilidad fundamental para el bibliometrista contemporáneo que busca producir evaluaciones científicas robustas y multidimensionales.

La evolución del software bibliométrico refleja y simultáneamente impulsa el desarrollo teórico de la disciplina. Las herramientas actuales incorporan avances como el procesamiento de lenguaje natural para análisis de contenido o algoritmos de *machine learning* para detección de temas emergentes. Esta simbiosis entre desarrollo tecnológico y sofisticación metodológica ha transformado la bibliometría desde un ejercicio descriptivo básico hacia una ciencia analítica predictiva. El investigador moderno debe entender tanto los fundamentos bibliométricos como las capacidades técnicas del software para diseñar estudios que aprovechen todo el potencial analítico disponible en el ecosistema digital contemporáneo.

Este libro no pretende abordar de manera profunda las herramientas que se utilizan para la creación de bibliometrías, pero si dotar al lector de los conocimientos mínimos para manejarlas, así como guiarlo en caso de que este desee profundizar en las mismas.

3.1.1. Clasificación de herramientas y criterios de selección

El ecosistema actual de software bibliométrico presenta una diversidad significativa que responde a distintas necesidades metodológicas y niveles de experticia. Esta variedad puede organizarse en categorías claramente definidas según su complejidad técnica y objetivos analíticos particulares. La comprensión de esta taxonomía fundamental constituye el primer paso hacia una selección adecuada que optimice los recursos de investigación disponibles.[1]

En la base de la pirámide de herramientas encontramos el software de análisis básico, diseñado para usuarios que requieren cálculos rápidos de indicadores fundamentales. Estas soluciones priorizan la accesibilidad sobre la profundidad analítica, ofreciendo interfaces intuitivas que minimizan la curva de aprendizaje. Su principal ventaja radica en la capacidad de generar resultados inmediatos sin requerir conocimientos técnicos especializados, aunque su flexibilidad metodológica resulta limitada para investigaciones complejas.

Las plataformas de análisis integral representan el siguiente nivel de sofisticación tecnológica, integrando múltiples funcionalidades en entornos unificados. Desarrolladas frecuentemente como suites completas, permiten ejecutar flujos de trabajo bibliométricos completos sin cambiar de entorno. Su arquitectura modular facilita desde análisis descriptivos básicos hasta técnicas avanzadas de minería de texto y detección de patrones temáticos en grandes volúmenes de literatura científica.

Las herramientas de visualización especializada ocupan un espacio particular dentro del ecosistema, concentrándose en la representación gráfica de redes y estructuras de conocimiento. Su valor agregado reside en algoritmos específicos para *layout* de redes y técnicas cartográficas que transforman datos bibliométricos en mapas interpretables. Estas soluciones suelen complementar en vez de remplazar el análisis cuantitativo realizado en otras plataformas.

En la cúspide de la complejidad técnica se sitúan las soluciones personalizables mediante programación, que otorgan control total sobre cada aspecto metodológico. Este enfoque basado en código sacrifica inmediatez por flexibilidad absoluta, permitiendo implementar técnicas novedosas no disponibles en herramientas comerciales. Representa la opción preferida para investigación bibliométrica de vanguardia que requiere desarrollos metodológicos originales.

La selección entre estas categorías debe guiarse por criterios específicos que trascienden las preferencias personales. El volumen de datos constituye una consideración primordial, pues herramientas diseñadas para metaanálisis de miles de registros difieren significativamente de aquellas optimizadas para estudios bibliográficos focalizados. La escalabilidad del software determina su idoneidad para proyectos de envergadura variable.

Los recursos computacionales disponibles representan otro factor decisivo frecuentemente subestimado en la planificación investigadora. Algunas herramientas de visualización avanzada demandan capacidades de procesamiento gráfico sustanciales, mientras las soluciones basadas en web externalizan estos requerimientos técnicos. La evaluación realista de la infraestructura tecnológica accesible previene cuellos de botella durante la fase analítica.

La interoperabilidad con fuentes de datos específicas merece especial consideración ante la fragmentación actual del ecosistema bibliográfico. La compatibilidad nativa con Scopus, WoS, Dimensions o PubMed condiciona directamente la viabilidad técnica de muchos proyectos. Las soluciones que ofrecen conectores múltiples y estandarizados reducen significativamente los esfuerzos de preprocesamiento y normalización.

La curva de aprendizaje asociada a cada herramienta debe ponderarse contra el tiempo disponible y la experticia del equipo investigador. Mientras algunas plataformas modernas priorizan la experiencia de usuario mediante interfaces gráficas intuitivas, otras exigen conocimientos especializados que pueden ralentizar las etapas iniciales del proyecto. La formación existente y la capacidad constituyen variables críticas en esta ecuación.

La capacidad de reproducción y transparencia metodológica emerge como criterio decisivo para investigación de alto impacto. Las herramientas que permiten exportar flujos de trabajo completos y scripts ejecutables facilitan la verificación por pares y la replicabilidad estudios. Esta característica resulta particularmente valiosa en contextos de evaluación científica donde la auditabilidad constituye un requisito fundamental.

3.2. Preparación del entorno de trabajo para R

Unos de los grandes gigantes en la investigación bibliométrica son las librerías RBibliometrix(1) y PyBibx,(2) de los lenguajes R y Python respectivamente. Ambos lenguajes para ser usados requieren de configuraciones e instalaciones previas. Las que se abordaran a continuación.

3.2.1. Instalación y configuración de R y RStudio

La preparación del entorno R, para ejecutar código en R, inicia con la descarga de la última versión de R desde el Comprehensive R Archive Network (CRAN) (https://cran.r-project.org/), seleccionando el instalador correspondiente al sistema operativo. Tras completar la instalación básica, se procede a descargar RStudio Desktop (https://posit.co/products/open-source/rstudio), el entorno de desarrollo integrado que optimiza el trabajo con R. La configuración inicial incluye la definición del directorio de trabajo, ajuste de preferencias de apariencia y establecimiento de opciones de rendimiento según las capacidades del equipo, estas se van seleccionando a medida que se muestran en las opciones del propio instalador. Esta base sólida garantiza un entorno estable para la ejecución de paquetes bibliométricos.

Un concepto sumamente importante que se debe conocer a la hora de trabajar con estos lenguajes de programación es el de librería y/o paquete, que es una colección de código pre escrito que amplían las capacidades básicas del lenguaje. Funcionan como un conjunto de herramientas especializadas, ofreciendo funciones y métodos para tareas específicas sin que el desarrollador tenga que crearlas desde cero. Por ejemplo, una librería para ciencia de datos puede contener algoritmos para análisis estadístico o gráficos. Esto no solo acelera enormemente el proceso de desarrollo, sino que garantiza confiabilidad y eficiencia, al aprovechar código ya probado por la comunidad. En esencia, son componentes fundamentales que permiten construir aplicaciones complejas de manera modular y eficaz.

La gestión de paquetes en R constituye una etapa crucial para el análisis bibliométrico. Se recomienda iniciar con la instalación de paquetes esenciales como Bibliometrix, bibliometR y tidyverse desde el repositorio oficial de CRAN, este proceso será abordado en más detalles posteriormente. La configuración de librerías personales evita conflictos entre versiones y facilita la portabilidad de los proyectos. Esta meticulosa preparación previene errores durante los procesos analíticos posteriores.

3.2.2 Herramientas especializadas en R

Bibliometrix: Instalación y configuración completa

La instalación de Bibliometrix en RStudio se realiza mediante el comando install.packages ("bibliometrix"). Tras la instalación, la carga de la librería con library (bibliometrix) habilita todas sus funcionalidades para el análisis bibliométrico integral.

El paquete Bibliometrix ofrece tres interfaces diferenciadas para adaptarse a diversos perfiles de usuarios. La función ***biblioshiny()*** activa una interfaz gráfica web ideal para principiantes o análisis exploratorios rápidos. Para usuarios intermedios, las funciones básicas como ***convert2df()*** y ***biblioAnalysis()*** proporcionan control directo sobre el flujo de trabajo. Usuarios avanzados pueden acceder a todas las funciones mediante programación directa en R, permitiendo personalizaciones metodológicas sofisticadas. Esta flexibilidad hace de Bibliometrix una solución completa para todo tipo de proyectos bibliométricos. En este texto se usarán las dos últimas opciones, para un mayor control sobre el contenido generado.

Figura 3.1. Instalación de R Bibliometrix

3.2.3. Funcionalidades principales y análisis básico

Bibliometrix despliega un repertorio analítico exhaustivo que comienza con el análisis descriptivo básico de la producción científica. La función ***biblioAnalysis*** genera automáticamente indicadores fundamentales como crecimiento temporal, autores más productivos y revistas principales. La identificación de colaboraciones mediante ***networkPlot*** revela patrones de coautoría a nivel individual, institucional y nacional. Estas funcionalidades descriptivas proporcionan la base contextual necesaria antes de emprender análisis más especializados como el mapeo conceptual o el estudio de citas.

El análisis de contenido y tendencias temáticas representa una de las capacidades más poderosas de Bibliometrix. La función ***conceptualStructure*** mediante análisis de co-palabras identifica clústeres temáticos y su evolución temporal. El mapeo de redes de co-citación mediante ***cocMatrix*** revela la estructura intelectual del campo de estudio. La función ***thematicMap*** visualiza los temas según su grado de desarrollo y relevancia, distinguiendo entre temas emergentes, básicos, de nicho o motor. Estas herramientas facilitan la identificación de fronteras del conocimiento y oportunidades de investigación. Como usarlo e interpretarlo se verá con más detalle en capítulos posteriores.

El ecosistema R para bibliometría se enriquece con librerías especializadas que complementan Bibliometrix. Una de estas es ***bibliometR***, que ofrece funcionalidades adicionales para el análisis de redes de citación y la detección de comunidades. ***RefManageR*** facilita la gestión avanzada de referencias bibliográficas en diferentes estilos de citación. Existen otras, pero en la actualidad la más utilizada es ***RBibliometric***, que será el foco de atención en este libro.

3.3. Trabajo con Google Colab para PyBibX

Google Colab representa la plataforma ideal para implementar PyBibX, eliminando los problemas de configuración local. El acceso gratuito mediante una cuenta de Google proporciona inmediatamente un entorno Python preconfigurado con las principales librerías científicas. La conexión a recursos computacionales mejorados, incluyendo GPUs y TPUs, acelera

significativamente el procesamiento de grandes *datasets* bibliométricos. Esta infraestructura en la nube garantiza consistencia entre colaboradores y simplifica la reproducibilidad de los análisis.

Como Google Colab corre en la web no se necesita configuración previa del equipo, solo tener acceso a internet desde el navegador del dispositivo. Solo será necesario Instalar las librerías requeridas, al igual que en el caso de ***RBibliometric***. Esto es mediante el comando ***pip install pybibx***.

Figura 3.2. Proyecto en Google Colab de PyBibx.

La documentación oficial de pybibx (https://pypi.org/project/pybibx/) ofrece una lista de códigos según el tipo de archivo que se vaya a importar para hacer el análisis, por lo que no será necesario aprender cada código o escribirlo manualmente, solo abrir el enlace según el archivo y empezar a trabajar.

Hay otras alternativas a Colab como el uso de Anaconda, pero en este texto se utilizará la misma Colab pues la gestión de sesiones en este optimiza el trabajo con PyBibX en proyectos bibliométricos extensos, la conexión a *runtime* con alta RAM previene fallos durante el procesamiento de corpus documentales voluminosos, la programación de ejecuciones periódicas mediante ***cron jobs*** permite mantener actualizados los análisis sin intervención manual constante. La exportación automática de resultados a Google Drive asegura la persistencia de los hallazgos más allá de la duración de la sesión activa en la plataforma.

Tanto en R Bibliometrix como en Pybibx luego de importar las librerías es necesario cargar el archivo con los datos para el estudio, o set de datos, cada una de estas librerías tiene su forma particular de importarlas y posteriormente analizarlas.

R Bibliometrix

En R (Bibliometrix), se utiliza la función ***convert2df()*** para leer el archivo exportado de bases de datos como Scopus o WoS y convertirlo en un marco de datos (*dataframe*) con el que la librería puede trabajar.

```
# Cargar la librería
  library(bibliometrix)

# Importar y convertir el archivo descargado desde Scopus/WoS en un dataset llamado mi_
dataset
  mi_dataset <- convert2df("mi_archivo_scopus.bib", dbsource = "scopus", format = "bibtex")
```

Nota: Lo que esta luego de # son comentarios, aunque se pongan en RStudio o Colab no serán cargados como código.

Pybibx

En Python (Pybibx) con Google Colab, primero se debe subir el archivo al entorno y luego utilizar la función load del módulo *dataset* para cargarlo en un objeto Collection, que es la estructura central de datos. Esto puede requerir otras importaciones de librerías. La página oficial de Pybibx ofrece una lista de ejemplos de código para cada tipo de archivos de datos, por ejemplos si se exportaron como .bib de Scopus, o txt de PubMed, entre otros según sea su fuente.

```
# Instalar e importar la librería (en Google Colab)
  !pip install pybibx
  from pybibx import dataset

# Subir el archivo .bib o .csv a Colab
  from google.colab import files
  uploaded = files.upload()

# Cargar el dataset en una Colección llamada mi_dataset
 mi_dataset = dataset.load(list(uploaded.keys())[0])
```

No es necesario saber programar para utilizar este libro, los códigos que se muestran pueden copiarse tal cual y funcionara.

3.4. Publish or Perish: Análisis con Google Scholar

3.4.1. Instalación, configuración y tipos de búsqueda

La instalación de *Publish or Perish* comienza con su descarga gratuita desde el sitio oficial de Harzing.com (https://harzing.com/), seleccionando la versión compatible con el sistema operativo del usuario. El proceso de instalación es sencillo y no requiere configuración técnica avanzada, aunque es fundamental permitir las conexiones de red para que el software acceda a los datos de Google Scholar. Tras la instalación, la interfaz principal muestra opciones claramente organizadas que permiten comenzar a realizar búsquedas bibliométricas inmediatamente después de la configuración inicial.

La configuración óptima del software implica personalizar las preferencias según las necesidades específicas de investigación. Entre los ajustes más relevantes se incluye la selección del motor de búsqueda principal, donde Google Scholar representa la opción predeterminada más utilizada. Es crucial configurar el tiempo de espera entre consultas para evitar bloqueos por parte de Google, estableciendo intervalos que simulen el comportamiento humano. La personalización de los formatos de exportación y los límites de resultados completa la preparación para un uso eficiente y sostenible de la herramienta.

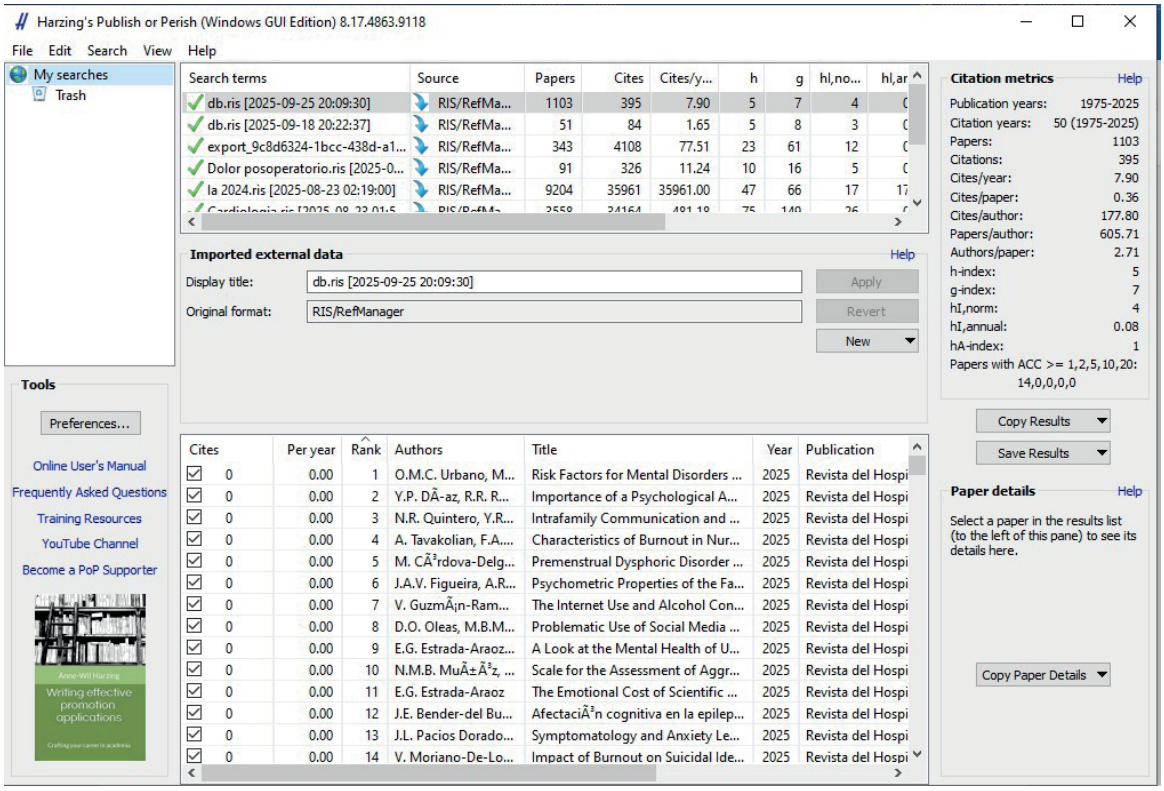


Figura 3.3. Vista de inicio de Harizing's Publish or Perish

Los tipos de búsqueda disponibles en *Publish or Perish* se adaptan a diferentes objetivos bibliométricos. La búsqueda por autor permite localizar y analizar el impacto de investigadores específicos utilizando diferentes variaciones de sus nombres. La búsqueda por artículo facilita el seguimiento de citas de publicaciones individuales mediante la combinación de título, autor y año. La búsqueda general por términos de investigación ofrece una visión amplia de la literatura sobre temas particulares, mientras la búsqueda por revista evalúa el impacto de publicaciones periódicas específicas.

Cada tipo de búsqueda requiere estrategias específicas para optimizar la recuperación de datos. Para autores, es esencial probar diferentes combinaciones de nombres e iniciales para contrarrestar las limitaciones de Google Scholar en la desambiguación. En búsquedas temáticas, la combinación estratégica de operadores booleanos y frases exactas mejora significativamente la precisión de los resultados. La configuración de filtros por año de publicación, idioma y tipo de documento refina los conjuntos de resultados antes del análisis, ahorrando tiempo en el procesamiento posterior de la información recuperada.

El software ofrece capacidades avanzadas de búsqueda que superan las limitaciones de la interfaz web de Google Scholar. La posibilidad de realizar búsquedas en múltiples idiomas amplía el alcance de la recuperación bibliográfica más allá de la literatura anglocéntrica. La función de búsqueda por afiliación institucional, aunque limitada por la cobertura de Google Scholar, proporciona perspectivas valiosas sobre la productividad de diferentes centros de investigación. Estas funcionalidades convierten a *Publish or Perish* en una herramienta indispensable para análisis bibliométricos complementarios a las bases de datos comerciales.

La eficacia de las búsquedas depende en gran medida del entendimiento de las limitaciones inherentes de Google Scholar. La cobertura irregular entre disciplinas, la inclusión de documentos no arbitrados y la duplicación de registros requieren una interpretación cuidadosa de los resultados. El software mitiga parcialmente estos problemas mediante algoritmos de limpieza y des duplicación, pero el juicio crítico del investigador sigue siendo esencial para validar la calidad de los datos obtenidos antes de su análisis bibliométrico formal.

Publish or Perish amplía significativamente sus capacidades al integrar conexiones directas con Scopus, Web of Science, PubMed y otros mediante credenciales institucionales válidas. Esta funcionalidad transforma la herramienta en un puente entre las métricas de Google Scholar y las bases de datos, permitiendo análisis comparativos únicos. Los usuarios pueden autenticarse mediante suscripciones institucionales para acceder a datos normalizados y de mayor calidad bibliográfica. Esta integración posibilita cruzar información entre fuentes, identificando discrepancias y complementariedades en la cobertura de citas para una evaluación más comprehensiva del impacto investigador.

Figura 3.4. Bloque de importación de los datos de Publish or Perish.

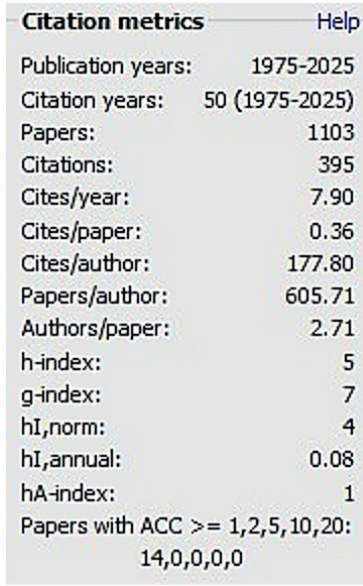
La capacidad de importar datos desde múltiples fuentes constituye una ventaja distintiva de *Publish or Perish* frente a otras herramientas bibliométricas. El software acepta formatos estándar como RIS, BibTeX y CSV provenientes de gestores de referencias o exportaciones de otras plataformas. Esta interoperabilidad facilita la consolidación de información bibliográfica dispersa en un único entorno de análisis métrico. Los investigadores pueden combinar *datasets* de diferentes procedencias para crear colecciones personalizadas que reflejen completamente su producción científica, superando las limitaciones de cobertura de cada fuente individual.

El proceso de importación permite enriquecer los metadatos básicos con indicadores de impacto calculados por *Publish or Perish*. Al cargar referencias desde Zotero, Mendeley o exportaciones de Scopus, el software automáticamente busca las citas correspondientes en Google Scholar y otras fuentes disponibles. Esta funcionalidad es particularmente valiosa para actualizar colecciones bibliográficas existentes con datos de citación recientes sin necesidad de búsquedas manuales. El resultado es un dataset integrado que preserva los metadatos originales mientras añade las métricas de impacto más actualizadas.

3.4.2. Extracción de datos y métricas proporcionadas

El sistema de *Publish or Perish* realiza búsquedas exhaustivas en Google Scholar, Scopus y Web of Science, capturando metadatos esenciales como títulos, autores, afiliaciones, años de publicación y fuentes. La herramienta procesa automáticamente las redes de citación para calcular indicadores avanzados, identificando tanto las citas recibidas como las tendencias temporales de impacto. Este proceso integral asegura la obtención de *datasets* completos para análisis bibliométricos rigurosos.

Entre las métricas principales que calcula automáticamente se incluyen el índice h, índice g y el índice h individual, proporcionando diferentes perspectivas sobre el impacto investigador. El software genera además el número total de citas, citas por año, citas por artículo y el promedio de citas por trabajo. Estos indicadores se complementan con estadísticas temporales que revelan patrones de productividad y evolución del impacto a lo largo de la carrera investigadora. Cada métrica se presenta con su cálculo detallado para facilitar su interpretación contextual.



Citation metrics	Help
Publication years:	1975-2025
Citation years:	50 (1975-2025)
Papers:	1103
Citations:	395
Cites/year:	7.90
Cites/paper:	0.36
Cites/author:	177.80
Papers/author:	605.71
Authors/paper:	2.71
h-index:	5
g-index:	7
hI,norm:	4
hI,annual:	0.08
hA-index:	1
Papers with ACC >= 1,2,5,10,20:	14,0,0,0,0

Figura 3.5. Bloque de resultados del análisis de *Publish or Perish*

La herramienta ofrece análisis comparativos sofisticados que permiten evaluar investigadores, instituciones o revistas contra *benchmarks* disciplinares. El sistema calcula percentiles de impacto, posiciones relativas y tendencias comparativas mediante algoritmos normalizados. Esta funcionalidad es particularmente valiosa para estudios bibliométricos que requieren contextualizar el desempeño dentro de campos específicos de conocimiento. Los resultados incluyen visualizaciones claras que facilitan la identificación de fortalezas y áreas de mejora en los perfiles investigadores analizados.

Los datos extraídos incluyen información cualitativa fundamental para interpretar correctamente las métricas cuantitativas. El software captura los títulos completos, resúmenes cuando están disponibles, y las referencias citadas más relevantes, permitiendo comprender las razones detrás del impacto medido, diferenciando entre citas de reconocimiento, metodológicas o controversias académicas. La integración de dimensiones cualitativas y cuantitativas enriquece sustancialmente el análisis bibliométrico resultante.

3.4.3. Exportación e integración con otras herramientas

Publish or Perish ofrece capacidades avanzadas de exportación que facilitan la interoperabilidad con el ecosistema bibliométrico. Los usuarios pueden exportar resultados completos en formatos CSV, RIS o BibTeX, preservando todos los metadatos y métricas calculadas. Esta flexibilidad permite integrar los datos con herramientas especializadas como VOSviewer para visualización de redes, Bibliometrix para análisis estadísticos avanzados, o CitNetExplorer

para estudiar redes de citación. La exportación mantiene la estructura relacional de los datos, asegurando que las conexiones entre autores, artículos y citas permanezcan intactas durante la transferencia.

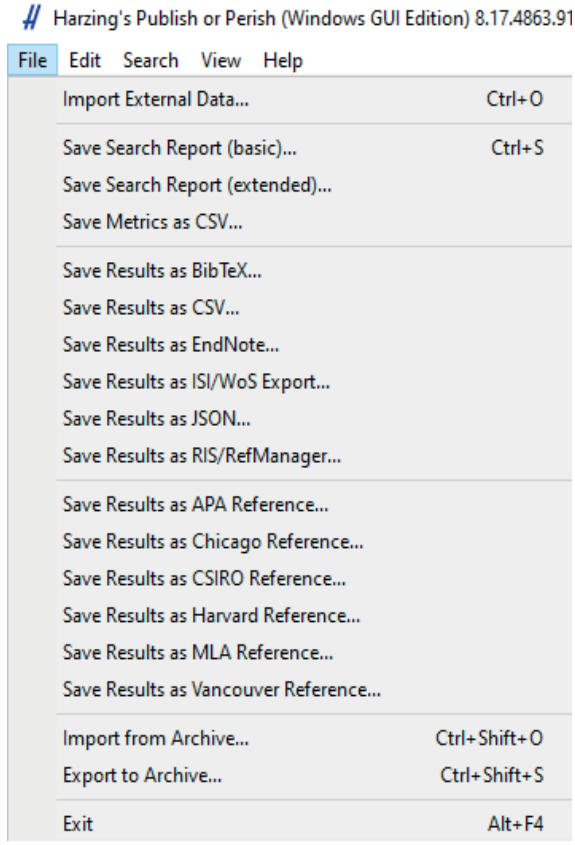


Figura 3.6. Menú de exportación de *Publish or Perish*

La integración con gestores bibliográficos representa una funcionalidad particularmente valiosa para investigadores. Los datos exportados en formato RIS pueden importarse directamente en Zotero, Mendeley o EndNote, enriqueciendo las bibliotecas personales con las métricas de impacto calculadas. Esta característica elimina la necesidad de procesos manuales de actualización y garantiza que las colecciones bibliográficas mantengan información métrica actualizada. La sincronización bidireccional permite trabajar simultáneamente en la gestión de referencias y el análisis bibliométrico dentro de flujos de trabajo cohesionados.

Para análisis cuantitativos avanzados, la exportación en formato CSV proporciona compatibilidad inmediata con software estadístico como R, Python o SPSS. Las tablas resultantes incluyen columnas normalizadas para cada variable bibliométrica, facilitando su procesamiento mediante scripts personalizados. Investigadores pueden combinar estos datos con información adicional proveniente de otras fuentes, creando *datasets* enriquecidos para modelos multivariados. Esta capacidad transforma a *Publish or Perish* en una herramienta de extracción inicial dentro de pipelines analíticos complejos.

Las opciones de exportación incluyen configuraciones personalizables para adaptarse a requisitos específicos de investigación. Los usuarios pueden seleccionar subconjuntos de métricas, filtrar por rangos temporales o elegir formatos de codificación específicos. Esta

granularidad asegura que los datos transferidos a otras herramientas contengan exactamente la información necesaria para cada tipo de análisis posterior. La combinación de flexibilidad y precisión en la exportación posiciona a *Publish or Perish* como un componente fundamental en flujos de trabajo bibliométricos integrados.

3.4.4. Limitaciones y buenas prácticas de uso

Publish or Perish presenta limitaciones significativas que los usuarios deben reconocer para interpretar correctamente los resultados. La dependencia principal de Google Scholar introduce sesgos inherentes relacionados con la cobertura irregular entre disciplinas, la inclusión de documentos no arbitrados y problemas de duplicación de registros. La herramienta carece de mecanismos robustos de desambiguación automática de autores, lo que puede distorsionar los perfiles individuales cuando existen investigadores homónimos. Además, las métricas calculadas reflejan principalmente impacto cuantitativo sin considerar dimensiones cualitativas como el prestigio de las revistas o el tipo de contribución científica.

Entre las buenas prácticas esenciales se encuentra la verificación cruzada sistemática de resultados con bases de datos complementarias como Scopus o Web of Science. Los usuarios deben implementar estrategias de búsqueda conservadoras, utilizando múltiples variaciones de nombres para autores y validando manualmente los perfiles más relevantes. Es crucial contextualizar las métricas dentro de las normas disciplinares específicas, reconociendo que valores idénticos del índice h poseen significados diferentes entre física teórica y filosofía. La transparencia metodológica exige documentar todas las decisiones de búsqueda y limpieza de datos en los informes finales.

La interpretación responsable de las métricas requiere comprender sus fundamentos metodológicos y limitaciones técnicas. Los usuarios deben complementar las métricas cuantitativas con evaluación cualitativa, revisando los trabajos más citados para determinar su real contribución al campo. Esta aproximación balanceada previene conclusiones erróneas basadas exclusivamente en indicadores numéricos sin contexto interpretativo.

La gestión ética de los datos obtenidos implica respetar los límites de uso establecidos por las fuentes y evitar la sobrecarga de sus servidores. Las búsquedas masivas deben programarse con intervalos razonables entre consultas, mientras el almacenamiento de resultados debe seguir protocolos de seguridad estándar. Los usuarios institucionales deben establecer directrices claras sobre el uso apropiado de las métricas en evaluaciones académicas, alineándose con los principios de la Declaración DORA para evitar reduccionismos en la valoración de la investigación.

3.5. Herramientas de visualización avanzada

La visualización de datos bibliométricos representa una etapa crucial en la interpretación y comunicación de resultados, transformando complejos conjuntos de datos en representaciones gráficas intuitivas que revelan patrones, relaciones y tendencias subyacentes en la literatura científica. Estas herramientas especializadas permiten a investigadores y evaluadores identificar redes de colaboración, estructuras temáticas y dinámicas de conocimiento que permanecerían ocultas en tablas y listados convencionales, facilitando tanto el análisis especializado como la divulgación efectiva de hallazgos bibliométricos a audiencias multidisciplinares.

3.5.1. VOSviewer: instalación y creación de mapas científicos

La instalación de VOSviewer comienza con la descarga gratuita desde el sitio oficial del Centro de Estudios de Ciencia y Tecnología de la Universidad de Leiden, disponible para Windows, Mac OS y Linux (https://www.VOSviewer.com/download). El proceso de instalación es sencillo y

no requiere configuración técnica avanzada, aunque es recomendable verificar que el sistema cuente con la última versión de Java para garantizar el funcionamiento óptimo de todas las funcionalidades. Una vez instalado, el entorno de trabajo se presenta organizado en módulos lógicos que guían al usuario a través del flujo completo de creación de mapas científicos.

La creación de mapas en VOSviewer inicia con la importación de datos desde fuentes bibliográficas como Scopus, Web of Science o PubMed, procesados previamente en formatos compatibles, en el próximo capitulo se describirá el proceso de obtención de los mismos. El software ofrece tres tipos principales de análisis: redes de coautoría para visualizar colaboraciones científicas, redes de co-citación para revelar la estructura intelectual de un campo, y mapas de coocurrencia de términos para identificar temas de investigación. Cada tipo de análisis emplea algoritmos específicos de *layout* y *clustering* que optimizan la representación visual de las relaciones bibliométricas.

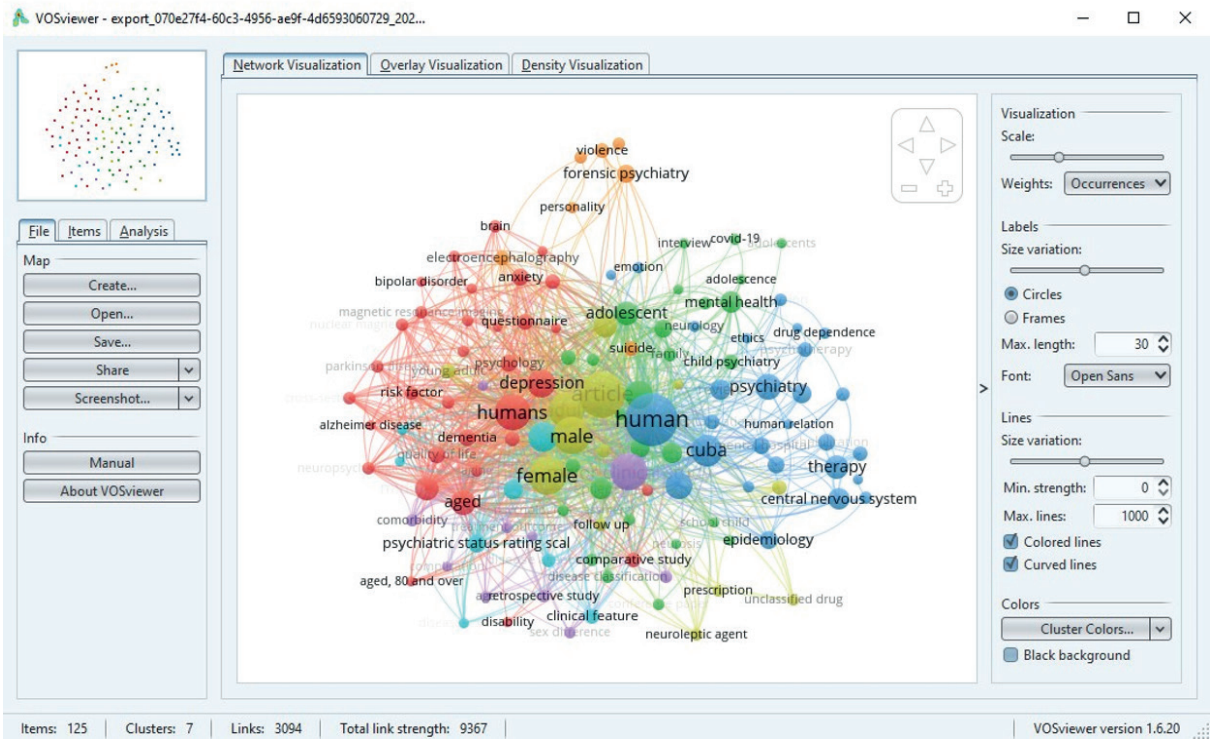


Figura 3.7. Vista principal de VOSviewer, con proyecto graficado

El proceso de mapeo implica varias etapas de configuración donde el usuario define parámetros clave como el umbral mínimo de ocurrencias, el método de normalización y el algoritmo de agrupamiento. VOSviewer aplica técnicas avanzadas de reducción de dimensionalidad mediante el algoritmo VOS (*Visualization of Similarities*), que posiciona los elementos en un espacio bidimensional preservando sus relaciones de similitud. La herramienta genera automáticamente clústeres diferenciados por colores que representan comunidades temáticas o grupos de colaboración, facilitando la identificación visual de patrones estructurales.

La personalización de los mapas permite ajustar múltiples aspectos visuales como tamaños de nodos, grosores de enlaces, fuentes de texto y esquemas de color. Los nodos pueden dimensionarse

según diferentes métricas como frecuencia de aparición, número de citas o centralidad en la red, mientras el grosor de los enlaces refleja la fuerza de las relaciones entre elementos. Estas opciones de personalización permiten adaptar la visualización a las necesidades específicas de cada análisis y audiencia objetivo, mejorando la claridad comunicativa de los resultados.

La interpretación de los mapas generados requiere comprender tanto la posición relativa de los elementos como su agrupamiento en clústeres temáticos. La distancia entre nodos indica su grado de relación temática o colaborativa, mientras la distribución espacial revela la estructura global del campo de estudio. Los mapas permiten identificar temas emergentes, autores centrales, instituciones líderes y patrones de colaboración internacional, proporcionando perspectivas valiosas para la evaluación científica y la planificación de investigación futura.

VOSviewer incluye funcionalidades avanzadas como el análisis temporal mediante mapas superpuestos, la exportación en formatos vectoriales para publicaciones académicas, y herramientas de zoom y filtrado para explorar áreas específicas de interés. La integración con otros softwares bibliométricos mediante formatos de intercambio estándar completa un ecosistema de visualización robusto que respalda investigaciones bibliométricas desde el análisis exploratorio inicial hasta la presentación final de resultados complejos.

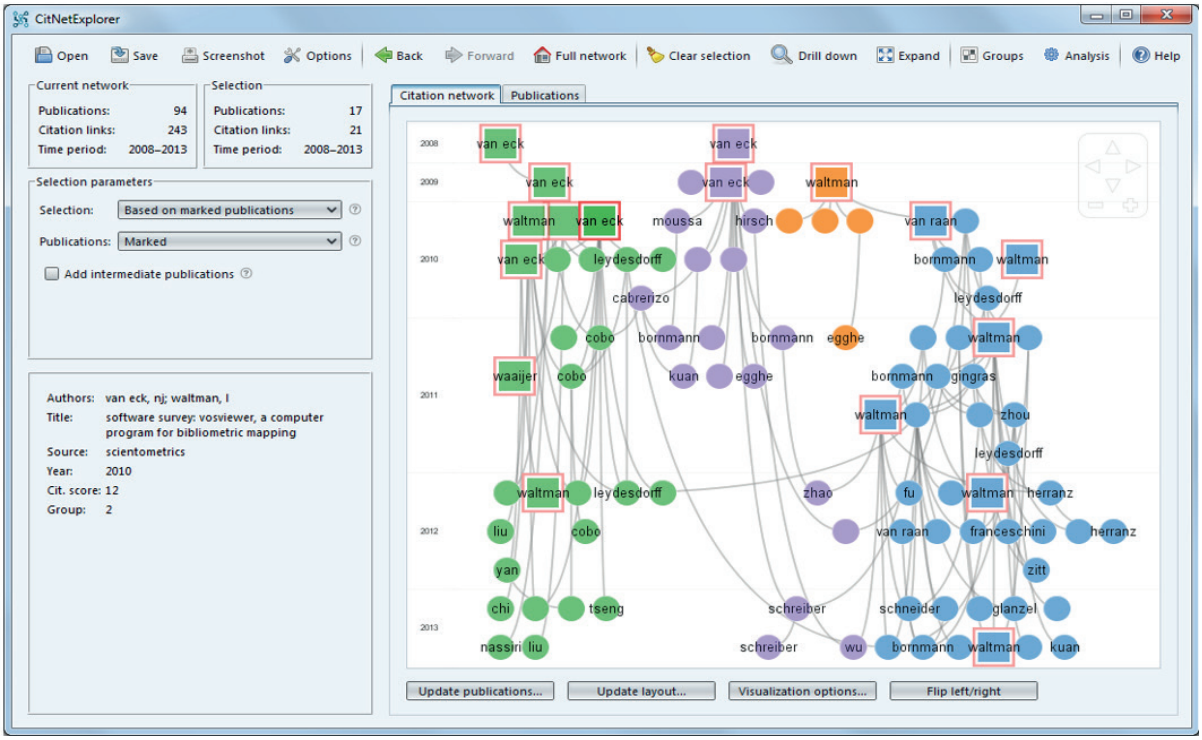
3.5.2. CitNetExplorer: análisis de redes de citación

CitNetExplorer se especializa en el análisis y visualización de redes de citación, permitiendo explorar la estructura y evolución de campos científicos a través de las relaciones entre publicaciones. Desarrollado por la Universidad de Leiden, este software permite identificar publicaciones fundamentales, trazar el desarrollo de líneas de investigación y analizar patrones de citación a lo largo del tiempo. Su enfoque único facilita el estudio de la herencia intelectual y las trayectorias de conocimiento dentro de disciplinas científicas específicas.

La instalación del software es directa, requiriendo solamente la descarga desde el sitio oficial y la descompresión del archivo ZIP en el directorio deseado (https://www.citnetexplorer.nl/download). CitNetExplorer no necesita instalación convencional, ejecutándose directamente desde el archivo JAR con Java Runtime Environment 8 o superior. Esta portabilidad permite usar la herramienta en diferentes sistemas sin procesos de configuración complejos, aunque se recomienda verificar los permisos de ejecución en sistemas operativos basados en Unix para garantizar el funcionamiento correcto.

El proceso de análisis comienza con la importación de datos desde Web of Science o Scopus, utilizando formatos de exportación específicos que preservan las referencias citadas. La herramienta construye automáticamente la red de citación, donde los nodos representan publicaciones y los enlaces las relaciones de citación. Los algoritmos de visualización organizan temporalmente las publicaciones, mostrando claramente la evolución del campo y permitiendo identificar trabajos seminales que funcionan como puntos de conexión entre diferentes líneas de investigación.

Las funcionalidades de exploración incluyen herramientas de zoom y filtrado temporal que permiten concentrarse en períodos específicos. Los usuarios pueden seleccionar publicaciones clave para visualizar su red local de citaciones, identificando trabajos predecesores y sucesores. El análisis de caminos de citación revela cómo las ideas científicas se transmiten y transforman a través de diferentes publicaciones, proporcionando perspectivas sobre la dinámica de difusión del conocimiento dentro de la comunidad académica.



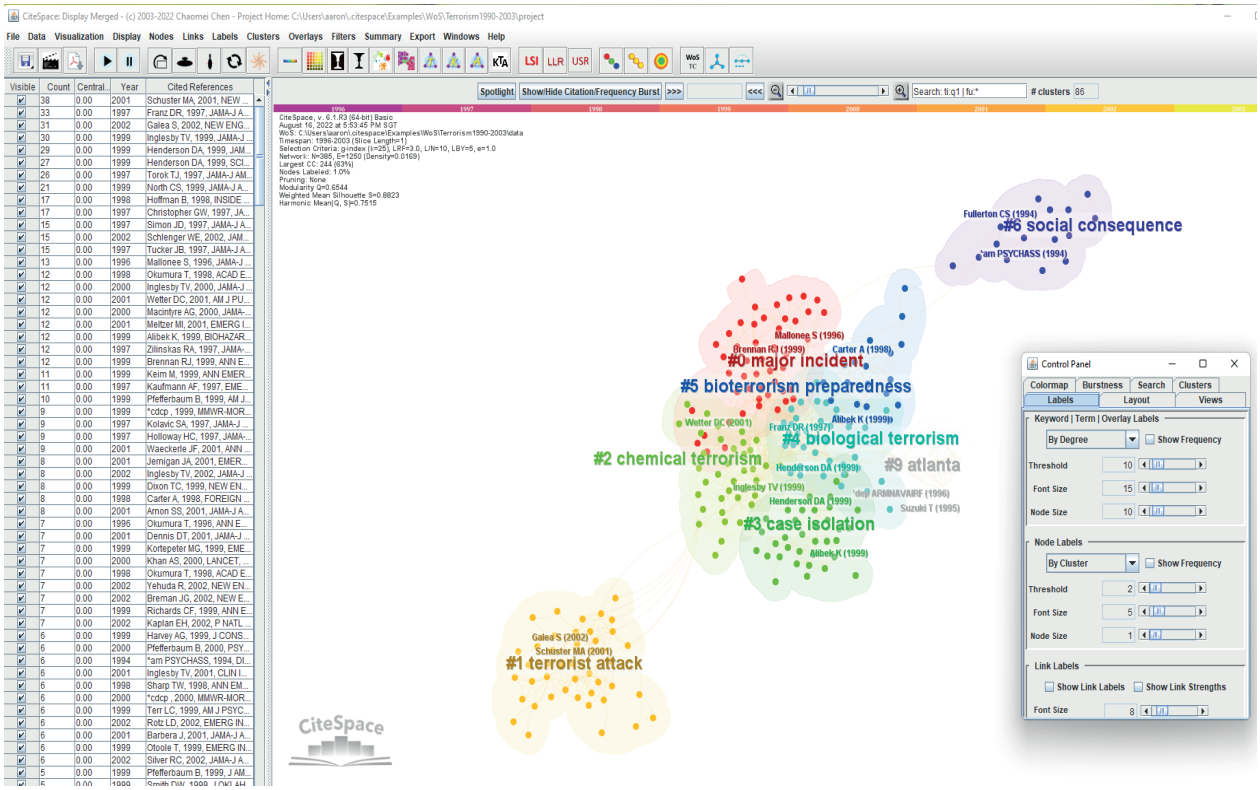
Fuente: Sitio oficial de CitNetExplorer (https://www.citnetexplorer.nl)
Figura 3.8. Vista principal de CitNetExplorer, con proyecto graficado

La identificación de clústeres temáticos se realiza mediante algoritmos que detectan comunidades de publicaciones altamente conectadas. Estos grupos representan subcampos o líneas de investigación específicas, cuya evolución puede rastrearse temporalmente. CitNetExplorer permite analizar cómo emergen, convergen o divergen diferentes tradiciones intelectuales, ofreciendo perspectivas únicas sobre la estructuración del conocimiento científico que complementan los análisis proporcionados por otras herramientas de visualización bibliométrica.

Las opciones de exportación incluyen la generación de reportes detallados con métricas de red, imágenes de alta resolución para publicaciones, y datos procesados para análisis complementarios. La herramienta permite guardar sesiones completas de trabajo, facilitando la continuidad en investigaciones extensas. Esta capacidad de preservar el estado del análisis resulta particularmente valiosa cuando se trabajan con grandes volúmenes de datos o se desarrollan estudios longitudinales complejos sobre evolución científica.

3.5.3. CiteSpace: análisis de patrones temporales

CiteSpace se distingue como herramienta especializada en el análisis de patrones temporales y la detección de tendencias emergentes en la literatura científica. Desarrollado por el Dr. Chaomei Chen, este software permite mapear la evolución de dominios científicos mediante técnicas avanzadas de minería de datos y visualización dinámica. Su capacidad para identificar «*bursts*» de citación y puntos de inflexión en el desarrollo científico lo convierte en instrumento invaluable para estudios de prospectiva tecnológica y análisis diacrónicos de la investigación.



Fuente: Sitio oficial de CiteSpace (https://www.java.com/en/download)
Figura 3.9. Vista principal de CiteSpace, con proyecto graficado

La instalación requiere instalar previamente Java Runtime Environment (https://www.java.com/en/download) y se realiza mediante archivo JAR ejecutable disponible en el sitio oficial (https://citespace.podia.com/). El proceso incluye la configuración de parámetros de memoria para optimizar el rendimiento con grandes volúmenes de datos. CiteSpace organiza su interfaz en paneles modulares que gestionan secuencialmente la importación de datos, el procesamiento analítico y la visualización de resultados, manteniendo un flujo de trabajo estructurado, aunque con curva de aprendizaje considerable para usuarios noveles.

El análisis temporal se fundamenta en la construcción de redes de citación segmentadas por períodos definidos por el usuario. El software aplica algoritmos de detección de comunidades para identificar clústeres temáticos y calcula métricas de centralidad para ubicar publicaciones puente entre diferentes áreas del conocimiento. La superposición de redes consecutivas permite visualizar la emergencia, convergencia o desaparición de líneas de investigación a lo largo del tiempo, revelando dinámicas de cambio científico.

La función de detección de «*bursts*» de citación identifica publicaciones que experimentan incrementos súbitos en su frecuencia de citación, señalando contribuciones potencialmente revolucionarias o temas de investigación emergentes. Estos *bursts* se visualizan mediante anillos de color en los nodos de la red, donde la intensidad y timing del color indican la magnitud y duración del período de alta citación. Esta característica permite identificar rápidamente trabajos que han marcado puntos de inflexión en su campo.

Las visualizaciones de paisajes científicos emplean modelos topográficos donde las elevaciones representan densidad de conexiones y los valles indican discontinuidades temáticas. CiteSpace genera mapas evolutivos que muestran la deriva de los clústeres temáticos a través del tiempo, complementados con líneas de tendencia que proyectan desarrollos futuros. Estas representaciones facilitan la identificación de oportunidades de investigación y áreas con potencial de crecimiento e innovación.

La exportación de resultados incluye reportes métricos detallados, animaciones de evolución temporal y datos procesados para análisis secundarios. CiteSpace genera tablas de clústeres temáticos con sus publicaciones fundamentales, métricas de cohesión y etiquetas representativas extraídas mediante algoritmos de procesamiento de lenguaje natural. Esta documentación exhaustiva respalda la interpretación de los patrones detectados y facilita la comunicación de hallazgos complejos sobre dinámicas temporales en la ciencia.

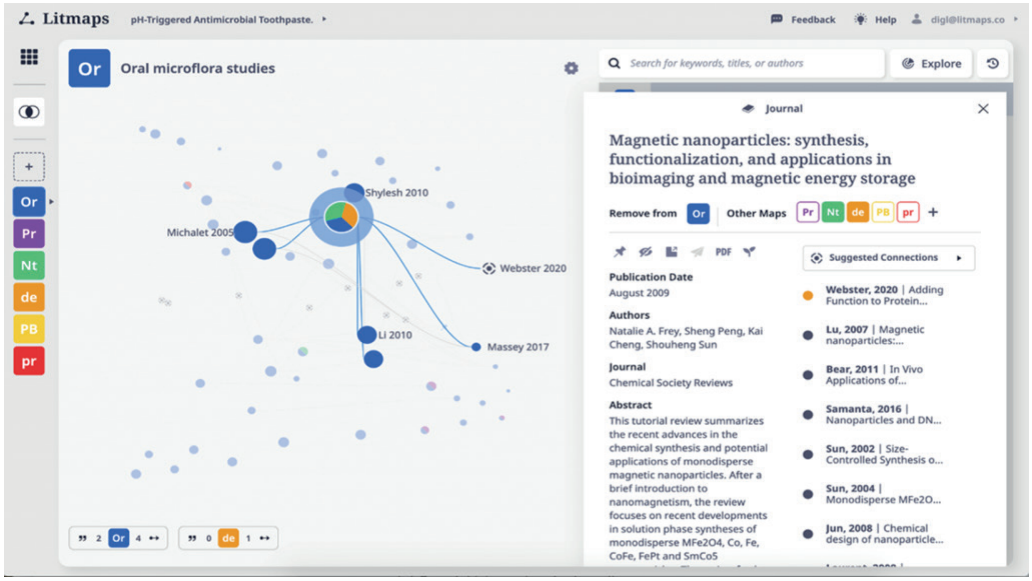
3.5.4. Herramientas online (Litmaps, ResearchRabbit)

Las herramientas online representan la evolución más reciente en visualización bibliométrica, ofreciendo accesibilidad inmediata y funcionalidades colaborativas en la nube. Litmaps y ResearchRabbit emergen como plataformas innovadoras que transforman la exploración literatura mediante interfaces intuitivas y algoritmos de recomendación avanzados. Estas soluciones web permiten descubrir conexiones ocultas entre publicaciones, realizar seguimiento de desarrollos científicos en tiempo real y colaborar en equipos de investigación distribuidos geográficamente, marcando una transición significativa hacia la bibliometría colaborativa y en tiempo real.

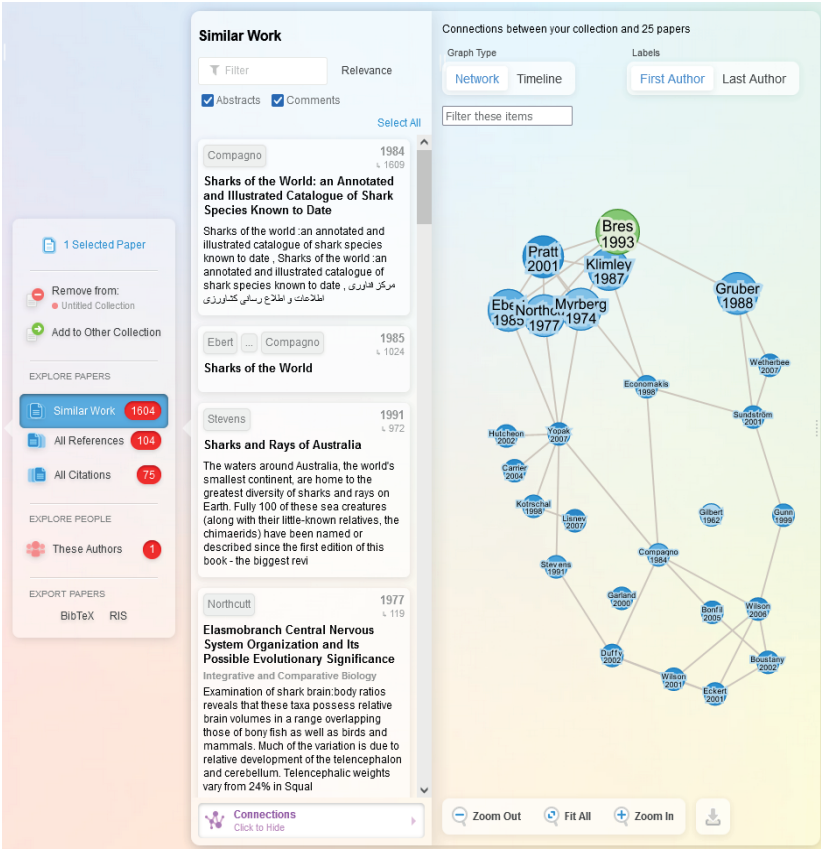
Litmaps (https://www.litmaps.com/) se destaca por su capacidad para generar mapas de citación interactivos que revelan las relaciones entre publicaciones de manera dinámica. La plataforma permite visualizar tanto las referencias citadas como las citas posteriores de un artículo seminal, creando redes de conocimiento que se expanden orgánicamente. Su algoritmo de «*Seed Maps*» identifica automáticamente las publicaciones más relevantes sobre un tema, mientras la función de alertas notifica sobre nuevas investigaciones relacionadas. Esta aproximación facilita la exploración literaria serendípica y el descubrimiento de conexiones interdisciplinarias inesperadas.

ResearchRabbit (https://researchrabbitapp.com/) adopta un enfoque diferente, funcionando como «Spotify para la investigación académica» mediante recomendaciones personalizadas basadas en preferencias de lectura. La herramienta permite crear colecciones de artículos y recibe sugerencias cada vez más afinadas mediante algoritmos de aprendizaje automático. Su interfaz de tabla de tiempo visualiza la evolución histórica de líneas de investigación, mientras las funciones de colaboración permiten compartir y comentar colecciones con equipos de investigación. Esta aproximación centrada en el usuario transforma la exploración literaria en experiencia personalizada y acumulativa.

La implementación de estas herramientas requiere solamente navegador web moderno y registro gratuito, eliminando barreras de instalación y compatibilidad de sistemas operativos. Ambas plataformas sincronizan datos en la nube, permitiendo acceso desde múltiples dispositivos y la recuperación automática de sesiones de trabajo. La integración con gestores de referencias como Zotero y Mendeley facilita la exportación de descubrimientos relevantes, mientras las conexiones con bases de datos principales aseguran cobertura comprehensiva de literatura actualizada.



Fuente: Sitio oficial de Litmaps (https://www.litmaps.com/)
Figura 3.10. Vista principal de Litmaps, con proyecto graficado



Fuente: Sitio oficial de ResearchRabbit (https://researchrabbitapp.com/)
Figura 3.11. Vista principal de ResearchRabbit, con proyecto graficado

Entre las limitaciones prácticas se encuentra la dependencia de conexión internet estable y consideraciones sobre privacidad de datos de investigación. La naturaleza cerrada de sus algoritmos puede dificultar la validación metodológica de las recomendaciones, y las versiones gratuitas típicamente imponen límites en el volumen de análisis. No obstante, su usabilidad inmediata y capacidad para reducir el tiempo de exploración literaria las convierte en complementos valiosos para el kit de herramientas bibliométricas modernas, particularmente en etapas iniciales de investigación y para investigadores en formación.

3.6. Selección de herramientas

La selección adecuada de herramientas bibliométricas requiere una evaluación sistemática que considere los objetivos específicos de investigación, los recursos disponibles y las características técnicas de cada software. Este capítulo proporciona un marco comparativo integral para orientar la elección de herramientas según diferentes escenarios de análisis, presentando tanto las capacidades individuales como las sinergias potenciales entre plataformas complementarias. La estrategia óptima frecuentemente combina múltiples herramientas en flujos de trabajo integrados que aprovechan las fortalezas distintivas de cada una.

Para el ***mapeo de redes de colaboración científica***, VOSviewer constituye la opción preferida gracias a sus algoritmos de layout optimizados para visualizar relaciones entre autores e instituciones. Cuando el análisis requiere identificación de comunidades y patrones estructurales en grandes redes, la combinación de Bibliometrix para el procesamiento estadístico y VOSviewer para la visualización produce resultados particularmente robustos. En proyectos que priorizan la claridad comunicativa sobre el detalle analítico, las herramientas online como ResearchRabbit ofrecen visualizaciones inmediatamente comprensibles para audiencias no especializadas.

El ***análisis de redes de citación y evolución temporal*** encuentra en CiteSpace su herramienta más potente, especialmente para detectar puntos de inflexión y tendencias emergentes en dominios científicos en rápida evolución. Cuando el objetivo principal es rastrear el desarrollo histórico de ideas específicas, CitNetExplorer proporciona capacidades diacrónicas superiores para visualizar trayectorias de conocimiento. Para estudios de prospectiva tecnológica que requieren identificar áreas de oportunidad investigadora, la combinación de CiteSpace con el análisis de «*bursts*» de Litmaps ofrece perspectivas complementarias valiosas.

En contextos de ***evaluación científica individual o institucional***, *Publish or Perish* permite cálculos rápidos de indicadores básicos a partir de Google Scholar, mientras Bibliometrix ofrece análisis más comprehensivos integrando múltiples fuentes de datos. Para comités de evaluación que requieren métricas normalizadas y comparables, la combinación de ambas herramientas proporciona tanto inmediatez como profundidad analítica. En entornos institucionales con recursos técnicos limitados, las herramientas online presentan la ventaja de requerir mínima infraestructura y capacitación previa.

La integración efectiva entre herramientas bibliométricas sigue el principio de modularidad, donde cada software contribuye sus fortalezas específicas a un flujo de trabajo coherente. Una estrategia común combina la extracción de datos con *Publish or Perish*, el procesamiento estadístico con Bibliometrix o PyBibx, y la visualización avanzada con VOSviewer o CiteSpace. Este enfoque de flujo secuencial de extracción y procesamiento de datos maximiza las capacidades individuales mientras mitiga las limitaciones particulares de cada herramienta, produciendo análisis más comprehensivos que los posibles con cualquier plataforma individual.

La interoperabilidad técnica se facilita mediante formatos de intercambio estandarizados

como RIS, BibTeX y CSV que permiten transferir datos entre herramientas con mínima pérdida de información. Scripts personalizados en R o Python pueden automatizar estas conversiones, especialmente útiles cuando se integran herramientas basadas en diferentes ecosistemas tecnológicos. La documentación meticulosa de los parámetros utilizados en cada etapa asegura la reproducibilidad del análisis integrado y facilita la identificación de fuentes de discrepancia entre resultados obtenidos con diferentes herramientas.

Las estrategias de integración deben considerar tanto los aspectos técnicos como los recursos humanos disponibles. Equipos con experticia en programación pueden implementar flujos de trabajo altamente personalizados que combinen Bibliometrix con análisis complementarios en R o Python. Para grupos con menores capacidades técnicas, la combinación de herramientas con interfaces gráficas como VOSviewer y *Publish or Perish* ofrece un balance efectivo entre capacidades analíticas y usabilidad práctica. La formación cruzada en herramientas complementarias representa una inversión estratégica que amplía significativamente las capacidades analíticas del equipo investigador.

Recapitulando

- Las herramientas bibliométricas permiten extraer, procesar, visualizar y analizar datos científicos provenientes de bases especializadas.
- Su uso sistemático mejora la precisión y reproducibilidad de los estudios bibliométricos.
- Existen tres tipos principales de herramientas:
  - De recuperación y extracción de datos.
  - De análisis y procesamiento estadístico.
  - De visualización y mapeo de información científica.
- Las bases de datos bibliográficas más utilizadas son Web of Science (WoS), Scopus, Dimensions, PubMed, Lens y Google Scholar.
- Web of Science ofrece indicadores clásicos como el *Journal Impact Factor* y es fundamental para estudios longitudinales.
- Scopus, de Elsevier, incluye el indicador *CiteScore* y el sistema *SCImago Journal Rank* (SJR).
- Google Scholar es una fuente abierta, útil para análisis de cobertura amplia, aunque con menor control de calidad.
- Entre las herramientas de análisis destacan VOSviewer, Bibliometrix (R), CiteSpace, SciMAT y Gephi.
- VOSviewer es ampliamente usado para construir mapas de coocurrencia de palabras, coautorías y co-citaciones.
- Bibliometrix, paquete de R, permite un análisis estadístico avanzado y la exportación a Biblioshiny, su interfaz web interactiva.
- CiteSpace se centra en la detección de tendencias y clústeres temáticos mediante redes temporales de citación.
- SciMAT se utiliza para análisis evolutivos de la producción científica, visualizando la dinámica de los temas a lo largo del tiempo.
- Gephi permite representar redes complejas de colaboración y co-citación con alta personalización visual.
- También se emplean herramientas estadísticas y de minería de texto (Excel, SPSS, R, Python) para procesar datos exportados.
- La combinación de varias herramientas incrementa la validez del análisis y permite perspectivas complementarias.
- La selección de la herramienta depende del objetivo del estudio, tamaño de la

muestra y nivel de detalle deseado.

- Es fundamental mantener criterios de transparencia y trazabilidad en los procesos de extracción y limpieza de datos.
- La visualización bibliométrica facilita la interpretación de resultados y la comunicación científica.
- Las herramientas deben ser utilizadas con criterios éticos y metodológicos coherentes, evitando manipulación de indicadores.
- El dominio de estas herramientas constituye una competencia esencial para investigadores, analistas y gestores de ciencia.

Preguntas de autoevaluación

1. ¿Cuáles son las tres categorías principales de herramientas bibliométricas?
2. ¿Qué funciones cumplen las bases de datos como Web of Science y Scopus?
3. ¿Qué indicador caracteriza principalmente a Web of Science?
4. ¿Qué sistema de métricas ofrece Scopus además del CiteScore?
5. ¿Qué ventajas y limitaciones tiene Google Scholar para análisis bibliométricos?
6. ¿Para qué se utiliza VOSviewer en los estudios de redes científicas?
7. ¿Qué características diferencian a Bibliometrix de otras herramientas?
8. ¿Qué aporta CiteSpace al estudio de las tendencias temáticas?
9. ¿Qué importancia tiene la visualización en el análisis bibliométrico?
10. ¿Por qué es importante documentar con transparencia los procesos de extracción y análisis de datos?

BIBLIOGRAFÍA

1. Moed HF. Citation analysis in research evaluation. Dordrecht: Springer; 2005. https://doi.org/10.1007/1-4020-3714-7

2. Thelwall M. Web indicators for research evaluation: a practical guide. San Rafael (CA): Morgan & Claypool; 2016. https://doi.org/10.2200/S00733ED1V01Y201602ICR048

3. Sugimoto CR, Larivière V. Measuring research: what everyone needs to know. Oxford: Oxford University Press; 2018. https://global.oup.com/academic/product/measuring-research-9780190640125

4. Ding Y, Rousseau R, Wolfram D. Measuring scholarly impact: methods and practice. Cham: Springer; 2014. https://doi.org/10.1007/978-3-319-10377-8

5. Börner K, Chen C, Boyack KW. Visualizing knowledge domains. In: Cronin B, Sugimoto CR, editors. Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. Cambridge (MA): MIT Press; 2014. p. 197–228. https://doi.org/10.7551/mitpress/9780262026792.003.0010

6. Chen C. CiteSpace: a practical guide for mapping scientific literature. Hauppauge (NY): Nova Science Publishers; 2016. ISBN 9781634842847

REFERENCIAS BIBLIOGRÁFICAS

1. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 1 de noviembre de 2017;11(4):959-75.

2. Pereira V, Basilio MP, Santos CHT. PyBibX – a Python library for bibliometric and scientometric

analysis powered with artificial intelligence tools. Data Technologies and Applications. 2025;59(2):302-37.