Chapter 6 / Capítulo 6

Applied bibliometrics. From data to publication (Spanish Edition)
ISBN: 978-9915-9680-6-3
DOI: 10.62486/978-9915-9680-6-3.ch06

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

CORRELATION GRAPHS / GRAFOS DE CORRELACIÓN

6.1. Fundamentos teóricos de los grafos bibliométricos

Los grafos de correlación representan una herramienta fundamental para el análisis de relaciones en la literatura científica, permitiendo visualizar y cuantificar conexiones entre entidades académicas. Basados en la teoría de grafos matemática, estas estructuras transforman datos bibliográficos en redes donde los elementos del sistema investigador se interconectan mediante relaciones significativas. La potencia analítica de los grafos bibliométricos reside en su capacidad para revelar patrones estructurales subyacentes en grandes volúmenes de datos, facilitando la identificación de comunidades intelectuales, tendencias emergentes y dinámicas de colaboración que permanecerían ocultas en análisis tabulares convencionales.

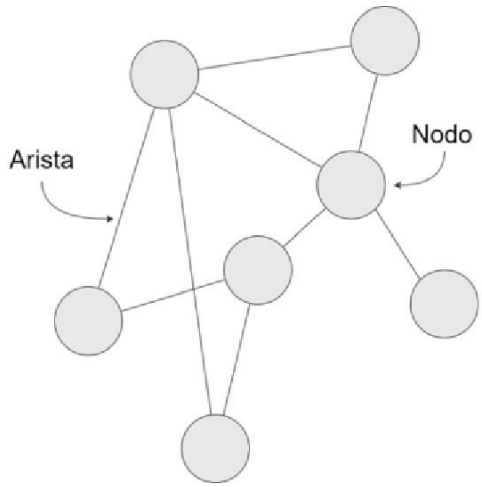


Figura 6.1. Estructura general de un grafo

6.1.1. Nodos: autores, palabras clave y revistas como unidades de análisis

Los nodos constituyen los elementos fundamentales del grafo, representando las entidades académicas cuyo estudio revela la estructura del conocimiento. Cuando los nodos representan autores, el grafo visualiza comunidades de colaboración, donde la posición relativa indica centralidad en redes de coautoría y la proximidad sugiere afinidad temática o institucional. Los investigadores altamente conectados funcionan como puentes entre diferentes grupos, mientras los nodos periféricos pueden indicar especialización temática o aislamiento académico. La densidad de conexiones alrededor de un autor específico refleja su grado de integración en la comunidad científica, proporcionando perspectivas sobre estrategias de colaboración y liderazgo intelectual.

En los mapas basados en palabras clave, los nodos capturan conceptos y temas de investigación, transformando el contenido intelectual en estructuras espaciales interpretables. La frecuencia de aparición de términos se traduce en tamaño nodal, mientras la co-ocurrencia en los mismos documentos establece conexiones entre conceptos. Esta aproximación revela la arquitectura conceptual de un campo científico, identificando núcleos temáticos consolidados, áreas emergentes y vacíos de investigación. Los clústeres terminológicos naturales emergen mediante algoritmos de detección de comunidades, mostrando cómo se organiza el conocimiento y cómo diferentes subcampos se relacionan entre sí dentro de un dominio disciplinar.

Las revistas científicas como nodos permiten analizar la estructura de comunicación

académica y los flujos de conocimiento entre diferentes especialidades. Los grafos de citación entre revistas muestran relaciones de influencia intelectual, donde la dirección de las citas indica flujos de conocimiento y la fuerza de las conexiones refleja proximidad disciplinar. Este enfoque identifica revistas centrales que funcionan como centros de difusión, publicaciones puente que conectan diferentes campos, y periferias especializadas. El análisis diacrónico de estas redes puede revelar la evolución de las fronteras disciplinarias y el surgimiento de nuevas áreas interdisciplinarias de investigación.

6.1.2. Aristas: peso y dirección como medidas de relación

Las aristas constituyen el elemento conectivo del grafo, cuantificando las relaciones entre nodos mediante dos dimensiones críticas: peso y dirección. El peso de una arista refleja la intensidad de la conexión, calculándose mediante diferentes métricas según el tipo de análisis. En redes de coautoría, el peso puede representar el número de publicaciones conjuntas; en redes de co-palabras, la frecuencia de co-ocurrencia terminológica; en redes de citación, el volumen de citas entre entidades. Esta ponderación permite distinguir entre conexiones ocasionales y relaciones sólidas y sostenidas, esencial para identificar colaboraciones estratégicas y núcleos temáticos consolidados.

La dirección de las aristas introduce una dimensión temporal y causal en el análisis, particularmente crucial en estudios de citación e influencia intelectual. En las redes de citación entre autores o revistas, las aristas dirigidas indican flujos de conocimiento, mostrando quién cita a quién y revelando patrones de influencia académica. Esta direccionalidad permite identificar trabajos seminales que reciben muchas citas, pero citan poco (nodos sumidero), versus trabajos de síntesis que citan extensamente y son ampliamente citados (nodos). El análisis de centralidad en redes dirigidas proporciona medidas más matizadas de influencia que consideran tanto el impacto recibido como la capacidad de difusión de conocimiento.

La interacción entre peso y dirección crea capas analíticas adicionales para la interpretación de grafos bibliométricos. Una arista pesada y bidireccional entre dos autores sugiere una colaboración intensa y recíproca, mientras una arista ligera y unidireccional puede indicar influencia incidental o reconocimiento puntual. En el análisis de co-citación, la fuerza y reciprocidad de las conexiones revelan proximidad intelectual y pertenencia a escuelas de pensamiento compartidas. La detección de comunidades en grafos ponderados y dirigidos identifica grupos con alta cohesión interna y patrones característicos de conexión externa, esencial para comprender la estructura social e intelectual de los campos científicos.

Los programas de visualización bibliométrica incorporan características distintivas que determinan la significación gráfica y capacidad interpretativa de los grafos generados. VOSviewer, por ejemplo, emplea algoritmos de disposición basados en atracción y repulsión que posicionan los nodos según su similitud, creando agrupaciones naturales donde la distancia visual representa proximidad temática o colaborativa. CiteSpace, en contraste, utiliza representaciones temporales que muestran la evolución de las redes mediante líneas de tiempo, donde la posición vertical indica tiempo de publicación y la horizontal refleja relaciones de citación. Estas diferencias algorítmicas producen visualizaciones con distinta capacidad para revelar patrones específicos, haciendo esencial comprender los principios de representación de cada herramienta.

La adición de clústeres o agrupamientos representa una funcionalidad avanzada que enriquece sustancialmente la interpretación de los grafos. Estos conglomerados se generan mediante algoritmos de detección de comunidades que identifican subgrupos de nodos con

alta conectividad interna y menor conexión externa, típicamente diferenciados por colores distintivos. En análisis de co-palabras, los clústeres revelan temas de investigación consolidados; en redes de coautoría, muestran grupos de colaboración; en redes de citación, identifican escuelas de pensamiento o paradigmas compartidos. La etiquetación automática de clústeres mediante términos representativos extraídos de los nodos constituye una funcionalidad clave que facilita la interpretación inmediata de la estructura subyacente del campo de estudio.

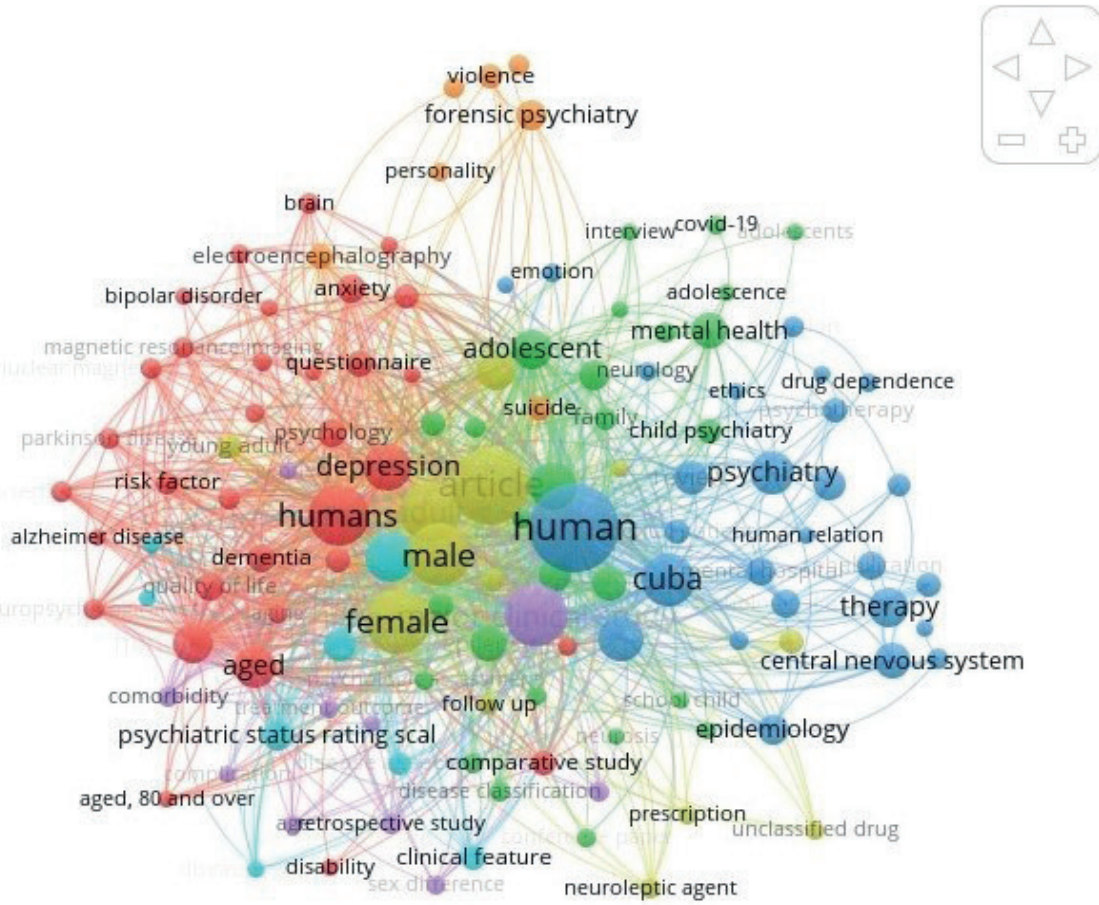


Figura 6.2. Ejemplo de grafo de co-ocurrencia de términos o palabras

Las herramientas más sofisticadas incorporan elementos adicionales como anillos concéntricos que representan dimensiones temporales, tamaños nodales proporcionales a métricas de impacto, y espesores de aristas que reflejan intensidad de relación. Algunos programas como CitNetExplorer añaden líneas de tiempo que muestran la evolución diacrónica de las conexiones, mientras SciMAT introduce mapas estratégicos que posicionan temas según su densidad y centralidad. Estas capas de información transforman los grafos de simples representaciones estructurales a herramientas analíticas multidimensionales que capturan simultáneamente relaciones, intensidad, evolución temporal e impacto, requiriendo del usuario comprensión de estas convenciones visuales para extraer perspectivas significativas de las visualizaciones generadas.

6.2. Algoritmos clave para construcción de grafos bibliométricos

Los algoritmos de construcción de grafos representan el núcleo metodológico del análisis de redes bibliométricas, transformando datos bibliográficos en estructuras relacionales significativas. Cada algoritmo opera bajo principios conceptuales distintos y responde a diferentes preguntas de investigación, requiriendo del analista una comprensión profunda de sus fundamentos teóricos y limitaciones prácticas. La selección adecuada del algoritmo determina no solo la validez de los resultados sino también la capacidad de extraer perspectivas significativas sobre la estructura y dinámica del conocimiento científico. Este capítulo examina en profundidad tres algoritmos fundamentales, co-ocurrencia, acoplamiento bibliográfico y co-citación, desentrañando sus mecanismos operativos, aplicaciones características y contextos de uso óptimos.

6.2.1. Análisis de co-ocurrencia: mapeando la estructura conceptual

El análisis de co-ocurrencia se fundamenta en el principio de que la aparición conjunta de términos en documentos académicos revela proximidad conceptual y temática. Este algoritmo construye redes donde los nodos representan conceptos, extraídos típicamente de palabras clave, títulos o resúmenes, y las aristas reflejan su co-ocurrencia en un mismo documento. La implementación técnica involucra múltiples etapas de procesamiento: inicialmente se normaliza el texto mediante lematización o *stemming* para unificar variantes morfológicas; posteriormente se filtran términos no significativos mediante listas de palabras vacías o *stopwords* personalizadas por dominio; finalmente se construye una matriz de co-ocurrencia documental donde cada celda registra la frecuencia de aparición conjunta.

La medición de la fuerza asociativa entre términos requiere aplicar coeficientes de normalización avanzados. El índice de Jaccard, calculado como el cociente entre la co-ocurrencia observada y la unión de frecuencias individuales, resulta particularmente efectivo para corregir el sesgo hacia términos muy frecuentes. Alternativamente, el coeficiente de similitud coseno mide el ángulo entre vectores de frecuencia en espacios multidimensionales, ofreciendo robustez en análisis de grandes volúmenes de datos. Para contextos donde la rareza terminológica contiene información valiosa, medidas como la asociación específica ponderada preservan conexiones entre términos especializados. La selección del coeficiente adecuado depende críticamente de la distribución terminológica del corpus analizado y de los objetivos específicos de investigación.

Las aplicaciones del análisis de co-ocurrencia abarcan desde la identificación de nichos de investigación emergentes hasta el mapeo de la estructura cognitiva de disciplinas consolidadas. En vigilancia tecnológica, permite detectar convergencias entre campos previamente disjuntos que indican áreas de innovación potencial. En estudios de interdisciplinariedad, revela puentes conceptuales entre dominios epistemológicos distintos. La evolución temporal de las redes de co-ocurrencia, obtenida mediante análisis diacrónicos segmentados por períodos, muestra la dinámica de formación, consolidación y disolución de temas de investigación, proporcionando perspectivas únicas sobre los procesos de cambio conceptual en la ciencia.

6.2.2. Acoplamiento bibliográfico: conexiones a través de referencias compartidas

El acoplamiento bibliográfico establece relaciones entre documentos basándose en su perfil compartido de citas, operando bajo el principio de que trabajos que citan fuentes comunes probablemente comparten marcos teóricos, metodológicos o conceptuales. A diferencia de otros algoritmos, el acoplamiento bibliográfico genera redes estáticas cuyas conexiones quedan fijadas en el momento de publicación, sin evolucionar con el tiempo. Esta característica lo hace especialmente valioso para analizar literatura reciente que aún no ha acumulado suficientes

citas para otros tipos de análisis relacional.

La implementación técnica requiere construir una matriz documentos-referencias donde cada fila representa un documento del corpus y cada columna una referencia citada. La matriz se transforma luego en una red de similitud mediante multiplicación matricial y aplicación de medidas de asociación. El índice de acoplamiento simple, basado en el conteo crudo de referencias compartidas, tiende a favorecer documentos con listas de referencias extensas, por lo que en la práctica se prefieren medidas normalizadas como el índice de Salton, que divide el número de referencias compartidas por la raíz cuadrada del producto de las referencias totales de cada documento. Para análisis especializados, el índice de proximidad bibliográfica pondera las referencias según su antigüedad, otorgando mayor importancia a citas recientes.

Las aplicaciones del acoplamiento bibliográfico son particularmente relevantes en contextos de evaluación de literatura contemporánea y mapeo de fronteras científicas activas. En revisiones sistemáticas de avances recientes, permite identificar grupos de publicaciones con enfoques afines sin esperar a que se establezcan tradiciones de citación. En estudios de interdisciplinariedad, revela documentos que funcionan como puentes entre campos distintos a través de sus patrones de citación mixtos. La principal limitación metodológica reside en su sensibilidad a estilos individuales de citación y su incapacidad para capturar influencias indirectas o evoluciones posteriores en las relaciones intelectuales.

6.2.3. Análisis de co-citación: la percepción colectiva de la comunidad científica

El análisis de co-citación se fundamenta en el principio de que la citación conjunta de dos documentos por trabajos posteriores refleja una relación percibida por la comunidad científica. A diferencia del acoplamiento bibliográfico, que examina citas desde los documentos fuente, la co-citación analiza citas desde los documentos que citan, capturando así una evaluación colectiva y dinámica de las relaciones intelectuales. Este algoritmo genera redes evolutivas donde la fuerza de conexión entre documentos puede aumentar, disminuir o reconfigurarse con el tiempo, reflejando cambios en los consensos científicos sobre la estructura del conocimiento.

La construcción de redes de co-citación implica fases técnicas complejas que incluyen la identificación de pares co-citados, el cálculo de frecuencias de co-citación, y la aplicación de umbrales mínimos para incluir conexiones significativas. La normalización de la fuerza de co-citación típicamente emplea el índice de co-citación estándar, simple conteo de citantes comunes, aunque implementaciones avanzadas utilizan el coeficiente de correlación de Pearson para capturar similitudes en patrones de co-citación a lo largo del tiempo. Para análisis diacrónicos, se segmentan los datos en ventanas temporales sucesivas, permitiendo rastrear la evolución de las agrupaciones intelectuales y la emergencia de nuevas líneas de investigación.

Las aplicaciones del análisis de co-citación son especialmente potentes en estudios históricos de la ciencia y análisis de la evolución de paradigmas intelectuales. Permite identificar trabajos seminales que han mantenido relevancia a lo largo del tiempo, detectar fusiones entre tradiciones intelectuales previamente separadas, y analizar procesos de especialización y fragmentación disciplinaria. En evaluación científica, proporciona indicadores robustos de impacto intelectual duradero más allá de las métricas de citación convencionales. Entre sus limitaciones se encuentra el retraso temporal inherente hasta que los documentos acumulan suficientes citas para análisis significativo y su posible reforzamiento de canonizaciones establecidas en detrimento de contribuciones marginales, pero potencialmente transformadoras.

Grafos de co-citación de autores

Los grafos de co-citación de autores representan las conexiones entre investigadores que son citados conjuntamente en las referencias de otros trabajos. Cuando dos autores aparecen frecuentemente citados en las mismas publicaciones, se infiere que sus contribuciones pertenecen a un marco teórico o temático común. La intensidad de la co-citación refleja su influencia compartida en un campo de estudio, permitiendo identificar escuelas de pensamiento, líderes intelectuales y la estructura invisible de las corrientes científicas. Un análisis temporal de estos grafos puede revelar la evolución de paradigmas y el surgimiento de nuevos enfoques.

Grafos de co-citación de Documentos

Estos grafos mapean las relaciones entre publicaciones que son citadas de manera simultánea por otros artículos. Cada nodo representa un documento, y las aristas simbolizan su co-ocurrencia en las secciones bibliográficas de trabajos posteriores. Esta estructura permite identificar los cimientos literarios de un campo: los artículos seminales que forman el núcleo teórico, así como los trabajos periféricos que conectan áreas distintas. La densidad de conexiones alrededor de un documento señala su centralidad en la construcción del conocimiento disciplinar.

Grafos de co-citación de Países e Instituciones

A nivel macro, los grafos de co-citación pueden aplicarse a países o instituciones, donde los nodos representan estas entidades y las aristas reflejan la frecuencia con que sus investigaciones son citadas conjuntamente. Esto indica una proximidad temática o metodológica entre sus sistemas científicos. Dos países con alta co-citación suelen especializarse en nichos de investigación similares o colaborar intensamente. Estos grafos son útiles para analizar la competitividad internacional, identificar clústeres de excelencia y diseñar estrategias de posicionamiento global en ciencia y tecnología.

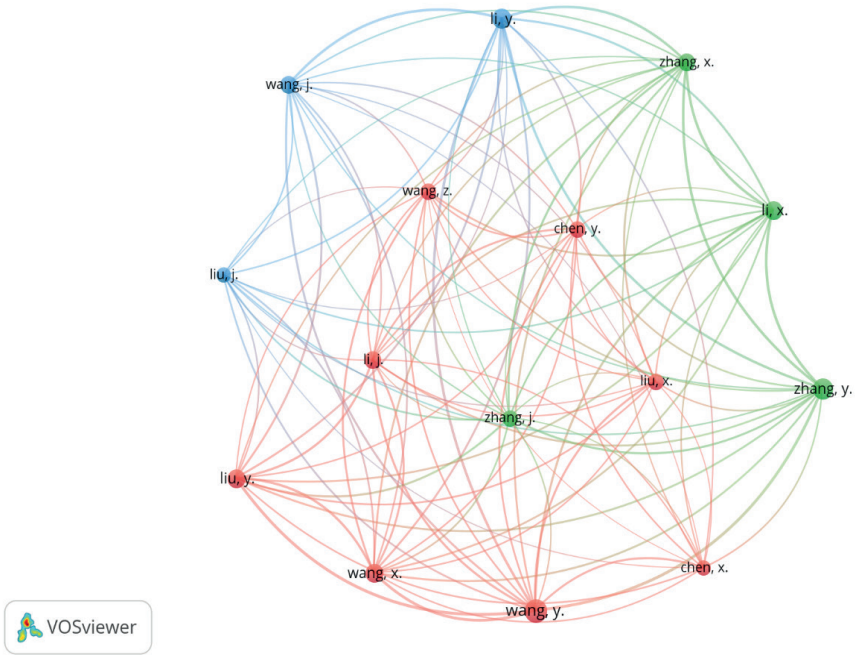


Figura 6.3. Ejemplo de grafo de co-citación

6.2.4. Integración metodológica y perspectivas comparativas

La selección entre estos algoritmos debe guiarse por consideraciones epistemológicas, temporales y pragmáticas específicas. El análisis de co-ocurrencia resulta óptimo para explorar la estructura conceptual inmediata de un campo, particularmente en disciplinas donde la terminología especializada codifica eficientemente el contenido intelectual. El acoplamiento bibliográfico ofrece ventajas decisivas para el análisis de literatura reciente y la identificación de alineamientos intelectuales contemporáneos. El análisis de co-citación, por su parte, proporciona profundidad histórica y captura evaluaciones colectivas consolidadas sobre relaciones intelectuales.

Las aproximaciones metodológicas más sofisticadas combinan múltiples algoritmos en diseños triangulados que mitigan las limitaciones individuales de cada método. Un diseño típico podría emplear co-ocurrencia para identificar temas emergentes, acoplamiento bibliográfico para analizar su estructuración actual, y co-citación para contextualizar su desarrollo histórico. Esta integración multi-método permite distinguir entre conexiones conceptuales inmediatas (co-ocurrencia), alineamientos intelectuales contemporáneos (acoplamiento bibliográfico) y percepciones consolidadas de relación (co-citación), produciendo análisis bibliométricos más ricos y metodológicamente robustos.

La comprensión profunda de estos algoritmos, sus fundamentos teóricos, implementaciones prácticas y sesgos característicos, constituye una competencia esencial para cualquier investigador que aspire a producir análisis bibliométricos rigurosos y significativos. Lejos de ser herramientas intercambiables, cada algoritmo ilumina dimensiones distintas de la compleja topografía del conocimiento científico, requiriendo selección consciente y aplicación contextualizada para revelar todo su potencial analítico.

6.3. Herramientas visuales

Las herramientas especializadas que se utilizan para construir grafos en el área de las bibliometrías son VOSviewer, CiteNetExplorer y CiteSpace, y R Bibliometrix y Pybibx aunque estas últimas no son especializadas en construcción de Grafos. El procedimiento en cada una es el siguiente:

6.3.1. VOSviewer: creación de mapas temáticos

1. Exportar datos bibliográficos completos desde Scopus, WoS o PubMed en formatos compatibles (RIS, CSV, EndNote)
2. Iniciar VOSviewer y seleccionar “Create” → “Create a map based on bibliographic data”
3. Elegir el tipo de análisis según el objetivo investigador:
    - Co-occurrence para análisis de términos y conceptos
    - Co-authorship para redes de colaboración científica
    - Co-citation para estructuras intelectuales basadas en citación
    - Bibliographic coupling para relaciones a través de referencias compartidas
4. Configurar parámetros de umbral según el volumen de datos:
    - Número mínimo de ocurrencias de un término (típicamente 5-10)
    - Número mínimo de documentos por autor (2-5 para análisis de colaboración)
    - Número mínimo de citas por referencia (10-20 para co-citación)
5. Aplicar el algoritmo VOS de agrupamiento y visualización
6. Personalizar la visualización final:
    - Ajustar tamaño de nodos según métricas de frecuencia o impacto
    - Modificar esquema de colores para diferenciar clústeres

- Configurar etiquetas y niveles de zoom para optimizar legibilidad
- Aplicar suavizado de redes para reducir superposición visual

La particularidad fundamental de VOSviewer reside en su algoritmo de visualización basado en técnicas de minimización de funciones de estrés multidimensional, donde la distancia entre nodos representa directamente su similitud calculada mediante medidas de asociación normalizadas. Esta herramienta genera mapas donde los clústeres emergen naturalmente como agrupaciones espaciales densas, diferenciadas cromáticamente para facilitar su identificación inmediata. La representación visual prioriza la interpretabilidad temática sobre la precisión métrica absoluta, haciendo de VOSviewer la opción preferida para análisis exploratorios y comunicación de resultados a audiencias no especializadas.

6.3.2. CitNetExplorer: creación de mapas de citación

1. Preparar datos de citación completos desde Web of Science en formato de exportación estándar
2. Cargar el archivo de datos mediante "File" → "Open Citation Network"
3. Configurar los parámetros de filtrado temporal y por impacto:
    - Definir el rango de años a analizar
    - Establecer número mínimo de citas para inclusión de documentos
    - Especificar criterios de selección de documentos seminales
4. Generar la red de citación completa con todas las interconexiones
5. Aplicar herramientas de exploración específicas:
    - Utilizar función de zoom para examinar períodos temporales específicos
    - Aplicar filtros dinámicos por número de citas o año de publicación
    - Emplear la herramienta de selección de componentes conectados
6. Realizar análisis de trayectorias citacionales:
    - Identificar documentos fundacionales y sus líneas de descendencia
    - Trazar caminos de desarrollo intelectual entre publicaciones
    - Analizar patrones de convergencia y divergencia temática

CitNetExplorer se distingue por su representación temporal explícita donde el eje vertical codifica inequívocamente el año de publicación, creando una línea de tiempo visual que revela la evolución diacrónica de las relaciones de citación. Esta herramienta especializada permite rastrear el desarrollo histórico de ideas científicas a través de las conexiones citacionales entre publicaciones, mostrando cómo los conceptos se transmiten, transforman y ramifican a lo largo del tiempo. La capacidad de animar la evolución de la red año por año proporciona perspectivas únicas sobre los patrones de herencia intelectual y los momentos de cambio paradigmático en el desarrollo científico.

6.3.3. CiteSpace: detección de "Bursts" de investigación

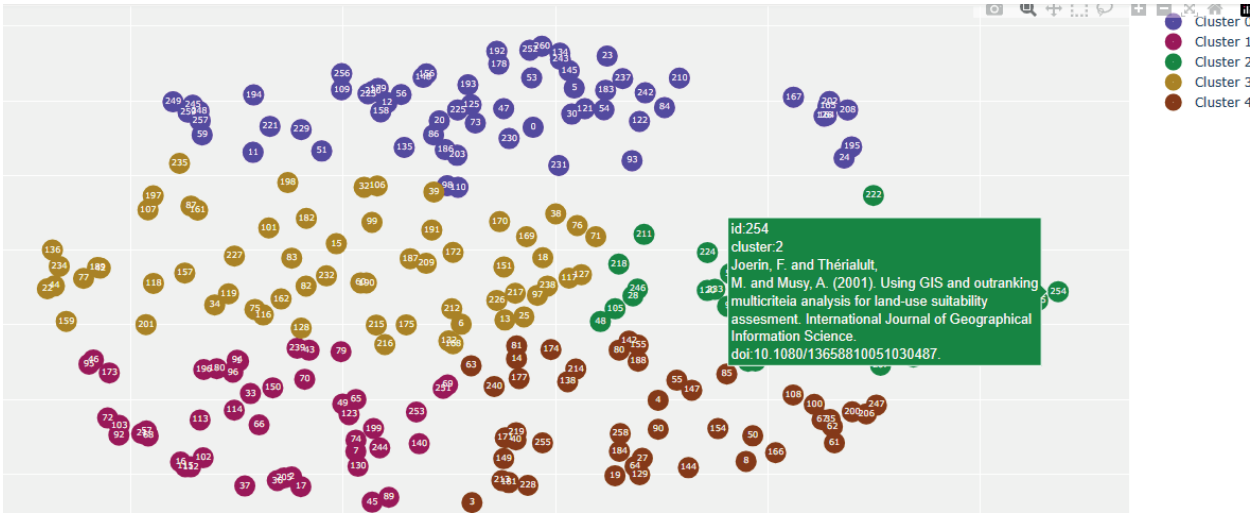
1. Descargar e instalar CiteSpace con Java Runtime Environment actualizado
2. Crear nuevo proyecto y configurar estructura de directorios para datos y resultados
3. Importar y convertir datos bibliográficos desde WoS o Scopus al formato nativo de CiteSpace
4. Configurar exhaustivamente los parámetros de análisis temporal:
    - Dividir el período de estudio en segmentos temporales (1-3 años recomendado)
    - Establecer criterios de selección por percentil (Top 10%, 20%, etc.) o por top N por segmento
    - Definir estrategias de poda de redes (Pathfinder, Minimum Spanning Tree, o ninguna)

5. Ejecutar el algoritmo de detección de bursts mediante:
    - Seleccionar entidades para análisis (términos, referencias, autores)
    - Configurar parámetros del algoritmo de Kleinberg para sensibilidad de detección
    - Especificar umbrales mínimos de duración e intensidad de bursts
6. Generar y analizar visualizaciones integradas:
    - Interpretar anillos concéntricos de citación anual en los nodos
    - Identificar clústeres temáticos mediante análisis de etiquetas automáticas
    - Analizar métricas de centralidad y conexiones para nodos puente

La singularidad de CiteSpace radica en su capacidad para integrar múltiples dimensiones analíticas en una única visualización: la posición espacial indica relaciones de similitud, los anillos concéntricos muestran patrones temporales de citación, los colores diferencian clústeres temáticos, y los anillos rojos destacan períodos de *burst* de actividad citacional. Esta herramienta emplea algoritmos especializados para la detección de puntos de inflexión en la literatura científica, identificando no solo temas emergentes sino también momentos específicos de aceleración en la atención investigadora. La representación visual resultante proporciona una cartografía dinámica de la evolución científica que captura simultáneamente estructura espacial, desarrollo temporal y puntos de inflexión intelectual.

6.3.4. Grafos con PyBibx

Tanto R Bibliometrix como Pybibx ofrecen funcionalidades para la construcción de grafos de coautoría, cada uno con sus particularidades algorítmicas. Pybibx se caracteriza por implementar un procedimiento de construcción más sencillo y con variantes flexibles. Una de estas representaciones genera un grafo donde los nodos, que representan artículos individuales referenciados por su ID única, interactúan entre sí sin la necesidad de que existan aristas explícitas que los conecten. Esta visualización alternativa permite analizar la proximidad o coexistencia de publicaciones dentro de un espacio conceptual o temporal definido por el investigador, ofreciendo una perspectiva diferente a la de una red tradicional de enlaces directos.



Fuente: Elaboración propia (Pybibx)

Figura 6.4. Ejemplo de grafo de co-ocurrencia de términos o palabras

Más allá de las representaciones sin aristas, estas herramientas también permiten generar grafos relacionales clásicos. En estos grafos, las conexiones o aristas entre los nodos sí están explícitamente definidas, representando vínculos de citación o co-citación entre los artículos individuales. A diferencia de los mapas que agrupan elementos en clústeres temáticos, esta visualización se centra en mostrar la red de conexiones directas y la estructura de la red en su estado más puro, sin aplicar algoritmos de agrupamiento. Esto permite al analista identificar patrones de conexión, artículos puente o la densidad de la red sin la influencia de una agrupación automática previa.

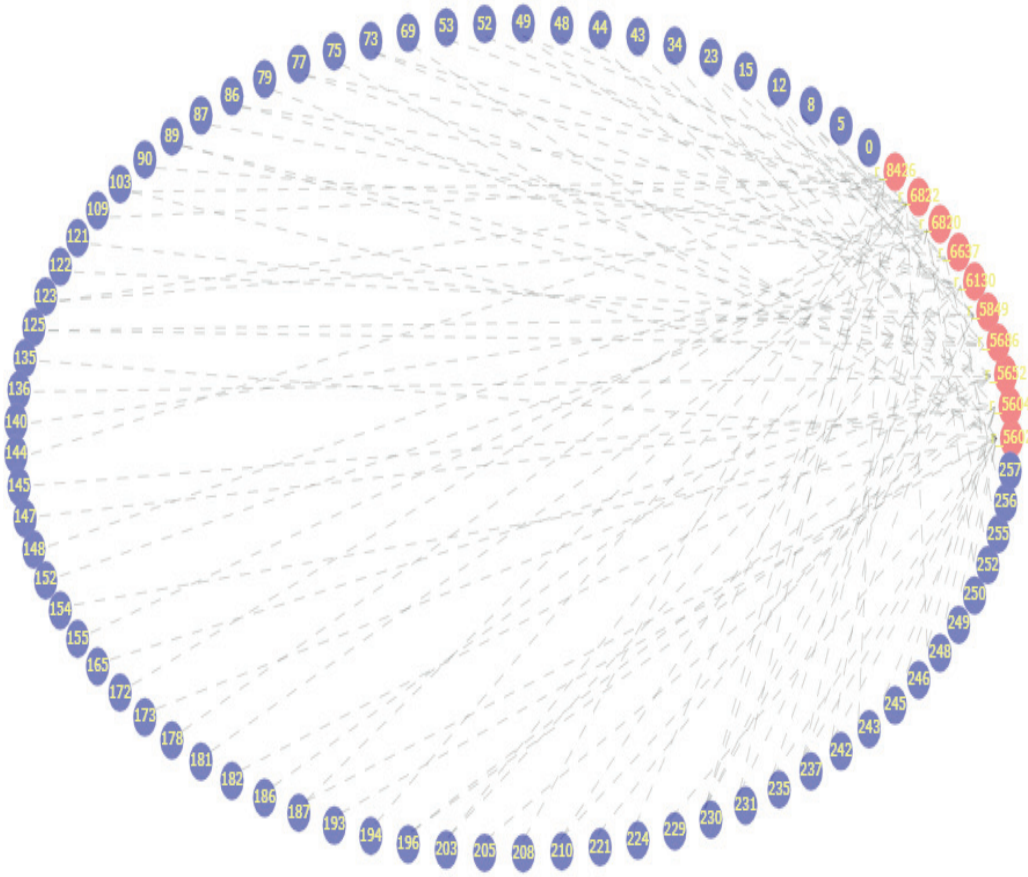


Figura 6.5. Ejemplo de grafo de co-ocurrencia de términos o palabras

Una de las visualizaciones más elocuentes para analizar la dimensión espacial de la investigación es el grafo superpuesto sobre un mapa geopolítico. En esta representación, los nodos se posicionan en las coordenadas geográficas de las afiliaciones de los autores, ilustrando la distribución global de la producción científica. Las aristas, que conectan estas instituciones, representan los vínculos de coautoría, permitiendo visualizar de forma inmediata los flujos y la intensidad de la colaboración internacional. Este grafo resulta indispensable para identificar centros de gravedad científica, patrones de colaboración regional y la proyección geográfica de las redes de investigación, ofreciendo una capa de análisis contextual sumamente valiosa.

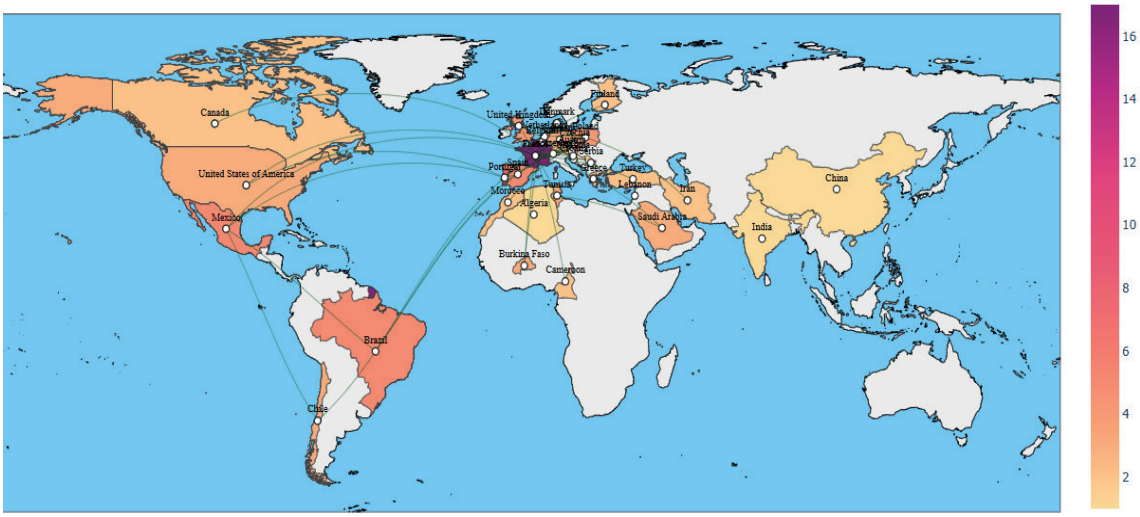


Figura 6.5. Ejemplo de grafo de co-citación con distribución geográfica de los autores

PyBibx ofrece otros tipos de grafos que pueden consultarse en su documentación oficial y en los ejemplos de la misma.

6.3.5. Grafos con R Bibliometrix

R Bibliometrix puede crear grafos de co-citación de autores, de acoplamiento bibliográfico entre documentos y de colaboración entre países e instituciones basados en relaciones de coautoría. Para crear grafos en R Bibliometrix se procede de la siguiente manera:

1. Instalar paquetes:

```
install.packages("bibliometrix")
install.packages("tidyverse")
```

2. Cargar librerías

```
library(bibliometrix)
library(tidyverse)
```

3. Cargar y procesar datos

```
file <- "archivo.bib"
M <- convert2df(file, dbsource = "wos", format = "bibtex")
```

4. *Crear red de co-ocurrencia*

```
NetMatrix <- biblioNetwork(M, analysis = "co-occurrence",
                           network = "author_keywords",
                           sep = ";")
```

5. Crear gráfico

```
net <- networkPlot(NetMatrix,
                   n = 50,  # Número de términos a mostrar
                   type = «fruchterman»,  # Layout
                   Title = «Co-ocurrencia de Términos - Author Keywords»,
                   labelsize = 0,8,
```

size = 5,
remove.isolates = TRUE,
clúster = «walktrap») # Método de *clustering*

Todos estos se visualizan en el panel de *plot* o pueden exportarse a imagen usando:
png("coocurrencia_terminos.png", width = 1200, height = 900, res = 150)
print(net)

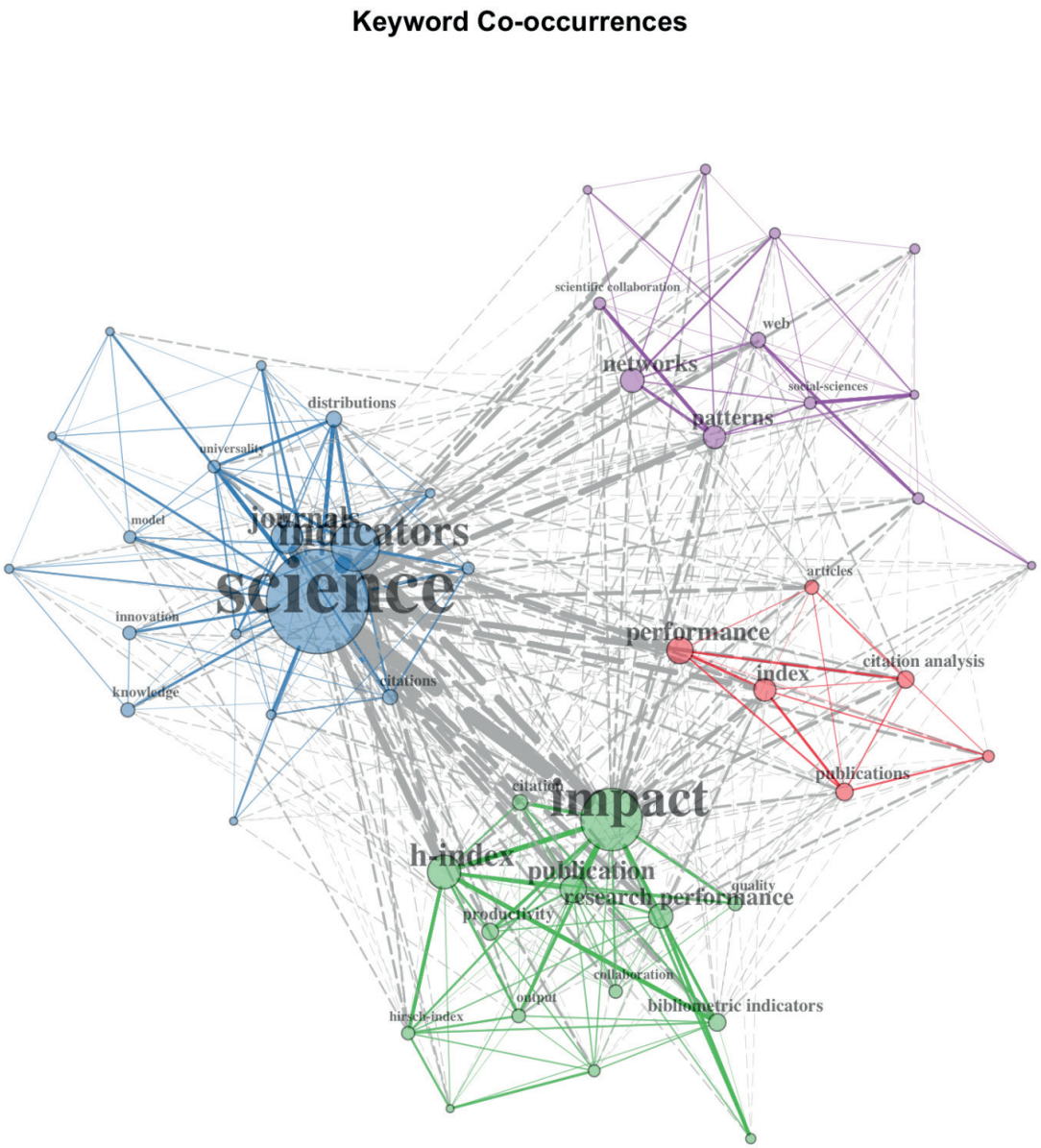


Figura 6.6. Ejemplo de grafo de co-ocurrencia

6.4. Interpretación de redes

6.4.1. Identificación de clústeres temáticos

La identificación de clústeres temáticos representa el proceso analítico mediante el cual se descubren y delimitan comunidades conceptuales dentro de una red bibliométrica. Estos conglomerados emergen naturalmente de los algoritmos de detección de comunidades que identifican subgrupos de nodos con alta conectividad interna y relativamente baja conexión externa. En la práctica, cada clúster encapsula un dominio temático coherente, donde los nodos, sean términos, autores o publicaciones, comparten características semánticas, metodológicas o teóricas significativas. La interpretación de estos clústeres requiere un análisis multidimensional que considere no solo la composición interna del grupo sino también sus relaciones con otros clústeres y su posición dentro de la estructura global de la red.

El proceso de interpretación comienza con el examen de las etiquetas representativas que los algoritmos asignan automáticamente a cada clúster, típicamente extraídas de los términos más frecuentes o centrales dentro del grupo. Sin embargo, esta aproximación automática debe complementarse con una evaluación cualitativa que examine publicaciones representativas de cada clúster para comprender su contenido sustantivo. La validación de la coherencia temática interna se realiza mediante el análisis de los documentos fundacionales, aquellos con mayor centralidad dentro del clúster, y la identificación de los conceptos nucleares que definen la identidad temática del grupo. Esta aproximación dual, cuantitativa y cualitativa, permite trascender la mera descripción estructural para alcanzar una comprensión profunda de los significados intelectuales encapsulados en cada agrupamiento.

La posición relativa de los clústeres dentro del mapa global proporciona información crucial sobre la estructura del campo de estudio. Clústeres próximos espacialmente típicamente comparten marcos conceptuales, metodologías o aplicaciones, mientras que clústeres distantes representan tradiciones intelectuales diferenciadas o especializaciones distantes. Las conexiones entre clústeres, manifestadas como enlaces puente, señalan interfaces temáticas potencialmente fértiles para investigación interdisciplinaria. El análisis diacrónico de la evolución de estos clústeres revela dinámicas de fragmentación disciplinaria, fusión de paradigmas o emergencia de nuevos campos híbridos, proporcionando perspectivas valiosas sobre la trayectoria de desarrollo del conocimiento en el dominio estudiado.

6.4.2. Centralidad vs. densidad

La centralidad y la densidad constituyen dos dimensiones analíticas fundamentales que capturan aspectos complementarios de la estructura y dinámica de las redes bibliométricas. La centralidad mide la posición estratégica de un nodo dentro de la red global, identificando elementos que funcionan como conectores entre diferentes regiones del grafo. Por contraste, la densidad cuantifica el grado de interconexión interna dentro de un clúster o subred específica, reflejando la cohesión y desarrollo maduro de una comunidad temática. La interpretación conjunta de estas métricas permite una caracterización sofisticada de la arquitectura intelectual del campo de estudio y de los roles específicos que diferentes elementos desempeñan dentro de dicho ecosistema cognitivo.

En la práctica analítica, la centralidad se manifiesta en múltiples formas, cada una revelando aspectos distintos de la influencia estructural. La centralidad de grado identifica nodos con muchas conexiones directas, típicamente conceptos fundamentales o autores muy colaborativos. La intermediación señala nodos que conectan diferentes clústeres, funcionando como puentes conceptuales entre comunidades temáticas separadas. La cercanía detecta nodos que pueden alcanzar rápidamente el resto de la red, indicando conceptos o autores con amplia

influencia difusa. Por ejemplo, en un mapa de co-palabras sobre inteligencia artificial, términos como "*machine learning*" mostrarían alta centralidad de grado, mientras "*explainable AI*" podría exhibir alta intermediación al conectar clústeres de ética computacional y algoritmos de aprendizaje.

La densidad, por su parte, caracteriza el desarrollo interno de los clústeres temáticos. Clústeres de alta densidad, con numerosas conexiones internas, representan campos maduros y altamente estructurados, donde los conceptos constitutivos han establecido relaciones bien definidas y estables. Clústeres de baja densidad sugieren áreas emergentes o en desarrollo, donde las relaciones conceptuales aún son incipientes y la estructura interna está en formación. En un análisis de autoría, un clúster denso indicaría un grupo de colaboración consolidado con múltiples proyectos conjuntos, mientras un clúster poco denso señalaría una red emergente de colaboración con interacciones más esporádicas o bilaterales.

La combinación de estas dimensiones en matrices estratégicas, como el análisis de centralidad versus densidad, proporciona un marco poderoso para la caracterización tipológica de clústeres y nodos. Por ejemplo, los temas motores aparecen como clústeres con alta densidad y alta centralidad, representando áreas consolidadas que impulsan el desarrollo del campo. Los temas de nicho muestran alta densidad, pero baja centralidad, indicando especializaciones maduras pero aisladas. Los temas emergentes exhiben baja densidad, pero alta centralidad, señalando conceptos puente con potencial de desarrollo futuro. Los temas periféricos presentan baja densidad y baja centralidad, representando especializaciones incipientes o áreas en declive. Esta tipología permite priorizar estratégicamente áreas de investigación según objetivos específicos de desarrollo científico o política investigadora.

En esta fase de interpretación, el dominio técnico de la bibliometría resulta necesario, pero no suficiente para extraer significado sustantivo de las redes generadas. El investigador debe poseer conocimientos profundos sobre el dominio específico de estudio para establecer conexiones significativas entre términos aparentemente dispares y reconocer relaciones conceptuales que trascienden las meras coincidencias estadísticas. Esta comprensión disciplinar especializada permite discernir entre asociaciones terminológicas superficiales y vínculos intelectuales profundos, entre modas terminológicas pasajeras y conceptos fundacionales con verdadera capacidad estructurante. La pericia temática se convierte así en el lente indispensable a través del cual los patrones cuantitativos adquieren significado cualitativo y relevancia intelectual.

La capacidad de abstracción emerge como competencia crítica para trascender la visualización inmediata y construir modelos mentales que expliquen la arquitectura subyacente del conocimiento representado. Esta habilidad permite integrar hallazgos fragmentarios en narrativas coherentes sobre la estructura y dinámica del campo de estudio, identificando no solo lo que las redes muestran explícitamente sino también lo que sugieren implícitamente a través de sus vacíos, asimetrías y patrones relacionales. El investigador debe moverse constantemente entre el nivel micro de términos individuales y conexiones específicas, el nivel meso de clústeres temáticos y sus interrelaciones, y el nivel macro de la estructura global del campo, sintetizando perspectivas desde múltiples escalas de análisis en una comprensión integrada y jerárquicamente organizada.

La interpretación final, por tanto, representa una síntesis creativa donde el rigor métrico se combina con la sensibilidad disciplinar y la capacidad inferencial del investigador. Este proceso transforma datos relacionales en conocimiento accionable, identificando oportunidades de investigación, revelando interfaces fértiles entre especialidades, y proponiendo narrativas

explicativas sobre la evolución y estado actual del campo. La calidad de esta interpretación depende fundamentalmente de la capacidad del investigador para ejercer juicio informado, contextualizar hallazgos dentro de tradiciones intelectuales más amplias, y comunicar perspectivas complejas de manera accesible sin sacrificar profundidad analítica o precisión conceptual.

Un ejemplo de grafo y su interpretación son los siguientes:

Contexto: Resultado de búsqueda en SCOPUS de: *(TITLE-ABS-KEY (cardiolog* OR cardiac OR heart OR coronary OR "myocardial infarction" OR arrhythmia OR echocardiography OR hypertension)) AND (AFFILCOUNTRY (cuba))*

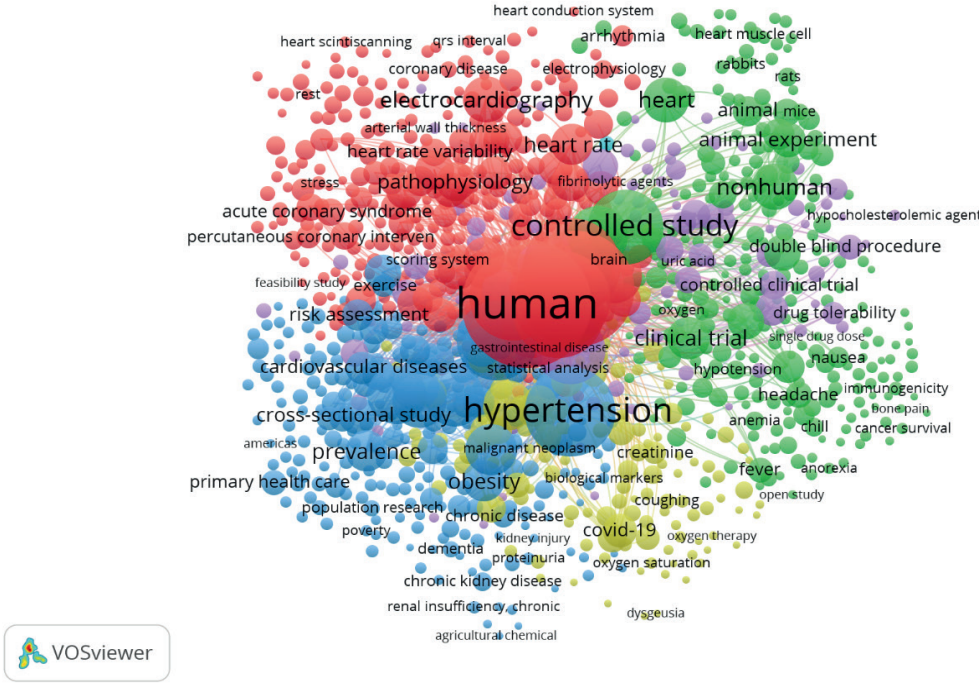


Figura 6.6. Ejemplo de grafo de co-ocurrencia

Posible interpretación

Clúster rojo (parte superior izquierda): Este clúster se centra en la cardiología clínica y la electrofisiología. Incluye términos como "*heart*," "*electrocardiography*," "*arrhythmia*," "*heart rate*," "*coronary disease*," y "*acute coronary syndrome*." Esto muestra un enfoque en la investigación de enfermedades coronarias, arritmias y el uso de técnicas de diagnóstico como el electrocardiograma. Las conexiones con "*human*" y "*controlled study*" demuestran que estos estudios se realizan en pacientes.

Clúster azul (parte inferior izquierda): Este clúster se enfoca en la epidemiología y los factores de riesgo de las enfermedades cardiovasculares. Términos clave son "*hypertension*," "*cardiovascular diseases*," "*prevalence*," "*cross-sectional study*," "*primary health care*," y "*obesity*." Esto indica un fuerte interés en la salud pública, la prevalencia de enfermedades y su relación con otros factores de riesgo en la población.

Clúster verde (parte superior derecha): Este clúster está relacionado con la investigación experimental y farmacológica. Los términos "*nonhuman*," "*animal experiment*," "*rats*," "*rabbits*," "*double blind procedure*," "*drug tolerability*," y "*drug efficacy*" son prominentes. Esto sugiere una línea de investigación que utiliza modelos animales y ensayos clínicos controlados para probar la eficacia y la seguridad de nuevos tratamientos.

Clúster amarillo (centro): Este es el núcleo de la investigación. Aunque no es tan grande, contiene los términos más centrales como "*human*", "*controlled study*", "*clinical trial*", y "*statistical analysis*". Este clúster actúa como el conector entre todos los demás, lo que confirma que la investigación en cardiología en Cuba se basa principalmente en estudios clínicos y controlados en seres humanos.

En resumen, los gráficos revelan que la investigación en cardiología en Cuba es multifacética, con un fuerte énfasis en la investigación clínica de enfermedades cardiovasculares y la hipertensión. Se abordan temas como la epidemiología, la electrofisiología, y la evaluación de tratamientos, además de una línea de investigación experimental en modelos animales.

Recapitulando

- Los grafos de correlación son representaciones visuales de las relaciones entre variables o elementos bibliométricos (autores, palabras clave, instituciones, revistas, etc.).
- Se basan en los principios de la teoría de redes y permiten analizar la estructura y dinámica del conocimiento científico.
- Un grafo está compuesto por nodos (entidades) y aristas (enlaces) que reflejan relaciones o asociaciones entre dichos elementos.
- En bibliometría, los nodos pueden representar autores, artículos, palabras clave o países, y las aristas su grado de correlación, coocurrencia o co-citación.
- El análisis de grafos permite identificar patrones de colaboración, afinidad temática y estructuras cognitivas en la producción científica.
- La fuerza de correlación entre dos nodos se cuantifica mediante medidas estadísticas como el coeficiente de correlación de Pearson o Spearman.
- Existen diversos tipos de redes:
    - Redes de coautoría (colaboración entre investigadores).
    - Redes de co-citación (artículos citados conjuntamente).
    - Redes de co-palabras o coocurrencia (términos que aparecen juntos).
- Los grafos se pueden construir a partir de matrices de correlación o similitud, derivadas de los datos bibliográficos.
- La densidad de una red indica el grado general de conexión entre los nodos.
- La centralidad (*degree, betweenness, closeness*) mide la importancia relativa de un nodo dentro de la red.
- Los clústeres o comunidades representan grupos de nodos altamente interconectados, que suelen corresponder a temas o subcampos científicos.
- El análisis de modularidad permite identificar esas comunidades mediante algoritmos de agrupamiento.
- Las herramientas más utilizadas para generar grafos de correlación son VOSviewer, Gephi, CiteSpace, BibExcel y Pajek.
- VOSviewer construye mapas de similitud y calcula distancias entre elementos según su grado de coocurrencia.
- Gephi ofrece funciones avanzadas de visualización interactiva y análisis estructural de redes grandes.

- El color y tamaño de los nodos suelen representar la intensidad de correlación y el peso de las conexiones.
- Los grafos permiten observar la evolución temporal de los temas y la emergencia de nuevas áreas de investigación.
- Una interpretación adecuada requiere combinar el análisis cuantitativo (estadístico) y cualitativo (semántico) de los resultados.
- Los grafos de correlación contribuyen a la vigilancia científica, la identificación de líderes de opinión y la detección de vacíos temáticos.
- Su correcta aplicación exige rigurosidad metodológica, selección cuidadosa de los datos y visualizaciones éticas y comprensibles.

Preguntas de autoevaluación

1. ¿Qué representan los nodos y las aristas en un grafo de correlación?
2. ¿Qué tipo de información se puede analizar mediante grafos en estudios bibliométricos?
3. ¿Qué diferencia existe entre una red de coautoría y una de co-citación?
4. ¿Qué mide el coeficiente de correlación en la construcción de un grafo?
5. ¿Qué indica la densidad de una red bibliométrica?
6. ¿Qué tipos de centralidad existen y qué representa cada una?
7. ¿Qué significa la existencia de un clúster en un grafo de correlación?
8. ¿Qué herramientas permiten construir grafos de correlación y visualizar redes científicas?
9. ¿Qué información transmite el color y tamaño de los nodos en un grafo?
10. ¿Por qué es importante combinar análisis cuantitativo y cualitativo al interpretar redes bibliométricas?

BIBLIOGRAFÍA

1. Newman MEJ. Networks: an introduction. Oxford: Oxford University Press; 2010. ISBN 9780199206650

2. Barabási AL. Network science. Cambridge: Cambridge University Press; 2016. ISBN 9781107076266

3. 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

4. Chen C. Mapping scientific frontiers: the quest for knowledge visualization. 2nd ed. London: Springer; 2017. https://doi.org/10.1007/978-3-319-32043-4

5. Leydesdorff L, Welbers K. The semantic mapping of science: co-words, co-authors, and co-citations. J Informetrics. 2011;5(1):1–14. https://doi.org/10.1016/j.joi.2010.10.002

6. Hanneman RA, Riddle M. Introduction to social network methods. Riverside (CA): University of California; 2005. https://faculty.ucr.edu/~hanneman/nettext/