

Chapter 8 / Capítulo 8

Applied bibliometrics. From data to publication (Spanish Edition)
ISBN: 978-9915-9680-6-3
DOI: 10.62486/978-9915-9680-6-3.ch08

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

SANKEY DIAGRAMS: FLOWS AND RELATIONSHIPS BETWEEN CONCEPTS / MAPAS DE SANKEY: FLUJOS Y RELACIONES ENTRE CONCEPTOS

Los mapas de Sankey representan una herramienta visual distintiva que captura flujos de intensidad variable entre diferentes categorías o estados en un sistema bibliométrico. A diferencia de los grafos tradicionales que muestran relaciones de adyacencia o conectividad, los diagramas de Sankey visualizan transferencias cuantificadas de recursos, atención investigadora o influencia intelectual mediante flechas de ancho proporcional a la magnitud del flujo. Esta representación resulta particularmente poderosa para analizar trayectorias temporales, transferencias interdisciplinarias y redistribuciones de enfoque investigador a lo largo del tiempo.

8.1. Interpretación

La interpretación adecuada de estos mapas requiere comprender que el ancho de las conexiones codifica valores cuantitativos específicos, mientras la disposición espacial revela la estructura de las relaciones entre los elementos del sistema.

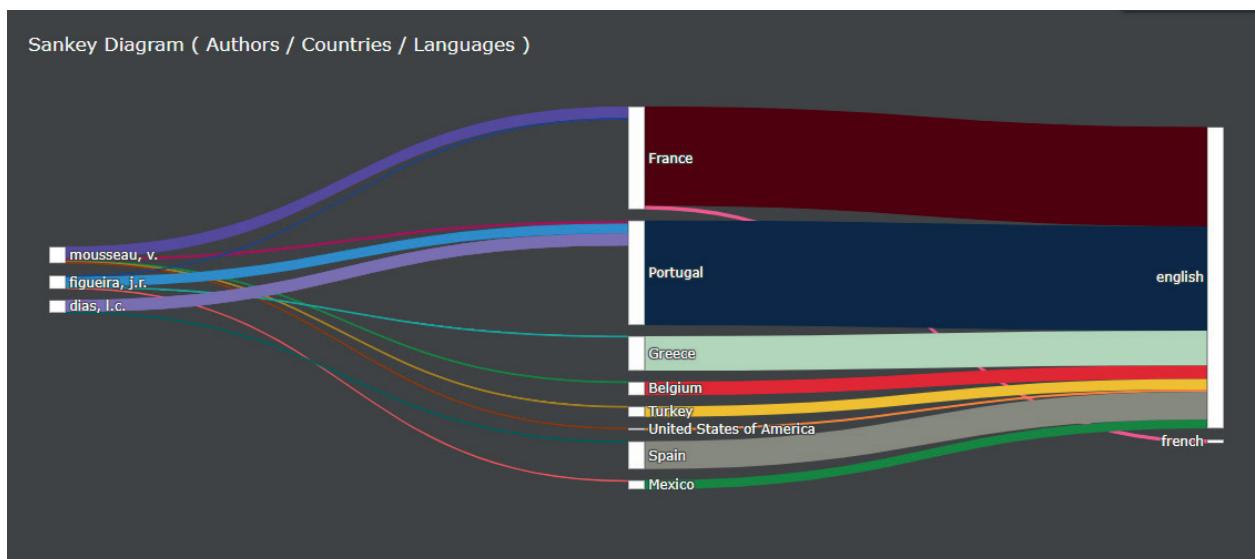


Figura 8.1. Ejemplo de diagrama de Sankey

La lectura de un diagrama de Sankey bibliométrico implica descomponer la visualización en tres componentes estructurales principales: los nodos que representan categorías discretas (campos disciplinares, instituciones, países), los flujos que conectan estos nodos indicando transferencias o relaciones, y la dirección que establece la orientación temporal o causal de dichas transferencias. La interpretación comienza identificando los flujos dominantes, aquellas conexiones más anchas que representan las relaciones más intensas, para comprender las principales vías de intercambio intelectual o colaboración institucional. Posteriormente, el análisis se desplaza hacia los flujos secundarios que, aunque menos intensos, pueden revelar conexiones emergentes o nichos especializados con potencial de desarrollo futuro. Finalmente, la identificación de nodos aislados o con escasas conexiones señala áreas desconectadas del mainstream investigador u oportunidades para establecer nuevas colaboraciones.

La dimensión temporal en los mapas de Sankey añade profundidad analítica al permitir

rastrear evoluciones diacrónicas en los patrones de flujo. Cuando el diagrama representa períodos sucesivos, los cambios en el grosor de las conexiones entre diferentes segmentos temporales revelan dinámicas de consolidación, diversificación o reorientación de líneas de investigación. El fortalecimiento progresivo de ciertos flujos indica la emergencia de nuevas especialidades o la intensificación de colaboraciones establecidas, mientras el adelgazamiento o desaparición de otros sugiere el declive de paradigmas o la ruptura de alianzas investigadoras. Esta capacidad para visualizar trayectorias hace de los mapas de Sankey una herramienta excepcional para el análisis de la movilidad científica, la evolución de intereses investigadores y la reconfiguración de fronteras disciplinarias.

La interpretación contextualizada trasciende el análisis cuantitativo inmediato para integrar factores explicativos que dan significado a los patrones observados. Un flujo intenso entre dos campos disciplinarios puede reflejar tanto la fructífera integración de marcos conceptuales como la mera apropiación instrumental de metodologías sin verdadera síntesis intelectual. La concentración de flujos en torno a ciertos nodos puede indicar liderazgo intelectual genuino o simplemente efectos de escala derivados del tamaño de comunidades investigadoras. El analista experto distingue estas posibilidades mediante triangulación con evidencia cualitativa y conocimiento especializado del dominio, transformando el diagrama de Sankey de representación descriptiva en herramienta diagnóstica para comprender las dinámicas de cambio en el paisaje científico.

Las aplicaciones bibliométricas específicas de los mapas de Sankey abarcan desde el análisis de movilidad académica, trazando trayectorias de investigadores entre instituciones o países, hasta el estudio de la evolución temática, mostrando cómo los conceptos migran entre diferentes campos disciplinarios. En evaluación de políticas científicas, permiten visualizar flujos de financiamiento entre agencias, áreas temáticas y equipos investigadores. En estudios de interdisciplinariedad, revelan patrones de citación entre campos distantes o la integración de metodologías diversas. La versatilidad de esta representación la convierte en un instrumento particularmente valioso para comunicar hallazgos complejos a audiencias no especializadas, traduciendo dinámicas investigadoras abstractas en narrativas visuales intuitivas sobre los flujos que dan forma al conocimiento contemporáneo.

8.2. Tipos de diagramas Sankey

Los mapas de Sankey de evolución temática visualizan cómo los conceptos y temas de investigación migran entre diferentes campos disciplinarios o cómo se transforman a lo largo del tiempo. Estos diagramas muestran la emergencia, consolidación y declive de líneas de investigación, revelando procesos de especialización, fusión de paradigmas o surgimiento de áreas interdisciplinarias. Su construcción se basa en el análisis diacrónico de términos extraídos de títulos, resúmenes y palabras clave, segmentados en períodos temporales sucesivos donde se identifican las transiciones terminológicas más significativas mediante medidas de similitud semántica y análisis de co-ocurrencia evolutiva.

Los diagramas de flujo de citación representan transferencias de influencia intelectual entre diferentes campos, instituciones o grupos de investigación. A diferencia de las redes de citación convencionales que muestran conexiones puntuales, los Sankey de citación capturan flujos netos de influencia, mostrando cómo ciertas áreas actúan como exportadoras netas de conocimiento mientras otras funcionan principalmente como importadoras. Este enfoque resulta particularmente valioso para estudiar relaciones interdisciplinarias, donde los flujos de citación pueden revelar asimetrías en el intercambio intelectual entre campos consolidados y emergentes, o entre diferentes tradiciones epistemológicas.

Los mapas de Sankey de colaboración científica visualizan patrones de cooperación entre instituciones, países o disciplinas, mostrando no solo la existencia de colaboraciones sino también su intensidad y evolución temporal. Estos diagramas permiten identificar consorcios estables de investigación, patrones emergentes de cooperación sur-sur o norte-sur, y la dinámica de formación y disolución de redes colaborativas. Su construcción típicamente utiliza datos de coautoría segmentados por períodos, donde los flujos representan el volumen de publicaciones conjuntas entre diferentes entidades, permitiendo analizar cómo las estrategias de colaboración institucional evolucionan en respuesta a programas de financiamiento, políticas científicas o desarrollos tecnológicos.

Los diagramas de flujo de financiamiento representan la distribución de recursos económicos entre diferentes áreas temáticas, instituciones o tipos de investigación, mostrando cómo los fondos se asignan, transfieren y concentran en el ecosistema científico. Estos mapas son particularmente útiles para evaluar el impacto de políticas de financiamiento, identificar desajustes entre prioridades declaradas y patrones reales de asignación, y analizar la eficiencia en la distribución de recursos para investigación. Su construcción requiere integrar datos de convocatorias, proyectos concedidos y outputs científicos, creando una trazabilidad completa desde las fuentes de financiamiento hasta los productos de investigación generados.

Los Sankey de producción científica por sectores visualizan la distribución de outputs investigadores entre diferentes tipos de organizaciones (universidades, centros de investigación, empresas, hospitalares) y su evolución temporal. Estos diagramas capturan tendencias en la especialización sectorial de la investigación, procesos de transferencia de conocimiento desde academia hacia industria, y cambios en el rol relativo de diferentes actores en el sistema de innovación. Su implementación requiere clasificar consistentemente las afiliaciones institucionales por sector y analizar su evolución temporal, revelando patrones de diversificación o especialización en la producción científica de diferentes tipos de organizaciones.

Cada tipo de diagrama Sankey bibliométrico responde a preguntas de investigación específicas y requiere estrategias particulares de preparación de datos, procesamiento y validación. La selección del tipo apropiado debe guiarse por los objetivos analíticos, la disponibilidad de datos con la granularidad y cobertura temporal necesarias, y las necesidades comunicativas del estudio. Independientemente del tipo específico, todos comparten la capacidad de transformar datos complejos de relaciones bibliométricas en narrativas visuales intuitivas sobre los flujos que estructuran la dinámica del conocimiento científico contemporáneo.

8.3. Construcción

8.3.1. Implementación en R Bibliometrix

La construcción de mapas de Sankey en Bibliometrix se realiza mediante la función *sankeyPlot()*, que transforma matrices de transición entre categorías bibliométricas en diagramas de flujo interactivos. El proceso inicia con la preparación de una matriz de transición donde las filas representan el estado origen y las columnas el estado destino, con los valores celulares indicando la frecuencia o intensidad del flujo. Para análisis temáticos temporales, esta matriz captura la migración de conceptos entre períodos consecutivos, mientras para estudios de colaboración puede representar movilidad institucional o cambios en patrones de coautoría.

La implementación técnica requiere primero procesar los datos bibliográficos mediante *biblioAnalysis()* para obtener la estructura básica, luego utilizar *coupling()* o *cocMatrix()* para generar matrices de asociación específicas. La función *sankeyPlot()* acepta parámetros de personalización como *node.width* para ajustar el ancho de los nodos, *node.pad* para el

espaciado entre elementos, y **units** para especificar la métrica representada en los flujos. Bibliometrix integra la librería *networkD3* para renderizar visualizaciones interactivas donde los usuarios pueden arrastrar nodos, hacer *hover* sobre flujos para ver valores exactos, y filtrar conexiones por umbrales de intensidad.

Para análisis diacrónicos avanzados, Bibliometrix permite segmentar los datos en períodos temporales mediante **timeslice** y generar matrices de transición sucesivas que muestran la evolución de flujos. La función **themeEvolution()** complementa este análisis identificando trayectorias temáticas específicas que luego pueden visualizarse como mapas de Sankey especializados. La exportación de resultados incluye formatos HTML para preservar la interactividad, así como imágenes vectoriales SVG para publicaciones académicas, manteniendo la claridad visual en diferentes escalas y resoluciones.

El código para generar un mapa Sankey es:

1. Instalar y cargar paquetes
`install.packages("bibliometrix")
install.packages("plotly")
install.packages("networkD3")`

`library(bibliometrix)
library(plotly)
library(networkD3)
library(dplyr)`

2. Cargar datos
`data(management, package = "bibliometrix")
M <- management`

3. Crear datos para Sankey de evolución temporal
`years <- c(2015, 2020) # Rango de años
sankey_data <- thematicEvolution(M),
years = years,
n = 10, # Número de términos por año
field = "ID") # "ID" para Keywords Plus, "DE" para Author Keywords`

4. Generar Sankey plot
`sankey_plot <- plotThematicEvolution(sankey_data$Nodes, sankey_data$Edges)
sankey_plot`

8.3.2 Implementación en PyBibX

PyBibX aborda la construcción de mapas de Sankey mediante la clase *SankeyDiagram*, que ofrece un control granular sobre todos los aspectos de la visualización. La implementación básica requiere crear una instancia del diagrama especificando los datos de flujo en formato de lista de tuplas o diccionario anidado, donde cada entrada define un flujo individual con su origen, destino y magnitud. El método **generate()** procesa estos datos y aplica algoritmos de posicionamiento automático que minimizan el cruce de conexiones y optimizan la legibilidad general del diagrama.

La personalización avanzada en PyBibX incluye ajustes estéticos mediante parámetros como **color_palette** para definir esquemas cromáticos consistentes con el dominio de estudio, **node_**

alignment para controlar la disposición vertical de los nodos, y **flow_opacity** para manejar la superposición visual de flujos múltiples. La integración nativa con Plotly permite crear visualizaciones interactivas donde los usuarios pueden explorar flujos específicos, reorganizar nodos dinámicamente, y aplicar filtros en tiempo real según diferentes criterios de magnitud o categoría.

Para análisis bibliométricos especializados, PyBibX incluye métodos específicos como **author_mobility_sankey()** que transforma datos de afiliación institucional en mapas de movilidad académica, y **concept_evolution_sankey()** que visualiza la migración terminológica entre períodos de investigación. La herramienta también ofrece funciones de preprocesamiento como **normalize_flows()** para ajustar magnitudes según diferentes métricas de impacto, y **cluster_flows()** para agrupar conexiones menores y simplificar visualizaciones complejas sin perder información sustantiva.

Recapitulando

- Los diagramas de Sankey muestran flujos dirigidos entre categorías donde el ancho de la banda es proporcional a la magnitud del flujo.
- Permiten visualizar transferencias temáticas, citacionales o de financiación.
- Cada enlace necesita datos origen-destino y una magnitud.
- La direccionalidad del flujo debe ser clara mediante flechas o disposición.
- La normalización evita saturación visual y mejora comparabilidad.
- Escalas logarítmicas ayudan cuando los flujos son heterogéneos.
- Limpieza y unificación de categorías son esenciales antes del análisis.
- Sankeys jerárquicos muestran subcategorías dentro de flujos mayores.
- El orden de los nodos afecta la legibilidad; se deben minimizar cruces.
- Los colores deben distinguir grupos y contar con leyenda explicativa.
- Los Sankey deben incluir unidades y sumas verificables.
- Software popular: R (ggalluvial), Python (plotly), D3.js.
- La interactividad mejora la exploración de flujos complejos.
- En series temporales, se usan paneles por período o animaciones.
- Verificar outliers y errores de agregación aumenta la validez.
- Agrupar categorías menores en “otros” facilita claridad visual.
- Complementan grafos y mapas de calor mostrando rutas dinámicas.
- Documentar criterios de agregación garantiza reproducibilidad.
- Los Sankey deben acompañarse de tablas numéricas.
- Interpretar los flujos en su contexto disciplinar evita conclusiones erróneas.

Preguntas de autoevaluación

1. ¿Qué representa el ancho de una banda en un Sankey?
2. ¿Qué datos mínimos requiere un Sankey?
3. ¿Cómo mostrarías evolución temporal mediante Sankey?
4. ¿Por qué se normalizan las categorías antes de construirlo?
5. ¿Qué ventajas tiene sobre un grafo estático?
6. ¿Cuándo usarías escala logarítmica?
7. ¿Qué efectos produce un orden de nodos incorrecto?
8. ¿Cómo reducir sobrecarga visual en Sankey complejos?
9. ¿Qué herramientas pueden emplearse?
10. ¿Por qué es vital documentar criterios de agregación?

BIBLIOGRAFÍA

1. Tufte ER. *The visual display of quantitative information*. 2nd ed. Cheshire (CT): Graphics Press; 2001. ISBN 9780961392147
2. Healy K. *Data visualization: a practical introduction*. Princeton (NJ): Princeton University Press; 2018. ISBN 9780691181622
3. Murray S. *Interactive data visualization for the web: an introduction to designing with D3*. 2nd ed. Sebastopol (CA): O'Reilly Media; 2017. ISBN 9781491921289
4. Knaflic CN. *Storytelling with data: a data visualization guide for business professionals*. Hoboken (NJ): Wiley; 2015. <https://doi.org/10.1002/9781119055259>