

Applied bibliometrics. From data to publication

Spanish Edition

AG
EDITOR

Applied bibliometrics. From data to publication (Spanish Edition)
First Edition

Annier Jesús Fajardo Quesada

Eduardo Antonio Hernández González

René Herrero Pacheco

Lisannia Virgen Beritán Yero

Copyright Page

© 2025. The authors. This is an open access book, distributed under the terms of a Creative Commons Attribution 4.0 license (<https://creativecommons.org/licenses/by/4.0>) that allows use, distribution, and reproduction in any medium provided that the original work is properly cited.

This AG Editor imprint is published by AG Editor.

The registered company is **AG Editor SAS, Montevideo, Uruguay.**

For more information, see AG Editor's Open Access Policy: <https://www.ageditor.org/editorial-policies.php>

ISBN (Spanish edition, eBook): 978-9915-9680-6-3

ISBN (English edition, eBook): 978-9915-9680-7-0

This title is available in both print (softcover) and digital (open access PDF) formats.

DOI: 10.62486/ 978-9915-9680-6-3

Publisher: Javier González Argote

Chief Executive Officer: editorial@ageditor.org

Editorial Director: Emanuel Maldonado

Editorial Coordinators: William Castillo González; Karina Maldonado

Production Manager: Adrian Vitón Castillo

Legal Deposit: National Library of Uruguay — Law No. 13.835/1970 and Decree No. 694/1971

ISBN Record: National ISBN Agency (Uruguay) — Filing No. 58241

Cataloging Data

Cataloging-in-Publication Data (CIP):

AG Editor

Applied bibliometrics. From data to publication (Spanish Edition) / Annier Jesús Fajardo Quesada, Eduardo Antonio Hernández González, René Herrero Pacheco, Lisannia Virgen Beritán Yero. – Montevideo: AG Editor, 2025.

First Edition.

ISBN (Spanish edition, eBook): 978-9915-9680-6-3

ISBN (English edition, eBook): 978-9915-9680-7-0

THEMA classification codes:

GLF — Information technology, Internet and electronic resources in libraries

Editorial Notice and Acknowledgments

The publication of this book is part of AG Editor's mission to promote open, ethical, and rigorous scientific communication across all areas of knowledge.

All books published by AG Editor undergo a double-blind peer-review process and a technical evaluation in accordance with the publisher's editorial policies, aligned with COPE and ICMJE standards.

AG Editor acknowledges the valuable collaboration of authors, reviewers, designers, and production teams who made this publication possible.

“La verdad es la verdad, dígala Agamenón o su porquero”

El Porquero de Agamenón

La Bibliometría como Piloto en la Tormenta de la Ciencia Abierta.

Vivimos en la era de la explosión informativa. Cada 6 segundos, se publica un nuevo artículo científico en el mundo. Solo en 2024, Scopus registró más de 5 millones de estudios indexados. Ante este diluvio de conocimiento, investigadores, instituciones y gobiernos enfrentan un desafío crítico: ¿cómo navegar este océano sin naufragar en la irrelevancia o la redundancia?

La bibliometría emerge no como lujo académico, sino como herramienta de supervivencia intelectual. Esta disciplina, que analiza cuantitativamente la producción y diseminación del conocimiento, se ha transformado en el radar esencial para:

- Priorizar recursos en universidades con presupuestos menguantes
- Detectar fraudes científicos en campos como la medicina

Pero hoy su papel es aún más vital. La ciencia abierta, con su mandato de transparencia, acceso libre y reproducibilidad, ha convertido a la bibliometría en guardiana de la integridad académica. Cuando el Manifiesto de Leiden (2015) y la Declaración DORA (2012) exigen evaluar investigaciones por su impacto real, no por el factor de impacto de sus revistas, son las métricas responsables las que permiten cumplir esta promesa.

Este libro nace para empoderarte en ese cambio de paradigma. No como teórico distante, sino como guía práctica que transforma datos en decisiones. Aquí no encontrarás fórmulas abstractas, sino protocolos probados para:

- Identificar líneas de investigación emergentes antes que competidores
- Demostrar el impacto social de tu trabajo más allá de las citas
- Evitar los sesgos que distorsionan rankings universitarios

Esta obra fue cincelada para dos perfiles de exploradores del conocimiento:

1. Investigadores

Para ti que escribes artículos en la madrugada:

¿Cómo demostrar que tu trabajo sobre nanotecnología aplicada a cultivos merece financiamiento?

¿De qué modo identificar colaboradores clave en tu país para tu proyecto?

Aquí aprenderás a:

- Usar VOSviewer para mapear redes de coautoría
- Calcular tu índice-h corregido por disciplina
- Documentar tu impacto con métricas alternativas (ej. políticas públicas derivadas)

2. Estudiantes

Para ti que inicia tu viaje académico:

¿Cómo elegir un supervisor con trayectoria real (no solo fama)?

¿Qué revistas priorizar en tu primera publicación?

Encontrarás:

- Tutoriales en Google Colab para análisis sin instalar software
- Instrucciones paso a paso con Bibliometrix
- Alertas sobre depredadores bibliométricos (journals que inflan métricas)

Del Machete al Satélite

Este libro es una expedición metodológica diseñada en cinco etapas progresivas

Parte I: Fundamentos de Navegación

Capítulo 1 te equipa con el GPS conceptual: diferencias entre bibliometría, cienciometría y altimetría.

Capítulo 2 es tu brújula histórica: desde los índices de citas de Garfield (1955) hasta la revolución de los metadatos abiertos.

Parte II: Preparando la Travesía

Capítulo 3 son tu caja de herramientas digitales: instalación de Python/R incluso en computadores lentos.

Capítulo 4 es tu kit de supervivencia: obtención ética de datos en Scopus, WoS y PubMed.

Parte III: Explorando Territorios Complejos

Capítulo 5 despliega tu mapa topográfico: cálculo crítico de 15 índices bibliométricos (con códigos replicables).

Capítulo 6, 7 y 8 es tu dron cartográfico: creación e interpretación de grafos con VOSviewer y CiteSpace.

Capítulo 9 integra el resto de indicadores para potenciar tus resultados

Parte IV: Comunicando Hallazgos

Capítulo 10 y 11 es tu bitácora de viaje: cómo redactar métodos, resultados y discusiones que superen revisiones rigurosas.

Parte V: Futuro y recursos y aplicaciones

Capítulo 12 y 13 proyecta tu telescopio hacia IA generativa y web.

Apéndices son tu maletín de emergencia: glosario y sitios de interés.

El Compromiso: Rigor sin Dogma

Este manual rechaza dos extremos peligrosos: el fetichismo métrico que reduce ciencia a números y el desprecio anticuantitativo que juzga “deshumanizador” medir impacto.

En su lugar, te ofrecemos un enfoque tridimensional:

- Técnico (códigos en Python/R actualizados a 2025)
- Crítico (alertas sobre sesgos de género, idioma y geografía)
- Ético (alineado con la ciencia abierta y evaluación responsable)

Al final de este viaje, no serás solo un usuario pasivo de métricas. Serás un navegante capaz de:

- Crear estudios bibliométricos válidos
- Interpretar gráficos complejos con ojo clínico
- Debatir políticas científicas con evidencia dura

La ciencia del siglo XXI exige más que productores de datos: requiere intérpretes de significados. Esta es tu brújula para no perder el rumbo.

Annier Jesús Fajardo Quesada
La Habana, octubre de 2025.

PART I: INTRODUCTION AND FOUNDATIONS / PARTE I: INTRODUCCIÓN Y FUNDAMENTOS	8
CHAPTER 1 / CAPÍTULO 1	9
1. INTRODUCTION TO BIBLIOMETRICS / INTRODUCCIÓN A LA BIBLIOMETRÍA	10
1.1. Definition and Objectives / Definición y objetivos	10
1.1.1. Objectives: Beyond Counting Citations / Objetivos: Más allá de contar citas	11
1.1.2. Ethical Limits: What Bibliometrics Should NOT Be / Límites éticos: lo que la bibliometría NO debe ser	13
1.2. Classification of Bibliometrics / Clasificación de las bibliometrías	14
1.3. Key Differences: Bibliometrics vs. Scientometrics vs. Altmetrics / Diferencias clave: bibliometría vs. cienciometría vs. Altmetría	15
1.3.1. Bibliometrics: The Analysis of What Is Published / Bibliometría: el análisis de lo publicado	15
1.3.2. Scientometrics: The Science of Science / Cienciometría: la ciencia de la ciencia	16
1.3.3. Altmetrics: Impact in the Digital Society / Altmetría: el impacto en la sociedad digital	17
1.4. Practical Applications / Aplicaciones prácticas	20
1.4.1. Research Policies / Políticas de investigación	21
1.4.2. Bibliometrics as a Social Thermometer / Bibliometría como termómetro social	22
1.5. Ethics and Good Practices (DORA Declaration, Leiden Manifesto) / Ética y buenas prácticas (Declaración DORA, Manifiesto de Leiden)	23
1.5.1. Ethical Risks: When Numbers Reinforce Inequalities / Riesgos éticos: cuando los números refuerzan desigualdades	23
CHAPTER 2 / CAPÍTULO 2	31
2. BRIEF HISTORICAL OVERVIEW / BREVE RECUENTO HISTÓRICO	
2.1. Evolution: From Garfield (SCI 1960) to the Digital Era / Evolución: de Garfield (SCI 1960) a la era digital	32
2.2. The Digital Revolution (1990-2010): Expansion and Criticism / La revolución digital (1990-2010): expansión y críticas	35
2.3. The Modern Era (2010-Present): Complexity and Democratization / La era moderna (2010-presente): complejidad y democratización	36
2.4. Historical Lessons / Lecciones históricas	37
PART II: DATA PREPARATION / PARTE II: PREPARACIÓN DE DATOS	40
CHAPTER 3 / CAPÍTULO 3	41
3. TOOLS FOR BIBLIOMETRIC ANALYSIS / HERRAMIENTAS PARA EL ANÁLISIS BIBLIOMÉTRICO	42
3.1. Introduction to Bibliometric Tools / Introducción a las herramientas bibliométricas	42
3.1.1. Classification of Tools and Selection Criteria / Clasificación de herramientas y criterios de selección	42
3.2. Setting Up the R Environment / Preparación del entorno de trabajo para R	44
3.2.1. Installing and Configuring R and RStudio / Instalación y configuración de R y RStudio	44
3.2.2. Specialized Tools in R / Herramientas especializadas en R	44
3.2.3. Main Features and Basic Analysis / Funcionalidades principales y análisis básico	45
3.3. Working with Google Colab for PyBibX / Trabajo con Google Colab para PyBibX	45
3.4. Publish or Perish: Analysis with Google Scholar / Publish or Perish: análisis con Google Scholar	47
3.4.1. Installation, Configuration, and Search Types / Instalación, configuración y tipos de búsqueda	47
3.4.2. Data Extraction and Provided Metrics / Extracción de datos y métricas proporcionadas	49

3.4.3. Export and Integration with Other Tools / Exportación e integración con otras herramientas	50
3.4.4. Limitations and Best Practices / Limitaciones y buenas prácticas de uso.....	52
3.5. Advanced Visualization Tools / Herramientas de visualización avanzada	52
3.5.1. VOSviewer: Installation and Scientific Map Creation / VOSviewer: instalación y creación de mapas científicos.....	53
3.5.2. CitNetExplorer: Citation Network Analysis / CitNetExplorer: análisis de redes de citación.....	54
3.5.3. CiteSpace: Temporal Pattern Analysis / CiteSpace: análisis de patrones temporales	55
3.5.4. Online Tools (Litmaps, ResearchRabbit) / Herramientas online (Litmaps, ResearchRabbit)	57
3.6. Tool Selection / Selección de herramientas	59
CHAPTER 4 / CAPÍTULO 4.....	63
4. DATA COLLECTION / OBTENCIÓN DE DATOS	64
4.1. Search Strategies for Bibliometric Studies / Estrategias de búsqueda para estudios bibliométricos	64
4.1.1. Search strategy design / Diseño de la estrategia de búsqueda	64
4.1.2. Construction of search chains / Construcción de cadenas de búsqueda	65
4.1.3. Specific techniques by database / Técnicas específicas por base de datos	66
4.1.4. Validation and documentation: ensuring rigor in bibliometric research / Validación y documentación: garantizando rigor en la búsqueda bibliométrica	68
4.1.5. Common errors and solutions / Errores comunes y soluciones.....	66
4.2. Alternative Tools for Bibliometric Search: Broadening Horizons / Herramientas alternativas para búsqueda bibliométrica: ampliando horizontes.....	70
4.2.1. Google Scholar: Practical Advantages and Limitations / Google Scholar: ventajas y limitaciones prácticas.....	70
4.2.2. Harzing's Publish or Perish: A Comprehensive Bibliometric Tool / Harzing's Publish or Perish: una herramienta integral para el análisis bibliométrico	71
4.2.3. Reference Managers for Advanced Bibliometric Analysis / Gestores de referencias para análisis bibliométrico avanzado.....	73
4.3. Export Formats in Bibliometrics: Interoperability and Analysis / Formatos de exportación en bibliometría: interoperabilidad y análisis	73
4.3.1. Format Conversion: Tools and Uses in Bibliometric Software / Conversión entre formatos: herramientas y usos en software bibliométrico	73
PART III: ADVANCED ANALYSIS / PARTE III: ANÁLISIS AVANZADO	77
CHAPTER 5 / CAPÍTULO 5.....	78
5. BIBLIOMETRIC INDICATORS / ÍNDICES BIBLIOMÉTRICOS	79
5.1. Classification / Clasificación	79
5.2. Temporal Evolution of Bibliometric Indicators / Evolución temporal de los indicadores bibliométricos.....	82
5.3. Combining Bibliometric Indicators with Contextual Variables / Combinación de indicadores bibliométricos con variables contextuales.....	84
5.4. Practical Calculation / Cálculo práctico	85
5.4.1. With R/Bibliometrix (biblioAnalysis()) / Con R/Bibliometrix (biblioAnalysis())	85
5.4.2. Alternatives: Publish or Perish, Excel (Formulas) / Alternativas: Publish or Perish, Excel (fórmulas)	86
5.5. Critical Interpretation / Interpretación crítica	87

CHAPTER 6 / CAPÍTULO 6	91
6. CORRELATION GRAPHS / GRAFOS DE CORRELACIÓN	92
6.1. Theoretical Foundations of Bibliometric Graphs / Fundamentos teóricos de los grafos bibliométricos	92
6.1.1. Nodes: Authors, Keywords, and Journals as Units of Analysis / Nodos: autores, palabras clave y revistas como unidades de análisis	92
6.1.2. Edges: Weight and Direction as Relationship Measures / Aristas: peso y dirección como medidas de relación	93
6.2. Key Algorithms for Bibliometric Graph Construction / Algoritmos clave para construcción de grafos bibliométricos	95
6.2.1. Co-occurrence Analysis: Mapping Conceptual Structure / Análisis de co-ocurrencia: mapeando la estructura conceptual	95
6.2.2. Bibliographic Coupling: Connections via Shared References / Acoplamiento bibliográfico: conexiones a través de referencias compartidas	95
6.2.3. Co-citation Analysis: Collective Perception of the Scientific Community / Análisis de co-citación: la percepción colectiva de la comunidad científica	96
6.2.4. Methodological Integration and Comparative Perspectives / Integración metodológica y perspectivas comparativas	98
6.3. Visual Tools / Herramientas visuales	98
6.3.1. VOSviewer: Thematic Map Creation / VOSviewer: creación de mapas temáticos	98
6.3.2. CiteNetExplorer: Citation Map Creation / CiteNetExplorer: creación de mapas de citación	99
6.3.3. CiteSpace: Detecting Research “Bursts” / CiteSpace: detección de “bursts” de investigación	99
6.3.4. Graphs with PyBibX / Grafos con PyBibX	100
6.3.5. Graphs with R Bibliometrix / Grafos con R Bibliometrix	102
6.4. Network Interpretation / Interpretación de redes	104
6.4.1. Identification of Thematic Clusters / Identificación de clústeres temáticos	104
6.4.2. Centrality vs. Density (Practical Examples) / Centralidad vs. densidad (ejemplos prácticos)	104
CHAPTER 7 / CAPÍTULO 7	
7. HEAT MAPS: REPRESENTING RESEARCH DENSITY / MAPAS DE CALOR: REPRESENTANDO LA DENSIDAD DE LA INVESTIGACIÓN	110
7.1. Interpretation of Heat Maps in Bibliometric Analysis / Interpretación de mapas de calor en análisis bibliométrico	110
7.2. Algorithm for Building Bibliometric Heat Maps / Algoritmo de construcción de mapas de calor bibliométricos	111
7.3. Creation of Bibliometric Heat Maps / Creación de mapas de calor bibliométricos	112
CHAPTER 8 / CAPÍTULO 8	115
8. SANKEY MAPS: FLOWS AND RELATIONSHIPS BETWEEN CONCEPTS / MAPAS DE SANKEY: FLUJOS Y RELACIONES ENTRE CONCEPTOS	116
8.1. Interpretation / Interpretación	116
8.2. Types of Sankey Diagrams / Tipos de diagramas Sankey	117
8.3. Construction / Construcción	118
8.3.1. Implementation in R Bibliometrix / Implementación en R Bibliometrix	118
8.3.2. Implementation in PyBibX / Implementación en PyBibX	119
CHAPTER 9 / CAPÍTULO 9	122
9. OTHER RESULTS / OTROS RESULTADOS	123
9.1. Interpretation of Language in Scientific Production / Interpretación del idioma en la producción	123

científica	123
9.2. Interpretation of the Role of Scientific Journals / Interpretación del rol de las revistas científicas	123
9.3. Interpretation of Country Distribution / Interpretación de la distribución por países.....	124
9.4. Interpretation of Gender in Scientific Authorship / Interpretación del género en la autoría científica	125
9.5. Interpretation of Document Types / Interpretación de la tipología documental	125
9.6. Interpretation of Keyword Analysis / Interpretación del análisis de palabras clave	126
9.7. Analysis of Institutional Collaboration Networks / Análisis de redes de colaboración institucional	126
9.8. Funding Patterns in Scientific Production / Patrones de financiamiento en la producción científica	126
9.9. Combinations / Combinaciones.....	127
9.10. Visualization Strategies for Different Bibliometric Dimensions / Estrategias de visualización para diferentes dimensiones bibliométricas.....	128

PART V: PUBLISHING AND COMMUNICATION / PARTE V: PUBLICAR Y COMUNICAR 131

CHAPTER 10 / CAPÍTULO 10 132

10. ACADEMIC WRITING / REDACCIÓN ACADÉMICA..... 133

10.1. Concept of Bibliometrics / Idea de bibliometría	133
10.2. Writing the Methods / Redacción de los métodos	134
10.2.1. PRISMA Diagram for Transparency / Diagrama PRISMA para transparencia	134
10.2.2. Ethical Protocols / Protocolos éticos	134
10.3. Writing the Results / Redacción de los resultados	135
10.3.1. Selecting Visualizations for the Audience / Selección de visualizaciones por audiencia	136
10.3.2. Dynamic Tables for Comparisons / Tablas dinámicas para comparativas	137
10.4. Writing the Discussion / Redacción de las discusiones.....	137
10.4.1. Linking with Science Policy / Vinculación con políticas científicas	138
10.4.2. Methodological Limitations / Limitaciones metodológicas.....	138
10.4.3. Integration of Multidisciplinary Perspectives / Integración de perspectivas multidisciplinares.....	139
10.5. Best Practices in Data Visualization / Buenas prácticas en visualización de datos	140
10.6. Other Sections in Bibliometric Studies / Otras secciones en estudios bibliométricos	141

CHAPTER 11 / CAPÍTULO 11 143

11. PUBLISHING AND COMMUNICATING THE ARTICLE / PUBLICAR Y COMUNICAR EL ARTÍCULO 144


11.1. Selecting the Journal / Selección de la revista	144
11.2. Adapting to the Journal's Guidelines / Adaptar a la norma de la revista seleccionada	146
11.3. Presenting the Information / Presentación de la información.....	147
11.3.1. Slide Presentations / Presentación en diapositivas.....	147
11.3.2. Reproducible Reports with RMarkdown and Google Colab / Informes reproducibles con RMarkdown y Google Colab	148
11.3.3. Interactive Dashboards with Shiny, Dash, and Streamlit / Paneles interactivos con Shiny, Dash y Streamlit	149
11.3.4. Scientific Posters for Conferences / Pósteres científicos para congresos	150
11.3.5. Executive Summaries for Decision-Makers / Resúmenes ejecutivos para tomadores de decisiones	150

PART VI: APPLIED BIBLIOMETRICS / PARTE VI: BIBLIOMETRÍA APLICADA ...	153
CHAPTER 12 / CAPÍTULO 12	154
12. EMERGING TRENDS / TENDENCIAS EMERGENTES	155
12.1. Generative AI: GPT in Scientific Text Mining / IA generativa: GPT en minería de textos científicos.....	155
12.2. Blockchain for Decentralized Metrics / Blockchain para métricas descentralizadas	156
12.3. Image and Multimodal Data Analysis / Análisis de imágenes y datos multimodales.....	158
12.4. Ethics and Transparency in the AI Era / Ética y transparencia en la era de la IA.....	160
CHAPTER 13 / CAPÍTULO 13	163
13. ACADEMIA, GOVERNMENT, AND INDUSTRY / ACADEMIA, GOBIERNO E INDUSTRIA..	164
13.1. Sectoral Differences in Objectives and Indicators / Diferencias en objetivos e indicadores por sector.....	164
13.2. Evidence-Based National Science Policies / Políticas científicas nacionales basadas en evidencia.....	165
13.3. Institutional Ranking and Evaluation / Ranking y evaluación institucional.....	167
13.4. University-Industry Knowledge Transfer / Transferencia de conocimiento universidad-empresa	168
13.5. Bibliometrics for Assessing Social Impact / Bibliometría para la evaluación de impacto social.....	169
13.6. Intellectual Property Management Based on Metrics / Gestión de propiedad intelectual basada en métricas.....	170
APPENDICES / APÉNDICES.....	I
I. Technical Glossary / Glosario técnico.....	I
II. Additional Terms / Términos adicionales	III
III. Resources and Useful Links / Recursos y enlaces de interés	III

Part I / Parte I

INTRODUCTION AND FOUNDATIONS

INTRODUCCIÓN Y FUNDAMENTOS



Chapter 1 / Capítulo 1

Applied bibliometrics. From data to publication (Spanish Edition)

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch01

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

INTRODUCTION TO BIBLIOMETRICS / INTRODUCCIÓN A LA BIBLIOMETRÍA

1.1 Definición y Objetivos de la Bibliometría

La Ciencia que Mide el Conocimiento

La bibliometría constituye una disciplina científica que aplica métodos cuantitativos para analizar sistemáticamente la producción, difusión e impacto del conocimiento académico. El término, que deriva de las voces griegas *'biblion'* (libro) y *'metron'* (medida), ha evolucionado para abarcar el estudio de diversos productos académicos contemporáneos, desde artículos científicos y patentes hasta tesis doctorales y conjuntos de datos. Esta aproximación metodológica representa una alternativa robusta a las revisiones bibliográficas tradicionales, destacándose por tres atributos fundamentales que garantizan su rigor científico.⁽¹⁾

La objetividad metodológica se sustenta en el análisis de evidencias empíricas concretas, como patrones de citación, redes de colaboración e indicadores de impacto normalizados. Esta característica permite superar las limitaciones de las evaluaciones puramente cualitativas, basándose en datos verificables y métricas estandarizadas. La escalabilidad operativa posibilita procesar volúmenes masivos de literatura científica mediante herramientas computacionales especializadas, analizando miles de documentos en tiempos reducidos. Finalmente, la reproducibilidad analítica se garantiza mediante protocolos estandarizados, como las directrices PRISMA para la selección sistemática de estudios, que permiten la verificación independiente de los resultados.^(1,2)

1.1.1 Objetivos: Más Allá de Contar Citas

La bibliometría no se limita a medir citas; es una disciplina analítica que busca entender, evaluar y optimizar la producción científica. Sus cinco objetivos fundamentales permiten a investigadores, instituciones y políticas científicas tomar decisiones basadas en evidencia.⁽³⁾

Evaluar el impacto científico

El objetivo primordial de evaluar el impacto científico ha evolucionado significativamente, trascendiendo la visión reduccionista que lo equiparaba de forma exclusiva al recuento de citas bibliográficas. Si bien este indicador tradicional permanece como un elemento esencial, al ofrecer una prueba tangible de la asimilación de una investigación por parte de la comunidad especializada y su contribución al avance del conocimiento disciplinar, la concepción contemporánea del impacto es notablemente más rica y matizada.

La bibliometría moderna ha comprendido que el impacto científico genuino no es un fenómeno unidimensional, sino una onda expansiva que se manifiesta de múltiples maneras. Por ello, su evaluación ahora integra un espectro más amplio de métricas que capturan la influencia de una investigación a lo largo de diversos ejes. Por un lado, las altmétricas actúan como un termómetro de la utilidad práctica y la transferencia del conocimiento más allá de la torre de marfil académica. Un artículo que fundamenta una guía clínica o una regulación ambiental posee un impacto científico profundo, aunque su número de citas pueda ser moderado.

Por otro lado, herramientas como Dimensions han perfeccionado este análisis al permitir una integración sinérgica de las citas tradicionales con datos procedentes de patentes, proyectos de financiación y conjuntos de datos. Esta integración no solo enriquece la evaluación, sino que permite trazar el ciclo de vida de la investigación: desde la idea original, pasando por su desarrollo (financiado por proyectos), hasta su protección (patentes) y, finalmente, su diseminación e influencia (citas y uso en nuevas investigaciones). De este modo, se puede discernir si un trabajo, más que ser simplemente citado, ha servido como piedra angular para

un nuevo campo de desarrollo tecnológico o una línea de investigación fructífera.

En definitiva, el objetivo actual de evaluar el impacto científico ya no se satisface con contar ecos en la literatura, sino con comprender la naturaleza y el alcance de la contribución. Se trata de diferenciar entre la mera visibilidad y una influencia sustantiva que acelera el descubrimiento, informa la aplicación y, en última instancia, consolida el legado de una investigación en el edificio del conocimiento humano.

Mapear estructuras de conocimiento

Más allá de medir el impacto, la bibliometría posee una capacidad única para cartografiar el vasto y dinámico territorio del conocimiento científico. La producción académica no es un conjunto aislado de *papers*, sino un ecosistema orgánico en constante evolución, donde ideas, disciplinas e investigadores se interconectan formando estructuras complejas. El objetivo aquí es revelar estas arquitecturas invisibles, identificando las corrientes intelectuales que ganan fuerza, las que se agotan y los espacios vacíos donde la innovación puede florecer.

Técnicas como el análisis de co-palabras, que identifica los conceptos que frecuentemente aparecen juntos, o el acoplamiento bibliográfico, que relaciona documentos que comparten referencias comunes, actúan como potentes lentes para observar este paisaje. Estas metodologías no se limitan a listar temas; detectan patrones, revelan la anatomía de las disciplinas y muestran cómo surgen, convergen o se fragmentan los campos de estudio.

La verdadera potencia de este análisis se materializa en herramientas de visualización como VOSviewer o CiteSpace, que se verán en capítulos posteriores. Estos programas transforman los datos masivos de publicaciones en mapas visuales intuitivos, donde los clústeres temáticos se agrupan en nubes de colores y la proximidad entre nodos indica la fuerza de su relación. Un vistazo a estos mapas puede desvelar tendencias emergentes antes de que se consoliden, señalar colaboraciones interdisciplinarias fructíferas o, por el contrario, poner de manifiesto áreas de investigación saturadas donde es difícil hacer una contribución original.

Este mapeo no es un ejercicio meramente descriptivo; es una herramienta estratégica de primer orden. Para el investigador individual, funciona como una brújula que señala “nichos de oportunidad” o gaps en la literatura, guiando así el planteamiento de preguntas de investigación verdaderamente novedosas. Para las instituciones y agencias de financiación, estos mapas ofrecen una base objetiva para diseñar políticas científicas, asignar recursos de manera inteligente y fomentar colaboraciones que fortalezcan ecosistemas de investigación con mayor potencial.

Optimizar recursos académicos

En un contexto de recursos finitos y competencia creciente, la gestión de la ciencia no puede depender únicamente de intuiciones o inercias históricas. La bibliometría emerge aquí como una herramienta fundamental para la administración estratégica, aportando el rigor de los datos a decisiones críticas sobre financiación, contratación y planificación institucional. Su objetivo es garantizar que la inversión en investigación, ya sea pública o privada, se dirija hacia donde pueda generar el máximo rendimiento científico y social.

El análisis bibliométrico permite identificar áreas de conocimiento con un dinamismo excepcional, tasas de crecimiento prometedoras o un potencial de transferencia tangible a la sociedad. De esta forma, las agencias de financiación, como la National Science Foundation (NSF), pueden priorizar subsidios en campos emergentes con mayor probabilidad de impactar en

el futuro. Al mismo tiempo, estos datos ayudan a evitar la dispersión de esfuerzos en líneas de investigación saturadas o de escaso alcance. Un dato revelador que subraya esta necesidad de eficiencia es que, según un estudio de PLOS ONE (2024), aproximadamente el 60 % de los artículos científicos nunca son citados. Esta cifra invita a una reflexión profunda sobre la productividad académica y cuestiona la eficiencia de algunas líneas de investigación, impulsando un enfoque más cualitativo y estratégico.⁽⁴⁾

Más allá de las áreas temáticas, la bibliometría permite evaluar el capital investigador con una perspectiva multidimensional. Las instituciones ya no se fijan únicamente en el volumen de publicaciones de un candidato. Analizan métricas de influencia, como el número de citas, y, de manera crucial, cartografían sus redes de colaboración. Identificar investigadores con redes internacionales sólidas y consolidadas se ha convertido en un factor clave, ya que esta ciencia colaborativa suele traducirse en proyectos más ambiciosos, innovadores y visibles.

En definitiva, la bibliometría ha transformado la gobernanza científica. Provee a universidades y centros de investigación de un “panel de control” basado en evidencia para tomar decisiones informadas: desde contratar al talento que mejor potencie su ecosistema investigador, hasta reorientar sus departamentos hacia nichos de excelencia. Se trata, en esencia, de sustituir la conjetura por el análisis, optimizando cada recurso para fortalecer el conjunto del sistema de I+D+i.

Detectar irregularidades

En su faceta más crítica, la bibliometría trasciende su función evaluadora para convertirse en un instrumento esencial de vigilancia y control de calidad dentro del ecosistema científico. Ante la creciente presión por publicar y los incentivos perversos que pueden derivarse de un sistema excesivamente cuantitativo, surge la tentación de manipular los indicadores de rendimiento. Es aquí donde las técnicas bibliométricas se erigen como un “policía científico”, detectando y disuadiendo prácticas que comprometen la integridad académica.

Entre las irregularidades más comunes se encuentran las autocitas excesivas, donde un autor o grupo cita de manera desproporcionada su propio trabajo para inflar artificialmente su impacto; la publicación en revistas depredadoras, caracterizadas por procesos de revisión opacos y patrones anómalos de citas; y la formación de cárteles de citas (*citation cartels*), redes de revistas o autores que pactan citarse mutuamente para elevar sus métricas de forma artificial.⁽⁵⁾

Para combatir estas distorsiones, se han desarrollado herramientas algorítmicas sofisticadas. Plataformas como *Scimago Journal Rank* monitorizan el comportamiento de las revistas, señalando aquellas con perfiles de citación sospechosos, como un crecimiento anómalo y repentino de su factor de impacto. De forma más avanzada, algoritmos desarrollados en entornos como Python, entre ellos el *Citation Cartels Detector*, son capaces de analizar grandes volúmenes de datos para identificar clústeres de publicaciones que muestran patrones de citas circulares o mutuamente beneficiosas que no responden a un mérito científico genuino.⁽⁵⁾

La aplicación rigurosa de estos análisis es vital para proteger la equidad en la evaluación científica. Permite a las agencias de financiación, comités de promoción y editores de revistas tomar decisiones basadas en datos depurados, premiando el mérito real y aislando las prácticas fraudulentas. De este modo, la bibliometría no solo mide la excelencia, sino que también contribuye activamente a defenderla, asegurando que los recursos y reconocimientos se asignen con justicia y transparencia.

Promover la ciencia abierta

La irrupción del acceso abierto (OA, por sus siglas en inglés) ha reconfigurado radicalmente los cimientos de la comunicación científica, democratizando el conocimiento al eliminar las barreras de pago para lectores. La bibliometría, en este contexto, se ha erigido como la disciplina indispensable para cuantificar el verdadero alcance de esta transformación, ofreciendo perspectivas cruciales que van más allá del simple recuento de publicaciones.

Al analizar la distribución geográfica del OA dorado (donde el artículo se publica de forma inmediata y gratuita en la revista), los datos revelan un liderazgo claro de países europeos, como el Reino Unido, Alemania y los Países Bajos, gracias a fuertes mandatos institucionales y financieros. Les siguen de cerca naciones como Brasil e Indonesia, que han impulsado potentes repositorios nacionales. Este mapa bibliométrico no solo muestra adopción, sino también una clara división geoeconómica en la capacidad para asumir los costes de publicación (APCs), lo que plantea un desafío crucial para la equidad global.

El impacto del OA es particularmente significativo en disciplinas como las Humanidades. Tradicionalmente confinadas a un ciclo de comunicación más lento y con menores índices de citas, el acceso abierto ha impulsado su visibilidad de manera extraordinaria. Al eliminar el muro de pago, la investigación humanística alcanza a un público mucho más amplio: educadores, profesionales del sector cultural, periodistas y ciudadanos interesados. La bibliometría moderna, a través de las altmétricas, captura esta influencia ampliada, demostrando que el impacto en estos campos se mide no solo por citas, sino por su capacidad para permear el debate público y la cultura.

La llamada “ventaja de citas del OA”, respaldada por estudios que indican que los artículos en acceso abierto pueden recibir hasta un 50 % más de citas, por ejemplo, es un poderoso argumento para su implementación. Sin embargo, esta ventaja no es uniforme. Su magnitud varía sustancialmente según la disciplina y la región, lo que subraya la necesidad de políticas diferenciadas.

Para navegar este panorama complejo, herramientas como *Unpaywall* y el *Open Access Panel* de la Unión Europea se han vuelto esenciales. Estas plataformas rastrean en tiempo real el estado de millones de artículos, permitiendo visualizar tendencias, identificar barreras y evaluar la eficacia de las políticas nacionales. Así, la bibliometría proporciona la evidencia necesaria para diseñar estrategias que no solo impulsen el acceso abierto, sino que lo hagan de manera inteligente y equitativa, cerrando brechas y asegurando que el conocimiento sea un bien común, no un privilegio.⁽⁶⁾

1.1.2 Límites éticos: Lo que la bibliometría NO debe ser

La bibliometría es una herramienta poderosa, pero su uso acrítico puede generar distorsiones en la evaluación científica. Reconocer sus límites éticos es esencial para evitar que los indicadores cuantitativos se conviertan en fines en sí mismos, perpetuando desigualdades o sustituyendo el análisis cualitativo. A continuación, se desarrollan tres principios clave que todo investigador o institución debe considerar al aplicar métricas bibliométricas.

No es un fin en sí misma: las métricas deben servir al conocimiento, no reemplazar el juicio experto

Las métricas son medios, no fines. Reducir el valor de una investigación a su número de citas o índice h ignora dimensiones clave como su relevancia social, originalidad o rigor metodológico. Por ejemplo, estudios teóricos en filosofía pueden tener menos impacto inmediato que trabajos

aplicados en ingeniería, pero su influencia a largo plazo puede ser profunda. La evaluación académica debe combinar indicadores cuantitativos con revisiones por pares cualitativas. Casos como el *San Francisco Declaration on Research Assessment* (DORA) advierten contra el uso simplista de métricas en decisiones de contratación o financiamiento, promoviendo en su lugar una evaluación holística.

No es neutral: refleja sesgos de género, idioma y geografía

Los datos bibliométricos no son objetivos; reproducen las desigualdades estructurales de la ciencia. Por ejemplo:

- Sesgo de género: Solo el 33 % de los autores citados en inteligencia artificial son mujeres (estudio de *Nature*, 2023), lo que invisibiliza contribuciones clave.⁽⁷⁾
- Sesgo lingüístico: El 90 % de las publicaciones en Scopus están en inglés, marginando investigaciones en otros idiomas, especialmente en humanidades y ciencias sociales.⁽⁸⁾
- Sesgo geográfico: Las bases de datos privilegian revistas de Europa y EE.UU., subrepresentando a África, América Latina y Asia, incluso en áreas como biodiversidad o medicina tropical.⁽⁹⁾

Ignorar estos sesgos perpetúa ciclos de exclusión. Soluciones como *Cite Black Women* o el *Latin American Citation Index* buscan corregir estos desequilibrios, pero se requiere un esfuerzo consciente para diversificar las fuentes de análisis.

No es infalible

Los errores en citas y la manipulación distorsionan los resultados. Las métricas bibliométricas no son perfectas y pueden estar contaminadas por prácticas cuestionables o fallas sistémicas como citas incorrectas, autocitas infladas y revistas depredadoras.

Herramientas como *Citation Context Analysis* o algoritmos anti-manipulación ayudan a detectar estos problemas, pero la transparencia y la auditoría humana siguen siendo indispensables.

Esta guía práctica no solo instruye en el cálculo de indicadores bibliométricos, sino que fundamentalmente promueve su cuestionamiento sistemático, asegurando que la bibliometría sirva para enriquecer la práctica científica sin simplificarla o distorsionarla mediante aplicaciones reduccionistas. El equilibrio entre aprovechamiento analítico y conciencia crítica representa la clave para una implementación ética y efectiva de estas herramientas en el ecosistema investigador contemporáneo.

“La bibliometría es como un termómetro: mide la fiebre de la ciencia, pero no diagnostica la enfermedad ni receta la cura”.

Adaptado de Loet Leydesdorff.

Preguntas que la Bibliometría Responde

1. ¿Quiénes son los actores clave en mi campo? → Análisis de coautoría
2. ¿Está mi investigación alineada con tendencias globales? → Mapas temáticos
3. ¿Cómo demostrar mi impacto a evaluadores no expertos? → Índices h + altmétricas

1.2. Clasificación de las bibliometrías

La bibliometría, como cualquier disciplina científica, requiere de un divisiones para no perderse en la inmensidad de los datos. Estas permiten ordenar y dar sentido al vasto universo de publicaciones e indicadores. Lejos de ser una mera taxonomía abstracta, esta organización

es fundamental para aplicar la herramienta correcta a cada pregunta de investigación, ya sea medir el impacto de una revista, la productividad de un grupo o la solidez de una teoría. Sin este orden, estaríamos ante un caos de números sin narrativa.

Según el objeto de evaluación

La bibliometría se estructura en torno a distintos objetos de análisis, siendo de revistas científicas y de investigadores los más destacados. Para el primer caso, se utilizan indicadores de prestigio como el Factor de Impacto (JCR) o el *SCImago Journal Rank* (SJR), los cuales miden la influencia media de los artículos publicados en una revista y son empleados frecuentemente para la toma de decisiones en políticas editoriales y de evaluación. En el caso de los autores y sus grupos de investigación, se aplican métricas como el índice h o el índice i10, diseñadas para sintetizar en una única cifra su productividad y el impacto percibido de su trabajo, ofreciendo así una visión de su trayectoria y relevancia dentro de la comunidad académica. Esta distinción es crucial para no confundir el impacto de una publicación periódica con el mérito individual de quien publica en ella.⁽¹⁰⁾

Según la naturaleza de la medida

Una perspectiva complementaria organiza los indicadores según la cualidad que miden. Las métricas de productividad, como el número total de publicaciones, ofrecen una cuantificación básica del output investigador, aunque sin ponderar su alcance o significancia. Un nivel de análisis más profundo lo proporcionan las métricas de impacto o visibilidad, cuyo ejemplo más directo son los recuentos de citas, que buscan reflejar la utilidad y la adopción de los hallazgos científicos por parte de la comunidad académica. Finalmente, las métricas de colaboración, a menudo representadas mediante mapas de redes de coautoría, trascienden lo cuantitativo para examinar los patrones de cooperación intelectual, revelando la estructura y la dinámica de las relaciones entre investigadores, instituciones y países.⁽¹¹⁾

Según la dimensión temporal

Este texto introduce la temporalidad como un criterio de clasificación fundamental, argumentando que la ventana de tiempo analizada condiciona sustancialmente la interpretación de cualquier indicador. Los análisis de series temporales constituyen el enfoque más robusto, ya que permiten observar la evolución de los indicadores a lo largo de varios períodos. Como se desarrollará en capítulos posteriores, los indicadores bibliométricos pueden combinarse en series temporales para identificar tendencias de crecimiento, puntos de inflexión en un campo de estudio o la consolidación del impacto de un investigador, proporcionando una narrativa dinámica y contextualizada. Por contraste, un análisis de periodo temporal corto, como un solo año, se limita a ofrecer una instantánea estática. Esta visión fragmentaria es incapaz de distinguir entre una tendencia sostenida y una fluctuación puntual, resultando generalmente insuficiente para evaluar la solidez de una línea de investigación o la trayectoria de un científico.

1.3. Diferencias clave: bibliometría vs. scientometría vs. Altimetría.

Tres Lentes para Medir la Ciencia

En el ecosistema de la investigación cuantitativa, bibliometría, scientometría y altimetría son disciplinas hermanas con enfoques complementarios. Mientras todas miden el impacto del conocimiento, lo hacen desde perspectivas distintas. Esta sección desentraña sus diferencias fundamentales con ejemplos prácticos y aplicaciones típicas.

1.3.1. Bibliometría: el análisis de lo publicado

La bibliometría es el estudio cuantitativo de documentos académicos, como artículos, libros y patentes, y sus relaciones mediante citas, coautorías y palabras clave, abordando a lo largo

de este libro todos sus elementos fundamentales. Se centra principalmente en publicaciones formales de revistas indexadas y emplea indicadores clásicos como conteo de citas, índice h y factor de impacto para medir impacto y visibilidad. Entre sus aplicaciones prácticas destacan la evaluación de la productividad individual de investigadores, la identificación de revistas centrales en disciplinas específicas y el mapeo de colaboraciones científicas a través del análisis de redes de coautoría, ofreciendo así herramientas objetivas para comprender y optimizar la dinámica de la producción científica.

1.3.2. Cienciometría: La Ciencia de la Ciencia

La cienciometría es el estudio cuantitativo de la ciencia como sistema social, analizando no solo su producción documental, sino también sus patrones de crecimiento, colaboración e impacto. A diferencia de la bibliometría, que se centra en métricas de publicaciones y citas, la cienciometría adopta una visión más amplia, examinando cómo se estructura, evoluciona y se financia la actividad científica.

Enfoque principal: Redes de conocimiento y leyes bibliométricas

La cienciometría estudia las conexiones entre disciplinas y el surgimiento de nuevas áreas de investigación mediante el uso de diversas herramientas metodológicas. Entre ellas destaca el análisis de redes de colaboración, que permite visualizar y comprender cómo investigadores, instituciones y países interactúan en la generación de conocimiento científico. Este enfoque revela patrones de cooperación y flujos de información que configuran el panorama investigativo.

Además, la cienciometría se apoya en leyes bibliométricas fundamentales para explicar fenómenos científicos. La Ley de Lotka describe la distribución asimétrica de la productividad académica, demostrando que un reducido grupo de investigadores genera la mayor parte de la producción científica. Por otro lado, la Ley de Bradford ayuda a identificar el núcleo limitado de revistas que concentran las publicaciones más relevantes dentro de un campo específico.

Estos principios teóricos y metodológicos permiten descubrir dinámicas ocultas en el mundo científico, como la concentración del conocimiento en ciertas instituciones líderes o la aparición de disciplinas emergentes que se desprenden de áreas tradicionales. A través de este enfoque analítico, la cienciometría proporciona una comprensión más profunda de la estructura y evolución del conocimiento científico.

Aplicaciones prácticas: desde la predicción de tendencias hasta políticas de I+D

La cienciometría se ha consolidado como una herramienta estratégica para la toma de decisiones en política científica, ofreciendo perspectivas valiosas sobre dinámicas investigativas que trascienden el análisis bibliométrico tradicional. Un aspecto clave de su aplicación es la identificación temprana de disciplinas emergentes, donde técnicas como el análisis de co-citaciones han demostrado su capacidad predictiva. Este fue el caso de la inteligencia artificial aplicada a diagnóstico médico, donde patrones de citación cruzada entre disciplinas anticiparon su adopción clínica años antes de que se manifestara en la literatura especializada.

En el ámbito de la política científica, la cienciometría proporciona metodologías robustas para evaluar el retorno de inversión en I+D. Países líderes en innovación como Corea del Sur han implementado sistemas cienciométricos avanzados que correlacionan indicadores de producción científica con métricas económicas, permitiendo cuantificar el impacto de subsidios públicos en áreas estratégicas como la nanotecnología. Estos modelos han demostrado ser particularmente útiles para optimizar la asignación de recursos en sectores de alto potencial tecnológico.

Otro campo donde la cienciometría aporta valor único es en el estudio de la movilidad científica global. A través del análisis de patrones de publicación y afiliación institucional, ha sido posible mapear flujos de talento investigador con precisión sin precedentes. Los datos revelan fenómenos complejos como la diáspora científica india, donde cerca del 40 % de los investigadores que publican en colaboración con instituciones estadounidenses permanecen en EE.UU., según un reciente estudio publicado en *Nature* (2023). Estos hallazgos tienen implicaciones profundas para el diseño de políticas de retención de talento y programas de repatriación en países en desarrollo.

Desafíos y sesgos en el análisis cienciométrico

Aunque la cienciometría ofrece una lente poderosa para comprender la empresa científica, es crucial reconocer que esta lente no es perfectamente transparente. Su potencial analítico convive con una serie de limitaciones estructurales y metodológicas que exigen una interpretación cautelosa y crítica de sus resultados.

En primer lugar, la disciplina depende por completo de la calidad y amplitud de los macrodatos provenientes de bases de datos globales como Scopus y Web of Science. Estas plataformas, pese a su exhaustividad, no son espejos neutrales de la producción global de conocimiento. Tienen una subrepresentación sistemática de las investigaciones procedentes de regiones en desarrollo, de publicaciones en idiomas distintos al inglés y de revistas locales, lo que distorsiona el mapa global de la ciencia y perpetúa un sesgo geolingüístico.

En segundo término, la sofisticación de las técnicas cienciométricas actúa como un arma de doble filo. Modelos matemáticos complejos, como el análisis de redes para cartografiar colaboraciones o la aplicación de la Ley de Price para modelar el crecimiento exponencial de la literatura, requieren una especialización estadística avanzada. Una aplicación incorrecta o una interpretación superficial de estos modelos puede llevar a conclusiones erróneas, generando un espejismo de precisión cuantitativa que oculta simplificaciones o malentendidos conceptuales.

Finalmente, y de manera más profunda, la cienciometría hereda y, en ocasiones, amplifica, los sesgos inherentes al propio sistema científico. Los datos que analiza no se generan en un vacío equitativo; reflejan las desigualdades estructurales en financiación, el predominio del inglés como lingua franca y las dinámicas de poder académico. Por lo tanto, un mapa de colaboraciones puede mostrar no solo la excelencia, sino también la exclusión; y un indicador de impacto puede estar midiendo, en parte, la visibilidad preexistente de una institución o país.

En conclusión, la cienciometría es una herramienta de diagnóstico formidable, pero no un oráculo. Su verdadero valor no reside en la aceptación acrítica de sus métricas, sino en la capacidad de los investigadores, gestores y policy-makers para comprender sus supuestos, reconocer sus sesgos y utilizar sus hallazgos como una guía informada, y nunca como un veredicto definitivo.

1.3.3. Altmetría: el Impacto en la sociedad digital

En un mundo donde la ciencia se discute en Twitter, se aplica en políticas públicas y se viraliza en podcasts, las métricas tradicionales como el *Journal Impact Factor* o el conteo de citas ya no son suficientes para capturar la influencia real de la investigación. Aquí es donde entra la altmetría (*alternative metrics*), que rastrea el impacto de los estudios en espacios digitales y no académicos, ofreciendo una visión más amplia de su relevancia social.

Enfoque principal: ciencia en la esfera pública

La altmetría aborda cuestiones fundamentales sobre el impacto social de la investigación científica más allá del ámbito académico tradicional. Esta disciplina analiza cómo se utiliza y discute el conocimiento científico en diversos espacios públicos y digitales, proporcionando una visión más amplia del alcance real del trabajo investigativo.

Las menciones en redes sociales como Twitter, Facebook o Reddit reflejan el nivel de interacción y discusión pública que genera una investigación. Cuando los hallazgos científicos aparecen en Wikipedia o son tratados en blogs especializados, esto demuestra su penetración en la cultura general y su influencia en la divulgación del conocimiento. Más significativamente aún, las referencias en documentos gubernamentales o informes de organizaciones no gubernamentales señalan el impacto concreto de la investigación en la formulación de políticas y la toma de decisiones.

Para medir estos impactos, la altmetría emplea plataformas especializadas como PlumX, Altmetric.com y Dimensions, que recopilan y analizan diversos tipos de datos. Estos incluyen desde las descargas en plataformas académicas como ResearchGate y Academia.edu, que muestran el interés directo de la comunidad investigadora, hasta la cobertura en medios de comunicación masiva como la BBC o *The New York Times*, que indican la relevancia pública del trabajo. Además, el uso de investigaciones en cursos online de plataformas como Coursera o edX evidencia su incorporación en procesos educativos formales e informales.

Estos indicadores altmétricos complementan las métricas tradicionales de citación, ofreciendo una visión multidimensional del impacto científico que abarca desde la academia hasta la sociedad en general, pasando por la política, la educación y los medios de comunicación.

Aplicaciones prácticas: desde políticas públicas hasta divulgación

La altmetría se ha convertido en una herramienta fundamental para evaluar el impacto real de la investigación científica en diversos ámbitos. Uno de sus principales valores radica en su capacidad para medir el impacto social inmediato de estudios sobre temas urgentes como el cambio climático, la salud pública o la equidad de género, los cuales suelen generar un intenso debate en plataformas digitales y medios de comunicación mucho antes de acumular citaciones académicas tradicionales.

Esta disciplina resulta particularmente valiosa para cuantificar la efectividad de las estrategias de divulgación científica, permitiendo medir el alcance de iniciativas como campañas en redes sociales. Cuando una universidad comparte un descubrimiento a través de TikTok, las métricas altmétricas de shares, likes y comentarios ofrecen datos concretos sobre su penetración en la sociedad, información que las citas académicas no podrían proporcionar en etapas tempranas.

Además, la altmetría juega un papel crucial en la justificación de financiamiento para proyectos de investigación. La presencia de un artículo científico en documentos de política pública, como un plan estatal de energías renovables, o su uso por parte de organizaciones no gubernamentales, demuestra de manera tangible la utilidad práctica de la investigación. Este tipo de evidencia resulta fundamental para convencer a gobiernos y organismos financiadores sobre el valor aplicado del trabajo científico, especialmente en áreas donde el impacto social es tan importante como el académico.

El reverso de la moneda: los desafíos intrínsecos de las altmétricas

La promesa de las altmétricas de capturar el impacto social en tiempo real es, sin duda,

revolucionaria. Sin embargo, este dinamismo conlleva una serie de desafíos fundamentales que es necesario reconocer para evitar una interpretación ingenua de sus datos.

Uno de los riesgos más significativos reside en la calidad y autenticidad de las fuentes. La viralidad no es sinónimo de rigor. Un tuit puede acumular miles de likes por un titular sensacionalista, y campañas de bots o debates altamente polarizados pueden inflar artificialmente las menciones, creando un espejismo de relevancia que no se sustenta en una discusión académica o seria. Distinguir entre el ruido y la señal, entre la manipulación y el engagement genuino, se convierte en una tarea crítica.

Asimismo, las altmétricas presentan un sesgo disciplinar pronunciado. Temas de gran repercusión mediática e inmediata aplicabilidad, como un avance médico o una catástrofe natural, dominan de forma natural el espacio digital. Por el contrario, disciplinas que operan en ciclos de debate más largos y especializados, como la filosofía teórica o las matemáticas puras, generan de forma natural una huella altmétrica mucho más tenue. Esto no significa que su impacto social sea menor, sino que se manifiesta de formas más sutiles y menos cuantificables a través de estos canales, pudiendo penalizarlas en evaluaciones que prioricen estos indicadores.

Finalmente, la falta de estandarización metodológica plantea un obstáculo para la comparabilidad. Plataformas como Altmetric.com y PlumX emplean algoritmos propios, ponderan de manera diferente las diversas fuentes (por ejemplo, las vistas en Wikipedia frente a las menciones en prensa) y definen su “puntuación” de forma única. La ausencia de un protocolo unificado significa que un mismo artículo puede tener puntuaciones altmétricas radicalmente distintas en cada plataforma, lo que dificulta la creación de benchmarks fiables y socava la consistencia en la evaluación.

En definitiva, las altmétricas no deben reemplazar a los indicadores tradicionales, sino complementarlos. Su valor máximo se alcanza cuando se las aborda con una mirada crítica, entendiendo sus sesgos y empleándolas para contar una historia más rica y matizada sobre la diseminación del conocimiento en la sociedad, y no como un simple número definitivo.

Ejemplo de caso integrado: un estudio sobre vacunas

Bibliometría:

- Análisis: 5,000 artículos sobre vacunas en Scopus.
- Hallazgo: El 60 % de las publicaciones proviene de EE.UU., China y Reino Unido.

Cienciometría:

- Análisis: Patrones de colaboración entre países.
- Hallazgo: Los equipos internacionales producen un 40 % más de patentes.

Altmetría:

- Análisis: Menciones en Facebook y noticias.
- Hallazgo: Los artículos con resúmenes en lenguaje sencillo tienen 3x más impacto público.

Tres herramientas, un propósito

Estas disciplinas no son excluyentes, sino complementarias:

- Bibliometría responde “¿Quién cita a quién?”.
- Cienciometría explica “¿Por qué la ciencia crece así?”.
- Altmetría revela “¿Cómo la sociedad usa este conocimiento?”.

“Usar solo bibliometría es como medir un océano con un vaso: necesitas altimetría para ver las olas y cienciometría para entender las corrientes.”

1.4. Aplicaciones prácticas: evaluación científica y políticas de investigación

De los datos a las decisiones

La bibliometría trasciende el mero conteo de citas para convertirse en una herramienta estratégica en la gestión del conocimiento. Este epígrafe explora cómo gobiernos, universidades y agencias de financiamiento utilizan indicadores bibliométricos para evaluar ciencia y diseñar políticas, con ejemplos concretos y controversias éticas.

Evaluación científica

En el capítulo 5 se desarrollarán con más profundidad los índices y métodos de evaluación. Aquí se muestran algunos de sus usos.

Evaluación de investigadores

La evaluación de investigadores mediante indicadores presenta un complejo equilibrio entre la medición objetiva del impacto científico y la preservación de la integridad académica. Entre los indicadores más utilizados destaca el índice *h* y sus variantes (*g*, *m*), que buscan capturar simultáneamente la productividad y el impacto de un investigador, aunque con limitaciones inherentes. La citación normalizada por campo emerge como una herramienta más refinada, permitiendo comparaciones justas entre disciplinas con diferentes dinámicas de publicación y citación. Complementariamente, el porcentaje de publicaciones en el percentil superior de citación (top 10 %) ofrece una perspectiva valiosa para identificar investigación verdaderamente transformadora en cada área del conocimiento.

Sin embargo, estos sistemas de evaluación no están exentos de polémicas significativas. Más problemático aún resulta el fetichismo del índice *h*, fenómeno que ha generado consecuencias no deseadas en el ecosistema científico: por un lado, desincentiva la investigación arriesgada y disruptiva que suele tener ciclos de maduración más largos; por otro, promueve una cultura de publicación incremental en detrimento de contribuciones fundamentales pero menos frecuentes.⁽¹²⁾ Estas tensiones entre métricas cuantitativas y calidad científica real continúan generando intensos debates sobre cómo evaluar el mérito investigador sin sacrificar la diversidad epistemológica y la innovación radical.

Caso real:

La Universidad de Leiden (Países Bajos) usa citación normalizada para evitar sesgos contra investigadores en campos con baja tasa de citas (ej: matemáticas vs. biomedicina).⁽¹³⁾

Evaluación de revistas

La evaluación del prestigio y la influencia de las revistas académicas es una piedra angular de la bibliometría, pero ha evolucionado para superar la dependencia exclusiva de métricas unívocas. El *Journal Impact Factor* (JIF), calculado por *Clarivate Analytics*, ha sido durante décadas el indicador hegemónico. Mide la frecuencia con la que los artículos de una revista publicados en un bienio son citados en un año concreto. Sin embargo, su metodología ha sido objeto de críticas sustanciales: presenta un sesgo inherente hacia las revistas en inglés y de ciencias naturales, marginando publicaciones de alta calidad en humanidades, ciencias sociales o revistas regionales, cuyas dinámicas de citación son más pausadas.

Como respuesta a estas limitaciones, han surgido indicadores alternativos que ofrecen perspectivas más matizadas. El *SCImago Journal Rank* (SJR), basado en la base de datos Scopus,

introduce un concepto clave: la transferencia de prestigio. No todas las citas valen lo mismo; una cita proveniente de una revista de alto prestigio tiene más peso que una de una publicación menor. Este enfoque, inspirado en el algoritmo de Google, ayuda a identificar revistas que son referentes influyentes dentro de sus nichos específicos, incluso si su número total de citas no es el más alto.

Complementariamente, el *CiteScore* (también de Elsevier/Scopus) amplía la ventana temporal y el tipo de documentos considerados en el cálculo. Incluye no solo artículos de investigación, sino también revisiones, conferencias, capítulos de libros y notas, lo que lo hace especialmente relevante para disciplinas como la ingeniería o las ciencias de la computación, donde las actas de congresos tienen un valor seminal. Esta mayor inclusividad proporciona una visión más representativa del impacto de una revista en su ecosistema disciplinar completo.

Evaluación de instituciones

La evaluación bibliométrica de universidades y centros de investigación ha trascendido el simple conteo de publicaciones para adoptar métricas avanzadas que reflejen la calidad, la internacionalización y el liderazgo científico.

Dos indicadores son particularmente reveladores:

- **Grado de Colaboración Internacional:** calculado como el porcentaje de publicaciones de una institución que tienen coautores extranjeros. Un índice alto es un fuerte indicador de la integración de la institución en redes globales de conocimiento, su capacidad para atraer talento internacional y su acceso a proyectos de gran envergadura.
- **Índice de Excelencia Investigadora:** mide el porcentaje de artículos de una institución que se encuentran entre el 10 % más citado del mundo en sus respectivos campos. Esta métrica no se limita a cuantificar la producción, sino que identifica la capacidad de una institución para generar investigación de vanguardia y alto impacto, es decir, para situarse en la frontera del conocimiento.

En conjunto, estos instrumentos, tanto para revistas como para instituciones, facilitan una transición crucial: pasar de una cultura de la cantidad a una cultura del impacto y la excelencia, siempre que se interpreten con conciencia de sus contextos y limitaciones.

1.4.1 Políticas de investigación

Como se ha demostrado en epígrafes anteriores, la bibliometría trasciende su función descriptiva para convertirse en un instrumento fundamental en la gobernanza de la ciencia. Al proporcionar una evidencia sólida y cuantitativa, permite a gobiernos e instituciones diseñar políticas de investigación más inteligentes, estratégicas y equitativas. Lejos de ser una mera herramienta de evaluación a posteriori, la bibliometría informa la toma de decisiones a priori, guiando la distribución de recursos, fomentando colaboraciones de alto impacto y acelerando la transición hacia una ciencia más abierta y colaborativa. A continuación, se detallan sus tres aplicaciones clave en la gestión de políticas científicas.

Distribución de financiamiento: invertir con estrategia, no por inercia

Las agencias de financiación han dejado atrás la era de las asignaciones basadas únicamente en la intuición o el prestigio histórico. Hoy, los análisis bibliométricos ofrecen un mapa dinámico para invertir en lo que realmente importa. Mediante técnicas como el mapeo de ciencia, es posible identificar áreas emergentes con crecimiento exponencial en publicaciones y citas, señalando dónde una inversión tendrá el máximo rendimiento. De forma complementaria, el análisis de redes de citas y la identificación de brechas de conocimiento revelan temas científicos

descuidados que, sin embargo, son cruciales para el avance interdisciplinario o para abordar desafíos sociales urgentes, permitiendo dirigir fondos de manera proactiva hacia estos vacíos.

Diseño de programas de colaboración: conectar talentos globales

La ciencia contemporánea es inherentemente colaborativa. La bibliometría actúa como un catalizador para diseñar estas alianzas de forma estratégica. El análisis de coautoría permite identificar instituciones y países con intereses de investigación complementarios y fortalezas sinérgicas, allanando el camino para convocatorias de proyectos conjuntos. El análisis de la movilidad científica, rastreando la tasa de investigadores que migran entre regiones, proporciona datos cruciales para políticas de retención y atracción de talento. El dato de que aproximadamente el 35 % de los científicos africanos especializados en IA trabajan en EE.UU, esto es un pilar bien establecido en los estudios de política científica y es mencionado por numerosos organismos como la OCDE y la UNESCO.⁽¹⁴⁾

Políticas de ciencia abierta: medir para democratizar

Como se discutió previamente, el acceso abierto es una fuerza transformadora. La bibliometría proporciona los indicadores necesarios para medir el avance hacia este objetivo y diseñar políticas efectivas que democratizen el conocimiento. Plataformas como el Open Access Panel de la UE, mencionado antes, permiten monitorizar en tiempo real el porcentaje de publicaciones en acceso abierto de un país o institución, el predominio de las distintas rutas (dorada, verde) y las brechas existentes entre disciplinas y regiones. Estos datos son esenciales para ajustar mandatos, negociar con editoriales y asegurar que la transición hacia la ciencia abierta sea inclusiva y reduzca, en lugar de ampliar, las desigualdades en el acceso al conocimiento.

La bibliometría ha dejado de ser un simple espejo que refleja la actividad científica. Al proporcionar la evidencia para una toma de decisiones estratégica, se ha convertido en un motor que no solo describe la ciencia, sino que la transforma, ayudando a construir un ecosistema de investigación más eficiente, conectado y abierto.

1.4.2 Bibliometría como termómetro social

La bibliometría dista de ser una herramienta neutral: constituye un reflejo de los valores, prioridades y sesgos inherentes al ecosistema científico contemporáneo. Su verdadero poder transformador radica en tres dimensiones críticas:

En primer lugar, funciona como brújula para la optimización de recursos escasos. Al cuantificar el impacto demostrable de la investigación, ya sea mediante citaciones, indicadores de transferencia tecnológica o métricas de política pública, permite asignar financiamiento con base en evidencia más que en intuiciones o tradiciones académicas. Este enfoque resulta particularmente valioso en contextos donde cada inversión en I+D debe justificarse ante sociedades que demandan retornos tangibles.

Simultáneamente, la bibliometría bien aplicada puede democratizar los procesos evaluativos. Frente a sistemas tradicionales basados en redes de influencia o reputación institucional, las métricas estandarizadas ofrecen, en teoría, un lenguaje común y reproducible.

Sin embargo, su potencial más revolucionario reside en su capacidad para desnudar y combatir desigualdades estructurales. Los análisis bibliométricos han expuesto con crudeza la concentración geográfica de la producción científica, brechas de género persistentes (las investigadoras reciben sistemáticamente menos citas en campos como IA o física teórica y el

menosprecio métrico hacia disciplinas como las humanidades o las ciencias sociales.

El desafío ético consiste en utilizar estos diagnósticos no para perpetuar jerarquías existentes, sino para diseñar políticas correctivas, como los criterios de evaluación diferenciados por disciplina o las cuotas de financiamiento para regiones subrepresentadas.

Esta tensión entre reproducción y transformación del status quo convierte a la bibliometría en un campo de batalla crucial para definir qué tipo de ciencia queremos construir: una que premie solo la productividad inmediata, o una que valore también la diversidad epistemológica y el impacto social a largo plazo. La presente guía proporcionará las herramientas para navegar este dilema con rigor metodológico y conciencia crítica.

“Las métricas son como un bisturí: en manos expertas salvan vidas (científicas); en manos inexpertas, mutilan carreras.”

1.5. Ética y buenas prácticas en bibliometría

1.5.1 Riesgos éticos: cuando los números refuerzan desigualdades

La bibliometría, con su poder de cuantificar, ha aportado una apariencia de objetividad a la evaluación científica. Sin embargo, es fundamental reconocer que estos indicadores no operan en un vacío neutral. Por el contrario, están impregnados de los sesgos y dinámicas de poder preexistentes en el sistema académico, pudiendo perpetuarlos y amplificarlos. Esta sección expone los principales riesgos éticos de una aplicación acrítica de la bibliometría y explora las estrategias emergentes para construir una evaluación científica más justa y representativa.

El “Efecto Mateo” en la Ciencia

Este fenómeno, bautizado por el sociólogo Robert K. Merton a partir de un pasaje bíblico, describe con crudeza cómo la ventaja inicial atrae ventajas acumulativas. En la ciencia, se materializa en la Ley de Price, que postula que una minoría (alrededor del 20 % de los investigadores) concentra la inmensa mayoría (aproximadamente el 80 %) de las citas y el reconocimiento.⁽¹⁵⁾

Este ciclo virtuoso para unos pocos se convierte en una trampa para muchos. Jóvenes investigadores brillantes, instituciones de países en desarrollo o científicos de campos marginales luchan denodadamente por alcanzar visibilidad, incluso cuando producen trabajo de la más alta calidad. El sistema, al premiar principalmente a quienes ya son conocidos, puede sofocar la innovación que surge desde los márgenes.

La aplicación rígida de estos indicadores también puede distorsionar los incentivos, premiando la productividad rápida y la citabilidad predecible por encima de la investigación arriesgada, de largo alcance o con impacto social no inmediatamente cuantificable. Asimismo, existe el riesgo de excluir formas valiosas de conocimiento, como los saberes locales o las contribuciones en lenguas distintas al inglés, que no se ajustan a los canales formales de publicación indexada.

Frente a este panorama, el movimiento por una cienciometría responsable gana fuerza, abogando por el uso contextualizado de los indicadores, la evaluación cualitativa por pares y el desarrollo de métricas que capturen la diversidad de contribuciones al ecosistema del conocimiento. La bibliometría debe ser un sirviente de la ciencia, no su amo.

La bibliometría también revela y, en ocasiones, refuerza, profundas desigualdades estructurales. Un claro ejemplo es el sesgo de género, como se ha mencionado anteriormente.

La crisis de la evaluación cuantitativa

La aplicación rígida de los indicadores bibliométricos ha generado distorsiones profundas en la evaluación científica, donde cifras arbitrarias pueden llegar a pesar más que el mérito intelectual real. Esta crisis se manifiesta cuando comités de promoción establecen barreras numéricas inflexibles, como exigir un Factor de Impacto superior a 20, que terminan desvirtuando el sentido mismo de la evaluación al ignorar la calidad intrínseca, la originalidad y el aporte concreto de una investigación. Estos abusos en el uso de las métricas, donde un número opaca el valor real del trabajo, fueron lo que motivó la creación de iniciativas globales como DORA (Declaración de San Francisco sobre la Evaluación de la Investigación), que buscan restaurar el juicio experto y cualitativo como eje central de la evaluación científica.⁽¹⁶⁾

Los dos pilares de la bibliometría responsable

Declaración DORA

“Evaluar la ciencia por su contenido, no por el contenedor”

La Declaración DORA (*San Francisco Declaration on Research Assessment*), establecida en 2013, marca un punto de inflexión en la evaluación científica al proponer un cambio de paradigma: valorar la investigación por su contenido real y no por el prestigio del medio donde se publica. Esta iniciativa surge como respuesta crítica a la sobrevaloración del Factor de Impacto (FI) de las revistas, promoviendo en su lugar una evaluación más holística y justa del trabajo investigador. Sus principios fundamentales incluyen la prohibición expresa de utilizar el FI como medida de calidad para investigadores individuales, el reconocimiento de la diversidad de productos científicos, desde conjuntos de datos y software hasta actividades de divulgación, y la exigencia de máxima transparencia en los métodos de cálculo y fuentes de datos utilizados para cualquier evaluación.⁽¹⁶⁾

Su adopción masiva refleja un consenso creciente sobre la necesidad de reformar los sistemas de evaluación científica, desplazando el foco desde métricas superficiales hacia una valoración cualitativa y contextualizada de las contribuciones al conocimiento. La declaración ha impulsado además el desarrollo de nuevas prácticas evaluativas que reconocen el valor social de la investigación y fomentan la ciencia abierta, marcando el camino hacia un ecosistema académico más equitativo y centrado en el verdadero impacto del trabajo científico.

Manifiesto de Leiden

“10 Mandamientos para el Uso de Indicadores”

El Manifiesto de Leiden (2015), que propone un conjunto de principios fundamentales para el uso adecuado de los indicadores bibliométricos, enfatiza la importancia de ajustar los análisis según la disciplina en la que se apliquen.⁽¹⁷⁾

De acuerdo con este principio, un índice h de 12 podría considerarse excelente en el campo de la filosofía, pero se consideraría bajo si se aplica al ámbito de la biología, debido a las diferencias inherentes en las dinámicas de publicación y citación entre distintas áreas del conocimiento. Además, se subraya la necesidad de priorizar la calidad sobre la cantidad al evaluar el impacto de la investigación. Así, un artículo que sea verdaderamente innovador y revolucionario debería tener más peso que varios estudios triviales que solo aportan a la acumulación de citas sin generar un verdadero avance en la disciplina. Finalmente, el manifiesto recalca la importancia de auditar las bases de datos utilizadas para estos análisis, ya que muchas de ellas, como es el caso de Scopus, pueden pasar por alto una gran parte de la producción académica, especialmente en áreas como las humanidades, donde el 80 % de los libros relevantes pueden quedar excluidos del análisis.

Estos principios, en conjunto, buscan hacer un uso más riguroso y reflexivo de los indicadores bibliométricos, con el fin de evitar distorsiones en la evaluación de la ciencia.

Los 7 pecados capitales de la bibliometría

La bibliometría nació para iluminar los caminos de la ciencia, pero en su aplicación desmedida ha terminado por crear nuevos laberintos. Lo que comenzó como una herramienta de orientación se ha convertido en muchos casos en un fin en sí mismo, distorsionando los valores fundamentales de la investigación. Identificar estos “pecados capitales” de la evaluación bibliométrica es el primer paso para recuperar su sentido original: servir a la ciencia, no dominarla.

El **fetichismo del factor de impacto** representa quizá la desviación más evidente. Hemos convertido un número creado para ayudar en la selección de revistas en un oráculo incuestionable. Esta idolatría ha generado consecuencias paradójicas: mientras más obsesionados estamos con el factor de impacto, menos nos fijamos en lo que realmente importa, la calidad intrínseca de la investigación. La solución no está en abandonar las métricas, sino en elegirlas con sabiduría, optando por indicadores más matizados como el SJR o el CiteScore, que ofrecen una visión más contextual y menos susceptible a la manipulación.

Las **autocitas infladas** nos hablan de una triste realidad: en el afán por escalar posiciones, algunos investigadores han convertido las citas en una moneda de cambio en lugar de un genuino reconocimiento intelectual. Cuando las referencias dejan de ser un diálogo académico para convertirse en una estrategia de posicionamiento, la propia esencia de la ciencia como empresa colaborativa se resquebraja. Establecer límites razonables, como ese 20 % propuesto por el prestigioso Centro de Leiden, no es solo una cuestión técnica: es un recordatorio de que la integridad debe prevalecer sobre el oportunismo.

La **ignorancia de los contextos disciplinares** equivale a juzgar a un pez por su capacidad para trepar árboles. Las humanidades, las ciencias sociales y las artes tienen ritmos, formatos y tradiciones de comunicación que no caben en los moldes diseñados para las ciencias naturales. Exigirles que jueguen con las mismas reglas no solo es injusto, es científicamente incorrecto. La verdadera excelencia se mide dentro de cada ecosistema disciplinar, no mediante comparaciones forzadas que solo generan frustración y homogenización.

El **uso irreflexivo de los rankings institucionales** ha creado un efecto de “riqueza genera riqueza” que ahonda las brechas globales. Las instituciones del Sur Global, por más talento que tengan, luchan contra un sistema de evaluación que premia precisamente los recursos de los que carecen. Los rankings no miden solo calidad, miden también privilegios históricos y ventajas estructurales. Por eso necesitamos indicadores que sepan contextualizar, que comprendan que la excelencia en Nairobi no se parece, ni debe parecerse, a la excelencia en Harvard.

La **exclusión de los libros** como fuente de evaluación equivale a amputar buena parte del pensamiento más profundo de las humanidades y ciencias sociales. Mientras premiamos artículos breves y resultados inmediatos, penalizamos la reflexión pausada, el estudio minucioso, la obra que tarda años en madurar. Incorporar herramientas como Google Scholar que reconocen estos formatos no es una concesión, es un acto de justicia epistemológica.

La **obesidad métrica** nos ha llevado a creer que más datos significan mejor evaluación. Pero la experiencia nos muestra lo contrario: cuando todo se mide, nada se comprende. La proliferación interminable de indicadores no nos ha acercado a la verdad, nos ha distraído de ella. La elegancia está en la simplicidad: cinco indicadores bien elegidos pueden decirnos

mucho más que cincuenta mal interpretados.

La **opacidad algorítmica** constituye quizá el desafío más insidioso. Evaluamos con herramientas cuyos criterios desconocemos, confiando en cajas negras que pueden esconder los mismos sesgos que decimos combatir. La transparencia no es aquí solo una virtud, es una condición de supervivencia. Sin auditorías públicas, sin escrutinio abierto, la evaluación bibliométrica se convierte en un dogma incuestionable.

La corrección de estos excesos no representa un retroceso, sino una maduración necesaria. Nos invita a construir una bibliometría a escala humana, rigurosa pero comprensiva, ambiciosa pero humilde. Una que sepa que detrás de cada número hay un investigador, detrás de cada ranking hay una institución, y detrás de cada métrica hay una decisión ética sobre qué ciencia queremos y para quién.

Marco de Buenas Prácticas (UNESCO, 2021)

En respuesta a las crecientes distorsiones en los sistemas de evaluación científica, la UNESCO ha establecido un marco de buenas prácticas que busca reequilibrar la forma en que entendemos y valoramos la producción de conocimiento. Este marco, fruto del consenso global, propone transformaciones concretas en los tres pilares del ecosistema académico: evaluación, autoría y edición.⁽¹⁸⁾

Para los evaluadores, el marco propone una fórmula equilibrada: combinar métricas tradicionales con la revisión por pares cualitativa en una proporción 70/30. Esta aproximación reconoce el valor de los datos bibliométricos, pero sitúa en el centro del proceso el criterio experto. La novedad radica en cómo se integran las métricas: se insta a incorporar sistemáticamente la altmetría para capturar el impacto social de la investigación, reconociendo así que el valor del conocimiento trasciende las citas académicas y se manifiesta en su capacidad para influir en políticas públicas, transformar prácticas sociales o enriquecer el debate cultural.

En el ámbito de la autoría, el marco aborda dos prácticas particularmente dañinas. Por un lado, desincentiva el *“salami slicing”*, esa tendencia a segmentar artificialmente una investigación en el menor número de publicaciones posibles, recordando que la integridad intelectual debe primar sobre la productividad numérica. Por otro lado, establece un umbral ético para el autocitado, recomendando que no supere el 15 % del total de citas. Este límite no busca limitar la construcción coherente de una trayectoria investigadora, sino asegurar que el trabajo se valide a través del reconocimiento de la comunidad académica en su conjunto, preservando así la objetividad en la evaluación.

Finalmente, el marco dirige su mirada a los editores, actores clave en la preservación de la integridad científica. Les insta a practicar una transparencia radical en sus algoritmos de selección editorial, haciendo públicos los criterios que determinan qué se publica y qué no. Simultáneamente, les exige erradicar el *“guest authorship”*, la inclusión de autores fantasmas por razones de conveniencia o prestigio, mediante la implementación rigurosa de criterios de autoría basados en contribuciones reales. Estas medidas buscan devolver la confianza a un sistema editorial cuya credibilidad se ha visto amenazada por prácticas opacas.

En conjunto, estas directrices representan un giro hacia una ciencia más humana y transparente. No se limitan a corregir abusos, sino que proponen una nueva filosofía de evaluación donde la calidad, la integridad y el impacto social se entrelazan para dar lugar a un ecosistema científico más robusto y, en definitiva, más fiel a su misión de servicio a la sociedad.

Caso Práctico: Cómo Implementar DORA en un Departamento

Paso 1: Capacitación en métricas responsables (4 horas).

Paso 2: Reemplazar FI por:

- Índice h normalizado por disciplina.
- Porcentaje de artículos en top 10 % de citas.

Paso 3: Introducir evaluación cualitativa mediante:

- 2 informes de pares externos.
- 1 carta de impacto social.

Un caso real:

Resultado en UC Davis (2023):

- 32 % más de mujeres promovidas.
- 25 % aumento en investigaciones interdisciplinarias.

Buenas prácticas en 4 pasos

Esta guía recomienda:

1. Auditar sesgos: Usar *Gender Citation Gap Analyzer* (tool de Python) para detectar disparidades.
2. Triangular métricas: Ej: Si un investigador tiene pocas citas pero alto impacto en políticas (medido vía *Overton*), valorar ambos.
3. Capacitar evaluadores: Talleres para evitar juicios basados en prejuicios (ej: desestimar revistas en español).
4. Exigir transparencia: Que las universidades publiquen sus criterios bibliométricos.

La meta no es abandonar los datos, sino usarlos para amplificar voces silenciadas y recompensar ciencia con sentido social.

Hacia una cultura de evaluación justa

La bibliometría ética no rechaza los números, sino su uso simplista. Como resume *Ludo Waltman* (coautor del Manifiesto de Leiden):

“Cuando las métricas se convierten en fines, corrompen la ciencia. Cuando son medios, la mejoran.”

El Capítulo 2 explorará la historia de la bibliometría, desde los índices de citas de Garfield hasta la inteligencia artificial.

Recapitulando

- La bibliometría es la disciplina que aplica métodos cuantitativos para medir la producción, difusión e impacto del conocimiento científico.
- Surge como parte de las ciencias de la información y la documentación, vinculada a la evaluación de la actividad investigadora.
- Su finalidad es evaluar y comprender la dinámica de la ciencia, no solo cuantificarla.
- Se basa en el análisis de publicaciones y citas, consideradas trazas medibles del avance del conocimiento.
- La bibliometría moderna combina estadística, informática y teoría de la comunicación científica.
- Proporciona indicadores objetivos y comparables, fundamentales para políticas científicas y gestión universitaria.
- Los principales objetos de análisis son autores, instituciones, revistas, países y

temáticas.

- Los indicadores básicos incluyen: productividad (número de publicaciones), impacto (número de citas) y colaboración (coautorías).
- Existen indicadores compuestos como el índice h, índice g, factor de impacto y SJR.
- La bibliometría no sustituye la evaluación cualitativa, sino que la complementa mediante evidencia empírica verificable.
- Su desarrollo histórico parte de los años 60 con Garfield y el Science Citation Index.
- Ha evolucionado hacia la cienciometría y la altmetría, ampliando su alcance al análisis de redes y medios digitales.
- Es una herramienta esencial para medir visibilidad, colaboración e influencia científica.
- Se aplica en ranking institucional, evaluación de revistas, análisis de tendencias y vigilancia tecnológica.
- La bibliometría es útil tanto para evaluadores (agencias, comités, universidades) como para investigadores individuales.
- Sus resultados deben interpretarse con cautela: las cifras no siempre reflejan calidad o relevancia científica.
- La interpretación correcta requiere contextualización disciplinar, temporal y lingüística.
- Entre sus limitaciones destacan los sesgos de cobertura, idioma y autocitación.
- Las métricas deben usarse de forma ética y transparente, evitando el “fetichismo del número”.
- La bibliometría, bien aplicada, fortalece la rendición de cuentas, la planificación científica y la ciencia abierta.

Preguntas de Autoevaluación

1. ¿Cómo se define la bibliometría y cuál es su principal propósito?
2. ¿Qué elementos constituyen la base empírica del análisis bibliométrico?
3. ¿Qué disciplinas confluyen en el desarrollo moderno de la bibliometría?
4. ¿Cuáles son los tres tipos principales de indicadores bibliométricos?
5. ¿Qué diferencia hay entre los indicadores simples (como número de citas) y los compuestos (como índice h)?
6. ¿Por qué la bibliometría debe complementarse con la evaluación cualitativa?
7. ¿Qué papel jugó Eugene Garfield en la historia de la bibliometría?
8. ¿Qué sesgos o limitaciones pueden afectar los resultados bibliométricos?
9. ¿En qué ámbitos institucionales se aplica la bibliometría actualmente?
10. ¿Qué principios éticos deben regir el uso de los indicadores bibliométricos?

BIBLIOGRAFÍA

1. Moed HF. Citation analysis in research evaluation. Dordrecht: Springer; 2005. doi:10.1007/1-4020-3714-7.
2. Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. J Doc. 2008;64(1):45-80. doi:10.1108/00220410810844150.
3. Thelwall M. Web indicators for research evaluation: a practical guide. San Rafael (CA): Morgan & Claypool Publishers; 2016. doi:10.2200/S00733ED1V01Y201602ICR048.
4. Leydesdorff L, Milojević S. Scientometrics. In: Wright JD, editor. International encyclopedia of the social & behavioral sciences. 2nd ed. Amsterdam: Elsevier; 2015. p. 322-7. doi:10.1016/

B978-0-08-097086-8.85017-7.

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

REFERENCIAS BIBLIOGRÁFICAS

1. Lim WM, Kumar S. Guidelines for interpreting the results of bibliometric analysis: a sensemaking approach. *Glob Bus Organ Excell*. 2024;43(2):17-26.
2. Yepes-Nuñez JJ, Urrútia G, Romero-García M, Alonso-Fernández S. Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Rev Esp Cardiol*. 2021;74(9):790-9.
3. Passas I. Bibliometric analysis: the main steps. *Encyclopedia*. 2024;4(2):1014-25.
4. Maddi A, Da Silva JAT. Beyond authorship: analyzing contributions in PLOS ONE and the challenges of appropriate attribution. *J Data Inf Sci*. 2024;9(3):88-115.
5. Kojaku S, Livan G, Masuda N. Detecting anomalous citation groups in journal networks. *Sci Rep*. 2021;11(1):1-11.
6. Diprose JP, Hosking R, Rigoni R, Roelofs A, Chien TY, Napier K, et al. A user-friendly dashboard for tracking global open access performance. *J Electron Publ*. 2023;26(1):17-61.
7. Oza A. Citations show gender bias – and the reasons are surprising. *Nature*. 2023 Dec 22.
8. Zhang X, Li J, Gu Z, Gao X. How does language learning contribute to individual growth in a multilingual world? A systematic review. *J Multiling Multicult Dev*. 2025 Sep 30.
9. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*. 2016;106(1):213-28.
10. García-Villar C, García-Santos JM. Indicadores bibliométricos para evaluar la actividad científica. *Radiología*. 2021;63(3):228-35.
11. Peralta González MJ, Maylín I, Guzmán F, Gregorio O, Li C. Criterios, clasificaciones y tendencias de los indicadores bibliométricos en la evaluación de la ciencia. *Rev Cub Inf Cienc Salud*. 2015;26(3):290-309.
12. Rousseau R, Hu X. Metrics: a fetish for high-profile journals. *Nature*. 2012;490(7420):343.
13. Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, Van Eck NJ, et al. The Leiden ranking 2011/2012: data collection, indicators, and interpretation. *J Am Soc Inf Sci Technol*. 2012;63(12):2419-32.
14. Directorate for Education and Skills [Internet]. Paris: Organisation for Economic Co-operation and Development (OECD); c2025 [cited 2025 Oct 25]. Available from: <https://www.oecd.org/en/about/directorates/directorate-for-education-and-skills.html>

15. De Solla Price DJ. Networks of scientific papers. *Science*. 1965;149(3683):510-5.
16. The American Society for Cell Biology. San Francisco Declaration on Research Assessment (DORA). 2012 Dec 16. Available from: <https://sfedora.org/>
17. Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I. Bibliometrics: the Leiden manifesto for research metrics. *Nature*. 2015;520(7548):429-31.
18. UNESCO. Recomendación sobre la ciencia abierta. París: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura; 2021.



Chapter 2 / Capítulo 2

Applied bibliometrics. From data to publication (Spanish Edition)

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch02

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

BRIEF HISTORICAL OVERVIEW / BREVE RECuento HISTÓRICO

En 1955, un joven químico llamado Eugene Garfield publicó un artículo titulado “*Citation Indexes for Science*”, donde proponía un sistema para rastrear el impacto de las investigaciones. Nadie imaginaba que esta idea revolucionaría la ciencia, dando origen a la bibliometría moderna.⁽¹⁾

2.1. La Era Analógica (1960-1990): Los Cimientos

Science Citation Index (SCI) - 1960

En 1960, el *Institute for Scientific Information* (ISI), fundado por Eugene Garfield, revolucionó la evaluación científica con el lanzamiento del *Science Citation Index* (SCI), el primer índice de citas comercializado a gran escala. Publicado originalmente en formato impreso, este sistema permitió, por primera vez, rastrear de manera sistemática cómo los artículos científicos se vinculaban entre sí a través de sus referencias bibliográficas.⁽²⁾

El SCI se basaba en un principio sencillo pero poderoso: “*las citas son conexiones de conocimiento*”. Su metodología incluía rastreo de referencias cruzadas entre 600 revistas científicas seleccionadas (principalmente de EE.UU. y Europa), indexación inversa (en lugar de solo listar autores o temas, el SCI permitía buscar qué artículos citaban a un trabajo determinado, revelando su influencia posterior), cobertura multidisciplinaria (aunque enfocado en ciencias naturales, sentó las bases para futuros índices en ciencias sociales y humanidades)

El SCI transformó la forma de medir el impacto científico, antes del SCI, la productividad se medía por número de publicaciones. El índice introdujo la idea de que un artículo podía ser influyente incluso si su autor publicaba poco (ej: el paper de Watson y Crick sobre el ADN tuvo pocas publicaciones previas, pero cambió la biología). Identificó trabajos altamente citados que definieron campos enteros (ej: el artículo de 1953 de Miller sobre el origen de la vida). A pesar de sus limitaciones, como el sesgo anglófono, el enfoque occidental y el acceso restringido a instituciones privilegiadas, su influencia perdura a través de plataformas contemporáneas como Web of Science, mientras Google Scholar y Scopus han adoptado su lógica fundamental de citas cruzadas.⁽³⁾

El SCI no solo midió la ciencia: la hizo más transparente y conectada. Su historia nos recuerda que hasta las herramientas más disruptivas deben usarse con conciencia crítica.

Dato curioso:

El SCI original ocupaba 5 metros lineales de estanterías y solo instituciones élite como Harvard podían costearlo.⁽⁴⁾

Leyes Fundamentales (Años 70-80). Los pilares matemáticos de la bibliometría

Durante las décadas de 1970 y 1980, la bibliometría consolidó su rigor científico gracias a la validación y aplicación masiva de dos leyes empíricas formuladas décadas antes: la Ley de Lotka (1926) y la Ley de Bradford (1934). Estas leyes, aunque desarrolladas inicialmente para describir patrones en la literatura científica, se convirtieron en herramientas esenciales para entender la productividad de los investigadores y la distribución del conocimiento en las revistas académicas.

Ley de Lotka: la desigualdad en la productividad científica

La Ley de Lotka, formulada por Alfred J. Lotka en 1926 pero masificada durante el auge cuantitativo de los años setenta, constituye uno de los hallazgos fundamentales sobre la distribución desigual de la productividad científica. Este principio establece que aproximadamente

el 10 % de los investigadores produce el 50 % de las publicaciones académicas, revelando un patrón de concentración del output científico que trasciende disciplinas y períodos históricos. La ley representa la aplicación específica del principio de Pareto al dominio científico, confirmando que la producción de conocimiento sigue distribuciones asimétricas donde pocos actores generan la mayor parte de los resultados investigativos.⁽⁵⁾

La evidencia empírica contemporánea valida persistentemente esta distribución desigual a través de múltiples campos del conocimiento. Esta distribución potencia el «efecto Mateo» en ciencia, donde investigadores establecidos con acceso a redes colaborativas, recursos institucionales y capital simbólico acumulado tienden a publicar con mayor frecuencia, reforzando así sus ventajas competitivas iniciales.

Las aplicaciones prácticas de la Ley de Lotka en evaluación académica son múltiples y significativas. Permite identificar investigadores altamente productivos para decisiones de contratación, promoción o asignación de financiamiento, proporcionando un punto de referencia cuantitativo para evaluar desempeños excepcionales. Simultáneamente, sirve como herramienta diagnóstica para detectar desigualdades estructurales en sistemas científicos específicos, informando políticas destinadas a democratizar las oportunidades de publicación y visibilidad académica.

Sin embargo, la ley presenta limitaciones metodológicas importantes en contextos contemporáneos. Su aplicabilidad resulta más robusta en disciplinas STEM que en humanidades, donde predominan la autoría única y formatos de publicación diversos que escapan a las métricas tradicionales. Tampoco anticipó la explosión de coautorías masivas característica de la ciencia de grandes equipos, donde publicaciones en genómica o física experimental pueden incluir miles de autores, redistribuyendo las dinámicas de productividad individual. Estas limitaciones subrayan la necesidad de contextualizar la ley dentro de los cambios paradigmáticos en las prácticas de autoría y colaboración científica del siglo XXI.

Ley de Bradford: el núcleo de las revistas clave

La Ley de Bradford (1943) representa un principio fundamental en bibliometría que describe la distribución desigual de literatura científica relevante entre las revistas académicas. Establece que un reducido núcleo de publicaciones periódicas concentra la mayoría de los artículos significativos en cualquier campo del conocimiento, mientras una periferia más extensa alberga contribuciones dispersas. Este patrón de concentración refleja cómo se estructura la comunicación especializada en la ciencia contemporánea, donde ciertas revistas funcionan como canales privilegiados para la diseminación de investigación de alto impacto.
⁽⁶⁾ La evidencia empírica confirma consistentemente este fenómeno de concentración entre múltiples disciplinas.

Las aplicaciones prácticas de esta ley son particularmente valiosas para la gestión de recursos de información. Las bibliotecas académicas utilizan análisis de zonas de Bradford para optimizar suscripciones, focalizando recursos limitados en el núcleo de revistas que maximiza el acceso a literatura relevante. Simultáneamente, la identificación de “zonas desérticas”, campos donde el conocimiento está ampliamente disperso entre numerosas fuentes, alerta sobre la necesidad de estrategias de búsqueda más comprehensivas, especialmente en áreas interdisciplinarias o estudios culturales específicos.

Sin embargo, la Ley de Bradford enfrenta limitaciones significativas en el ecosistema científico contemporáneo. Presenta un marcado sesgo anglófono, la mayoría de los artículos

publicados en las grandes bases de datos son en inglés, marginando contribuciones valiosas en otros idiomas. Además, la emergencia de la ciencia abierta y los repositorios de *preprints* está transformando sustancialmente estos patrones de concentración. En inteligencia artificial, por ejemplo, arXiv ha afectado el modelo tradicional al crear un canal alternativo de diseminación que compite con las revistas establecidas, demostrando la creciente fluidez en los patrones de comunicación científica.

Esta evolución hacia modelos más distribuidos sugiere que, aunque el principio de concentración sigue siendo relevante, sus manifestaciones concretas se están transformando rápidamente. La bibliometría contemporánea debe desarrollar herramientas capaces de capturar estas nuevas dinámicas mientras mantiene la utilidad analítica del concepto original de Bradford para comprender la estructura de la comunicación científica.

Limitaciones generales de estas leyes

Aunque revolucionarias, ambas leyes tienen problemas en la era digital:

1. Cobertura sesgada: Se basaron en datos de EE.UU. y Europa, ignorando la producción científica de Asia, África y América Latina. Ejemplo: La Ley de Bradford no predice bien el núcleo de revistas en agronomía africana, donde la investigación se publica en revistas locales.
2. Procesamiento manual: En los 70-80, los conteos se hacían a mano, lo que generaba errores (ej: duplicación de autores con nombres similares).
3. Contexto actual: Hoy, algoritmos como los de Scopus o Dimensions permiten análisis más dinámicos, pero las leyes siguen siendo la base teórica.

¿Por qué siguen importando en el siglo XXI?

Estas leyes bibliométricas históricas mantienen una sorprendente relevancia en el ecosistema científico digital del siglo XXI, demostrando que los patrones fundamentales de la comunicación académica trascienden los cambios tecnológicos. La Ley de Lotka continúa informando el desarrollo de modernos sistemas de evaluación, donde el Índice h y sus variantes incorporan su comprensión sobre la distribución desigual de la productividad científica. Esta perspectiva contextualiza métricas individuales dentro de patrones sistémicos más amplios, previniendo interpretaciones simplistas del desempeño investigador y reconociendo la naturaleza inherentemente asimétrica de la producción de conocimiento.

En el ámbito de las políticas científicas, la Ley de Bradford proporciona un marco analítico crucial para navegar la expansión exponencial de la comunicación académica. Su principio de concentración ayuda a identificar revistas depredadoras que operan fuera de los núcleos legítimos de cada disciplina, ofreciendo un criterio objetivo para distinguir canales de comunicación confiables. Esta aplicación resulta particularmente valiosa en contextos de acceso abierto, donde la proliferación de editoriales cuestionables requiere mecanismos robustos para garantizar la integridad de la comunicación científica.

La influencia más profunda de estas leyes se manifiesta en los fundamentos algorítmicos que sustentan las herramientas digitales contemporáneas. El algoritmo PageRank de Google Scholar, por ejemplo, implementa computacionalmente el principio de Bradford al asignar mayor peso a las citas provenientes de fuentes consideradas “núcleos” de autoridad académica. Simultáneamente, los sistemas de recomendación y descubrimiento de literatura científica incorporan entendimiento derivado de Lotka para priorizar contenido de investigadores altamente productivos e influyentes.

La vigencia de estos principios demuestra que, aunque las tecnologías de comunicación científica han experimentado transformaciones radicales, los patrones subyacentes de producción y diseminación del conocimiento mantienen regularidades estructurales observables. Esta continuidad convierte a las leyes bibliométricas clásicas en herramientas esenciales para comprender tanto las permanencias como las transformaciones en la dinámica científica contemporánea, proporcionando un marco interpretativo para navegar la complejidad del actual ecosistema de investigación.

2.2. La Revolución Digital (1990-2010): expansión y críticas

La Era en que la Bibliometría se Hizo Global (y Polémica)

La década de 1990 marcó el inicio de la bibliometría digital, transformando radicalmente cómo se mide y gestiona la ciencia. Con la llegada de Internet, los índices de citas dejaron atrás los formatos impresos y adoptaron plataformas en línea, ampliando su alcance, pero también generando nuevos desafíos éticos y metodológicos.

Web of Science (WoS) - 1997: El SCI se vuelve digital

Fue la versión digital del *Science Citation Index (SCI)*, lanzada por el *Institute for Scientific Information (ISI)*. Permitió por primera vez búsquedas en tiempo real y análisis avanzados de citas. Pasó de 600 revistas (en papel) a 8 000 revistas indexadas, incluyendo ciencias sociales (*Social Sciences Citation Index*) y artes (*Arts & Humanities Citation Index*). Introdujo funciones como *Citation Reports* (para calcular el impacto de autores) y *Journal Impact Factor (JIF)* en línea (antes solo disponible en el *Journal Citation Reports* impreso).⁽⁷⁾

El JIF, creado en 1975 para evaluar revistas, comenzó a aplicarse erróneamente a investigadores individuales. Universidades exigían publicar en “revistas Q1” para contrataciones, ignorando la calidad real de los artículos.

Scopus (2004): el competidor que desafió la hegemonía

Desarrollado por Elsevier como una alternativa directa al Web of Science (WoS), Scopus amplió el horizonte de la indexación científica al incorporar una mayor cantidad de revistas no anglófonas, con un fuerte enfoque en publicaciones europeas y asiáticas. Entre sus aportes más relevantes, introdujo los perfiles de autor, sistemas pioneros en la desambiguación de nombres que anticiparon herramientas como ORCID. También incorporó métricas alternativas como el *CiteScore*, concebido como un indicador complementario al *Journal Impact Factor (JIF)*.

Sin embargo, Scopus no estuvo exento de críticas. Se le ha señalado por presentar un sesgo comercial, favoreciendo en sus rankings a publicaciones asociadas a Elsevier. Además, su negativa a indexar *preprints* lo situó en desventaja frente a sistemas más abiertos como Google Scholar, al excluir de su base de datos una parte significativa de la literatura científica informal o “gris”.

Google Scholar (2004): la disrupción democrática

El lanzamiento de Google Scholar representó una transformación profunda en el acceso al conocimiento científico. Su sistema de indexación automática permitió incorporar no solo artículos revisados por pares, sino también *preprints*, como los de arXiv, libros, tesis y documentos técnicos, muchos de los cuales habían permanecido fuera del radar de las bases tradicionales. A diferencia de WoS y Scopus, Google Scholar se ofreció de forma gratuita, lo que democratizó el acceso a la literatura científica global. Su capacidad de capturar contenidos en idiomas y formatos tradicionalmente marginados, como estudios en español de América Latina, le otorgó una cobertura sin precedentes. Además, con el desarrollo de *Google Scholar Citations*,

los investigadores pudieron crear perfiles públicos para visibilizar su producción académica y sus métricas personales.

No obstante, esta apertura también generó controversias. Uno de los principales cuestionamientos ha sido la opacidad de sus algoritmos, ya que no se informa cómo se jerarquizan los resultados, lo que ha dado lugar a estrategias para manipular la visibilidad de ciertos trabajos mediante el uso intencional de palabras clave. Por otro lado, la calidad de los documentos indexados es motivo de debate, dado que Google Scholar incluye revistas depredadoras y materiales sin revisión por pares.

Impacto y críticas de esta revolución

La irrupción de plataformas como Scopus y Google Scholar marcó un antes y un después en la forma en que se accede, evalúa y difunde el conocimiento científico. Entre sus principales legados se encuentra la globalización de la ciencia, al otorgar visibilidad a investigaciones provenientes de regiones históricamente marginadas, como América Latina, Asia o África, más allá de los centros tradicionales de producción académica en Estados Unidos y Europa. Además, la digitalización masiva de citas permitió el desarrollo de nuevas métricas, siendo el índice *h*, propuesto por Hirsch en 2005, una de las más influyentes al medir simultáneamente la productividad y el impacto de un investigador.

No obstante, persisten problemas estructurales que estas plataformas no han logrado resolver. A pesar de las críticas acumuladas, el factor de impacto (JIF) continúa siendo utilizado como criterio decisivo en procesos de evaluación y financiamiento, como sucede en agencias nacionales como la ANID en Chile.⁽⁸⁾ Esta persistencia refuerza una lógica cuantitativa que muchas veces ignora la calidad o el contexto del conocimiento producido. Por otro lado, la brecha digital sigue siendo un obstáculo importante: muchas instituciones en países de ingresos bajos o medianos no pueden costear el acceso a bases de datos comerciales como WoS o Scopus, lo que las obliga a depender de alternativas gratuitas como Google Scholar, con todas las limitaciones y riesgos que esto implica en términos de confiabilidad. Finalmente, el auge de las herramientas digitales ha dado lugar a nuevas formas de manipulación de métricas, como el *citation stacking*, una práctica que consiste en crear círculos de citas fraudulentas para inflar artificialmente el impacto de ciertas publicaciones, distorsionando así los procesos de evaluación académica.

2.3 La Era Moderna (2010-Presente): complejidad y democratización

Altmetría(2010)

A partir de 2010, la bibliometría comenzó a expandirse hacia formas más amplias de evaluación del impacto científico, dando origen a la altmetría. Este enfoque busca medir el impacto más allá del conteo de citas académicas, incorporando indicadores como menciones en redes sociales, medios de comunicación, blogs, Wikipedia y documentos de política pública. El lanzamiento de Altmetric.com en 2011 marcó un punto de inflexión, al ofrecer herramientas capaces de rastrear en tiempo real la circulación digital de artículos científicos.

Bibliometría Open Source(2017)

En 2017, la democratización de la bibliometría dio un salto significativo con la aparición de herramientas de código abierto, que facilitaron el acceso a análisis avanzados sin los altos costos asociados a plataformas comerciales. Entre las más destacadas se encuentran Bibliometrix, un paquete desarrollado en R que permite realizar análisis bibliométricos sofisticados, y VOSviewer, una herramienta ampliamente utilizada para la visualización de redes de coautoría, co-citación o términos clave. Estas plataformas permitieron que investigadores de instituciones con recursos limitados pudieran acceder a metodologías de análisis robustas, contribuyendo a

una descentralización del conocimiento bibliométrico.

Inteligencia Artificial y Minería de Textos(2020)

En la década de 2020, la integración de la inteligencia artificial revolucionó la manera en que se procesa y analiza la literatura científica. Modelos como GPT-4 abrieron la posibilidad de resumir tendencias a partir de millones de *abstracts* de forma automatizada y eficiente, permitiendo un análisis sintético de campos de investigación completos en tiempo récord. Paralelamente, plataformas como Dimensions.ai comenzaron a integrar bases de datos que conectan publicaciones científicas con patentes, subvenciones y resultados de investigación, ofreciendo una perspectiva mucho más completa del ciclo de producción y transferencia del conocimiento. Esta convergencia entre IA, minería de textos y ciencia abierta ha ampliado el alcance y la profundidad de la bibliometría contemporánea, posicionándola como una herramienta clave en la toma de decisiones científicas y de política pública.

Reto actual:

Cómo evitar que la IA genere “artículos zombis” (textos bien citados, pero sin sustento real).

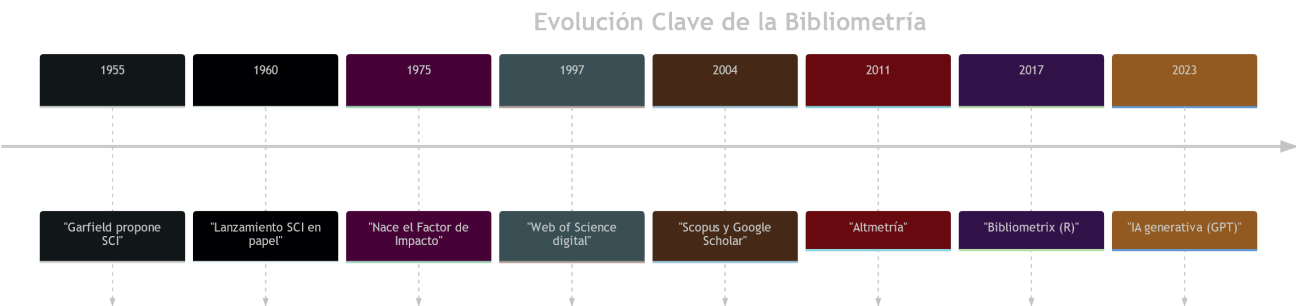


Figura 2.1. Línea de tiempo visual

2.4 Lecciones Históricas

1. De lo analógico a lo digital: Los saltos tecnológicos han democratizado el acceso, pero introducido nuevos sesgos (ej: sobrecarga informativa).
2. De lo cuantitativo a lo cualitativo: Las críticas al FI llevaron a métricas responsables (DORA, Leiden).
3. De lo académico a lo social: La altmetría amplió la noción de impacto científico.

“La bibliometría ya no es solo contar citas, sino entender cómo el conocimiento fluye —y a veces se estanca— en la sociedad.”

La Sección II explorará las herramientas más usadas para la construcción de bibliometrías así como obtener los datos para crear los estudios bibliométricos, de una manera práctica.

Recapitulando

- La bibliometría tiene sus orígenes en la primera mitad del siglo XX, cuando surgieron los primeros intentos de cuantificar la producción científica.
- El término “bibliometría” fue acuñado por Paul Otlet (1934) y consolidado por Alan Pritchard (1969).

- Su desarrollo estuvo influenciado por los avances en documentación, estadística e informática.
- En los años 1920-1940 se publicaron los primeros estudios sobre productividad de autores y distribución de artículos.
- Lotka (1926) formuló la Ley de Productividad Científica, que describe la frecuencia desigual con que los investigadores publican.
- Bradford (1934) propuso su Ley de Dispersión, que explica cómo los artículos relevantes se concentran en un número reducido de revistas núcleo.
- Zipf (1949) introdujo la Ley de Frecuencia de Palabras, base para el análisis de términos y coocurrencias.
- En la década de 1950-1960, Eugene Garfield fundó el Institute for Scientific Information (ISI) y desarrolló el Science Citation Index (SCI).
- El SCI revolucionó la medición de la ciencia al permitir rastrear redes de citación entre publicaciones.
- En los años 70-80, el campo se institucionalizó con la aparición de la revista *Scientometrics* y los primeros congresos internacionales.
- En esta etapa, la bibliometría se amplió hacia la cienciometría, enfocada en la dinámica y política de la ciencia.
- A partir de los años 90, el uso de bases de datos electrónicas y software especializado permitió análisis más amplios y precisos.
- El surgimiento de Internet y los motores de búsqueda transformó radicalmente el acceso y la recopilación de información científica.
- Con la llegada de Google Scholar (2004) y Scopus (2004), se diversificaron las fuentes bibliométricas y los indicadores de impacto.
- Desde 2010, emergen la altmetría y la webmetría, orientadas al impacto social y digital de la ciencia.
- En esta etapa moderna, la bibliometría se integra con la inteligencia artificial, el big data y la ciencia abierta.
- Los métodos actuales permiten mapear redes de colaboración, tendencias temáticas y estructuras cognitivas.
- El desarrollo histórico refleja una transición desde un enfoque cuantitativo simple a una visión multidimensional y ética de la evaluación científica.
- La bibliometría contemporánea combina la tradición de Garfield, Lotka y Bradford con herramientas digitales avanzadas.
- En síntesis, su evolución ha convertido a la bibliometría en un pilar esencial de la investigación, la gestión del conocimiento y la política científica.

Preguntas de autoevaluación

1. ¿Quién acuñó el término “bibliometría” y en qué año se consolidó su uso?
2. ¿Qué describe la Ley de Lotka en relación con la productividad científica?
3. ¿Cuál es el principio central de la Ley de Bradford?
4. ¿Qué aportó Zipf al desarrollo de la bibliometría?
5. ¿Qué importancia tuvo Eugene Garfield en la historia de la disciplina?
6. ¿Qué papel desempeñó la creación del Science Citation Index en los años 60?
7. ¿Por qué la década de los 70-80 marcó la institucionalización de la bibliometría?
8. ¿Cómo influyeron las bases de datos electrónicas e Internet en la evolución del campo?
9. ¿Qué aportes introdujo la altmetría a partir de 2010?
10. ¿Qué caracteriza a la bibliometría contemporánea respecto a sus primeras etapas?

BIBLIOGRAFÍA

1. 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>
2. Glänzel W. Bibliometrics as a research field: A course on theory and application of bibliometric indicators. Leuven: KU Leuven; 2003. https://www.kuleuven.be/metaforum/docs/pdf/lecture_glanzel_bibliometrics.pdf
3. Waltman L. A review of the literature on citation impact indicators. J Informetrics. 2016;10(2):365-91. <https://doi.org/10.1016/j.joi.2016.02.007>
4. Cronin B, Sugimoto CR, editors. Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact. Cambridge (MA): MIT Press; 2014. <https://doi.org/10.7551/mitpress/9780262026792.001.0001>
5. Hood WW, Wilson CS. The literature of bibliometrics, scientometrics, and informetrics. Scientometrics. 2001;52(2):291-314. <https://doi.org/10.1023/A:1017919924342>

REFERENCIAS BIBLIOGRÁFICAS

1. Garfield E. Citation indexes for science. Science. 1955;122(3159):108-11.
2. Garfield E. "Science Citation Index"—A new dimension in indexing. Science. 1964;144(3619):649-54.
3. Martín-Martín A, Thelwall M, Orduna-Malea E, Delgado López-Cózar E. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics. 2021;126(1):871-906. <https://doi.org/10.1007/s11192-020-03690-4>
4. Garfield E. The evolution of the science citation index. International Microbiology. 2007;10(1):65-9.
5. Lotka AJ. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences. 1926;16(12):317-23.
6. Bradford SC. Classic paper: Sources of information on specific subjects. Collection Management. 1976;1(3-4):95-103. https://doi.org/10.1300/J105v01n03_06
7. Patsopoulos NA, Analatos AA, Ioannidis JPA. Relative citation impact of various study designs in the health sciences. JAMA. 2005;293(19):2362-6. <https://doi.org/10.1001/jama.293.19.2362>
8. Agencia Nacional de Investigación y Desarrollo (ANID). Concurso de Proyectos Fondecyt Regular 2024. 2024. <https://anid.cl/concursos/concurso-de-proyectos-fondecyt-regular-2024/>

Part II / Parte II

DATA PREPARATION

PREPARACIÓN DE DATOS



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 sí 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*.



```
[1] # Restart the session afther this cell to avoid Google Colab errors
!pip install --upgrade --force-reinstall numpy==1.26.4 pandas

[1] !pip install pybibx
!pip install tabulate

[1] # Dowload .bib file
!wget https://github.com/Valdecy/pyBibX/raw/main/assets/bibs/scopus.bib

[1] # Required Libraries
import numpy as np
import pandas as pd
import textwrap

from google.colab import data_table
from tabulate import tabulate
from prettytable import PrettyTable
from pybibx.base import pbx_probe
```

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.

Harzing's Publish or Perish (Windows GUI Edition) 8.17.4863.9118

File Edit Search View Help

My searches
Trash

Search terms	Source	Papers	Cites	Cites/y...	h	g	hl, no...	hl, ar
db.ris [2025-09-25 20:09:30]	RIS/RefMa...	1103	395	7.90	5	7	4	
db.ris [2025-09-18 20:22:37]	RIS/RefMa...	51	84	1.65	5	8	3	
export_9c8d6324-1bcc-438d-a1...	RIS/RefMa...	343	4108	77.51	23	61	12	
Dolor posoperatorio.ris [2025-0...	RIS/RefMa...	91	326	11.24	10	16	5	
la 2024.ris [2025-08-23 02:19:00]	RIS/RefMa...	9204	35961	35961.00	47	66	17	17
Cardiologia.ris [2025-08-23 01:15...	RIS/RefMa...	2559	24164	401.10	75	140	26	

Imported external data

Display title: db.ris [2025-09-25 20:09:30] Apply

Original format: RIS/RefManager Revert

New

Cites	Per year	Rank	Authors	Title	Year	Publication
<input checked="" type="checkbox"/>	0	0.00	1	O.M.C. Urbano, M...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	2	Y.P. DÁ-az, R.R. R...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	3	N.R. Quintero, Y.R...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	4	A. Tavakolian, F.A...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	5	M. CÃrdova-Delg...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	6	J.A.V. Figueira, A.R...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	7	V. GuzmÃn-Ram...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	8	D.O. Oleas, M.B.M...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	9	E.G. Estrada-Araoz...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	10	N.M.B. MuÃ±oz...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	11	E.G. Estrada-Araoz	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	12	J.E. Bender-del Bu...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	13	J.L. Pacios Dorado...	2025	Revista del Hospi
<input checked="" type="checkbox"/>	0	0.00	14	V. Moriano-De-Lo...	2025	Revista del Hospi

Citation metrics

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

hl, norm: 4

hl, annual: 0.08

hA-index: 1

Papers with ACC >= 1,2,5,10,20: 14,0,0,0,0

Copy Results

Save Results

Paper details

Select a paper in the results list (to the left of this pane) to see its details here.

Copy Paper Details

Tools

Preferences...

Online User's Manual

Frequently Asked Questions

Training Resources

YouTube Channel

Become a PoP Supporter

Writing effective promotion applications

Guiding your career in academia

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.

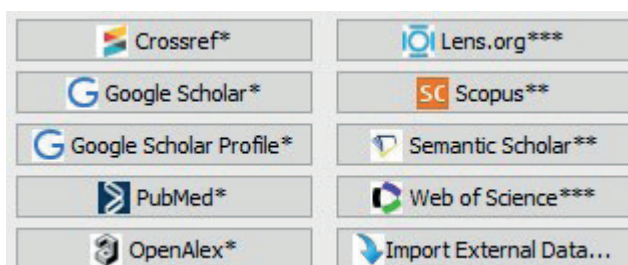


Figura 3.4. Bloque de importación de los datos de Publish or Perish.

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.

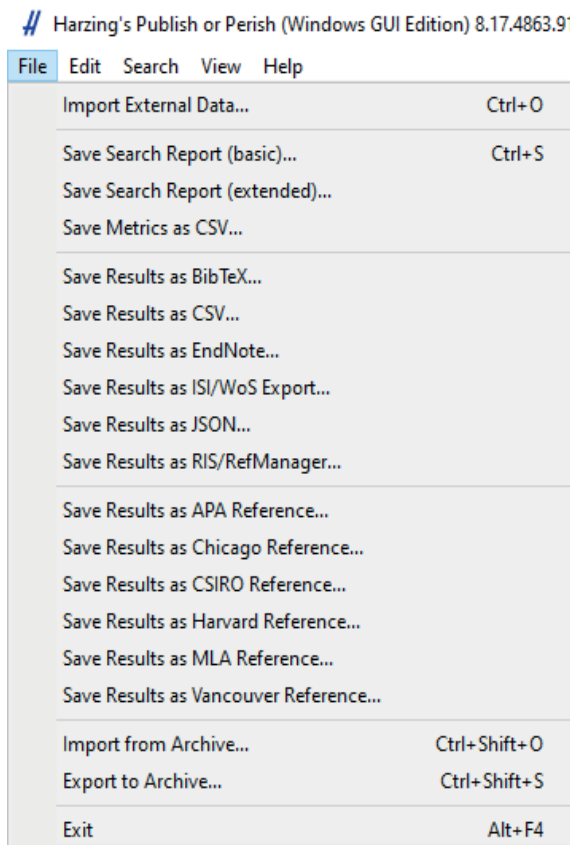


Figura 3.6. Menú de exportación de *Publish or Perish*

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 capítulo 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.

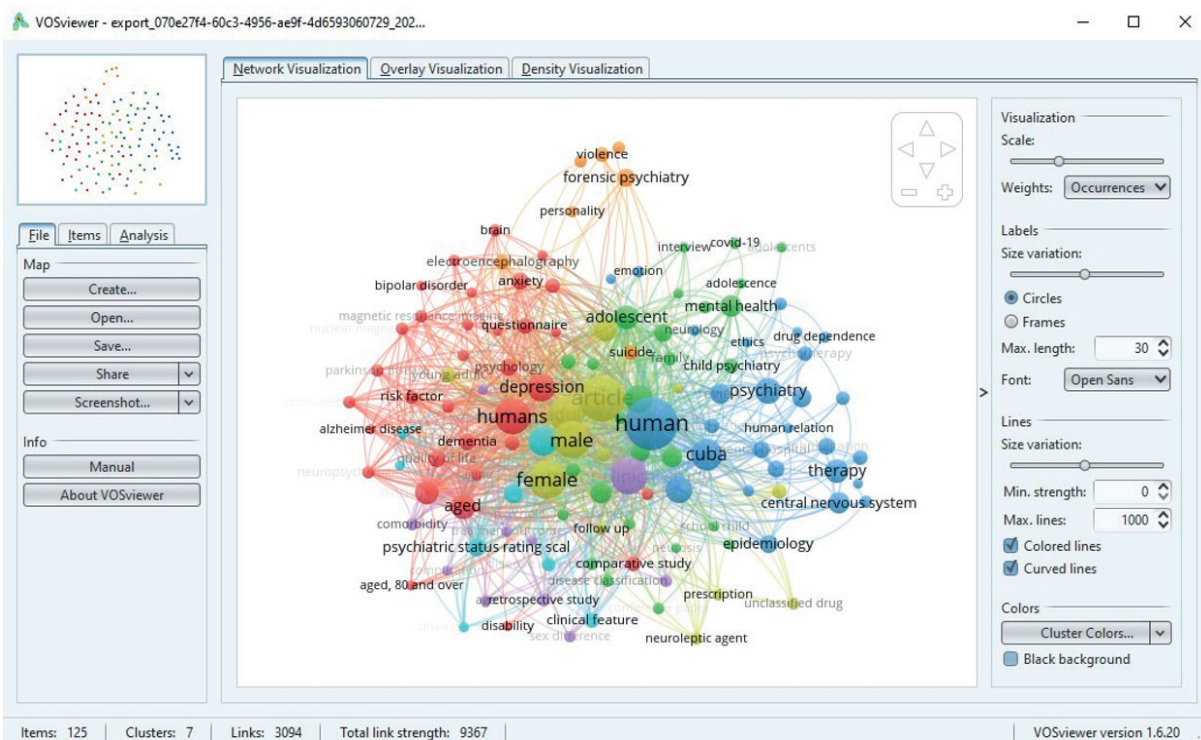


Figura 3.7. Vista principal de VOSviewer, con proyecto graficado

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.

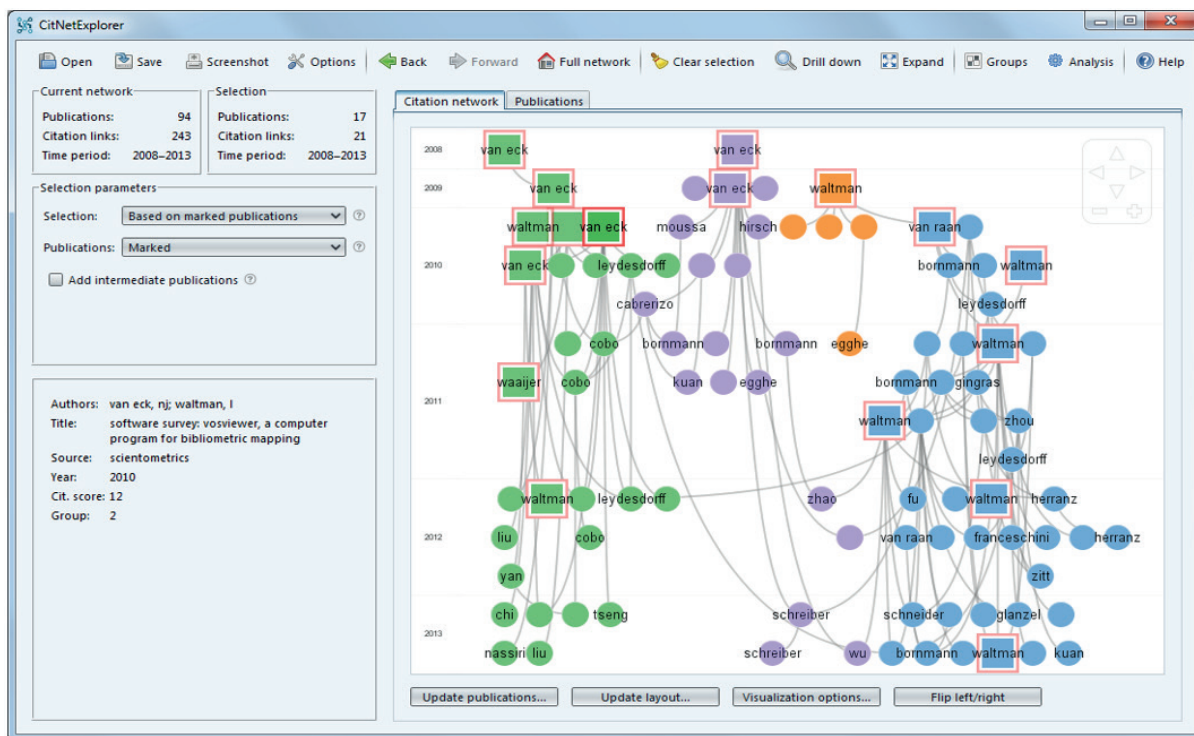
3.5.2. CitNetExplorer: análisis de redes de citación

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>)

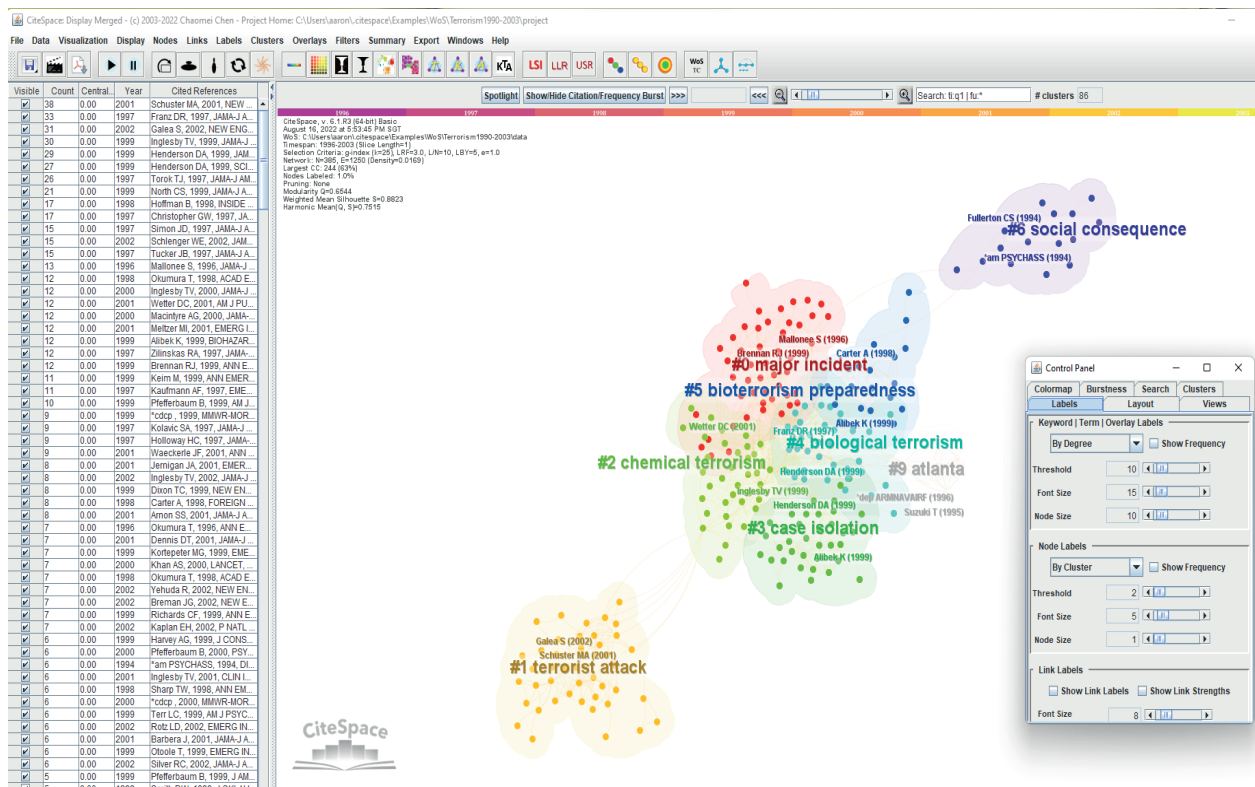
Figura 3.8. Vista principal de CitNetExplorer, con proyecto graficado

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.

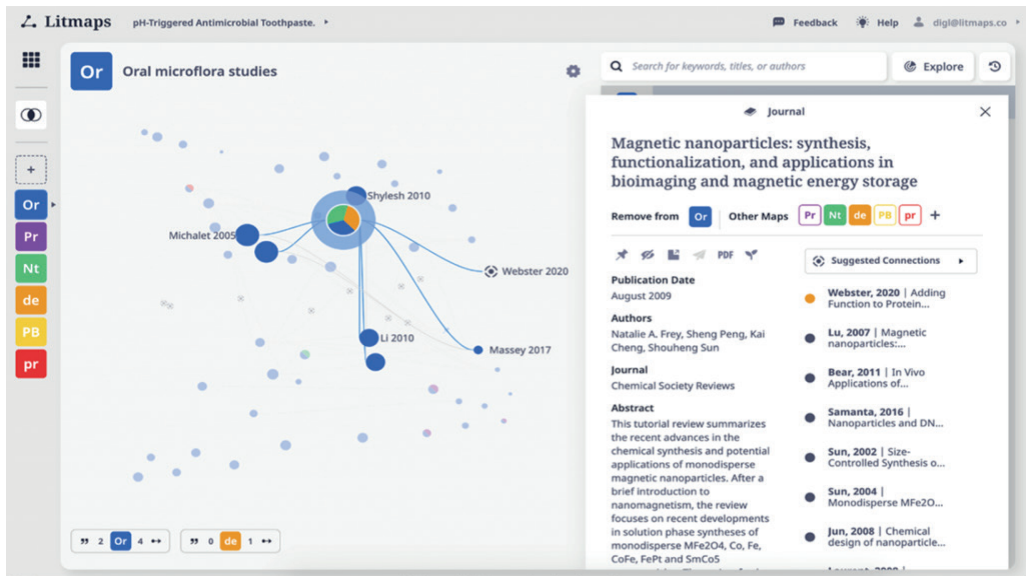
3.5.4. Herramientas online (Litmaps, ResearchRabbit)

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 literaria 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.

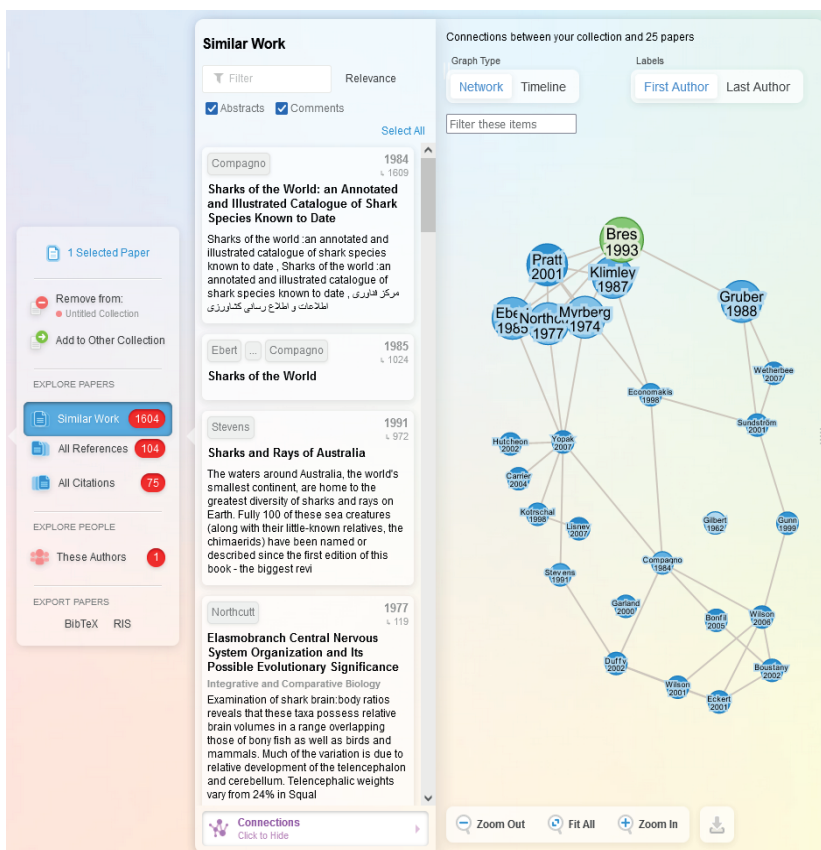
Litmaps (<https://www.litmaps.com/>) se destaca por su capacidad para generar mapas de citación interactivos que revelan las relaciones entre publicaciones de manera dinámica. La plataforma permite visualizar tanto las referencias citadas como las citas posteriores de un artículo seminal, creando redes 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.



Chapter 4 / Capítulo 4

Applied bibliometrics. From data to publication (Spanish Edition)

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch04

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

DATA COLLECTION / OBTENCIÓN DE DATOS

Del Caos al Conocimiento Organizado

4.1. Estrategias de búsqueda para estudios bibliométricos

En el ámbito de la bibliometría, la calidad de los resultados depende fundamentalmente de una estrategia de búsqueda rigurosa y bien planificada. Una aproximación deficiente en esta fase inicial puede comprometer todo el estudio, generando sesgos significativos, omitiendo literatura fundamental o incorporando material irrelevante que distorsiona los análisis. Las consecuencias de un diseño inadecuado son particularmente graves en este campo, donde la integridad de los hallazgos está directamente vinculada a la exhaustividad y precisión del corpus documental recuperado.

Este capítulo ofrece un marco metodológico sistemático para desarrollar cadenas de búsqueda óptimas en las principales plataformas bibliográficas (Web of Science, Scopus, PubMed, entre otras). El enfoque presentado combina principios teóricos con aplicaciones prácticas:

- Técnicas avanzadas para la formulación de ecuaciones de búsqueda balanceadas
- Estrategias de validación para asegurar cobertura completa sin sacrificar precisión
- Ejemplos concretos extraídos de diversas disciplinas científicas
- Errores frecuentes y cómo evitarlos, basados en análisis de casos reales

El método propuesto enfatiza la necesidad de adaptar cada búsqueda a las particularidades del sistema documental de cada campo de estudio, considerando factores como: la terminología disciplinar específica, las características de indización de cada base de datos, y los sesgos inherentes a los diferentes sistemas de recuperación de información.

Mediante este enfoque estructurado, los investigadores podrán construir conjuntos documentales robustos que sirvan como base fiable para análisis bibliométricos avanzados, desde estudios de mapeo científico hasta evaluaciones de impacto. La metodología presentada ha sido validada empíricamente en múltiples proyectos de investigación, demostrando su eficacia para producir resultados representativos y replicables en diversas áreas del conocimiento.

4.1.1. Diseño de la estrategia de búsqueda

Definir el alcance: la base de una búsqueda bibliométrica efectiva

El primer paso crítico en cualquier estudio bibliométrico es establecer con precisión los límites y objetivos de la investigación. Este proceso de delimitación requiere responder preguntas fundamentales que determinarán todo el diseño metodológico posterior.

La pregunta de investigación central debe formularse con claridad, ya que definirá el tipo de análisis a realizar. ¿El objetivo es identificar tendencias temporales en un campo? ¿Evaluar el impacto de ciertas instituciones? ¿Mapear redes de colaboración internacional? Cada uno de estos propósitos exigirá estrategias de búsqueda diferentes.

Los criterios de inclusión constituyen los filtros esenciales para garantizar relevancia y manejabilidad de los resultados:

- El período temporal debe justificarse según la evolución del campo (ej: en áreas emergentes, los últimos 5 años pueden ser suficientes; en disciplinas consolidadas, puede necesitarse una década o más).
- Los tipos documentales seleccionados (artículos, revisiones, patentes, tesis) deben alinearse con las prácticas de comunicación de cada disciplina.

- La decisión sobre idiomas implica un equilibrio entre exhaustividad (incluir contribuciones en múltiples lenguas) y factibilidad (el predominio del inglés en muchas áreas científicas).

Un ejemplo aplicado, por ejemplo, para estudiar “*La producción científica sobre inteligencia artificial aplicada a medicina entre 2015-2025*”, una delimitación adecuada podría especificar:

- Solo artículos completos en revistas con revisión por pares
- Documentos indexados en las principales bases de datos (WoS, Scopus)
- Publicaciones en inglés (dado su predominio en estas disciplinas técnicas)
- Exclusión de patentes y literatura gris para mantener el foco en investigación básica

Esta fase de planificación, aunque a menudo subestimada, es determinante. Una gran parte de los errores en bibliometrías se originan en una definición imprecisa del alcance inicial. Una delimitación clara evita posteriormente la inclusión de ruido documental o la omisión de literatura clave, problemas que pueden comprometer la validez de los hallazgos.

Seleccionar bases de datos

Teniendo en cuenta el alcance planificado se debe seleccionar la base de datos sobre la cual se hará la búsqueda, por lo que son necesarios conocimientos básicos sobre el contenido y bondades de cada base de datos. Cada base tiene ventajas y sesgos:

Tabla 4.1. Características de las principales bases de datos científicas			
Base de Datos	Cobertura	Filtros Útiles	Limitaciones
Scopus	+25,000 revistas, fuerte en STEM	Análisis de citas integrado	Costoso, débil en humanidades
Web of Science	+20,000 revistas, núcleo histórico	Índice h incorporado	Sesgo hacia EE.UU./ Europa
PubMed	+30M registros en biomedicina	Filtros por tipo de estudio (ensayos clínicos)	Sin datos de citas completos
Google Scholar	Libre, incluye libros y grises	Búsqueda semántica	Sin herramientas de análisis
Dimensions	Integra publicaciones y patentes	Datos de financiamiento	Cobertura desigual por región

Es oportuno explicar que, para evitar los sesgos/limitaciones de las bases de datos, pueden emplearse varias en un mismo estudio (ej: Scopus + PubMed).

4.1.2. Construcción de cadenas de búsqueda

Una vez definido con claridad cuál es el tema sobre el cual se hará la búsqueda se deben seleccionar las palabras más representativas de la investigación.

Palabras clave y operadores booleanos

La efectividad de una búsqueda de información para una investigación bibliométrica depende, en gran medida, de la correcta selección de los términos clave y su combinación mediante operadores lógicos. El operador AND restringe los resultados, conectando conceptos que deben aparecer simultáneamente (ejemplo: «*machine learning*» AND «*healthcare*» recupera solo documentos que mencionen ambos temas). Por el contrario, OR amplía el alcance al incluir sinónimos o variantes conceptuales (ejemplo: «*deep learning*» OR «*neural networks*»). Para excluir temas no relevantes, se emplea NOT, aunque debe usarse con precaución para

no eliminar literatura útil (ejemplo: «cancer» NOT «breast cancer» podría omitir estudios transversales importantes).

Las comillas garantizan búsquedas exactas de frases completas, evitando resultados dispersos («climate change» es más preciso que climate AND change). Los comodines capturan variantes morfológicas de una raíz común (ejemplo: educat incluye «education», «educative», «educational»).

Ejemplo de cadena optimizada:

(«artificial intelligence» OR «AI») AND («medical diagnosis» OR «clinical decision») NOT «robotics»

Esta consulta recupera documentos sobre IA aplicada a diagnóstico médico, excluyendo aquellos centrados en robótica.

Los operadores anteriores son comunes para la mayoría de bases de datos y buscadores, pero hay algunas bases de datos que incorporan nuevos en aras de ganar en especificidad de la búsqueda. De todas formas la lógica que siguen estos operadores es la misma independientemente de donde se implementen.

Filtros avanzados para precisión

Muchas veces solo con lo anterior no basta para recuperar el contenido requerido o trazado, o abarca mucho más de lo que se planeó. Las bases de datos permiten afinar búsquedas mediante filtros específicos:

- Por campo: En Scopus, **TITLE-ABS-KEY ()** busca en título, resumen y palabras clave; en PubMed, **TI/AB** cubre título y resumen. Limitar la búsqueda solo al título (**TITLE()**) subestima la literatura relevante, ya que muchos conceptos clave pueden aparecer únicamente en el resumen.
- Por impacto: Filtros como **CITES > 50 (WoS)** identifican trabajos altamente citados, útiles para estudios de influencia científica.
- Por institución: **AFFIL("Harvard University")** localiza la producción de una universidad específica, esencial para análisis de colaboraciones.

Error frecuente y solución:

Problema: Usar solo **TITLE()** puede perder documentos relevantes

Solución: Combinar siempre título, resumen y palabras clave (**TITLE-ABS-KEY()** en Scopus o **TI/AB/KW** en WoS).

4.1.3. Técnicas específicas por base de datos

En Scopus, la sintaxis avanzada permite filtrar con precisión mediante operadores como **TITLE-ABS-KEY()**, que abarca título, resumen y palabras clave simultáneamente. Por ejemplo, una consulta como **TITLE-ABS-KEY(«blockchain» AND «supply chain») AND PUBYEAR > 2015** recupera exclusivamente publicaciones recientes sobre aplicaciones de blockchain en cadena de suministro.

Web of Science ofrece funcionalidades únicas para el análisis de redes institucionales. Su sintaxis **TS=()** para términos de búsqueda combinada con **AD=()** para afiliaciones permite mapear colaboraciones entre centros de investigación específicos. Una consulta como **TS=(«quantum computing») AND AD=(«MIT» OR «Stanford»)** identifica la producción conjunta de estas universidades en computación cuántica.

Para investigaciones biomédicas, PubMed se distingue por su sistema de vocabulario controlado MeSH (Medical Subject Headings). Este tesoro jerárquico resuelve problemas de sinonimia y ambigüedad terminológica. Una búsqueda como «COVID-19»[Mesh] AND «Vaccines»[Mesh] AND «clinical trials»[Publication Type] garantiza cobertura completa sobre ensayos clínicos de vacunas contra COVID-19, evitando los sesgos de las búsquedas por texto libre. La precisión de PubMed lo hace indispensable para revisiones sistemáticas en ciencias de la salud.

Cada plataforma exige estrategias específicas para optimizar precisión y recall. Scopus y WoS priorizan control terminológico y capacidades analíticas integradas, mientras PubMed ofrece vocabulario estandarizado y Google Scholar amplitud documental a costa de transparencia metodológica. La elección entre ellas debe basarse en los objetivos del estudio: análisis de citación y colaboración (WoS/Scopus), revisiones clínicas rigurosas (PubMed) o identificación de literatura no convencional (Google Scholar). Un diseño de investigación robusto frecuentemente combina múltiples bases para compensar sus respectivas limitaciones.

Recientemente, herramientas como Dimensions han emergido como alternativas prometedoras, integrando publicaciones, patentes y datasets en un solo entorno con API abierta. Su sintaxis flexible (ej: search publications for «CRISPR» where year > 2020) y cobertura multidisciplinar están redefiniendo los estándares para búsquedas bibliométricas complejas. Sin embargo, la adopción generalizada de estos nuevos sistemas aún enfrenta desafíos de interoperabilidad con las métricas tradicionales.

En resumen, los operadores y filtros específicos por bases de datos son:

Scopus

Operadores booleanos y de búsqueda avanzada:

- AND, OR, AND NOT → Combinan términos.
- → Truncamiento (ej.: comput busca “computer”, “computing”, etc.).
- “ “ → Búsqueda exacta (ej.: “machine learning”).
- W/n → Palabras dentro de n términos de distancia (ej.: AI W/3 healthcare).
- PRE/n → Primera palabra antes de la segunda, con n términos de distancia (ej.: digital PRE/2 transformation).

Filtros avanzados:

- Año: PUBYEAR > 2020
- Tipo de documento: DOCTYPE(ar) (artículos) o DOCTYPE(re) (revisiones).
- Campo específico:
 - TITLE-ABS-KEY(“blockchain”) → Busca en título, resumen o palabras clave.
 - AFFIL(“Harvard University”) → Por afiliación.
- Índices de citas: REFERENCES() (ej.: REFERENCES(123456789)).

Ejemplo combinado: TITLE-ABS-KEY(“quantum computing”) AND PUBYEAR > 2018 AND AFFIL(“MIT”) AND DOCTYPE(ar)

Web of Science (WoS)

Operadores y sintaxis:

- AND, OR, NOT → Operadores estándar.
- → Truncamiento (ej.: bio para “biology”, “biotechnology”).
- “ “ → Frase exacta.

- NEAR/x → Términos cercanos (ej.: climate NEAR/2 change).
- SAME → Términos en el mismo campo (útil para direcciones: UNIV COLORADO SAME BOULDER).

Filtros avanzados:

- Categorías WC: WC=(“Computer Science”).
- Año: PY=2020-2023.
- Tipo de documento: DT=(“Article”).
- Búsqueda por campo:
 - TS=(“neural network”) → Tema (título, resumen, palabras clave).
 - AU=(“Smith J”) → Autor.
 - OG=(“Stanford University”) → Organización.

Ejemplo combinado: TS=(“CRISPR” AND “gene editing”) AND PY=(2020-2023) AND WC=(“Biotechnology”) AND DT=(“Review”)

PubMed

Operadores y sintaxis:

- AND, OR, NOT → Operadores booleanos.
- → Truncamiento (ej.: inmun para “immune”, “immunity”).
- “ “ → Frase exacta.
- [Field] → Búsqueda por campo (ej.: «COVID-19»[Title]).

Filtros avanzados (usando etiquetas):

- Tipo de artículo: “review”[Publication Type].
- Fecha: “2020/01/01”[Date - Publication] : “2023”[Date - Publication].
- Campos específicos:
 - “deep learning”[Title/Abstract].
 - “NIH”[Affiliation].
 - - “1AU”[Author] → Autor principal.
- - Filtros predefinidos:
 - “free full text”[Filter].
 - “clinical trial”[Filter].

Ejemplo combinado: “artificial intelligence”[Title/Abstract] AND “diagnosis”[Title] AND “2020”[Date - Publication] AND “systematic review”[Publication Type]

4.1.4. Validación y documentación: garantizando rigor en la búsqueda bibliométrica

La fase de validación es fundamental para asegurar que los resultados de una búsqueda bibliométrica sean exhaustivos, precisos y reproducibles. Un proceso riguroso comienza con una prueba piloto, donde se analizan críticamente los primeros 50-100 registros obtenidos. Este muestreo inicial permite identificar problemas frecuentes: términos demasiado amplios que generan ruido documental o, por el contrario, cadenas de búsqueda demasiado restrictivas que omiten literatura relevante. La iteración es clave aquí, cada ajuste en los operadores booleanos o en la selección de términos debe probarse hasta alcanzar un equilibrio óptimo entre precisión y sensibilidad.

Para garantizar transparencia metodológica, el protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) ofrece un marco estandarizado, aplicable también a estudios bibliométricos. Su enfoque sistemático requiere documentar detalladamente todo

el proceso de selección: el número total de registros identificados en cada base de datos, los criterios de exclusión aplicados (duplicados, documentos fuera del alcance temporal o temático), y las decisiones tomadas en cada fase de filtrado. El diagrama de flujo PRISMA se convierte así en una herramienta visual poderosa que resume el proceso de depuración del corpus documental, permitiendo a otros investigadores evaluar la solidez metodológica del estudio.⁽¹⁾

La documentación detallada debe incluir, además: la sintaxis exacta de todas las cadenas de búsqueda utilizadas (incluyendo variantes descartadas), las fechas de ejecución de las consultas (importante dado el carácter dinámico de las bases de datos), y cualquier limitación conocida en la cobertura de las fuentes utilizadas. Tener claridad de todo esto es crucial desde la fase inicial de la investigación pues como se verá en capítulos siguientes todo esto debe quedar detallado en el informe final.

Los investigadores experimentados suelen complementar este proceso con validación externa, consultando a expertos temáticos para verificar que la búsqueda no haya omitido contribuciones seminales, o comparando sus resultados con revisiones sistemáticas existentes en el área. Este paso adicional ayuda a detectar posibles sesgos en la estrategia de búsqueda que podrían afectar las conclusiones del estudio. La combinación de validación interna (mediante protocolos estandarizados) y externa (a través de juicio experto) constituye la mejor garantía para producir resultados bibliométricos robustos y confiables.

Este enfoque metódico transforma la búsqueda bibliográfica de un mero procedimiento técnico en un proceso investigativo riguroso, donde cada decisión está justificada y documentada. La inversión de tiempo en esta fase crítica se traduce directamente en la calidad y credibilidad de los hallazgos posteriores, evitando los sesgos y limitaciones que frecuentemente afectan a los estudios bibliométricos realizados sin esta sistematicidad.

Esta reducción progresiva debe documentarse meticulosamente, especificando los motivos de exclusión en cada etapa. La trazabilidad completa del proceso no solo fortalece la validez interna del estudio, sino que facilita su replicación o actualización futura.

4.1.5. Errores comunes y soluciones

Una búsqueda demasiado amplia, que arroja más de 10 000 registros, suele ser inmanejable para análisis bibliométricos detallados y puede incluir gran cantidad de ruido documental. Este problema ocurre frecuentemente al utilizar términos genéricos sin filtros adecuados. La solución pasa por incorporar restricciones estratégicas como límites temporales (ejemplo: últimos 10 años), filtrado por tipo de documento (solo artículos de investigación) o enfoque en campos específicos (título y resumen en lugar de texto completo).

En el extremo opuesto, una búsqueda demasiado estrecha puede dejar fuera contribuciones seminales por usar terminología muy específica. Esto es particularmente problemático en campos emergentes donde los conceptos aún no han estandarizado su nomenclatura. Para evitarlo, es esencial construir cadenas de búsqueda que incorporen sinónimos y variantes terminológicas mediante operadores OR, además de consultar tesauros disciplinares como MeSH en PubMed o los términos controlados de Scopus.

Un error frecuente en ciertas disciplinas es ignorar libros y actas de conferencias, lo que genera un sesgo sistemático contra áreas como humanidades o ciencias computacionales respectivamente. Las bases tradicionales (WoS/Scopus) tienen cobertura limitada en estos

formatos, por lo que debe complementarse la búsqueda con Google Scholar para libros o Dimensions para proceedings. En estudios interdisciplinarios, esta omisión puede significar perder gran parte de la literatura relevante.

La falta de documentación de la estrategia de búsqueda compromete la reproducibilidad del estudio, un principio fundamental en investigación. La solución implica guardar sistemáticamente no solo la cadena final, sino todas sus iteraciones, especificando: plataforma utilizada, fecha de búsqueda, filtros aplicados y número de resultados obtenidos en cada paso. Formatos estándar como los diagramas PRISMA o plantillas de protocolos de revisión sistemática pueden adaptarse para este fin. Herramientas como Zotero o Mendeley permiten exportar los registros de búsqueda con metadatos completos para auditoría posterior.

Estos problemas y sus soluciones ilustran cómo el diseño de búsquedas bibliométricas es un proceso iterativo que requiere equilibrio entre exhaustividad y precisión. La experiencia muestra que dedicar el 30 % del tiempo total del estudio a perfeccionar y validar la estrategia de búsqueda puede mejorar significativamente la calidad de los resultados finales. La incorporación de revisiones por pares expertos en la fase de diseño metodológico ayuda a identificar potenciales omisiones o sesgos antes de ejecutar las búsquedas definitivas.

El uso creciente de técnicas de *machine learning* para el *screening* automatizado de grandes volúmenes de resultados está transformando este proceso, pero la construcción inicial de cadenas de búsqueda precisas sigue siendo fundamental. Plataformas como ASReview o Rayyan pueden luego ayudar en la fase de selección, pero dependen críticamente de una estrategia de recuperación documental bien diseñada desde el inicio. Esta sinergia entre métodos tradicionales y nuevas tecnologías marca el camino hacia búsquedas bibliométricas más eficientes y completas.

4.2 Herramientas alternativas para búsqueda bibliométrica: ampliando horizontes

Aunque Scopus y Web of Science siguen siendo referentes en análisis bibliométrico, existen alternativas valiosas que democratizan el acceso a la información científica. Google Scholar, con su amplia cobertura de literatura gris, junto con herramientas como *Publish or Perish* y gestores de referencias avanzados, ofrecen posibilidades únicas para investigadores con recursos limitados o que trabajan con formatos no convencionales

4.2.1 Google Scholar: ventajas y limitaciones prácticas

La principal fortaleza de Google Scholar radica en su capacidad para indexar contenido que frecuentemente queda fuera de las bases comerciales: desde *preprints* en arXiv hasta tesis doctorales y documentos de trabajo. Esta exhaustividad lo hace particularmente útil para estudios en humanidades o disciplinas emergentes donde la publicación formal puede ser más lenta. Sin embargo, su uso sistemático presenta desafíos técnicos significativos. El límite aproximado de 100 consultas diarias antes del bloqueo IP obliga a estrategias de muestreo cuidadosas, mientras que la falta de metadatos estructurados complica los análisis bibliométricos avanzados.

Para investigadores con habilidades técnicas, la librería “scholarly” de Python ofrece un balance entre automatización y respeto a los términos de servicio. El proceso básico implica: instalación mediante pip, configuración de consultas específicas, y extracción controlada de datos. Un *script* podría incluir:

```
```python
import csv
from scholarly import scholarly
```

```
search_query = scholarly.search_pubs("blockchain in healthcare")
with open('resultados.csv', 'w', newline="", encoding='utf-8') as file:
 writer = csv.writer(file)
 writer.writerow(["Título", "Autores", "Año", "Citas"])
 for i in range(50): Límite responsable
 try:
 pub = next(search_query)
 writer.writerow([
 pub['bib'].get('title', ''),
 pub['bib'].get('author', ''),
 pub['bib'].get('year', ''),
 pub.get('num_citations', 0)
])
 except StopIteration:
 break
 ...
```

Este enfoque incluye salvaguardas clave: límite de registros, manejo de errores, y almacenamiento en formato CSV para análisis posterior. Para usuarios sin programación, extensiones como Scholarcy ofrecen funcionalidad básica de extracción directamente desde el navegador, aunque con menos personalización.

El *scraping* de datos académicos debe seguir principios claros: intervalos entre consultas que simulen comportamiento humano, respeto a los robots.txt de cada sitio, y uso exclusivo para fines de investigación. Es crucial recordar que muchos documentos en Google Scholar están protegidos por derechos de autor, por lo que la extracción masiva de textos completos sin permiso constituye violación ética. Las mejores prácticas recomiendan:

- Limitar las extracciones a metadatos básicos (título, autores, citas)
- Almacenar solo la información necesaria para el análisis
- Citar adecuadamente todas las fuentes recuperadas
- Verificar periódicamente los términos de servicio de las plataformas

Estas herramientas alternativas funcionan mejor como complemento, no reemplazo, de las bases tradicionales. Un flujo de trabajo robusto podría comenzar con búsquedas sistemáticas en Scopus/WoS, seguido de un barrido en Google Scholar para capturar literatura adicional, finalizando con gestores como Zotero o Mendeley para organización y de duplicación. Esta triangulación metodológica mitiga los sesgos inherentes de cada plataforma mientras maximiza la cobertura documental.

#### 4.2.2 Harzing's Publish or Perish

*Publish or Perish* (PoP) no solo extrae datos de Google Scholar, sino que también ofrece integración con bases de datos académicas clave como Scopus, Web of Science (WoS) y PubMed, ampliando significativamente su utilidad para investigadores que requieren análisis más estructurados. Esta capacidad multiplataforma lo convierte en una solución versátil para estudios bibliométricos comparativos, permitiendo contrastar métricas de impacto entre diferentes fuentes y detectar posibles discrepancias en la cobertura de citas. En el capítulo 3 se abordó en detalle como instalar esta herramienta así como sus ventajas y particularidades. De manera general, para hacer una búsqueda con ella se procede de la siguiente manera.

1. Descarga e instalación:

- Disponible en [<https://harzing.com/resources/publish-or-perish>](<https://harzing.com/resources/publish-or-perish>) (Windows/Mac/Linux).

2. Primera búsqueda:

- Abre el software y haz clic en “New Query”.
- Ingresa términos de búsqueda (ej: `”artificial intelligence education”`).

4. Exportar resultados:

- Ve a File > Save As y elige CSV/Excel para análisis en Bibliometrix o Excel.

### 4.2.3 Gestores de referencias para análisis bibliométrico avanzado

Los gestores de referencias como Zotero y Mendeley han evolucionado más allá de simples organizadores bibliográficos, convirtiéndose en herramientas poderosas para la extracción y procesamiento masivo de datos bibliométricos. Su capacidad para capturar metadatos estructurados directamente desde portales académicos o archivos PDF los hace indispensables en flujos de trabajo de investigación cuantitativa.

Zotero destaca por su integración perfecta con bases de datos como Scopus y Web of Science. Al utilizar su conector de navegador, los investigadores pueden importar cientos de referencias con un solo clic, preservando información crítica como DOI, afiliaciones institucionales y campos de citación. El proceso de exportación a CSV (a través del menú contextual) genera tablas listas para su análisis en software especializado, aunque requiere atención durante la selección de campos personalizados para asegurar que todos los metadatos relevantes sean incluidos.

Mendeley ofrece ventajas distintas, particularmente en el manejo de colecciones de PDFs existentes. Su motor de reconocimiento de documentos puede extraer metadatos incluso de archivos sin información embebida, aunque con precisión variable. La función “**Verify Document Details**” permite corregir errores comúnmente encontrados en artículos más antiguos o de revistas menos establecidas. Para proyectos bibliométricos, su capacidad de exportación en formato RIS resulta especialmente valiosa, ya que este estándar conserva relaciones entre documentos que formatos tabulares como CSV podrían perder.

Un desafío recurrente con ambas herramientas es la calidad inconsistente de metadatos. Estudios recientes muestran que hasta el 15 % de los artículos en gestores de referencias requieren corrección manual de campos básicos como año de publicación o lista completa de autores. Aquí, APIs como CrossRef se vuelven complementos esenciales. El ejemplo en Python demostrado (utilizando la librería *habanero*) ilustra cómo automatizar la recuperación de metadatos faltantes mediante consultas basadas en títulos o fragmentos de texto. Este enfoque híbrido -combinando extracción automática con verificación programática- puede mejorar significativamente la eficiencia en la preparación de datos bibliométricos.

Consideraciones técnicas avanzadas emergen al trabajar con grandes volúmenes de referencias. Zotero permite la ejecución de scripts personalizados a través de su API interna (Zotero API), posibilitando tareas como:

- Normalización automática de nombres institucionales
- Detección y fusión de duplicados basada en similitud semántica
- Extracción de citas entre documentos mediante análisis de referencias

Mientras tanto, Mendeley Desktop (a diferencia de su versión web) ofrece acceso directo a su



base de datos SQLite local, permitiendo consultas complejas que superan las limitaciones de la interfaz gráfica. Investigadores con conocimientos de SQL pueden así realizar transformaciones avanzadas de datos antes de la exportación final.

La elección entre estas herramientas depende frecuentemente del flujo de trabajo específico. Zotero brilla en proyectos colaborativos gracias a su sincronización en la nube y permisos granulares, mientras que Mendeley puede preferirse en entornos donde la gestión de PDFs masiva es prioritaria. Ambos, sin embargo, comparten el potencial de servir como puente entre la recolección inicial de literatura y el análisis bibliométrico sofisticado en plataformas como VOSviewer o Bibliometrix.

Estos gestores pueden utilizarse para extraer datos de PDFs o incluso de la web para posteriormente ser utilizados en los estudios bibliométricos.

### **4.3 Formatos de exportación en bibliometría: interoperabilidad y análisis**

La selección del formato de exportación en bibliometría representa una decisión crítica que impacta directamente en la calidad y profundidad del análisis posterior. El formato RIS (*Research Information Systems*), desarrollado originalmente para bases de datos académicas, sobresale por su capacidad de preservar metadatos complejos como las relaciones de citación y las estructuras jerárquicas de autoría. Este formato mantiene un equilibrio único entre legibilidad humana y procesamiento automatizado, siendo particularmente valioso cuando se trabaja con herramientas de análisis de redes como VOSviewer. Su estructura de etiquetas estandarizadas (como AU para autores o PY para año de publicación) permite una transición fluida entre diferentes plataformas sin pérdida de información crítica.

BibTeX, por su parte, ofrece ventajas distintas en entornos donde prima la escritura académica técnica. Su sintaxis minimalista y su integración nativa con sistemas LaTeX lo hacen insustituible para investigadores que necesitan gestionar referencias mientras redactan. Sin embargo, esta misma simplicidad se convierte en limitación al abordar análisis bibliométricos avanzados, donde se requieren metadatos más ricos sobre citaciones, afiliaciones institucionales o indicadores de impacto. Herramientas modernas como PyBibX han extendido su utilidad al permitir cierto grado de procesamiento bibliométrico, pero aún requieren transformaciones adicionales para alcanzar el nivel de detalle que proporcionan otros formatos.

El formato CSV representa el puente más directo hacia el análisis cuantitativo avanzado. Su estructura tabular se adapta perfectamente a entornos estadísticos como R (a través de Bibliometrix) o Python (con Pandas), donde la manipulación de grandes volúmenes de datos exige flexibilidad. No obstante, esta misma adaptabilidad conlleva riesgos: las exportaciones automáticas desde algunas plataformas pueden aplanar estructuras de datos complejas, como listas de autores o jerarquías de instituciones, en cadenas de texto que requieren posterior procesamiento. La clave para aprovechar CSV en bibliometría está en garantizar que la exportación inicial capture todos los metadatos relevantes en columnas bien definidas.

#### **4.3.1 Conversión entre formatos: herramientas y usos en software bibliométrico**

La realidad del trabajo bibliométrico contemporáneo exige frecuentemente navegar entre múltiples formatos, cada uno optimizado para diferentes etapas del flujo de investigación. Las conversiones entre RIS, BibTeX y CSV no son meras traducciones técnicas, sino procesos que pueden ampliar o limitar las posibilidades analíticas. Herramientas como Zotero realizan estas transformaciones de manera aceptable para usos básicos, pero los proyectos de envergadura requieren enfoques más sofisticados.

En el ecosistema actual de software bibliométrico, cada herramienta principal ha desarrollado preferencias y capacidades específicas para manejar estos formatos. VOSviewer, por ejemplo, ha optimizado su motor para procesar archivos RIS manteniendo intactas las relaciones entre documentos, lo que es esencial para construir redes de co-citación precisas. CiteSpace, con su enfoque en análisis temporales, demanda estructuras de datos que preserven las cronologías de citación, algo que los formatos plain text derivados de Web of Science proporcionan naturalmente.

Bibliometrix en R representa quizás el punto más flexible en este espectro, capaz de ingerir múltiples formatos pero con una clara ventaja cuando se alimenta con RIS. Su función `convert2df()` no solo interpreta los metadatos estándar, sino que aplica algoritmos de limpieza y normalización que preparan los datos para análisis complejos como el mapeo de co-palabras o el cálculo de indicadores de colaboración internacional.

Para los investigadores que trabajan con pipelines personalizados, Python ofrece librerías como PyBibX que puentean las limitaciones de BibTeX para análisis bibliométricos. Estas herramientas permiten, por ejemplo, extraer redes semánticas desde conjuntos de referencias originalmente destinados a la composición de documentos académicos. No obstante, el proceso requiere frecuentemente etapas intermedias de transformación, especialmente cuando los datos provienen de múltiples fuentes con diferentes estándares de metadatos.

Las decisiones sobre formatos y conversiones repercuten directamente en la validez de los hallazgos bibliométricos. Un caso ilustrativo es el análisis de colaboraciones institucionales: mientras RIS preserva típicamente la estructura completa de afiliaciones, las conversiones automáticas a CSV pueden fragmentar esta información en columnas separadas, introduciendo artefactos en los mapas de redes generados por herramientas como CitNetExplorer.

La experiencia recomienda adoptar un enfoque estratificado para la gestión de formatos. En la fase inicial de recolección, RIS suele ser la opción más segura, garantizando que ningún metadato relevante se pierda durante la exportación desde las bases de datos. Para las etapas de limpieza y normalización, CSV brinda la flexibilidad necesaria para manipular grandes volúmenes de datos con herramientas estadísticas. Finalmente, en la fase de visualización y análisis especializado, muchos investigadores optan por regresar a formatos como RIS o a estructuras nativas del software bibliométrico utilizado.

Este ciclo iterativo subraya un principio fundamental: en bibliometría avanzada, los formatos de intercambio de datos no son meros contenedores pasivos, sino componentes activos que condicionan las posibilidades analíticas. La creciente interoperabilidad entre herramientas, impulsada por estándares como OpenAPI y formatos como JSON-LD, promete simplificar estos flujos de trabajo en el futuro. Sin embargo, hasta que estas innovaciones se adopten masivamente, el conocimiento profundo de RIS, BibTeX y CSV sigue siendo una habilidad esencial para cualquier investigador que busque extraer perspectivas significativas de los datos bibliográficos.

Ya está toda la información, pero ahora, ¿qué hacer con ella? En la próxima sección se abordará todo el tratamiento de los datos para obtener resultados e interpretaciones de alto nivel.

## **Rescapitulando**

- La obtención de datos bibliométricos es la primera fase práctica de un estudio y

define la validez de todo el análisis posterior.

- Las principales fuentes de información son: Scopus, Web of Science (WoS) y PubMed, por su cobertura, estructura y exportabilidad de metadatos.
- En Scopus, la búsqueda se optimiza usando operadores booleanos (AND, OR, NOT), filtros por año, país, área o afiliación, y códigos SOURCE-ID.
- En Web of Science, se emplean las colecciones *Core Collection* e *Emerging Sources* y las categorías temáticas para refinar resultados.
- En PubMed, se utilizan términos controlados MeSH, límites clínicos y estrategias de combinación entre título/resumen/palabras clave.
- Las herramientas alternativas amplían la obtención de datos cuando las bases tradicionales son insuficientes o restringidas. Entre ellas destacan:
  - Scholar (scraping ético) → útil para ampliar cobertura pero con cautela por ruido y duplicados.
  - Harzing's Publish or Perish → permite extraer y analizar métricas de Google Scholar con filtros de Google personalizables.
  - Gestores de referencias (Zotero, Mendeley) → facilitan la exportación masiva de metadatos y conversión de formatos.
- Los formatos más usados para exportar registros son:
  - RIS → compatible con la mayoría de softwares bibliográficos.
  - BibTeX → estándar para análisis en R o Python.
  - CSV → permite manipulación directa en Excel o software estadístico.
- Es fundamental mantener la coherencia del formato durante el proceso, evitando pérdida de campos (autores, afiliación, DOI, país, año).
- Se recomienda conservar la versión original de cada *dataset* y documentar el flujo de obtención de datos (fecha, fuente, filtros y número final de registros).
- La ética en la obtención de datos prohíbe prácticas como *scraping* masivo sin autorización o el uso de datos con derechos restringidos.
- Se promueve la reproducibilidad mediante la descripción clara de los criterios de búsqueda y la publicación del *dataset* o código empleado.
- El capítulo vincula la obtención de datos con las siguientes etapas metodológicas: limpieza, análisis y comunicación de resultados.
- En esta fase se definen los datos pertinentes: variables bibliográficas (autor, año, revista, país, citas, afiliación, tipo de documento).
- Los errores más comunes en esta etapa son: duplicados, registros incompletos, errores en nombres de autores y confusión entre fuentes.
- El control de calidad incluye verificar consistencia, eliminar duplicados y homogeneizar nombres mediante normalización.
- La documentación completa del proceso de búsqueda aumenta la transparencia y credibilidad del estudio bibliométrico.
- Se sugiere acompañar esta fase con capturas de pantalla o scripts reproducibles (ej. en Google Colab o RMarkdown).
- El capítulo enfatiza que sin una base de datos sólida, las métricas y visualizaciones posteriores pierden fiabilidad.
- En síntesis, esta parte enseña a construir el “corazón empírico” del estudio: una base bibliográfica ética, limpia, completa y normalizada.

## Preguntas de autoevaluación

1. ¿Por qué la obtención de datos es considerada la fase crítica en un estudio bibliométrico?
2. ¿Qué características diferencian a Scopus, Web of Science y PubMed como fuentes

de datos?

3. ¿Qué función cumplen los operadores booleanos en las búsquedas bibliográficas?
4. ¿Qué ventajas y limitaciones presenta el uso de Google Scholar mediante *scraping* ético?
5. ¿Para qué sirve el programa *Harzing's Publish or Perish* en la bibliometría?
6. ¿Qué diferencia existe entre los formatos RIS, BibTeX y CSV al exportar registros?
7. ¿Qué buenas prácticas deben seguirse para asegurar la reproducibilidad de las búsquedas bibliográficas?
8. ¿Cuáles son los principales errores que pueden afectar la calidad del *dataset* bibliométrico?
9. ¿Qué criterios éticos deben respetarse en la recolección de datos científicos?
10. ¿Por qué la documentación y normalización del *dataset* son esenciales antes del análisis bibliométrico?

## 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 Publishers; 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, editors. Measuring scholarly impact: methods and practice. Cham (Switzerland): Springer; 2014. <https://doi.org/10.1007/978-3-319-10377-8>
5. Cronin B, Sugimoto CR, editors. Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. Cambridge (MA): MIT Press; 2014. <https://doi.org/10.7551/mitpress/9780262026792.001.0001>
6. Cooper HM, Hedges LV. Research synthesis and meta-analysis: a step-by-step approach. 5th ed. Los Angeles (CA): SAGE Publications; 2019.

## REFERENCIAS BIBLIOGRÁFICAS

1. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Systematic Reviews. 2022; 18(2):e1230.

# Part III / Parte III

ADVANCED ANALYSIS

ANÁLISIS AVANZADO



## Chapter 5 / Capítulo 5

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch05

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

## BIBLIOMETRIC INDICES / ÍNDICES BIBLIOMÉTRICOS

### 5.1. Clasificación

Los índices bibliométricos constituyen el núcleo del análisis cuantitativo en estudios de ciencia, tecnología e innovación, proporcionando medidas estandarizadas para evaluar el impacto, la productividad y la influencia de la producción científica. Su correcta clasificación e interpretación resulta fundamental para extraer conclusiones válidas y evitar los frecuentes errores derivados de aplicaciones mecánicas sin contextualización adecuada. Este capítulo presenta un marco integral para comprender la taxonomía, cálculo e interpretación de los principales indicadores bibliométricos utilizados en la evaluación científica contemporánea.

**Los Indicadores de Producción Científica** cuantifican el volumen de resultados investigativos generados, ofreciendo una base inicial para evaluar la actividad científica. La métrica más elemental es el número total de publicaciones. Un recuento bajo puede ser síntoma de una productividad limitada o de una estrategia deliberada de publicar exclusivamente en revistas de alto impacto, lo que conlleva procesos de revisión más extensos. Por el contrario, un volumen elevado sugiere una alta productividad, pero también puede alertar sobre una posible fragmentación de resultados, donde los hallazgos de un solo estudio se dividen en múltiples artículos. El estándar de lo que se considera “alto” o “bajo” varía enormemente según la disciplina y la etapa de la carrera.

Para refinar la medición, el índice de productividad individual ajusta la contribución de un investigador considerando factores como su posición en la lista de autores y el tipo de contribución realizada. Un valor bajo en este índice indica una participación mayoritaria como coautor en posiciones secundarias, sugiriendo un rol de apoyo en los proyectos. Un valor alto, por el contrario, denota una autoría principal frecuente (como primer o último autor), reflejando un liderazgo intelectual sostenido. Aunque no existe un umbral universal, en muchas disciplinas un valor superior a 0,7 en escalas normalizadas se considera una contribución sustancial y liderazgo.<sup>(1)</sup>

El índice de colaboración calcula el grado de trabajo en equipo mediante el promedio de coautores por publicación. Un valor bajo, cercano a 1, es característico de campos tradicionalmente solitarios como la filosofía o las humanidades. Un valor alto, que puede superar la decena en áreas como la física de partículas o la genómica, refleja una investigación eminentemente colaborativa. La interpretación de este indicador debe considerar las normas disciplinares, ya que la autoría múltiple responde a tradiciones científicas y necesidades metodológicas profundamente diferentes entre campos de conocimiento.<sup>(2)</sup>

**Los Indicadores de Impacto Basados en Citas** evalúan la influencia e importancia relativa de las publicaciones científicas, midiendo cómo son recibidas y utilizadas por la comunidad académica a través de las referencias que reciben. El número total de citas proporciona una medida cruda del impacto acumulado. Sin embargo, su interpretación requiere una normalización temporal y disciplinar. Un valor bajo puede indicar una investigación poco influyente o altamente especializada, mientras que un valor alto sugiere contribuciones fundamentales que han resonado en el campo. No existe un estándar universal, pues un número de citas considerado bajo en inmunología podría ser excepcionalmente alto en filosofía.

El índice-h es un indicador bibliométrico que busca equilibrar la productividad de un investigador con el impacto de su trabajo. Su cálculo se realiza ordenando las publicaciones de un autor de mayor a menor número de citas recibidas. El índice-h corresponde al punto en el



que el número de orden de un artículo (h) coincide con el número de citas que ha recibido ese mismo artículo. Por ejemplo, un índice-h de 10 significa que el autor tiene 10 artículos que han sido citados al menos 10 veces cada uno.

Un valor bajo puede ser indicativo de una carrera científica incipiente o de la especialización en un área de estudio con una tasa de citación generalmente reducida. Por el contrario, un índice-h elevado suele reflejar una trayectoria consolidada y la producción constante de trabajos que han influido significativamente en su campo.

Para realizar una comparación más justa entre investigadores de distintas áreas, se utiliza el índice-h corregido por disciplina. Su cálculo se realiza en varios pasos. Primero, se identifica la disciplina principal del investigador. Luego, se obtiene el índice-h promedio de todos los investigadores en esa misma disciplina, utilizando una base de datos bibliográfica normalizada. Finalmente, se divide el índice-h individual del investigador entre este promedio de referencia. El resultado es un valor que indica si su impacto está por encima o por debajo de la media de su campo especializado.

Como referencia, Hirsch propuso que, para científicos en activo, un índice-h aproximadamente igual a los años de carrera (m) es característico de un investigador “exitoso”. Un valor alrededor de 2m correspondería a “investigadores sobresalientes”, mientras que un  $h \approx 3m$  se asociaría a “científicos verdaderamente únicos”, siempre considerando las particularidades de cada disciplina.<sup>(3)</sup>

Para complementarlo, el índice g otorga mayor peso a las publicaciones excepcionalmente citadas, siendo más sensible a la presencia de trabajos seminales dentro de la producción de un autor. Un valor del índice g que es significativamente más alto que el índice h sugiere que el investigador cuenta con uno o varios artículos altamente citados que han tenido un impacto extraordinario. Finalmente, el índice m ajusta el índice h por los años de carrera investigadora, proporcionando una medida de productividad impactante anualizada. Un valor de m superior a 1 se considera muy bueno, ya que indica que el investigador genera, en promedio, más de un artículo “h” por año.<sup>(3)</sup>

El índice m ajusta el índice h por años de carrera investigadora, donde valores bajos sugieren impacto decreciente y valores altos indican productividad sostenida.

**Los Indicadores de Colaboración Científica** analizan las redes y patrones de cooperación entre investigadores, instituciones y países, revelando las dinámicas sociales detrás de la producción de conocimiento. El índice de cooperación calcula el porcentaje de publicaciones con múltiples autores. Valores bajos pueden indicar un trabajo individualista o ser característicos de campos tradicionalmente solitarios. Por el contrario, valores altos, a menudo por encima del 80 % en ciencias experimentales, reflejan una investigación eminentemente colaborativa, propia de la ciencia moderna donde la autoría colectiva es la norma para abordar problemas complejos.

El índice de colaboración internacional mide la proporción de trabajos con coautores extranjeros. Un valor bajo es a menudo indicativo de un cierto aislamiento científico o de un enfoque de investigación con relevancia principalmente local. Un valor alto, en cambio, demuestra una activa integración en redes globales de conocimiento y suele correlacionarse con un mayor impacto. En evaluaciones institucionales, se suele aspirar a que este índice supere el 30-40 %, considerándose un valor alto por encima del 50 %. Asimismo, el índice de colaboración institucional evalúa la diversidad de afiliaciones en las publicaciones, revelando la capacidad de

establecer alianzas estratégicas más allá de la propia institución.

**Los Indicadores de Calidad de Fuentes** se concentran en evaluar el prestigio y la influencia de los canales de difusión científica, como las revistas académicas, asumiendo que la calidad del vehículo de publicación refleja, hasta cierto punto, la calidad de los artículos que contiene. El Factor de Impacto (FI), calculado anualmente en el *Journal Citation Reports*, mide la frecuencia promedio con que los artículos de una revista son citados en un período específico. Valores bajos pueden corresponder a revistas muy especializadas o de alcance regional. Valores altos, en cambio, reflejan típicamente a las revistas líderes en sus campos.

Para superar las limitaciones del FI, se han desarrollado métricas alternativas. El *SCImago Journal Rank* (SJR) incorpora el prestigio de las revistas citantes, siendo más sensible a la calidad de las citas recibidas. El *CiteScore* emplea una ventana de citación más amplia. Por su parte, el *Source Normalized Impact per Paper* (SNIP) corrige las diferencias en prácticas de citación entre disciplinas, donde valores inferiores a 1 indican un impacto por debajo del promedio de su campo y valores superiores a 1 señalan un impacto por encima de la media. Estos indicadores deben utilizarse críticamente, evitando extrapolaciones inadecuadas a nivel de artículo o investigador individual.<sup>(4)</sup>

**Los Indicadores de Influencia y Liderazgo** evalúan el papel específico y la contribución real de un investigador dentro de las redes científicas, yendo más allá de las métricas de volumen o impacto crudo. El índice h individual excluye las autocitas para medir estrictamente el impacto externo y el reconocimiento por parte de pares ajenos al círculo inmediato del investigador. Un valor bajo en este índice puede indicar una alta dependencia de las autocitas para mantener un perfil de citas, o un impacto limitado. Un valor alto, por el contrario, refleja un reconocimiento externo sólido.

El índice de liderazgo calcula la proporción de publicaciones donde el investigador aparece como autor de correspondencia o primer autor, roles que típicamente denotan una contribución intelectual principal. Valores bajos sugieren roles secundarios o de apoyo en grandes colaboraciones, mientras que valores altos indican un liderazgo activo en proyectos. En etapas avanzadas de la carrera, se espera que este índice se mantenga por encima del 0,5 para reflejar independencia investigadora. Complementariamente, el índice de originalidad mide la diversidad de las fuentes citadas, donde un valor alto refleja interdisciplinariedad. Una alternativa a este, para la diversidad de fuentes citadas (interdisciplinariedad): El Índice de Rao-Stirling es una métrica establecida para medir la diversidad e interdisciplinariedad basada en las referencias citadas.<sup>(5)</sup>

**Los Indicadores Compuestos y Normalizados** integran múltiples dimensiones de la actividad científica en medidas integrales que permiten comparaciones más justas y matizadas entre investigadores, instituciones o países, teniendo en cuenta las particularidades de cada campo. El *Field-Weighted Citation Impact* (FWCI) es un indicador clave que compara las citas recibidas por un conjunto de publicaciones contra el promedio mundial en sus respectivos campos específicos. Un valor de 1 representa exactamente el impacto esperado para la media mundial. Valores inferiores a 1 indican un impacto por debajo del esperado, mientras que valores superiores a 1 señalan un impacto por encima del esperado.<sup>(6)</sup>

Los indicadores de excelencia identifican el porcentaje de publicaciones de un investigador que se encuentran entre el 10 % más citado a nivel mundial en sus respectivos años y campos. Este es un umbral de alto impacto. Valores bajos, por ejemplo, inferiores al 5 %, sugieren que

las contribuciones se sitúan en el rango promedio. Por el contrario, valores altos, superiores al 10 % o especialmente al 15 %, indican una producción consistentemente influyente y de élite, demostrando una capacidad recurrente para generar investigación que define el estado del arte en su área. Estos indicadores son particularmente útiles para identificar desempeños excepcionales.<sup>(7)</sup>

Nuevos Indicadores Emergentes amplían el espectro métrico tradicional hacia dimensiones complementarias, capturando aspectos como la visibilidad inmediata, la sostenibilidad del impacto y la diversidad de la huella investigadora. El índice i10, popularizado por Google Scholar, contabiliza de forma sencilla el número de publicaciones con al menos 10 citas. Proporciona una medida inmediata de trabajos moderadamente influyentes. Un valor bajo puede indicar un predominio de publicaciones muy recientes que aún no han acumulado citas, o una especialización en campos de nicho. Un valor alto sugiere varios trabajos que han logrado una penetración significativa en la literatura.

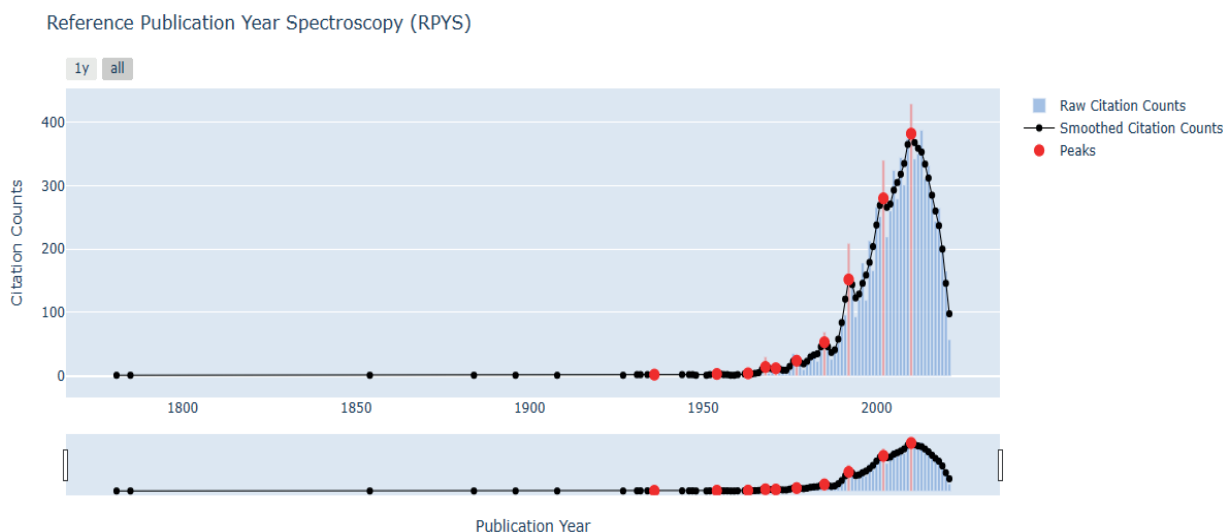
El índice de sostenibilidad de impacto evalúa la persistencia temporal de las citas recibidas, analizando la curva de decadencia de las citas a lo largo del tiempo. Valores decrecientes de forma pronunciada pueden señalar una obsolescencia temprana o que la investigación responde a tendencias pasajeras. Valores estables o de lento declive indican, en cambio, una relevancia prolongada y que las contribuciones mantienen su utilidad para la comunidad científica varios años después de su publicación, lo que es signo de un trabajo fundamental. Este índice es crucial para distinguir entre el impacto pasajero y el perdurable.

El índice de diversidad temática es otro que pertenece a este grupo de indicadores y calcula la variedad de áreas científicas en las que un investigador publica. Valores bajos reflejan una alta especialización temática, con un enfoque profundo en un campo bien definido. Valores altos indican versatilidad investigadora y una capacidad para contribuir a múltiples disciplinas, lo que puede ser una ventaja en entornos interdisciplinarios y es cada vez más valorado en la resolución de problemas complejos que requieren perspectivas integradoras. Un investigador con un perfil diversificado suele tener un índice superior a 0,5 en escalas de entropía temática.

La interpretación responsable de cualquier índice bibliométrico requiere comprender sus fundamentos metodológicos, limitaciones técnicas y contexto disciplinar específico. Ningún indicador individual captura la multidimensionalidad del impacto científico, por lo que su uso combinado y crítico resulta esencial para evaluaciones balanceadas. La transparencia en los métodos de cálculo, la selección de indicadores apropiados para cada propósito evaluativo, y la consideración de factores cualitativos complementarios constituyen principios irrenunciables en la aplicación ética de la bibliometría a la evaluación científica contemporánea. Cada índice debe entenderse como una lente particular que revela aspectos específicos del complejo fenómeno de la comunicación científica, donde su verdadero valor emerge del análisis integrado y contextualizado de múltiples perspectivas métricas.

## **5.2. Evolución Temporal de los Indicadores Bibliométricos**

La evolución temporal de los indicadores bibliométricos proporciona una perspectiva dinámica crucial para evaluar trayectorias investigadoras, trascendiendo la mera fotografía estática que ofrecen los valores puntuales. El análisis longitudinal revela patrones de desarrollo profesional, estrategias de publicación y el impacto sostenido de las contribuciones científicas. La interpretación de esta evolución varía sustancialmente según el indicador específico, requiriendo una contextualización cuidadosa dentro de la etapa de carrera del investigador y las normas de su disciplina.



**Figura 5.1.** Ejemplo de grafica de citas/publicaciones x año en una espectroscopia

En los indicadores de producción, una trayectoria ascendente en el número de publicaciones anuales sugiere una productividad en expansión, posiblemente asociada a la consolidación de un grupo de investigación o a la obtención de proyectos financiados. Por el contrario, un descenso sostenido puede indicar una transición hacia roles más administrativos, una estrategia deliberada de calidad sobre cantidad, o dificultades para mantener actividad competitiva. La evolución del índice de colaboración, cuando muestra un incremento progresivo, suele reflejar una integración creciente en redes científicas más amplias y complejas.

La trayectoria del índice  $h$  es particularmente informativa. Un crecimiento lineal o acelerado en los primeros años de carrera señala una exitosa consolidación científica. Durante la etapa de mitad de carrera, se espera un crecimiento sostenido, mientras que una estabilización en fases avanzadas puede ser natural. No obstante, un estancamiento prematuro o un descenso podrían sugerir una pérdida de relevancia científica. El índice  $m$ , al ajustar por años de actividad, permite identificar si un investigador mantiene un ritmo de producción impactante constante o si este se diluye con el tiempo.

En los indicadores de impacto, la evolución del *Field-Weighted Citation Impact* (FWCI) revela cómo cambia la recepción de la obra de un investigador en relación con su campo. Una tendencia alcista indica una creciente influencia y que sus trabajos recientes son recibidos con mayor interés que los anteriores. Una tendencia a la baja, especialmente si cae por debajo de 1, puede señalar que la investigación no se mantiene a la vanguardia del campo. El porcentaje de publicaciones en el top 10 % de citas debe analizarse por ventanas temporales para identificar si la excelencia se mantiene, mejora o se reduce.

La sostenibilidad del impacto se visualiza mediante la curva de citas acumuladas a lo largo del tiempo para diferentes publicaciones. Los trabajos seminales muestran una curva de crecimiento constante y pronunciada, indicando relevancia prolongada. Los trabajos de actualidad crecen rápidamente, pero pueden saturarse pronto, indicando un interés temporal. Un perfil de citas plano sugiere una influencia limitada. El análisis de la evolución de la colaboración internacional, mostrando un aumento en su proporción, apunta hacia una progresiva internacionalización del

perfil investigador, algo altamente valorado en contextos de evaluación global.

### **5.3. Combinación de Indicadores Bibliométricos con Variables Contextuales**

La potencia analítica de los indicadores bibliométricos se multiplica cuando se combinan sistemáticamente con variables contextuales. Estas integraciones trascienden la evaluación unidimensional, permitiendo diseccionar la actividad científica para responder preguntas complejas sobre equidad, movilidad, especialización y transferencia del conocimiento. Cada combinación persigue un fin analítico específico, revelando patrones estructurales y dinámicas subyacentes que de otra forma permanecerían ocultas en los promedios agregados.

La combinación del factor de impacto de la revista con la posición en la autoría y el sexo del investigador se utiliza para diagnosticar posibles sesgos de género en la comunicación científica. Esta triangulación permite investigar si existen diferencias sistemáticas en la visibilidad y el liderazgo, analizando si, por ejemplo, los autores masculinos tienden a publicar como primeros o últimos autores en revistas de mayor prestigio con más frecuencia que sus colegas femeninas. Un hallazgo recurrente en algunos campos es la sobrerrepresentación de hombres en la autoría de correspondencia, lo que señala barreras en el acceso a posiciones de liderazgo intelectual. Esta aproximación es fundamental para diseñar políticas basadas en evidencia que fomenten la equidad en la ciencia.

La relación entre el país de afiliación, el índice de colaboración internacional y el impacto normalizado (FWCI) se emplea para evaluar la posición geoestratégica de los sistemas nacionales de ciencia y tecnología. Un país con una alta tasa de colaboración internacional pero un FWCI moderado podría indicar una estrategia de internacionalización exitosa que, sin embargo, no se traduce aún en un liderazgo científico global, sugiriendo un papel más de participante que de impulsor. Por el contrario, una nación con un FWCI muy superior a 1 y una colaboración internacional selectiva sugiere una posición de vanguardia y autonomía científica. Esta combinación es crucial para los organismos de financiación que buscan calibrar sus políticas de cooperación internacional.

La integración de la antigüedad del investigador (años desde su primera publicación) con el índice de liderazgo y la diversidad temática permite estudiar las trayectorias profesionales y la evolución de los intereses científicos. Una antigüedad alta con un índice de liderazgo bajo puede indicar una carrera desarrollada principalmente en roles de apoyo, mientras que la misma antigüedad con una diversidad temática creciente sugiere una evolución hacia intereses interdisciplinarios. Esta combinación ayuda a las instituciones a comprender y apoyar las diferentes trayectorias y patrones de movilidad intelectual dentro de su plantilla investigadora.

La combinación del tipo de documento (artículo, revisión, patente, procedimiento de conferencia) con el índice de colaboración industrial y las citas recibidas en patentes arroja luz sobre los mecanismos de transferencia de conocimiento e innovación. Un alto porcentaje de publicaciones en procedimientos de conferencias junto con una elevada colaboración industrial sugiere una investigación orientada al desarrollo y la aplicación inmediata. En cambio, un perfil dominado por artículos de investigación básica y revisiones, incluso con alto impacto académico, puede indicar una desconexión con los sectores productivos. Esta métrica compuesta es vital para las políticas de innovación.

Cruzar los indicadores de especialización temática con la procedencia institucional (universidad, organismo público de investigación, hospital, empresa) y el índice de centralidad en redes de coautoría permite cartografiar ecosistemas de conocimiento. Una institución con un

índice de especialización alto y una centralidad baja actúa como un nodo especializado, pero potencialmente aislado. La misma institución con alta centralidad se configura como un centro de referencia obligada en ese nicho. Estas combinaciones estratégicas son indispensables para la planificación y la toma de decisiones informadas en política científica a nivel macro y meso.

## 5.4. Cálculo práctico

### 5.4.1. Con R/Bibliometrix (**biblioAnalysis()**)

El cálculo de índices bibliométricos mediante R y el paquete Bibliometrix representa una de las metodologías más robustas y completas disponibles actualmente. La función **biblioAnalysis()** constituye el núcleo del proceso analítico, procesando set de datos bibliográficos completos para generar un objeto que contiene más de 30 indicadores diferentes. La implementación práctica inicia con la carga del set de datos en formato marco de datos, como se vio en epígrafes anteriores, seguida de la aplicación directa de la función principal: **results <- biblioAnalysis(dataframe)**. Esta operación ejecuta automáticamente todos los cálculos necesarios, desde los indicadores básicos de productividad hasta métricas avanzadas de colaboración e impacto y devuelve el resultado como texto plano.

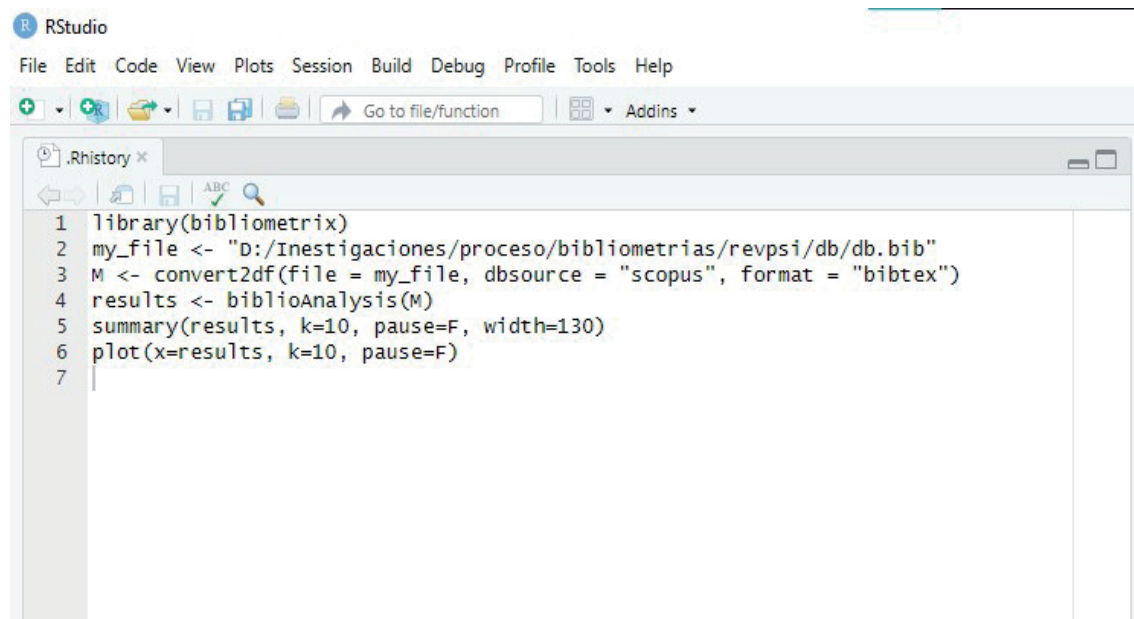


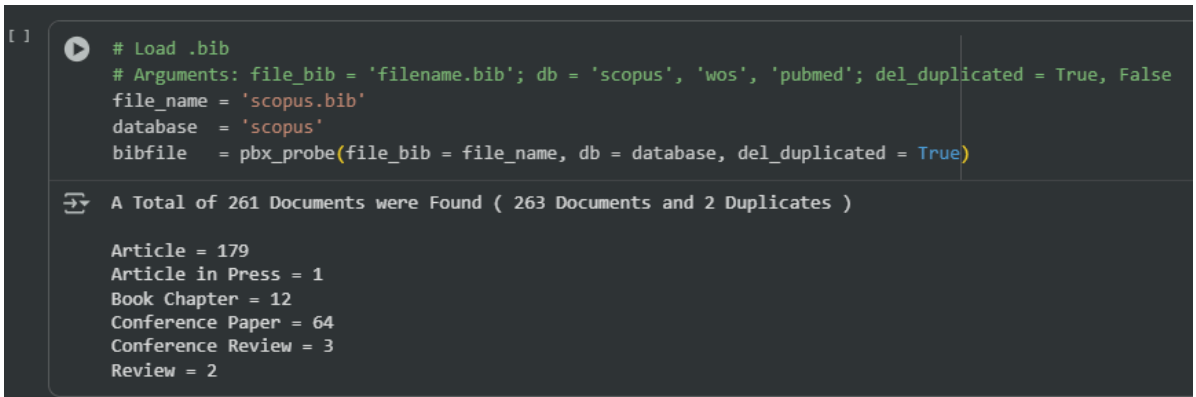
Figura 5.2. Proyecto básico en R Bibliometrix

La extracción de resultados específicos se realiza mediante funciones complementarias que acceden al objeto generado por **biblioAnalysis()**. Para obtener el índice h y variantes, se utiliza **summary(results)\$h.index**, mientras las estadísticas de colaboración internacional se recuperan con **summary(results)\$Collaboration**. La función **plot(results)** genera visualizaciones inmediatas de la distribución temporal de publicaciones y citas, las que se pueden ver en el panel de la esquina inferior derecha de R Studio.

La personalización de los cálculos en Bibliometrix permite adaptar los análisis a necesidades específicas de investigación. Mediante parámetros como **sep = “;”** en la función inicial, se ajustan los separadores de autores e instituciones según las características del set de datos. Para análisis comparativos entre períodos, la función **timeslice** divide la serie temporal en segmentos específicos, calculando indicadores evolutivos. La integración con otras librerías de R



como *igraph* enriquece las posibilidades analíticas, permitiendo calcular índices de centralidad personalizados y realizar análisis de comunidades en redes complejas de coautoría y citación.



```
[] # Load .bib
Arguments: file_bib = 'filename.bib'; db = 'scopus', 'wos', 'pubmed'; del_duplicated = True, False
file_name = 'scopus.bib'
database = 'scopus'
bibfile = pbx_probe(file_bib = file_name, db = database, del_duplicated = True)

A Total of 261 Documents were Found (263 Documents and 2 Duplicates)

Article = 179
Article in Press = 1
Book Chapter = 12
Conference Paper = 64
Conference Review = 3
Review = 2
```

Figura 5.3. Carga del set de datos en PyBibX

El núcleo analítico de PyBibX reside en la clase **MetricsCalculator**, que implementa métodos especializados para cada categoría de indicadores. Tras cargar los datos, se instancia el calculador mediante `calculator = MetricsCalculator(parsed_data)` y se invocan métodos específicos como `calculator.h_index()` para el índice h, `calculator.g_index()` para el índice g, y `calculator.m_index()` para la versión normalizada por tiempo. Para análisis de colaboración, `calculator.collaboration_metrics()` genera indicadores de coautoría multi nivel, mientras `calculator.citation_analysis()` proporciona estadísticas detalladas sobre distribución de citas. Cada método retorna no solo el valor numérico sino también metadatos contextuales que facilitan la interpretación de resultados.

Las capacidades avanzadas de PyBibX incluyen análisis temporal mediante la función `temporal_analysis()`, que segmenta los datos por períodos definidos por el usuario y calcula la evolución de indicadores. Para identificar trabajos seminales, `burst_detection()` aplica algoritmos de detección de picos de citación. La integración con pandas permite convertir resultados en marcos de datos para análisis posteriores, mientras las utilidades de visualización nativas generan gráficos de evolución temporal y redes de colaboración. Esta combinación de capacidades métricas y de visualización posiciona a PyBibX como una herramienta particularmente versátil para investigadores que requieren análisis bibliométricos completos dentro del ecosistema Python.

En el ejemplo de código disponible en la documentación oficial de pybibx están los códigos necesarios para correr todos estos índices, por lo que como se dijo anteriormente, no será necesario ni memorizarlos ni escribirlos manualmente.

#### 5.4.2. Alternativas: Publish or Perish y Excel (fórmulas)

Para investigadores que requieren soluciones inmediatas sin programación, *Publish or Perish* constituye una alternativa. Una vez cargados los datos este mostrara en un panel derecho diferentes índices como el h, g, m, entre otros.

Aunque las hojas de cálculo en Excel inicialmente no están diseñadas para análisis bibliométrico, ofrece capacidades mediante la aplicación estratégica de fórmulas y funciones.



El cálculo del índice h puede implementarse ordenando las citas por artículo de forma descendente y aplicando la fórmula  $=MAX(FILA(A1:A100)*(A1:A100 \geq FILA(A1:A100)))$  como matriz, donde A1:A100 contiene el número de citas por publicación.

Para el índice g, la aproximación requiere  $=MAX(FILA(A1:A100)*(A1:A100^2 \geq SUMA(A1:A100)))$ , mientras el índice i10 se calcula simplemente con  $=CONTAR.SI(A1:A100;">=10")$ .

Estas implementaciones, aunque básicas, proporcionan una verificación independiente de resultados obtenidos mediante herramientas especializadas. Sin embargo, las capacidades analíticas de Excel se expanden significativamente mediante el uso combinado de funciones estadísticas y tablas dinámicas.

Para análisis de colaboración, las funciones *CONTAR.SI* y *SI* permiten calcular porcentajes de coautoría internacional e institucional mediante fórmulas como  $=CONTAR.SI(rango\_afiliaciones;"*país*")/CONTAR.A(rango\_afiliaciones)$ .

Las tablas dinámicas facilitan agregaciones por autor, institución o período temporal, mientras los gráficos integrados proveen visualizaciones inmediatas de distribuciones de citas y patrones de productividad. Para usuarios avanzados, la implementación de macros en VBA permite automatizar cálculos recurrentes y generar reportes estandarizados, transformando Excel en una herramienta bibliométrica sorprendentemente competente para proyectos de escala moderada.

## 5.5. Interpretación crítica

La interpretación de índices bibliométricos debe reconocer fundamentalmente las profundas diferencias en culturas de publicación y citación entre disciplinas académicas. En ciencias duras como física o biomedicina, los ciclos de publicación son rápidos, con alta prevalencia de autoría múltiple y tasas de citación elevadas que reflejan tanto el impacto genuino como prácticas establecidas de citación rutinaria. Un índice h de 20 en física de partículas podría considerarse moderado, mientras el mismo valor en humanidades representaría una influencia excepcional. Esta disparidad surge de diferencias estructurales: las ciencias generan más publicaciones anuales por investigador y desarrollan tradiciones de citación que priorizan la actualización constante de referencias.

Las humanidades y ciencias sociales operan bajo paradigmas radicalmente diferentes, donde los libros constituyen el formato principal de comunicación científica y los ciclos de evaluación son sustancialmente más largos. La monografía especializada, con su desarrollo argumental profundo y sostenido, raramente recibe el reconocimiento métrico proporcional a su influencia intelectual real cuando se aplican indicadores diseñados para artículos de revistas científicas.<sup>(8)</sup>

La ingeniería y tecnologías aplicadas presentan patrones híbridos donde la citación académica convive con otros indicios de impacto como patentes, desarrollos tecnológicos y transferencia al sector productivo. Un investigador en ingeniería podría tener un índice h modesto mientras genera innovaciones con significativa repercusión industrial, creando una desconexión peligrosa entre métricas académicas e impacto real. Estas disciplinas frecuentemente publican en conferencias especializadas cuyas citas no son completamente capturadas por las bases de datos bibliográficas tradicionales, subestimando sistemáticamente su impacto intelectual.

Las ciencias de la salud evidencian sesgos particulares derivados de la concentración de citas en revisiones sistemáticas y meta-análisis, que reciben desproporcionadamente más citas que

los estudios primarios. Un investigador clínico dedicado a estudios longitudinales complejos, pero con muestras limitadas podría aparecer como menos influyente que colegas que publican revisiones frecuentes, distorsionando la evaluación de contribuciones sustantivas al avance del conocimiento médico. La especialización extrema en subcampos con diferentes tamaños de comunidad académica introduce adicionales variaciones que invalidan comparaciones directas.

La evaluación interdisciplinaria representa el desafío más complejo, donde investigadores que trabajan en fronteras entre campos enfrentan doble penalización: sus publicaciones pueden ser menos citadas en cada disciplina individual mientras quedan fuera de los núcleos centrales que concentran las citas. Un neurocientífico computacional, por ejemplo, podría publicar en revistas de neurociencia y de informática, dividiendo su impacto entre dos comunidades con diferentes prácticas de citación y sin alcanzar umbrales de visibilidad crítica en ninguna por separado, a pesar de contribuciones potencialmente transformadoras en la intersección disciplinaria.

La normalización disciplinar intenta corregir estos sesgos mediante comparativas dentro de campos definidos, pero enfrenta limitaciones metodológicas significativas. Indicadores como el *Field-Weighted Citation Impact* (FWCI) utilizan categorías amplias que frecuentemente agrupan subdisciplinas con culturas de publicación heterogéneas. Un filósofo analítico y un historiador de la filosofía comparten categoría, pero operan en ecosistemas de citación sustancialmente diferentes, invalidando en la práctica muchas correcciones estadísticas. La solución reside en complementar las métricas cuantitativas con evaluación cualitativa contextualizada por expertos disciplinares que comprenden estas sutiles pero cruciales diferencias.

## Recapitulando

- Los índices bibliométricos son medidas cuantitativas utilizadas para evaluar la productividad, impacto e influencia científica.
- Constituyen la base para comparar autores, revistas, instituciones o países dentro de un mismo campo de conocimiento.
- Se derivan principalmente de los registros de publicaciones y citas en bases de datos como Web of Science, Scopus y Google Scholar.
- Los indicadores se agrupan en tres grandes categorías:
  - Productividad científica (cantidad de publicaciones).
  - Impacto o influencia (citaciones recibidas).
  - Colaboración y redes (coautorías, afiliaciones).
- El Factor de Impacto (FI), creado por Eugene Garfield, mide el promedio de citas que reciben los artículos de una revista en dos años.
- El *CiteScore* (Elsevier) evalúa las citas recibidas en un período de cuatro años, cubriendo más revistas que el FI.
- El *SCImago Journal Rank* (SJR) pondera las citas según la relevancia de las revistas emisoras, basado en el algoritmo PageRank.
- El *Source Normalized Impact per Paper* (SNIP) ajusta el impacto de citas según las diferencias entre disciplinas.
- A nivel de autor, los indicadores más utilizados son el índice h y el índice g.
- El índice h refleja el equilibrio entre productividad y citación: un autor tiene índice h si ha publicado h artículos con al menos h citas cada uno.
- El índice g, propuesto por Egghe, da mayor peso a los artículos más citados.
- Existen variantes como h5, hm, hg, hl o hla, aplicadas para ajustar diferencias de edad académica o coautoría.
- Para las revistas científicas, los indicadores más comunes incluyen FI, SJR,

*CiteScore*, SNIP y *Eigenfactor*.

- El *Eigenfactor* mide la importancia de una revista considerando la estructura completa de la red de citaciones.
- Los índices de colaboración se calculan mediante coautorías, redes institucionales y análisis de afiliaciones.
- En evaluación institucional o nacional se usan métricas agregadas (por ejemplo, número total de citas o publicaciones por campo).
- Los indicadores deben interpretarse en contexto, evitando comparaciones entre disciplinas con diferentes hábitos de citación.
- Se recomienda el uso de indicadores normalizados y métricas complementarias (altmétricas, descargas, visibilidad social).
- Los índices bibliométricos, aunque útiles, presentan limitaciones éticas y metodológicas cuando se usan como únicos criterios de evaluación.
- Una evaluación responsable combina indicadores cuantitativos y revisión cualitativa, conforme a la Declaración de San Francisco (DORA) y el Leiden Manifesto.

### Preguntas de autoevaluación

1. ¿Cuál es la función principal de los índices bibliométricos?
2. ¿En qué tres grandes categorías se agrupan los indicadores bibliométricos?
3. ¿Quién creó el Factor de Impacto y cuál es su principio básico?
4. ¿Qué diferencia al *CiteScore* del Factor de Impacto?
5. ¿Cómo pondera las citas el indicador SJR?
6. ¿Qué mide el índice h y cómo se interpreta su valor?
7. ¿Qué aporta el índice g respecto al índice h?
8. ¿Qué caracteriza al indicador SNIP en comparación con otros índices?
9. ¿Qué principios promueven DORA y el Leiden Manifesto en la evaluación científica?
10. ¿Por qué es necesario contextualizar los resultados de los índices bibliométricos?

### BIBLIOGRAFÍA

1. Moed HF. Citation analysis in research evaluation. Dordrecht: Springer; 2005. <https://doi.org/10.1007/1-4020-3714-7>
2. Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. *J Doc.* 2008;64(1):45-80. <https://doi.org/10.1108/00220410810844150>
3. Gingras Y. Bibliometrics and research evaluation: uses and abuses. Cambridge (MA): MIT Press; 2016. ISBN 9780262337663
4. Egghe L. Theory and practice of the g-index. *Scientometrics.* 2006;69(1):131-52. <https://doi.org/10.1007/s11192-006-0144-7>
5. Leydesdorff L, Bornmann L, Mutz R, Opthof T. Turning the tables on citation analysis one more time: principles for comparing sets of documents. *J Informetrics.* 2019;13(3):1020-30. <https://doi.org/10.1016/j.joi.2019.05.010>

### REFERENCIAS BIBLIOGRÁFICAS

1. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing.* 2015;28(2):151-5. <https://doi.org/10.1087/20150211>

2. Bornmann L, Bauer J. Evaluation of the highly-cited researchers' database for a country: proposals for meaningful analyses on the example of Germany. *Scientometrics*. 2015;105(3):1997-2003. <https://doi.org/10.1007/s11192-015-1754-3>
3. Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*. 2005;102(46):16569-72. <https://doi.org/10.1073/pnas.0507655102>
4. Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AFJ. Towards a new crown indicator: some theoretical considerations. *Journal of Informetrics*. 2011;5(1):37-47. <https://doi.org/10.1016/j.joi.2010.08.001>
5. Stirling A. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*. 2007;4(15):707-19. <https://doi.org/10.1098/rsif.2007.0213>
6. Purkayastha A, Palmaro E, Falk-Krzesinski HJ, Baas J. Comparison of two article-level, field-independent citation metrics: field-weighted citation impact (FWCI) and relative citation ratio (RCR). *Journal of Informetrics*. 2019;13(2):635-42. <https://doi.org/10.1016/j.joi.2019.02.005>
7. Elsevier. SciVal metric - Pure. 2025. [https://helpcenter.pure.elsevier.com/en\\_US/metrics-in-pure/scival-metric](https://helpcenter.pure.elsevier.com/en_US/metrics-in-pure/scival-metric)
8. Kulczycki E, Engels TCE, Pölönen J, Bruun K, Dušková M, Guns R, et al. Publication patterns in the social sciences and humanities: evidence from eight European countries. *Scientometrics*. 2018;116(1):463-86. <https://doi.org/10.1007/s11192-018-2711-0>

## Chapter 6 / Capítulo 6

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

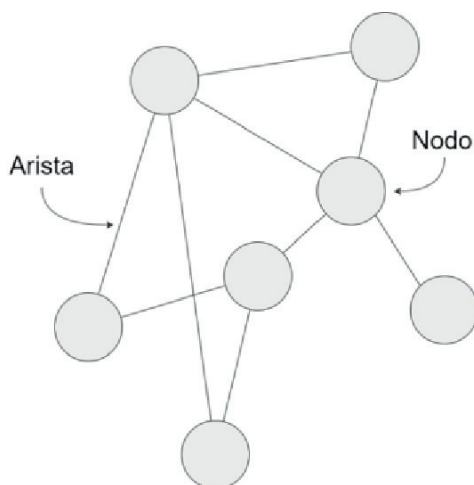
DOI: 10.62486/978-9915-9680-6-3.ch06

©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





## 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.

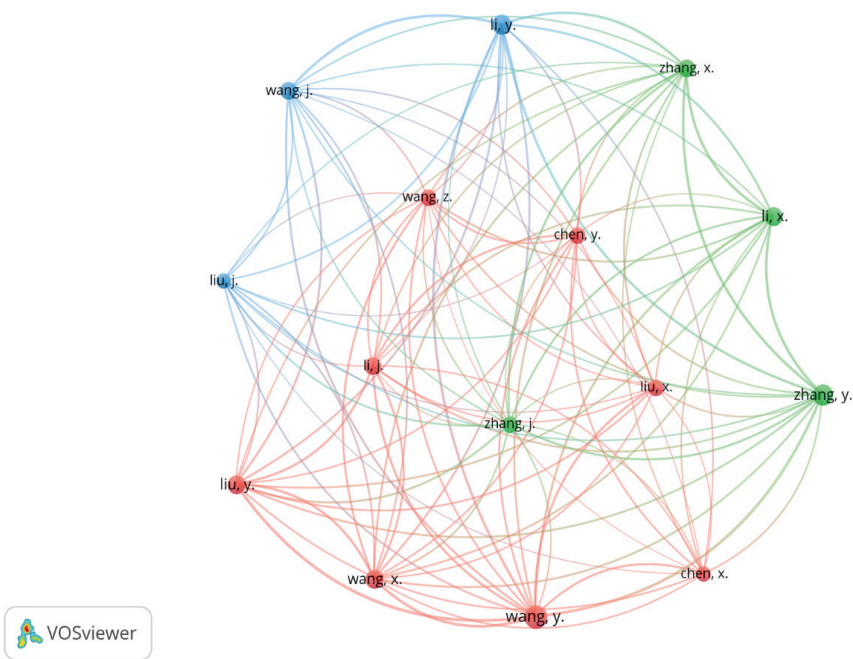


Figura 6.3. Ejemplo de grafo de co-citación

#### **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

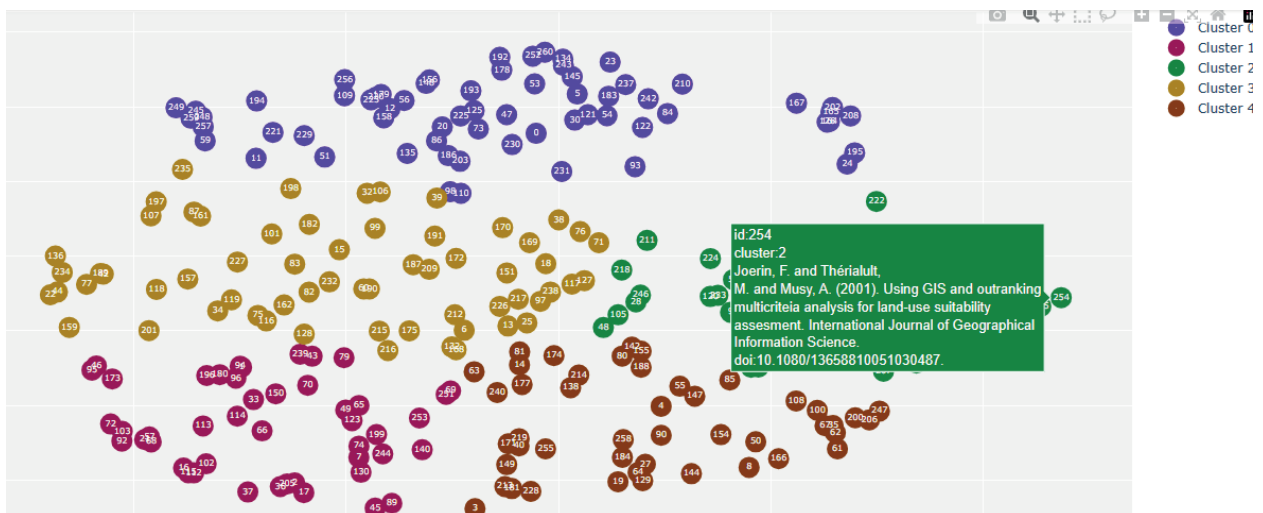
1. Descargar e instalar CiteSpace con Java Runtime Environment actualizado
2. Crear nuevo proyecto y configurar estructura de directorios para datos y resultados
3. Importar y convertir datos bibliográficos desde WoS o Scopus al formato nativo de CiteSpace
4. Configurar exhaustivamente los parámetros 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.

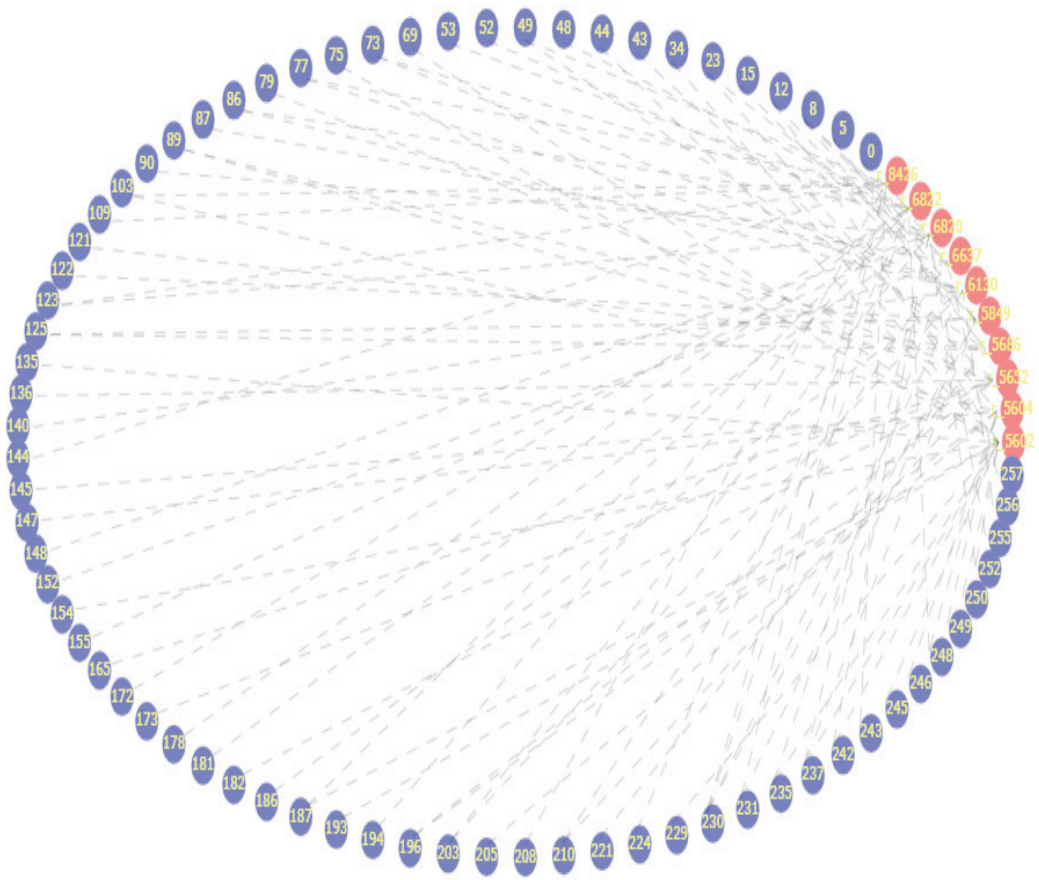


Fuente: Elaboración propia (Pybibx)

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

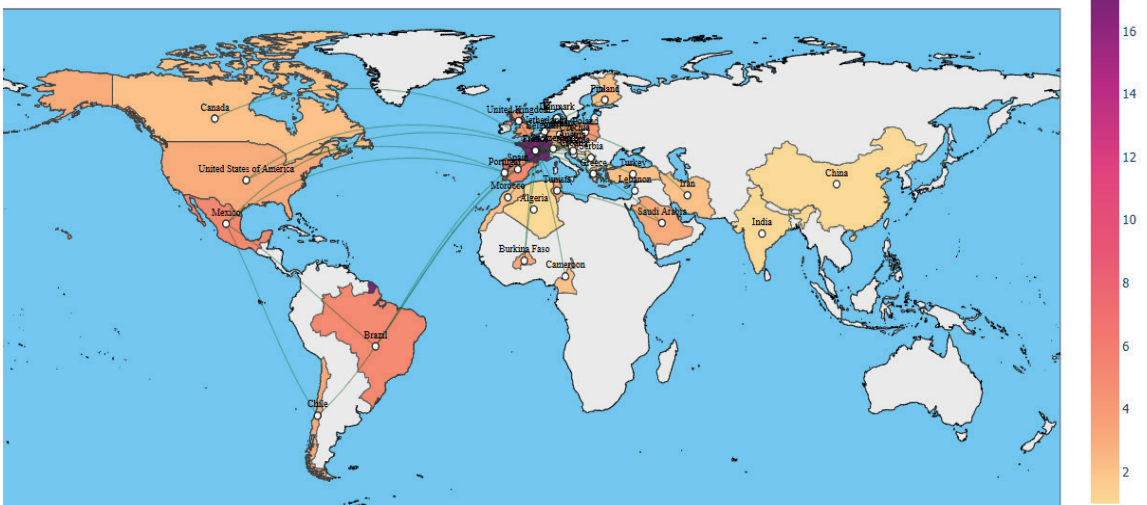


Más allá de las representaciones sin aristas, estas herramientas también permiten generar grafos relacionales clásicos. En estos grafos, las conexiones o aristas entre los nodos sí están explícitamente definidas, representando vínculos de citación o co-citación entre los artículos individuales. A diferencia 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)`

### Keyword Co-occurrences

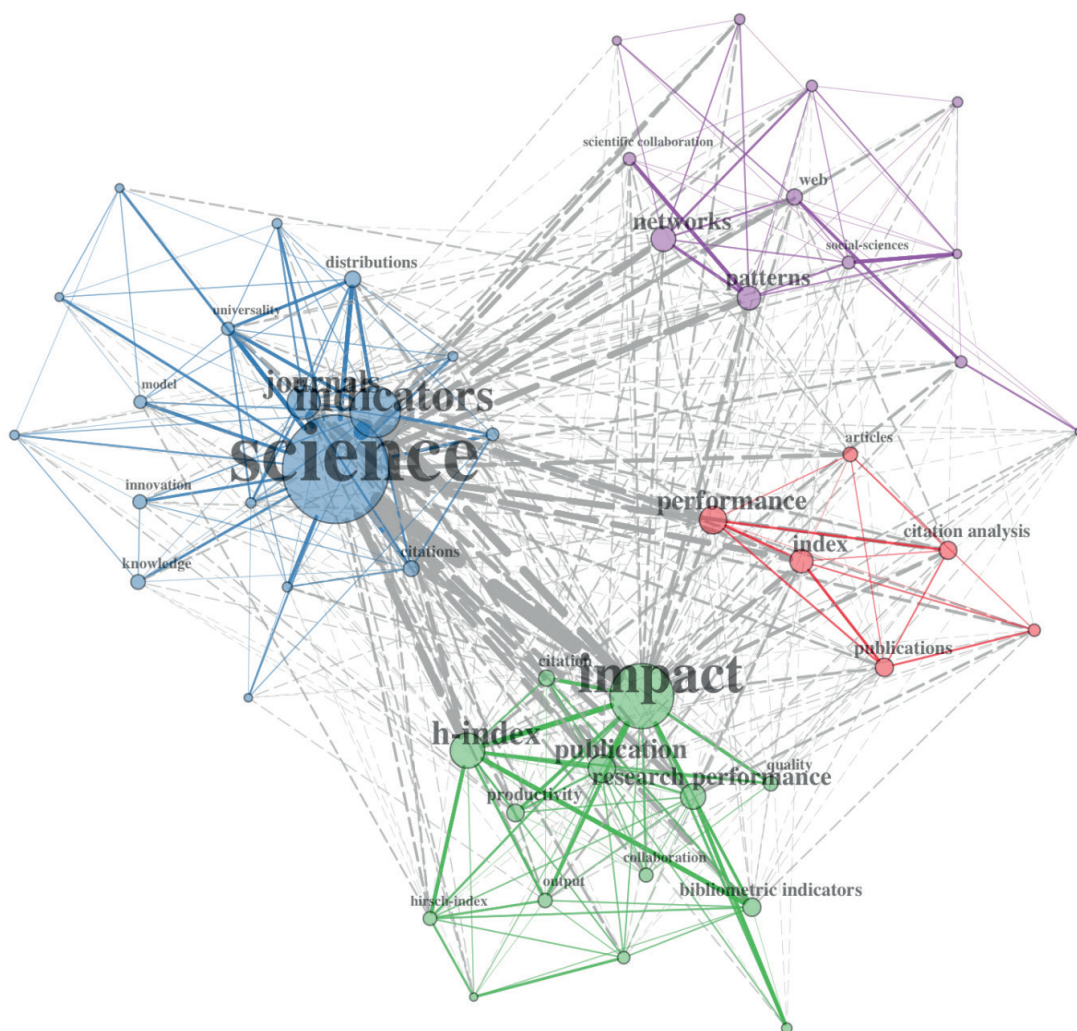


Figura 6.6. Ejemplo de grafo de co-ocurrencia

## **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.

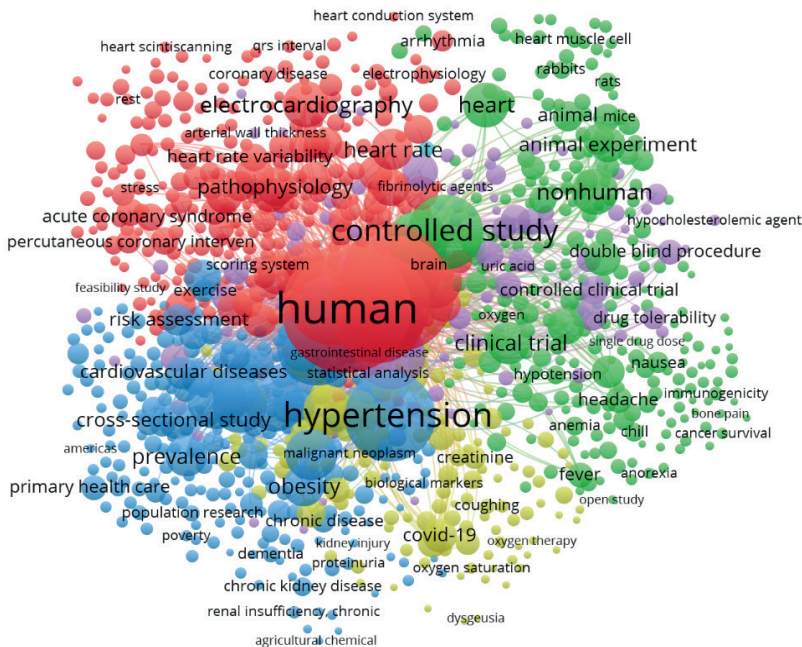
La interpretación final, por tanto, representa una síntesis creativa donde el rigor métrico se combina con la sensibilidad disciplinar y la capacidad inferencial del investigador. Este proceso transforma datos relacionales en conocimiento accionable, identificando oportunidades de investigación, revelando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



## Chapter 7 / Capítulo 7

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch07

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

Los mapas de calor constituyen una herramienta visual poderosa que transforma datos



1.  $\frac{1}{2}$  2.  $\frac{1}{3}$  3.  $\frac{1}{4}$  4.  $\frac{1}{5}$  5.  $\frac{1}{6}$  6.  $\frac{1}{7}$  7.  $\frac{1}{8}$  8.  $\frac{1}{9}$  9.  $\frac{1}{10}$  10.  $\frac{1}{11}$  11.  $\frac{1}{12}$  12.  $\frac{1}{13}$  13.  $\frac{1}{14}$  14.  $\frac{1}{15}$  15.  $\frac{1}{16}$  16.  $\frac{1}{17}$  17.  $\frac{1}{18}$  18.  $\frac{1}{19}$  19.  $\frac{1}{20}$  20.  $\frac{1}{21}$  21.  $\frac{1}{22}$  22.  $\frac{1}{23}$  23.  $\frac{1}{24}$  24.  $\frac{1}{25}$  25.  $\frac{1}{26}$  26.  $\frac{1}{27}$  27.  $\frac{1}{28}$  28.  $\frac{1}{29}$  29.  $\frac{1}{30}$  30.  $\frac{1}{31}$  31.  $\frac{1}{32}$  32.  $\frac{1}{33}$  33.  $\frac{1}{34}$  34.  $\frac{1}{35}$  35.  $\frac{1}{36}$  36.  $\frac{1}{37}$  37.  $\frac{1}{38}$  38.  $\frac{1}{39}$  39.  $\frac{1}{40}$  40.  $\frac{1}{41}$  41.  $\frac{1}{42}$  42.  $\frac{1}{43}$  43.  $\frac{1}{44}$  44.  $\frac{1}{45}$  45.  $\frac{1}{46}$  46.  $\frac{1}{47}$  47.  $\frac{1}{48}$  48.  $\frac{1}{49}$  49.  $\frac{1}{50}$  50.  $\frac{1}{51}$  51.  $\frac{1}{52}$  52.  $\frac{1}{53}$  53.  $\frac{1}{54}$  54.  $\frac{1}{55}$  55.  $\frac{1}{56}$  56.  $\frac{1}{57}$  57.  $\frac{1}{58}$  58.  $\frac{1}{59}$  59.  $\frac{1}{60}$  60.  $\frac{1}{61}$  61.  $\frac{1}{62}$  62.  $\frac{1}{63}$  63.  $\frac{1}{64}$  64.  $\frac{1}{65}$  65.  $\frac{1}{66}$  66.  $\frac{1}{67}$  67.  $\frac{1}{68}$  68.  $\frac{1}{69}$  69.  $\frac{1}{70}$  70.  $\frac{1}{71}$  71.  $\frac{1}{72}$  72.  $\frac{1}{73}$  73.  $\frac{1}{74}$  74.  $\frac{1}{75}$  75.  $\frac{1}{76}$  76.  $\frac{1}{77}$  77.  $\frac{1}{78}$  78.  $\frac{1}{79}$  79.  $\frac{1}{80}$  80.  $\frac{1}{81}$  81.  $\frac{1}{82}$  82.  $\frac{1}{83}$  83.  $\frac{1}{84}$  84.  $\frac{1}{85}$  85.  $\frac{1}{86}$  86.  $\frac{1}{87}$  87.  $\frac{1}{88}$  88.  $\frac{1}{89}$  89.  $\frac{1}{90}$  90.  $\frac{1}{91}$  91.  $\frac{1}{92}$  92.  $\frac{1}{93}$  93.  $\frac{1}{94}$  94.  $\frac{1}{95}$  95.  $\frac{1}{96}$  96.  $\frac{1}{97}$  97.  $\frac{1}{98}$  98.  $\frac{1}{99}$  99.  $\frac{1}{100}$  100.  $\frac{1}{101}$  101.  $\frac{1}{102}$  102.  $\frac{1}{103}$  103.  $\frac{1}{104}$  104.  $\frac{1}{105}$  105.  $\frac{1}{106}$  106.  $\frac{1}{107}$  107.  $\frac{1}{108}$  108.  $\frac{1}{109}$  109.  $\frac{1}{110}$  110.  $\frac{1}{111}$  111.  $\frac{1}{112}$  112.  $\frac{1}{113}$  113.  $\frac{1}{114}$  114.  $\frac{1}{115}$  115.  $\frac{1}{116}$  116.  $\frac{1}{117}$  117.  $\frac{1}{118}$  118.  $\frac{1}{119}$  119.  $\frac{1}{120}$  120.  $\frac{1}{121}$  121.  $\frac{1}{122}$  122.  $\frac{1}{123}$  123.  $\frac{1}{124}$  124.  $\frac{1}{125}$  125.  $\frac{1}{126}$  126.  $\frac{1}{127}$  127.  $\frac{1}{128}$  128.  $\frac{1}{129}$  129.  $\frac{1}{130}$  130.  $\frac{1}{131}$  131.  $\frac{1}{132}$  132.  $\frac{1}{133}$  133.  $\frac{1}{134}$  134.  $\frac{1}{135}$  135.  $\frac{1}{136}$  136.  $\frac{1}{137}$  137.  $\frac{1}{138}$  138.  $\frac{1}{139}$  139.  $\frac{1}{140}$  140.  $\frac{1}{141}$  141.  $\frac{1}{142}$  142.  $\frac{1}{143}$  143.  $\frac{1}{144}$  144.  $\frac{1}{145}$  145.  $\frac{1}{146}$  146.  $\frac{1}{147}$  147.  $\frac{1}{148}$  148.  $\frac{1}{149}$  149.  $\frac{1}{150}$  150.  $\frac{1}{151}$  151.  $\frac{1}{152}$  152.  $\frac{1}{153}$  153.  $\frac{1}{154}$  154.  $\frac{1}{155}$  155.  $\frac{1}{156}$  156.  $\frac{1}{157}$  157.  $\frac{1}{158}$  158.  $\frac{1}{159}$  159.  $\frac{1}{160}$  160.  $\frac{1}{161}$  161.  $\frac{1}{162}$  162.  $\frac{1}{163}$  163.  $\frac{1}{164}$  164.  $\frac{1}{165}$  165.  $\frac{1}{166}$  166.  $\frac{1}{167}$  167.  $\frac{1}{168}$  168.  $\frac{1}{169}$  169.  $\frac{1}{170}$  170.  $\frac{1}{171}$  171.  $\frac{1}{172}$  172.  $\frac{1}{173}$  173.  $\frac{1}{174}$  174.  $\frac{1}{175}$  175.  $\frac{1}{176}$  176.  $\frac{1}{177}$  177.  $\frac{1}{178}$  178.  $\frac{1}{179}$  179.  $\frac{1}{180}$  180.  $\frac{1}{181}$  181.  $\frac{1}{182}$  182.  $\frac{1}{183}$  183.  $\frac{1}{184}$  184.  $\frac{1}{185}$  185.  $\frac{1}{186}$  186.  $\frac{1}{187}$  187.  $\frac{1}{188}$  188.  $\frac{1}{189}$  189.  $\frac{1}{190}$  190.  $\frac{1}{191}$  191.  $\frac{1}{192}$  192.  $\frac{1}{193}$  193.  $\frac{1}{194}$  194.  $\frac{1}{195}$  195.  $\frac{1}{196}$  196.  $\frac{1}{197}$  197.  $\frac{1}{198}$  198.  $\frac{1}{199}$  199.  $\frac{1}{200}$  200.  $\frac{1}{201}$  201.  $\frac{1}{202}$  202.  $\frac{1}{203}$  203.  $\frac{1}{204}$  204.  $\frac{1}{205}$  205.  $\frac{1}{206}$  206.  $\frac{1}{207}$  207.  $\frac{1}{208}$  208.  $\frac{1}{209}$  209.  $\frac{1}{210}$  210.  $\frac{1}{211}$  211.  $\frac{1}{212}$  212.  $\frac{1}{213}$  213.  $\frac{1}{214}$  214.  $\frac{1}{215}$  215.  $\frac{1}{216}$  216.  $\frac{1}{217}$  217.  $\frac{1}{218}$  218.  $\frac{1}{219}$  219.  $\frac{1}{220}$  220.  $\frac{1}{221}$  221.  $\frac{1}{222}$  222.  $\frac{1}{223}$  223.  $\frac{1}{224}$  224.  $\frac{1}{225}$  225.  $\frac{1}{226}$  226.  $\frac{1}{227}$  227.  $\frac{1}{228}$  228.  $\frac{1}{229}$  229.  $\frac{1}{230}$  230.  $\frac{1}{231}$  231.  $\frac{1}{232}$  232.  $\frac{1}{233}$  233.  $\frac{1}{234}$  234.  $\frac{1}{235}$  235.  $\frac{1}{236}$  236.  $\frac{1}{237}$  237.  $\frac{1}{238}$  238.  $\frac{1}{239}$  239.  $\frac{1}{240}$  240.

con redes estructurales, revelando cómo las áreas de alta densidad se relacionan con nodos centrales y puentes interdisciplinarios.

La interpretación contextualizada trasciende el análisis visual inmediato para integrar factores socio-institucionales, epistemológicos y de política científica que explican los patrones observados. Un área de alta densidad investigadora puede reflejar tanto el vigor intelectual de un campo promisorio como el efecto de financiamiento concentrado o modas académicas pasajeras. Las zonas frías pueden indicar tanto fronteras no exploradas con potencial innovador como callejones sin salida intelectuales o áreas con barreras de entrada metodológicas. El analista experto distingue estas posibilidades mediante triangulación con otras fuentes de evidencia y conocimiento disciplinar especializado, transformando el mapa de calor de mera descripción distribucional en herramienta diagnóstica para la planificación estratégica de la investigación.

## **7.2 Algoritmo de construcción de mapas de calor bibliométricos**

La construcción metodológicamente rigurosa de mapas de calor bibliométricos inicia con la definición precisa de las dimensiones a representar, típicamente ejes temáticos, temporales, geográficos o institucionales, y la selección de la métrica de intensidad apropiada para el objetivo investigador. Las opciones métricas abarcan desde indicadores básicos de productividad (número de publicaciones) hasta medidas sofisticadas de impacto (citaciones normalizadas por campo) o especialización (índices de concentración temática). La elección de esta métrica determina fundamentalmente el tipo de patrones que el mapa revelará, requiriendo alineación cuidadosa con las preguntas de investigación que motivan el análisis.

El procesamiento de datos para construcción de mapas de calor involucra etapas sucesivas de agregación, normalización y suavizado. La agregación transforma datos individuales de publicaciones en valores resumidos para las celdas de la matriz multidimensional que constituirá el mapa. La normalización ajusta estos valores crudos para permitir comparaciones significativas entre dominios con diferentes tamaños, prácticas de publicación o tradiciones citacionales, esencial cuando se analizan campos interdisciplinarios. El suavizado aplica algoritmos de interpolación para crear transiciones graduales entre celdas adyacentes, mejorando la legibilidad visual, pero introduciendo potenciales artefactos que el analista debe reconocer y controlar.

La codificación cromática representa la etapa donde los valores numéricos se transforman en experiencias visuales interpretables. La selección de paletas de color debe considerar principios de percepción visual, accesibilidad para personas con daltonismo, y convenciones disciplinares establecidas. Las paletas secuenciales, con variaciones de luminosidad de un solo color, son ideales para representar magnitudes unidireccionales, mientras las paletas divergentes, con dos colores contrastantes, capturan adecuadamente desviaciones de un punto de referencia central. La elección de puntos de quiebre entre intervalos de color puede destacar u ocultar patrones críticos, requiriendo justificación metodológica explícita basada en distribuciones estadísticas naturales de los datos o umbrales sustantivos significativos para el campo de estudio.

La validación interpretativa cierra el proceso constructivo, asegurando que los patrones visuales emergentes correspondan a fenómenos reales del ecosistema investigador y no a artefactos metodológicos. Esta validación involucra pruebas de sensibilidad a decisiones técnicas (umbrales, algoritmos de suavizado, paletas), triangulación con otras representaciones (redes, series temporales), y contraste con conocimiento experto del dominio. El mapa de calor maduro trasciende así su función descriptiva inicial para convertirse en interface interactiva de exploración analítica, una herramienta dinámica que permite formular y verificar hipótesis sobre

la estructura y evolución del conocimiento científico mediante la manipulación sistemática de parámetros de visualización y la iteración entre diferentes escalas de análisis.

### 7.3. Creación de mapas de calor bibliométricos

La creación de mapas de densidad en VOSviewer se realiza de manera integrada con la generación de grafos, requiriendo simplemente cambiar la visualización en la pestaña “Items” a la opción “*Density Visualization*”. Esta transición transforma la representación de red en un mapa de calor donde las áreas de color intenso indican regiones de alta concentración de elementos, manteniendo la misma disposición espacial obtenida en el análisis de redes. La densidad se calcula mediante funciones esenciales que suavizan la distribución puntual de los nodos, creando gradientes continuos donde el color, desde azul (baja densidad) hasta rojo (alta densidad), revela aglomeraciones temáticas o colaborativas. Los usuarios pueden ajustar la sensibilidad del mapa mediante el parámetro de suavizado y personalizar la paleta cromática para optimizar la legibilidad según las características específicas del set de datos analizado.

En CitNetExplorer, los mapas de densidad se generan mediante la función “*Clúster Density Visualization*” que permite observar la concentración de publicaciones en diferentes regiones de la red de citación. El proceso implica calcular la densidad de conexiones alrededor de cada nodo y representarla mediante un mapa térmico donde las áreas de mayor actividad investigadora aparecen en tonos cálidos. Esta aproximación resulta particularmente valiosa para identificar períodos de intensa actividad citacional dentro de trayectorias de investigación específicas, revelando momentos paradigmáticos en la evolución de campos científicos.

CiteSpace implementa mapas de densidad a través de su módulo “*Spectral Density Mapping*” que combina algoritmos de detección espectral con representaciones térmicas. El usuario debe activar la opción “*Show Density*” en el panel de control de visualización, lo que genera una capa adicional de color que se superpone al mapa de redes convencional. Esta herramienta permite identificar clústeres de alta cohesión interna y fronteras entre diferentes escuelas de pensamiento, con la particularidad de que la densidad se calcula considerando simultáneamente la proximidad espacial y la similitud temática entre elementos.

En el ecosistema de R con **Bibliometrix**, la creación de mapas de densidad se realiza mediante la función ***termDensity()*** para análisis temáticos o ***authorDensity()*** para estudios de colaboración. El proceso requiere primero generar la matriz de co-ocurrencia o colaboración y luego aplicar la función ***heatmap()*** sobre la matriz normalizada. La personalización avanzada incluye ajustar parámetros de suavizado mediante núcleos gaussianos y definir paletas de color específicas para diferentes rangos de densidad. Esta aproximación basada en programación ofrece máxima flexibilidad analítica, pero requiere competencias técnicas en manipulación de matrices y visualización en R.

PyBibX en Python aborda los mapas de densidad mediante su módulo ***density\_analysis*** que implementa algoritmos de estimación de densidad por núcleo sobre distribuciones bibliométricas. El flujo de trabajo típico implica cargar el set de datos bibliográficos, calcular las coordenadas multidimensionales mediante reducción de dimensionalidad, y luego generar el mapa de calor con ***plot\_density\_map()***. La librería permite ajustar el ancho de banda del núcleo para controlar el nivel de suavizado y exportar los mapas en formatos vectoriales para publicaciones de alta calidad. Esta implementación resulta especialmente poderosa para análisis de grandes volúmenes de datos donde se requiere procesamiento eficiente y personalización avanzada de visualizaciones.

## **Recapitulando**

- Los mapas de calor (*heat maps*) son representaciones gráficas que muestran la intensidad o densidad de un fenómeno mediante gradientes de color.
- En bibliometría, permiten visualizar zonas temáticas con mayor o menor concentración de publicaciones o citas.
- Se basan en la distribución de frecuencias o correlaciones dentro de matrices de coocurrencia, co-citación o colaboración.
- Cada celda o punto del mapa refleja el nivel de actividad investigadora, medido por número de documentos, citas o vínculos entre términos.
- El color actúa como variable visual: los tonos más cálidos (rojos, naranjas) indican mayor densidad, y los más fríos (azules, verdes) menor actividad.
- Los mapas de calor son útiles para detectar áreas emergentes, consolidadas o en declive dentro de un campo científico.
- También permiten observar la interrelación entre temas o autores, revelando clústeres de alta concentración investigadora.
- Se aplican tanto en análisis temático (palabras clave) como en análisis de colaboración (autores, países, instituciones).
- Los datos de entrada se obtienen de bases como Scopus, Web of Science o Dimensions, exportados en formato compatible (CSV, RIS, BibTeX).
- La elaboración de un mapa de calor requiere una matriz numérica de similitud o frecuencia.
- El proceso incluye normalización de datos, elección de la escala cromática y configuración del rango de densidad.
- Las herramientas más utilizadas para crear mapas de calor bibliométricos son VOSviewer, Bibliometrix (R/Biblioshiny), CiteSpace, Gephi y Excel con complementos estadísticos.
- En VOSviewer, los mapas de densidad representan zonas de mayor concentración de elementos (nodos) con colores más intensos.
- Los mapas de densidad temática facilitan la identificación de núcleos conceptuales y relaciones entre áreas del conocimiento.
- Los mapas institucionales o geográficos muestran la distribución espacial de la producción científica y colaboración internacional.
- Estos mapas permiten comparar el desempeño científico entre regiones o campos disciplinarios.
- El uso de escalas adecuadas evita distorsiones visuales y mejora la interpretación de la densidad de datos.
- Los mapas de calor deben ir acompañados de leyendas claras y metadatos completos que expliquen los criterios de color y escala.
- Una interpretación rigurosa combina la lectura visual con el análisis estadístico complementario (frecuencias, correlaciones, centralidades).
- Bien empleados, los mapas de calor se convierten en una herramienta poderosa para comunicar resultados bibliométricos de manera visual, intuitiva y comparativa.

## **Preguntas de autoevaluación**

1. ¿Qué representan los colores en un mapa de calor aplicado a bibliometría?
2. ¿Qué tipo de información cuantitativa se visualiza en un heat map bibliométrico?
3. ¿Qué diferencia hay entre un mapa de calor temático y uno institucional?
4. ¿Qué bases de datos suelen utilizarse para generar los datos de entrada?
5. ¿Qué papel cumple la normalización en la elaboración del mapa?
6. ¿Qué programas permiten crear mapas de calor de densidad investigadora?

7. ¿Cómo interpreta VOSviewer las zonas más cálidas del mapa?
8. ¿Qué precauciones deben tomarse al elegir la escala de color?
9. ¿Por qué es importante incluir una leyenda explicativa y metadatos?
10. ¿Qué aporta el uso de mapas de calor a la comprensión visual de la producción científica?

## **BIBLIOGRAFÍA**

1. 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>
2. Börner K. Atlas of science: visualizing what we know. Cambridge (MA): MIT Press; 2010. ISBN 9780262014454
3. Börner K, Polley DE. Visual perspectives: a practical guide to making sense of data. Cambridge (MA): MIT Press; 2014. ISBN 9780262027898
4. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84(2):523-38. <https://doi.org/10.1007/s11192-009-0146-3>
5. Klavans R, Boyack KW. Toward a consensus map of science. *J Assoc Inf Sci Technol*. 2009;60(3):455-76. <https://doi.org/10.1002/asi.20991>



## 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.

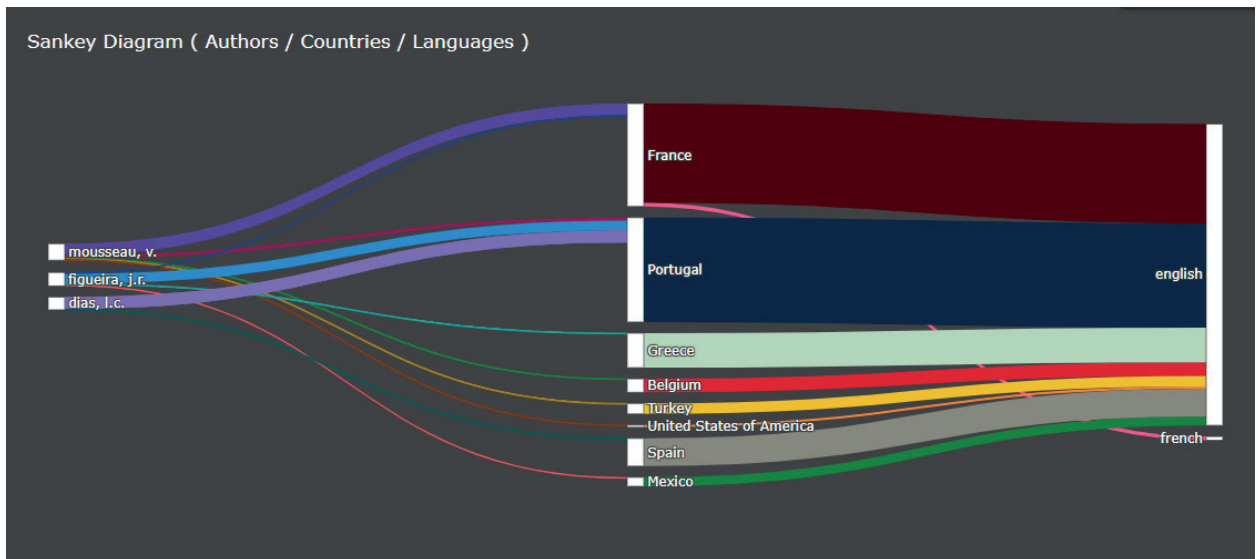


Figura 8.1. Ejemplo de diagrama de Sankey

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 disciplinares.

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 disciplinares 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 disciplinares. 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 disciplinares 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, hospitales) 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. Knaflíc CN. Storytelling with data: a data visualization guide for business professionals. Hoboken (NJ): Wiley; 2015. <https://doi.org/10.1002/9781119055259>





## Chapter 9 / Capítulo 9

*Applied bibliometrics. From data to publication (Spanish Edition)*

**ISBN:** 978-9915-9680-6-3

**DOI:** 10.62486/978-9915-9680-6-3.ch09

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

## OTHER RESULTS / OTROS RESULTADOS

### *Análisis de dimensiones complementarias en bibliometría*

Más allá de los indicadores convencionales de producción e impacto, el análisis bibliométrico engloba dimensiones contextuales esenciales que revelan patrones estructurales del ecosistema científico. Este capítulo examina tres dimensiones frecuentemente subestimadas pero críticas para una comprensión integral de la dinámica investigadora: la distribución lingüística de la producción científica, el rol de las revistas como canales de comunicación especializada, y el posicionamiento geopolítico del conocimiento. Cada una de estas dimensiones aporta perspectivas únicas sobre cómo se organiza, comunica y distribuye el conocimiento científico a través de diferentes barreras culturales, institucionales y geográficas.

### **9.1. Interpretación del idioma en la producción científica**

El análisis de la distribución lingüística en la literatura científica revela profundas asimetrías en la geopolítica del conocimiento. El dominio absoluto del inglés como *lingua franca científica*, que típicamente concentra entre el 80 % y 95 % de las publicaciones indexadas en bases de datos internacionales, refleja no solo prácticas de estandarización comunicativa sino también estructuras de poder académico consolidadas. Esta hegemonía lingüística tiene implicaciones significativas para la visibilidad, accesibilidad e impacto de la investigación producida en otros idiomas, creando un sesgo sistemático que margina contribuciones científicas valiosas desarrolladas en contextos lingüísticos periféricos. La interpretación de estos patrones requiere contextualización histórica que considere cómo la expansión del inglés como idioma científico se vinculó con procesos de globalización académica y con la concentración de recursos editoriales en países angloparlantes.<sup>(1)</sup>

La distribución multilingüe de la producción científica varía sustancialmente entre disciplinas, reflejando tradiciones intelectuales y audiencias objetivo diferenciadas. En campos como la medicina clínica o las ciencias ambientales, donde la investigación tiene aplicación local inmediata, es más frecuente encontrar publicaciones en idiomas distintos al inglés que comunican hallazgos relevantes para contextos específicos. Por el contrario, en disciplinas como la física o las matemáticas, la estandarización en inglés es casi absoluta. Estas diferencias disciplinarias tienen implicaciones importantes para las políticas de evaluación científica, que deben reconocer la legitimidad de publicaciones en múltiples idiomas cuando responden a necesidades comunicativas específicas de cada campo de conocimiento.

La evolución temporal de los patrones lingüísticos muestra tendencias contradictorias: por un lado, una consolidación progresiva del inglés como idioma científico universal; por otro, resurgencias periódicas de publicaciones en idiomas locales impulsadas por políticas de ciencia abierta y reconocimiento de la diversidad epistemológica. El análisis diacrónico de estas tendencias permite identificar momentos de cambio en las prácticas comunicativas científicas, usualmente vinculados con transformaciones en las políticas editoriales, el surgimiento de repositorios multilingües, o cambios en los sistemas de evaluación académica que valoran diferencialmente las publicaciones en diferentes idiomas.

### **9.2. Interpretación del rol de las revistas científicas**

Las revistas académicas constituyen instituciones fundamentales que estructuran los campos científicos mediante sus funciones de validación, diseminación y construcción de comunidades. El análisis bibliométrico de las revistas trasciende las métricas de impacto simplistas para examinar cómo estas configuran trayectorias intelectuales, consagran marcos teóricos y establecen fronteras disciplinarias. La distribución de la producción científica típicamente

sigue patrones de concentración donde un núcleo pequeño de revistas captura atención desproporcionada, mientras una periferia extensa alberga contribuciones especializadas con audiencias segmentadas.

La arquitectura invisible de los campos científicos se revela mediante el mapeo de redes de revistas a través de análisis de co-citación y acoplamiento bibliográfico. Estos análisis identifican clústeres de publicaciones que representan subcampos epistemológicos o tradiciones metodológicas distintivas, mostrando cómo se organiza la comunicación especializada dentro de cada disciplina. La posición relativa de una revista dentro de estas redes proporciona perspectivas más matizadas sobre su influencia intelectual que cualquier métrica unidimensional, revelando su papel como puente entre comunidades, bastión de ortodoxias o espacio de innovación disruptiva.

La ecología de revistas está experimentando transformaciones profundas impulsadas por la digitalización, la ciencia abierta y nuevas economías de la atención académica. El surgimiento de mega revistas interdisciplinarias desafía taxonomías disciplinarias establecidas, mientras la fragmentación temática crea espacios para publicaciones ultraespecializadas. Simultáneamente, la transición hacia modelos de acceso abierto está reconfigurando relaciones de poder entre editores, autores, instituciones y sociedades científicas, con consecuencias aún inciertas para la calidad, diversidad y sostenibilidad de la comunicación académica.

Los indicadores de salud del ecosistema de revistas deben considerar dimensiones más allá del impacto citacional, incluyendo diversidad geográfica y lingüística, equidad en procesos editoriales, y sostenibilidad económica. El análisis bibliométrico avanzado permite diagnosticar distorsiones como la endogamia citacional, la homogenización temática y la concentración excesiva, informando políticas para un ecosistema de publicaciones más robusto, diverso y responsivo a las necesidades evolutivas de las comunidades científicas.

### **9.3. Interpretación de la distribución por países**

La distribución geográfica de la producción científica revela profundas desigualdades en la capacidad investigadora a nivel global, con un grupo pequeño de países concentrando la mayoría de las publicaciones y citas de alto impacto. Estos patrones de concentración reflejan históricas asimetrías en la distribución de recursos para investigación, infraestructura científica capitalizada y capacidad de formación de capital humano avanzado. Sin embargo, el análisis diacrónico muestra procesos de convergencia gradual impulsados por políticas científicas agresivas en economías emergentes y por la creciente internacionalización de la colaboración investigadora.

Los patrones de especialización temática por países constituyen una dimensión analítica particularmente reveladora sobre ventajas comparativas en diferentes sistemas nacionales de innovación. Algunos países desarrollan perfiles de publicación marcadamente especializados en campos donde poseen ventajas competitivas basadas en recursos naturales, tradiciones intelectuales consolidadas o clústeres industriales específicos. Otros muestran perfiles más diversificados que reflejan estrategias deliberadas de desarrollo de capacidades científicas generales. El análisis de estas especializaciones relativas permite identificar nichos de excelencia y oportunidades para colaboraciones internacionales complementarias.

Las redes de colaboración internacional representan quizás la dimensión más dinámica de la geopolítica del conocimiento contemporáneo. El análisis bibliométrico de coautorías internacionales revela patrones complejos de asociación preferencial que típicamente reflejan

proximidades lingüísticas, históricas o geográficas, pero también la emergencia de nuevas alianzas estratégicas basadas en complementariedades temáticas o recursos compartidos. La posición de los países dentro de estas redes globales de colaboración, como núcleos centrales, puentes periféricos o actores aislados, condiciona significativamente su capacidad para acceder a fronteras del conocimiento y participar en innovaciones transformadoras.

#### **9.4. Interpretación del género en la autoría científica**

El análisis de género en la autoría científica revela patrones profundos sobre la participación diferencial en la producción de conocimiento. Las distribuciones por género varían significativamente entre disciplinas, reflejando historias particulares de inclusión/exclusión, culturas disciplinares específicas y diferentes trayectorias de profesionalización. En campos como la enfermería o la educación, la autoría femenina suele ser mayoritaria, mientras en ingenierías o física persisten marcadas subrepresentaciones. Estas diferencias disciplinarias deben interpretarse considerando factores históricos, sociales e institucionales que han moldeado el acceso y permanencia de diferentes géneros en cada campo.

La posición en la autoría constituye una dimensión analítica crucial, ya que refleja jerarquías en la contribución intelectual y liderazgo en proyectos investigadores. Estudios consistentes muestran que las mujeres están sobrerrepresentadas en posiciones intermedias de autoría y subrepresentadas como primeras autoras o autoras de correspondencia, particularmente en colaboraciones internacionales de alto impacto. Estas disparidades señalan posibles barreras en el reconocimiento de liderazgo científico y en la asignación de roles centrales en proyectos de investigación, aspectos que merecen atención en políticas de equidad género en ciencia.

La evolución temporal de la participación por género muestra trayectorias complejas con progresos notables pero también persistentes estancamientos. Mientras algunas disciplinas han experimentado convergencias significativas, otras mantienen brechas persistentes, particularmente en los niveles más altos de productividad e impacto. El análisis diacrónico permite identificar puntos de inflexión vinculados a políticas de equidad, cambios culturales o intervenciones institucionales, proporcionando evidencia valiosa para diseñar estrategias efectivas de promoción de la igualdad de género en el sistema científico.

#### **9.5. Interpretación de la tipología documental**

La diversidad tipológica de las publicaciones científicas refleja la pluralidad de géneros discursivos y formas de comunicación en cada disciplina. Los artículos de investigación empírica dominan en ciencias naturales y médicas, mientras en humanidades y ciencias sociales los libros y capítulos de libro mantienen un lugar central. Estas diferencias tipológicas responden a tradiciones epistemológicas, prácticas de validación del conocimiento y estructuras de recompensa académica específicas de cada campo, y deben considerarse críticamente en cualquier ejercicio evaluativo.

El análisis de la distribución tipológica a lo largo del tiempo revela transformaciones en las prácticas de comunicación científica. La creciente predominancia de los artículos de investigación en muchas disciplinas refleja procesos de homogenización impulsados por sistemas de evaluación basados en métricas de revistas. Sin embargo, resurgen formas alternativas como *preprints*, sets de datos, software y otros productos de investigación que responden a movimientos de ciencia abierta y reconocen la diversidad de contribuciones al avance del conocimiento.

El impacto relativo de diferentes tipologías documentales varía sustancialmente entre disciplinas y subcampos. Mientras en algunas áreas los artículos de revisión sistemática reciben

mayor atención citacional, en otras son las contribuciones metodológicas o teóricas las más influyentes. Comprender estas variaciones es esencial para evaluaciones bibliométricas justas que reconozcan el valor diferencial de distintos tipos de contribución científica según los criterios de validación propios de cada comunidad epistémica.

### **9.6. Interpretación del análisis de palabras clave**

El análisis de palabras clave proporciona una ventana privilegiada a la estructura conceptual y cognitiva de los campos científicos. Las frecuencias de términos revelan los conceptos centrales que definen un dominio de investigación, mientras las co-ocurrencias muestran las conexiones semánticas que estructuran el espacio conceptual. Este análisis permite identificar núcleos temáticos consolidados, interfaces entre subcampos y espacios conceptuales vacíos que representan oportunidades para investigación innovadora.

La evolución terminológica captura procesos de cambio conceptual, emergencia de nuevos paradigmas y transformaciones en los marcos teóricos de las disciplinas. El análisis diacrónico de palabras clave permite rastrear la migración de conceptos entre campos, la emergencia de nuevos vocabularios especializados y la desaparición de términos obsoletos. Estos patrones de cambio terminológico reflejan dinámicas más profundas de reconfiguración de los paisajes cognitivos de las disciplinas.

El análisis de diversidad terminológica ofrece perspectivas sobre el grado de especialización o interdisciplinariedad de un campo de investigación. Campos altamente especializados muestran vocabularios técnicos densos y compartidos, mientras áreas interdisciplinarias presentan mayor heterogeneidad terminológica y préstamos conceptuales de múltiples dominios. Esta dimensión analítica resulta particularmente valiosa para estudiar procesos de fragmentación disciplinaria, emergencia de campos híbridos y dinámicas de convergencia científica.

### **9.7. Análisis de redes de colaboración institucional**

Las redes de colaboración institucional revelan la arquitectura organizacional del sistema científico, mostrando cómo instituciones de diferente tipo (universidades, centros de investigación, empresas, hospitales) se conectan para producir conocimiento. El análisis de estas redes permite identificar instituciones líderes que funcionan como centros de colaboración, patrones de asociación preferencial basados en proximidad geográfica o complementariedad temática, y la emergencia de consorcios estratégicos que reconfiguran los paisajes investigadores.

La evolución de las redes de colaboración muestra tendencias hacia una creciente interconexión institucional, pero con patrones marcados por desigualdades estructurales. Instituciones de países centrales típicamente ocupan posiciones más centrales en redes globales, mientras instituciones periféricas enfrentan mayores barreras para integrarse en colaboraciones internacionales. El análisis diacrónico de estas redes permite evaluar el impacto de políticas científicas diseñadas para promover colaboraciones más inclusivas y diversas.

La diversidad institucional en las colaboraciones constituye un indicador de salud del ecosistema investigador. Colaboraciones que involucran múltiples tipos de instituciones (academia-industria-gobierno) típicamente generan conocimientos con mayor potencial de impacto social y económico. El análisis de esta diversidad colaborativa informa políticas destinadas a fortalecer los vínculos entre diferentes sectores del sistema de innovación.

### **9.8. Patrones de financiamiento en la producción científica**

El análisis de las agencias y programas de financiamiento revela la infraestructura económica

que sustenta la investigación científica. La distribución de las fuentes de financiamiento muestra la centralidad de ciertas agencias en el apoyo a la investigación, las prioridades temáticas que guían las inversiones en I+D, y la concentración o diversificación de las fuentes de apoyo a la ciencia. Este análisis proporciona evidencia valiosa para evaluar la efectividad de políticas de financiamiento científico.

El impacto diferencial de diferentes tipos de financiamiento constituye una dimensión analítica de creciente interés. Estudios bibliométricos muestran que investigaciones con apoyo competitivo típicamente alcanzan mayor visibilidad e impacto que aquellas sin financiamiento explícito. Además, el monto, duración y tipo de financiamiento (individual vs. colaborativo, básico vs. aplicado) se correlacionan con diferentes patrones de productividad e impacto, información crucial para optimizar estrategias de inversión en investigación.

La transparencia en las fuentes de financiamiento ha aumentado significativamente en las últimas décadas, facilitando el análisis de las relaciones entre inversión en I+D y resultados científicos. Esta transparencia permite trazar conexiones entre prioridades de financiamiento, temas de investigación emergentes, contribuyendo a una rendición de cuentas más robusta de los sistemas de apoyo a la ciencia y a una mejor alineación entre inversión científica y necesidades sociales.

## **9.9. Combinaciones**

La verdadera potencia del análisis bibliométrico contemporáneo emerge cuando se integran múltiples dimensiones complementarias para construir narrativas analíticas multidimensionales. Estas combinaciones permiten trascender las limitaciones de los análisis unidimensionales, revelando interacciones complejas entre diferentes aspectos del ecosistema científico. La integración sistemática de dimensiones como género, tipología documental, patrones de colaboración y financiamiento genera perspectivas que serían imposibles de obtener examinando cada dimensión por separado, facilitando una comprensión más holística y matizada de la dinámica investigadora.

La combinación de análisis de género con estudios de colaboración institucional revela patrones cruciales sobre la participación diferencial en redes científicas. Por ejemplo, se puede examinar si las investigadoras participan en redes de colaboración con las mismas características estructurales que sus colegas masculinos, o si existen diferencias sistemáticas en el tipo de instituciones con las que colaboran. Esta integración permite identificar si ciertas configuraciones colaborativas, como consorcios internacionales multinstitucionales, presentan mayores o menores niveles de equidad de género, información valiosa para diseñar políticas de colaboración científica más inclusivas.

La intersección entre análisis tipológico y evolución temporal desvela transformaciones profundas en las prácticas de comunicación científica. Al combinar estas dimensiones, podemos rastrear no solo cambios en los volúmenes de producción sino también transformaciones en los géneros discursivos predominantes en diferentes períodos. Esta integración permite identificar momentos de cambio paradigmático donde emergen nuevas formas de comunicación científica, o procesos de homogenización donde ciertos tipos documentales desplazan a otros, con implicaciones para la diversidad epistemológica de las disciplinas.

La triangulación de análisis de palabras clave con datos de financiamiento revela las conexiones entre prioridades de financiación y la evolución de agendas investigadoras. Esta combinación permite examinar cómo la introducción de nuevos programas de financiamiento

se refleja en la emergencia de términos especializados en la literatura, o cómo cambios en las prioridades de agencias financiadoras reorientan los focos temáticos de comunidades científicas enteras. Este análisis es particularmente valioso para evaluar el impacto de políticas científicas en la configuración de paisajes cognitivos disciplinares.

La integración de dimensiones geográficas, temáticas y de colaboración genera mapas complejos de la geopolítica del conocimiento. Al combinar estas dimensiones, se puede identificar no solo qué países producen conocimiento en qué áreas, sino también cómo se estructuran las redes globales de colaboración alrededor de diferentes especialidades temáticas. Esta integración revela patrones de dependencia cognitiva, nichos de excelencia regional, y asimetrías en la capacidad de diferentes países para influir en agendas investigadoras globales.

El análisis multivariado de dimensiones complementarias enfrenta desafíos metodológicos significativos, particularmente en el manejo de interacciones complejas y la identificación de relaciones causales. Sin embargo, aproximaciones como el análisis de correspondencias múltiples, los modelos de ecuaciones estructurales y las técnicas de *machine learning* permiten explorar sistemáticamente estas interrelaciones. El desarrollo de paneles interactivos que permiten cruzar dinámicamente diferentes dimensiones representa una innovación prometedora para la exploración visual de estas complejidades.

La interpretación de análisis integrados requiere una sensibilidad particular a las especificidades disciplinares y contextuales. Patrones que aparecen como universales pueden desvanecerse al examinar disciplinas individuales, mientras relaciones que parecen marginales a nivel agregado pueden resultar cruciales en campos específicos. Esta sensibilidad contextual es esencial para evitar generalizaciones indebidas y para producir conocimientos que respeten la diversidad epistemológica del universo científico.

Las combinaciones de dimensiones abren posibilidades para nuevas métricas sintéticas que capturen aspectos multidimensionales de la producción científica. Indicadores de diversidad cognitiva, equidad en colaboraciones, o alineación entre financiamiento y resultados emergen naturalmente de estas integraciones, ofreciendo herramientas más sofisticadas para la evaluación científica.

La obtención de datos para análisis bibliométricos multidimensionales representa un proceso altamente contextual que varía sustancialmente según los objetivos específicos de investigación y las preguntas analíticas planteadas. Cada investigador debe diseñar una estrategia de recolección personalizada que responda a sus necesidades particulares, partiendo típicamente de dos fuentes principales: los archivos CSV exportados directamente desde bases de datos bibliográficas como Scopus, Web of Science o Dimensions, o los resultados crudos generados por herramientas especializadas como *Publish or Perish*, Bibliometrix o VOSviewer. Esta elección inicial condiciona profundamente el posterior procesamiento analítico, ya que cada fuente presenta formatos, estructuras y niveles de complejidad significativamente diferentes.

Ya están todos los resultados listos, junto con la idea de que significa cada uno, el siguiente paso es comunicarlo. En la siguiente sección se abordará como redactar una bibliometría, así como preparar la presentación de las mismas.

## 9.10. Estrategias de visualización para diferentes dimensiones bibliométrica

La selección de representaciones visuales adecuadas trasciende lo meramente estético para convertirse en una decisión metodológica fundamental que impacta directamente en la



comprensión e interpretación de los hallazgos bibliométricos. Cada tipo de análisis demanda formatos específicos que resalten sus patrones característicos esenciales, facilitando la extracción de elementos significativos de conjuntos de datos complejos. La efectividad comunicativa de un estudio bibliométrico depende críticamente de esta elección estratégica, donde la correspondencia entre estructura de datos y formato visual determina la capacidad del receptor para decodificar la información presentada.

Para el análisis de distribución lingüística en la producción científica, los gráficos de barras verticales u horizontales ofrecen una solución óptima al permitir comparaciones inmediatas entre frecuencias de publicación por idioma. Esta representación visualiza con claridad el predominio del inglés como *lingua franca* científica frente a otros idiomas, mientras posibilita identificar tendencias temporales mediante series históricas. La incorporación de codificación por colores según familias lingüísticas o regiones geográficas añade capas analíticas adicionales, transformando una simple distribución cuantitativa en una herramienta para explorar hegemonías culturales en la comunicación científica global.

La representación de distribución geográfica encuentra en los mapas coropléticos su expresión más intuitiva y eficaz, transformando datos tabulares en narrativas espaciales inmediatamente comprensibles. Esta modalidad cartográfica asigna intensidades cromáticas según la densidad de producción científica por países o regiones, revelando patrones de concentración del conocimiento y asimetrías Norte-Sur con impacto visual inmediato. La superposición con indicadores de colaboración internacional mediante flujos dirigidos crea visualizaciones multidimensionales que capturan simultáneamente productividad individual y redes de cooperación, ofreciendo una perspectiva integrada del panorama científico global.

Las herramientas bibliométricas especializadas como Bibliometrix en R y PyBibX en Python generan automáticamente visualizaciones avanzadas que integran múltiples dimensiones abordadas en el presente capítulo, integrándolas entre sí o entre los indicadores bibliométricos como lo son la tasa de citación bruta o anual.

La elección de qué tipo de herramienta y representación a usar está dada por las necesidades de lo que se quiere comunicar, siempre teniendo en cuenta que se debe usar la forma de representación más sencilla e intuitiva posible (texto plano -> tabla -> representación gráfica), sin repetir información.

## Recapitulando

- Las dimensiones contextuales complementan la comprensión de resultados bibliométricos.
- El idioma de publicación influye en la visibilidad internacional.
- La tipología documental determina representatividad y peso.
- Es fundamental normalizar afiliaciones institucionales.
- El análisis de género requiere métodos éticos y transparentes.
- Las inferencias de género deben estar sustentadas y documentadas.
- Los análisis por país permiten identificar desigualdades regionales.
- Vincular financiamiento con producción ayuda a evaluar impacto.
- La colaboración industria-academia se mide en coautorías y patentes.
- El análisis por revista muestra concentración editorial.
- Cruzar dimensiones revela desigualdades estructurales.
- Los datos faltantes limitan la precisión.
- Fuentes complementarias (Crossref, OpenAlex) mejoran metadatos.

- Visualizaciones geográficas facilitan comparaciones globales.
- Analizar idiomas en títulos/resúmenes mide accesibilidad.
- La cobertura de bases de datos influye en sesgos.
- La normalización por campo y año es esencial para comparar.
- Se debe reportar incertidumbre y vacíos de datos.
- Es necesario respetar confidencialidad de subgrupos.
- Incorporar variables adicionales enriquece la interpretación.

### **Preguntas de autoevaluación**

1. ¿Por qué es útil estudiar el idioma de publicación?
2. ¿Qué sesgos surgen del uso exclusivo de Scopus o WoS?
3. ¿Cómo se normalizan afiliaciones?
4. ¿Qué precauciones éticas hay en inferir género?
5. ¿Cómo vincularías datos de financiamiento con temas?
6. ¿Qué implicaciones tiene la tipología documental?
7. ¿Cómo evitar errores por datos faltantes?
8. ¿Por qué usar fuentes complementarias?
9. ¿Qué aporta cruzar dimensiones (país, género)?
10. ¿Qué visualizaciones son más efectivas?

### **BIBLIOGRAFÍA**

1. Gingras Y. *Bibliometrics and research evaluation: uses and abuses*. Cambridge (MA): MIT Press; 2016. ISBN 9780262337663

2. De Bellis N. *Bibliometrics and citation analysis: from the Science Citation Index to cybermetrics*. Lanham (MD): Scarecrow Press; 2009. ISBN 9780810867130

3. Rousseau R, Egghe L, Guns R. *Becoming metric-wise: a bibliometric guide for researchers*. Cambridge (MA): Chandos/Elsevier; 2018. ISBN 9780081024744

4. Waltman L. A review of the literature on citation impact indicators. *J Informetrics*. 2016;10(2):365-91. <https://doi.org/10.1016/j.joi.2016.02.007>

### **REFERENCIAS BIBLIOGRÁFICAS**

1. UNESCO. The race against time for smarter development | 2021 Science Report [Internet]. Disponible en: <https://www.unesco.org/reports/science/2021/en>

# Part IV / Parte IV

PUBLISHING AND COMMUNICATION

PUBLICAR Y COMUNICAR



## Chapter 10 / Capítulo 10

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch10

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

## ACADEMIC WRITING / REDACCIÓN ACADÉMICA

La culminación de cualquier investigación bibliométrica reside en su capacidad para comunicar hallazgos de manera clara, rigurosa y persuasiva. Este capítulo se concentra en los aspectos prácticos de la redacción académica, la estructuración de artículos y la comunicación efectiva de resultados bibliométricos a diferentes audiencias. Desde la conceptualización inicial del estudio hasta la presentación final de resultados, cada capítulo proporciona herramientas concretas para transformar análisis de datos en contribuciones científicas significativas que enriquezcan nuestra comprensión de la dinámica del conocimiento.

### 10.1. Idea de bibliometría

Un estudio bibliométrico se distingue de otras formas de investigación por su enfoque cuantitativo en el análisis de la producción, diseminación e impacto del conocimiento científico a través de métricas y técnicas estadísticas específicas. La concepción de una investigación bibliométrica, como toda investigación científica, empieza con la identificación de un vacío en la comprensión de la estructura o dinámica de un campo científico, donde el análisis de patrones en la literatura puede ofrecer soluciones. A diferencia de las revisiones sistemáticas que sintetizan hallazgos sustantivos, o los estudios cualitativos que exploran significados y experiencias, la bibliometría se concentra en los patrones macroscópicos que emergen del conjunto de publicaciones científicas, utilizando indicadores cuantitativos para revelar tendencias, relaciones y estructuras subyacentes.

La delimitación clara del objeto de estudio constituye el primer paso crucial para definir un proyecto bibliométrico. Esto implica especificar con precisión el corpus documental a analizar, los períodos temporales considerados, las fuentes de datos utilizadas y las limitaciones inherentes a estas decisiones metodológicas. Un estudio bibliométrico bien definido se caracteriza por preguntas de investigación que no pueden responderse mediante la mera lectura de literatura, sino que requieren el análisis sistemático de grandes volúmenes de datos bibliográficos para revelar patrones que trascienden la experiencia subjetiva individual. Por ejemplo, mientras un investigador podría tener intuiciones sobre las tendencias en su campo, la bibliometría permite cuantificar estas tendencias, identificar factores influyentes y descubrir relaciones no evidentes a simple vista.

Los elementos que distinguen un estudio bibliométrico de otros tipos de artículos incluyen el uso sistemático de indicadores cuantitativos, el análisis de patrones relacionales mediante técnicas de red, la identificación de trayectorias temporales, y la contextualización de hallazgos dentro de marcos teóricos sobre la dinámica científica. Un proyecto se convierte en bibliométrico cuando su objetivo principal es comprender la arquitectura del conocimiento, las redes de colaboración, los patrones de citación, o la evolución temática, más que avanzar el conocimiento sustantivo dentro de un campo particular.

Esta orientación metodológica específica determina tanto el diseño del estudio como la estructura del manuscrito resultante, requiriendo secciones metodológicas detalladas sobre la obtención y procesamiento de datos, y una presentación de resultados que equilibre el rigor cuantitativo con la interpretación sustantiva significativa.

Las investigaciones bibliométricas responden a brechas del conocimiento, o sea, tienen un fin cognoscitivo, por lo que el problema científico siempre estará en torno a una ausencia del conocimiento sobre tendencias investigativas. Esto es de suma relevancia pues sobre el mismo debe plantearse el objetivo y dirigir todo el artículo. Todo lo anterior queda implícito en la

sección INTRODUCCIÓN del artículo.

## 10.2. Redacción de los métodos

La sección de MÉTODOS en un estudio bibliométrico debe proporcionar una descripción exhaustiva que permita la replicación completa de la investigación. Esta transparencia metodológica constituye un pilar fundamental de la integridad científica en bibliometría, donde decisiones aparentemente menores durante la recolección y procesamiento de datos pueden influir significativamente en los resultados. La redacción debe equilibrar el detalle técnico necesario para la reproducibilidad con la claridad expositiva que facilite la comprensión por parte de lectores de diferentes áreas. Unos métodos bien documentados no solo validan la credibilidad del estudio, sino que también contribuyen al avance metodológico de la disciplina al permitir la comparación y mejora iterativa de técnicas de análisis.

### 10.2.1. Diagrama PRISMA para transparencia

El diagrama PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (<https://www.prisma-statement.org/>) sistematiza el proceso de selección documental mediante cuatro fases secuenciales: identificación, screening, elegibilidad e inclusión. En bibliometría, la fase de identificación detalla las bases consultadas, estrategias de búsqueda y registros obtenidos, incluyendo duplicados. El screening aplica criterios basados en metadatos como tipo documental, idioma o accesibilidad. La elegibilidad evalúa textos completos según relevancia temática y calidad de metadatos. La fase final especifica el corpus analizado y los criterios de exclusión aplicados. Esta estructura garantiza transparencia en la construcción del set de datos bibliográfico, fundamental para la validez del estudio.

La adaptación bibliométrica de PRISMA incorpora elementos específicos como los protocolos de extracción y limpieza de metadatos. Esto incluye procesos de normalización de autores, estandarización de afiliaciones y unificación terminológica. También debe documentar los algoritmos para resolver ambigüedades en nombres institucionales y la gestión de registros duplicados entre bases de datos. Estos elementos ampliados son cruciales para garantizar la consistencia del corpus analítico. La visualización del flujo debe reflejar estas etapas adicionales, mostrando cómo los datos brutos se transforman en el set de datos final listo para análisis bibliométrico.

La reproductibilidad técnica completa el marco PRISMA mediante la documentación exhaustiva de herramientas y parámetros. Esto incluye versiones de software bibliométrico, configuraciones específicas de análisis y scripts de procesamiento. Los umbrales aplicados en análisis de redes, los algoritmos de *clustering* utilizados y los criterios de normalización de indicadores deben especificarse detalladamente. La disponibilidad de código y set de datos mediante repositorios accesibles permite la verificación independiente de resultados. Esta transparencia metodológica posibilita la replicación exacta del estudio y facilita comparaciones entre investigaciones bibliométricas similares.

### 10.2.2. Protocolos éticos

Los aspectos éticos en bibliometría abarcan consideraciones sobre el *scraping* de datos, el uso de información sensible, y la responsabilidad en la interpretación y comunicación de resultados. La redacción debe declarar explícitamente el cumplimiento de los términos de servicio de las bases de datos utilizadas, las medidas implementadas para evitar la sobrecarga de servidores durante la extracción de datos, y los protocolos seguidos para garantizar el almacenamiento seguro y uso responsable de la información obtenida.

Cuando se analicen datos individuales de investigadores, deben describirse los procedimientos para preservar el anonimato y evitar usos potencialmente lesivos, especialmente en contextos evaluativos donde los resultados podrían afectar trayectorias profesionales. La declaración de conflictos de interés y fuentes de financiamiento completa los aspectos éticos básicos que deben abordarse en la sección de métodos.

La metodología de los estudios bibliométricos también puede apoyarse en un diagrama de flujo de artículos en el que se especifique el número de artículos encontrados, y de ahí el flujo de filtros o criterios de selección hasta que quede la cantidad o muestra con la que se trabajará. Cada etapa del flujo debe registrar el número de documentos incluidos y excluidos, junto con los criterios específicos aplicados en cada filtro, lo que permite visualizar de inmediato cómo la muestra inicial se reduce progresivamente hasta conformar el corpus bibliográfico definitivo para el análisis. Esto es recomendable hacerlo en los estudios en los que no se trabaje con la totalidad de artículos encontrados, o sea, en los que se aplique muestreo, muestreo que debe declararse.

### 10.3. Redacción de los resultados

La presentación de resultados bibliométricos debe seguir una secuencia lógica que guíe al lector desde los patrones generales hacia los análisis específicos. Por esto se sugiere siempre empezar con los indicadores de productividad básicos: número total de publicaciones, distribución temporal y tasas de crecimiento anual, así como los criterios complementarios como tipología de artículo, países etc. Estos datos fundamentales establecen el contexto cuantitativo esencial para interpretar análisis posteriores más complejos. Estos resultados pueden representarse en tablas o gráficos de evolución temporal que destaquen períodos de aceleración o desaceleración en la actividad investigadora.

Se continua con el análisis de autores e instituciones más productivas, empleando tablas que jerarquicen los principales contribuyentes según métricas de productividad e impacto. Incluyendo el número de publicaciones por autor, las instituciones con mayor producción y los países más activos. Estas tablas deben presentarse con columnas bien definidas que muestren rangos, nombres, número de documentos y citas totales.

Los indicadores de impacto y citación conforman la tercera sección de resultados, donde se presentan distribuciones de citas, índices h e indicadores normalizados por campo. El empleo de gráficos de barras muestra citas por documento, mientras los *boxplots* visualizan distribuciones de citas y las tablas de percentiles establecen comparaciones normalizadas. Estos elementos visuales destacan las publicaciones más influyentes y los patrones de citación predominantes, requiriendo especial atención a las escalas y rangos para evitar distorsiones en la interpretación.

El análisis de redes y colaboraciones representa un componente visual crítico que se presenta mediante figuras de grafos anotadas. La secuencia inicia con redes de coautoría mostrando los principales clústeres colaborativos, seguido de mapas de co-palabras que revelan la estructura temática del campo. Cada figura incluye leyendas detalladas explicando el significado de colores, tamaños de nodo y grosores de enlace, manteniendo resolución suficiente para garantizar legibilidad incluso en impresiones a tamaño artículo.

Los mapas de densidad y calor se reservan para mostrar patrones espaciotemporales y concentraciones temáticas. Estos elementos visuales resultan particularmente efectivos para comunicar la evolución diacrónica de temas o la distribución geográfica de la producción científica. El diseño asegura que las paletas de color sean intuitivas y accesibles, utilizando



gradientes que progresan de bajas a altas concentraciones de manera lógica y consistentemente interpretable.

La aplicación de los diagramas de Sankey muestra movilidad temática, cambios en patrones de colaboración o evolución en focos de investigación. El diseño de estos diagramas mantiene suficiente espacio entre flujos para garantizar legibilidad, utilizando colores distintivos para diferentes trayectorias. La complejidad de estos gráficos requiere una explicación detallada en leyendas y referencias en el texto principal.

Los análisis de tendencias temporales ocupan una sección separada donde se presentan series cronológicas de conceptos emergentes o indicadores evolutivos. Los gráficos de línea muestran trayectorias temáticas mientras los gráficos de área visualizan la prominencia relativa de diferentes temas a través del tiempo. Estos elementos incluyen puntos de referencia significativos como lanzamientos de políticas científicas o hitos tecnológicos que contextualizan los cambios observados.

La sección final presenta resultados especializados como análisis de *burst*, patrones de citación o estudios de interdisciplinariedad. Estas visualizaciones complejas se reservan para el final, una vez comprendidos los patrones fundamentales. Su inclusión como figuras complementarias profundiza aspectos específicos del estudio, asegurando que cada visualización avanzada mantenga una justificación clara en los objetivos de investigación.

Algunas veces hay resultados poco trascendentes para el cumplimiento del objetivo planteado en el primer momento, por lo que se deberá prescindir de ellos para darle prioridad a los que de verdad dan peso a la investigación. Así mismo el orden puede cambiar en dependencia del flujo de información que se desee seguir, siempre garantizando que se ira de los elementos más generales y básicos a los más específicos y avanzados.

La numeración de tablas y figuras sigue estrictamente el orden de aparición en el texto, con referencias cruzadas que guían fluidamente entre diferentes elementos visuales. Cada tabla y figura se diseña para comprensión independiente, con títulos descriptivos y leyendas autoexplicativas. Este cuidado en la presentación visual diferencia los estudios bibliométricos profesionales de los meros ejercicios cuantitativos, elevando la calidad comunicativa del manuscrito académico.

### 10.3.1. Selección de visualizaciones por audiencia

La elección de visualizaciones en estudios bibliométricos debe adaptarse específicamente al conocimiento técnico y los intereses de la audiencia objetivo. Para comités de evaluación y gestores científicos, se priorizan gráficos de indicadores básicos como evolución temporal de publicaciones y citas, diagramas de redes de colaboración simplificados, y mapas de calor de productividad institucional. Estas visualizaciones comunican eficientemente información esencial sobre productividad, impacto y colaboraciones sin requerir conocimientos especializados en bibliometría, facilitando procesos decisorios basados en evidencia cuantitativa robusta pero accesible.

Para audiencias académicas especializadas en el campo de estudio, las visualizaciones pueden incorporar mayor complejidad técnica y terminología específica. Los mapas de co-palabras detallados, los análisis de co-citación y los diagramas de evolución temática resultan apropiados para estos contextos. La presentación para especialistas puede incluir parámetros metodológicos avanzados, algoritmos de *clustering* específicos y medidas de centralidad de

redes, siempre que estos elementos técnicos aporten valor interpretativo sustantivo a los hallazgos del estudio.

Cuando el estudio bibliométrico se dirige a tribunales editoriales o audiencias interdisciplinarias, las visualizaciones deben maximizar la claridad inmediata y el impacto comunicativo. Estas visualizaciones transforman datos complejos en narrativas intuitivas sobre la dinámica del campo investigador, facilitando la transferencia de conocimiento bibliométrico a contextos de política científica y planificación estratégica.

### **10.3.2. Tablas para comparativas**

Las tablas constituyen el formato óptimo para presentar comparativas sistemáticas entre entidades bibliométricas como autores, instituciones, revistas o países. El diseño tabular debe organizar la información según criterios claros de ordenamiento, típicamente por métricas de productividad o impacto descendente. Cada tabla debe incluir columnas para el ranking, nombre de la entidad, número de publicaciones, citas totales, e indicadores normalizados como el índice h o el impacto promedio por documento, permitiendo comparaciones multidimensionales inmediatas.

La estructuración jerárquica de tablas facilita el análisis progresivo de resultados comparativos. Las tablas iniciales presentan rankings generales de productividad, mientras tablas subsiguientes muestran desgloses por períodos temporales, especialidades temáticas o tipos de colaboración. Esta organización permite identificar no solo las entidades más productivas globalmente, sino también patrones de especialización, trayectorias de crecimiento diferencial y cambios en liderazgo relativo a lo largo del tiempo. Cada tabla debe acompañarse de breves análisis textuales que destaquen los hallazgos comparativos más relevantes.

## **10.4. Redacción de las discusiones**

La sección de discusión constituye el componente analítico central donde los hallazgos bibliométricos adquieren significado sustantivo mediante su interpretación contextualizada. Esta transición desde la descripción cuantitativa hacia la explicación cualitativa requiere integrar sistemáticamente evidencias empíricas con marcos teóricos existentes sobre la dinámica científica. La construcción de argumentos interpretativos se sustenta en la triangulación metódica entre hallazgos propios, literatura especializada y teorías de la ciencia, estableciendo conexiones significativas entre patrones observados y procesos subyacentes en el ecosistema investigador. La discusión debe evitar tanto la mera repetición de resultados como la especulación no fundamentada, equilibrando rigor interpretativo con relevancia disciplinar.

La articulación entre resultados y literatura previa representa un proceso metodológico que requiere estrategias específicas de contraste documental. La técnica de mapeo conceptual sistemático identifica publicaciones clave que abordan fenómenos similares en contextos comparables, estableciendo diálogos críticos con estudios antecedentes. La selección de literatura para contrastar debe incluir investigaciones bibliométricas previas sobre el mismo campo temático, estudios cualitativos que exploren dimensiones no capturadas por métricas cuantitativas, y trabajos teóricos que proporcionen marcos explicativos para interpretar los patrones observados. Esta triangulación documental enriquece sustancialmente la profundidad interpretativa de la discusión.

La identificación de consistencias e inconsistencias con la literatura existente debe realizarse mediante análisis comparativo estructurado. Cuando los resultados confirman hallazgos previos, la discusión debe explorar los mecanismos subyacentes que explicarían estas regularidades

transcontextuales. Frente a discrepancias con estudios anteriores, el análisis debe examinar factores metodológicos, diferencias temporales o particularidades contextuales que podrían dar cuenta de las variaciones observadas. Este abordaje comparativo transforma las simples coincidencias o diferencias en oportunidades para avanzar en la comprensión teórica de los fenómenos bibliométricos analizados.

#### **10.4.1. Vinculación con políticas científicas**

La discusión de implicaciones para políticas científicas constituye un componente esencial que conecta el análisis bibliométrico con aplicaciones prácticas en gestión de la investigación. La interpretación de resultados debe identificar cómo los patrones observados informan el diseño, implementación o evaluación de intervenciones en el sistema científico. Por ejemplo, la identificación de áreas temáticas subrepresentadas puede sustentar recomendaciones para programas de financiamiento específicos, mientras el análisis de redes de colaboración puede informar estrategias de internacionalización. Esta vinculación entre evidencia bibliométrica y acción política añade relevancia sustantiva al estudio.

La elaboración de recomendaciones de política requiere un cuidadoso equilibrio entre evidencia empírica y consideraciones contextuales. Las propuestas deben derivarse lógicamente de los resultados, evitando extrapolaciones indebidas más allá de lo que los datos permiten afirmar. La discusión debe reconocer explícitamente la naturaleza parcial y complementaria de la evidencia bibliométrica dentro de procesos complejos de toma de decisiones políticas. Esta aproximación rigurosa fortalece la credibilidad del estudio y su potencial impacto en la mejora de sistemas de investigación.

La identificación de mecanismos causales plausibles que conecten intervenciones políticas con patrones bibliométricos observados enriquece significativamente la discusión. Por ejemplo, cambios en patrones de colaboración internacional pueden relacionarse con programas específicos de movilidad científica, mientras variaciones en productividad temática pueden asociarse a prioridades de financiamiento establecidas. Esta elaboración teórica trasciende la mera correlación para proponer explicaciones causales fundamentadas que informen futuras intervenciones de política científica.

La contextualización de hallazgos dentro de marcos de política científica existentes añade profundidad analítica a la discusión. El contraste con documentos de planificación científica nacionales e internacionales, informes de organismos como la UNESCO o la OCDE, y evaluaciones de programas de financiamiento específicos permite situar los resultados dentro de tendencias políticas más amplias. Esta contextualización facilita la identificación de desajustes entre objetivos declarados de política científica y patrones reales observados en la actividad investigadora.

La elaboración de escenarios prospectivos basados en tendencias bibliométricas identificadas representa una estrategia discursiva particularmente valiosa para la planificación científica. La proyección de trayectorias temáticas, la identificación de oportunidades emergentes de especialización y el mapeo de capacidades científicas futuras transforman el análisis retrospectivo típico de la bibliometría en herramienta para la anticipación estratégica. Estos ejercicios prospectivos deben fundamentarse claramente en los datos presentados y reconocer sus limitaciones predictivas.

#### **10.4.2. Limitaciones metodológicas**

El reconocimiento exhaustivo de limitaciones metodológicas constituye un elemento de

rigor intelectual en la redacción de discusiones bibliométricas. La cobertura parcial de bases de datos representa una limitación fundamental que debe caracterizarse cuantitativamente cuando es posible, especificando los sesgos geográficos, disciplinares y tipológicos introducidos por la selección de fuentes. La discusión debe evaluar cómo estas limitaciones afectan la validez externa de los hallazgos y delimitar apropiadamente el alcance de las conclusiones. Esta transparencia metodológica fortalece la credibilidad del estudio.

Las decisiones metodológicas en el procesamiento de datos introducen limitaciones adicionales que merecen examen crítico en la discusión. Los algoritmos de desambiguación de autores, los criterios de normalización terminológica y los umbrales en análisis de redes constituyen fuentes potenciales de sesgo que afectan los resultados obtenidos. La discusión debe explorar cómo diferentes decisiones metodológicas habrían podido alterar los patrones observados, demostrando conciencia sobre la naturaleza construida de los datos bibliométricos. Esta reflexividad metodológica caracteriza la investigación bibliométrica de alta calidad.

La identificación de limitaciones debe complementarse con la propuesta de direcciones para investigación futura que superen las restricciones metodológicas actuales. La discusión puede sugerir estrategias para validación de hallazgos mediante triangulación con otras fuentes de datos, métodos alternativos de análisis o réplicas en diferentes contextos temporales o disciplinares. Esta proyección hacia futuros avances metodológicos posiciona el estudio dentro de un programa de investigación en evolución más que como un ejercicio aislado.

El análisis del impacto diferencial de limitaciones metodológicas en distintos tipos de hallazgos añade sofisticación al examen crítico. Algunas limitaciones afectan principalmente a análisis cuantitativos de productividad, mientras otras distorsionan especialmente las estructuras relacionales en análisis de redes. Esta caracterización diferenciada permite a los lectores evaluar qué hallazgos poseen mayor robustez metodológica y cuáles deben interpretarse con mayores cautelas contextuales.

#### **10.4.3. Integración de perspectivas multidisciplinarias**

La incorporación de perspectivas teóricas desde diferentes disciplinas enriquece sustancialmente la interpretación de hallazgos bibliométricos. Los marcos conceptuales desde la sociología de la ciencia, la economía de la innovación y los estudios de ciencia, tecnología y sociedad proporcionan lentes interpretativos valiosos para comprender los patrones observados. Esta integración multidisciplinaria trasciende la mera descripción bibliométrica para avanzar hacia explicaciones sustantivas sobre la dinámica social e institucional subyacente a los datos cuantitativos.

La triangulación con evidencias cualitativas sobre los fenómenos estudiados representa una estrategia particularmente poderosa para enriquecer la discusión bibliométrica. La incorporación de entrevistas con investigadores, análisis de documentos de política científica o estudios etnográficos de laboratorios, cuando están disponibles, permite contextualizar los patrones cuantitativos dentro de procesos sociales concretos. Esta integración metodológica supera las limitaciones inherentes de los enfoques puramente cuantitativos en bibliometría.

Examinar las implicaciones epistemológicas de los hallazgos bibliométricos añade profundidad teórica a la discusión. El análisis de cómo los patrones observados reflejan o desafían concepciones establecidas sobre la producción de conocimiento, la estructura de las disciplinas científicas o la dinámica del cambio paradigmático conecta la bibliometría con debates fundamentales en filosofía de la ciencia. Esta elevación del diálogo intelectual maximiza la contribución teórica

del estudio.

### **10.5. Buenas prácticas en visualización de datos**

La visualización efectiva de datos bibliométricos requiere la aplicación sistemática de principios de diseño gráfico que equilibren precisión analítica con claridad comunicativa. La selección de colores debe priorizar paletas perceptualmente uniformes, garantizando accesibilidad para personas con daltonismo mediante herramientas de verificación específicas. Los esquemas cromáticos deben aplicarse consistentemente a través de todas las visualizaciones del estudio, utilizando convenciones intuitivas donde colores cálidos representen valores altos y colores fríos indiquen valores bajos. Esta coherencia visual facilita la interpretación comparativa entre diferentes figuras y gráficos.

La jerarquización visual constituye un principio fundamental para guiar la atención del lector hacia los elementos más significativos de cada visualización. El uso estratégico de tamaño, contraste y posición destaca patrones clave sin distorsionar la integridad de los datos subyacentes. Las etiquetas textuales deben ubicarse cuidadosamente para maximizar legibilidad sin ocultar información importante, empleando conectores cuando sea necesario para asociar texto con elementos gráficos específicos. Esta organización jerárquica transforma visualizaciones complejas en narrativas visuales intuitivas.

La elección entre formatos de visualización debe fundamentarse en la naturaleza de los datos y los objetivos comunicativos específicos. Los gráficos de barras resultan ideales para comparar magnitudes entre categorías discretas, mientras los gráficos de línea muestran efectivamente tendencias temporales. Las visualizaciones de redes capturan relaciones estructurales, y los mapas de calor representan densidades o intensidades entre dimensiones continuas. Cada selección formativa debe justificarse por su capacidad para comunicar eficientemente los hallazgos centrales del análisis.

La escalabilidad de las visualizaciones merece consideración especial en el contexto de publicaciones académicas. Las figuras deben diseñarse para mantener legibilidad tanto en formatos impresos como digitales, utilizando resoluciones suficientes para permitir zoom sin pérdida de calidad. El balance entre nivel de detalle y claridad visual se optimiza mediante técnicas de simplificación progresiva que presentan información general inicialmente, permitiendo exploración de detalles mediante visualizaciones complementarias o interactivas.

Contextualizar las visualizaciones mediante elementos de referencia apropiados mejora sustancialmente su interpretabilidad. Las escalas claramente marcadas, las líneas de referencia significativas y las anotaciones explicativas situacional los datos dentro de marcos interpretativos relevantes. Esta contextualización debe incluir indicadores de significación estadística cuando aplique, así como puntos de comparación con promedios disciplinares o valores de referencia establecidos en la literatura especializada.

La transparencia metodológica en visualización exige la documentación completa de todas las transformaciones aplicadas a los datos crudos. Los procedimientos de suavizado, agregación o normalización deben explicarse detalladamente, permitiendo a los lectores comprender cómo las decisiones de procesamiento afectan los patrones visualizados. Esta transparencia se extiende a la declaración de software y parámetros específicos utilizados para generar cada visualización, facilitando la replicación y verificación independiente.

La evaluación iterativa de usabilidad mediante pruebas con representantes de la audiencia

objetivo identifica oportunidades de mejora en las visualizaciones. Este proceso de refinamiento continuo asegura que los elementos visuales comuniquen efectivamente los hallazgos entendidos a lectores con diferentes niveles de familiaridad con técnicas bibliométricas. La incorporación de retroalimentación sobre aspectos como interpretación de colores, comprensión de símbolos y navegación de visualizaciones interactivas optimiza la efectividad comunicativa final.

La accesibilidad universal representa un principio ético fundamental en visualización de datos bibliométricos. El diseño debe considerar necesidades diversas mediante la provisión de descripciones textuales alternativas, la garantía de suficiente contraste cromático y la evitación de dependencia exclusiva en el color para transmitir información crítica. Estas consideraciones de accesibilidad amplían el alcance impacto del estudio y alinean la práctica bibliométrica con estándares contemporáneos de inclusividad académica.

### **10.6. Otras secciones en estudios bibliométricos**

Los estudios bibliométricos pueden incorporar diversas secciones adicionales según los requisitos específicos de cada revista académica. Las conclusiones representan el componente más común, sintetizando los hallazgos principales y su relevancia para el campo de estudio. Esta sección debe ofrecer una visión concisa que trascienda la mera repetición de resultados, integrando las implicaciones teóricas, metodológicas y prácticas identificadas en la discusión. Las conclusiones efectivas establecen conexiones claras entre los objetivos iniciales y los logros del estudio, proyectando además direcciones significativas para investigación futura.

Los agradecimientos constituyen otra sección frecuente que reconoce contribuciones no suficientes para la autoría y fuentes de financiamiento. Esta sección debe especificar claramente el papel de cada colaborador mencionado y declarar cualquier posible conflicto de interés. El reconocimiento de apoyo técnico, asesoramiento especializado o acceso a infraestructuras de investigación enriquece la transparencia académica del estudio. Las fuentes de financiamiento se detallan mediante identificadores persistentes cuando están disponibles, facilitando la trazabilidad de apoyos institucionales.

Los apéndices y material suplementario albergan contenido que complementa la narrativa principal sin interrumpir su flujo argumental. Esta sección puede incluir metodologías detalladas, algoritmos de procesamiento, visualizaciones adicionales o conjuntos de datos extensos. La organización del material suplementario debe seguir una estructura lógica con referencias cruzadas claras desde el texto principal. Cada elemento incluido justifica su presencia mediante su valor para replicación o profundización del análisis presentado.

Algunas revistas especializadas requieren secciones específicas como implicaciones prácticas, resúmenes gráficos o declaraciones de disponibilidad de datos. Las implicaciones prácticas trasladan hallazgos bibliométricos a recomendaciones concretas para diferentes partes interesadas del sistema científico. Los resúmenes gráficos sintetizan visualmente las contribuciones principales del estudio, facilitando su diseminación entre canales académicos y profesionales. Las declaraciones de disponibilidad de datos detallan condiciones de acceso al set de datos utilizados, alineándose con los principios de ciencia abierta.

La estandarización de estas secciones adicionales varía significativamente entre disciplinas y revistas académicas. La revisión cuidadosa de las directrices para autores previa a la sumisión del manuscrito asegura el cumplimiento de requisitos específicos. La adaptación de la estructura del estudio a estos requerimientos editoriales optimiza su evaluación durante el proceso de revisión por pares. El siguiente capítulo examinará en profundidad los criterios de selección de



revistas destino y estrategias para navegar el proceso de publicación académica en bibliometría.

## **Recapitulando**

- La estructura IMRyD mejora claridad y coherencia.
- Métodos deben incluir fuentes, filtros y fechas de recolección.
- Publicar datos y código promueve transparencia.
- Las figuras deben ser autoexplicativas y reproducibles.
- El resumen debe incluir objetivo, método y resultados clave.
- Tablas suplementarias mantienen orden en el cuerpo del texto.
- La normalización debe ser explicada.
- Reconocer limitaciones aumenta credibilidad.
- Discusión debe relacionar hallazgos con literatura previa.
- RMarkdown y Jupyter integran texto y análisis.
- Material suplementario aumenta reproducibilidad.
- Claridad, precisión y estilo conciso son esenciales.
- Evitar sobreinterpretación de correlaciones.
- Seguir las normas de cada revista.
- Responder revisiones con profesionalismo.
- Declarar conflictos de interés.
- Incluir ORCID y afiliaciones normalizadas.
- Agradecer colaboraciones técnicas.
- Adaptar lenguaje según audiencia.
- Organizar archivos de trabajo sistemáticamente.

## **Preguntas de autoevaluación**

1. ¿Qué secciones conforman la estructura IMRyD?
2. ¿Qué debe contener la sección de Métodos?
3. ¿Por qué publicar datos y código?
4. ¿Qué caracteriza una figura autoexplicativa?
5. ¿Cómo estructurar un resumen efectivo?
6. ¿Cómo describir limitaciones en un manuscrito?
7. ¿Qué ventajas ofrece RMarkdown?
8. ¿Qué importancia tiene declarar conflictos?
9. ¿Cómo mejorar visibilidad del artículo?
10. ¿Qué incluir en material suplementario?

## **BIBLIOGRAFÍA**

1. Schimel J. Writing science: how to write papers that get cited and proposals that get funded. Oxford: Oxford University Press; 2012. ISBN 9780199760237
2. Alley M. The craft of scientific writing. 4th ed. New York (NY): Springer; 2018. ISBN 9783319695437
3. Day RA, Gastel B. How to write and publish a scientific paper. 6th ed. Santa Barbara (CA): Greenwood; 2012. ISBN 9781619250002
4. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: Wiley; 2009. ISBN 9780470057247





## Chapter 11 / Capítulo 11

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch11

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

## PUBLISH AND COMMUNICATE THE ARTICLE / PUBLICAR Y COMUNICAR EL ARTÍCULO

Una investigación bibliométrica solo alcanza su verdadera culminación cuando se completa el ciclo de publicación y comunicación científica. Este proceso transforma el trabajo individual en una contribución colectiva al acervo del conocimiento, permitiendo que la comunidad académica acceda, critique, valide y construya sobre sus hallazgos. La publicación formal constituye el mecanismo institucionalizado que garantiza la preservación, diseminación y legitimación del conocimiento generado, trascendiendo las limitaciones temporales y geográficas del contexto investigador original. Sin este paso esencial, incluso el estudio bibliométrico más riguroso y novedoso permanece como un ejercicio académico incompleto, privado de su potencial para influir en el desarrollo del campo.

La comunicación efectiva de los resultados bibliométricos representa un imperativo epistemológico que va más allá del mero requisito académico. Es a través del proceso de publicación que los hallazgos se someten al escrutinio crítico de pares especializados, se integran en las conversaciones disciplinares en curso y se convierten en fundamento para investigaciones futuras. La validación progresiva de un estudio bibliométrico ocurre precisamente a través de su utilización como sustrato teórico o metodológico por otros investigadores, quienes replican, extienden o cuestionan sus abordajes y conclusiones. Este proceso dialéctico de construcción colectiva del conocimiento confiere a la publicación científica su carácter esencial dentro del ecosistema investigador contemporáneo.

La selección del canal de publicación adecuado constituye una decisión estratégica que influye significativamente en el alcance, impacto y legitimación del estudio bibliométrico. Diferentes revistas y formatos de comunicación alcanzan audiencias distintas, operan bajo estándares metodológicos diversos y poseen niveles variables de prestigio e influencia dentro de la comunidad bibliométrica y las disciplinas aplicadas. Esta decisión inicial debe alinearse cuidadosamente con los objetivos de la investigación, las características metodológicas del estudio y las contribuciones sustantivas que se aspiran realizar al campo del conocimiento. El proceso de selección implica considerar múltiples dimensiones que se explorarán detalladamente en los siguientes epígrafes.

### 11.1. Selección de la revista

La identificación de revistas potenciales para la publicación de un estudio bibliométrico requiere un análisis sistemático de múltiples criterios interrelacionados. El alcance temático de la revista constituye el primer filtro esencial, determinando si la publicación se especializa en estudios bibliométricos, acepta investigaciones metodológicas en un campo disciplinar específico o privilegia aplicaciones bibliométricas en áreas temáticas particulares. La revisión de números recientes permite evaluar la alineación entre el manuscrito y el perfil editorial actual, identificando posibles cambios en las preferencias temáticas o metodológicas del equipo editorial. Esta evaluación inicial previene sumisiones inapropiadas que serían rechazadas inmediatamente por falta de ajuste al *scope* journalístico.

El análisis del prestigio e impacto de las revistas potenciales implica considerar métricas cuantitativas junto con evaluaciones cualitativas de influencia disciplinar. Los indicadores bibliométricos tradicionales como el factor de impacto, el *CiteScore* o el *SJR* proporcionan medidas comparativas de visibilidad e influencia, pero deben complementarse con evaluaciones de prestigio percibido dentro de la comunidad bibliométrica específica. La indexación en bases de datos relevantes como Web of Science, Scopus o bases de datos especializadas constituye

otro criterio fundamental, ya que determina la accesibilidad y descubribilidad del artículo publicado. Este análisis multidimensional del impacto potencial permite priorizar revistas que maximicen la visibilidad del estudio entre las audiencias target.

La evaluación de los tiempos editoriales y tasas de aceptación representa una consideración práctica crucial con implicaciones significativas para la diseminación oportuna de la investigación. Los tiempos promedios desde sumisión hasta decisión editorial y desde aceptación hasta publicación afectan directamente la actualidad de los hallazgos en campos de rápida evolución. La consulta de estadísticas editoriales públicas, cuando están disponibles, o la estimación a través del examen de fechas en artículos recientes proporciona puntos de interés valiosos sobre la eficiencia operativa de cada revista. Esta evaluación debe balancearse con consideraciones sobre calidad e impacto, reconociendo que procesos rigurosos de revisión típicamente requieren plazos más extensos.

El examen de los requisitos metodológicos y estándares de calidad esperados por cada revista identifica ajustes necesarios antes de la sumisión. Algunas publicaciones enfatizan innovación metodológica en bibliometría, mientras otras privilegian aplicaciones sustantivas en campos disciplinares específicos. La revisión de guías para autores y artículos recientes publicados revela expectativas sobre profundidad analítica, sofisticación metodológica y contribución al campo. Esta preparación proactiva aumenta significativamente las probabilidades de éxito durante el proceso de revisión por pares al asegurar que el manuscrito cumple con los estándares establecidos por la revista target.

La consideración de políticas de acceso abierto y costos de publicación asociados completa el análisis de viabilidad para la selección *journalística*. Las opciones de acceso dorado, híbrido o verde presentan diferentes ventajas en términos de alcance de lectores, cumplimiento de mandatos de financiadores y costos para los autores. La evaluación de exenciones potenciales, apoyos institucionales o inclusiones en acuerdos transformativos permite una decisión informada sobre la sostenibilidad económica de la publicación. Esta dimensión práctica adquiere especial relevancia en contextos de recursos limitados o cuando existen múltiples autores con diferentes capacidades de financiamiento.

La consulta de experiencias de colegas y revisiones en plataformas académicas proporciona *insights* cualitativos valiosos sobre el proceso editorial de revistas específicas. Las evaluaciones sobre la calidad de las revisiones, la equidad en las decisiones editoriales y el trato profesional a los autores complementan los criterios cuantitativos en la toma de decisión. Esta inteligencia colectiva, disponible a través de plataformas como Scopus, Web of Science o foros académicos especializados, ayuda a anticipar desafíos potenciales y seleccionar revistas con procesos editoriales rigurosos pero justos.

La elaboración de una lista priorizada de revistas potenciales, típicamente entre tres y cinco opciones, constituye la culminación del proceso de selección. Esta estrategia escalonada permite avanzar secuencialmente hacia publicaciones de mayor impacto si ocurren rechazos, minimizando pérdidas de tiempo entre sumisiones. La lista debe reflejar un balance realista entre aspiraciones de impacto y probabilidades de aceptación, considerando la novedad, solidez metodológica y contribución específica del estudio bibliométrico. Esta planificación metódica transforma la selección de revista de una decisión arbitraria en una estrategia deliberada para maximizar el impacto y alcance de la investigación bibliométrica.

## **11.2. Adaptar a la norma de la revista seleccionada**

La adaptación del manuscrito bibliométrico a las normas específicas de la revista target constituye un proceso meticuloso que trasciende el mero formato técnico. Esta adaptación implica comprender profundamente las expectativas editoriales en términos de estructura, estilo discursivo y enfoque metodológico característicos de cada publicación. El proceso inicia con el estudio detallado de las instrucciones para autores, donde se especifican requisitos sobre longitud máxima, estructura de secciones, formato de referencias y convenciones de citación. Esta fase preparatoria resulta crucial para evitar rechazos administrativos y demostrar respeto por los estándares establecidos por la publicación.

La reestructuración del contenido según los patrones organizativos preferidos por la revista representa el primer paso sustantivo en la adaptación. Algunas publicaciones privilegian estructuras IMRaD tradicionales, mientras otras favorecen formatos más flexibles que priorizan la narrativa sobre la rigidez seccional. El análisis de artículos recientemente publicados en la revista revela estos patrones estructurales implícitos, permitiendo alinear la presentación del estudio bibliométrico con las expectativas del equipo editorial. Esta adaptación estructural facilita la evaluación durante el proceso de revisión al presentar la información en formatos familiares para los editores y revisores.

El ajuste del estilo discursivo y nivel técnico según el perfil de la audiencia target asegura la comunicabilidad efectiva del manuscrito. Revistas especializadas en bibliometría permiten un lenguaje técnico denso y presuponen familiaridad con conceptos metodológicos avanzados, mientras publicaciones disciplinares requieren mayor contextualización de métodos bibliométricos y traducción de hallazgos a implicaciones sustantivas para el campo. Esta adaptación lingüística afecta desde la elección de terminología hasta el nivel de detalle en la explicación de procedimientos metodológicos, buscando siempre el equilibrio entre precisión técnica y accesibilidad para la audiencia prevista.

La reformulación de tablas, figuras y elementos visuales según los requisitos técnicos y estéticos de la revista garantiza su efectividad comunicativa en el contexto de publicación específico. Las especificaciones sobre formatos de archivo, resoluciones, esquemas de color y convenciones de etiquetado deben implementarse meticulosamente durante esta fase. La revisión de visualizaciones en números recientes proporciona guía adicional sobre estilos preferidos, nivel de complejidad visual aceptable y estrategias efectivas de integración texto-figura características de la publicación. Esta atención al detalle visual demuestra profesionalismo y facilita la producción editorial final.

La adaptación del sistema de referencias y citas a las normas bibliográficas específicas de la revista completa la estandarización formal del manuscrito. La verificación minuciosa de cada referencia según el estilo requerido, incluyendo formato de autores, títulos, fuentes y elementos de puntuación, constituye un aspecto aparentemente menor pero crítico para la percepción de calidad y meticulosidad del trabajo. La utilización de gestores bibliográficos facilita esta transición entre estilos, pero requiere posterior verificación manual para detectar y corregir inconsistencias automáticas. Esta precisión bibliográfica refleja el rigor académico característico de investigaciones de alta calidad.

La adecuación del resumen y palabras clave a las convenciones específicas de la revista optimiza la visibilidad y recuperación del artículo una vez publicado. Algunas publicaciones requieren resúmenes estructurados con secciones predeterminadas, mientras otras prefieren formatos narrativos tradicionales. La selección de palabras clave debe reflejar tanto la

terminología técnica bibliométrica como los términos disciplinares relevantes para la audiencia objetivo, facilitando el cruce de fronteras disciplinares cuando sea apropiado. Esta optimización de metadatos maximiza el impacto potencial del estudio al asegurar su visibilidad para las comunidades académicas más relevantes.

Revisar exhaustivamente de aspectos éticos y declaraciones requeridas completa el proceso de adaptación a las normas. La verificación del cumplimiento con políticas específicas sobre autoría, conflictos de interés, disponibilidad de datos y aprobaciones éticas, cuando aplican, constituye un requisito fundamental para la aceptación editorial. La inclusión de declaraciones estandarizadas en formatos y ubicaciones especificadas por la revista demuestra compromiso con la integridad académica y facilita la evaluación de cumplimiento normativo durante el proceso editorial. Esta atención al detalle ético complementa la adaptación técnica y sustantiva del manuscrito.

La implementación de todas las adaptaciones requiere una revisión final minuciosa que asegure coherencia interna a través de todo el manuscrito transformado. Los cambios en estructura, estilo o formato pueden introducir inconsistencias que afectan la fluidez argumental y la claridad expositiva. Esta revisión holística verifica que las modificaciones realizadas han mejorado, no simplemente alterado, la presentación del estudio bibliométrico, manteniendo su integridad intelectual mientras optimizan su ajuste a las expectativas y estándares de la revista seleccionada. Este proceso iterativo de refinamiento asegura que el manuscrito representa la mejor posible versión de la investigación para el contexto de publicación específico.

### **11.3. Presentación de la información**

Comunicar efectivamente de resultados bibliométricos trasciende la publicación escrita para incluir diversas modalidades de presentación que amplían el alcance e impacto de la investigación. Cada formato de presentación requiere estrategias específicas de adaptación que consideren las características de la audiencia, el contexto de diseminación y los objetivos comunicativos particulares. La versatilidad en la presentación permite dirigir los hallazgos bibliométricos a múltiples partes interesadas, desde comités científicos especializados hasta tomadores de decisiones con formación técnica limitada en bibliometría. Esta flexibilidad comunicativa resulta esencial para maximizar la utilidad práctica y el impacto social de la investigación bibliométrica.

#### **11.3.1. Presentación en diapositivas**

Las presentaciones en diapositivas constituyen el formato más extendido para la comunicación oral de resultados bibliométricos en congresos, seminarios y reuniones institucionales. El diseño efectivo de diapositivas equilibra contenido sustantivo con claridad visual, utilizando el principio de minimalismo informativo para evitar la saturación cognitiva del auditorio. Cada diapositiva debe transmitir una idea principal mediante combinaciones estratégicas de texto conciso, visualizaciones impactantes y elementos gráficos que guíen la atención hacia los hallazgos más relevantes. Esta economía visual facilita la comprensión auditiva simultánea durante la exposición oral.

La narrativa de la presentación debe estructurarse según una progresión lógica que contextualice, evidencie e interprete los hallazgos bibliométricos. La secuencia típica inicia con la motivación y preguntas de investigación, continúa con la metodología de forma sucinta, presenta resultados clave mediante visualizaciones progresivas y culmina con implicaciones y conclusiones. Las transiciones entre secciones deben establecer conexiones explícitas que mantengan la coherencia argumental a través de toda la presentación. Esta estructura narrativa

transforma datos complejos en una historia investigadora comprensible y memorable.

La adaptación del nivel técnico al perfil específico de la audiencia determina el éxito comunicativo de la presentación. Para audiencias especializadas en bibliometría, puede profundizarse en aspectos metodológicos y técnicos, mientras para audiencias disciplinares deben privilegiarse las implicaciones sustantivas para el campo. El lenguaje debe modularse según este criterio, evitando jerga técnica innecesaria o, por el contrario, proporcionando definiciones esenciales cuando se presenten conceptos bibliométricos especializados. Esta sensibilidad audiencia asegura que el mensaje central trascienda las barreras disciplinares.

La integración efectiva de elementos visuales requiere consideraciones específicas para el formato de proyección. Las visualizaciones deben simplificarse respecto a las versiones publicadas, eliminando detalles secundarios y ampliando elementos críticos para garantizar legibilidad a distancia. El uso estratégico de animaciones controladas puede revelar progresivamente capas de complejidad en visualizaciones multidimensionales, guiando la atención del auditorio a través de los patrones más significativos. Esta optimización visual para proyección diferencia las diapositivas efectivas de la mera transferencia de figuras desde el manuscrito escrito.

La preparación para la interacción con el auditorio completa el diseño de presentaciones exitosas. La anticipación de preguntas frecuentes, la preparación de diapositivas complementarias con información adicional y el dominio de posibles objeciones metodológicas constituyen elementos esenciales para manejar competentemente sesiones de preguntas y respuestas. Esta preparación integral transforma la presentación de resultados bibliométricos de un monólogo informativo en un diálogo académico productivo que enriquece la interpretación y contextualización de los hallazgos presentados.

### **11.3.2. Informes reproducibles con RMarkdown y Google Colab**

Los informes reproducibles representan un estándar contemporáneo de transparencia metodológica en investigación bibliométrica, integrando código, resultados y narrativa interpretativa en documentos autocontenidos. RMarkdown permite generar informes dinámicos que ejecutan análisis bibliométricos directamente desde el código R, actualizando automáticamente resultados y visualizaciones ante modificaciones en los datos o parámetros analíticos. Esta integración garantiza la consistencia entre los análisis reportados y el código subyacente, facilitando la verificación y extensión del estudio por otros investigadores.

La estructuración de informes reproducibles sigue principios de organización modular que separan claramente las fases de carga de datos, preprocesamiento, análisis y visualización. Cada sección del informe combina fragmentos de código ejecutable con texto explicativo que contextualiza las operaciones realizadas e interpreta los resultados generados. Esta integración de código y prosa narrativa transforma el informe de un mero reporte estático en un documento ejecutable que documenta exhaustivamente el flujo analítico completo del estudio bibliométrico.

Google Colab ofrece una alternativa basada en Python para la creación de informes reproducibles accesibles a través de navegadores web sin requerir instalación local de software. La capacidad de ejecutar código en infraestructura en la nube facilita el procesamiento de grandes volúmenes de datos bibliográficos y el compartir fácilmente notebooks interactivos con colaboradores. La integración con Google Drive simplifica la gestión de sets de datos bibliométricos y la versionización de los análisis realizados, particularmente valiosa en proyectos colaborativos distribuidos.

La implementación de principios de reproducibilidad en informes bibliométricos incluye la gestión explícita de dependencias, la documentación de versiones de software y la semilla de aleatorización para análisis que involucren componentes estocásticos. Estas prácticas permiten la replicación exacta de los análisis meses o años después de su ejecución inicial, dirigiendo preocupaciones fundamentales sobre la sostenibilidad temporal de la investigación computacional. Esta atención al detalle técnico diferencia los informes profesionalmente reproducibles de meros scripts académicos.

La publicación de informes reproducibles junto con los manuscritos tradicionales enriquece sustancialmente el ecosistema de comunicación bibliométrica. Plataformas como GitHub, Zenodo o repositorios institucionales especializados facilitan el archivo permanente y citación formal de estos recursos complementarios. Esta práctica emergente establece nuevos estándares de transparencia metodológica en el campo, permitiendo a la comunidad bibliométrica construir de manera más eficiente sobre trabajo previo al acceder directamente a las implementaciones computacionales subyacentes.

### **11.3.3. Paneles interactivos con Shiny, Dash y Streamlit**

Los paneles interactivos transforman la visualización de resultados bibliométricos de representaciones estáticas a experiencias exploratorias dinámicas que permiten a usuarios personalizar perspectivas analíticas según intereses específicos. Shiny (R), Dash (Python) y Streamlit (Python) constituyen los *frameworks* predominantes para desarrollar aplicaciones web interactivas que comunican hallazgos bibliométricos a través de interfaces accesibles vía navegadora. Estas herramientas permiten crear paneles donde usuarios pueden filtrar datos, ajustar parámetros y alternar entre visualizaciones sin requerir conocimientos de programación.

El diseño efectivo de paneles bibliométricos equilibra capacidades exploratorias avanzadas con interfaces intuitivas que minimizan la curva de aprendizaje para usuarios finales. La organización típica incluye paneles de control con widgets de filtrado (deslizadores, menús desplegables, selectores de rango), áreas de visualización principal que responden dinámicamente a las interacciones del usuario, y paneles informativos que contextualizan los datos mostrados. Esta arquitectura modular guía a usuarios a través del espacio analítico sin abrumarlos con complejidad innecesaria.

La implementación de paneles bibliométricos requiere consideraciones específicas sobre el volumen y estructura de datos subyacentes. La optimización del rendimiento a través de técnicas como pre-agregación de métricas, muestreo estratégico para visualizaciones de grandes sets de datos e implementación de almacenamiento en caché inteligente asegura experiencias de usuario responsivas incluso con corpus bibliográficos extensos. Estas consideraciones técnicas resultan críticas para el éxito de paneles destinados a explorar colecciones documentales de escala considerable típicas en bibliometría.

La personalización de paneles según perfiles específicos de usuarios maximiza su utilidad práctica para diferentes partes interesadas del sistema científico. Paneles para gestores de investigación pueden enfatizar indicadores de productividad y colaboración institucional, mientras versiones para investigadores pueden priorizar herramientas de exploración temática y identificación de tendencias emergentes. Esta adaptación contextual transforma paneles genéricos en herramientas especializadas que direccionan necesidades informativas concretas de diferentes comunidades usuarias.

El despliegue y mantenimiento de paneles interactivos constituye una consideración operativa



esencial para su sostenibilidad temporal. Plataformas como shinyapps.io, Heroku o servidores institucionales proporcionan infraestructura de hosting con diferentes ventajas en términos de costo, control y escalabilidad. La implementación de estrategias de actualización automática de datos, monitorización de uso y evolución iterativa basada en la retroalimentación de usuarios asegura que los paneles mantengan relevancia y utilidad más allá del ciclo de vida inicial del proyecto de investigación.

#### **11.3.4. Pósteres científicos para congresos**

Los posters científicos representan un formato intermedio entre publicaciones escritas y presentaciones orales, ideal para recibir retroalimentación focalizado durante eventos académicos. El diseño efectivo de póster bibliométricos equilibra densidad informativa con claridad visual, organizando el contenido en columnas lógicas que guíen al espectador a través de la narrativa investigadora. El uso estratégico de jerarquías tipográficas, elementos visuales de alto impacto y espacio en blanco suficiente diferencia póster profesionales de meras compilaciones abrumadoras de texto y figuras.

La adaptación de visualizaciones bibliométricas al formato de póster requiere modificaciones específicas para garantizar legibilidad a distancias cortas. Las figuras deben simplificarse respecto a versiones publicadas, ampliando elementos críticos y utilizando paletas de color de alto contraste optimizadas para impresión. La inclusión de títulos descriptivos y leyendas autoexplicativas permite que cada visualización comunique efectivamente incluso sin mediación verbal inmediata del investigador. Esta autonomía visual resulta crucial durante sesiones de póster con alta circulación de asistentes.

La preparación de materiales complementarios enriquece la interacción durante sesiones de póster. Resúmenes extendidos, tarjetas de contacto con enlaces a recursos digitales y versiones reducidas del póster para distribución facilitan el seguimiento posterior de conversaciones iniciadas durante el evento. Esta capa adicional de recursos transforma la presentación de póster de un evento aislado en el inicio de posibles colaboraciones futuras basadas en los hallazgos bibliométricos presentados.

#### **11.3.5. Resúmenes ejecutivos para tomadores de decisiones**

Los resúmenes ejecutivos trasladan hallazgos bibliométricos a formatos accesibles para audiencias de gestión y política científica con limitaciones de tiempo y formación técnica especializada. Estos documentos condensan resultados complejos en narrativas concisas que destacan implicaciones prácticas, oportunidades identificadas y recomendaciones accionables. El lenguaje debe privilegiar claridad sobre tecnicismo, traduciendo conceptos bibliométricos especializados a consecuencias comprensibles para la planificación y evaluación de sistemas de investigación.

La estructuración de resúmenes ejecutivos sigue convenciones establecidas que priorizan mensajes clave en secciones iniciales, soportando por evidencias sintéticas en secciones subsiguientes. La inclusión de visualizaciones de alto impacto, métricas contextualizadas mediante comparativas con puntos de referencia relevantes y destacados de hallazgos contraintuitivos aumenta la efectividad comunicativa de estos documentos. Esta economía expositiva diferencia resúmenes ejecutivos efectivos de meras simplificaciones superficiales de análisis bibliométricos complejos.

La distribución estratégica de resúmenes ejecutivos completa el proceso de transferencia a audiencias no académicas. La identificación de canales formales e informales dentro de

estructuras de gobernanza científica, la adaptación de formatos según protocolos institucionales específicos y el seguimiento para evaluar utilización de los hallazgos maximizan el impacto potencial de la investigación bibliométrica en procesos de toma de decisiones. Esta proyección práctica trasciende la comunicación académica tradicional para influir directamente en la evolución de sistemas de investigación.

Al concluir esta parte dedicada a la publicación y comunicación de estudios bibliométricos, se establecen las bases para que la investigación trascienda el ámbito académico tradicional y genere impacto tangible en la sociedad. La cuidadosa selección de revistas, la adaptación a normativas editoriales y el dominio de múltiples formatos de divulgación constituyen competencias esenciales que garantizan la efectiva transferencia del conocimiento bibliométrico hacia diferentes audiencias y contextos de aplicación. Es precisamente esta capacidad de conectar la rigurosidad metodológica con las necesidades del mundo real la que da paso a la siguiente dimensión exploratoria: la aplicación práctica de la bibliometría en escenarios emergentes y su creciente influencia en los ecosistemas de ciencia, tecnología e innovación.

### **Recapitulando**

- Elegir la revista según su alcance, audiencia y política editorial.
- Leer las instrucciones para autores antes de enviar el manuscrito.
- Redactar una carta de presentación clara y concisa dirigida al editor.
- Utilizar *preprints* para difusión temprana cuando la política editorial lo permita.
- Seleccionar revisores sugeridos de forma ética y responsable.
- Comprender los tipos de revisión por pares y sus implicaciones.
- Depositar datos y código en repositorios con DOI para reproducibilidad.
- Monitorear el impacto mediante altmetrics y descargas.
- Adaptar figuras y materiales gráficos para diferentes audiencias.
- Preparar un resumen de divulgación para prensa o redes.
- Etiquetar adecuadamente los metadatos del artículo.
- Evaluar costos y opciones de acceso abierto.
- Mantener versiones de manuscrito y correspondencia archivadas.
- Promover ciencia abierta mediante datos y código compartidos.
- Monitorear el rendimiento de difusión en redes académicas.
- Participar en foros y congresos para incrementar visibilidad.
- Cuidar el cumplimiento de políticas de ética y derechos de autor.
- Evitar autopromoción engañosa o manipulación de métricas.
- Documentar la estrategia de comunicación post-publicación.
- Evaluar impacto académico y social como parte del proceso de cierre.

### **Preguntas de autoevaluación**

1. ¿Qué criterios ayudan a elegir la revista adecuada?
2. ¿Qué elementos debe tener una carta de presentación?
3. ¿Qué ventajas y riesgos tiene publicar un preprint?
4. ¿Cómo se eligen los revisores sugeridos de manera ética?
5. ¿Qué beneficios aporta depositar datos y código en repositorios?
6. ¿Qué mide altmetrics y cómo se diferencia de las citas tradicionales?
7. ¿Qué cuidados éticos deben tenerse en la difusión pública de resultados?
8. ¿Qué tipos de revisión por pares existen?
9. ¿Cómo incrementar visibilidad sin violar políticas editoriales?
10. ¿Por qué conviene conservar todas las versiones y correspondencias editoriales?


## **BIBLIOGRAFÍA**

1. Gasparyan AY, Yessirkepov M, Voronov AA, Gerasimov AN, Kostyukova EI, Kitash GD. Scientific publishing: process, practices, and ethics. Cham: Springer; 2019. ISBN 9783030218097
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. ISBN 9780190640125
4. Tennant JP, Ross-Hellauer T, et al., editors. The state of the art in scholarly communication. Cambridge (MA): MIT Press; 2020. ISBN 9780262539371

# Part V / Parte V

APPLIED BIBLIOMETRICS

BIBLIOMETRÍA APLICADA



## Chapter 12 / Capítulo 12

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch12

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

## EMERGING TRENDS / TENDENCIAS EMERGENTES

El panorama bibliométrico contemporáneo experimenta una transformación radical impulsada por la convergencia de tecnologías disruptivas que expanden sustancialmente las fronteras de lo analizable e interpretable. Estas innovaciones tecnológicas no solo optimizan metodologías establecidas, sino que fundamentalmente redefinen las preguntas de investigación que pueden formularse y responderse a través del análisis cuantitativo de la literatura científica. Desde la inteligencia artificial generativa hasta las tecnologías de registro distribuido, las tendencias emergentes prometen superar limitaciones históricas del campo mientras introducen nuevos desafíos éticos y metodológicos que la comunidad bibliométrica debe abordar con rigor y anticipación. Este capítulo examina críticamente estas tendencias transformadoras, evaluando tanto su potencial analítico como sus implicaciones para el futuro de la evaluación científica.

### 12.1. IA generativa: GPT en minería de textos científicos

Los modelos de lenguaje de gran escala como GPT representan una innovación disruptiva en el análisis bibliométrico al permitir la comprensión contextual de textos científicos a escala masiva. A diferencia de los métodos tradicionales de minería de texto basados en frecuencia de términos o análisis de co-ocurrencia, estos modelos capturan relaciones semánticas complejas y matices interpretativos que anteriormente requerían intervención humana especializada. Esta capacidad transforma sustancialmente el análisis de contenido bibliométrico, permitiendo desde la clasificación automática de artículos según dimensiones epistemológicas complejas hasta la identificación de patrones argumentativos y retóricos que trascienden la mera presencia de términos específicos. La aplicación de estos modelos a corpus bibliográficos completos abre posibilidades analíticas sin precedentes para comprender la evolución discursiva de los campos científicos.

La implementación de IA generativa en bibliometría permite la extracción automatizada de dimensiones analíticas tradicionalmente inaccesibles a través de métodos cuantitativos convencionales. Los modelos pueden identificar automáticamente contribuciones teóricas, innovaciones metodológicas y hallazgos empíricos clave en publicaciones científicas, clasificando artículos según taxonomías complejas sin requerir diccionarios terminológicos predefinidos. Esta flexibilidad analítica resulta particularmente valiosa para estudiar campos interdisciplinarios donde los vocabularios especializados evolucionan rápidamente y donde las contribuciones intelectuales trascienden las categorizaciones disciplinarias establecidas. La capacidad de estos sistemas para generar resúmenes analíticos sintetizando hallazgos de múltiples publicaciones amplifica aún más su utilidad para revisiones bibliométricas de grandes volúmenes de literatura.

El análisis de tendencias temáticas mediante IA generativa supera significativamente las limitaciones de los métodos basados en palabras clave al capturar evoluciones conceptuales y semánticas más allá de la mera frecuencia terminológica. Los modelos pueden rastrear cómo ciertos conceptos adquieren significados diferentes en distintos contextos temporales o disciplinarios, identificando procesos de recontextualización teórica o apropiación conceptual entre campos. Esta capacidad para mapear la deriva semántica de ideas científicas proporciona perspectivas únicas sobre la dinámica intelectual de los campos de investigación, revelando cómo los marcos conceptuales se transforman a través de su circulación al lado de diferentes comunidades epistémicas. Este análisis diacrónico de la evolución semántica constituye una contribución metodológica fundamental de la IA generativa a la bibliometría contemporánea.

La generación automática de hipótesis de investigación representa otra aplicación transformadora donde los sistemas de IA analizan patrones en la literatura científica para

identificar vacíos de conocimiento y oportunidades de investigación promisorias. Al procesar corpus documentales completos, estos sistemas pueden detectar conexiones no evidentes entre hallazgos aparentemente no relacionados, sugerir combinaciones innovadoras de enfoques metodológicos o identificar problemas de investigación infraexplorados con alto potencial de impacto. Esta capacidad de síntesis creativa al lado de dominios de conocimiento tradicionalmente separados posiciona a la IA generativa como herramienta valiosa para la planificación estratégica de la investigación y la identificación de fronteras científicas emergentes.

La evaluación de la calidad argumentativa y solidez metodológica mediante IA generativa introduce dimensiones cualitativas previamente inaccesibles a escala bibliométrica. Los modelos pueden analizar sistemáticamente la estructura argumental de las publicaciones, evaluar la adecuación entre preguntas de investigación y métodos empleados, e identificar limitaciones metodológicas recurrentes en literaturas específicas. Esta capacidad para evaluar dimensiones de calidad más allá de las citas recibidas ofrece alternativas prometedoras para sistemas de evaluación científica que trasciendan las métricas de impacto tradicionales, aunque introduce importantes desafíos de validación y transparencia algorítmica.

La implementación de IA generativa en bibliometría enfrenta desafíos significativos relacionados con la transparencia metodológica, los sesgos en los datos de entrenamiento y la reproducibilidad de los análisis. Los modelos de lenguaje exhiben tendencias a generar alucinaciones factuales, reproducir sesgos presentes en sus datos de entrenamiento y mostrar inconsistencia en respuestas al otro lado de diferentes solicitudes. Estos desafíos requieren el desarrollo de protocolos de validación rigurosos, la documentación exhaustiva de parámetros y *prompts* utilizados, y la implementación de mecanismos de explicabilidad que permitan comprender los fundamentos de las clasificaciones y análisis generados automáticamente.

La integración de IA generativa con métodos bibliométricos tradicionales establece un paradigma analítico híbrido que combina las fortalezas de ambos enfoques. Las técnicas establecidas de análisis de citas y redes proporcionan validación cuantitativa para las perspectivas cualitativas generados por IA, mientras el análisis semántico avanzado contextualiza y enriquece la interpretación de patrones estructurales identificados a través de métodos tradicionales. Esta integración metodológica produce análisis bibliométricos más robustos y matizados que aprovechan lo mejor de ambos paradigmas analíticos, estableciendo nuevos estándares para la comprensión integral de la dinámica científica.

El futuro desarrollo de IA generativa aplicada a bibliometría probablemente se orientará hacia modelos especializados en dominios científicos específicos, entrenados en corpus disciplinarios comprehensivos y capaces de comprender convenciones retóricas y epistemológicas particulares de cada campo. La evolución hacia sistemas multimodales que integren análisis de texto, imágenes, datos y código en publicaciones científicas ampliará aún más las fronteras del análisis bibliométrico, permitiendo una comprensión más integral de la comunicación científica contemporánea. Esta especialización progresiva condiciona limitaciones actuales de los modelos generales mientras maximiza su utilidad para aplicaciones bibliométricas específicas.

## 12.2. *Blockchain* para métricas descentralizadas

La tecnología *blockchain* surge como una solución transformadora para abordar desafíos fundamentales de transparencia, confianza y descentralización en los sistemas de métricas académicas. A diferencia de las bases de datos centralizadas tradicionales, los registros distribuidos permiten crear sistemas de métricas incorruptibles donde cada transacción académica, citas, revisiones, contribuciones, queda registrada de manera inmutable y verificable



por cualquier participante de la red. Esta arquitectura descentralizada elimina puntos únicos de fallo y reduce la dependencia de intermediarios institucionales, estableciendo las bases para un ecosistema de métricas más resiliente y confiable. La aplicación de *blockchain* en bibliometría representa un cambio paradigmático hacia sistemas de evaluación científica más transparentes y resistentes a la manipulación.

La implementación de *blockchain* permite crear registros descentralizados de contribuciones académicas que capturan la totalidad del proceso investigador, no solo sus productos finales. Cada contribución, desde la conceptualización inicial y diseño metodológico hasta el análisis de datos y redacción, puede registrarse con marca temporal de manera inalterable en la cadena de bloques, creando trazas auditables de autoría y contribución intelectual. Este enfoque resuelve problemas crónicos de atribución de autoría y reconocimiento de contribuciones no tradicionales, particularmente valioso en investigación colaborativa donde múltiples investigadores contribuyen en diferentes capacidades a lo largo del ciclo de investigación. La granularidad en el registro de contribuciones permite métricas más justas y precisas que reflejan la naturaleza compleja y colaborativa de la ciencia contemporánea.

Los sistemas de métricas basados en *blockchain* introducen mecanismos innovadores para verificación de citas y prevención de manipulación métrica. Cada cita puede registrarse como una transacción en la cadena, creando un historial transparente e inmutable de relaciones de influencia intelectual. Este registro distribuido permite detectar patrones de citación circular, redes de auto citas coordinadas y otras prácticas manipulativas que distorsionan las métricas tradicionales. La verificación descentralizada mediante consenso de red elimina la dependencia de bases de datos centralizadas cuyos algoritmos de procesamiento y limpieza frecuentemente operan como cajas negras inescrutables para los usuarios finales.

Los tokens académicos y sistemas de reputación descentralizados representan otra aplicación disruptiva de *blockchain* en bibliometría. Estos sistemas permiten a investigadores acumular reputación basada en contribuciones verificadas registradas en la cadena, creando capital académico portable e independiente de instituciones específicas. Los mecanismos de tokenización pueden alinear incentivos individuales con objetivos colectivos del sistema científico, recompensando comportamientos académicos valiosos como revisiones por pares rigurosas, compartir datos de investigación o replicar estudios previos. Este enfoque transforma fundamentalmente los sistemas de recompensa académica al hacer explícitos y cuantificables valores tradicionalmente implícitos en la cultura científica.

La gestión descentralizada de identidades académicas mediante *blockchain* resuelve problemas persistentes de desambiguación y portabilidad de perfiles investigador. Cada investigador puede mantener una identidad auto soberana registrada en la cadena que consolida sus contribuciones a través de diferentes instituciones, plataformas y períodos temporales. Esta identidad persistente e independiente de proveedores específicos facilita la movilidad académica y elimina la fragmentación métrica que ocurre cuando investigadores cambian de afiliación institucional. La integración con sistemas ORCID existentes y otros identificadores académicos crea un ecosistema de identidad más robusto y centrado en el usuario.

Los contratos inteligentes permiten automatizar procesos de evaluación científica basados en métricas verificadas en la cadena. Estos protocolos auto ejecutables pueden implementar criterios de evaluación transparentes y automatizar procesos como revisión por pares, asignación de fondos o promociones académicas basadas en métricas objetivas registradas en *blockchain*. La automatización mediante contratos inteligentes reduce la carga administrativa en evaluación

científica mientras aumenta la consistencia y transparencia de los procesos decisorios. Sin embargo, esta automatización requiere diseño cuidadoso para evitar codificar sesgos existentes o crear sistemas excesivamente rígidos que no capturen dimensiones cualitativas importantes de la excelencia científica.

La implementación de métricas descentralizadas enfrenta desafíos significativos de escalabilidad, adopción y gobernanza. El registro de cada transacción académica en la cadena genera volúmenes masivos de datos que deben procesarse eficientemente, mientras la adopción por comunidades académicas diversas requiere superar inercias institucionales y estandarizar protocolos a través de diferentes disciplinas y regiones. Los modelos de gobernanza para estos sistemas descentralizados deben balancear participación comunitaria con eficiencia decisional, evitando tanto la captura por intereses particulares como la parálisis por exceso de deliberación. Estos desafíos prácticos requieren atención cuidadosa durante el diseño e implementación de sistemas bibliométricos basados en *blockchain*.

La integración de *blockchain* con otras tecnologías emergentes como inteligencia artificial e internet de las cosas crea ecosistemas métricos aún más sofisticados y comprehensivos. Los sistemas de IA pueden analizar los datos inmutables registrados en la cadena para generar conocimientos más confiables, mientras dispositivos *Internet of Things* (IoT) pueden automatizar el registro de actividades investigadoras en laboratorios y entornos de campo. Esta convergencia tecnológica potencia las capacidades de cada tecnología individual mientras mitiga sus limitaciones específicas, creando infraestructuras métricas más robustas, transparentes y alineadas con las prácticas reales de investigación científica en diferentes dominios y metodologías.

### 12.3. Análisis de imágenes y datos multimodales

El análisis de imágenes y datos multimodales representa una frontera emergente en bibliometría que expande significativamente el alcance de lo cuantificable en la comunicación científica. Tradicionalmente, los estudios bibliométricos se han centrado predominantemente en el análisis de texto, dejando de lado el rico contenido visual que acompaña a las publicaciones académicas contemporáneas. Las imágenes, gráficos, diagramas y visualizaciones constituyen componentes esenciales del discurso científico moderno, transmitiendo información compleja que frecuentemente no puede expresarse adecuadamente mediante solo texto. La capacidad de analizar sistemáticamente estos elementos visuales abre nuevas dimensiones para comprender cómo se construye y comunica el conocimiento en diferentes disciplinas científicas.

Las técnicas de visión artificial y aprendizaje profundo permiten extraer información significativa de elementos visuales en publicaciones científicas a escala masiva. Los algoritmos pueden identificar y clasificar automáticamente diferentes tipos de imágenes, desde gráficos estadísticos y diagramas de flujo hasta microfotografías y visualizaciones tridimensionales, estableciendo patrones de uso según disciplinas, metodologías o épocas específicas. Esta capacidad analítica transforma las imágenes de meras ilustraciones en datos cuantificables que revelan preferencias representacionales, estándares de visualización y evoluciones en las prácticas de comunicación visual dentro de diferentes comunidades científicas. El análisis automatizado de estos elementos visuales complementa los enfoques textuales tradicionales, ofreciendo una perspectiva más comprehensiva de la retórica visual en la ciencia.

El análisis multimodal integra información proveniente de diferentes formatos, texto, imágenes, tablas, ecuaciones, para generar entendimientos más ricos y contextualizados del contenido científico. En lugar de analizar cada modalidad por separado, los enfoques

multimodales capturan las interrelaciones entre diferentes formas de representación del conocimiento, revelando cómo los investigadores combinan recursos semióticos diversos para construir argumentos científicos persuasivos. Esta integración permite, por ejemplo, analizar cómo las descripciones textuales se relacionan con las visualizaciones correspondientes, o cómo las ecuaciones matemáticas se complementan con explicaciones narrativas y representaciones gráficas. La bibliometría multimodal captura así la naturaleza esencialmente híbrida de la comunicación científica contemporánea.

La detección de tendencias metodológicas a través del análisis de imágenes constituye una aplicación particularmente valiosa para el mapeo científico. Ciertos tipos de imágenes funcionan como indicadores confiables de enfoques metodológicos específicos, las microfotografías electrónicas sugieren ciertas técnicas de laboratorio, mientras los diagramas de circuitos señalan aproximaciones experimentales particulares en ingeniería. El análisis diacrónico de la prevalencia de diferentes tipos de imágenes puede revelar la adopción, consolidación o declive de metodologías de investigación, proporcionando perspectivas sobre la evolución de las prácticas científicas que complementan los análisis basados en terminología textual. Esta capacidad para rastrear metodologías a través de sus huellas visuales representa una contribución única del análisis multimodal a la bibliometría.

La evaluación de la calidad y claridad visual en publicaciones científicas emerge como otra aplicación prometedora del análisis multimodal. Los algoritmos pueden evaluar aspectos como la resolución de imágenes, la adecuación de las escalas, la claridad de las etiquetas y la efectividad comunicativa de diferentes tipos de visualizaciones. Esta evaluación automatizada de la calidad visual permite estudios a escala de la evolución de los estándares de visualización científica, la identificación de mejores prácticas en comunicación visual y el análisis de cómo la calidad visual se correlaciona con el impacto científico medido a través de citas y otros indicadores. Esta dimensión cualitativa tradicionalmente subjetiva se transforma así en objeto de análisis cuantitativo sistemático.

El análisis de datos multimodales enfrenta desafíos metodológicos significativos relacionados con la estandarización, interpretación y validación de resultados. La enorme diversidad de formatos visuales en disciplinas científicas complica el desarrollo de taxonomías y algoritmos universalmente aplicables. Además, la interpretación de elementos visuales frecuentemente requiere conocimiento especializado del dominio, haciendo difícil la automatización completa del análisis sin perder matices disciplinares cruciales. La validación de los resultados del análisis multimodal requiere estrategias innovadoras que combinen evaluación automática con verificación por expertos humanos, estableciendo estándares de confiabilidad para esta nueva frontera bibliométrica.

La integración del análisis multimodal con otras tendencias emergentes como la IA generativa y *blockchain* crea sinergias particularmente poderosas. Los modelos de lenguaje multimodal pueden generar descripciones textuales de contenido visual o crear visualizaciones a partir de descripciones textuales, facilitando la accesibilidad y búsqueda de contenido científico multimodal. Simultáneamente, *blockchain* puede proporcionar la infraestructura para registrar y verificar la procedencia de imágenes y datos multimodales, direccionando preocupaciones crecientes sobre la manipulación de imágenes científicas. Estas integraciones tecnológicas amplifican el potencial transformador del análisis multimodal en bibliometría.

El futuro desarrollo del análisis bibliométrico multimodal probablemente evolucionará hacia sistemas capaces de capturar dimensiones cada vez más sofisticadas de la comunicación científica

integrada. La capacidad de analizar no solo imágenes estáticas sino también visualizaciones interactivas, videos científicos y representaciones inmersivas en realidad virtual expandirá aún más las fronteras de lo analizable.

#### **12.4. Ética y transparencia en la era de la IA**

La implementación de sistemas de inteligencia artificial en bibliometría plantea desafíos éticos fundamentales que requieren marcos de gobernanza robustos. La opacidad algorítmica característica de muchos modelos contemporáneos genera riesgos significativos de reproducir y amplificar sesgos históricos en la evaluación científica, perpetuando desigualdades estructurales. Esta falta de transparencia socava la legitimidad de los sistemas de evaluación bibliométrica basados en inteligencia artificial, exigiendo aproximaciones metodológicas que prioricen la explicabilidad y auditabilidad de los procesos decisorios para mantener la confianza de la comunidad académica en estas herramientas emergentes.

Garantizar la transparencia metodológica constituye un pilar esencial mediante la documentación exhaustiva de datos de entrenamiento, arquitecturas de modelos y procedimientos de validación empleados. Los metadatos deben incluir información detallada sobre la composición demográfica y geográfica de los conjuntos de datos utilizados, permitiendo identificar brechas representacionales que puedan distorsionar los resultados. Los investigadores tienen la responsabilidad ética de divulgar tanto capacidades como limitaciones conocidas de sus sistemas, evitando extrapolaciones indebidas más allá de las condiciones de desarrollo y validación originales, lo que facilita la evaluación crítica por pares.

La protección de datos personales emerge como consideración crítica cuando los sistemas procesan información sensible de investigadores, incluyendo trayectorias profesionales y redes de colaboración. Los marcos éticos deben asegurar el cumplimiento normativo, implementar principios de minimización en la recolección y establecer protocolos claros de consentimiento informado. Estas salvaguardas resultan particularmente importantes en contextos evaluativos donde los resultados pueden afectar oportunidades profesionales y asignación de recursos, requiriendo balances cuidadosos entre utilidad analítica y protección de la privacidad individual en el ecosistema investigador.

La rendición de cuentas por decisiones algorítmicas representa otro desafío central, exigiendo cadenas claras de responsabilidad cuando sistemas automatizados generan rankings o evalúan mérito científico. Deben establecerse mecanismos de apelación efectivos para casos donde investigadores o instituciones consideren evaluaciones injustas, complementados con comités de supervisión ética que incorporen diversidad disciplinaria y geográfica. Esta estructura de gobernanza distribuida permite identificar riesgos operativos y establecer protocolos para el uso responsable que mantengan la confianza institucional en estos sistemas emergentes de evaluación bibliométrica.

Mitigar proactivamente los sesgos algorítmicos demanda estrategias sistemáticas que incluyan desarrollo de técnicas especializadas, auditorías regulares de equidad y diversificación de equipos de desarrollo. La evaluación continua del impacto diferencial entre grupos demográficos, instituciones y regiones geográficas permite identificar y corregir disparidades emergentes antes de su consolidación como injusticias estructurales. Este enfoque preventivo resulta esencial para construir sistemas bibliométricos que reflejen los valores de equidad e inclusión propios de la empresa científica contemporánea en su diversidad epistemológica y metodológica.

La sostenibilidad ecológica completa el panorama ético, considerando los requerimientos computacionales masivos que generan huellas de carbono significativas durante el entrenamiento y despliegue de modelos. Los desarrolladores deben priorizar eficiencia energética, seleccionar arquitecturas balanceadas y transparentar costos ambientales, alineando la innovación tecnológica con los imperativos de responsabilidad ambiental que caracterizan la ciencia contemporánea. Esta conciencia ecológica representa una dimensión esencial de la ética aplicada al desarrollo de herramientas bibliométricas con inteligencia artificial en el contexto actual de crisis climática global.

La gobernanza participativa involucra múltiples actores-investigadores, gestores, representantes institucionales- en el diseño y supervisión de sistemas, asegurando que diversas perspectivas informen su desarrollo. Establecer procesos deliberativos abiertos sobre valores incorporados, compensaciones entre objetivos y estándares de rendición de cuentas construye consensos sociales alrededor del uso apropiado de estas tecnologías. Esta inclusividad mitiga riesgos de imposición tecnocrática y garantiza que los sistemas reflejen los valores plurales de la comunidad científica global en su conjunto diverso y multicultural.

La formación en ética digital desarrolla usuarios críticos capaces de interpretar resultados algorítmicos mediante programas que enfatizan el juicio humano como complemento esencial de evaluaciones automatizadas. Cultivar escepticismo informado hacia resultados contraintuitivos o potencialmente sesgados previene aplicaciones acríticas, mientras se fortalecen capacidades comunitarias para aprovechar beneficios y mitigar riesgos. Esta alfabetización constituye una inversión necesaria en el capital humano que utilizará estas herramientas, asegurando su implementación responsable en los diversos contextos institucionales y disciplinares donde se aplicarán.

## Recapitulando

- La ciencia abierta promueve transparencia y reproducibilidad.
- Las plataformas OpenAlex y Crossref democratizan el acceso a metadatos.
- La inteligencia artificial facilita el análisis y predicción de tendencias.
- La altmetría amplía la medición del impacto más allá de las citas.
- El uso de paneles interactivos mejora la visualización dinámica.
- La transparencia algorítmica es fundamental en bibliometría basada en IA.
- Big data permite analizar producción científica en tiempo real.
- Los indicadores de equidad evalúan diversidad de género, idioma y geografía.
- Integrar patentes, publicaciones y financiamiento mejora evaluación de innovación.
- Herramientas reproducibles como R y Python facilitan replicación.
- Riesgos: manipulación algorítmica y *gaming* de métricas.
- Manifiestos como DORA y Leiden impulsan evaluación responsable.
- Los repositorios abiertos fortalecen la ciencia colaborativa.
- La interoperabilidad de datos (DOI, ORCID) mejora trazabilidad.
- Las métricas alternativas reflejan impacto social y mediático.
- Las paneles web permiten seguimiento continuo de indicadores.
- Los informes métricos evolucionan hacia visualizaciones narrativas.
- La ética de IA se convierte en prioridad en investigación.
- La bibliometría predictiva combina algoritmos y vigilancia de tendencias.
- La capacitación en herramientas de análisis se vuelve esencial en el futuro.

## Preguntas de autoevaluación

1. ¿Qué cambios aporta la IA al análisis bibliométrico?

2. ¿Qué diferencia existe entre altmetría y citas tradicionales?
3. ¿Qué fuentes abiertas son relevantes para nuevos estudios?
4. ¿Cuáles son los principales riesgos éticos de la IA en evaluación científica?
5. ¿Qué promueve la Declaración DORA?
6. ¿Por qué es importante la interoperabilidad de datos?
7. ¿Qué ventajas ofrecen los paneles interactivos?
8. ¿Qué implica el concepto de evaluación responsable?
9. ¿Qué herramientas son claves para análisis reproducibles?
10. ¿Cómo puede la bibliometría predecir tendencias emergentes?

## **BIBLIOGRAFÍA**

1. Sugimoto CR, Larivière V. Measuring research: what everyone needs to know. Oxford: Oxford University Press; 2018. ISBN 9780190640125
2. Thelwall M. Web indicators for research evaluation: a practical guide. San Rafael (CA): Morgan & Claypool; 2016. <https://doi.org/10.2200/S00733ED1V01Y201602ICR048>
3. Gingras Y. Bibliometrics and research evaluation: uses and abuses. Cambridge (MA): MIT Press; 2016. ISBN 9780262337663
4. Wilsdon J, et al. The metric tide: independent review of the role of metrics in research assessment and management. London: HEFCE; 2015. ISBN 9780992492124



## Chapter 13 / Capítulo 13

*Applied bibliometrics. From data to publication (Spanish Edition)*

ISBN: 978-9915-9680-6-3

DOI: 10.62486/978-9915-9680-6-3.ch13

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



## **ACADEMY, GOVERNMENT, AND INDUSTRY / ACADEMIA, GOBIERNO E INDUSTRIA**

El análisis bibliométrico ha trascendido su aplicación tradicional en el ámbito académico para convertirse en una herramienta estratégica fundamental en los tres sectores clave del sistema de innovación. Esta expansión responde a la creciente necesidad de evidencias cuantitativas para la toma de decisiones en política científica, gestión de investigación y desarrollo y estrategias de innovación empresarial. Cada sector presenta objetivos distintivos, lógicas institucionales particulares y criterios de valoración específicos que condicionan profundamente cómo se utiliza e interpreta la información bibliométrica. Comprender estas diferencias sectoriales resulta esencial para diseñar sistemas de evaluación adecuados.

La creciente interconexión entre estos tres ámbitos mediante colaboraciones público-privadas, proyectos de investigación traslacional y políticas de innovación basadas en evidencia ha generado demandas complejas sobre los sistemas bibliométricos. Estas demandas requieren indicadores capaces de capturar no solo la excelencia académica tradicional sino también el impacto social, la transferencia tecnológica y la creación de valor económico. Este capítulo examina críticamente las particularidades del uso bibliométrico en cada sector, las tensiones que emergen en sus interfaces y las oportunidades para desarrollar enfoques integrados que respeten las misiones distintivas de cada ámbito.

### **13.1. Diferencias en objetivos e indicadores por sector**

La academia tradicionalmente ha privilegiado indicadores de excelencia científica centrados en la producción de conocimiento básico y su validación mediante el sistema de revisión por pares. Los indicadores bibliométricos más utilizados incluyen citas, factores de impacto de revistas e índices de colaboración internacional, reflejando valores centrales como la originalidad, rigor metodológico y contribución al avance disciplinar. Sin embargo, persisten tensiones significativas entre diferentes subsectores académicos, donde humanidades y ciencias sociales muestran patrones de publicación sustancialmente diferentes a ciencias naturales. Estas diferencias internas complican la estandarización de criterios evaluativos incluso dentro del mismo sector académico.

El sector gubernamental emplea la bibliometría principalmente para informar políticas científicas, asignar recursos estratégicamente y evaluar el retorno de inversión en investigación y desarrollo. Los indicadores privilegiados incluyen volumen de producción científica, participación en redes internacionales, especialización temática comparativa y contribución a objetivos nacionales de desarrollo. Aquí predominan las métricas de desempeño agregado a nivel institucional o nacional, con menor énfasis en el impacto individual. La tensión fundamental reside en balancear indicadores de excelencia internacional con relevancia nacional, especialmente en contextos donde la internacionalización científica puede entrar en conflicto con necesidades locales.

La industria utiliza enfoques bibliométricos radicalmente diferentes, orientados principalmente a inteligencia competitiva, vigilancia tecnológica y gestión de propiedad intelectual. Los indicadores típicos incluyen análisis de patentes, mapeo de entornos tecnológicos, identificación de actores clave y detección de oportunidades de innovación. El horizonte temporal es considerablemente más corto que en academia, con énfasis en aplicabilidad comercial inmediata y protección de ventajas competitivas. La tensión característica surge entre la necesidad de secretismo empresarial y las prácticas de ciencia abierta, creando desafíos metodológicos para capturar contribuciones que no se publican en literatura científica tradicional.

Las colaboraciones entre sectores generan necesidades específicas de indicadores híbridos que capturen contribuciones tanto al avance del conocimiento como al desarrollo tecnológico. Estas asociaciones requieren métricas que documenten coautorías academia-industria, citas de literatura científica en patentes, y movilidad de investigadores entre sectores. Sin embargo, persisten desafíos significativos de medición, particularmente para capturar contribuciones no publicadas, conocimiento tácito transferido, e impactos indirectos en cadenas de valor. Desarrollar indicadores significativos para estas interfaces constituye una frontera metodológica crucial para la bibliometría contemporánea.

La evolución de los sistemas de evaluación refleja presiones institucionales divergentes en los tres sectores. Mientras la academia enfrenta críticas por los factores de impacto, el sector gubernamental experimenta demandas de mayor rendición de cuentas y el sector industrial requiere herramientas más ágiles para navegar entornos tecnológicos en rápida evolución. Estas presiones están generando innovaciones en indicadores como métricas alternativas, mediciones de impacto social e indicadores de innovación responsable. El riesgo de adoptar mecánicamente indicadores diseñados para otro sector subraya la importancia de desarrollar marcos evaluativos contextualizados.

La armonización de indicadores entre sectores representa un desafío permanente que requiere balances cuidadosos entre estandarización necesaria para comparabilidad y flexibilidad esencial para relevancia contextual. Iniciativas internacionales buscan establecer lenguajes comunes mientras reconocen diferencias legítimas en objetivos sectoriales. Los avances recientes en el análisis de grandes conjuntos de datos ofrecen oportunidades para desarrollar indicadores más matizados que capturen dimensiones múltiples de valor en diferentes contextos de aplicación. Esta evolución hacia enfoques multidimensionales promete superar limitaciones de sistemas tradicionales mientras respeta la diversidad de misiones que caracterizan los ecosistemas modernos de innovación.

### **13.2. Políticas científicas nacionales basadas en evidencia**

Las políticas científicas nacionales fundamentadas en evidencia bibliométrica representan un avance significativo en la gestión estratégica de los sistemas de investigación. Estas políticas utilizan análisis cuantitativos de la producción científica para identificar fortalezas nacionales, áreas de oportunidad y brechas de conocimiento que requieren atención prioritaria. La bibliometría proporciona a los tomadores de decisiones indicadores objetivos sobre el desempeño del sistema científico, la especialización temática del país y su posición relativa en el contexto internacional. Este enfoque basado en datos permite optimizar la asignación de recursos hacia áreas con mayor potencial de impacto y alinear las inversiones en investigación con las prioridades nacionales de desarrollo.

Los diagnósticos bibliométricos integrales constituyen la base para diseñar políticas científicas efectivas y contextualizadas. Estos análisis examinan la productividad investigadora por regiones, instituciones y disciplinas, identificando clústeres de excelencia y nichos emergentes. La evaluación de redes de colaboración internacional revela alianzas estratégicas y oportunidades para fortalecer la inserción global del país en la ciencia mundial. La detección de tendencias temáticas mediante análisis diacrónico permite anticipar campos de futuro desarrollo e invertir prospectivamente en áreas con potencial de crecimiento. Estos diagnósticos multidimensionales proporcionan una radiografía completa del sistema científico nacional.

La bibliometría informa críticamente los procesos de planificación científica mediante la

identificación de capacidades instaladas y ventajas comparativas. El análisis de especialización temática relativa permite concentrar recursos en áreas donde el país muestra fortalezas distintivas o puede desarrollar ventajas competitivas. La evaluación de impacto científico mediante citas y otros indicadores de influencia intelectual guía la priorización de financiamiento hacia investigaciones de mayor calidad y relevancia internacional. Estos criterios bibliométricos, combinados con consideraciones de relevancia social, permiten diseñar agendas de investigación nacionales balanceadas entre excelencia científica y pertinencia nacional.

El monitoreo y evaluación de políticas científicas implementadas representa otra aplicación crucial de la bibliometría en la gestión pública de la ciencia. El seguimiento bibliométrico continuo permite medir el progreso hacia objetivos establecidos, evaluar el retorno de inversión en programas específicos y realizar ajustes basados en evidencia durante la implementación. Los estudios de impacto posterior a la implementación de políticas revelan cambios en patrones de publicación, colaboración e influencia científica atribuibles a intervenciones específicas. Esta evaluación continua facilita el aprendizaje institucional y la mejora iterativa de las políticas científicas nacionales.

La comparación internacional mediante puntos de interés bibliométricos sitúa el desempeño nacional en contexto global. Estos análisis comparativos identifican países con características similares que han implementado políticas exitosas, proporcionando referentes para el diseño de estrategias nacionales. La evaluación de la competitividad científica internacional mediante indicadores normalizados permite establecer metas realistas y monitorear el progreso relativo del país. Estas comparaciones informadas evitan el aislamiento cognitivo y facilitan el aprendizaje de mejores prácticas internacionales adaptadas al contexto nacional específico.

Los sistemas de información científica nacionales integran múltiples fuentes bibliométricas para proporcionar paneles comprehensivos a tomadores de decisiones. Estas plataformas consolidan datos de producción científica, patentes, colaboraciones e impacto en formatos accesibles y visualmente intuitivos. La interoperabilidad entre diferentes bases de datos nacionales e internacionales permite análisis integrados que capturan dimensiones múltiples del desempeño científico. Estas herramientas de visualización facilitan la interpretación de datos complejos por audiencias no especializadas, democratizando el acceso a información estratégica para la política científica.

Los desafíos metodológicos en la aplicación de bibliometría a políticas científicas incluyen la adecuada contextualización de indicadores internacionales a realidades nacionales específicas. Es crucial adaptar las métricas globales para capturar dimensiones de relevancia local, impacto regional y contribución al desarrollo nacional. La complementación de indicadores bibliométricos con datos cualitativos y estudios de caso enriquece el análisis y previene reduccionismos cuantitativos. Este enfoque mixto produce evaluaciones más balanceadas que informan políticas científicas integrales y socialmente responsables.

La evolución hacia políticas científicas basadas en evidencia bibliométrica representa un avance hacia una gestión más transparente, responsable y efectiva de los sistemas nacionales de investigación. La creciente sofisticación de las herramientas analíticas permite desarrollar políticas más precisas, mejor focalizadas y evaluables mediante indicadores robustos. Este enfoque fundamentado en datos contribuye a legitimar las decisiones de política científica ante la ciudadanía y la comunidad académica, construyendo consensos alrededor de prioridades nacionales basadas en diagnósticos objetivos y análisis rigurosos del panorama científico.

### 13.3. Ranking y evaluación institucional

Como se mencionó en el capítulo 1, los rankings institucionales basados en indicadores bibliométricos se han convertido en herramientas influyentes para la evaluación comparativa de universidades y centros de investigación. Estos sistemas emplean métricas estandarizadas que permiten situar el desempeño de las instituciones en contextos nacionales e internacionales, facilitando la identificación de fortalezas relativas y áreas de mejora. Sin embargo, la proliferación de metodologías diversas y los diferentes énfasis en los criterios de evaluación generan resultados que deben interpretarse con comprensión de sus fundamentos metodológicos. La transparencia en los procedimientos de cálculo y la claridad sobre los objetivos de cada ranking resultan esenciales para su uso adecuado en la gestión institucional.

Los principales sistemas de ranking globales como el *SCImago Institutions Rankings*, el *CWTS Leiden Ranking* y el *QS World University Rankings* emplean aproximaciones metodológicas distintivas que reflejan diferentes concepciones de la excelencia institucional. Mientras algunos privilegian indicadores de productividad e impacto científico, otros incorporan dimensiones de reputación académica, internacionalización o impacto social. Estas diferencias metodológicas explican las variaciones en las posiciones que una misma institución puede ocupar en diferentes clasificaciones. La comprensión profunda de los indicadores específicos que componen cada ranking permite a las instituciones identificar estrategias de mejora alineadas con sus misiones particulares y contextos específicos.

La evaluación institucional basada en bibliometría debe considerar cuidadosamente las diferencias disciplinares para evitar sesgos contra áreas con prácticas de publicación y citación distintivas. Las humanidades, las artes y ciertas ciencias sociales muestran patrones bibliométricos diferentes a las ciencias naturales y las ingenierías, lo que requiere ajustes metodológicos en los sistemas de evaluación. La normalización de indicadores por campo del conocimiento, el uso de percentiles en lugar de valores absolutos y la consideración de productos académicos diversos más allá de los artículos científicos constituyen prácticas necesarias para evaluaciones institucionales equitativas. Estos refinamientos metodológicos permiten comparaciones más justas entre instituciones con perfiles disciplinares diferentes.

El uso responsable de rankings en la gestión institucional requiere evitar la obsesión por las posiciones numéricas y centrarse en el análisis de tendencias y el progreso relativo. El seguimiento diacrónico del desempeño en indicadores específicos proporciona información más valiosa para la planificación estratégica que la mera atención a la posición en clasificaciones. La identificación de instituciones comparables que sirvan como referentes permite establecer metas realistas y aprender de experiencias exitosas. Este enfoque analítico transforma los rankings de instrumentos de competición en herramientas de mejora continua y aprendizaje institucional.

La complementación de indicadores bibliométricos con otras fuentes de información enriquece sustancialmente la evaluación institucional. La integración de datos sobre innovación, transferencia tecnológica, impacto social y satisfacción de estudiantes proporciona una visión más comprehensiva del desempeño institucional. Los sistemas de evaluación multidimensionales que balancean métricas de investigación con indicadores de docencia, vinculación y gestión permiten capturar la complejidad de las misiones universitarias contemporáneas. Esta aproximación integral previene distorsiones en la asignación de recursos y incentivos dentro de las instituciones de educación superior.

La crítica fundamentada a los sistemas de ranking constituye una práctica saludable que

impulsa mejoras metodológicas y usos más responsables. Las limitaciones en la cobertura de bases de datos, los sesgos geográficos y lingüísticos, y la simplificación excesiva de realidades institucionales complejas representan objeciones válidas que deben considerarse en la interpretación de resultados. El desarrollo de rankings alternativos con enfoques innovadores, como los que priorizan la sostenibilidad, la inclusión o la innovación docente, enriquece el ecosistema de evaluación y responde a demandas sociales emergentes. Esta diversificación de aproximaciones evaluativas beneficia a todo el sistema de educación superior.

La evaluación institucional basada en evidencias bibliométricas, cuando se practica con rigor metodológico y conciencia de sus limitaciones, contribuye a la transparencia y mejora de los sistemas de educación superior e investigación. El desafío reside en desarrollar culturas institucionales que aprovechen las perspectivas proporcionadas por estos instrumentos sin caer en reduccionismos cuantitativos o en la obsesión por posicionamientos rankings. El equilibrio entre el uso informado de métricas y la preservación de la diversidad institucional y la autonomía académica representa la clave para una evaluación que realmente sirva al progreso de la ciencia y la educación superior.

### **13.4. Transferencia de conocimiento universidad-empresa**

La transferencia de conocimiento entre universidades y empresas representa un proceso fundamental para convertir el conocimiento científico en innovación y desarrollo económico. Este flujo bidireccional permite que los avances generados en la academia encuentren aplicaciones prácticas en el sector productivo, mientras las necesidades empresariales informan nuevas direcciones de investigación universitaria. La bibliometría ofrece herramientas valiosas para medir, analizar y optimizar estos procesos de transferencia, proporcionando indicadores sobre la intensidad, características e impacto de las colaboraciones entre ambos sectores. El estudio de estas interacciones mediante métricas específicas permite diseñar políticas más efectivas de vinculación tecnológica.

Las publicaciones científicas de múltiple autoría entre investigadores universitarios y profesionales de empresas constituyen un indicador directo de colaboración en investigación y desarrollo. El análisis de estas coautorías revela patrones de especialización temática, intensidad colaborativa y evolución temporal de las relaciones universidad-empresa. La identificación de instituciones y empresas más activas en estas colaboraciones permite reconocer mejores prácticas y modelos exitosos de vinculación. El examen de las redes de coautoría muestra cómo se estructuran estos ecosistemas de colaboración y qué actores funcionan como puentes entre ambos sectores, facilitando el flujo de conocimiento.

Las citas de literatura científica en patentes representan otro indicador bibliométrico crucial para medir la transferencia de conocimiento científico al sector productivo. Este análisis permite rastrear cómo los hallazgos de la investigación básica y aplicada alimentan el desarrollo de nuevas tecnologías y productos. La identificación de publicaciones científicas particularmente influyentes en el ámbito patentario revela áreas donde la investigación universitaria tiene mayor impacto en la innovación empresarial. El estudio de los campos tecnológicos que más intensamente utilizan conocimiento científico reciente ayuda a priorizar áreas de investigación con alto potencial de transferencia tecnológica.

Los indicadores de movilidad investigadora entre universidad y empresa complementan el análisis de las colaboraciones, mostrando cómo el capital humano facilita la transferencia de conocimiento. El seguimiento de trayectorias profesionales que alternan entre ambos sectores revela patrones de circulación de expertos. La identificación de investigadores que

mantienen productividad científica durante etapas en empresas demuestra la compatibilidad entre actividad investigadora y desarrollo tecnológico industrial. Esta movilidad favorece la creación de redes informales y la transferencia de conocimiento tácito, dimensiones cruciales del proceso innovador.

La evaluación del impacto económico y social de la transferencia requiere combinar indicadores bibliométricos con otras fuentes de datos. La creación de *spin-offs* universitarias, los contratos de investigación con empresas y la participación en consorcios público-privados representan resultados tangibles de la transferencia que pueden correlacionarse con indicadores bibliométricos. El análisis integrado de estas métricas permite desarrollar modelos más comprehensivos para evaluar el retorno de la inversión en investigación universitaria y su contribución al ecosistema de innovación regional y nacional.

Los desafíos metodológicos en la medición de la transferencia universidad-empresa incluyen la captura de formas no convencionales de colaboración y la atribución causal entre investigación y resultados de innovación. Muchas interacciones valiosas, como consultorías, formación especializada o uso compartido de equipos, no quedan reflejadas en indicadores bibliométricos tradicionales. La complementación con estudios cualitativos, encuestas a investigadores y análisis de casos específicos enriquece la comprensión de estos procesos complejos. Esta aproximación metodológica mixta produce evaluaciones más balanceadas y útiles para la gestión de la transferencia de conocimiento.

La evolución hacia indicadores más sofisticados de transferencia universidad-empresa representa una frontera importante para la bibliometría aplicada. El desarrollo de métricas que capturen la bidireccionalidad del flujo de conocimiento, la diversidad de mecanismos de transferencia y los impactos a diferentes plazos permitirá una gestión más estratégica de estas colaboraciones. Esta especialización responde a la creciente importancia de la innovación basada en conocimiento para la competitividad empresarial y el desarrollo socioeconómico, posicionando a la bibliometría como herramienta esencial para la política de ciencia, tecnología e innovación.

### **13.5. Bibliometría para la evaluación de impacto social**

La evaluación del impacto social de la investigación mediante herramientas bibliométricas representa un avance significativo hacia sistemas de valoración más comprehensivos. Tradicionalmente, la bibliometría se ha centrado en medir el impacto académico a través de las citas, pero existe una creciente demanda por capturar cómo la investigación influye en la sociedad más allá del ámbito académico. Esta evolución responde a la necesidad de demostrar el valor social de la inversión pública en ciencia y de alinear la actividad investigadora con los desafíos sociales urgentes. El desarrollo de indicadores de impacto social constituye un campo emergente que complementa las métricas tradicionales de excelencia científica.

Las menciones en documentos de políticas públicas representan un indicador valioso para medir la influencia de la investigación en la toma de decisiones gubernamentales. El análisis de cómo las publicaciones científicas son citadas en legislación, informes parlamentarios y documentos de planificación estratégica revela el camino del conocimiento desde la academia hacia la esfera pública. La identificación de investigaciones particularmente influyentes en políticas específicas permite entender qué características facilitan la utilización del conocimiento científico en la gestión pública. Este análisis informa estrategias para hacer la investigación más accesible y relevante para los tomadores de decisiones.



La presencia en medios de comunicación y redes sociales ofrece otra dimensión para evaluar el alcance social de la investigación. Las métricas alternativas capturan cómo los hallazgos científicos son discutidos, compartidos y comentados en plataformas digitales y medios tradicionales. Este análisis revela qué temas de investigación generan mayor interés público y cómo la ciencia se integra en el debate social. La identificación de investigadores e instituciones que comunican efectivamente su trabajo al público general proporciona modelos para mejorar la divulgación científica y el compromiso con la sociedad civil.

Las colaboraciones con organizaciones no gubernamentales y comunidades locales representan otra faceta del impacto social medible a través de indicadores bibliométricos adaptados. La coautoría con representantes de la sociedad civil, la investigación participativa y los proyectos co-diseñados con comunidades muestran formas de investigación socialmente arraigada. El análisis de estas colaboraciones revela modelos de producción de conocimiento más inclusivos y responsables socialmente. La identificación de prácticas exitosas de investigación-acción participativa informa estrategias para construir puentes más sólidos entre la academia y la sociedad.

La contribución a los Objetivos de Desarrollo Sostenible constituye un marco emergente para evaluar el impacto social de la investigación. El mapeo de publicaciones científicas con metas específicas de desarrollo sostenible permite visualizar cómo la investigación aborda desafíos globales como el cambio climático, la reducción de desigualdades o la salud pública. Este análisis identifica brechas temáticas donde se necesita más investigación orientada a problemas sociales urgentes. La evaluación de la contribución investigadora a agendas globales de desarrollo proporciona una perspectiva contextualizada del valor social de la ciencia.

Los desafíos metodológicos en la medición del impacto social incluyen la atribución causal, la diversidad temporal de los impactos y la consideración de contextos locales. Los efectos sociales de la investigación pueden manifestarse años después de su publicación y a través de caminos indirectos difíciles de rastrear. La complementación de indicadores cuantitativos con estudios de caso, narrativas de impacto y metodologías participativas enriquece la evaluación. Esta aproximación mixta captura mejor la complejidad del impacto social y evita reduccionismos en su medición.

La evolución hacia sistemas de evaluación que valoren equilibradamente el impacto académico y social representa una transformación necesaria en la cultura investigadora. El desarrollo de marcos multidimensionales que reconozcan diferentes tipos de contribuciones al avance del conocimiento y al bienestar social promueve una ciencia más diversa y relevante. Esta evolución responde a demandas sociales por una investigación más responsable y alineada con los valores y necesidades de las comunidades a las que sirve, posicionando a la bibliometría como herramienta para una ciencia más consciente de su papel social.

### **13.6. Gestión de propiedad intelectual basada en métricas**

La gestión estratégica de la propiedad intelectual ha incorporado progresivamente herramientas bibliométricas para optimizar decisiones sobre protección, comercialización y desarrollo tecnológico. Estas métricas permiten evaluar el potencial de innovación, identificar oportunidades de patentamiento y priorizar inversiones en investigación con mayor probabilidad de generar derechos de propiedad intelectual valiosos. El análisis bibliométrico aplicado a la propiedad intelectual trasciende el enfoque tradicional centrado en contar patentes, desarrollando indicadores más sofisticados que capturan dimensiones de calidad, impacto y posicionamiento competitivo. Esta aproximación cuantitativa complementa la experticia legal



tradicional en gestión de tecnología.

El análisis de citas entre patentes y publicaciones científicas constituye una métrica fundamental para medir la transferencia de conocimiento entre ciencia y tecnología. Las citas de literatura científica en documentos patentarios indican cómo la investigación básica alimenta el desarrollo tecnológico, mientras las citas de patentes en artículos científicos revelan flujos inversos donde la tecnología influye en la investigación. La identificación de publicaciones particularmente citadas en patentes señala investigación con alto potencial de aplicación tecnológica. El mapeo de estas redes de citación cruzada permite identificar campos donde la ciencia y la tecnología interactúan más intensamente, informando estrategias de investigación orientada a la innovación.

Los indicadores de valor patentario basados en métricas bibliométricas ofrecen perspectivas sobre el potencial comercial y tecnológico de las invenciones. La amplitud geográfica de la protección, medida a través del número de países donde se solicita una patente, indica la percepción de valor comercial global. Las citas recibidas de patentes posteriores reflejan influencia tecnológica e importancia para el desarrollo posterior del campo. La densidad de reivindicaciones proporciona indicadores adicionales de valor. Estos indicadores combinados permiten priorizar portafolios de propiedad intelectual y tomar decisiones informadas sobre mantenimiento, licenciamiento o abandono de derechos.

El análisis de paisajes tecnológicos mediante técnicas bibliométricas permite mapear competitividad y oportunidades de innovación. La agregación y análisis de grandes conjuntos de datos patentarios revela tendencias tecnológicas, actores clave y espacios blancos en el desarrollo tecnológico. La identificación de clústeres tecnológicos emergentes y tecnologías convergentes informa estrategias de posicionamiento competitivo. El análisis de la diversidad tecnológica de actores específicos permite identificar oportunidades de colaboración o adquisición. Estas herramientas apoyan decisiones estratégicas de investigación y desarrollo al proporcionar inteligencia competitiva basada en datos.

La evaluación del desempeño innovador de instituciones mediante métricas de propiedad intelectual requiere ajustes metodológicos específicos. La normalización por tamaño institucional, especialización temática e intensidad de investigación permite comparaciones más justas. La integración de indicadores patentarios con métricas de publicación científica proporciona una visión más completa del desempeño en innovación. El análisis de la eficiencia en la transferencia, medido a través de la relación entre inversión en investigación y generación de propiedad intelectual, informa sobre la productividad del sistema de innovación institucional. Estas evaluaciones integrales apoyan la gestión estratégica de la investigación e innovación.

Los desafíos metodológicos en la aplicación de bibliometría a la propiedad intelectual incluyen diferencias en prácticas patentarias entre sectores y países, y la variabilidad temporal en estrategias de protección. Las tecnologías farmacéuticas muestran patrones diferentes a las tecnologías de la información, mientras las estrategias patentarias varían significativamente entre empresas y universidades. La complementación con análisis cualitativos, estudios de caso y experticia sectorial específico enriquece la interpretación de métricas. Esta aproximación contextualizada evita conclusiones erróneas basadas en comparaciones directas entre realidades institucionales o sectoriales diferentes.

La integración de sistemas de información sobre propiedad intelectual con plataformas bibliométricas representa la frontera actual en gestión tecnológica basada en datos. El

desarrollo de paneles que combinen métricas de publicación, patentes, colaboración e impacto proporciona herramientas poderosas para la toma de decisiones. La aplicación de técnicas de inteligencia artificial para el análisis predictivo del potencial de innovación permite identificar oportunidades antes de que sean evidentes. Esta evolución hacia sistemas integrales de gestión del conocimiento posiciona a la bibliometría como componente esencial en los sistemas modernos de gestión de innovación tecnológica.

Has alcanzado ahora el dominio fundamental de la bibliometría, dispones de las herramientas metodológicas, las competencias analíticas y la perspectiva crítica necesarias para emprender investigaciones bibliométricas rigurosas y significativas. Este conocimiento te posiciona en el umbral de un vasto territorio por explorar, donde cada estudio representará tanto una aplicación práctica de lo aprendido como una oportunidad para contribuir al avance de esta disciplina en constante evolución. El verdadero viaje comienza en este momento: lleva estos cimientos hacia nuevas preguntas de investigación, contextos innovadores y contribuciones originales que expandan las fronteras del análisis cuantitativo de la ciencia, siempre equilibrando el rigor métrico con la profundidad interpretativa y la responsabilidad ética que caracterizan a los bibliometristas destacados.

## Recapitulando

- El modelo Triple Hélice describe la interacción universidad-industria-gobierno.
- Las co-publicaciones son indicadores de colaboración científica.
- Los sistemas de innovación nacional miden la eficiencia tecnológica.
- La transferencia tecnológica requiere oficinas de enlace universitarias.
- Las políticas públicas orientan prioridades de investigación.
- Las patentes reflejan aplicación práctica del conocimiento.
- Las universidades emprendedoras promueven *spin-offs* y startups.
- Los indicadores combinados (publicaciones + patentes) muestran innovación.
- La gobernanza ética evita conflictos de interés.
- La transparencia en propiedad intelectual garantiza confianza.
- Los incentivos fiscales fomentan colaboración público-privada.
- La evaluación equilibrada valora investigación básica y aplicada.
- Los parques tecnológicos impulsan transferencia del conocimiento.
- Las redes internacionales fortalecen la innovación global.
- Los acuerdos universidad-industria requieren regulación clara.
- La inversión en I+D correlaciona con competitividad nacional.
- El impacto socioeconómico es un criterio emergente en evaluación.
- Los sistemas de medición deben incorporar valor social.
- Las alianzas equitativas con países en desarrollo reducen brechas.
- La Triple Hélice evoluciona hacia modelos cuádruple hélice con sociedad civil.

## Preguntas de autoevaluación

1. ¿Qué describe el modelo Triple Helix?
2. ¿Qué indicadores combinados puedes usar para medir transferencia universidad-industria?
3. ¿Qué riesgos éticos existen en la colaboración academia-industria?
4. ¿Qué instrumentos de política pública fomentan la innovación (subsidios, créditos, incentivos fiscales)?
5. ¿Cómo puedes medir el impacto socio-económico de la investigación científica?
6. ¿Qué rol juegan las *spin-offs* y parques tecnológicos en la transferencia de conocimiento?

7. ¿Por qué es importante distinguir entre investigación fundamental y aplicada en sistemas de evaluación?
8. ¿Qué factores definen una universidad emprendedora?
9. ¿Qué criterios permiten evaluar colaboraciones internacionales equitativas con países en desarrollo?
10. ¿Qué métricas capturan la triple hélice (publicaciones + patentes + contratos)?

## **BIBLIOGRAFÍA**

1. Etzkowitz H. The triple helix: university-industry-government innovation and entrepreneurship. London: Routledge; 2008. ISBN 9780415432305
2. Clark BR. Creating entrepreneurial universities: organizational pathways of transformation. Oxford: Pergamon/Elsevier; 1998. ISBN 9780080436270
3. Nelson RR. National innovation systems: a comparative analysis. Oxford: Oxford University Press; 1993. ISBN 9780195076179
4. Simon D, Kuhlmann S, Stamm J, Canzler W, editors. Handbook on science and public policy. Cheltenham: Edward Elgar; 2019. ISBN 9781788973039

### I. GLOSARIO TÉCNICO

#### A

**Altmetría:** Métricas alternativas que miden el impacto de la investigación a través las citas tradicionales, incluyendo menciones en redes sociales, políticas y medios.

**Análisis de cocitación:** Técnica que identifica relaciones entre documentos basado en la frecuencia con que son citados juntos.

**Análisis de coautoría:** Estudio de patrones de colaboración entre investigadores a través de publicaciones conjuntas.

**Autocita:** Cita que un autor realiza a sus propios trabajos anteriores.

#### B

**Bibliometría:** Disciplina que aplica métodos estadísticos y matemáticos para analizar la producción y diseminación del conocimiento científico.

**Burst de citación:** Período en el que un documento experimenta un aumento significativo en su frecuencia de citación.

#### C

**Ciencia abierta:** Movimiento que promueve el acceso libre a publicaciones, datos y métodos de investigación.

**Cienciometría:** Estudio cuantitativo de la ciencia como sistema social y económico.

**Cita:** Referencia a un trabajo previo en una publicación científica.

**CiteScore:** Métrica de impacto de revistas calculada por Scopus basado en citas recibidas por documentos publicados.

**Clúster temático:** Grupo de conceptos, autores o documentos altamente conectados en un mapa bibliométrico.

#### D

**Desambiguación:** Proceso de distinguir entre entidades con nombres similares (autores, instituciones).

**Diagrama de Sankey:** Visualización que muestra flujos entre diferentes grupos o categorías.

#### E

**Ethical scraping:** Extracción de datos de fuentes web respetando términos de servicio y limitaciones técnicas.

**Excelencia científica:** Porcentaje de publicaciones de una institución que se encuentran entre el 10% más citado a nivel mundial.

#### F

**Factor de impacto:** Métrica que mide la frecuencia promedio con que los artículos de una revista son citados en un período específico.

**FWCI (Field-Weighted Citation Impact):** Indicador que compara las citas recibidas por un conjunto de documentos con el promedio mundial en sus campos.

#### G

**Google Scholar:** Motor de búsqueda gratuito que indexa literatura académica de múltiples fuentes.

## H

**H-index:** Índice que combina productividad e impacto, donde un investigador tiene índice h si h de sus artículos tienen al menos h citas cada uno.

## I

**Índice g:** Variante del índice h que da más peso a publicaciones muy citadas.

**Índice m:** Índice h normalizado por años de carrera investigadora.

**Interdisciplinariedad:** Grado en que una investigación integra métodos o conceptos de múltiples disciplinas.

## J

**Journal Impact Factor:** Véase Factor de impacto.

## L

**Ley de Bradford:** Ley bibliométrica que describe la distribución desigual de literatura relevante a través de revistas científicas.

**Ley de Lotka:** Ley que describe la distribución de productividad entre autores científicos.

## M

**Mapa de calor:** Visualización que usa colores para representar densidades o intensidades de fenómenos bibliométricos.

**Mapa científico:** Representación visual de la estructura intelectual o social de un campo de investigación.

**MeSH (Medical Subject Headings):** Vocabulario controlado utilizado para indexar artículos en PubMed.

**Métricas responsables:** Uso ético y contextualizado de indicadores bibliométricos.

## N

**Normalización:** Ajuste de indicadores bibliométricos para permitir comparaciones entre diferentes campos o períodos.

## P

**PubMed:** Base de datos bibliográfica gratuita desarrollada por la National Library of Medicine.

**Pybliometrics:** Librería de Python para acceder y analizar datos de Scopus.

**Palabras clave:** Términos que describen el contenido esencial de un documento.

## R

**Red de colaboración:** Representación gráfica de las relaciones cooperativas entre investigadores o instituciones.

**Reproducibilidad:** Capacidad de replicar un análisis bibliométrico utilizando los mismos datos y métodos.

## S

**Scopus:** Base de datos bibliográfica comercial mantenida por Elsevier.

**SCImago Journal Rank (SJR):** Indicador de prestigio de revistas que pondera las citas por prestigio de la fuente citante.

**SciELO:** Biblioteca electrónica que indexa revistas científicas de América Latina y otros países.

## T

**Tesaurus:** Vocabulario controlado y estructurado utilizado para la indización y recuperación de información.

## V

**VOSviewer:** Software para construir y visualizar redes bibliométricas.

**Visualización:** Representación gráfica de datos bibliométricos para facilitar su interpretación.

## W

**Web of Science:** Base de datos bibliográfica comercial sostenida por Clarivate Analytics.

## II. TÉRMINOS ADICIONALES

**Academic footprint:** Huella académica que representa el impacto y visibilidad de un investigador.

**Bibliographic coupling:** Técnica que relaciona documentos basado en sus referencias comunes.

**Citation analysis:** Análisis de patrones de citación para evaluar impacto e influencia.

**Co-word analysis:** Análisis de co-ocurrencia de términos para identificar estructuras conceptuales.

**Data mining:** Extracción de patrones y conocimiento desde grandes conjuntos de datos bibliográficos.

**Domain analysis:** Estudio de la estructura y dinámica de un campo científico específico.

**Ethical bibliometrics:** Aplicación de principios éticos en la práctica bibliométrica.

**Gender gap:** Brecha de género en productividad, impacto o reconocimiento científico.

**Knowledge diffusion:** Proceso de diseminación del conocimiento científico through publicaciones y citas.

**Literature mapping:** Técnicas para cartografiar el territorio intelectual de un campo de estudio.

**Metric invariants:** Propiedades bibliométricas que se mantienen constantes a través de diferentes contextos.

**Open citations:** Movimiento para hacer accesibles los datos de citación.

**Predatory publishing:** Prácticas editoriales que priorizan ganancias sobre la calidad académica.

**Quantum metrics:** Nueva generación de métricas basadas en análisis complejos de datos.

**Research assessment:** Evaluación de la calidad e impacto de la investigación usando indicadores bibliométricos.

**Science governance:** Uso de métricas para informar políticas y gestión científica.

**Triangulation:** Uso de múltiples métodos o fuentes para validar hallazgos bibliométricos.

## III. RECURSOS Y ENLACES DE INTERÉS

### Normas y Estándares

#### *Declaraciones y Marcos Éticos*

- Declaración DORA (San Francisco Declaration on Research Assessment): <https://sf.dora.org/>
- Manifiesto de Leiden sobre Indicadores de Investigación: <https://www.leidenmanifesto.org/>
- Principios de Hong Kong sobre Evaluación de Investigadores: <http://www.hkprinciples.org/>

#### *Estándares Bibliométricos*

- ISO 31-19: Indicadores de Ciencia y Tecnología: <https://www.iso.org>
- NISO Alternative Assessment Metrics Initiative: <https://www.niso.org/standards-committees/altmetrics>
- COUNTER Code of Practice para Métricas: <https://www.projectcounter.org/>

### Documentación Oficial de Herramientas

#### *Software de Análisis Bibliométrico*

- Bibliometrix (R) - Documentación Oficial: <https://www.bibliometrix.org/>
- VOSviewer - Manual y Tutoriales: <https://www.VOSviewer.com/documentation>

- CiteSpace - Guía del Usuario: <http://cluster.cis.drexel.edu/~cchen/citespace/>
- CitNetExplorer - Tutoriales: <https://www.citnetexplorer.nl/>

### *Plataformas de Programación*

- R para Bibliometría - CRAN Task View: <https://cran.r-project.org/web/views/>
- Python Bibliometric Libraries - PyPI: <https://pypi.org/>
- Google Colab - Guías: <https://colab.research.google.com/>

## **Descarga de Aplicaciones**

### *Herramientas Principales*

- VOSviewer - Descarga: <https://www.VOSviewer.com/download>
- CiteSpace - Descarga: <http://cluster.cis.drexel.edu/~cchen/citespace/download/>
- CitNetExplorer - Descarga: <https://www.citnetexplorer.nl/download.php>
- Publish or Perish - Descarga: <https://harzing.com/resources/publish-or-perish>

### *Entornos de Desarrollo*

- RStudio - Descarga: <https://www.rstudio.com/products/rstudio/download/>
- Anaconda (Python) - Distribución: <https://www.anaconda.com/products/distribution>
- Jupyter Notebook: <https://jupyter.org/install>

## **Bases de Datos y APIs**

### *Acceso a Datos Bibliográficos*

- Scopus API Documentation: <https://dev.elsevier.com/>
- Web of Science API: <https://developer.clarivate.com/apis/wos>
- Dimensions API: <https://docs.dimensions.ai/dsl/>
- CrossRef API: <https://www.crossref.org/documentation/retrieve-metadata/>
- OpenAlex API: <https://docs.openalex.org/>

## **Listas de Comprobación y Guías**

### *Evaluación de Investigación*

- Lista de Comprobación PRISMA para Revisiones Sistemáticas: <http://www.prisma-statement.org/>
- Guía para Métricas Responsables - LERU: <https://www.leru.org/publications>
- Marco de Evaluación de Investigación - UKRI: <https://www.ukri.org/>

### *Publicación Científica*

- Directrices COPE para Ética en Publicaciones: <https://publicationethics.org/>
- Guías para Autores - ICMJE: <http://www.icmje.org/>
- Checklist STROBE para Estudios Observacionales: <https://www.strobe-statement.org/>

## **Repositorios de Datos**

- Zenodo - Repositorio de Datos de Investigación: <https://zenodo.org/>
- Figshare - Plataforma de Datos Abiertos: <https://figshare.com/>
- GitHub - Código y Scripts: <https://github.com/>

## **Comunidades y Foros**

### *Discusión y Soporte*

- ResearchGate - Comunidad Científica: <https://www.researchgate.net/>
- Stack Overflow - Soporte Técnico: <https://stackoverflow.com/>



### *Asociaciones Profesionales*

- ISSI - International Society for Scientometrics and Informetrics: <https://www.issi-society.org/>
- ASIS&T - Association for Information Science and Technology: <https://www.asist.org/>
- LIBER - Asociación Europea de Bibliotecas de Investigación: <https://libereurope.eu/>

### **Herramientas en Línea**

#### *Análisis y Visualización*

- Litmaps - Mapeo de Literatura: <https://www.litmaps.com>
- ResearchRabbit - Descubrimiento de Literatura: <https://www.researchrabbit.ai/>
- Connected Papers - Exploración Visual: <https://www.connectedpapers.com/>

#### *Gestión de Referencias*

- Zotero - Gestor de Referencias: <https://www.zotero.org/>
- Mendeley - Gestor de Referencias: <https://www.mendeley.com/>
- EndNote - Software de Referencias: <https://endnote.com/>

## **ABOUT THE AUTHORS / SOBRE LOS AUTORES**

Annier Jesús Fajardo Quesada

<https://orcid.org/0000-0002-2071-3716>

Universidad de Ciencias Médicas de Granma. Granma, Cuba.

annierfq01@gmail.com

Eduardo Antonio Hernández González

<https://orcid.org/0000-0001-7325-6099>

Universidad de Ciencias Médicas de Pinar del Río. Pinar del Río, Cuba.

René Herrero Pacheco

<https://orcid.org/0000-0002-9450-1572>

Universidad de Ciencias Médicas de Granma. Granma, Cuba.

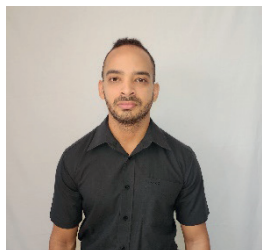
Lisannia Virgen Beritán Yero

<https://orcid.org/0009-0004-6992-9915>

Universidad de Ciencias Médicas de Pinar del Río. Pinar del Río, Cuba.

## ABOUT THE AUTHOR / SOBRE EL AUTOR

---



**Annier Jesús Fajardo Quesada** nació en Jiguaní, provincia de Granma, el 1.º de enero de 2001. Cursó estudios en el Instituto Preuniversitario Vocacional de Ciencias Exactas “Silberto Álvarez Aroche”, donde se destacó en concursos provinciales y nacionales de Programación Competitiva, obteniendo primeros lugares a nivel provincial y medallas de bronce y oro a nivel nacional. Formó parte de la preselección nacional durante dos años consecutivos y participó en olimpiadas regionales y copas nacionales, alcanzando diversas distinciones.

Posteriormente ingresó a la **Universidad de Ciencias Médicas de Granma**, donde, gracias a sus resultados en investigación y publicaciones, fue seleccionado como **director de la Revista Científica Estudiantil Dos de Diciembre**. Durante su trayectoria universitaria participó en múltiples eventos científicos nacionales e internacionales, consolidando una línea de investigación centrada en el **desarrollo de software aplicado a la gestión editorial y la educación médica**.

Se desempeñó como **profesor de Matemática y Bioestadística** en la Facultad de Ciencias Médicas de Bayamo y como **desarrollador web freelance**. En 2025 ingresó a la **Unión de Informáticos de Cuba**.

Se le otorgó el Internado Vertical en **Anestesiología y Reanimación**, donde desarrolla su proyecto de **tesis doctoral**, enfocada en el **diseño de índices predictivos aplicados a la práctica anestésica**.

## **ADDITIONAL INFORMATION / INFORMACIÓN ADICIONAL**

### **Declaración de responsabilidad**

La responsabilidad ética y legal del contenido de esta obra recae exclusivamente en sus autores, quienes garantizan el cumplimiento de la normativa vigente en materia de propiedad intelectual y derechos de autor. La editorial no se hace responsable por las opiniones, resultados o interpretaciones expresadas, ni por el uso que terceros hagan de este material

### **Declaración de Conflicto de Intereses**

Los autores declaran no tener conflicto de intereses de carácter personal, comercial, institucional ni financiero con relación a esta obra.

### **Financiación**

Esta obra no ha recibido financiación específica de organismos públicos, privados ni sin ánimo de lucro.