

Contra la bibliometría “rápida y sucia”: aspectos para valorar la complejidad en los análisis bibliométricos

A critique of ‘quick and dirty’ bibliometrics: issues for assessing complexity in bibliometric analysis

Rafael Repiso; Álvaro Cabezas-Clavijo



Rafael Repiso

Universidad de Granada, España
<https://orcid.org/0000-0002-2803-7505>
repiso@ugr.es



Álvaro Cabezas-Clavijo

Universidad Internacional de La Rioja,
España
<https://orcid.org/0000-0001-9641-8855>
alvaro.cabezas@unir.net

Cómo citar este artículo

Repiso, R., & Cabezas-Clavijo, Á. (2025). Contra la bibliometría “rápida y sucia”: aspectos para valorar la complejidad en los análisis bibliométricos. *Revista Panamericana de Comunicación*, 7(1), 3419. <https://doi.org/10.21555/rpc.v7i1.3419>

Recibido: 22 - 04 - 2025

Aceptado: 05 - 05 - 2025

Publicado en línea: 13-05-2025

Resumen

El presente trabajo denuncia el auge indiscriminado de estudios bibliométricos de baja calidad que, aprovechando la accesibilidad de bases de datos como Web of Science o Scopus y el uso de herramientas automatizadas, proliferan en revistas científicas de diversas disciplinas. Esta tendencia, facilitada por la ausencia de conocimientos especializados de los revisores de revista y por el interés en incrementar rápidamente la productividad académica, ha derivado en la publicación de investigaciones metodológicamente pobres, con escasa elaboración conceptual y valor analítico limitado. Frente a este panorama, los autores reivindican la bibliometría como un campo de alta especialización que exige comprensión teórica, competencia metodológica e interpretación crítica de los datos. Con base en estos principios, se propone una aproximación metodológica para valorar la complejidad de los estudios bibliométricos a partir de seis dimensiones: tamaño de la población, origen y fuente de los datos, forma de recolección, grado de normalización, tipos de análisis empleados y herramientas utilizadas. Esta perspectiva metodológica pretende ofrecer a editores e investigadores un marco para identificar investigaciones sustantivas y distinguirlas de aquellas realizadas con poco esfuerzo, criterio y contexto teórico.

Keywords: Bibliometría; Metodología; Calidad; Bases de datos científicas; Análisis de datos; Visualización de datos.

Abstract

This paper denounces the indiscriminate rise of low-quality bibliometric studies. These studies have proliferated across scientific journals in a wide range of disciplines, taking advantage of the accessibility of databases such as Web of Science

and Scopus and the use of automated tools. This trend, driven by a lack of specialised knowledge among journal reviewers and a desire to rapidly increase academic output, has led to the publication of studies with methodological shortcomings, characterised by limited conceptual development and low analytical value. In response to this situation, the authors advocate for bibliometrics as a highly specialised field that requires theoretical understanding, methodological expertise, and critical interpretation of data. The paper proposes a methodological framework for assessing the complexity of bibliometric studies. This framework is based on six dimensions: size of the population, origin and source of data, mode of data collection, degree of data normalisation, types of analysis employed, and analytical and visualisation tools used. This methodological perspective aims to provide editors and researchers with a framework for identifying substantive research and distinguishing it from studies produced with minimal effort, limited judgment, and weak theoretical grounding.

Keywords: Bibliometrics; Methodology; Quality; Scientific databases; Data analysis; Data visualisation.

La plaga de los análisis bibliométricos

Una plaga de trabajos de naturaleza bibliométrica asola las revistas científicas de toda temática y condición. La bibliometría se ha puesto de moda desde hace unos años y no es que se haya incrementado sustancialmente el número de expertos en el tema, sino que se han multiplicado exponencialmente los trabajos realizados por inexpertos en el tema, incluso por encargo. La realidad es que la producción de estudios bibliométricos se ha incrementado significativamente gracias a dos factores claves. El primero es la facilidad actual para recopilar datos, puesto que las grandes bases de datos científicas, Web of Science y Scopus principalmente, identifican y recopilan datos “limpios” en minutos, mientras que otras técnicas requieren días, semanas, meses e incluso años. El otro elemento fundamental es que el análisis de estos datos, al menos a nivel descriptivo básico (conteos de autores, revistas, o instituciones), es ofrecido por las propias bases de datos, mientras que programas como VosViewer o Bibliometrix permiten que cualquier persona sin necesidad de conocimientos específicos en Bibliometría genere indicadores y visualizaciones atractivas para el público. Estamos asistiendo a la recepción y publicación de trabajos bibliométricos “fáciles”, con poco mérito académico, poca elaboración y carentes de contexto o fundamentación académica.

Poco mérito hay en apropiarse de unos datos que ha procesado con facilidad y sin apenas intervención humana una aplicación informática. Dicho de otra forma, una investigación cuya búsqueda y análisis de datos se puede replicar en menos de un día, en muchos casos menos de una hora, no debería tener cabida en una revista científica seria. Este tipo de trabajos solo contribuyen a la proliferación de ciencia de bajo mérito e interés, aumentando el ruido en el campo. ¿Quién le dice que no a aumentar su producción científica de manera fácil y rápida? Y, por el contrario, ¿quién le dice que sí? Pues en muchos casos investigadores ávidos de producción y ajenos totalmente al ámbito cuantitativo, que ven en la bibliometría un recurso fácil para incrementar sus estadísticas de *papers* publicados.

Este grupo de migrantes académicos se caracteriza principalmente porque publican sus estudios bibliométricos en revistas ajenas al área de Documentación, pues suelen ser rechazados masivamente por las revistas especializadas, y por tanto evitan las cabeceras con revisores expertos. Han encontrado en las *megajournals* y en las revistas *open access* “marca blanca” de las grandes editoriales científicas un filón para publicar rápido y sin esfuerzo. Sus publicaciones son habitualmente la repetición de un mismo procedimiento metodológico a diferentes grupos.

No obstante, históricamente el área ha tenido un segundo grupo de “paracaidistas de la bibliometría” (término acuñado por Emilio Delgado-López-Cózar), y son aquellos investigadores

de áreas cuantitativas, principalmente de Ciencias, que realizan aplicaciones a los estudios en sus áreas. Se caracterizan porque su dominio de las técnicas cuantitativas supera con creces al promedio del área, utilizan la cienciometría como método, pero suelen ser exitosos en sus respectivas líneas de investigación que casi nunca tienen nada que ver con el área, publican en las mejores revistas de la disciplina y lo hacen realizando aportaciones significativas. No debemos olvidar que muchos de los padres fundadores del estudio cuantitativo aplicado a la comunicación científica provenían de áreas como la Física o la Química. Pero había una diferencia sustancial con el oportunismo masivo de los últimos años, y es que para estos gigantes la bibliometría era una segunda línea de investigación, una herramienta para profundizar en estudios evaluativos o descriptivos de grandes áreas, para algunos incluso un *hobbie* (¿de qué otra manera podemos calificar la propuesta del índice H por parte del físico Hirsch?), pero principalmente hacían aportaciones significativas a la bibliometría.

La bibliometría, lejos de ser una técnica mecánica y reductible a gráficos automatizados, constituye un campo riguroso y altamente especializado dentro de la investigación científica, que exige comprensión teórica, dominio metodológico, y una interpretación crítica de los datos. Su desarrollo ha estado históricamente vinculado a la evaluación del conocimiento científico y la construcción de políticas de ciencia, tecnología e innovación (Moed, 2005; Leydesdorff, 2001). Los estudios bibliométricos de calidad no se limitan a contar publicaciones, sino que abordan cuestiones complejas como la dinámica de la producción científica, la estructura de las redes de colaboración o la evolución temática de las disciplinas (Waltman y Van-Eck, 2015). Son pura ciencia social, ya que arrojan luz sobre tendencias y comportamientos del colectivo científico. Cuando se aplica con el debido rigor, la bibliometría permite iluminar patrones invisibles del sistema científico y contrastar hipótesis de gran valor estratégico. Por ello, reducirla a un ejercicio descriptivo superficial no solo empobrece su potencial, sino que desvirtúa su rol fundamental en la comprensión crítica de la ciencia.

La propia comunidad científica está poniendo de manifiesto la debilidad de muchos de estos estudios bibliométricos (Cabezas-Clavijo et al., 2023), proponiendo protocolos y marcos de trabajo que permitan la generación de análisis más robustos y de mayor calidad metodológica. Entre ellos, destacan *Guidance List for repOrting Bibliometric AnaLyses* (GLOBAL) (Ng et al., 2023), las directrices BIBLIO para el reporte de análisis bibliométricos en biomedicina (Montazeri et al., 2023), o el marco VALOR para evaluar estudios bibliométricos multi-fuente (Hoang, 2025), entre otros.

Cómo valorar la calidad de un artículo bibliométrico

Como se ha indicado anteriormente, hay tres aspectos fundamentales: la comprensión teórica, el dominio metodológico, y la interpretación crítica de los datos. Quizás la más fácil y objetivable de las tres dimensiones es la metodológica. Podemos decir ambigualmente que la comprensión teórica requiere una cultura científica en el estudio y aplicación de estas técnicas y por supuesto del objeto de estudio, algo que muy pocos expertos estarían en disposición de tener. Y la interpretación de los datos es una deriva de la comprensión teórica a los resultados concretos del estudio explicados principalmente en la discusión y las conclusiones. Es por ello por lo que este trabajo pretende acotar desde la vía metodológica el valor de un artículo bibliométrico, pues es más fácilmente objetivable si se identifican las partes intervinientes y se le da valor a su ejecución, en tiempo, conocimientos necesarios y complejidad.

No se puede decir que la complejidad de un trabajo guarde una relación directa con la importancia de sus resultados, ni una buena metodología garantiza buenos resultados. Además, hay trabajos muy importantes cuya principal aportación es de orden teórico. Pero si el aspecto que queremos evaluar es el metodológico, evaluar su complejidad es fundamental, al menos

para conocer el esfuerzo. ¿Cómo saber el valor de un artículo? Esta es una pregunta difícil de responder pues es muy parecida a la pregunta clave en Economía ¿cuál es el valor de las cosas? Tema de discusión desde Aristóteles que hablaba del valor de uso y valor de cambio de las cosas hasta los teóricos que creen que el valor es subjetivo y existen una miríada de motivos que aportan valor a las cosas. En este trabajo se pretende simplificar el problema desde la perspectiva de aquellos que creen en el valor objetivo de las cosas (David Ricardo, Adam Smith o Karl Marx entre otros), para los cuales el valor de algo proviene directamente del trabajo necesario para producir ese algo (tiempo, materiales, herramientas, esfuerzo...). Por tanto, pretendemos hacer un aporte desde la complejidad del trabajo, pues otras apreciaciones, como su impacto científico, económico, teórico, etc, son en el momento de la evaluación de un artículo, completamente inabordables.

A nuestro parecer, seis son los aspectos que definen la complejidad en una investigación bibliométrica:

- Tamaño de la población.
- Origen y fuentes de datos.
- Forma de obtención de los datos.
- Grado de normalización y corrección de los datos.
- Tipos de análisis.
- Programas de análisis y visualización de datos.

Tamaño de la población

El tamaño del corpus de documentos analizados condiciona directamente el diseño metodológico, la profundidad del análisis y las herramientas empleadas. En los estudios con muestras pequeñas (por ejemplo, menos de 100 documentos, [200 para ScienceDirect]), es habitual adoptar un enfoque cualitativo-cuantitativo, donde el análisis bibliométrico se complementa con una revisión crítica del contenido. De hecho, una de las tipologías que ha hecho fortuna en los últimos tiempos es el análisis bibliométrico acompañado de una revisión sistemática. Normalmente hay que desconfiar de estos análisis pues ni una ni otra técnica suelen aplicarse con el rigor necesario. Igualmente, cabe plantearse hasta qué punto es necesario realizar un estudio bibliométrico de una población tan pequeña. Cuando se cuenta con tan pocos datos se suelen llevar a cabo métodos de análisis manual o semiautomáticos, incluyendo estudios de coautoría, referencias citadas y revisión de la estructura argumental de los textos.

En poblaciones medianas (por ejemplo, entre 100 y 5.000 registros), ya es poco viable realizar análisis cualitativos, pero se amplían las posibilidades metodológicas de los análisis bibliométricos. Así, pueden aplicarse análisis de redes de coautoría, coocurrencia de palabras clave, cocitación y análisis de clústeres temáticos. En estudios de gran escala (más de 5.000 documentos), la complejidad se multiplica exponencialmente. Aquí es fundamental trabajar con técnicas avanzadas de minería de datos, aprendizaje automático, análisis multivariante, o incluso procesamiento de lenguaje natural (NLP) para clasificar, reducir dimensionalidades y detectar patrones significativos. La cantidad de datos también determina el soporte que se va a utilizar para su almacenamiento y tratamiento, en los casos más simples hojas de cálculo y en los más complejos sistemas de bases de datos interconectadas.

Origen y fuentes de datos

Las fuentes de datos no son repositorios de información neutros, sino que cada una representa una información diferente, con diferentes procesos de inclusión, selección, diferentes grados de normalización y control de autoridades. Incluso cada una tiene una complejidad diferente

de acceso y consulta y descarga de los datos, desde las que facilitan la descarga de todos los datos disponibles en campos normalizados en un solo archivo hasta las que sólo permiten la consulta de los datos, pero no su descarga, pasando por productos que permiten descargas parciales de los conjuntos de datos. Descargar 50.000 registros en una base de datos puede ser trabajo de cinco minutos en unos productos o de varios días en otros.

El uso de bases de datos tradicionales como Web of Science o Scopus no supone complejidad alguna en los análisis y es el enfoque más habitual en los estudios bibliométricos, sin embargo, el uso de otras fuentes de datos tanto globales (Dimensions, Crossref, OpenAlex) como regionales (Redalyc, Dialnet, Scielo) permite ofrecer nuevas perspectivas sobre la producción científica, y reflejar de manera más adecuada la bibliodiversidad del ecosistema académico.

Forma de obtención de los datos

Como decíamos, existen bases de datos que incluyen entre sus servicios la descarga de datos, algunas sujetas a limitaciones cuantitativas, para su reutilización y otras que no permiten la descarga de los datos, e incluso la complica, por ejemplo, descargar un catálogo de una biblioteca universitaria. En el caso de fuentes como Medline o Dimensions, la extracción de datos puede requerir conocimientos en el uso de APIs o en programación en R o Python, especialmente cuando se trabaja con grandes volúmenes de información. Por otro lado, plataformas de acceso abierto como OpenAlex han democratizado el acceso a datos académicos, pero su aprovechamiento pleno también exige familiaridad con estructuras de datos complejas (JSON) y procesamiento computacional avanzado. Esta diversidad de modos de acceso implica que el investigador bibliométrico debe no solo diseñar bien su estrategia de búsqueda (aspecto basal que, mal ejecutado, puede arruinar cualquier análisis bibliométrico), sino también poseer o coordinar competencias técnicas para la recolección eficiente y fiable de datos, por tanto, en el uso de las fuentes y formas de obtención de datos se puede identificar la calidad y complejidad de un trabajo.

Grado de normalización y corrección de los datos

El grado de normalización y corrección constituye un aspecto crítico en los estudios bibliométricos, cuya calidad y validez dependen en gran medida de la exactitud y estandarización de los datos. Productos como Web of Science o Scopus presentan una estructura relativamente ordenada, especialmente en la cobertura anglosajona, si bien aún se localizan inconsistencias en nombres de autores, instituciones, palabras clave o referencias citadas. No obstante, esta falta de normalización se da con mayor frecuencia en bases de datos con menos recursos, o cuando se realizan estudios de gran escala o en áreas multidisciplinares. Estas variaciones pueden generar duplicidades, agrupar incorrectamente documentos o desdibujar patrones reales de colaboración e impacto. La tarea de normalización implica procedimientos complejos como la unificación de variantes de nombres (por ejemplo, “Univ Autonoma Madrid” vs. “Universidad Autónoma de Madrid”), la desambiguación de autores, la homogeneización de términos clave y, en muchos casos, la revisión manual de campos que los algoritmos automáticos no resuelven con precisión. Este proceso, aunque invisible para el lector final, es uno de los momentos de mayor carga técnica y epistemológica del análisis bibliométrico, pues una base mal depurada puede conducir a conclusiones erróneas o irrelevantes. Cuando se va a hacer un estudio bibliométrico de una base de datos poco utilizada en bibliometría se debe comprobar la consistencia de los datos pues probablemente haya que invertir en homogeneizar autoridades y cualquier elemento que se pretenda analizar, agregando un trabajo de normalización que puede ser automatizado en gran medida, si se tiene conocimiento para ello. Se trata de un aspecto crítico, cuya complejidad se eleva si se usan varias fuentes para el estudio bibliométrico, dado que cada productor tiene distintos sistemas de estructuración del contenido, y de tratamiento de datos.

Tipos de análisis

El tipo de análisis empleado en los estudios bibliométricos determina no solo la profundidad del trabajo, sino también su capacidad explicativa y de manera derivada su valor científico. Muchos estudios se limitan a aplicar estadística descriptiva, ofreciendo recuentos básicos de publicaciones, citas, autores o palabras clave, lo cual es útil para una primera aproximación (muchas veces ofrecida de manera directa por la propia base de datos), pero insuficiente para interpretar dinámicas complejas. Un segundo nivel, más elaborado, incorpora estadística bivariable, como correlaciones entre variables o comparaciones por áreas, países o periodos, que permiten establecer relaciones simples pero significativas. Por su parte, los estudios de mayor sofisticación recurren a la estadística multivariante, incluyendo análisis factoriales, clústeres o análisis de componentes principales, lo que permite identificar patrones latentes, agrupar temáticamente la producción científica o explorar estructuras complejas del conocimiento. A ello se suma el uso de herramientas sociométricas, propias del análisis de redes sociales, aplicadas al estudio de coautoría, cocitación, copalabras o vínculos institucionales. Estos enfoques, basados en grafos y medidas de tendencias estructural, ofrecen una representación relacional del sistema científico que va mucho más allá de la simple contabilidad.

La bibliometría, entendida como disciplina compleja y rigurosa, debe trascender los recuentos y explorar relaciones profundas entre los actores, discursos y dinámicas de producción científica, y para ello es imprescindible un dominio técnico y metodológico que distinga los estudios sustantivos de los meramente ilustrativos.

Programas de análisis y visualización de datos

La selección del software de análisis no es un detalle menor, sino una decisión metodológica que viene condicionada por los conocimientos que se poseen, y los recursos que se tienen, pero que a su vez determinan los tipos de análisis posibles y el nivel de profundidad y complejidad del estudio. Para realizar las principales funciones se utilizan programas genéricos ofimáticos como hojas de cálculo, programas específicos como el BibExcel o Bibliometrix, los desarrollos de Loet Leydesdorff o herramientas orientadas para generar redes como Sci2 Tool, CiteSpace o VOSViewer. Estas herramientas tienen diferentes características y pueden presentar un uso de extremadamente fácil a extremadamente difícil. Muchas herramientas son tan fáciles de usar que sin conocimientos generan resultados aparentes y vistosos, como es el caso de VOSViewer o facilitan tanto el análisis de los datos que permite realizar análisis sin la toma de decisiones adecuada, sin conocimiento riguroso, facilitando mucho ciertos tipos de análisis, como una herramienta “*plug-and-play*”. El grado más elevado de conocimiento se requiere en la programación propia (R, Phytion o Matlab), para la descarga, análisis de datos y visualización. La selección y uso de las herramientas son un signo inequívoco de las capacidades del investigador, de la complejidad del trabajo

Conclusiones

Tal y como ya señalan algunas revistas científicas, alarmadas por la cantidad de estudios bibliométricos triviales que reciben, la aplicación acrítica de las herramientas bibliométricas está dando lugar a resultados sin valor académico significativo (Hulland, 2024). Un buen investigador en bibliometría se caracteriza por poseer un amplio “maletín de herramientas” que le permite acceder a numerosas fuentes documentales, más allá de las manidas bases de datos de revistas Web of Science y Scopus, así como conocimientos de programas estadísticos y de visualización. Un abanico de posibilidades amplio siempre ofrece mayores posibilidades para investigar. La bibliometría, por tanto, ejercida con rigor, dista mucho de ser una técnica “rápida y sucia” de fácil réplica. Su verdadero valor radica en la complejidad metodológica que

implica desde la selección del corpus hasta la interpretación crítica de los resultados. Su valor como investigación está en aportar datos que no están al alcance de una simple consulta, poner luz, donde ninguna herramienta lo pone, no emular.

En este trabajo de corte metodológico hemos presentado seis factores para determinar la complejidad de un análisis bibliométrico: tamaño de la población; origen y fuentes de datos; forma de obtención de los datos; grado de normalización y corrección de los datos; tipos de análisis; y programas de análisis y visualización de datos. Esta perspectiva no pretende agotar todas las dimensiones del valor científico de una investigación, pero sí ofrecer un marco conceptual de trabajo para identificar trabajos sustantivos frente a aquellos que, amparados en la apariencia visual y la facilidad de ciertas aplicaciones, contribuyen más al ruido que al conocimiento. Nuestro objetivo es que un editor, un investigador ajeno a estas técnicas, sea capaz de aproximarse si no al valor del documento, a la complejidad y aporte metodológico de los trabajos. Reivindicar la bibliometría como un campo especializado es también defender la ética del esfuerzo en un entorno académico que a menudo premia la cantidad por encima de la calidad.

Contribución de los autores

Rafael Repiso: Análisis formal; Conceptualización; Escritura - borrador original; Escritura - revisión y edición.

Álvaro Cabezas-Clavijo: Análisis formal; Conceptualización; Escritura - borrador original; Escritura - revisión y edición.

Bibliografía

- Cabezas-Clavijo, A., Milanés-Guisado, Y., Alba-Ruiz, R., & Delgado-Vázquez, Á. M. (2023). The need to develop tailored tools for improving the quality of thematic bibliometric analyses: Evidence from papers published in Sustainability and Scientometrics. *Journal of Data and Information Science*, 8(4), 10-35. <https://doi.org/10.2478/jdis-2023-0021>
- Hoang, A. D. (2025). Evaluating bibliometrics reviews: A practical guide for peer review and critical reading. *Evaluation Review*, OnlineFirst. <https://doi.org/10.1177/0193841X251336839>
- Hulland, J. (2024). Bibliometric reviews-some guidelines. *Journal of the Academy of Marketing Science*, 52(4), 935-938. <https://doi.org/10.1007/s11747-024-01016-x>
- Leydesdorff, L. (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Universal Publishers.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer.
- Montazeri, A., Mohammadi, S., Hesari, P. M., Ghaemi, M., Riazi, H., & Sheikhi-Mobarakeh, Z. (2023). Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. *Systematic Reviews*, 12(1), 239. <https://doi.org/10.1186/s13643-023-02410-2>
- Ng, J. Y., Haustein, S., Ebrahimzadeh, S., Chen, C., Sabe, M., Solmi, M., & Moher, D. (2023). *Guidance List for repOrting Bibliometric AnaLyses (GLOBAL): A research protocol*. <https://osf.io/nvu6w>
- Waltman, L., & Van-Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872-894. <https://doi.org/10.1016/j.joi.2015.08.001>