

■ METODOLOGÍA PARA LA CONSTRUCCIÓN SISTEMÁTICA DE ANTECEDENTES DE INVESTIGACIÓN Y EL DESARROLLO DEL CAMPO

Stephanie Lifszyc¹

Luciano Levin²

Resumen

La construcción de los antecedentes de investigación y la caracterización de un campo a estudiar constituye uno de los principales desafíos en las etapas iniciales de la realización de una tesis e investigaciones académicas, especialmente en un contexto de crecimiento sostenido de la producción científica. Este trabajo propone una metodología sistemática para la construcción de antecedentes de investigación basada en un enfoque de bibliometría enriquecida. El enfoque integra procesos de organización, análisis y visualización de grandes corpus bibliográficos, y se complementa con estrategias de triangulación metodológica que incorporan técnicas cualitativas y herramientas de inteligencia artificial. A modo ilustrativo, se presenta un caso aplicado al análisis de la producción científica sobre dengue, que evidencia el potencial del método para construir antecedentes más robustos, críticos y contextualizados.

Palabras claves

BIBLIOMETRÍA ENRIQUECIDA, ANTECEDENTES DE INVESTIGACIÓN, METODOLOGÍA DE REVISIÓN DE LA LITERATURA, ACOPLAMIENTO BIBLIOGRÁFICO, ESTADO DEL ARTE.

■ METHODOLOGY FOR THE SYSTEMATIC CONSTRUCTION OF RESEARCH BACKGROUND

Abstract

The construction of research background and the characterization of the field represents one of the main challenges in the early stages of theses and academic research, particularly in a context of sustained growth in scientific production. This paper proposes a systematic methodology for building research background based on an enriched bibliometric approach. The approach integrates processes for organizing, analyzing, and visualizing large bibliographic corpora, and is complemented by methodological triangulation through qualitative methods and artificial intelligence tools. As an illustrative example, the paper presents a case study on scientific production related to dengue, demonstrating the potential of the proposed method to produce a more robust, critical, and context-sensitive research background.

Keywords

ENRICHED BIBLIOMETRICS, RESEARCH BACKGROUND, LITERATURE REVIEW METHODOLOGY, BIBLIOGRAPHIC COUPLING, STATE OF THE ART.

¹ Universidad Nacional de Río Negro. CITECDE. CONICET

² Universidad Nacional de Río Negro. CITECDE. CONICET

1. Introducción

Comenzar una investigación, sea un proyecto, una tesis de grado, maestría o doctorado puede ser un desafío significativo para cualquier estudiante o investigador. Implica adentrarse en un mundo hasta el momento desconocido para él o ella, o sólo parcialmente explorado, con el objetivo de construir nuevos conocimientos. Uno de los primeros pasos en esa travesía es la construcción del estado del arte o los antecedentes de investigación, una etapa fundamental que nos permite conocer lo que ya se ha producido sobre un determinado tema, identificar posibles aportes y familiarizarnos con nociones, conceptos y metodologías clave para formular nuestras propias preguntas de investigación.

La construcción de estos antecedentes de investigación, junto a la elaboración de los objetivos generales y específicos, suele ser una de las etapas más desafiantes para los estudiantes que inician su tesis e investigadores en etapas tempranas de sus carreras. Es muy desafiante porque implica reunir, organizar y comprender volúmenes muy grandes de información. Es desafiante porque implica estar mucho tiempo concentrado en una tarea específica. Implica convertirse en experto en un tema. Es en este punto donde los investigadores en formación y tesisistas invierten la mayor cantidad de tiempo y, con frecuencia, experimentan sentimientos de desmotivación que pueden llevar al abandono.

Esto ocurre, en gran medida, debido a la falta de una estructuración metodológica adecuada que facilite avances significativos en un contexto donde el acceso a la producción científica se incrementa. Asimismo, influye la ausencia o insuficiencia de una guía clara por parte de los equipos de dirección de tesis³.

En este capítulo proponemos una estrategia metodológica innovadora basada en lo que llamamos bibliometría enriquecida -es decir, la articulación de técnicas bibliométricas y herramientas cualitativas- como recurso para construir antecedentes de investigación bien estructurados. Este enfoque busca resolver una de las dificultades centrales para cualquier investigador en formación: comprender su posición dentro del campo en el que intenta insertarse.

Es importante destacar que esta propuesta no debe

interpretarse como una solución definitiva o mágica para la construcción de antecedentes de investigación. Tampoco pretende sustituir otros enfoques cualitativos e interpretativos, que siguen siendo esenciales para acercarse al complejo mundo de la investigación. No obstante, se presenta como un paso, cada vez más necesario, para organizar y gestionar el gran volumen de información disponible en la actualidad, un desafío particularmente complejo en las etapas iniciales de cualquier investigación.

2. ¿Qué son los antecedentes de investigación?

Antes de empezar, conviene establecer algunas definiciones. El universo metodológico se ha vuelto tan amplio y las propuestas conceptuales acerca de lo que significan conceptos como “metodología”, “estrategia metodológica”, “problema de investigación”, “antecedentes”, “contexto”, “estado de arte”, por nombrar sólo algunos, son tan heterogéneas que dan lugar a muchas confusiones. Es necesario evitar la polisemia que deviene de su uso tan diverso.

Es por ello que vamos a ofrecer algunas definiciones sencillas y operativas para poder elaborar un discurso que no presente confusiones importantes:

Problema de investigación: El problema de investigación es aquello que vamos a investigar. Dar cuenta del Problema de Investigación significa responderse, en forma completa y sin ambigüedades, a la simple pregunta; “¿Qué voy a investigar?”. La respuesta debe dar cuenta de los límites espaciales, temporales y conceptuales del objeto, cómo se circunscribe en un marco teórico y como se pretende abordar empíricamente el problema. Esto implica que también están incluidos en la definición del problema de investigación algunos aspectos metodológicos.

Estrategia Metodológica: Es el conjunto de pasos lógicos, organizados sistemática y temporalmente (aunque no necesariamente de forma secuencial), que permiten visualizar en el tiempo, el desarrollo de una investigación. Incluye la planificación de recursos (materiales y simbólicos) y la implementación de técnicas e instrumentos acordes al estudio.

Metodología: Es un conjunto de pasos más o menos

³ Para un análisis sobre este tema, ver (Wainerman & Levin, 2024)

estrictos que me permiten abordar una cuestión específica. Puede referirse a diferentes escalas y objetos, desde el análisis de documentos, la construcción de una base de datos, o su interpretación.

Técnicas de investigación: Son los procedimientos específicos dentro de una metodología. Incluyen métodos de recolección, análisis e interpretación de datos, adaptados a las necesidades del estudio.

Antecedentes de Investigación: En este trabajo consideramos que los antecedentes de investigación son todos aquellos trabajos que han abordado previamente algún aspecto sustantivo de la investigación. Los mismos pueden ser tanto teóricos, como el caso bajo análisis, pero también pueden ser antecedentes metodológicos, es decir trabajos que utilizaron los mismos métodos que yo utilizaré en mi propio trabajo. Los antecedentes deben estar demarcados en función de la especificidad de los elementos anteriores al caso que voy a investigar. Para delimitar qué puede formar parte de los Antecedentes de Investigación o no, formular -y responder- alguna/s de las siguientes preguntas:

- ¿Quién investigó antes que yo, el mismo objeto de investigación que yo quiero investigar?
- ¿Quién investigó antes que yo, el mismo “caso” que yo quiero investigar?
- ¿Quién utilizó antes que yo, la misma metodología que yo quiero aplicar? (puede que sea aplicada a otro objeto o caso)
- ¿Quién utilizó antes que yo, el mismo marco teórico (o partes de él), aplicado a objetos y/o casos similares?

Tal como sostiene Souza (2007), este recorrido permite “contar con un acervo teórico y metodológico organizado y jerarquizado de manera que se puedan determinar las constantes, las inconsistencias, las tendencias y los posibles núcleos problemáticos de interés, tanto científico como social, sobre los cuales se deben encaminar los nuevos proyectos de investigación”.

3. Propuestas metodológicas existentes para la construcción de antecedentes

En la construcción de antecedentes se han desarrollado diversas metodologías, cada una con particularidades, aportes y limitaciones específicas. Entre las más recono-

cidas se encuentra la revisión narrativa, considerada la modalidad tradicional o exploratoria. En este enfoque no se definen criterios explícitos de inclusión o exclusión, y la selección de los artículos se realiza de manera arbitraria y la principal limitación radica en que no garantiza exhaustividad ni objetividad en la recopilación de la información (Aisenberg Ferenhof & Fernandes, 2016; Cordeiro et al., 2007).

Por otro lado, la revisión sistemática se distingue por su rigor metodológico y planificación previa. Este enfoque implica la identificación, selección, análisis y síntesis de evidencia proveniente de múltiples estudios primarios relacionados con una pregunta específica. Para garantizar la validez, es necesaria la participación de al menos dos investigadores que discutan y acuerden qué estudios incluir, siguiendo un protocolo definido. Las estrategias implementadas buscan minimizar sesgos y errores aleatorios, lo que hace de este enfoque una síntesis exhaustiva y confiable de la literatura disponible (Cook et al., 1997; Higgins & Green, 2011).

La revisión integrativa, en cambio, tiene como propósito reunir y sintetizar resultados de investigaciones sobre un tema delimitado de manera sistemática y ordenada. Este método permite profundizar en el conocimiento del tema, identificar vacíos y proponer nuevas perspectivas de análisis, aunque su aplicación puede ser compleja cuando la literatura es heterogénea o dispersa.

Con el objetivo de superar las limitaciones observadas en estos enfoques, Aisenberg (2016) propuso el Systematic Search Flow (SSF), un procedimiento que organiza la revisión en cuatro fases sucesivas: la definición del protocolo, la fase de análisis, la fase de síntesis y la fase de escritura. Esta metodología busca hacer el proceso más replicable y aplicable tanto a revisiones sistemáticas como integrativas, facilitando la identificación de vacíos y la comparación crítica de estudios mediante herramientas como la Matriz del Conocimiento.

Adicionalmente, en el ámbito de la medicina basada en evidencia, Khan, Kleijnen y Kunz (2003) proponen un modelo sintético de cinco pasos para la realización de revisiones sistemáticas: formulación de una pregunta clara, búsqueda sistemática de estudios relevantes, evaluación de la calidad metodológica, síntesis de resultados y, finalmente, interpretación de los hallazgos. Este enfoque subraya la importancia de una búsqueda exhaustiva, sin

restricción de idioma ni formato, garantizando así la cobertura amplia de la literatura disponible.

En el campo de las ciencias sociales, la construcción rigurosa de los antecedentes se considera esencial para evitar la repetición de hallazgos y la circularidad en la producción de conocimiento. Este proceso permite formular nuevas perspectivas, aplicar teorías en contextos inéditos y explorar realidades aún no abordadas (Duque & Cervantes-Cervantes, 2019). Para que una revisión cumpla con estándares de rigor, es fundamental explicitar criterios de inclusión y exclusión, evaluar la calidad de las fuentes, sintetizar los hallazgos de manera imparcial y presentar un balance crítico del conocimiento disponible. No obstante, la abundancia de textos disponibles hace que, en la práctica, resulte difícil abarcar la totalidad de la literatura y mantener criterios claros y estructurados en la selección de fuentes.

4. Aportes de la bibliometría a la construcción de antecedentes

El uso del acoplamiento bibliográfico (*Bibliographic Coupling*) fue introducido por M. M. Kessler en 1963 y ha sido ampliamente utilizado para la clusterización de documentos a partir de la similitud entre sus listas de referencias. Esta técnica constituye una de las bases metodológicas de los mapas de la ciencia desarrollados, entre otros, por Loet Leydesdorff (2009) y Kevin Boyack (2005). Tradicionalmente, el acoplamiento bibliográfico ha sido aplicado principalmente con fines evaluativos, orientados al análisis de la estructura de campos científicos, la medición del impacto académico y la visualización de redes de conocimiento a gran escala.

Si bien el *Bibliographic Coupling* ha sido ampliamente utilizado en estudios de ciencia-métrica y mapeo de la ciencia para identificar estructuras cognitivas, dinámicas disciplinares y relaciones entre comunidades científicas, en este trabajo proponemos recuperar el acoplamiento bibliográfico como una herramienta clave para el análisis de literatura y la construcción del estado del arte, además de su potencial para caracterizar el campo de investigación.

A diferencia de las revisiones narrativas tradicionales, basadas en selecciones realizadas principalmente a criterio del investigador, la bibliometría permite abordar la revisión de literatura de manera sistemática, rigurosa y

reproducibile (Zupic & Čater, 2015). Este enfoque posibilita incorporar una porción significativamente más amplia del corpus teórico producido sobre un tema, así como integrar dimensiones cuantitativas al análisis de la literatura científica. De este modo, permite identificar los trabajos más representativos, reconocer tendencias de investigación y construir mapas del campo de estudio (Agostini & Nosella, 2019).

Estos aportes adquieren especial relevancia en el contexto actual de expansión acelerada de la producción científica. El crecimiento exponencial del volumen, disponibilidad y diversidad de publicaciones académicas genera un escenario de sobreabundancia informativa que dificulta la delimitación de antecedentes relevantes para una investigación específica. En este contexto, emergen al menos cuatro dimensiones problemáticas fundamentales:

- *Cantidad*: ¿Cuántos trabajos hay sobre el tema? ¿Cuán grande es la comunidad que produce conocimiento sobre este problema/objeto?
- *Calidad y relevancia*: ¿Cuáles son los trabajos más relevantes de todos los relevados? ¿Cuáles los más citados? ¿Cuáles los más actuales?
- *Contenido*: ¿Sobre qué núcleos temáticos versan estos trabajos? ¿Qué vínculos hay entre los mismos? ¿Con qué metodologías han investigado ese tema/objeto? ¿Cuáles son los “límites” conceptuales de mi objeto de investigación?
- *Dinámicas temporales* de los trabajos: ¿cuál es el límite temporal que debo considerar para construir los antecedentes? ¿Cuál ha sido la dinámica de producción de conocimiento sobre el tema? ¿Cómo se despliega esa dinámica en diferentes escalas (local, regional, global)?

Responder a estas preguntas de forma rigurosa constituye el primer paso para construir antecedentes sólidos. En este contexto, la bibliometría se configura como una herramienta estratégica para enfrentar estos desafíos. Su aplicación ofrece múltiples ventajas: por un lado, proporciona una visión general única, muy difícil de captar por otros métodos; por otro, permite identificar lagunas de conocimiento en grandes volúmenes de información, o áreas en las que el conocimiento ha sido producido con menor intensidad -lo que es muy útil en términos de la originalidad del trabajo que se está realizando. También permite ver conexiones que no son evidentes por medio de

otras estrategias, posibilitando la ampliación de las bases teóricas y metodológicas, pero también fertiliza el campo para el surgimiento de nuevas ideas. Finalmente, contribuye a posicionarse más rápidamente dentro de un campo de producción de conocimiento, reconociendo patrones, actores y dinámicas (Donthu et al., 2021).

Si bien son vastos los aportes de la bibliometría, hay una tendencia en el último tiempo a realizar estudios bibliométricos básicos, sin explorar la totalidad de posibilidades que esta metodología ofrece. Muchos de los estudios que encontramos se reducen a analizar la cantidad de trabajos producidos en relación a un tema, principales autores, pico de publicaciones, instituciones y países más productivos, e idiomas.

En este trabajo, proponemos el uso de una bibliometría enriquecida, utilizando los estudios mencionados previamente, pero sumándole los aportes de la clusterización y diferentes técnicas cualitativas, así como técnicas de iteración metodológica que permiten potenciar los hallazgos e interpretaciones.

5. Bibliometría enriquecida: descripción del método propuesto

5.1. Introducción

Toda construcción de antecedentes de investigación comienza con una búsqueda orientada, que requiere definir con claridad el objeto de búsqueda. No vamos a entrar aquí en detalles sobre este punto, pero resulta inevitable establecer claramente que, para iniciar la construcción de los antecedentes de investigación, es necesario que el problema de investigación esté bien definido. Esto no implica que el problema no se pueda modificar luego. De hecho, se espera que la elaboración de los antecedentes funcione como un 'punto de revisión' para afinar la formulación del problema a partir del análisis del estado del arte.

Esto contribuye a establecer el nivel de profundidad requerido, de modo que se mantenga la especificidad del enfoque y, al mismo tiempo, se potencie su originalidad y coherencia interna. Estas decisiones, por supuesto, estarán a cargo del investigador, quien ahora podrá fundamentarlas con evidencia de naturaleza bibliográfica.

5.2. Repositorios y bases de datos

Para realizar una revisión bibliográfica rigurosa y siste-

mática, es fundamental conocer y utilizar los principales repositorios y bases de datos académicas (nótese el plural) que permiten acceder a la producción científica actualizada y, si es posible, una combinación de ellos. A continuación, se describen algunas de las más utilizadas en investigaciones científicas y tecnológicas:

Web of Science: Plataforma de Clarivate Analytics que reúne una extensa colección de bases de datos bibliográficas desde el 1900 hasta la actualidad. Permite acceder a publicaciones científicas de todas las disciplinas, con funciones avanzadas para evaluar el rendimiento, la calidad y el impacto de la investigación. Entre sus funciones más destacadas, se puede mencionar la búsqueda avanzada y refinamiento por múltiples criterios (autor, institución, país, revista, palabras clave, año, etc.), índice de citas cruzadas entre artículos (puede verse quién cita a quién), herramientas de análisis bibliométrico, visualización de redes de colaboración entre autores, instituciones y países; y exportación de resultados a diferentes formatos.

SCOPUS: Es una base de datos multidisciplinaria de resúmenes y citas bibliográficas, desarrollada por Elsevier. Contiene más de 85 millones de registros, incluyendo artículos, actas de congresos, libros y capítulos. Entre sus funcionalidades se encuentran: Búsqueda rápida y búsqueda avanzada por campos específicos (autor, afiliación, título de publicación, etc.); Análisis detallado de citas y métricas; Herramientas visuales para analizar tendencias, redes de colaboración y mapas de ciencia; Perfiles de autor con historial de publicaciones, métricas y afiliaciones; y Exportación masiva de resultados para uso en bibliometría y software externo. La mayoría de las funciones requieren un acceso pago.

SCIELO: Scientific Electronic Library Online es una biblioteca electrónica que conforma una red iberoamericana de colecciones de revistas científicas en texto completo y con acceso abierto, libre y gratuito. El modelo recupera la ciencia de los países en desarrollo y otorga la visibilidad que la corriente principal no ofrece a la comunicación científica de la región. Como instrumento de evaluación, el sitio proporciona indicadores estadísticos de uso y factor de impacto de las revistas.

OpenAlex: OpenAlex es un catálogo gratuito y abierto de documentos académicos, autores e instituciones, inspirado en la histórica Biblioteca de Alejandría. Esta integración de múltiples fuentes asegura una cobertura

amplia y diversa que incluye artículos, libros, tesis, patentes y preprints. Permite exportar los datos en diferentes formatos

Repositorios institucionales: Los repositorios institucionales son plataformas digitales desarrolladas por universidades, centros de investigación y organismos gubernamentales, que almacenan y difunden la producción científica generada por sus integrantes. Estas plataformas pueden incluir tesis, informes técnicos, publicaciones académicas, etc. Son una fuente de acceso abierto y permiten recuperar literatura gris para el análisis de la ciencia en contextos específicos.

La selección de las bases de datos y repositorios no debe realizarse de manera arbitraria, sino en función de las características del problema de investigación, la escala del análisis y el tipo de producción científica que se busca recuperar. En este trabajo proponemos considerar, al menos, cuatro criterios principales para orientar dicha selección:

Cobertura temática y disciplinar: algunas bases presentan mayor representatividad en determinadas áreas del conocimiento. Por ejemplo, Web of Science y Scopus poseen una fuerte cobertura de revistas indexadas de circulación internacional, mientras que SciELO y repositorios institucionales permiten recuperar producción regional, literatura gris y trabajos menos visibles en circuitos centrales.

Cobertura geográfica y lingüística: utilización de múltiples fuentes permite reducir sesgos asociados a la concentración de publicaciones en determinados idiomas o regiones. Esto resulta especialmente relevante en investigaciones sobre problemáticas locales o regionales, donde gran parte de la producción puede no encontrarse indexada en bases comerciales internacionales.

Disponibilidad y calidad de metadatos: la elección también depende de la posibilidad de exportar registros completos y estructurados, incluyendo referencias bibliográficas, afiliaciones y palabras clave. Estos elementos son fundamentales para la posterior aplicación de técnicas bibliométricas y procesos de clusterización.

Si bien teóricamente recomendamos la utilización de múltiples bases de datos para ampliar la cobertura documental y reducir sesgos de indexación, en la práctica esto no siempre es posible debido a las limitaciones técnicas, la compatibilidad entre formatos y la calidad desigual de los metadatos disponibles. En este sentido, aconsejamos que

la elección de las fuentes responda a criterios metodológicos explícitos vinculados al objetivo de investigación, la escala del análisis y las posibilidades reales de integración y procesamiento de la información.

5.3. Diseño de la estrategia de búsqueda

Para asegurar una búsqueda completa sobre el tema en cuestión, es necesario contar con un conjunto acotado pero significativo de palabras, o conjuntos de palabras, que caractericen el tema. No existe una regla universal para esto, puesto que depende en gran medida del tema que uno esté considerando. Por ejemplo, hay temas que quedan definidos por palabras muy específicas, como es el caso de “Chagas”, pero hay otros temas cuyas palabras son polisémicas y las búsquedas arrojan resultados muy sucios. En esos casos conviene utilizar estrategias de asociaciones de palabras recurriendo a operadores booleanos y caracteres de truncamiento:

Los símbolos de truncamiento, que tienen por objeto recuperar las variaciones del plural y la ortografía en sitios específicos dentro de las palabras, o en terminaciones polimórficas. Los símbolos son:

- * (de cero a infinitos caracteres)
- ? (un carácter cualquiera)
- \$ (cero o un carácter).

Por ejemplo: Nano* (Nanociencia, Nanotecnología, Nanobots, etc.); Tr?panosoma (Trypanosoma, Tripanosoma); Trypanosoma\$ (Trypanosoma, Trypanosomas)

Los operadores booleanos se utilizan para combinar y/o unir términos o frases:

- AND: recupera ambos términos al mismo tiempo. Por ejemplo: Chagas AND enfermedad.
- OR: recupera cualquiera de los términos. Por ejemplo: Chagas OR “Trypanosoma cruzi”.
- AND NOT: recupera el primer término y excluye los que contienen el segundo. Por ejemplo: Chagas AND NOT vinchuca.
- COMILLAS (“”): Permite buscar frases exactas. Por ejemplo: “Dengue Fever”.

No obstante estas estrategias, la información recabada mediante palabras clave y estrategias de búsqueda basadas en booleanos y símbolos de truncamiento generalmente tienen errores. Existen metodologías de limpieza y agrupamiento para corregir estos desvíos de la información.

Como recomendación general para las búsquedas, es

mejor obtener muchos datos, aunque algunos no sean del todo adecuados, que generar conjuntos con datos faltantes. Lo que sobra se puede eliminar (aunque lleva trabajo), pero no se puede agregar lo que falta.

5.4. Organización de la información: Parsing, limpieza y organización en subconjuntos de metadatos

El parseo o parsing es un procedimiento de limpieza y organización de la información que permite ordenar la información en las unidades que el usuario necesite. Este procedimiento es fundamental para armar conjuntos que nos sean útiles en nuestra investigación. Existen muchas formas de limpiar y organizar la información, desde procedimientos manuales en planillas tipo Excel, hasta métodos más automatizados. En general, se requieren habilidades de programación para una limpieza eficiente.

Una buena estrategia para eliminar los datos que no corresponden a la búsqueda que nos interesa, es realizar procesos de clusterización de la información. Esto nos agrupará los textos según algún criterio de similitud escogido y dejará sin unificar aquellos textos que no tengan ninguna similitud. Hay muchas formas de clusterizar los datos: según palabras clave, autores, instituciones, co-citaciones, entre otras. En este trabajo proponemos utilizar el acoplamiento bibliográfico, metodología que está descripta más adelante. Los textos que hayan sido obtenidos por error, bajo un criterio común, se agruparán juntos y será sencillo descartarlos en forma conjunta.

Una vez realizada una primera estrategia de clusterización, y descartados los textos no clusterizados o los clústers de temas no relevantes, se pueden obtener las palabras clave más importantes de estos textos y realizar una nueva búsqueda, iterando así el proceso y completando la información con nueva información, esta vez, más adecuada. Esta nueva estrategia de búsqueda puede ser utilizada en otras bases de datos y repositorios.

Los clusters se pueden conformar mediante diferentes estrategias, es decir, utilizando como elemento de inclusión en un clúster diferentes datos, por ejemplo:

- **Coautorías:** Permite establecer la producción conjunta de artículos científicos entre los autores y visualizar redes de investigadores que trabajan en colaboración, identificar grupos consolidados, líderes en un área y la intensidad de vínculos académicos.
- **Cocitaciones:** Se construyen a partir de documentos que son citados juntos por un tercero. Si dos papers son mencionados en la misma publicación, se establece un vínculo entre ellos.
- **Palabras Clave:** Se elaboran a partir de la coocurrencia de palabras clave en los artículos. Esto permite identificar clústeres de temas recurrentes, áreas de investigación emergentes y la forma en que se estructuran los tópicos en un campo determinado.
- **Temas:** Son representaciones más amplias que combinan palabras clave, títulos, resúmenes y categorías temáticas. Ayudan a comprender la evolución conceptual y temática de un campo, mostrando cómo se agrupan y relacionan los ejes de investigación.
- **Instituciones:** Permiten analizar la producción y vínculos entre universidades, centros de investigación o agencias. A partir de ellos se pueden observar patrones de cooperación, jerarquías institucionales y distribución de recursos académicos.
- **Elementos heterogéneos:** Integran distintos niveles de análisis (autores, instituciones, palabras clave, temas, países, etc.) para observar relaciones más complejas y multidimensionales en la producción científica.
- **Referencias bibliográficas:** Analizar las referencias compartidas entre artículos es otra estrategia. Los papers que citan las mismas fuentes tienden a compartir enfoques metodológicos, marcos teóricos, objetos de estudio o problemas sociales. Existen varios métodos para analizar la co-referenciación, los dos más utilizados son, Co-citation (Small, 1973) y Bibliographic coupling (Kessler, 1963).

En este texto, nos enfocaremos en esta última metodología debido a la complejidad de información que nos proporciona. Si bien existen otras metodologías ampliamente utilizadas para la identificación de temáticas en grandes corpus documentales, el uso de *Bibliographic Coupling* presenta ventajas específicas para la construcción de antecedentes de investigación y la caracterización de campos científicos. A diferencia de las estrategias centradas exclusivamente en el contenido textual, el acoplamiento bibliográfico permite reconstruir relaciones cognitivas entre artículos a partir de las referencias que comparten, identificando comunidades de producción científica que

suelen compartir marcos teóricos, tradiciones metodológicas, problemas de investigación y/o debates conceptuales. Esto resulta especialmente útil en etapas exploratorias, donde el investigador aún no conoce completamente las categorías emergentes del campo. Asimismo, el enfoque basado en referencias compartidas reduce parcialmente algunos problemas asociados a la polisemia, sinonimia y variabilidad terminológica presentes en los análisis textuales. Otra ventaja relevante es que permite detectar estructuras relacionales del campo científico que no siempre son visibles mediante técnicas exclusivamente semánticas, favoreciendo una comprensión más contextualizada de las dinámicas de producción de conocimiento.

Además, algunos estudios comparativos (Boyack & Klavans, 2010) muestran además que el acoplamiento bibliográfico presenta un desempeño superior frente a otros métodos de análisis basados en citaciones para representar el frente de investigación. En comparación con la cocitación, que depende de que los artículos ya hayan sido citados y tiende a representar áreas más consolidadas, el acoplamiento bibliográfico permite identificar con mayor precisión líneas de investigación recientes y emergentes. Por su parte, la citación directa facilita detectar conexiones rápidas entre publicaciones, aunque suele generar agrupamientos más fragmentados y menos coherentes temáticamente.

5.5. Acomplamiento de referencias bibliográficas (Bibliographic coupling)

El acoplamiento bibliográfico, o *Bibliographic Coupling* (BC), es una técnica desarrollada por Kessler en 1963 que parte de la premisa de que el hecho de que dos artículos que compartan referencias, representa una buena medida de su similitud cognitiva (Grauwin & Jensen, 2011; Kessler, 1963).

Para ello, se crea una red usando como nodos los artículos obtenidos y su “acoplamiento bibliográfico” (BC) como enlaces. La fuerza del acoplamiento bibliográfico se define como el número de referencias compartidas entre dos artículos (Agostini & Nosella, 2019). Dependiendo del sistema que se utilice, es posible establecer un umbral mínimo de referencias compartidas (X) para considerar la existencia de un vínculo, lo que permite evitar conexiones artificiales generadas por referencias excesivamente comunes.

Como resultado, se obtiene una red compuesta por registros conectados por las referencias que comparten, mientras que, aquellos artículos que no comparten al menos X referencias son descartados en esta etapa. Esto también implica que un cierto porcentaje de los artículos no logra integrarse a la red analizada, lo que sugiere que sus referencias no los vinculan directamente con el núcleo temático del dominio de estudio convirtiéndose en una efectiva técnica de delimitación.

A continuación se utiliza un algoritmo (sugerimos utilizar Louvain (Blondel et al., 2008) para maximizar la modularidad e identificar un número de clústeres con más de Y artículos (por ejemplo, $Y=50$), un umbral que se establece en función del tamaño de la población. La modularidad calcula la fuerza de la división de la red en sus clústers y se caracteriza por dos parámetros clave. Por un lado, la densidad de los enlaces dentro de cada clúster y por otro, la densidad de enlaces entre clústers. Maximizar la modularidad garantiza una red fuerte y bien definida (o la mejor de entre las redes posibles).

Estos clústers, definidos por referencias compartidas, pueden estar a su vez vinculados entre sí con diferentes intensidades. Los que se encuentran vinculados más intensamente, representan, en general, los subcampos relevantes para la investigación en curso. Aquellos clústers sueltos, son aquellos destinados muchas veces a su eliminación debido a la falta de relevancia temática.

5.6. Estrategias de interpretación

Los clústers obtenidos a partir del análisis de acoplamiento bibliográfico, si bien agrupan artículos que comparten referencias comunes, no poseen un significado intrínseco sin una posterior interpretación cualitativa. Por ejemplo, un clúster puede contener artículos que se han agrupado por motivos muy diferentes: pueden contener artículos que utilicen una misma metodología o reunir artículos que recurran a los mismos referenciales teóricos y otros más a objetos de investigación similares. Adicionalmente, hay muchas características de los textos que conforman los clústers que es necesario conocer para poder darles sentido.

Para interpretar los clústers, se utilizan técnicas de extracción de datos para obtener conjuntos de información sistematizada sobre cada clúster, así como las medidas estadísticas generales que caracterizan a cada clúster respecto de la población general de textos analizados.

Estos conjuntos de datos, permiten asignar a cada clúster elementos que facilitan su interpretación, como la “antigüedad” promedio de los papers que conforman cada clúster, los países de los autores, las principales palabras clave, las principales palabras que conforman los títulos y los abstracts, las instituciones más frecuentes, entre otras.

En rigor, cada elemento de un paper puede ser extraído y analizado en el conjunto de un clúster para ayudar a interpretarlo. De este modo, se construyen conjuntos de datos asociados, al menos a los siguientes elementos:

- Autores
- Años de publicación
- Citas
- Palabras Clave
- Palabras de los Títulos
- Revistas
- Temas
- Países
- Instituciones
- Referencias

Para cada uno de estos conjuntos de datos se obtienen algunas estadísticas que nos permiten valorar la frecuencia de aparición de un elemento dentro de un clúster y también la especificidad de aparición en un clúster respecto al total de la red.

Al mismo tiempo, valorando frecuencia y significancia, uno puede elaborar un listado de los papers más representativos, más frecuentes y más citados de un clúster, para descargarlos y realizar una lectura más profunda.

De esta forma, la técnica de Bibliographic Coupling nos puede brindar nuevas conclusiones a la información brindada por los estudios bibliométricos anteriores y un panorama más complejo de la producción del campo. A diferencia de estudios bibliométricos que muchas veces se focalizan únicamente en la cantidad de la producción, idiomas en los que se escriben y principales autores, esta metodología permite analizar con mayor complejidad visualizar las principales discusiones que hay al respecto. Esta metodología no sólo permite diferenciar las áreas temáticas predominantes y los cambios en las dinámicas del campo, sino que también abre la puerta a identificar nichos de investigación menos explorados o emergentes, lo cual podría orientar futuras líneas de estudio.

5.7. Dinámicas temporales

Todos los análisis anteriores se pueden iterar temporalmente. Es posible establecer ventanas temporales de corte y realizar los procesos de clusterización en cada una de esas ventanas que se van moviendo a lo largo del período analizado. Es raro encontrar trabajos que realicen estas iteraciones, pues requieren de una capacidad de cálculo y un trabajo analítico más largo y complejo. Sin embargo, es una herramienta muy potente para analizar la forma en que se “mueve” el conocimiento en el tiempo.

5.8. Triangulación de datos: estrategias cualitativas y uso de inteligencia artificial

Es recomendable y necesario complementar los análisis bibliométricos con otras metodologías cualitativas. Esta articulación metodológica permite la triangulación de la información, aportando una mirada más compleja y contextualizada de los datos obtenidos. Esto es especialmente útil cuando los resultados bibliométricos generan dudas, para abrir nuevas hipótesis o para categorizar hallazgos difíciles de determinar. Estas estrategias permiten contrastar y profundizar los resultados cuantitativos desde la perspectiva de actores que forman parte activa del campo investigado.

Entre las estrategias cualitativas más relevantes para complementar estudios bibliométricos destacamos:

- Entrevistas: Las entrevistas permiten ahondar en cuestiones que no comprendemos, que queremos profundizar o como forma de contrastar las conclusiones a las que llegamos a partir del análisis de la información. Las entrevistas con investigadores, editores, evaluadores o referentes institucionales aportan información cualitativa clave sobre las prácticas, decisiones y dinámicas del campo que los indicadores no logran capturar.
- Grupos Focales: Según Powell, Single, y Loyd (1996), el grupo focal es “un grupo de individuos seleccionados y reunidos por el investigador con el propósito de discutir y comentar, desde su experiencia, sobre un tema que es el objeto de estudio de la investigación” (p. 499). Los grupos focales pueden utilizarse para presentar resultados preliminares y propiciar la reflexión colectiva en torno a ellos, recuperando así distintas perspectivas de actores del campo.

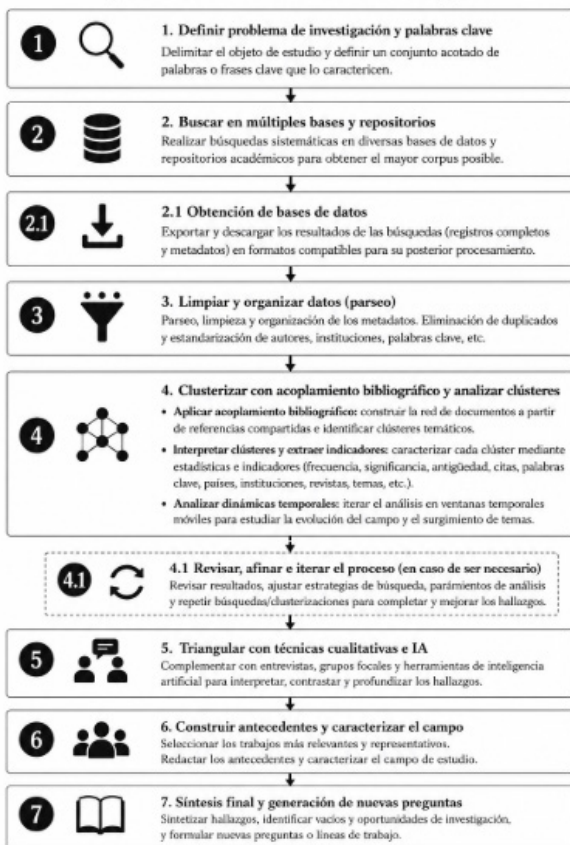
Complementariamente, el uso de herramientas de inteligencia artificial permite profundizar los análisis y contrastar los resultados construidos. Estas tecnologías resultan especialmente valiosas para la interpretación de *clusters*, dado que, por ejemplo, facilitan la identificación de los temas centrales abordados en los textos más representativos.

En síntesis, el uso combinado de bibliometría, estrategias cualitativas y herramientas de inteligencia artificial

permite no solo validar y enriquecer la interpretación de los datos, sino también abrir nuevas dimensiones de análisis, visibilizar tensiones y construir una mirada más compleja y situada sobre los antecedentes de investigación. Esta lógica de triangulación refuerza el enfoque de bibliometría enriquecida que proponemos, integrando lo cuantitativo con lo interpretativo para fortalecer la comprensión del campo.

Figura 1

Figura 1. Esquema de pasos del método de bibliometría enriquecida para la construcción de antecedentes de investigación



Nota. El paso 4.1 se realiza en caso de ser necesario y puede repetirse cuantas veces sea pertinente antes de continuar con los pasos 5–7.

6. Un ejemplo: Producción de conocimiento sobre el Dengue

Aquí proporcionamos un ejemplo de cómo se aplicó el método anteriormente descrito a un caso concreto: la construcción de los antecedentes de investigación sobre “Dengue” a nivel global.

La investigación sobre la que se quiere trabajar, tiene como objetivo analizar la producción de conocimiento sobre esta enfermedad, en el contexto de un país semi-periférico (Argentina), y entender los procesos asociados a la producción y uso de conocimientos. Discutir si el conocimiento que se produce sobre la enfermedad se encuentra alineado con la resolución del problema de salud a escala local, o si se encuentra alineado con intereses internacionales o con ninguno de ellos y se trata de cuestiones de ciencia básica, lejos de las aplicaciones a la salud, por ejemplo.

Por lo tanto, en la búsqueda de información que ayude a comprender esa problemática, resulta sustantivo comprender cómo se produce conocimiento en todo el mundo, cuáles son las temáticas relevantes, cuáles son las más actuales y cuáles se han ido abandonando. También resulta importante conocer cuáles son los países que más producen y en qué temas, las redes de autores y las disciplinas que se encuentran detrás de los conocimientos producidos, entre otras cuestiones.

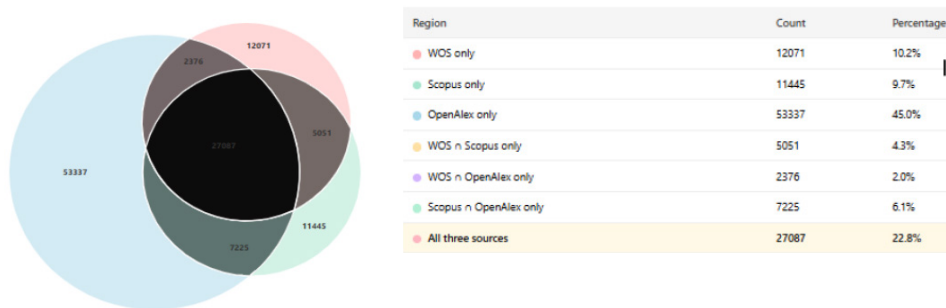
Para ello, se utilizó la bibliometría enriquecida, diseñando una búsqueda lo más amplia posible. Utilizamos la

palabra “Dengue”, que en este caso resulta lo suficientemente descriptiva, específica y amplia para captar la mayor cantidad de producción posible sobre el tema. Para este ejercicio, se buscó en la base de datos SCOPUS, en Título, Palabras clave y Abstracts, obteniendo un total de 47,431 artículos que no fueron filtrados bajo ningún otro criterio.

Si bien la propuesta metodológica presentada recomienda, siempre que sea posible, la combinación de múltiples bases de datos y repositorios, el caso aplicado sobre dengue se realizó exclusivamente a partir de registros recuperados en Scopus. Esta decisión responde a que se trata de una primera versión exploratoria del análisis; en futuras publicaciones se incorporarán otras bases de datos para ampliar la cobertura documental. La aplicación de las técnicas de bibliometría enriquecida no varía en su complejidad (aunque por supuesto los resultados serán diferentes). Somos conscientes de que esta elección presenta limitaciones, particularmente la posible subrepresentación de literatura regional, literatura gris y publicaciones no indexadas en bases comerciales internacionales. En este sentido, el caso presentado debe entenderse como una aplicación ilustrativa de la metodología propuesta y no como una representación exhaustiva de toda la producción científica sobre el dengue.

Proporcionamos no obstante los datos generales de las búsquedas en tres bases de datos, SCOPUS, Web of Science y Openalex, sólo para ilustrar la relación entre los datos obtenidos en cada una.

Figura 2. Distribución de los artículos en las bases de datos. Artículos totales: 118.592.



Volviendo a los datos obtenidos en SCOPUS, en este caso, se procesaron los datos mediante técnicas de parsing y se obtuvieron las frecuencias de los principales metadatos de la muestra completa y luego se procedió a la clusterización mediante técnicas de Bibliographic coupling. Para ello se utilizaron los servicios de Track your Knowledge (TyK), mediante el acceso a una licencia gratuita.

De este modo, primero se presenta una descripción de la muestra en su conjunto (47,431 arts), luego se realiza una descripción de los clusters obtenidos.

6.1. Descripción de la muestra en su conjunto:

- Base de datos: SCOPUS
- Búsqueda: “dengue” en título, resumen y palabra clave.
- Resultados: 47,371 documentos encontrados

A partir de un primer análisis de la frecuencia de los países y las instituciones a la que pertenecen los autores de los papers, podemos obtener la siguiente información:

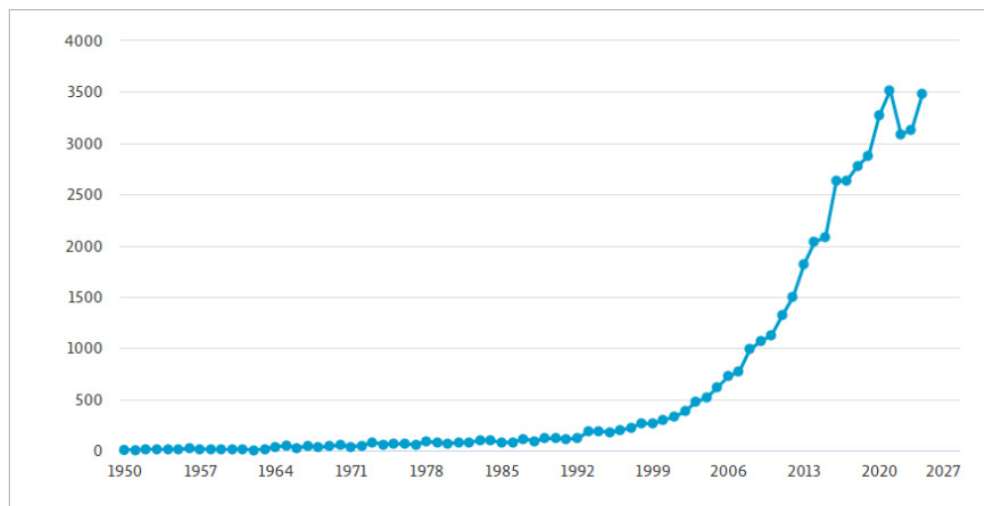
Tabla 1. Los 10 países e instituciones más frecuentes.

País	#	Institución	#
Estados Unidos	11153	Fundación Oswaldo Cruz	1393
India	6301	Mahidol University	1199
Brasil	4606	Universidade de Sao Paulo	888
Reino Unido	3291	University of Oxford	763
China	3005	National University of Singapore	733
Tailandia	2534	CDC	732
Francia	2473	NIH	661
Malasia	2003	CNRS	638
Australia	1994	Institut Pasteur, Paris	631
Indonesia	1793	London School of Hygiene & Tropical Medicine	631

Como se puede observar, Estados Unidos encabeza la lista de países con mayor producción en el tema, sin embargo, a nivel institucional, es una institución brasilera la que la encabeza. Este tipo de información métrica dispara interesantes preguntas acerca de las formas en la que se produce el conocimiento, quién lo produce, quién lo financia y

quién lo utiliza, que resultan insumos de investigación muy valiosos para profundizar los análisis.

En términos de la dinámica temporal de la producción, se puede ver en la figura 1 un crecimiento sostenido a partir de la década del 2000 con un crecimiento exponencial a partir de 2010.

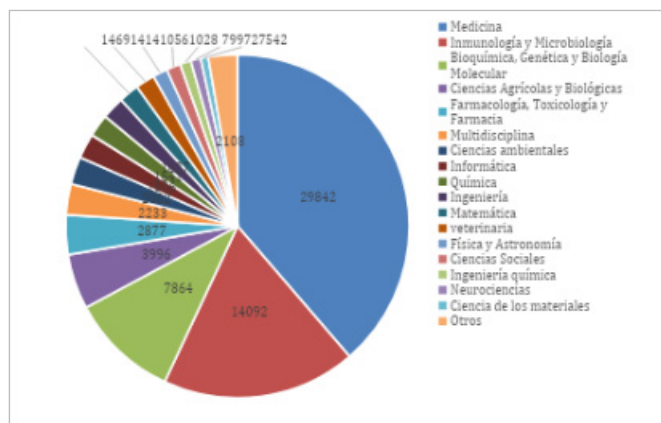
Figura 3. Dinámica temporal de la producción de Dengue en SCOPUS.

6.2. Análisis temático de la producción:

Como mencionamos en la descripción metodológica, existen muchas formas de analizar temáticamente la producción científica. Las bases de datos proporcionan una

primera información, que no es muy confiable, pero no obstante vale la pena revisar.

Así, a partir de la estructuración temática que ofrece SCOPUS, podemos tener una primera aproximación:

Figura 4. Estructuración temática de la producción según SCOPUS.

Por ejemplo, aquí se puede ver fácilmente que, según SCOPUS, casi el 40% de la producción es en el área de la Medicina, seguida muy de lejos (18%) por “Inmunología y Microbiología”. Debido a la naturaleza de la investigación

que se está realizando, resulta muy importante mirar con más detalle la producción de Ciencias Sociales, que si bien es pequeña (2%), son 1028 papers que es importante saber sobre qué temas son.

Figura 5. Dinámica de la producción de Dengue en Ciencias sociales en SCOPUS.

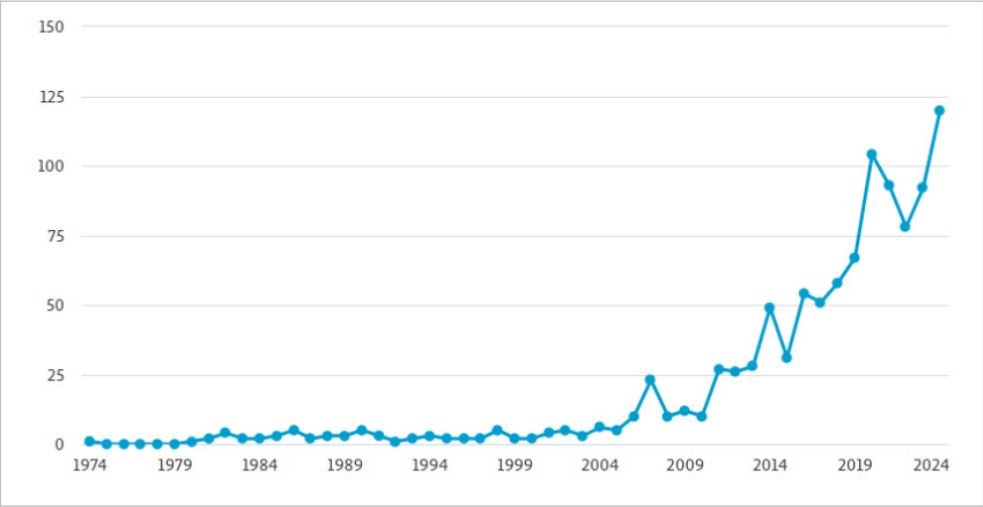


Tabla 2. Distribución temática de la producción sobre Dengue en Ciencias Sociales según SCOPUS.

Área temática	#	Área temática	#
Ciencias Sociales	995	Negocios, Gestión y Contabilidad	41
Medicina	378	Energía	38
Ciencias Ambientales	260	Ciencias Agrícolas y Biológicas	37
Informática	159	Economía, Econometría y Finanzas	37
Ciencias de la Tierra y Planetarias	103	Bioquímica, Genética	22
Ingeniería	98	Farmacología, Toxicología y Farmacia	20
Salud	76	Física y Astronomía	16
Artes y Humanidades	64	Química	13
Ciencias de la decisión	44	Ingeniería química	12
Matemática	44	Psicología	10
Enfermería	42	Otros	18

Lo anterior es simplemente un ejemplo de cómo se puede trabajar con datos métricos sencillos, obtenidos directamente de las bases de datos, para analizar campos y subcampos de forma más detallada. No se proveen aquí los análisis completos de esa información, pues lo que se busca es dejar en claro la estrategia metodológica, pero resulta evidente que la información es mucho más rica que lo expuesto.

Sin embargo, la información anterior no deja de tener los sesgos asociados a la forma en la que cada base estructura los datos.

A continuación, proporcionamos el análisis temático que surge de utilizar la metodología de Bibliographic coupling- de ahora en más BC- , propuesta en este texto.

6.3. Red de Clusters BC

A partir del uso de bibliographic coupling, se identificaron un total de 5 clusters principales y 40 subclusters, organizados a partir de las referencias compartidas entre las publicaciones. Esos clusters fueron etiquetados a priori utilizando la palabra más frecuente de cada clúster según:

Tabla 3. Clústers del caso “Dengue”.

Clúster	# de papers	Palabra Clave más frecuente
1	15,525	Aedes Aegypti
2	8,982	Dengue Virus
3	5,613	Fever
4	4,514	Dengue
5	2,958	Zika Virus

Figura 6. Red de Clústers BC (izq) y sub clusters BC (der).



A continuación se presenta una descripción interpretativa de cada cluster realizada en una combinación de técnicas que incluyeron:

- Revisión de los metadatos de cada clúster
- Revisión de los subclústers de cada cluster
- Lectura completa de los papers más representativos y

- más frecuentemente citados de cada clúster
 - Utilización de modelos GTP para la interpretación del contenido de cada clúster
- No se proporciona aquí una descripción detallada de cada subcluster por cuestiones de espacio.

6.4. Interpretación de clusters

A continuación se presenta una descripción interpretativa de cada uno de los clústers BC que incluye los datos métricos, temporales y el análisis de todos los metadatos y la lectura de los papers seleccionados en función de sus frecuencias y significancias dentro de cada clúster. Se utilizaron también técnicas de interpretación automatizada por modelos de grandes lenguajes que complementan las interpretaciones anteriores.

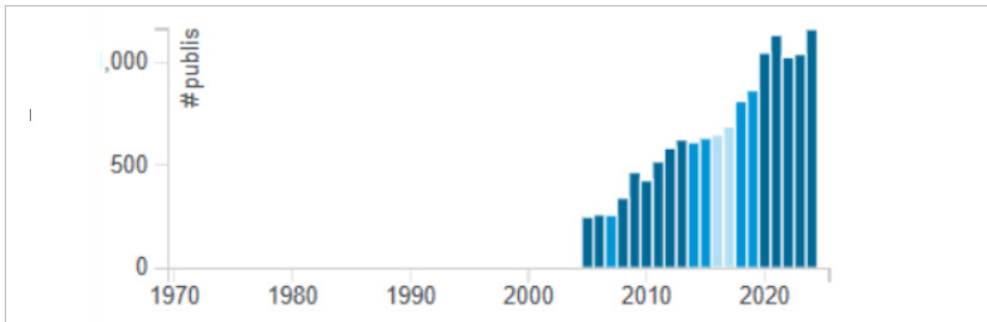
Se obtuvo también la dinámica temporal de los documentos de cada clúster, calculando la frecuencia esperada y la diferencia entre esta y la frecuencia en la población general de papers. Esta información, que en conjunto permiten realizar las interpretaciones

El primer clúster “*Aedes Aegypti*” proporciona un enfoque integral sobre el estudio del mosquito *Aedes aegypti*, destacando su importancia como vector en la propagación de enfermedades como el dengue, chikungunya y fiebre amarilla. Aquí se abordan aspectos fundamentales como su biología y ecología, así como las estrategias

para mitigar su impacto, como la resistencia a insecticidas y el uso de bacterias como *Wolbachia* y biopesticidas (por ejemplo, *Bacillus thuringiensis israelensis*). Además, se menciona el papel de *Aedes albopictus* como vector alternativo, lo que amplía el alcance de la investigación. Las palabras claves más representativas son: animales, *Aedes Aegypti*, *Aedes*, mosquito, larva, vector mosquito e insecticidas. Las disciplinas más representativas son la agricultura y ciencias biológicas, seguidas por las ciencias ambientales y veterinarias. agricultura y ciencias biológicas, seguidas por las ciencias ambientales y veterinarias. Su vínculo más fuerte es con el cluster “Dengue”, aunque es más débil en comparación a los demás clusters.

El clúster de “*Aedes Aegypti*” ha mantenido una producción constante desde 2014, con un incremento reciente en 2021-2023. Entre el 2016 y 2018 hay menos publicaciones de las esperadas. Se podría pensar que este primer puesto se debe al reconocimiento de que la eliminación del vector es hoy en día la forma más efectiva de disminuir la transmisión de la enfermedad.

Figura 7. Dinámica del clúster “AEDES AEGYPTI” (15525).



Se presenta la cantidad de papers por año del clúster AEDES AEGYPTI. Los colores representan la relación entre la frecuencia esperada en la población general y la encontrada en el clúster. Celeste claro indica menos papers que lo esperado, celeste indica la misma frecuencia que lo esperado y azul oscuro indica más papers que lo esperado.

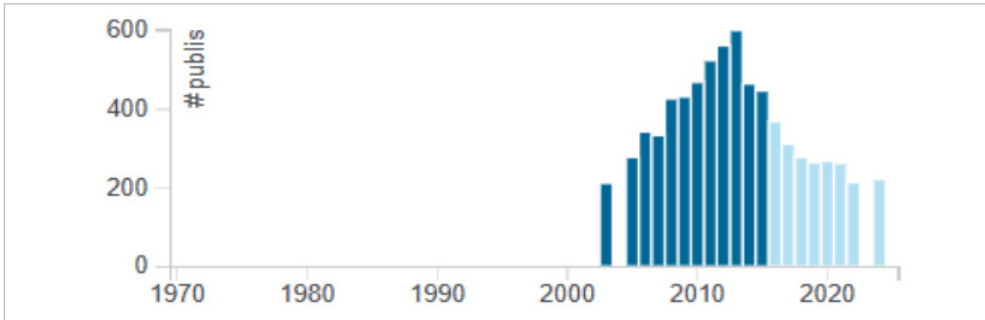
El segundo cluster, “Dengue Virus”, analiza la biología y patogénesis del virus del dengue (DENV), centrándose en aspectos clave como las proteínas de la envoltura viral, la filogenia y evolución del virus, su interacción con células humanas, el diagnóstico molecular mediante RT-PCR y las características clínicas de los serotipos

DENV-1 a DENV-4. Las palabras más representativas son humano, dengue, virus dengue, estudios controlados y anticuerpos. La disciplina más representativa es la inmunología y la microbiología. Si bien comparte una fuerte conexión con el clúster de Dengue, que aborda el virus desde una perspectiva epidemiológica y clínica, este en-

foque se distingue por centrarse exclusivamente en la biología molecular del virus, sin abordar estrategias de control o manejo del vector, lo cual es cubierto en otros clústeres.

Su desarrollo comenzó en 2003, alcanzó su auge en 2013 y desde entonces ha experimentado una desaceleración. Esto coincide con el aumento de relevancia del cluster “dengue”.

Figura 8. Dinámica del clúster “DENGUE VIRUS” (8982).



Se presenta la cantidad de papers por año del clúster DENGUE VIRUS. Los colores representan la relación entre la frecuencia esperada en la población general y la encontrada en el clúster. Celeste claro indica menos papers que lo esperado, celeste indica la misma frecuencia que lo esperado y azul oscuro indica más papers que lo esperado.

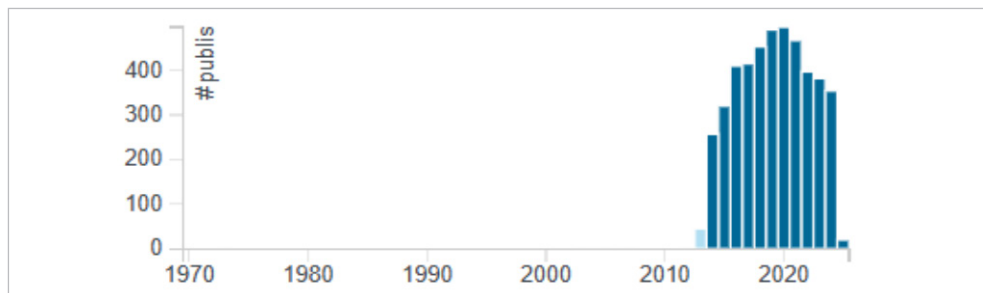
El cluster “Fever”, se centra en los aspectos clínicos y la etiología del dengue en humanos, abarcando estudios clínicos, reportes de casos y grandes investigaciones sobre la enfermedad. Se destacan temas como el impacto del dengue en el embarazo, la diferenciación con enfermedades similares como la leptospirosis, y complicaciones graves como el síndrome hemofagocítico. También se vincula con el estudio de otras enfermedades infecciosas, incluido el coronavirus. Las palabras más representativas son humano, dengue, estudio clínico, reporte de caso y artículo clínico. Su vínculo más fuerte es con el cluster “Dengue”. Las primeras publicaciones de este cluster se identifican en el año 2006. Comienza un crecimiento exponencial con un pico en el año 2024.

El cuarto cluster, “Dengue”, se enfoca en la epidemiología, prevención y control del dengue, abordando temas

como el desarrollo de vacunas, la relación entre serotipos y gravedad de la enfermedad (con DENV-2 más asociado a casos severos) y el uso de Wolbachia para reducir la transmisión. También analiza la incidencia global y la distribución geográfica del dengue. Tiene una conexión fuerte con el clúster de Dengue Virus y, en menor medida, con el de fiebre. A diferencia de otros clústeres, su enfoque principal está en estrategias de prevención y vacunación, más que en la biología del virus. Las palabras más representativas son dengue, humanos, virus del dengue, estudio controlado, virología, dengue virus, e inmunología. Su vínculo más fuerte es con el cluster “Dengue Virus” y, posteriormente, con el de “Fiebre”- observable a partir del grosor de la líneas que lo conectan.

Primera publicación es en el 2013 con pico en 2020 y 2021, con un descenso gradual desde entonces.

Figura 9. Dinámica del clúster “FEVER” (4514).

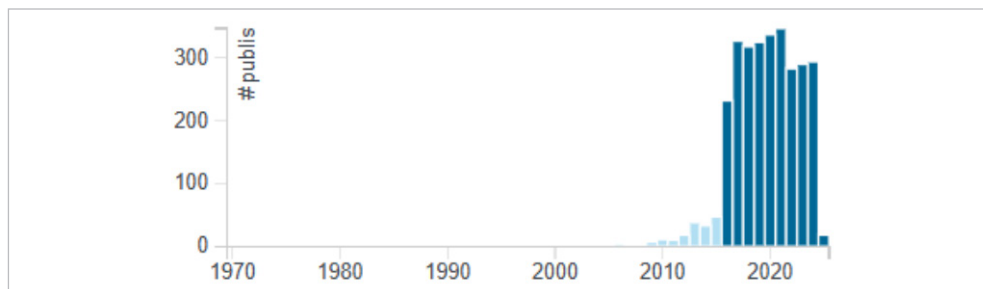


Se presenta la cantidad de papers por año del clúster FEVER. Los colores representan la relación entre la frecuencia esperada en la población general y la encontrada en el clúster. Celeste claro indica menos papers que lo esperado, celeste indica la misma frecuencia que lo esperado y azul oscuro indica más papers que lo esperado.

El clúster “Zika Virus”, el quinto, se centra en el estudio del virus del Zika, su transmisión, impacto en la salud humana y su relación con otras enfermedades como el dengue. Examina los factores ambientales y epidemiológicos que facilitan la propagación del virus, así como sus efectos en el sistema nervioso y su relación con condiciones graves como la microcefalia en recién nacidos y el síndrome de Guillain-Barré. También aborda estrategias de control y prevención, incluyendo la investigación sobre agentes antivirales y la influencia de bacterias como *Wolbachia* en

la reducción de la transmisión del virus. Las palabras más representativas son humanos, fiebre de zika, virus de zika, infección de virus de zika, virología, replicación del virus. Las disciplinas más significativas son la farmacología, toxicología y farmacéutica, seguido por química y neurociencia. Su vínculo más fuerte es con el cluster “Dengue”. Las primeras publicaciones datan de 2009, con un crecimiento exponencial a partir de 2016 y un pico en 2021. Desde entonces, se observa un descenso en la producción.

Figura 10. Dinámica del clúster “ZIKA VIRUS” (2958).



Se presenta la cantidad de papers por año del clúster FEVER. Los colores representan la relación entre la frecuencia esperada en la población general y la encontrada en el clúster. Celeste claro indica menos papers que lo esperado, celeste indica la misma frecuencia que lo esperado y azul oscuro indica más papers que lo esperado.

Como mencionamos en el apartado anterior, la interpretación del cluster no está dada sino que requiere el análisis de los subclusters y el vínculo con otros clusters, palabras

claves de los mismos, categorías temáticas y papers más representativos. Para ejemplificarlo, retomamos el caso del primer clúster.

Figura 11. Principales metadatos del clúster AEDES AEGYPTI.

Keywords	f(%)	σ
ARTICLE	58.74	-23.82
ANIMALS	51.10	48.85
AEDES AEGYPTI	46.67	73.89
ANIMAL	44.99	38.41
AEDES	42.81	54.73
NONHUMAN	40.14	14.07
FEMALE	29.35	-4.73
HUMAN	26.33	-57.91
CONTROLLED STUDY	25.19	-6.12
MOSQUITO	23.36	38.18
HUMANS	21.71	-39.99
DENGUE	21.31	-56.62
LARVA	17.14	40.87
MALE	15.74	-24.24
GENETICS	15.54	6.92
MOSQUITO VECTOR	13.71	31.50
PHYSIOLOGY	13.55	18.14
MOSQUITO VECTORS	12.99	31.05
INSECTICIDE	12.23	36.53

Title Words	f(%)	σ
AEDES	46.07	66.20
AEGYPTI	38.87	58.11
MOSQUITO	19.45	44.24
DENGUE	18.25	-63.55
DIPTERA	10.98	34.57
CULICIDAE	10.87	34.37
AGAINST	10.80	21.82
MOSQUITOES	10.24	26.97
VIRUS	7.98	-39.89
VECTOR	7.83	24.45

Country	f(%)	σ
United States	31.71	12.76
Brazil	13.38	11.59
India	11.90	-1.11
United Kingdom	7.23	2.76
France	6.31	4.77
China	5.71	-2.01
Australia	4.43	2.50
Thailand	4.40	-5.38
Mexico	3.69	6.42
Malaysia	3.61	-1.66

Author	f(%)	σ
Raikhel AS	0.71	9.78
Ritchie SA	0.68	8.18
Severson DW	0.68	9.82
Benelli G	0.60	9.84
Hoffmann AA	0.58	8.14
Failloux A-B	0.56	6.97
Govindarajan M	0.55	9.86
Scott TW	0.52	3.78
Chareonviriyaphap T	0.48	7.99
Tabanca N	0.48	8.96

Subject	f(%)	σ
Medicine	53.59	-34.81
Agricultural and Biological Sciences	34.19	41.99
Immunology and Microbiology	28.13	-11.17
Biochemistry, Genetics and Molecular	23.81	4.71
Environmental Science	7.87	17.19

Para cada clase de metadato (Keyword, Palabras de títulos, País, Autor, Tema), se presenta la frecuencia de aparición en el clúster y σ (significancia), que es una medida de la frecuencia de aparición en el clúster ponderada con la población total, mientras que valores positivos indican alta especificidad del elemento en ese clúster.

Si bien esto nos permite ver la particularidad de los temas abordados, es sumamente necesario también poder ahondar en los subcluster que contiene cada cluster para terminar de realizar un análisis más integrador del mismo.

Para complementar este análisis, es relevante también descargar la bibliografía más representativa del cluster. Esta se calcula del mismo modo, obteniendo los papers

mas citados y más representativos, midiendo la cantidad de veces que un paper se encuentra conectado con otros del cluster. En este caso, obtuvimos los 10 más citados y los 10 más representativos que se pueden descargar y proceder a su lectura completa para una comprensión definitiva del clúster.

Figura 12. Los papers más representativos y los 15 papers más citados del clúster AEDES AEGYPTI.

			Most Cited publications
<i>d₁₀₀</i>	Times Cited	Publication Ref	Publication Title
47	1105	Dudchenko O, 2017, <i>Science</i> (356)	De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds
2618	1460	Kraemer MUG, 2015, <i>eLife</i> (4)	The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus
795	1279	Brady OJ, 2012, <i>PLoS Neglected Tropical Diseases</i> (6)	Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus
1107	1274	Moreira LA, 2009, <i>Cell</i> (139)	A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium
342	1193	Agarwala R, 2016, <i>Nucleic Acids Research</i> (44)	Database resources of the National Center for Biotechnology Information
823	1181	Twissarkin KA, 2007, <i>PLoS Pathogens</i> (3)	A single mutation in Chikungunya virus affects vector specificity and epidemic potential
707	1124	Hoffmann AA, 2011, <i>Nature</i> (476)	Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission
975	993	Walker T, 2011, <i>Nature</i> (476)	The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations
444	911	Nene V, 2007, <i>Science</i> (316)	Genome sequence of Aedes aegypti, a major arbovirus vector
130	893	Wilm A, 2012, <i>Nucleic Acids Research</i> (40)	LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from...
99	883	Loo Y-M, 2008, <i>Journal of Virology</i> (82)	Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity
731	855	Benedict MQ, 2007, <i>Vector-Borne and Zoonotic Diseases</i> (7)	Spread of the tiger: Global risk of invasion by the mosquito Aedes albopictus
503	750	McMeniman CJ, 2009, <i>Science</i> (323)	Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti
146	736	Githeko AK, 2000, <i>Bulletin of the World Health Organization</i> (78)	Climate change and vector-borne diseases: A regional analysis
			Most Representative Publications
<i>d₁₀₀</i>	Times Cited	Publication Ref	Publication Title
2875	1	Unlu I, 2019, <i>CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources</i> (14)	Aedes albopictus in America: Current perspectives and future challenges
2860	18	Ritchie SA, 2021, <i>Ecology and Control of Vector-Borne Diseases</i> (6)	Insecticide-based approaches for dengue vector control
2730	74	Kaufman EH, 2017, <i>Journal of Infectious Diseases</i> (246)	Zika Virus Mosquito Vectors: Competence, Biology, and Vector Control
2618	1460	Kraemer MUG, 2015, <i>eLife</i> (4)	The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus
2556	19	Tangena J-AA, 2018, <i>PLoS ONE</i> (13)	Bionomics and insecticide resistance of the arboviral vector Aedes albopictus in northern Lao PDR
2490	265	Brady OJ, 2014, <i>Parasites and Vectors</i> (7)	Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus...
2291	254	Bouritzis K, 2014, <i>Acta Tropica</i> (132)	Harnessing mosquito-Wolbachia symbiosis for vector and disease control
2256	7	Rahim J, 2017, <i>Journal of the American Mosquito Control Association</i> (33)	Adulticidal Susceptibility Evaluation of Aedes albopictus Using New Diagnostic Doses in Penang Island, Malaysia
2228	84	Manni M, 2017, <i>PLoS Neglected Tropical Diseases</i> (11)	Genetic evidence for a worldwide chaotic dispersion pattern of the arbovirus vector, Aedes albopictus
2204	7	Camara DCP, 2022, <i>Tropical Medicine and Infectious Diseases</i> (7)	Entomological Surveillance of Aedes Mosquitoes: Comparison of Different Collection Methods in an Endemic Area in RIO de...
1997	26	Parry R, 2021, <i>Microorganisms</i> (9)	Uncovering the worldwide diversity and evolution of the virome of the mosquitoes aedes aegypti and aedes albopictus
1996	16	Bharati M, 2017, <i>Acta Tropica</i> (170)	Insecticide susceptibility status and major detoxifying enzymes activity in Aedes albopictus (Skuse), vector of dengue...
1987	14	Padilla-Torres SD, 2013, <i>PLoS ONE</i> (8)	Modeling Dengue Vector Dynamics under Imperfect Detection: Three Years of Site-Occupancy by Aedes aegypti and Aedes...
1979	34	Moretti R, 2018, <i>PLoS Neglected Tropical Diseases</i> (12)	Combining Wolbachia-induced sterility and virus protection to fight Aedes albopictus-borne viruses

Cuando la interpretación no es concluyente, o se requiere un análisis más profundo, se recomienda complementar con estrategias cualitativas, como entrevistas o grupos focales, que permiten contrastar los hallazgos con la perspectiva de actores del campo. Los datos anteriores proporcionan importantes elementos para elaborar entrevistas, listas de personas que entrevistar, instituciones que visitar y otro tipo de información útil en las investigaciones.

Así, la bibliometría enriquecida complejiza mucho los primeros análisis que realizamos con bibliometría básica en el primer momento de nuestra investigación. En ese entonces, retomamos diferentes estudios bibliométricos que analizaron la producción científica en diferentes repositorios y bases de datos con diferentes reportes

temporales. Los ejes de análisis solían ahondar en la cantidad de publicaciones del tema, año con mayor pico en las producciones, los países e instituciones científicas con mayor cantidad de publicaciones, y los idiomas más utilizados en las mismas.

6.5. Comparación con otros estudios bibliométricos similares

Para contrastar la calidad de los datos obtenidos y las interpretaciones proporcionadas en este ejemplo, realizamos una búsqueda sistematizada de otros estudios bibliométricos sobre Dengue, para evaluar si nuestros hallazgos eran de mejor calidad. Encontramos 4 trabajos relevantes. En el siguiente cuadro, presentamos los resultados obtenidos:

Tabla 4. Cuadros comparativo con otros estudios bibliométricos sobre dengue.

	A Scientometric and Bibliometric Analysis of World Scientific Production on Dengue in Microsoft Academic (1828-2019)	A Scientometric Mapping of World Publications	Dengue research: a bibliometric analysis of worldwide and Arab publications during 1872–2015 (Zyoud,2016)	Scopus (producción propia, 2025)	Open Alex (Producción Propia,2025)
Autor	José Federico Medrano	Raj Kumar Bhardwaj*	Sa'ed H. Zyoud	yo	yo
Fuente de datos	Microsoft Academic	Scopus	Scopus	Scopus	Open Alex
Año de análisis	2020	2014	2016	2025	2025
Periodo de análisis	1828-2019	2001-2012	1872-2015	Sin límite	Sin límite
Cantidad de publicaciones	34,955	9618	19,581	47,371	78,120
Pico de publicaciones	2016	2012	2014	2021	2016 con leve baja desde entonces
Países más productivos	EEUU, Brasil	EEUU, India	EEUU, India	EEUU, India	EEUU, India
Instituciones más productivas	1. Fundação Oswaldo Cruz (Brasil, 2,267 pubs) 2. CDC (EE. UU.) (1,725 pubs) 3. Instituto Pasteur (Francia, 1,343 pubs)	1. Mahidol University (Tailandia) 2. Fundação Oswaldo Cruz (Brasil)	1. Mahidol University (Tailandia, 620 pubs) 2. Fundação Oswaldo Cruz (Brasil, 495 pubs) 3. CDC (EE. UU.) (395 pubs)	1. Fundacao Oswaldo Cruz 2. Mahidol University	1. Politeknik Kesehatan Kemenkes Semarang (Indonesia) 2. Diponegoro University (Indonesia) 3. Airlangga University (Indonesia)
Idiomas				Inglés, portugues, indonesio.	Inglés, español, chino.
Mención a Ciencias Sociales	-	-	-	1018 de 47,371 (2% por ciento) resultados Pico en el 2024	9126 de 78120 (1% por ciento) resultados Indonesia lidera en esta área

En este análisis comparativo, y luego de consultar diversas fuentes, pudimos dar cuenta que tanto en Openalex (Lifszyc & Levin, 2025), Scopus (Levin & Lifszyc, 2025; Zyoud, 2016), como en Microsoft Academic (Medrano, 2020), se muestra a Estados Unidos como principal pro-

ductor de artículos académicos. Sin embargo, hay diferencias con respecto al segundo lugar: mientras que Microsoft Academic era Brasil, en OpenAlex y Scopus es India. En relación a las Ciencias Sociales, la tendencia se modifica y coloca a Indonesia en primer lugar. En Scopus, Brasil

toma el segundo lugar en vez de India. Estas diferencias pueden transformarse en preguntas para analizar la dinámica de producción de conocimiento. Medrano (2020) profundiza aún más este análisis y realiza un mapeo institucional de las producciones científicas en torno al dengue a partir de las producciones en Microsoft Academic entre los años 1828-2019.

En una primera instancia, el análisis se centró en dimensiones cuantitativas y en la identificación de cuestiones vinculadas a los repositorios y bases de datos. En cambio, en la segunda etapa fue posible profundizar en el estudio de las investigaciones sobre dengue, superando una mirada uniforme y reconociendo las principales discusiones que se desarrollan en el campo. A diferencia de publicaciones previas, que abordan los cambios a lo largo del tiempo considerando el corpus bibliográfico de manera global, este enfoque permite relevar las discusiones más relevantes en cada momento particular. La clusterización, además de diferenciar las áreas temáticas predominantes y los cambios en las dinámicas del campo, abre la posibilidad de identificar nichos de investigación menos explorados o emergentes, lo cual podría orientar futuras líneas de estudio.

7. Palabras finales

La construcción de antecedentes de investigación continúa siendo un obstáculo metodológico central en las etapas iniciales de tesis e investigaciones académicas. Frente al volumen creciente de información, se requieren estrategias más sistemáticas y escalables para acompañar los procesos de formulación de problemas.

En este artículo, se propuso una estrategia de bibliometría enriquecida que articula técnicas métricas con enfoques cualitativos, permitiendo construir antecedentes más contextualizados y robustos. A diferencia de las metodologías bibliométricas más restringidas, esta propuesta permite organizar grandes corpus teóricos con criterios replicables, identificar vacíos temáticos y conexiones no evidentes, y combinar la exploración cuantitativa con análisis cualitativos profundos. Estas estrategias se pueden adaptar a los múltiples campos temáticos y niveles de experiencia del investigador.

A pesar de sus ventajas, la bibliometría enriquecida implica desafíos técnicos y metodológicos relevantes:

se requiere el acceso a bases de datos especializadas, el procesamiento de grandes volúmenes de información, y la necesidad de competencias mixtas en análisis de datos, programación y métodos cualitativos. Su implementación requiere tiempo, recursos y aprendizaje técnico.

La potencialidad de la bibliometría enriquecida es especialmente valiosa en investigaciones interdisciplinarias, emergentes o poco sistematizadas. Al mismo tiempo, invita a repensar los modos de construir conocimiento en entornos académicos crecientemente automatizados, proponiendo una articulación equilibrada entre las herramientas digitales y la mirada interpretativa del investigador. De este modo, la bibliometría enriquecida no sólo permite organizar los saberes producidos de forma sistemática, sino que también constituye un punto de partida fértil para imaginar nuevas preguntas y explorar caminos de investigación innovadores.

8. Referencias

- Agostini, L., & Nosella, A. (2019). Inter-organizational relationships involving SMEs: A bibliographic investigation into the state of the art. *Long Range Planning*, 52(1), 1-31. <https://doi.org/10.1016/j.lrp.2017.12.003>
- Boyack, K.W. and Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. *J. Am. Soc. Inf. Sci.*, 61: 2389-2404. <https://doi.org/10.1002/asi.21419>
- Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, 126(5), 376-380. <https://doi.org/10.7326/0003-4819-126-5-199703010-00006>
- Cordeiro, A. M., Oliveira, G. M. D., Rentería, J. M., & Guimarães, C. A. (2007). Revisão sistemática: Uma revisão narrativa. *Revista do Colégio Brasileiro de Cirurgiões*, 34(6), 428-431. <https://doi.org/10.1590/S0100-69912007000600012>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Duque, P., & Cervantes-Cervantes, L.-S. (2019). Respon-

sabilidad Social Universitaria: Una revisión sistemática y análisis bibliométrico. *Estudios Gerenciales*, 35(153), 451-464.

Ferenhof, H.A. and Fernandes, R.F. (2016). Desmistificando a revisão de literatura como base para redação científica: método SFF. *Revista ACB: Biblioteconomia em Santa Catarina*, 21(3), pp.550-563.

Grauwin, S., & Jensen, P. (2011). Mapping scientific institutions. *Scientometrics*, 89(3), 943-954.

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10-25. <https://doi.org/10.1002/asi.5090140103>

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the royal society of medicine*, 96(3), 118-121.

Leydesdorff, L. and Rafols, I. (2009). A global map of science based on the ISI subject categories. *J. Am. Soc. Inf. Sci.*, 60: 348-362. <https://doi.org/10.1002/asi.20967>

Levin, L. G., & Lifszyc, S. (2025). *Metodología para la construcción sistemática de antecedentes de investigación*. Zenodo. <https://doi.org/10.5281/zenodo.17259447>

Medrano, J. (2020). *A Scientometric and Bibliometric Analysis of World Scientific Production on Dengue in Microsoft Academic (1828-2019) / Un Análisis Cienciométrico y Bibliométrico de la Producción Científica Mundial Sobre Dengue en Microsoft Academic (1828-2019)*. <https://doi.org/10.18687/LAC-CEI2020.1.1.625>

Powell, R. A., Single, H. M., & Lloyd, K. R. (1996). Focus groups in mental health research: Enhancing the validity of user and provider questionnaires. *International Journal of Social Psychiatry*, 42(3), 193-206. <https://doi.org/10.1177/002076409604200303>

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 25(4), 348-349.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.*, 24, 265-269. <https://doi.org/10.1002/asi.4630240406>

Souza, M. S. (2007). La centralidad del estado del arte en la investigación científica. *Tram[p]as de la Comunicación y la Cultura*, no. 51. <http://sedici.unlp.edu.ar/handle/10915/37676>

Wainerman, C., & Levin, L. G. (2024). La supervisión de tesis como objeto de estudio. Origen y desarrollo a través de sus publicaciones. *Ciencia, Docencia y Tecnología*, 35(72). <https://doi.org/10.33255/3572/2020>

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational research methods*, 18(3), 429-472.

Zyoud, S. H. (2016). Dengue research: A bibliometric analysis of worldwide and Arab publications during 1872-2015. *Virology Journal*, 13, 78. <https://doi.org/10.1186/s12985-016-0534-2>.