



Introducción a la Bibliometría

Conceptos básicos y desafíos

por

Claudio Bustos

Financiado por Agencia Nacional de Investigación y Desarrollo / FONDECYT Regular N° 1201681



Conceptos básicos sobre bibliometría

Desafíos en estudios bibliométricos

Conceptos básicos sobre bibliometría



Se define como **bibliometría** como el tratamiento matemático y estadístico de la productividad científica, que nos permite entender tanto la generación como los efectos de la actividad científica (Gallegos et al., 2020).

El análisis de la productividad se puede realizar a distintos niveles: individuos, grupos, instituciones, temas y/o países.

Se puede usar para conocer las dinámicas y la evolución de la productividad científica, y para tomar decisiones sobre qué priorizar en investigación.



- ❖ **Indicadores bibliométricos:** Corresponden a datos estadísticos que se obtienen a partir de los metadatos y el contenido de las publicaciones (Gómez y Bordons, 1996).
- ❖ **Metadatos:** Conjunto de datos que describen a un recurso o texto específico, que no necesariamente son parte de su contenido. Ej: Año de publicación, nombre de autores, país del editor de la revista.



Llerena y Arévalo (2021) indica que existen dos tipos básicos de indicadores:

- ❖ **Indicadores de actividad.** Contabilizan datos referidos directamente a las publicaciones. Ej: Número de autores por artículo, publicaciones por año, idioma de las publicaciones.
- ❖ **Indicadores de impacto.** Refieren a la incidencia de las publicaciones en la comunidad científica, generalmente a través del análisis de las citas. Ej: factor de impacto de la revista (IF), índice H, índice de inmediatez.

Existe un desarrollo importante de **índices de colaboración**, que miden el grado en que los artículos son creados por diversos autores / instituciones / países (Savanur y Skiranth, 2010).



- ❖ El **análisis de redes sociales (ARS)**, se desarrolla como una herramienta de medición y análisis de las estructuras sociales que nacen a partir de las relaciones de distintos actores sociales .
- ❖ El ARS se aplica en la bibliometría para el análisis de las relaciones entre investigadores, instituciones, países, etc.
- ❖ Como sustento teórico para el ARS se suele utilizar la **teoría de grafos**. Esta ofrece herramientas matemáticas que permiten cuantificar la importancia de cada nodo para determinar su influencia sobre el resto de la red (Sanz, 2003).



- ❖ Basados en lenguajes de programación
 - ❖ R: Bibliometrix (<https://www.bibliometrix.org/home/>)
 - ❖ Python: litstudy (<https://github.com/NLeSC/litstudy>)
- ❖ Basados en interface gráfica
 - ❖ VosViewer (<https://www.vosviewer.com>)
 - ❖ Citespace - máximo 1000 nodos versión libre (<https://citespace.podia.com/>).

Desafíos en estudios bibliométricos



- ❖ En una primera instancia, puede parecer que hacer un estudio bibliométrico es una actividad menor o relativamente fácil: sería exportar la información de la base de datos, procesar en un software especializado y listo.
- ❖ Si bien esto puede ser cierto para una aproximación informal a un área de estudio (ej: paquete [litstudy](#) de python), no es correcto para un estudio bibliométrico serio.
- ❖ Existen múltiples desafíos en un estudio bibliométrico, que lo pueden hacer más trabajoso que un estudio de fuente primaria.



- ❖ Dentro de cada base de datos, si bien hay intentos de estandarización, es común encontrar que la misma entidad (autor, afiliación, artículos) aparezca con distintos nombres / códigos.
- ❖ El proceso mediante el cuál se ubica y unifica a los duplicados se denomina **deduplicación**.
- ❖ En estudios pequeños, la deduplicación se puede hacer manual mediante diccionarios. En estudios grandes, se deben usar reglas programadas o técnicas de machine learning para apoyar el proceso.



- ❖ Si trabajar con un base ya implica mucho trabajo, usar varias genera aún más problemas de deduplicación.
- ❖ En el caso de las afiliaciones múltiples de un autor, algunas bases funden varias filiaciones en una sola. Por tanto, no se pueden fundir directamente las afiliaciones entre bases.
- ❖ **Ventaja:** Se puede validar la deduplicación, al comparar los resultados entre las distintas bases. Hemos logrado esto para los autores y las revistas, pero no las afiliaciones.



En nuestro estudio sobre la psicología en Colombia, en el periodo 2013-2022 en las bases de datos WoS, Scopus y Scielo, encontramos:

- ❖ Se enumerar 3711 nombres distintos de afiliaciones. Al deduplicar, se identifican 466 afiliaciones únicas. Por tanto, para cada afiliación real en promedio se generan 7.96 nombres diferentes.
- ❖ En el caso específico de la Universidad del Valle, se distinguen 75 nombres distintos.
- ❖ Se enumerar 19637 nombres de autores distintos. Tras la deduplicación, se identifican 14368 autores únicos. Por tanto, para cada autor se generan en promedio 1.37 nombres diferentes.



- ❖ Cualquier intento de clasificar a los autores, afiliaciones o temas de los textos por algún criterio no explícito requiere identificar a las categorías a las que pertenecen. Esto requiere ya sea clasificación manual de cada texto (muy costoso) o usar métodos de NLP - procesamiento de lenguaje natural.
- ❖ Realizar asociaciones entre indicadores requiere usualmente el uso de métodos estadísticos multivariados. Especialmente relevantes son aquellos de análisis de conglomerados / perfil o clase latente.
- ❖ Los métodos de análisis de red deben ser usados con un conocimiento profundo del área disciplinar o tema estudiado, para darle sentido a los grupos que se generen.



- ❖ **Género:** Se sabe que en disciplinas STEM, las mujeres tienen poca representación, lo que se refleja en publicaciones. ¿Qué ocurre en salud?
- ❖ **Internacionalización vs extranjerización:** Una investigación cerrada a las fronteras nacionales es limitada, pero tampoco es buena que sea completamente dirigida a los intereses foráneos. ¿Qué indicadores son apropiados para detectar esto?
- ❖ **Análisis del cambio en las redes:** Si bien hay software que apoyan el análisis de redes, no hay acuerdo sobre cómo analizar el cambio en las redes.



¡Gracias!



- Gallegos, M., Pérez-Acosta, A. M., Klappenbach, H., López-López, W., & Bregman, C. (2020). Los estudios bibliométricos en el campo de la psicología iberoamericana: Una revisión metabibliométrica. *Interdisciplinaria Revista de Psicología y Ciencias Afines*, 37(2), 95-115. <https://doi.org/10.16888/interd.2020.37.2.6>
- Llerena Paz, M. A., & Arévalo Avecillas, M. E. (2021). Indicadores bibliométricos: origen, definición y aplicaciones científicas en el Ecuador. *Espíritu Emprendedor TES*, 5(1), 130-153. <https://doi.org/10.33970/eetes.v5.n1.2021.253>
- Sanz Menéndez, L. (2003). Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes. *Apuntes de Ciencia y Tecnología*, 7, 21-29.
- Savanur, K., & Srikanth, R. (2010). Modified collaborative coefficient: A new measure for quantifying the degree of research collaboration. *Scientometrics*, 84(2), 365-371. <https://doi.org/10.1007/s11192-009-0100-4>