

Workshop Iberoamericano de estudios métricos de la actividad científica orientada a temas locales/regionales. La Plata, 21 y 22 de agosto de 2018

El análisis y representación del contenido de la producción científica desde una perspectiva informétrica: aportes metodológicos

Liberatore, Gustavo*, Depto. Ciencia de la Información. Fac. Humanidades. UNMdP, Mar del Plata, Argentina.

RESUMEN

El estudio de la producción científica orientada a una localización geográfica específica comporta algunos desafíos de índole metodológica en términos de su análisis, representación e interpretación. El enfoque del análisis de dominios propone y define una serie de herramientas para el abordaje de las áreas de conocimiento atendiendo a la complejidad que supone determinar sus contornos y fronteras. Desde esta perspectiva el presente trabajo propone un recorrido metodológico para el tratamiento del corpus documental publicado en la corriente principal en el área de la ciencia mediambiental (*environmental science*) aplicada a la Argentina en el período 2007-2016.

Palabras clave: Medio ambiente; Argentina; Análisis bibliométrico

INTRODUCCIÓN

El “paradigma del análisis de dominio” es un enfoque teórico de la Ciencia de la Información que establece que la mejor forma de comprender la organización de los campos de conocimiento es delimitando y estudiando las distintas “comunidades discursivas” que los componen. Dichas comunidades deben abordarse desde una perspectiva holística, a partir de los contextos históricos, sociales y culturales en que se constituyen. Desde esta perspectiva, el conocimiento puede ser organizado y representado dinámicamente desde múltiples variables: la organización social del trabajo, los patrones de cooperación, el lenguaje, las formas de comunicación, los sistemas de información y los criterios subjetivos de relevancia. (Hjorland y Albrechtsen, 1995; Hjorland, 2002).

Dentro de este marco se proponen un conjunto de métodos y técnicas empíricas orientadas al estudio y descripción de los campos de conocimiento a partir del análisis de las relaciones que los estructuran (conceptos, entidades) y de sus vinculaciones con otros campos (Smiraglia, 2015). De entre ellos se destaca el método bibliométrico como el de mayor aporte en la delimitación y estudio de los campos de conocimiento tomando como unidades de análisis a los autores, los contenidos y las citas. Complementariamente, distintas técnicas de visualización de información y de dominios científicos proporcionan formas de representación de la información complejas que aportan claves de interpretación ocultas para las técnicas y métodos más tradicionales (Díaz-Pérez, Moya-Anegón & Carrillo-Calvet, 2017).

Teniendo en cuenta esta perspectiva se pretende abordar una problemática específica de las investigaciones empíricas realizadas con base en la producción científica que se relaciona con la captura, análisis y representación de la investigación científica aplicada a entornos geográficos específicos. El desafío aquí

es delimitar y representar de manera combinada dos dominios diferentes: el geográfico y el disciplinar o científico.

Bajo este marco se tomó como objeto de estudio la producción científica indizada en la corriente principal sobre medio ambiente aplicada a la Argentina, en donde al menos uno de los autores de las publicaciones reportara una afiliación en dicho país. A partir de esta delimitación se fijaron una serie de interrogantes que guiaron esta investigación: ¿Cuáles son las áreas geográficas más estudiadas en la Argentina en relación al medio ambiente? ¿Cuáles son las temáticas mediambientales investigadas? ¿Cómo se distribuyen las temáticas en función de las áreas geográficas?

MATERIALES Y METODOLOGÍA

El corpus bibliográfico fue seleccionado de la base de datos SCOPUS tomando en consideración los artículos publicados en el periodo 2007-2016. Se recuperaron 3672 registros en base a la definición de la siguiente ecuación de búsqueda¹:

(TITLE-ABS-KEY (argentin) AND AFFILCOUNTRY (argentina)) AND PUBYEAR > 2006 AND PUBYEAR < 2017 AND (LIMIT-TO(SUBJAREA,"ENVI"))*

Luego de la exportación de los resultados en formato *BibTex* se procedió a analizar el contenido tomando en consideración el campo de palabras clave de autor (*author keywords*) a partir del cual se construyó un índice de 9689 entradas o expresiones. Con este resultado se aplicaron una serie de cálculos con vista a la representación de las variables temática y geográfica.

Dentro del conjunto seleccionado de palabras clave se contabilizaron las frecuencias absolutas y luego se aplicó el cálculo de co-ocurrencias (*co-word analysis*) con el objetivo de generar una matriz de datos para la representación de las relaciones correspondientes.

Para analizar y representar las áreas geográficas involucradas de Argentina se procedió a una detección manual de todas aquellas expresiones que remitieran a una ubicación geográfica específica. Se aislaron 514 palabras clave con este perfil. Para su organización se realizó una lematización geográfica agrupando a las distintas expresiones en regiones y accidentes geográficos y provincias. Como resultado se obtuvo una taxonomía *ad hoc* constituida por 14 regiones y accidentes geográficos y 22 provincias.

Para el cálculo de frecuencias y análisis de co-ocurrencias se utilizó el software BibExcel (Persson) en su versión 2016-02-20. En la tabulación de datos y generación de algunas gráficas se usó el paquete estadístico Excel en su versión 2013. Las matrices de co-ocurrencia fueron representadas por grafos derivados de la aplicación del análisis de redes sociales (ARS). Para la valoración de los componentes de la red se utilizaron medidas de centralidad (grado e intermediación) y de modularidad (análisis de comunidades o clusters). Para la visualización de datos, se ha utilizado el programa Gephi en su versión 0.9.2 (2017), seleccionando para la distribución el algoritmo ForceAtlas2.

¹ El área temática de búsqueda es la proporcionada por la clasificación interna de SCOPUS denominada "Environmental Science".

RESULTADOS Y DISCUSION

El diseño de una taxonomía que agrupara los lugares geográficos hacia donde estaban enfocadas las investigaciones permitió la representación, en primer lugar, de las áreas y sitios de mayor interés en el campo disciplinar estudiado. Así, se pudo verificar que en el grupo *regiones y accidentes geográficos* la Patagonia (18%) y la región Pampeana (7%) concentran la mayor cantidad de publicaciones (Figura 1). En el grupo *Provincias* la Ciudad Autónoma de Buenos Aires² (11%) y Córdoba (3%) aparecen como los espacios geográficos con mayor presencia.



Figura 1. Distribución porcentual de las regiones y accidentes geográficos estudiados en relación al total de publicaciones observadas (n=3672).

Luego de circunscribir el dominio geográfico el análisis se enfocó en el contenido temático de las publicaciones. A partir del cálculo de co-ocurrencias de las palabras clave se aplicó el análisis de redes sociales con la concurrencia de dos medidas específicas: centralidad de grado de los nodos y modularidad de red. Previamente, en la generación del primer grafo se procedió a la poda del nodo *Argentina* para evitar una representación egocéntrica y poco útil a los fines de esta investigación. Luego de equilibrar la red, la primera medida permitió observar cuáles eran las principales temáticas globales del corpus bibliográfico determinadas por el nivel de participación de los nodos (número de enlaces) dentro de la estructura de relaciones. En un segundo momento se aplicó el cálculo de modularidad que, como medida estructural de la red, permite observar si existen módulos o divisiones conformadas por nodos densamente interconectados entre sí y con conexiones más débiles con el resto de los módulos (Figura 2). Esto permitió comprobar que los

² Desde 1994 la Ciudad de Buenos Aires goza de un estatus administrativo similar al de las provincias que constituyen el territorio argentino por lo que, en esta división, se la considera como tal.

nodos con mayor peso o grado calculados previamente definen las principales comunidades o clusters temáticos.

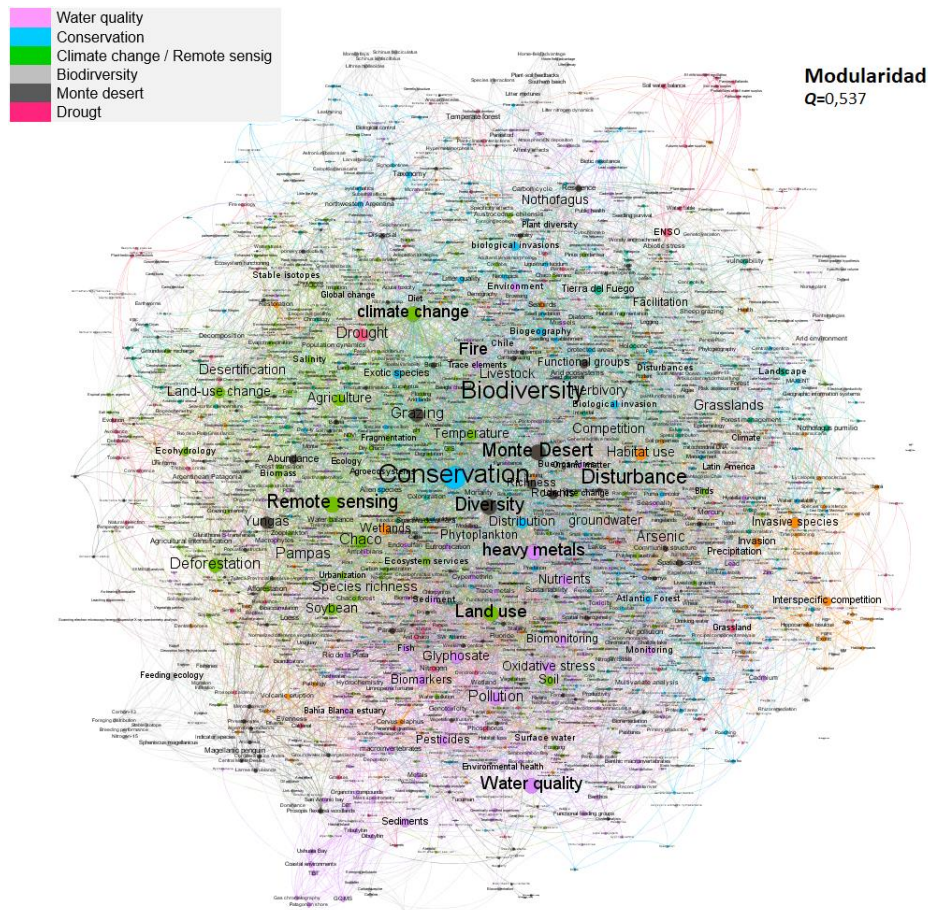


Figura 2. Principales clusters temáticos observados.

A partir de esta delimitación general en lo geográfico y temático se buscó cruzar ambos dominios con el fin de observar la distribución de los tópicos investigados en función de un área geográfica específica. Se tomó en consideración la *Región del Chaco*, a modo de ejemplo, para lo cual se procedió a aislar los registros correspondientes a esta área y realizar un cálculo de co-ocurrencia de las palabras claves y su representación mediante un grafo de red social (Figura 3). A la red resultante se le aplicó nuevamente el cálculo de modularidad a partir del cual pudo discriminarse los principales temas que concentran la atención en el campo medioambiental. Para una mejor visualización el grafo fue editado solamente con las principales etiquetas³ (temas) que representan a las comunidades o clusters observados. Así, la biodiversidad, el cambio climático, el agua y el impacto mediambiental constituyen los principales frentes de investigación, mientras que aparecen en foco dos subregiones, el Chaco árido y el semiárido, como áreas geográficas de interés.

³ Para una mejor comprensión han sido traducidas al español.

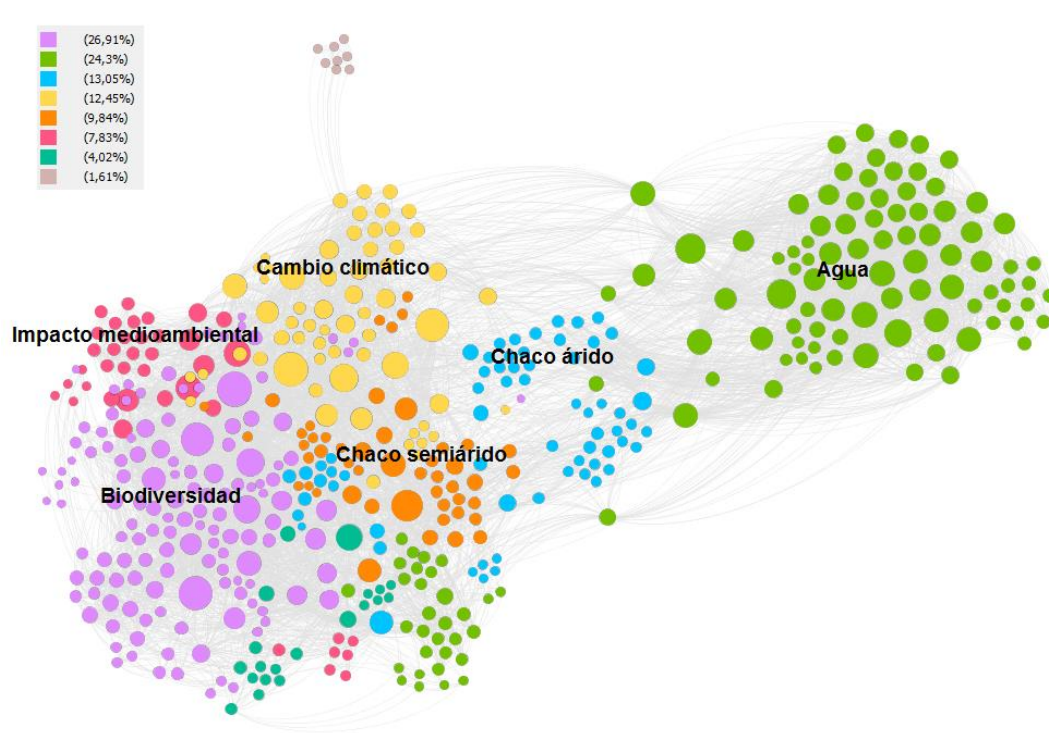


Figura 3. Principales temáticas estudiadas en la Región del Chaco.

CONCLUSIÓN

Dada la naturaleza metodológica de esta breve presentación respecto a las posibilidades o variantes para el análisis y representación de la investigación aplicada a espacios geográficos se postulan tres reflexiones sobre la experiencia presentada:

- La faceta geográfica en un análisis de contenido debe estar sujeta a modelos de representación estandarizados, tanto desde el punto de vista de la captura de datos (búsqueda) cuanto de sus modos de organización (taxonomías).
- Por el contrario, las temáticas vinculadas a los espacios geográficos se encuentran sujetas a formas discursivas asociadas a los distintos abordajes disciplinares y a la manera en que están relacionadas.
- Finalmente, son los interrogantes (intereses) fijados a priori los que demandarán las formas de representación y visualización más adecuadas para poder aportar respuestas concretas.

REFERENCIAS

- Díaz-Pérez, M., Moya-Anegón, F. & Carrillo-Calvet, H. A. (2017). Técnicas para la visualización de dominios científicos y tecnológicos. *Investigación Bibliotecológica*, Núm. especial de Bibliometría, 17-42.
- Hjørland, B. (2002) Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.
- Hjørland, B. & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal Of The American Society For Information Science*, 46(6), 400-425.
- Smiraglia, R. P. (2015). *Domain Analysis for Knowledge Organization: tools for ontology extraction*. Amsterdam: Elsevier-Chandos Publishing.