

Aplicación del Análisis de Redes Sociales (ARS) a la generación de mapas bibliométricos que caracterizan el uso de revistas por parte de investigadores. La experiencia de la Biblioteca del Centro de Investigaciones Ópticas (CONICET-CIC)

Lic. Claudia M. González
UNLP-FAHCE / CONICET
(claudiag@ciop.unlp.edu.ar)

Resumen

Se presenta en este trabajo una aproximación metodológica al estudio de uso de revistas por parte de los investigadores del Centro de Investigaciones Ópticas. Integrando la técnica de Análisis de Citaciones y de Análisis de Redes Sociales se realiza la construcción de mapas con gran capacidad de síntesis de información compleja. Se estima que su elaboración contribuiría con información valiosa a la hora de pensar la distribución equitativa de recursos económicos para compra de títulos.

1. Introducción

La identificación de los materiales que los usuarios encuentran esenciales para su investigación ha sido siempre el criterio directriz en el desarrollo de las colecciones de bibliotecas de centros de investigación. En este tipo de Unidades de Información, la colección siempre se alinea más directamente con las necesidades marcadas por los usuarios que con cualquier otro criterio que el bibliotecario pueda utilizar.

Las bibliotecas académicas y de investigación de Argentina cuentan desde el año 2002 con la Biblioteca Electrónica de Ciencia y Tecnología¹, portal de revistas que, gestionado por el Ministerio de Ciencia y Tecnología de la Nación, ofrece más de 9.000 títulos a las Universidades y organismos de investigación y desarrollo del país como CONICET, INTI, INTA, etc. Dado que los proveedores permiten el acceso solo a rangos de IP autorizados y que además el país presenta problemas de infraestructura en TICs, se presentan situaciones de enmascaramiento de redes pequeñas dentro de redes universitarias de mayor envergadura. Tal es el caso del Centro de Investigaciones Ópticas y la Universidad Nacional de La Plata.

¹ <http://biblioteca.mincyt.gov.ar>

La posibilidad de búsqueda y descarga directa del material por parte del usuario final hace que el registro tradicional de préstamo o pedido de localización que el bibliotecario realiza disminuya notablemente. Si además, por lo anterior, se hace técnicamente imposible distinguir por parte de los proveedores las descargas de material realizada por grupos específicos; la información brindada por sus sistemas (COUNTER² y otros) no aporta a las bibliotecas de institutos de investigación información de relevancia en este sentido. En suma, los bibliotecarios comienzan a perder fuentes de datos necesarias para la evaluación y desarrollo de sus propias colecciones y pierden capacidad de intervenir con criterio respaldado en información, en las decisiones tomadas por el consorcio nacional.

El empleo de la evidencia de uso reciente de los diferentes materiales como estimador de uso futuro ha sido tratado por Fussler y Simon [1] en relación con la disposición en el almacenamiento de los libros. Esto motivó que se realizaran estudios sobre el uso de las revistas para determinar su probable demanda, cuestión que siempre ha sido de vital interés para el desarrollo de colecciones en épocas de ofertas crecientes y presupuestos magros. Las tres medidas básicas de uso corresponden al uso en sala, a la circulación y a la citación. Un trabajo de D. Bleicic [2] demuestra que existe correlación entre los tres tipos de medidas y dado que los registros tradicionales llevados por los bibliotecarios argentinos se encuentran muy distantes de mostrar la realidad de uso, parece apropiado para este contexto aplicar la técnica de análisis de citaciones como parte de una metodología posible en la evaluación de colecciones. La aplicación de esta rama de la bibliometría en la determinación de patrones y frecuencias de las citas realizadas³ permite por ejemplo: encontrar las fuentes principales usadas en una disciplina, caracterizar los materiales por idioma, formas de literatura o fecha de publicación.

El uso de las citas como parte de la metodología de evaluación de colecciones, lo que se conoce como *citation checking*, es muy antiguo. Nisonger [3] menciona como primer antecedente al trabajo realizado en 1848 por un secretario del Smithsonian Institution, Charles Coffin Jewett, quien verificó la existencia en los catálogos de bibliotecas americanas de los materiales referenciados en 4 obras consideradas fundamentales para su época. En 1927, Gross y Gross [4], proponen la revisión de las referencias del volumen correspondiente al año 1926 de *The Journal of the American Chemical Society*, cómo

² <http://proyectcounter.org>

³ En este trabajo el término “citas” o “citación” se refiere a las entradas bibliográficas realizada en las notas finales, apartado de referencias o bibliografía que acompañan a un trabajo publicado y que se presume indican los materiales que el autor consultó.

manera de determinar las revistas que la Biblioteca del Pomona College debería adquirir si deseaban dar formación competitiva en el área de química a sus estudiantes con proyección a 10 años. A partir de allí, numerosos trabajos pueden citarse que dan cuenta de la aplicación de ésta técnica [5] [6] [7] [8] [9] [10]. Buenos trabajos de revisión que cubren distintos aspectos de la citación corresponden a Th. Nisonger [3], C. Urbano Salido [11] y recientemente Bornmann y Daniel [12].

Últimamente, la idea de que una imagen o gráfico sintetiza información facilitando la comprensión de un fenómeno complejo, también ha tenido lugar en la Ciencia de la Información. Diversos algoritmos de visualización se han apoyado sobre técnicas de reducción de dimensiones tales como el Análisis de Clusters, Análisis de Factores, Análisis de Componentes Principales, Escalamiento Multi-Dimensional, Análisis de Redes Sociales, Mapas Auto-organizativos [13]. Se han usado principalmente para desarrollar interfaces de recuperación de información y mostrar la estructura de dominios científicos. Particularmente, el Análisis de Redes Sociales (ARS) se utiliza para poner en evidencia estructuras complejas de relaciones como las que se dan en la Sociología de la Ciencia [14]. Por ejemplo, autores de trabajos (actores) que se relacionan entre sí (conexión) por una relación de co-citación. Para Wasserman y Faust [15] la principal característica de esta metodología es que las relaciones entre actores están por sobre los atributos individuales de estos y que dichas relaciones se manifiestan siguiendo patrones o regularidades que ponen de manifiesto una estructura susceptible de ser medida. El carácter central que tiene el uso de imágenes para mostrar este tipo de patrones, en gran medida, se debe al trabajo fundacional de Jacob Moreno, quien en 1932 crea el Sociograma.

2. Propósito

Este trabajo tiene como objetivo mostrar una aproximación metodológica basada en Análisis de Redes Sociales (ARS) a la elaboración de mapas bibliométricos que caractericen los materiales que los investigadores usaron para la elaboración de sus trabajos. Se busca brindar información relevante que pueda ser útil para el desarrollo de la colección de una biblioteca académica pequeña en la que intervienen diferentes líneas de investigación, a la vez que mostrar la disponibilidad de esos materiales en la Biblioteca Electrónica de CyT de Argentina.

3. Materiales y métodos

Se utilizó como fuente de información, la base de datos SCOPUS de Elsevier. Se buscaron los trabajos publicados por los investigadores del CIOp en el año 2007. Se descargaron los registros completos en formato TXT lo que permitió contar con la cita completa de cada trabajo en digital además de sus respectivas referencias.

Dado que interesa en este trabajo mostrar la información de uso discriminada por los diferentes grupos de investigación que funcionan en el Instituto, como primer paso se procedió a su normalización. La definición de las áreas temáticas y la propia conformación de los grupos no es tarea fácil, sin embargo, se arribó de manera consensuada a 6 categorías que permitirían clasificar de manera excluyente los trabajos. Se normalizaron sus nombres de la siguiente manera:

CIOpBE= Biospeckle

CIOpESAT= Espectroscopia Atómica

CIOpESOP= Espectroscopía Óptica

CIOpFO= Fibras Ópticas

CIOpPLME= Propagación de luz en medios estructurados

CIOpPLMTT=Propagación de luz en medios turbios y turbulentos

A partir del archivo TXT de la descarga se obtuvieron dos nuevos archivos. El primero se utilizó para generar el mapa de **Títulos donde se publica**. El segundo para generar el correspondiente a los **Títulos de revistas usados**. En ambos se separó cada registro con una línea en blanco y se borraron los datos que no interesaban para el análisis, dejando únicamente el título de la fuente de la cita en el primero y el título de cada una de las fuentes de la referencia en el segundo. En este último caso se identificó a los materiales que no fueran revistas con la siguiente nomenclatura:

L = Libro

T = Tesis

I = Informe

W = Recurso web

Además se identificaron las autocitas con el código AC. Esto se hizo a los fines de obtener una primera caracterización general de las proporciones de uso de los diferentes materiales por tipo. Las autocitas se marcaron para su posterior eliminación.

Una vez obtenidos los % generales se eliminaron los materiales que no eran revistas y las autocitas y se procedió a asignar en cada registro la sigla de grupo de procedencia.

Con apoyo de la herramienta de búsqueda y reemplazo de Word se normalizaron los títulos de las revistas de manera que no contuvieran espacios ni puntos de abreviatura. Un ejemplo del aspecto final de estos archivos es el siguiente:

1) Títulos donde se publica

CIOpESAT
RevBrasFis

CIOpPLMTT
ProcSPIE

2) Títulos de revistas usados

CIOpESAT
JPhysChemRefData
IEEEJQuantumElectronics

CIOpPLMTT
BellSystTechJ
BellSystTechJ
JStatPhys
PhysRevE

Estos archivos de texto se utilizaron para generar las respectivas matrices. Para ello se utilizó una GUI de MATLAB⁴ que indexa los archivos y genera un diccionario con cada término encontrado y los pares con sus frecuencias. Se revisó el diccionario y se detectaron errores ya que SCOPUS suele abreviar la fuente de las referencias de diferente manera. Se corrigió el archivo 2 y se volvió a generar la matriz. La GUI permite eliminar las fuentes que aparecen una sola vez (frecuencia 1). Esto se aplicó al segundo caso porque lo que se busca es generar un mapa que muestre solo lo más significativo.

El texto de la matriz resultado de MATLAB se trabajó con EXCEL, útil por su facilidad para encolumnar textos, con la finalidad de obtener el formato requerido por el archivo .NET requerido por el programa PAJEK⁵ generador de las redes. Con un procedimiento similar se generan los archivos .VEC para dar tamaño a los nodos según la frecuencia de uso. Se generó también un archivo .CLU de particiones que permitiera distinguir los nodos por colores y formas de acuerdo a las siguientes características:

Cuadrado amarillo	Diferentes grupos o líneas de investigación del CIOp
Circulo verde	Revistas disponibles en Bca. Elec. de C y T
Circulo rojo	Revistas que no están en CIOp ni en Bca. Elec. de C y T
Circulo negro	Revistas que compra la Bca. del CIOp
Circulo rosa	Revistas de acceso libre

⁴ Generador de matrices a partir de textos, desarrollada en MATLAB por D. Zeimpekis y E. Gallopoulos de la Universidad de Patras en Grecia. <http://scgroup.hpclab.ceid.upatras.gr/scgroup/Projects/TMG/>

⁵ Programa para Análisis de Redes desarrollado por Vladimir Batagelj y Andrej Mrvar de la Universidad de Ljubljana en Slovenia. <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

La red se representó aplicando el algoritmo de Kamada–Kawai [16] y se exportó en formato SVG para posteriormente ser trabajada con el programa Inkscape⁶ a los fines de mejorar su visualización.

4. Resultados

4.1. Análisis general

De la base de datos SCOPUS fueron recuperados 35 trabajos. La distribución según los diferentes grupos fue:

CIOpPLMTT	8	22,86%
CIOpPLME	8	22,86%
CIOpESOP	7	20,00%
CIOpESAT	5	14,29%
CIOpFO	4	11,43%
CIOpBE	3	8,57%
TOTAL	35	100,00%

Los 35 trabajos acumularon un total de 736 citas lo que da un promedio de 21 citas por trabajo. Discriminando por tipo de material, tenemos:

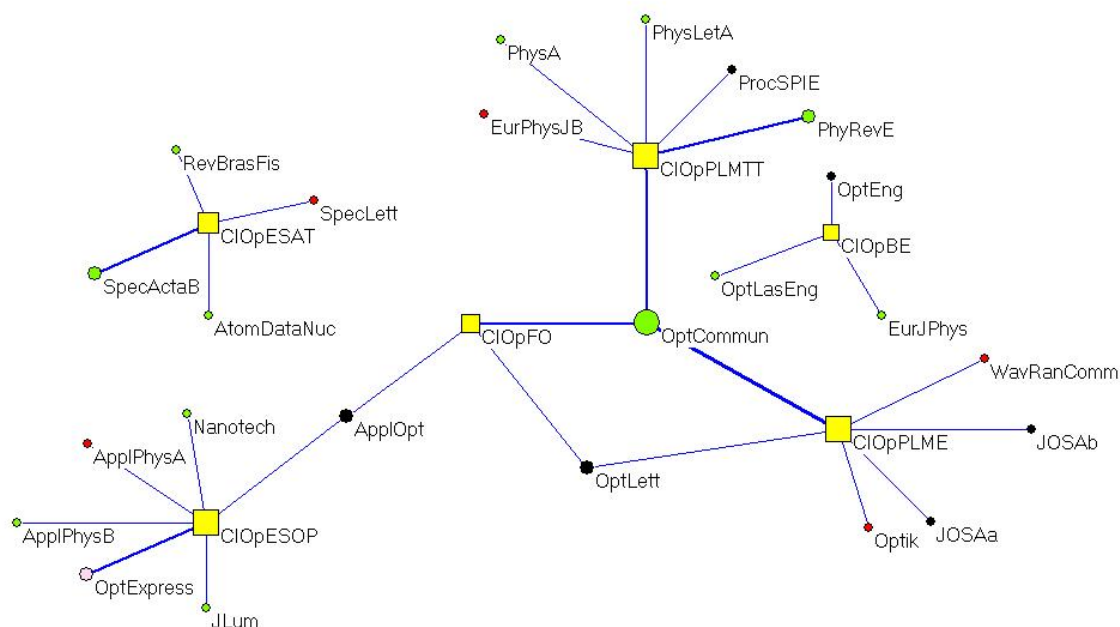
Revistas	638	86,68%
Libros	90	12,23%
Tesis	5	0,68%
Recursos Web	2	0,27%
Informes	1	0,14%

Del total de citas, un 21,20 % correspondió a autocitas, es decir referencias donde aparecía al menos 1 de los autores del trabajo en cuestión.

De las citas correspondientes a revistas, 135 son títulos diferentes, de los cuales, el 36,30% fue citado más de 1 vez.

⁶ Editor de gráficos vectoriales de código abierto. <http://www.inkscape.org/>

4.2. Títulos de revistas elegidos para publicar



En esta red se puede observar que 2 de los grupos son totalmente independientes en cuanto a las revistas elegidas para publicar. Espectroscopia Atómica (CIOpESAT) y Biospeckle (CIOpBE) no comparten ningún título con el resto. También se observa que la revista más elegida es Optics Communication y en menor medida Spectrochimica Acta B, Physics Review E, Optics Express, Applied Optics y Optics Letters. El grupo de Fibras Ópticas (CIOpFO) publica en revistas de óptica general y junto con el grupo de Biospeckle han tenido acceso a la totalidad de los títulos elegidos. Cabe aclarar que ambos son los grupos de menor producción.

5. Discusión

Si bien la aplicación de esta metodología permite obtener resultados clarificadores de una buena parte del consumo de información que realizan nuestros investigadores, también es necesario realizar algunas reflexiones para no perder de vista los alcances y limitaciones de este trabajo:

- En términos bibliotecarios el estudio de citaciones muestra un solo tipo de uso. Para tener más información se debería completar con el uso en sala y el registro de préstamo.
- La técnica de Análisis de Citas, tal como se lo aplicó en este estudio, tiene limitaciones derivadas, fundamentalmente, del hecho de que el autor no cita todo lo que usó. También es sabido que no todas las referencias usadas tienen el mismo valor. Está técnica basada en frecuencias tiende a homogeneizar y volver equivalentes a las unidades y los enlaces.
- La ventana temporal que se debería aplicar para que estos resultados tengan sentido en el desarrollo de colecciones también es digno de reflexión. La decisión de estimar las necesidades futuras mediante la observación en el uso de períodos relativamente cortos está apoyada en la consideración de que la probabilidad de uso tiende a cambiar con el tiempo. Por razones operativas en este trabajo se eligió un solo año, pero *a priori* se considera que es un periodo muy breve.
- La tarea de normalización de la información no se puede relegar. La calidad de la fuente de datos primaria es importante en términos de involucrar menor o mayor procesamiento. El contar con una fuente de datos que posea las referencias digitalizadas, como en nuestro caso las tiene SCOPUS, se percibe como la infraestructura básica imprescindible para estos tipos de estudio.
- Todo este trabajo tendrá sentido práctico para Argentina en la medida que, o bien todas las bibliotecas individualmente del sector de C y T lo incorporen en sus rutinas de evaluación e integren la información en una red nacional, o bien la Biblioteca Electrónica del MINCyT lo realice por áreas temáticas definidas. Se considera que es un buen camino para comenzar a generar información confiable y completa en pos de una política de desarrollo que involucró en el 2008 un presupuesto de U\$ 11.000.000.

6. Referencias

1. FUSSLER, H. H.; SIMON, J. L. Patterns in the Use of Books in Large Research Libraries. University of Chicago Library, 1961.
2. BLECIC, D. Measurements of journal use: an analysis of the correlations between three methods. *Bulletin of Medical Library Association*, 1999, vol. 87 (1), 20-25.
3. NISONGER, Th. E. A test of two citation checking techniques for evaluating political science collections in University Libraries. *Library Resources & Technical Services*, 1983, vol. 27, April/June, 163-176.
4. GROSS, P. L. K.; GROSS, E. M. College libraries and chemical education. *Science*, 1927, Oct.28, 385-389.
5. PAN, E. Journal citation as a predictor of journal usage in libraries. *Collection Management*, 1978, vol.2 (1), 29-38.
6. JOHNSON, B. Citation analysis of the Texas Tech University's statistics faculty: a study applied to collection development at the university library. *LIBRES: Library and Information Science Research*, 1996, vol.6 (3).
7. SYLVIA, M. J. Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies. *Collection Building*, 1998, vol.17 (1), 20-28.
8. SMITH, E. T. Assessing collection usefulness: an investigation of library ownership of the resources graduate students use. *College & Research Libraries*, 2003, vol. 64 (5), 344-355.
9. DINKINS, D. Circulation as assessment: collection development policies evaluated in terms of circulation at a small academic library. *College & Research Libraries*, 2003, vol. 64 (1), 46-53.
10. LEIDING, R. Using citation checking of undergraduate honors thesis bibliographies to evaluate library collections. *College & Research Libraries*, 2005, vol. 66 (5), 417-429.
11. URBANO SALIDO, C. El análisis de citas en trabajos de investigación como método para el estudio del uso de información en bibliotecas. *Anales de Documentación*, 2001, (4), 243-266.
12. BORNMANN, L.; DANIEL, H.D. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 2008, vol. 64 (1), 45-80.
13. VARGAS-QUESADA, B.; MOYA ANEGON, F. de. Visualizing the structure of science. Springer, 2007.

14. FREEMAN, L. C. The development of social network analysis: a study in the sociology of science. Vancouver, Empirical Press, 2004.
15. WASSERMAN, S.; FAUST, K. Social network analysis. New York, Cambridge University Press, 1994.
16. KAMADA, T.; KAWAI, S. An algorithm for drawing general undirected graphs. *Information processing letters*, 1989, vol. 31 (1), 7–15.