

VISUALIZACIÓN Y ANÁLISIS DE LA ESTRUCTURA DE LA BASE DE DATOS SCOPUS

ANTONIO J. GÓMEZ-NÚÑEZ

Universidad de Granada / Consejo Superior de Investigaciones Científicas - Unidad Asociada Grupo Scimago
anxusgo@gmail.com

BENJAMÍN VARGAS-QUESADA

Universidad de Granada / Consejo Superior de Investigaciones Científicas - Unidad Asociada Grupo Scimago
benjamin@ugr.es

TERESA MUÑOZ-ÉCIJA

Universidad de Granada / Consejo Superior de Investigaciones Científicas - Unidad Asociada Grupo Scimago
teresamunozecija@gmail.com

FÉLIX DE MOYA ANEGÓN

Universidad de Granada / Consejo Superior de Investigaciones Científicas - Unidad Asociada Grupo e Instituto de Políticas y Bienes Públicos
felix.demoya@cchs.csic.es

Resumen Introducción: La visualización de grandes redes de citación extraídas de bases de datos multidisciplinares como Web of Knowledge y Scopus es un tema de investigación recurrente en la investigación generada dentro de las ciencias de la información. La visualización de los elementos de la red y su agrupamiento en *clústeres* temáticos permite mapear la estructura de la investigación y la interrelación entre sus disciplinas, equiparables a los *clústeres* temáticos detectados.

Objetivos: Se pretende representar la estructura de Scopus en base a la extensa red de citación establecida entre las numerosas revistas Scopus incluidas en la plataforma Scimago Journal & Country Rank (SJR), que en nuestro estudio ascienden a 18891. Mediante técnicas de clustering y visualización, se procederá a la re-clasificación de las revistas.

Metodología: En base a la citación de trabajos, se obtuvieron listas de adyacencia agregadas a nivel de revistas para la *citación*, *co-citación* y *coupling*. Estas listas muestran parejas de revistas del SJR relacionadas mediante un valor numérico que expresa la fuerza de su relación. Las tres listas fueron integradas en una nueva resultante de su suma, y sus valores fueron normalizados mediante la *geo-similaridad*. Por último, se ejecutó el algoritmo de clustering de VOSviewer. Los *clústers* de revistas obtenidos se etiquetaron con las categorías originales del SJR junto con las palabras significativas más repetidas en los títulos.

Resultados y Discusión: El mapa resultante refleja la estructura de Scopus en función de un conjunto de categorías que representan el contenido temático de las revistas científicas incluidas en la base de datos. La reducción del conjunto de categorías en relación con el número inicial del SJR, así como el elevado número de cambios en la clasificación de las revistas sugiere un refinamiento y una optimización de la clasificación original.

Conclusiones: El *cienciograma* presentado constituye una representación fiable y precisa de la estructura de la investigación basada en revistas científicas, puesto que se fundamenta en la opinión de los expertos, reflejada por medio de sus citas.

Palabras-clave Clasificación. Visualización de información. Clustering.

Abstract Introduction: Visualization of big citation networks extracted from multidisciplinary databases as Web of Knowledge and Scopus is a recurrent topic in Library and Information Science research. Visualization and clustering of network items enable to map science and research structure on the basis of thematic clusters detected as well as their relations.

Objectives: We pretend to map Scopus database structure based on the extensive citation network derived from the full set of Scopus journals included in Scimago Journal & Country Rank (SJR) platform, which rise to 18891. We will re-classify the journals analysed using visualization and clustering techniques.

Method: Working from citation of papers we constructed three journal adjacency lists covering citation-based measures, namely, direct citation, co-citation and bibliographic coupling. These lists are showing journal couples related through a numeric value which express the strength of the relation. Then, the three lists were combined in a new one resulting from summing up their values which were later normalized through *geo-similarity* measure. Finally, VOSViewer clustering algorithm was executed and journal clusters obtained were labelled using original SJR category tags together with the most repeated significant words from journal titles. Results and Discussion: The resulting map reflects the Scopus structure through a set of categories that represents thematic content of scientific journals included in the database. The reduction of categories as well as the high number of shifts in journal classification originated from our method suggests a refinement and optimization of SJR journal original classification.

Conclusiones: The *sciencitogram* displayed arise like a reliable and accurate picture of science and research structure based on scientific journals, since it is built upon expert opinions, revealed by means of their citation patterns.

Keywords Classification. Information Visualization. Clustering.

Introducción

La visualización de información surge como una disciplina de enorme interés en el ámbito de la *Bibliometría* y de la *Cienciometría*, al proporcionar diferentes representaciones visuales y *cienciogramas* o mapas de la ciencia que facilitan el análisis de un dominio mostrando la estructura de la ciencia y de la investigación a través de las distintas disciplinas temáticas que la componen (elementos representados) junto con sus relaciones e interacciones (Moya-anegón et al., 2007). Principalmente, estos mapas se construyen a partir de la literatura científica compilada por las bases de datos, utilizando diferentes unidades de análisis (*papers*, revistas, categorías...) y distintas unidades de medida, tanto basadas en la citación y sus derivados (citación directa, co-citación, *coupling*...), como en el texto de las publicaciones. Pero además de mostrar la estructura disciplinar de la ciencia, permiten contemplar la evolución temporal de la investigación, detectar frentes de investigación, áreas de interdisciplinariedad, temas emergentes o en decadencia, etc.

Respecto a la bases de datos, en la actualidad destacan especialmente dos: Web of Knowledge (Wok) (Thomson Reuters, 2009) y Scopus (Elsevier, 2002), consideradas por la mayoría de la comunidad científica como las fuentes de información con una cobertura más exhaustiva y con mayor prestigio y reconocimiento a nivel internacional. Estas bases de datos, tienen carácter multidisciplinar y dan cabida a un elevado número de revistas científicas de prestigio de las que no sólo proporcionan información bibliográfica detallada, sino también índices de citas que permiten construir numerosos indicadores bibliométricos. Estos instrumentos, que pueden ser tanto cualitativos como cuantitativos, resultan de gran valor en tareas de evaluación de la ciencia y la investigación, y en especial, para los encargados de la toma de decisiones y el diseño de la política científica de los países.

Ahora bien, en el desarrollo y diseño de herramientas basadas en las publicaciones científicas albergadas en las bases de datos, conviene tener presente que la correcta clasificación de la literatura resulta de vital importancia para conseguir productos y resultados coherentes, fiables y sólidos. Por lo general, en la construcción de los *cienciogramas* subyacen procesos de asociación y distribución espacial de los ítems representados en función de su similaridad. La asociación puede calcularse, por ejemplo, en base a la co-ocurrencia de palabras significativas o al número de referencias bibliográficas compartidas. Utilizando técnicas estadísticas como el *clustering* o el *análisis factorial* es posible detectar grupos temáticos interrelacionados, que pueden interpretarse como un

reflejo de las diferentes disciplinas en las que puede descomponerse el conocimiento científico. En la actualidad, varias herramientas de visualización y análisis de redes, como Pajek (Batagelj & Mrvar, 1999) o VOSViewer (Eck & Waltman, 2010), integran diferentes algoritmos para la detección de *clústers* o *comunidades* dentro de una red, descomponiéndolas en grupos de ítems similares y fuertemente relacionados entre sí. Así, las herramientas de visualización aparecen también como una solución efectiva para la optimización y el refinamiento de la clasificación de la literatura en las bases de datos.

1 Clustering y visualización de información

El clustering emerge como una de las técnicas estadísticas más utilizadas en la clasificación e identificación de grupos temáticos. Son muchos los métodos de clustering que han sido utilizados con frecuencia por investigadores del ámbito de la visualización de la información con el fin de delinear la estructura del conocimiento científico y de la investigación. Para ello, resulta indispensable disponer de un esquema de clasificación consistente que represente de forma efectiva las diferentes disciplinas y/o subdisciplinas que integran la ciencia. Algunas propuestas significativas en el uso de clustering para la construcción de mapas de la ciencia (basados tanto en literatura de WoK como de Scopus) fueron recogidas y estudiadas en un trabajo de Klavans y Boyack (2009). Su propósito era desarrollar un mapa de la ciencia de consenso a partir de los mapas previamente analizados.

Numerosos investigadores han aplicado también algoritmos de clustering a matrices y redes de citación, co-citación y/o coupling de revistas. Chang y Chen (2011) proponen aplicar el método de *minimum span clustering* (MSC) a una matriz cuadrada de citación de aproximadamente 1.600 revistas del Social Science Citation Index (SSCI). Leydesdorff, Hammarfelt y Salah (2011) utilizaron el *algoritmo k-core* para representar las 25 categorías específicas del área de *Arts & Humanities Citation Index* y tratar de integrar su representación en un mapa global de la ciencia desarrollado previamente (Rafols, Porter, & Leydesdorff, 2010). Leydesdorff y Rafols (2012) publicaron un estudio en el que una matriz de citación compuesta de 9162 revistas extraídas del Science Citation Index Expanded de 2009 se utilizó para elaborar mapas interactivos. Entre los diferentes métodos llevados a cabo utilizaron diversos algoritmos de clustering para detectar grupos de revistas relacionadas e incluirlas en grupos temáticos bien definidos.

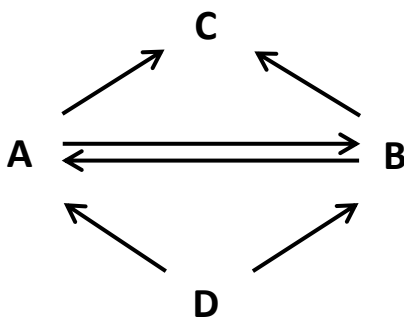
A lo largo de la cuantiosa literatura existente sobre clustering pueden encontrarse experimentos ejecutados a distintos niveles de agregación. A nivel de documento, Small (1999) desarrolló un mapa de la ciencia jerárquico mediante un método combinado de conteo fraccionalizado de los documentos citados, single-linkage clustering y ordenación bidimensional conforme a un proceso de triangulación geométrica. Moya-Anegón y otros (2004) propusieron la cocitación de categorías como unidad de análisis y representación, combinándola más adelante con técnicas de reducción del espacio como pathfinder networks (PFNET), y de identificación de grupos como análisis factorial (Vargas-Quesada & Moya-Anegón, 2007). Ahlgren y Colliander (2009) analizaron diferentes métodos de cálculo de similitud entre documentos en base a texto, coupling y combinación de ambos, así como varios métodos para representar y clasificar un pequeño conjunto de 43 documentos de la revista *Information Retrieval*. Para ello utilizaron *complete-linkage clustering* y compararon los clústers generados automáticamente con una clasificación previa realizada por expertos. De forma similar, Boyack y otros (2011) aplicaron una combinación de presentaciones gráficas y *average-link clustering* sobre varias matrices de similitud basadas en palabras significativas extraídas del título, resumen y descriptores del Medical Subject Headings (MeSH) de un total de más de 2 millones de artículos científicos extraídos de la base de datos Medline. Más recientemente Börner y otros (2012) presentaron una metodología para el diseño y actualización de un mapa de la ciencia y un sistema de clasificación elaborado para la Universidad de California, San Diego (UCSD) utilizando técnicas de clustering sobre matrices de similitud de revistas de Web of Science (WoS) y Scopus.

Por lo general, la ejecución de procedimientos de clustering sobre grandes redes y matrices de datos implica complejas operaciones de cálculo además de altos requerimientos de hardware y software. Además, la visualización de los datos debe ser lo más óptima y clara para su comprensión y manipulación. Programas como Pajek o VOSViewer surgen como excelentes herramientas para el análisis y la visualización de grandes masas de datos que, además, integran algoritmos de clustering para clasificar los datos analizados. En este trabajo, se utilizan ambas herramientas para llevar a cabo la representación de la estructura de la base de datos Scopus, utilizando Pajek para la preparación previa de los datos y VOSViewer para ejecutar la visualización final. Los detalles del proceso son pormenorizados en la siguiente sección de este trabajo.

2 Material y Método

En primer lugar, se diseñó una ventana temporal de dos años correspondiente a 2009 y 2010 y se recuperaron los datos de citación de un total de 18891 revistas de la plataforma Scimago Journal & Country Rank (SJR) (Scimago Lab, 2007). Esta plataforma aglutina todas las revistas contenidas en la base de datos Scopus y permite elaborar indicadores para la producción de análisis de dominio y rankings tanto de revistas como de países. Para este conjunto de datos sólo se contaron las citas incluidas en el período temporal comprendido entre 2000 y 2010. Las citas se calcularon a nivel de artículos y posteriormente fueron agrupadas por revistas. A partir de estos datos se construyeron tres listas de adyacencia, compuestas por parejas de revistas y un valor numérico que expresa la fuerza de su relación, representando medidas basadas en la citación, como son la citación directa, la co-citación y el coupling.

Por último, estas tres listas fueron integradas en una lista final mediante la suma de las tres medidas de citación mencionadas y siguiendo para ello el modelo propuesto por Persson (2010), que él mismo denominó *Weighted Direct Citation (WDC)*. Conforme a su trabajo, se presenta a continuación el diagrama utilizado por Persson para calcular el WDC. No obstante, se ha introducido una sensible modificación con respecto al original, al representar en nuestro caso los dos sentidos en que puede considerarse la citación directa.



Siguiendo el diagrama, la fórmula para la integración de las tres medidas de citación sería la siguiente:

$$c_{ij} = ABC + DAB + \max(AB, BA)$$

Donde ABC hace referencia al coupling, DAB a la co-citación, y AB o BA a la citación directa.

Seguidamente, se normalizaron los resultados de la red utilizando la *geo-similaridad*, medida cercana al coseno de Salton y que funciona dividiendo los elementos de la matriz por la media

geométrica de ambos elementos de la diagonal (Batagelj & Mrvar, 1999). Se detalla a continuación, la fórmula utilizada para su cálculo:

$$s_{ij} = c_{ij} / \sqrt{c_i * c_j}, c_i = \sum\{j: j \neq i: c_{ij}\}$$

La siguiente fase de nuestro método se corresponde con la ejecución del algoritmo de clustering sobre la red normalizada de revistas que integra las tres medidas basadas en la citación. VOSViewer permite llevar a cabo no sólo la visualización de información, sino también ejecutar un algoritmo de clustering que permite establecer una clasificación de los datos que posteriormente serán mapeados. En palabras de sus creadores (Waltman, Eck, & Noyons, 2010) este algoritmo incluye un parámetro de resolución capaz de detectar grupos o clústeres de pequeño tamaño si se proporciona un valor adecuado para configurarlo. También indican que un mayor parámetro de resolución implica un incremento paralelo del número de clústeres generados. Teniendo en cuenta estas consideraciones, decidimos ejecutar varias pruebas introduciendo distintos valores en el parámetro de resolución del algoritmo. De esa forma, podríamos obtener diferentes soluciones ofreciendo distintas descomposiciones (subredes) de la red de revistas y, por lo tanto, produciendo diferentes conjuntos de clústeres o comunidades temáticas. Nuestro objetivo final fue obtener un sistema de clasificación consistente para representar de forma eficiente las diferentes disciplinas de la ciencia y la investigación a partir de la literatura científica compilada por Scopus.

La Figura 1 muestra la correlación existente entre el valor del parámetro de resolución y el número de comunidades o clústeres proporcionados por el algoritmo de VOSViewer a lo largo de las diferentes pruebas ejecutadas. Atendiendo a nuestros propósitos de visualización, una solución aportando alrededor de 250-300 clústeres resultaría efectiva para representar la estructura de la ciencia eficazmente. Además, se estimó que con independencia de la solución final escogida, el tamaño mínimo de clúster no podría ser inferior a 10 revistas, garantizando de esta forma una serie de grupos temáticos con un mínimo de consistencia y delimitación. Una vez analizados los resultados de las diferentes pruebas, se consideró que el parámetro de resolución con valor 15, aportando 270 clústeres útiles con más de 10 revistas, resultaba ser una solución óptima para la elaboración del *cienciograma* que representaría la estructura de Scopus.

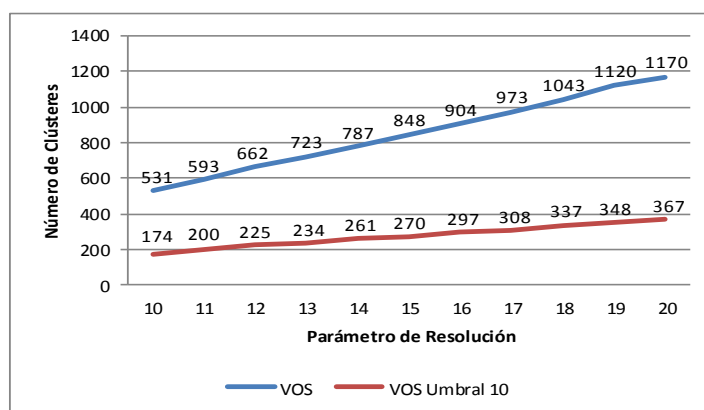


Figura 1: Número de clústeres por parámetro de resolución con y sin umbral 10

En la clasificación original de la plataforma SJR, sus 18891 revistas se distribuyen entre un total de 304 categorías temáticas, con un número medio de revistas por clúster de 62,14. Nuestra propuesta, en base al algoritmo de VOSViewer con parámetro de resolución 15 y un tamaño mínimo de clúster de 10 revistas, arrojó un número medio de revistas que asciende a 65,66 y que, por lo tanto, se antoja bastante similar a la distribución de revistas por categorías de la clasificación

original del SJR. Las distribuciones referentes al número de revistas clasificadas y al número medio de revistas por clúster a lo largo de los distintos parámetros de resolución configurados pueden observarse en las Figuras 2 y 3 respectivamente.

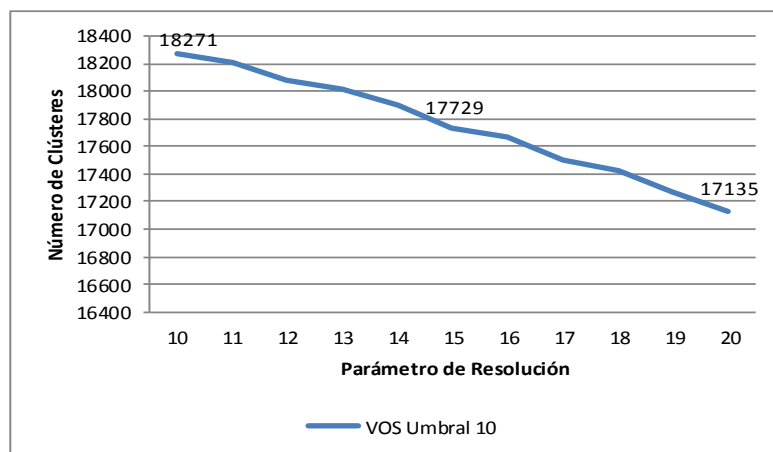


Figura 2: Número total de revistas clasificadas

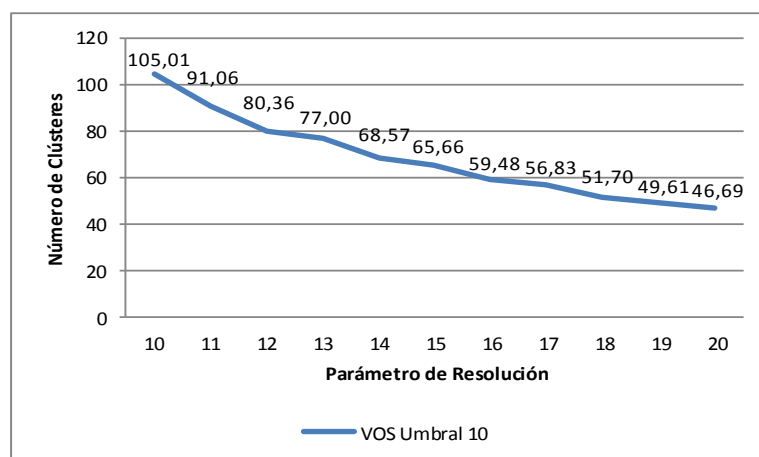


Figura 3: Número medio de revistas por clúster

Al mismo tiempo que el número de clústeres generados por el algoritmo se incrementa al aumentar el valor del parámetro de resolución, si aplicamos el umbral de 10 revistas como tamaño mínimo de clúster, comprobamos que se incrementa también el número de clústeres que no superan dicho umbral y que, por consiguiente, no son útiles para nuestros propósitos. En otras palabras, existe correlación positiva entre el incremento del valor del parámetro y el aumento del número de clústeres tanto que superan el umbral como que no. En relación con el número de revistas asignadas, tal como afirman los creadores de VOSViewer, todas las revistas son asignadas a alguna comunidad o grupo durante el procedimiento de clustering. No obstante, si comparamos el incremento del parámetro de resolución con el total de revistas asignadas a clústeres que superen el umbral de 10 como tamaño mínimo, entonces encontraremos que existe una correlación negativa entre ambas variables.

El último paso de nuestro método está relacionado con el etiquetado de los diferentes clústeres o comunidades detectadas por el algoritmo de VOSViewer. Para tal fin nos propusimos reutilizar los nombres o etiquetas de las categorías originales del SJR, asignando a cada clúster el nombre de las etiquetas más repetidas y derivadas de las revistas incluidas. La frecuencia de aparición de dichas etiquetas fueron transformadas en porcentajes y en pesos *tf-idf* (Salton & Buckley, 1988). A continuación, las categorías fueron ordenadas conforme a los pesos *tf-idf* y se seleccionaron sólo aquellas que representaban al menos un 33% del conjunto total de categorías

citadas por las revistas incluidas en cada clúster para su delimitación temática. En el proceso de etiquetado se desearon todas las categorías originales marcadas como *Misceláneas* o *Multidisciplinar*. Este hecho provocó que algunos clústeres tuvieran que ser etiquetados a posteriori, utilizando para ello una combinación de etiquetas de las categorías SJR originales junto con términos significativos extraídos del título de las revistas agrupadas en los clústeres.

Antes de continuar, conviene destacar dos asuntos importantes derivados del proceso de etiquetado desarrollado en nuestro método. En primer lugar, el uso de etiquetas de categorías prediseñadas dio lugar a la aparición de varios clústeres nombrados con exactamente las mismas etiquetas. Estos clústeres fueron fusionados en otro clúster nuevo, al considerarse que sus revistas abarcaban temas análogos y colindantes. Como consecuencia, el esquema temático inicial generado en base al algoritmo de VOSViewer derivó en un nuevo esquema temático basado en las etiquetas de las categorías y compuesto por un total de 219 categorías temáticas. En segundo lugar, habría que resaltar que la multi-asignación de revistas no proviene del funcionamiento del algoritmo propiamente dicho, sino que es una consecuencia de nuestro proceso de etiquetado, que permite asignar una revista a más de una categoría provocando un solapamiento en determinadas categorías temáticas. No obstante, esta multi-asignación no resulta demasiado elevada. Así, si observamos la Figura 4, descubrimos que alrededor de un 60% de las revistas fueron asignadas a una sola categoría temática, mientras que cerca de un 30% se asignaron a dos categorías, casi un 7% a tres categorías y un valor casi residual de revistas fueron asignadas a cuatro categorías.

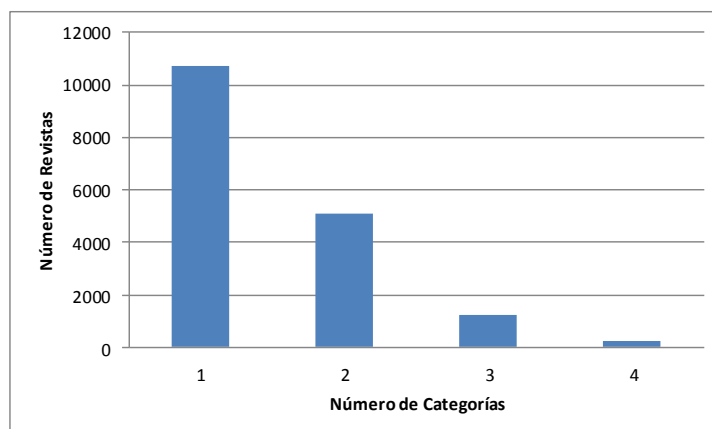


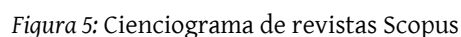
Figura 4: Asignación de revistas a categorías

3 Resultados y Discusión

La clasificación final del conjunto de revistas analizadas y sometidas al proceso de clustering comentado, puede consultarse en la siguiente dirección web:

http://www.scimago.es/benjamin/vos15_classification.pdf

Desde el punto de la visualización de la información a continuación mostramos y analizamos de forma breve, por las limitaciones lógicas de espacio, el cienciograma completo que representa la clasificación general obtenida, así como el de otras disciplinas que se pueden distinguir a simple vista.



Grosso modo, si nos fijamos en el cuerno que aparece en la parte inferior del cienciograma, de color marrón claro, observaremos como es en ese sitio donde se agrupan las revistas de Matemáticas. Siguiendo el sentido de las agujas del reloj, y por tanto a su izquierda, podemos observar de color rosa y verde a distintos tipos de Ingenierías, la Física de color amarillo, justo por encima de ella, y sobre esta última a la Agricultura. En el cuerno superior, y justo encima de la Agricultura, encontramos las Neurociencias, la Psicología, la Medicina y la Biología. Desde esta posición y siguiendo de nuevo el sentido de las agujas del reloj, podemos detectar la Sociología, la Lingüística, la Historia y la Literatura (extremo derecho del cienciograma). Siguiendo el cienciograma hasta nuestro punto de origen detectamos la Economía, la Documentación, y la Ciencias de la Computación, que conectan con las Matemáticas.

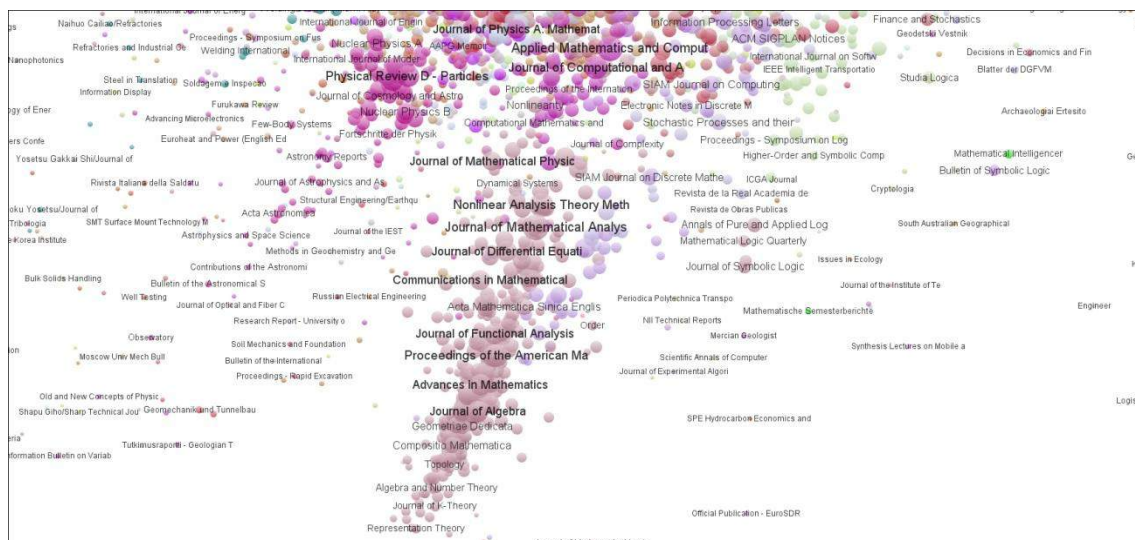


Figura 6: Cienciograma de revistas de Matemáticas de Scopus

La Figura 6 muestra claramente cómo se agrupan y se estructuran las revistas de Matemáticas en base a sus relaciones de citación. Si se observa de forma detenida, se puede ver como en la parte inferior aparecen las revistas de Matemáticas básicas y cómo, a medida que se asciende en el cienciograma, empiezan a aparecer las revistas de Matemáticas aplicadas, llegando a mezclarse con las de Ciencias de la Tierra, Física y Astronomía en la zona superior izquierda, y con las de Ciencias de la Computación en el área superior derecha.

Otras disciplinas, como ocurre con Library & Information Sciences (LIS), representada por revistas como *JASIST* o *Scientometrics* que se sitúan en el centro de la Figura 7 en color verde, no se muestran tan cohesionadas ni forman un clúster tan bien definido como sucedía con las Matemáticas en la figura anterior.

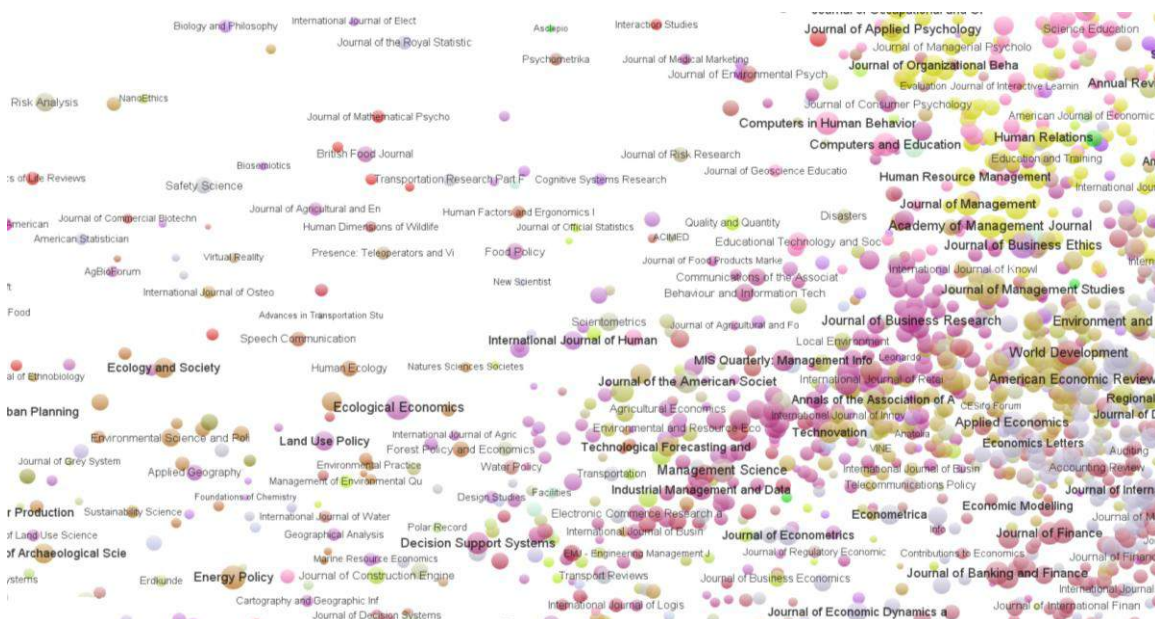


Figura 7: Cienciograma de revistas de Library & Information Science de Scopus

Esta dispersión de las revistas de LIS, al igual que ocurre con muchas otras disciplinas, se debe fundamentalmente a que muchas de estas revistas son interdisciplinarias y acaban publicando

contenidos que pertenecen a diferentes disciplinas científicas del mismo área. Desde el punto de vista de la visualización de la información, esto hace que las revistas de esta disciplina aparezcan cerca y, por tanto, mezcladas con otras de Ciencias de la Computación, Gestión, Planificación, etc. Este efecto disgregador de las revistas de una disciplina se ve además aumentado por la necesidad y obligatoriedad de tener que representar los cienciogramas en dos dimensiones (2D), puesto que su fin último es ser visualizados mediante una pantalla de ordenador de forma estática, o a través de un soporte como el papel. Esto provoca que la dimensión profundidad (Z) desaparezca, y que disciplinas que claramente están separadas, es decir, lejos de otras en un cienciograma en tres dimensiones, acaben estando juntas e incluso mezcladas, en otro de 2D.

Conclusiones

La visualización de la información es una herramienta muy potente para el análisis y corroboración de resultados en favor de la clasificación. No obstante, por si sola, no puede ni debe ser utilizada como única herramienta, pues la multidisciplinariedad e interdisciplinariedad de las unidades que se representan, en combinación con las limitaciones propias de un espacio de 2 dimensiones (papel o pantalla de ordenador) provocaran un falseamiento de los resultados. Por ello, al igual que hacemos en este trabajo, recomendamos generar esquemas de clasificación tradicionales apoyados y validados mediante técnicas de visualización de información.

La propuesta aquí realizada facilita la creación de un nuevo esquema de clasificación equilibrado en cuanto a la distribución de revistas por categorías, número de categorías útiles, y concentración moderada de revistas en las categorías o grupos temáticos con mayor poder de atracción. Así, en un trabajo desarrollado anteriormente (Gómez-Núñez, Vargas-Quesada, Moya-Anegón, & Glänzel, 2011) basado en un proceso iterativo de análisis de referencias bibliográficas citadas por las revistas del SJR, tan sólo cinco categorías resultaron suficientes para aglutinar un 25% de las 14166 revistas clasificadas. Con la propuesta de clasificación que presentamos aquí, se necesitan 18 categorías para alcanzar algo más de ese 25% de revistas clasificadas. Teniendo en cuenta que el número de revistas ahora es bastante más elevado (17729), este asunto adquiere aún mayor importancia, puesto que ese amplio margen podría favorecer todavía más las concentraciones de revistas en categorías con mayor poder de atracción.

Referencias

- AHLGREN, P.; COLLIANDER, C. (2009). Document-document similarity approaches and science mapping: Experimental comparison of five approaches. *Journal of Informetrics*, 3(1), 49-63. doi:10.1016/j.joi.2008.11.003.
- BATAGELJ, V.; MRVAR, A. (1999). *Pajek - Program for Large Network Analysis*. Retrieved from <http://pajek.imfm.si/doku.php>.
- BÖRNER, K.; KLAVANS, R.; PATEK, M.; ZOISS, A. M.; BIBERSTINE, J. R., et al. (2012). Design and update of a classification system: the UCSD map of science. *PloS one*, 7(7), e39464. doi:10.1371/journal.pone.0039464.
- BOYACK, K. W.; NEWMAN, D.; DUHON, R. J.; KLAVANS, R.; PATEK, M.; BIBERSTINE, J. R.; SCHIJVENAARS, B.; et al. (2011). Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PloS one*, 6(3), e18029. doi:10.1371/journal.pone.0018029.

- CHANG, Y. F.; CHEN, C. (2011). *Classification and Visualization of the Social Science Network by the Minimum Span Clustering Method*, 62(12), 2404–2413. doi:10.1002/asi.
- ECK, N. J. Van; WALTMAN, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. doi:10.1007/s11192-009-0146-3.
- ECK, N. J. Van; WALTMAN, L.; DEKKER, R.; BERG, J. Van Den. (2010). *A Comparison of Two Techniques for Bibliometric Mapping : Multidimensional Scaling and VOS*, 61(12), 2405–2416. doi:10.1002/asi.
- ELSEVIER. (2002). *Scopus*. Retrieved April 12, 2013, from <http://www.scopus.com/home.url>.
- GÓMEZ-NÚÑEZ, A. J.; VARGAS-QUESADA, B.; MOYA-ANEGÓN, F.; GLÄNZEL, W. (2011). Improving SCImago Journal & Country Rank (SJCR) subject classification through reference analysis. *Scientometrics*, 89(3), 741–758. doi:10.1007/s11192-011-0485-8.
- KLAVANS, R.; BOYACK, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3), 455–476. doi:10.1002/asi.20991.
- LEYDESDORFF, L.; HAMMARFELT, B.; SALAH, A. (2011). *The Structure of the Arts & Humanities Citation Index : A Mapping on the Basis of Aggregated Citations Among 1 , 157 Journals*, 62(12), 2414–2426. doi:10.1002/asi.
- LEYDESDORFF, L.; RAFOLS, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Informetrics*, 6(2), 318–332. doi:10.1016/j.joi.2011.11.003.
- MOYA-ANEGÓN, F.; VARGAS-QUESADA, B.; CHINCHILLA-RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; MUNOZ-FERNÁNDEZ, F. J.; HERRERO-SOLANA, V. (2007). *Visualizing the Marrow of Science*, 58(14), 2167–2179. doi:10.1002/asi.
- MOYA-ANEGÓN, F.; VARGAS-QUESADA, B.; HERRERO-SOLANA, V.; CHINCHILLA-RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; MUNOZ-FERNÁNDEZ, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61(1), 129–145.
- PERSSON, O. (2010). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, 4(3), 415–422. doi:10.1016/j.joi.2010.03.006.
- RAFOLS, I.; PORTER, A. L.; LEYDESDORFF, L. (2010). Science Overlay Maps : A New Tool for Research Policy and Library Management. *Journal of the American Society for Information Science and Technology*, 61(9), 1871–1887. doi:10.1002/asi.

- SALTON, G.; BUCKLEY, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
- SCIMAGO LAB. (2007). *Scimago Journal & Country Rank (SJR)*. Retrieved April 15, 2011, from <http://www.scimagojr.com/>.
- SMALL, H. (1999). *Visualizing Science by Citation Mapping*, 50(1973), 799–813.
- THOMSON REUTERS. (2009). *ISI Web of Knowledge*. Retrieved April 12, 2013, from <http://wokinfo.com/>.
- VARGAS-QUESADA, B.; MOYA-ANEGÓN, F. (2007). *Visualizing the structure of science*. New York: Springer.
- WALTMAN, L.; ECK, N. J. Van; NOYONS, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. doi:10.1016/j.joi.2010.07.002.