

Science Mapping Artificial Intelligence

Benjamín Vargas-Quesada
Departamento de Información y
Comunicación
Universidad de Granada.
Granada, Spain
e-mail: benjamin@ugr.es

Zaida Chinchilla Rodríguez
Instituto de Políticas y Bienes
Públicos (IPP). Consejo Superior de
Investigaciones Científicas (CSIC).
Madrid, Spain
e-mail: zaida.chinchilla@cchs.csic.es

Antonio Perianes-Rodríguez
Departamento de Biblioteconomía y
Documentación.
Universidad Carlos III
Madrid, Spain
e-mail: antonio.perianes@uc3m.es

Abstract— Mediante *science mapping*, se muestra la visualización de la estructura intelectual de la inteligencia artificial, a nivel mundial. Se identifican sus principales líneas de investigación a partir de las citas de sus trabajos, como interactúan entre sí, mediante la detección de los nodos con un alto grado de centralidad de *betweenness* o intermediación, así como el nivel de actividad, de cada una de ellas, por medio del uso de un indicador denominado *citation burst*. Se finaliza con un estudio longitudinal y evolutivo de las líneas detectadas.

Keywords— *Science Mapping; Inteligencia Artificial; Líneas de Investigación.*

I. INTRODUCCIÓN

Hace más de 50 años, que Moreno [1] propuso el uso de las redes sociales para analizar el comportamiento humano y detectar patrones e individuos relevantes dentro de las mismas. Hoy día, el concepto de red social ha cambiado, trasladándose la representación gráfica de las interacciones entre los individuos o entidades (grafos), a denominar a las compañías que ofrecen servicios de redes y medios sociales.

La aparición del *big data* ha permitido que grandes conjuntos de datos, que hasta hace poco, eran imposibles de procesar mediante el uso de herramientas tradicionales debido a su tamaño y complejidad, puedan ahora ser tratados y utilizados con fines analíticos y de predicción. La combinación del *big data*, *datamining*, bibliometría y redes sociales, ha dado lugar a lo que se ha denominado como *science mapping* [2], cuyo objetivo es visualizar la estructura y relaciones de la ciencia en general, y de las áreas científicas, disciplinas, líneas de investigación, autorías, etc., en particular. *Science mapping* ha sido utilizado para mostrar la columna vertebral de la ciencia a nivel mundial [3], representar y analizar la estructura de grandes bases de datos científicas [4] y, visualizar la estructura intelectual del grafeno, sus principales líneas de investigación y evolución mundial [5], entre otras importantes iniciativas. Hasta donde llega nuestro conocimiento, no hay ningún trabajo que visualice la estructura intelectual de la IA, ni que detecte e

identifique sus principales líneas de investigación, a partir de ella. Este trabajo persigue los siguientes objetivos:

- Visualizar y analizar la estructura de la red de la Inteligencia Artificial (IA).
- Identificar las principales líneas de investigación.
- Localizar los trabajos que actúan como conectores de las líneas de investigación.
- Detectar tendencias emergentes de investigación.
- Realizar un breve estudio longitudinal de las principales líneas de investigación.

II. MATERIAL Y MÉTODOS

Por medio de la consulta: $WC=(Computer\ Science, Artificial\ Intelligence)$ AND $TS=(Artificial\ Intelligence)$, filtrando por *papers, proceedings paper o reviews*, el 28 de mayo de 2018, descargamos de Web of Science 8.635 documentos publicados entre 1971 y 2017, conteniendo un total de 172.109 referencias. Para la visualización y análisis de los datos, hemos utilizado *CiteSpace*, que es un *software* gratuito diseñado específicamente para este tipo de tareas. En aras a nuestro objetivo principal, utilizamos las relaciones de citación de los documentos descargados, para establecer la estructura intelectual e identificar los clusters o líneas de investigación. Cada línea es etiquetada en base a los *keywords* de los artículos que las citan, como se indica en la Figura 1.

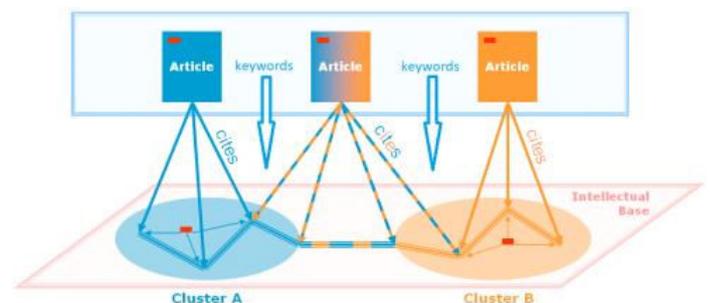


Fig. 1. Proceso para la detección y etiquetado de líneas de investigación

El uso de las citas, al igual que otras unidades de análisis de corte bibliométrico como son los autores, términos

extraídos de títulos y/o resúmenes, keywords, etc., podría ser considerado como una limitación, por las distorsiones que algunas de sus características inherentes como la intencionalidad, cobertura, popularidad, etc., pueden causar. Para salvar esta limitación, en cada visualización, se han establecido periodos de 5 años, y sólo se han representado los 50 documentos más citados de cada periodo. De esta forma se reduce en gran medida las distorsiones que se puedan producir, al representar la opinión consensuada de los autores, mediante sus citas.

III. RESULTADOS

Las visualizaciones y resultados aquí mostrados, pueden ser replicados mediante el programa *Citespace*, disponible en: <http://cluster.cis.drexel.edu/~cchen/citespace>, en combinación con los registros descargados de Web of Science, accesibles

en:

<https://www.dropbox.com/sh/5i0nb0npcsavcg4/AAA4uGb05fsCLN-L-18jekZca?dl=0>.

A. Estructura de la Red de IA

La red de IA a nivel mundial, cuenta con 382 nodos, 805 enlaces y una densidad de 0.011. El componente de mayor tamaño, Figura 2, está compuesto por un total de 287 nodos (75% de la red total). Mediante el método de Ward [6], se han identificado 8 clusters o líneas principales de investigación. La modularidad Q [7] de la red, es de 0.8066, indicando que está dividida en líneas de investigación altamente enlazadas y relacionadas entre sí. Su media Silhouette [8] es de 0.3387, sugiriendo, que la homogeneidad de dichas líneas es más bien baja.



Fig. 2. Science map mundial de IA y principales líneas de investigación

B. Principales Líneas de Investigación

Para su denominación, hemos recurrido a los *keywords* de los artículos fuente de cada línea, seleccionando el más representativo mediante el algoritmo likelihood ratio, por sus buenos resultados en cobertura y unicidad. Por orden de tamaño, las principales líneas de investigación son las siguientes: *Natural intelligence*, *Computational and artificial intelligence*, *Function aproximation*, *Case-based reasoning*, *Distributed artificial intelligence*, *Support vector machines*, *Heuristics*, y *Firefly algorithms*, como se puede ver en la Figura 2, precedidas por el símbolo #.

C. Trabajos Responsables de Conectar las Líneas de Investigación

La centralidad de intermediación (*betweenness*) de un nodo en una red, mide el grado en que el nodo es parte de *paths* que conectan otros nodos [9]. En este caso, con valores comprendidos entre 0 y 1, la utilizamos para detectar las referencias responsables de conectar distintas líneas de investigación. Las referencias con un *betweenness* alto, se muestran con un borde morado. Cuanto más grueso es el borde, mayor es su valor de intermediación, ver Figura 2. Por ejemplo, el trabajo de Russell SJ, 2003, *Artificial Intelligence*, tiene una centralidad de 0.37, la más alta de toda la red, y es el responsable de la conexión de las líneas de investigación *Natural intelligence*, *Computational and artificial intelligence*, *Case-base reasoning* y *Support vector machines*. Este análisis es extensible al resto de nodos de color rojo, en función de su tamaño.

D. Detectar tendencias emergentes de investigación

Citation burst (explosión de citas) [10], es un indicador utilizado para la detección de referencias y líneas de investigación activas, que reciben un elevado número de citas durante uno o varios años. La explosión de citas, proporciona la evidencia de que un trabajo concreto ha atraído un alto grado de atención por parte de sus colegas, reflejándolo

mediante sus citas, Figura 3. Si una línea de investigación contiene numerosas referencias de este tipo, se puede considerar que es muy activa y con una tendencia emergente. Si un nodo presenta un círculo rojo en su interior, Figura 4, indicará que atrae un gran número de citas. A mayor *citation burst*, mayor tamaño del círculo.

Top 25 References with the Strongest Citation Bursts

References	Year	Strength	Begin	End
KOLOUDNER J, 1993, CASE BASED REASONING, V0, P0	1993	11.5028	1996	2001
PEARL J, 1988, PROBABILISTIC REASON, V0, P0	1988	11.2276	1991	1996
GOLDBERG DE, 1989, GENETIC ALGORITHMS S, V0, P0	1989	10.5589	1991	1997
BROOKS RA, 1991, ARTIF INTELL, V47, P139, DOI	1991	8.8768	1995	1998
WOOLDRIDGE M, 1995, KNOWL ENG REV, V10, P115, DOI	1995	8.3438	1997	2003
BOND AH, 1988, READINGS DISTRIBUTED, V0, P0	1988	7.6561	1991	1995
WINOGRAD T, 1986, UNDERSTANDING COMPUT, V0, P0	1986	5.7848	1991	1994
WATERMAN D, 1986, GUIDE EXPERT SYSTEMS, V0, P0	1986	5.7848	1991	1994
RUMELHART DE, 1986, PARALLEL DISTRIBUTED, V1, P0	1986	5.2037	1991	1994
MINTON S, 1992, ARTIF INTELL, V58, P161, DOI	1992	4.9822	1997	2000
HAYESROTH B, 1985, ARTIF INTELL, V26, P251, DOI	1985	4.7946	1991	1993
HARNAD S, 1990, PHYSICA D, V42, P335, DOI	1990	4.6658	1993	1998
REITER R, 1987, ARTIF INTELL, V32, P57, DOI	1987	4.5838	1992	1995
NORVIG P, 1995, ARTIFICIAL INTELLIGE, V0, P0	1995	4.4763	1996	2003
GENESERETH MR, 1987, LOGICAL F ARTIFICIAL, V0, P0	1987	4.3618	1991	1995
RICH E, 1991, ARTIFICIAL INTELLIGE, V0, P0	1991	4.3304	1993	1999
KUIPERS B, 1986, ARTIF INTELL, V29, P289, DOI	1986	4.2375	1992	1994
RIESBECK CK, 1989, INSIDE CASE BASED RE, V0, P0	1989	4.2184	1992	1997
NEWELL A, 1990, UNIFIED THEORIES COG, V0, P0	1990	4.2055	1993	1995
GASSER L, 1991, ARTIF INTELL, V47, P107, DOI	1991	4.0773	1992	1999
DECHTER R, 1991, ARTIF INTELL, V49, P61, DOI	1991	4.0773	1992	1999
DEKLEER J, 1986, ARTIF INTELL, V28, P127, DOI	1986	4.0433	1991	1994
WELLMAN MP, 1993, J ARTIF INTELL RES, V1, P1	1993	4.0285	1996	1999
GASSER L, 1989, DISTRIBUTED ARTIFICI, V2, P0	1989	3.8146	1991	1995
SUCHMAN LA, 1987, PLANS SITUATED ACTIO, V0, P0	1987	3.8146	1991	1995

Fig. 3. Top 25 trabajos con mayor citation burst

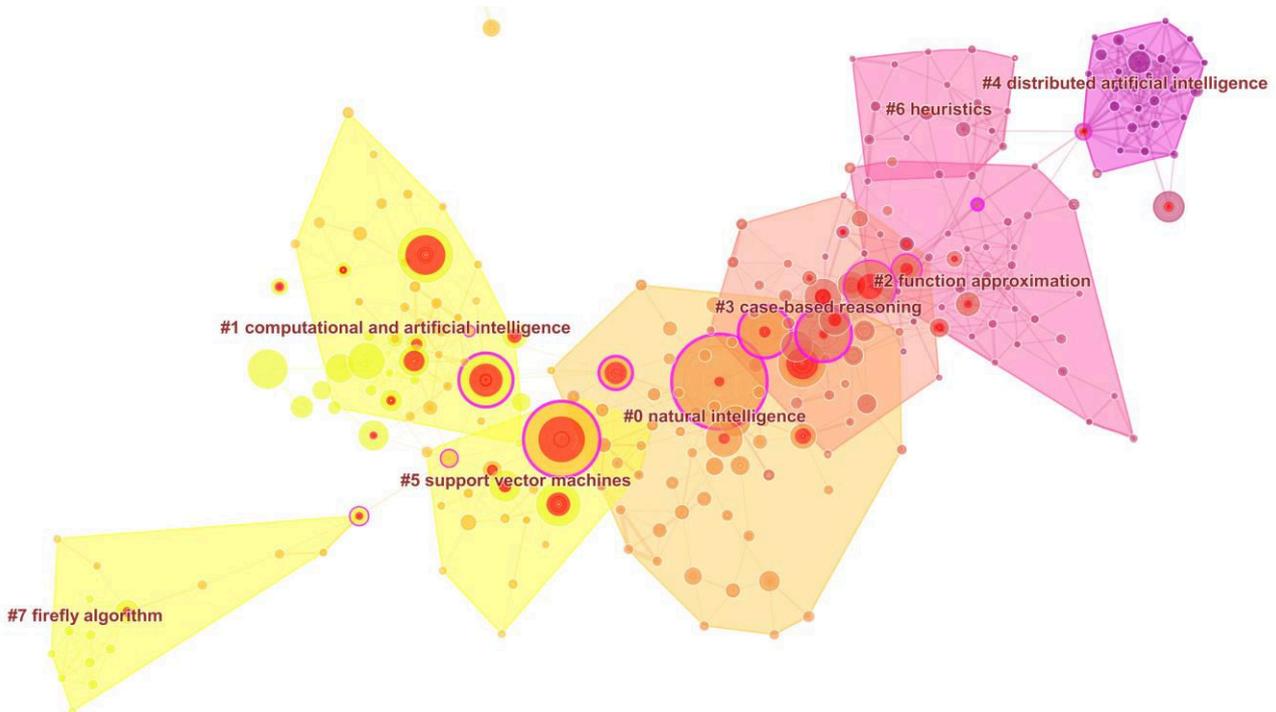


Fig. 4. Representación de los Top 25 trabajos con mayor citation burst

Por ejemplo, la línea de investigación 3 *Case-based reasoning*, contiene bastantes trabajos con un nivel de *citation burst* elevado, mostrando su alto nivel de actividad y emergencia. Así, el trabajo de Kolodner J, 1993, Case Based Reasoning, con un valor de 11,5028, seguido muy de cerca del de Pearl J, 1988, Probabilistic Reason, con 11,2276 de *citation burst*, respectivamente, ponen de manifiesto el amplio uso de estos trabajos y la relevancia de esa la línea de investigación.

E. Breve estudio longitudinal de las principales líneas de investigación etectar tendencias emergentes de investigación

Timeline view, Figura 5, es un tipo de representación que proporciona una visualización de la red por medio de líneas de tiempo horizontales. En este caso, se ha dividido en periodos

de cuatro años y la etiqueta de cada línea de investigación, aparece al final de la línea temporal, a la derecha. Las referencias citadas, se representan con círculos con un tamaño proporcional a la cantidad de citas recibidas en ese segmento de tiempo. Por lo tanto, un gran tamaño denota una referencia altamente citada. El color de cada referencia se corresponde con el segmento de tiempo en el que fueron citadas, pero se posicionan en base al año en que fueron publicadas. Además, se utilizan otro tipo de colores, explicados anteriormente, para resaltar atributos especiales como el *betweenness* o el *citation burst*. En este caso, las líneas que conectan dos referencias en la visualización, representan enlaces de cocitación, e indican la relación entre ellas, así como entre las líneas de investigación a las que pertenecen.

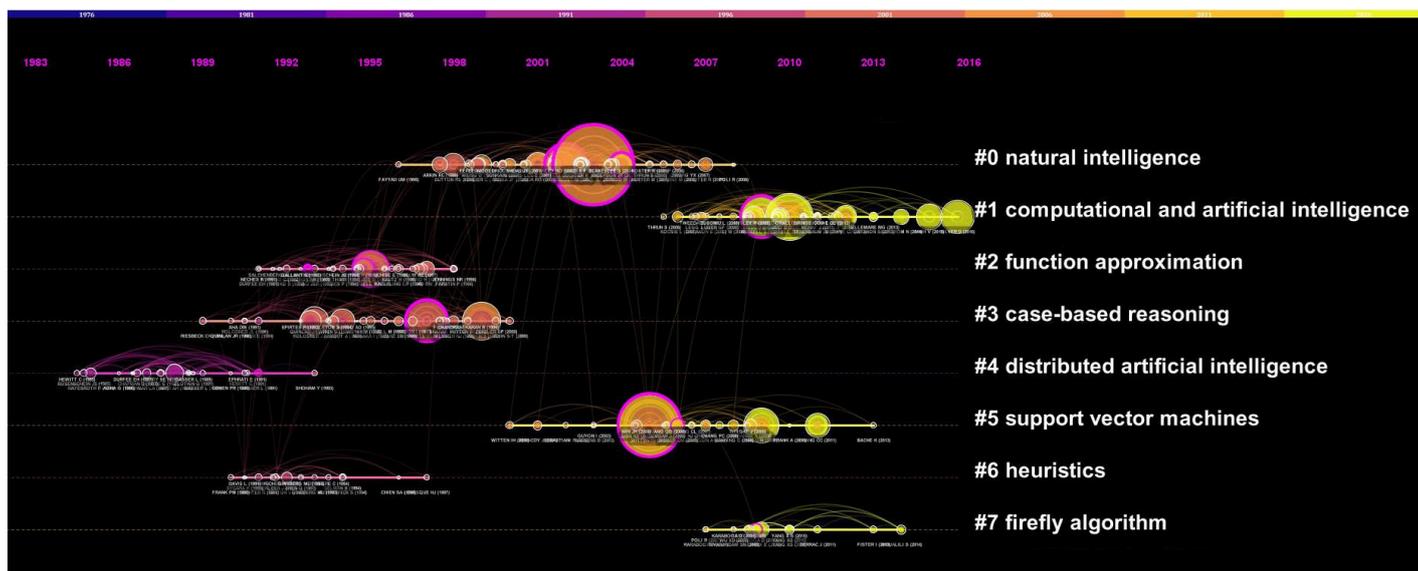


Fig. 5. *Timeline view* de las principales líneas de investigación

A nivel general y de forma breve, podemos decir que las líneas de investigación que se nutren de conocimientos más lejanos en el tiempo son *Distributed artificial intelligence* y *Heuristic*, pero sobre todo, la primera. Son líneas muy consolidadas, pero que no reciben muchas citas, como indica el tamaño de sus nodos, también son bastante lejanas en el tiempo, como muestra el color de sus nodos. Esto revela poca actividad, y una base de conocimiento con poca evolución.

Function approximation y *Case-base reasoning*, también son líneas consolidadas que se alimentan mucho de trabajos pretéritos. No obstante, sus citas son de fechas más recientes, y son puntos importantes de interconexión con otras líneas, como muestran las aureolas moradas de muchos de sus nodos. Son también bastante activas y atraen la atención de los investigadores, como podemos ver por su elevado número de nodos rojos.

Natural intelligence es la línea de investigación de mayor tamaño, la que más interrelaciona con el resto, como indica el contar con la referencia de mayor *betweenness*, así como otras con valores similares, a la vez que permanece bastante activa. Su número de citas no es muy elevado, pero sí es superior a las anteriores líneas de investigación.

Computational and artificial intelligence, *Support vector machines* y *Firefly Algorithm*, son las líneas con un mayor número de citas recientes. Sobre todo la primera.

Computational and artificial intelligence y *Support vector machines*, reciben un número elevado de citas a lo largo del tiempo, tienen trabajos con un elevado índice de *citation burst*, y cuentan con referencias con un alto grado de *betweenness*, permitiéndoles relacionarse entre sí y sobre todo, con *Natural intelligence*.



Firefly Algorithm es una línea joven, como se observa en su momento de aparición y las fechas de las citas que recibe, pero ya cuenta con varias referencias que le permiten ser una línea activa e interconectada con el resto.

IV. CONCLUSIONES

Mostramos por primera vez la red de la estructura intelectual de la IA a nivel mundial. Se han identificado ocho líneas principales de investigación altamente relacionadas entre sí, a la vez que poco homogéneas a nivel individual. Cinco de ellas están muy consolidadas y tres, son de más reciente aparición. Prácticamente, todas las líneas interactúan entre sí, pero sobre todo lo hace *Natural intelligence*, que es a su vez la de mayor tamaño. Son especialmente activas, por el número de citas que atraen: *Case-based reasoning*, *Computational and artificial intelligence* y *Support vector machines*.

BIBLIOGRAFIA

- [1] J.L. Moreno. Who shall survive? Beacon House, New York, 1953.
- [2] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera. "Science mapping software tools: Review, analysis, and cooperative study among tools". *Journal of the American Society for Information Science and Technology*, vol. 62, nº 7, pp. 1382-1402, 2011.
- [3] F. Moya-Anegón, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, E. Corera-Álvarez, F.J. Muñoz-Fernández, and V. Herrero-Solana, "Visualizing the marrow of science". *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 2167-2179, 2007.
- [4] A.J. Gómez-Núñez, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, V. Batagelj, and F. Moya-Anegón, "Visualization and analysis of SCImago Journal & Country Rank structure via journal clustering", *Aslib Journal of Information Management*, vol. 68, nº 5, pp. 607-627, 2016.
- [5] B. Vargas-Quesada, Z. Chinchilla-Rodríguez, and N. Rodríguez, "Identification and Visualization of the Intellectual Structure in Graphene Research", *Frontiers in Research Metrics and Analytics*, vol. 2, pp. 1-22, 2017.
- [6] J.H.Jr. Ward, "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, vol. 58, pp. 236-244, 1963.
- [7] C. Chen, "Predictive effects of structural variation on citation counts", *Journal of the American Society for Information Science and Technology*, vol. 63, nº 3, pp. 431-449, 2012.
- [8] C. Chen, F. Ibekwe-SanJuan, and J. Hou, "The structure and dynamics of co-citation clusters: a multiple-perspective co-citation analysis". *Journal of the American Society for Information Science and Technology*, vol. 61, nº 7, pp. 1386-1409, 2010.
- [9] L.C. Freeman, "A set of measuring centrality based on betweenness", *Sociometry*, vol. 40, pp. 35-41, 1977.
- [10] J. Kleinber, "Bursty and hierarchical structure in streams", 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Edmonton, Alberta, Canada, 2002, <http://www.cs.cornell.edu/home/kleinber/bhs.pdf>