

Los mapas auto-organizados para la evaluación de la investigación de tesis doctorales: el caso de la Didáctica de las Ciencias Sociales en España

Self-organizing maps for research evaluation of doctoral dissertations: the case of teaching Social Sciences in Spain

Curiel-Marín, Elvira⁽¹⁾; Passoni, Isabel⁽²⁾; Eva M. Olmedo-Moreno⁽¹⁾ & Fernández-Cano, Antonio⁽¹⁾

(1) Universidad de Granada (España) (2) Universidad Nacional del Mar del Plata (Argentina)

Abstract

This paper has as main objective to highlight the potential use of neural networks, self-organized maps type (SOM), as a clarifying tool in the treatment, analysis and visualization of scientometric data, specifically, in the case of the analysis of the Spanish doctoral theses in teaching Social Sciences, indexed in TESEO (Spanish national database of dissertations), and defended between 1976 and 2014. A census of 301 doctoral theses has been recovered, analyzed according to autonomous communities (Andalusia and Catalonia), five-year term groups, thematic categories and educational stages. In Andalusia, the production is highest in the five-year period 1986-1990 and 2001-2005. In Catalonia, the most productive five-year periods were 1991-1995, 1996-2000, 2001-2005 and 2006-2010. More agreement is needed in the nomenclature of the teaching Social Sciences area, as well as an update in the operation of the TESEO database. As a general conclusion, it can be inferred that the resulting SOM allow to update the understanding of the state of the art in the area based on the various variables considered. The potentiality of SOM as an exploratory approximation of multivariate data becomes evident.

Keywords

Scientometrics; Research Evaluation; Doctoral Theses; Self Organizing Maps; Neural Networks; Methodological Tools; Social Science Teaching

Resumen

El presente artículo tiene el objetivo de poner en valor el potencial del uso de redes neuronales del tipo mapas auto-organizados (SOM) como herramienta clarificadora en el tratamiento, análisis y visualización de datos cuantitativos, concretamente, en el caso del análisis de las tesis doctorales españolas en Didáctica de las Ciencias Sociales, indexadas en la base nacional española de tesis doctorales TESEO, defendidas entre 1976 y 2014. Se ha recuperado un censo de 301 tesis doctorales, clasificadas según las variables: comunidades autónomas (Andalucía y Cataluña), quinquenios de producción, categorías temáticas y etapas educativas. En Andalucía la producción es más alta en los quinquenios 1986-1990 y 2001-2005. En Cataluña, los quinquenios más productivos fueron 1991-1995, 1996-2000, 2001-2005 y 2006-2010. Como conclusión general se infiere que los SOM resultantes permiten actualizar la comprensión sobre el estado de la cuestión en la disciplina, en base a las diversas variables consideradas, que como “metáforas visuales” ayudan a revelar patrones ocultos en los datos. La potencialidad de los SOM como aproximación exploratoria de datos multivariados se hace manifiesta.

Palabras clave

Cuantiometría; Evaluación de la Investigación; Tesis Doctorales; Mapas Auto-organizados; Redes Neuronales; Herramientas Metodológicas; Didáctica de las Ciencias Sociales

**Reception
Date**

2018 May 03

**Approval
Date**

2018 June 18

**Publication
Date:**

2018 June 19

**Fecha de
recepción**

2018 Mayo 03

**Fecha de
aprobación**

2018 Junio 18

**Fecha de
publicación**

2018 Junio 19

Autor de contacto / Corresponding author

Curiel-Marín, Elvira. Facultad de Educación, Economía y Tecnología, Universidad de Granada. C/ Cortadura del Valle S/N, Ceuta (España). ecuriel@ugr.es

Hace ya varias décadas que la evaluación de la investigación se ha convertido en un área de gran interés, y, por tanto, de grandes avances dentro de la comunidad de expertos en el área, y por alusiones e interés directo hacia su trabajo, de todos los miembros de la comunidad científica. El auge y generalización de la aplicación de potentes técnicas cuantitativas es el pilar fundamental en el que se basan los actuales procesos de evaluación de instituciones, programas, docentes e investigadores. Algunos ejemplos son la aplicación de los cuartiles del *Journal Citation Report* de *Web of Science* (Clarivate Analytics) o del *Scimago Journal & Country Rank* (Scimago Lab) para la evaluación de sexenios de investigación o acreditaciones del Personal Docente e Investigador en las universidades y centros de investigación españoles.

Existen numerosos debates sobre las metodologías aplicadas en las evaluaciones de los resultados de investigación, y sobre el uso indebido de métricas e indicadores por agentes no expertos, su uso sin contextualizar o la complementariedad metodológica empleando métodos mixtos. Sin embargo, la comunidad científica ha asistido en numerosas ocasiones a un uso inapropiado de la interpretación de los índices e indicadores empleados en la evaluación de la investigación. Declaraciones internacionales como la Declaración de San Francisco sobre Evaluación de la Investigación (DORA, 2012) o el *Leiden Manifesto* (2015) pretenden visibilizar el problema del uso exagerado y no especializado de estas métricas.

Este debate es especialmente sensible entre los investigadores de Ciencias Sociales, Artes y Humanidades, que, con tradiciones de comunicación científica distintas a las Ciencias y Tecnología, sienten que en numerosas ocasiones se les juzga inadecuadamente al aplicárseles los mismos criterios que a sus compañeros. Van Leeuwen (2013) manifiesta que dada la cobertura actual de las muy conocidas bases de datos bibliométricas y el creciente foco del inglés

como lengua franca de la actividad académica moderna, las herramientas bibliométricas actuales son insuficientes para respaldar la evaluación real de los académicos en Ciencias Sociales, Humanidades y Ciencias Jurídicas, siendo necesario un enfoque más amplio con un abanico más grande de tipos de publicaciones considerados, pues las publicaciones no siempre están destinadas a un público meramente académico. Robinson-García, Van Leeuwen y Rafols (2016) recuerdan que los actuales contextos de evaluación de las políticas de investigación fueron diseñados para abordar las ciencias de la vida y la naturaleza, comunidades globales de investigación y el impacto científico, de modo que no se adaptan bien a la comunidad científica de las Ciencias Sociales y las Humanidades, a intereses locales, ni al impacto social desde un punto de vista amplio.

Uno de los retos a los que se enfrentan los expertos en cuantificación y en métodos de investigación, es la elaboración de técnicas de análisis y visualización de datos que sean aplicables y comprensibles para toda la comunidad científica. Se hace necesaria pues la continua actualización y aplicación de nuevas técnicas de tratamiento de datos surgidas de la colaboración entre distintas áreas de conocimiento. Este artículo ilustra como el uso de Redes Neuronales del tipo Auto-Organizadas es una valiosa herramienta para la evaluación de la investigación, ejemplificándolo en el caso de los trabajos doctorales en el área de Didáctica de las Ciencias Sociales en España.

Tanto la investigación cuantitativa con tesis doctorales, como el uso de mapas autoorganizados como metodología de investigación en Bibliometría (Noyons & Van Raan, 1998; Sotongo & Guzmán, 2001; Saavedra Fernández, Sotongo Aguilar & Guzmán Sánchez, 2002; White, 2015; Meireles & Cendón, 2017; Albo Hernández, Sánchez, Bouza Figueroa & Calero Ramos, 2018) resultan no ser aproximaciones novedosas al estudio de la ciencia, si bien consideramos que por su potencial, tanto como

documentos de gran valor como metodología de visualización de datos clarificadora, no se han explotado lo suficiente.

Los SOMs permiten visualizar vistas de baja dimensión de datos de alta dimensión, semejante a un escalado multidimensional, evidentemente multivariados. Sin embargo, su uso en investigación educativa es muy escaso. Una búsqueda en las bases de la Web of Science con la secuencia: TEMA: “self-organizing maps” and educat*” sólo arroja 25 artículos y la mayoría un tanto tangenciales al ámbito estrictamente educativo dado por revistas del campo de la educación. A destacar dos trabajos ampliamente educativos (Nielsen & Yeziarski, 2016; Pellicer et al. 2015).

La tesis doctoral como documento base para estudios bibliométricos

Las tesis doctorales han sido empleadas como base de numerosos estudios en España. En numerosas ocasiones se las considera como la materia prima de la ciencia. Sus funciones pueden ser muy diversas, aunque las principales son demostrar la competencia investigadora de un futuro doctor y generar conocimiento científico valioso (Fernández-Cano, Torralbo & Vallejo, 2008).

El uso de la tesis doctoral como base para estudios cuantitativos se remonta a bastantes años atrás, pero como señalan Andersen y Hammarfelt (2011) aún queda mucho por investigar, pues no se ha generalizado la aplicación de técnicas bibliométricas a éstas. En relación con el uso de tesis doctorales en cuantimetría:

Una de las primeras unidades históricas de la investigación cuantitativa está hoy en día en gran medida en desuso, como son las cifras de publicaciones de tesis doctorales. Derek J. de Solla Price utilizó esta unidad, junto con otras medidas, como un indicador de la adición de un nuevo investigador a un campo científico. (p. 372).

Algunos aspectos que son analizables para entender mejor la importancia de las tesis doctorales son (Jiménez-Contreras, Ruiz Pérez

& Delgado López-Cózar, 2014): su valor como fuente e indicador de las tendencias en investigación, su condición de período de formación de investigadores, su relevancia como productoras de resultados de investigación de alto nivel, y su interés para analizar las estructuras organizativas y de poder e influencia en las instituciones académicas.

Fernández-Cano, Torralbo y Vallejo (2012), en relación con la importancia y consideración de las tesis doctorales como documentos para estudios cuantitativos, subrayan la relevancia que tienen también para la comunidad científica a la que pertenece el doctorando, representando las tres funciones principales de las universidades: tutorización mediante el mentorazgo y supervisión, investigación desarrollada por el doctorando, y gestión de la información resultante.

Dentro de la investigación educativa española encontramos una prolífica literatura relacionada con revisiones bibliométricas de tesis doctorales. Destaca, Agudelo et al. (2003), que analizan la producción científica de tesis doctorales de Psicología en el periodo 1992-2001.

La didáctica específica sobre la que más se ha indagado en producción de tesis doctorales es la Didáctica de la Matemática. Ejemplos de ello son los trabajos realizados por Fernández-Cano, Torralbo y Vallejo. En Fernández-Cano, Torralbo y Vallejo (2012) se realiza un análisis de las series temporales de crecimiento científico usando los datos sobre tesis doctorales españolas en el periodo 1848 a 2009. En 2013, Ferreira-Villa, Pascual-García y Pol-Asmarats publicaron su estudio sobre tesis doctorales españolas sobre orientación en la base de datos TESEO, para los años 2001-2012. Fernández-Bautista, Torralbo y Fernández-Cano (2014) analizaron la producción longitudinal de las tesis doctorales en educación, defendidas en España, desde 1841 a 2012.

En relación con el área de Didáctica de las Ciencias Sociales, destacan dos trabajos que basan su estudio en el análisis de las tesis

doctorales, Prats (2002) explora el estado de la cuestión del área de Didáctica de las Ciencias Sociales en la Universidad Española; Liceras (2004) revisa la investigación relativa a la formación del profesorado en Didáctica de las Ciencias Sociales.

Potencial del uso de redes neuronales para la visualización de datos bibliométricos

Dentro de las tendencias crecientes en el tratamiento y análisis de datos, destaca el uso de los recursos que la Ciencia de Datos brinda a los investigadores. La enorme cantidad de información, no sólo bibliométrica, sino de todo tipo (redes sociales como Mendeley o Twitter, registros sanitarios, bases de datos de grandes multinacionales o gubernamentales) demanda un manejo de datos a través del uso de potentes procesadores que permitan la creación de modelos para el procesamiento de esa información.

La llamada Ciencia de Datos, hace referencia al conjunto de habilidades necesarias para el tratamiento de grandes cantidades de información, desde el manejo de software específico, pasando por conocimientos matemáticos y experiencia investigadora, hasta la habilidad de comprender, aplicar y explicar los diversos modelos que se generan. Dentro de la Ciencia de Datos, la Minería de Datos o *Data Mining*, como explica Casillas (2016), es el proceso de extracción de patrones de información (implícitos, no triviales, desconocidos y potencialmente útiles) a partir

de gran cantidad de datos. Estos patrones son utilizados para explicar los datos existentes, hacer predicciones, resumir las bases de datos masivas para facilitar las tomas de decisiones y para la visualización de datos a través de la extracción de estructuras. La Minería de Datos puede aplicarse a bases de datos relacionales, espaciales, temporales, documentales, multimedia; puede aplicarse a la *World Wide Web*, y a grandes volúmenes de datos, conociéndose como *Big Data* o *Social Big Data*. En la Figura 1 observamos un diagrama donde se sitúa la Minería de Datos en relación al proceso de extracción de conocimiento de las bases de datos (figura 1).

La minería de Datos (*Data Mining*) es un proceso de extracción de conocimiento útil y comprensible, previamente desconocido, a partir de grandes cantidades de datos almacenados en distintos formatos. Este proceso mediante el descubrimiento y cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil para la organización. La minería de datos es una idea basada en una simple analogía. El crecimiento de los repositorios de datos disponibles en las organizaciones ha creado montañas de datos, que representan un recurso de gran valor para la organización. No obstante, para extraer lo valioso de la montaña se debe excavar, o practicar la minería, para llegar a las “pepitas” de metal precioso, en nuestro caso, el conocimiento (Hernández Orallo et al 2004).

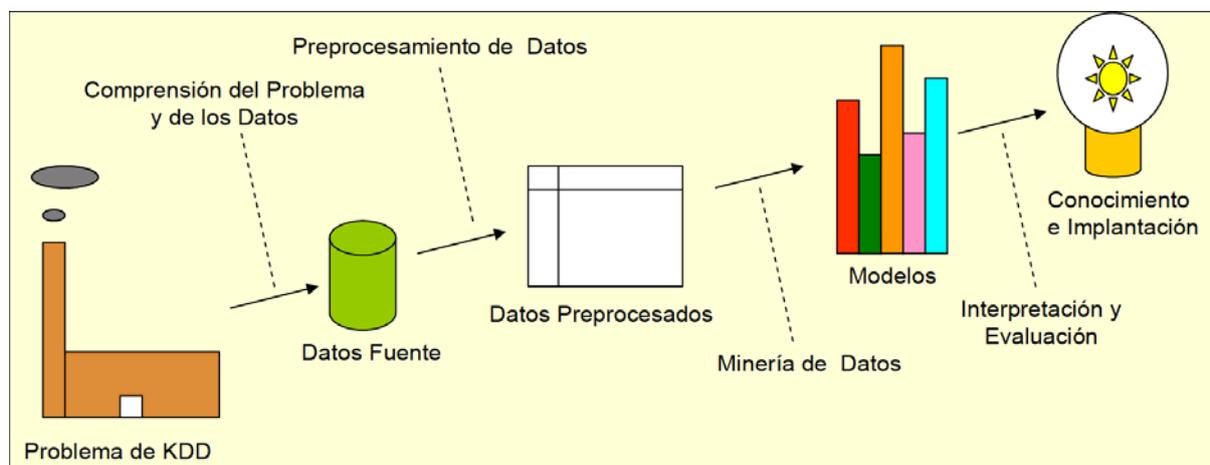


Figura 1. Diagrama del proceso *Knowledge Discovery from Databases* o KDD en el que se encuentra la Minería de Datos (Casillas, 2016)

Una de las opciones para abordar el problema de la representación con modelos es el uso de datos, conceptos y métodos empleados en el llamado Aprendizaje Automático o *Machine Learning*. Como define Passoni (2005):

Un método de aprendizaje automático aplicado a la modelización es un algoritmo que estima o induce un mapeo o dependencia aún desconocida entre las diversas variables del sistema (entradas y salidas). Se consideran “datos” para la modelización a las combinaciones de variables disponibles que se suponen “describen” o se presentan como relevantes para caracterizar el comportamiento del sistema.

Los grupos o técnicas comprendidos en la Inteligencia Computacional, esencialmente las Redes Neuronales Artificiales, los Sistemas Difusos y la Computación Evolutiva, han demostrado su eficacia en la generación de modelos a partir de datos, tanto para el análisis como para la predicción del comportamiento de sistemas complejos. En la actualidad existe una fuerte tendencia a desarrollar modelos híbridos, los que utilizan más de una de las técnicas de Inteligencia Computacional (p.110).

Los Mapas Auto-Organizados, también conocidos como *Self Organizing Maps* (SOM), son una red neuronal artificial de tipo no supervisada. Los Mapas autoorganizados de Kohonen constituyen una potente herramienta para el análisis exploratorio de datos, orientada al reconocimiento de patrones.

El análisis con Redes Neuronales llevado a cabo en esta investigación se basa en el trabajo realizado por Passoni (2003, 2013), empleando las redes auto-organizadas creadas por Kohonen (1990, 1993).

La Didáctica de las Ciencias Sociales en España

Dentro del paraguas de las Ciencias de la Educación, las didácticas específicas, como es el caso de la Didáctica de las Ciencias

Sociales, se dedican al estudio específico de la enseñanza y aprendizaje de esa rama. En base a esto, podríamos definir la Didáctica de las Ciencias Sociales como el área encargada del estudio de la parte relacionada con la enseñanza, aprendizaje, innovación, investigación, metodología, epistemología y evaluación, siempre relacionado con el proceso de enseñanza y aprendizaje, en cualquier contexto, de las disciplinas que, tras la reforma universitaria de 1983, conforman el currículo de Didáctica de las Ciencias Sociales en España: Didáctica de la Geografía y Didáctica de la Historia, que puede desglosarse también en Didáctica de la Historia del Arte. Las grandes disciplinas vertebradoras que se mantienen con entidad propia, por su trayectoria histórica en el currículo en nuestro país, en todos los niveles educativos desde primaria hasta bachillerato, son la Didáctica de la Geografía y la Didáctica de la Historia, ya sea agrupadas en áreas como Conocimiento del Medio, o escindidas como en el caso de la Historia del Arte (Curiel-Marín y Fernández-Cano, 2015).

Joaquín Prats Cuevas, catedrático de Didáctica de las Ciencias Sociales de la Universidad de Barcelona, y uno de los investigadores que más se ha interesado por la epistemología del área, (ver Prats 1997, 2003) distingue cinco líneas fundamentales para investigación en Didáctica de las Ciencias Sociales: Diseño y desarrollo curricular en sus diversas etapas, áreas y disciplinas educativas; Construcción de conceptos y elementos que centren el contenido relacional y polivalente de la Didáctica de las Ciencias Sociales; Estudios sobre comportamiento y desarrollo de la profesionalidad docente en lo referente a la enseñanza de la Historia, la Geografía y Ciencias Sociales; Investigaciones ligadas a las concepciones de la Historia y la Geografía u otras Ciencias Sociales entre el alumnado, y la evaluación de los aprendizajes; Investigaciones sobre Didáctica del Patrimonio. Estas líneas de investigación han servido como guía para la recopilación de las tesis doctorales del área indexadas en la base de datos TESEO, así como guía para la

conformación de las categorías temáticas para el análisis mediante Mapas Auto-Organizados.

Prats (2002) en un recorrido sobre el estado de la cuestión sobre la Didáctica de las Ciencias Sociales en la Universidad Española, repasa la profunda transformación de la disciplina, que desde su creación se ha establecido como un campo de conocimiento suficientemente diferenciado del resto de disciplinas educativas. En un primer momento, la constitución del área se produjo en un momento difícil, pues poco después de la creación de áreas de conocimiento en la universidad española, a mediados de 1980, se suprimió la especialidad de Ciencias Sociales de las Escuelas de Formación del Profesorado. Además, otras peculiaridades del área han dificultado su crecimiento, tal como la distribución interna de las categorías de profesorado (68% de titulares de escuela universitaria en 2002), o las titulaciones donde imparte docencia, ya que desde que desapareció la especialidad de Didáctica de las Ciencias Sociales en la antigua diplomatura de Magisterio, la presencia docente de esta disciplina se redujo notablemente. Por otro lado, el esfuerzo del colectivo docente e investigador especializado en Didáctica de las Ciencias Sociales ha dado lugar a la creación de publicaciones y reuniones periódicas fundamentales para la consolidación del área, tales como *Enseñanza de las Ciencias Sociales: Revista de Investigación*, la revista *Íber* y la *Asociación Universitaria de Profesorado de Didáctica de las Ciencias Sociales*.

Otro aspecto destacable ha sido el crecimiento del número de tesis doctorales defendidas en el área. Si bien, el área no cuenta con acuerdo en la nomenclatura de sus departamentos o palabras clave, llegando a no tener un Tesoro propio, dificultando así la indexación y recuperación de estos trabajos de investigación, se ha observado un crecimiento lineal ($R=0.97$) en la producción de tesis doctorales en el periodo comprendido entre 1976 y 2014 (Curiel-Marín, 2017), pasando de cero o una tesis por año al comienzo del

periodo observado a 15 o 20 tesis defendidas en los últimos años.

Método

Esta investigación es un estudio censal, ya que la muestra coincide con la población, habiendo recuperado todos los documentos disponibles para la investigación, en este caso, todas las tesis doctorales pertenecientes al área de Didáctica de las Ciencias Sociales, defendidas en España e indexadas en la base de datos TESEO, correspondiente a los años comprendidos entre 1976 y 2014. Para completar este censo se llevaron a cabo cuatro secuencias de búsqueda diferentes: En un primer momento se recopilaban las tesis doctorales en TESEO por descriptores y palabras clave. Posteriormente, se realizó una búsqueda exhaustiva en otras bases de datos, en concreto, en la base de datos de tesis doctorales de Histodidáctica, en Dialnet y en Tesis Doctorales en Red, siempre incorporando a la muestra aquellas tesis que también estaban indexadas en TESEO. En un tercer momento, se realizó otra búsqueda en TESEO, indagando cada departamento propio o afín al área de Didáctica de las Ciencias Sociales, de todas las universidades. Para completar el censo, se realizó una búsqueda por autores, según el listado de autores del área del índice bibliométrico *h Index Scholar* del Grupo de Investigación EC3 de la Universidad de Granada. Este proceso se describe con máximo detalle en Curiel-Marín (2017).

Tras este muestreo intencional, no probabilístico y censal, la muestra-población quedó finalmente conformada por 301 tesis doctorales.

Para el análisis se utilizaron los mapas autoorganizados de Kohonen, SOM (de sus siglas en inglés, Self Organizing Maps), (Kohonen 1982). Un SOM bidimensional es una grilla de celdas o neuronas de igual dimensión que los datos de entrada, conectados a todas las unidades de la grilla. La finalidad de este tipo de red neuronal es lograr un mapeo de los datos con el fin de

“descubrir” características de los mismos no observables fácilmente.

Un SOM es un tipo de red neuronal artificial que “aprende” iterativamente cuando un conjunto de datos le es presentado, siguiendo el paradigma de aprendizaje denominado “aprendizaje no supervisado”. En este tipo de aprendizaje no se dispone de una “salida esperada” con la que guiar el aprendizaje, también denominado “entrenamiento”, sino que el SOM intenta descubrir por sí mismo similitudes en los datos.

Se espera que, una vez entrenado un SOM, éste sea de utilidad para descubrir agrupamientos naturales de los datos, según su similaridad, y permita hacer un análisis detallado de cómo se produjeron estos agrupamientos y a qué tipos o categorías de datos representan. Un SOM es un arreglo de neuronas, que en este contexto también se denominan “celdas”. Cada celda tendrá celdas “vecinas” según su proximidad topográfica en el mapa. Puede pensarse que cada celda “guarda” un prototipo que representa uno o más datos del conjunto, debido a su similaridad. Así, a cada dato le corresponde una celda según su prototipo sea el más similar, y a esta celda se la llama “la celda representativa” para cada dato (en inglés se denomina *Best Matching Unit*).

Se denomina entrenamiento del SOM a un proceso iterativo en el que se le presentan de manera consecutiva los vectores “datos” y se realizan los ajustes en el estado interno de las celdas del mapa, tal que se establezca una similitud entre ellas y los datos. En consecuencia, cuando el SOM está correctamente entrenado es factible verificar dos características, que aportan a su potencialidad como método de visualización y análisis:

- a. dos datos similares tendrán celdas representativas iguales o muy cercanas dentro del mapa;
- b. el prototipo de la celda representativa será similar al dato.

Una vez que este proceso finaliza pueden visualizarse sus resultados de diversas maneras; dado un caso o dato y conocida una etiqueta que lo identifica, es posible representar esta etiqueta en el mapa según su celda más representativa. Se busca cuál es esta celda y se agrega una etiqueta sobre la misma. Entonces, las etiquetas que han quedado en celdas cercanas corresponden a datos similares, lo que permitirá extraer conclusiones.

Una interesante visualización de los resultados también puede obtenerse con el mapa de similaridades. En esta representación se visualizan en una escala de colores las similaridades entre datos prototipos de celdas vecinas. Así, similaridades altas (diferencias bajas) indican que dos celdas contienen prototipos similares y viceversa. Zonas de bajas diferencias indicarán zonas de agrupamiento de datos parecidos.

Finalmente, los mapas de componentes ofrecen la posibilidad de saber los valores de cada variable (atributo) en cada una de las celdas. Es posible así analizar si un determinado grupo de datos (descubierto según sus etiquetas) posee o no determinado atributo (representado en la variable). Para este trabajo el análisis mediante mapas auto-organizados, se llevó a cabo utilizando las librerías SOM Toolbox (Vatanen et al) que corren bajo MatLab® Versión 2013^a

En relación a los datos se generó una tabla en una hoja de cálculo con tesis recuperadas, incluyendo los diversos indicadores disponibles en TESEO para cada tesis indexada.

Se emplearon las siguientes variables de tipo dicotómico, es decir, caracterizadas por un “1” si estaban presentes o por un “0” si no lo estaban:

- Comunidades Autónomas: Andalucía, Aragón, Asturias, Islas Baleares, Islas Canarias, Cantabria, Castilla La Mancha, Castilla y León, Cataluña, Extremadura, Galicia, La Rioja, Madrid, Navarra, País Vasco, Murcia, Comunidad Valenciana.

- Quinquenios: Primer quinquenio (1976-1980), Segundo quinquenio (1981-1985), Tercer quinquenio (1986-1990), Cuarto quinquenio (1991-1995), Quinto quinquenio (1996-2000), Sexto quinquenio (2001-2005), Séptimo quinquenio (2006-2010) y Octavo quinquenio (2011-2014).
- Categorías temáticas: Didáctica de las Ciencias Sociales, Didáctica de la Historia, Didáctica de la Historia del Arte, Didáctica de la Geografía, Didáctica del Patrimonio, Educación Cívica, Educación Ambiental, Didáctica de la Religión, Educación Económica, Educación Matemática, Didáctica de la Filosofía, Otras didácticas.
- Etapas Educativas: Educación Infantil, Educación Primaria, Educación Secundaria, Educación Superior.

Tanto el instrumento de recogida de información, como el censo de tesis recuperadas han sido validados mediante triangulación por tres expertos nacionales e internacionales, del área de Métodos de Investigación, Didáctica de las Ciencias Sociales y del área de Estudios de la Ciencia y la Tecnología, mostrando su acuerdo unánime en la validez del instrumento y de los datos y sistema de categorías (variables) para la realización del estudio. La validez de las tesis doctorales recuperadas para esta investigación parte de la validez externa avalada por la inclusión de las tesis doctorales recuperadas de la base de datos TESEO, determinando la autenticidad del documento, y validez interna, basada en la fiabilidad de los documentos al extraerse la información de la ficha oficial que rellenan los propios doctorandos.

Así mismo, siguiendo a Fraenkel y Wallen (1990, citado en Vallejo, 2005), han sido controladas las siguientes amenazas a la validez del diseño (Curiel-Marín, 2017):

- Características del recolector de datos y expectativas: Las características personales del recolector de los datos pueden alterar el proceso de recogida de

datos. Esta amenaza ha sido controlada al contar con una base de datos previa validada (Curiel-Marín, 2013), y una recolectora experimentada que ha ido testeando sistemáticamente la recogida de los datos, siguiendo un proceso estandarizado.

- Mortalidad: La pérdida o falta de documentos es otra amenaza de este tipo de estudios. Al ser un estudio censal, se ha controlado en gran medida esta amenaza, si bien, queda en el aire que por fallos de actualización o de permiso de acceso a las fichas de la base de datos Teseo pueda haberse perdido algún documento.
- Efecto pretest o testing: Referente a las variaciones que se producen por influencia del pretest en los resultados finales. No se presenta este sesgo en este tipo de estudios, dado que no se aplica pre test alguno.
- Deterioro instrumental: No se produce al ser recuperada la ficha tal cual está en Teseo; de todos modos, se realizaron sesiones de duración moderada y se han hecho controles de réplica para verificar la correcta recolección de los datos.
- Una última amenaza a tener en cuenta en este tipo de estudios es la de un pobre análisis debido a los sesgos de supersimplificación y supergeneralización pero el uso de una metodología multivariada como los mapas auto-organizados permite incorporar toda una ingente cantidad de datos, operable tras esta técnica de minería.

Ante el gran número de posibilidades de exploración de las distintas variables mediante los mapas auto-organizados, se ha elegido entrenar los mapas para dos comunidades autónomas (Andalucía y Cataluña).

Resultados

A continuación, se presentan los resultados del análisis de las variables mediante mapas auto-organizados.

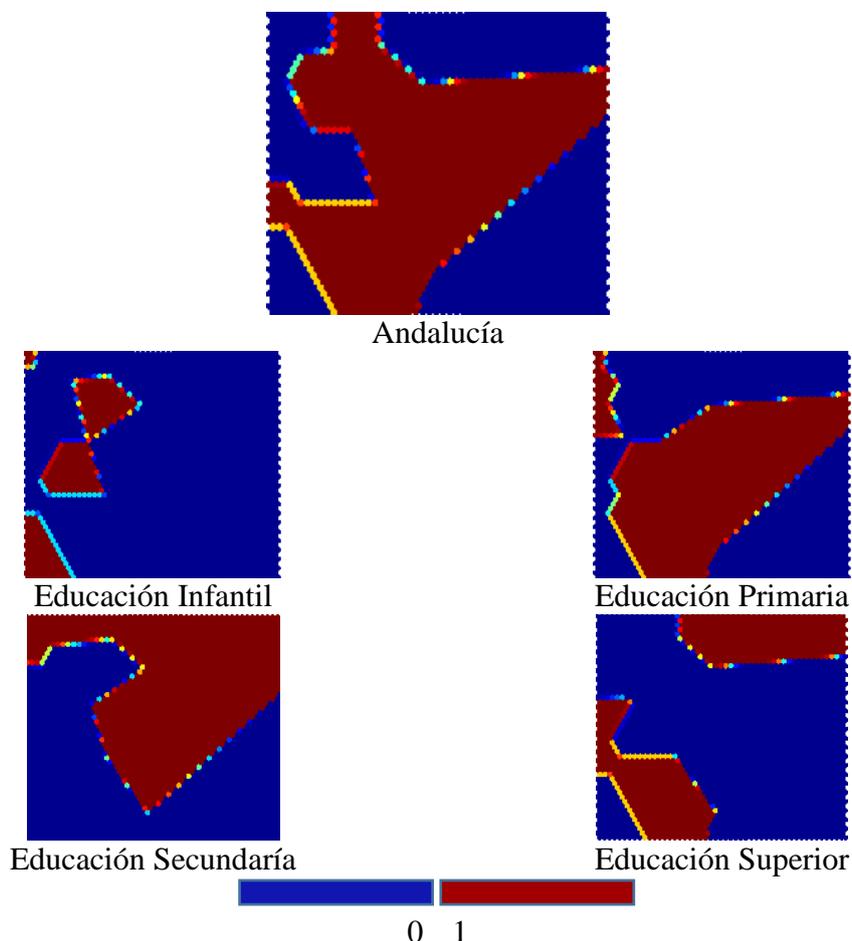


Figura 2. Mapa auto-organizado entrenado para la comunidad autónoma de Andalucía y etapas educativas

La Figura 2 muestra los mapas de variables generadas a partir de un Mapa Auto-Organizado entrenado con datos de la Comunidad Autónoma de Andalucía y las distintas etapas educativas. Los mapas de variables muestran la composición variable a variable de la estructura del SOM entrenado. De este modo, con una simple visualización, podemos observar que la mayor parte de las tesis doctorales que tratan o tienen en cuenta a esta comunidad autónoma y las distintas etapas

educativas, tratan sobre Educación Primaria y Educación Secundaria, al coincidir “topológicamente” en los mapas. Educación Infantil aparece, pero deja fuera parte (No coincide la esquina inferior izquierda del mapa ni el otro segmento trapezoidal del mapa). La Educación Superior queda también en cierta medida excluida en esta comunidad (Segmento de la esquina superior derecha del mapa).

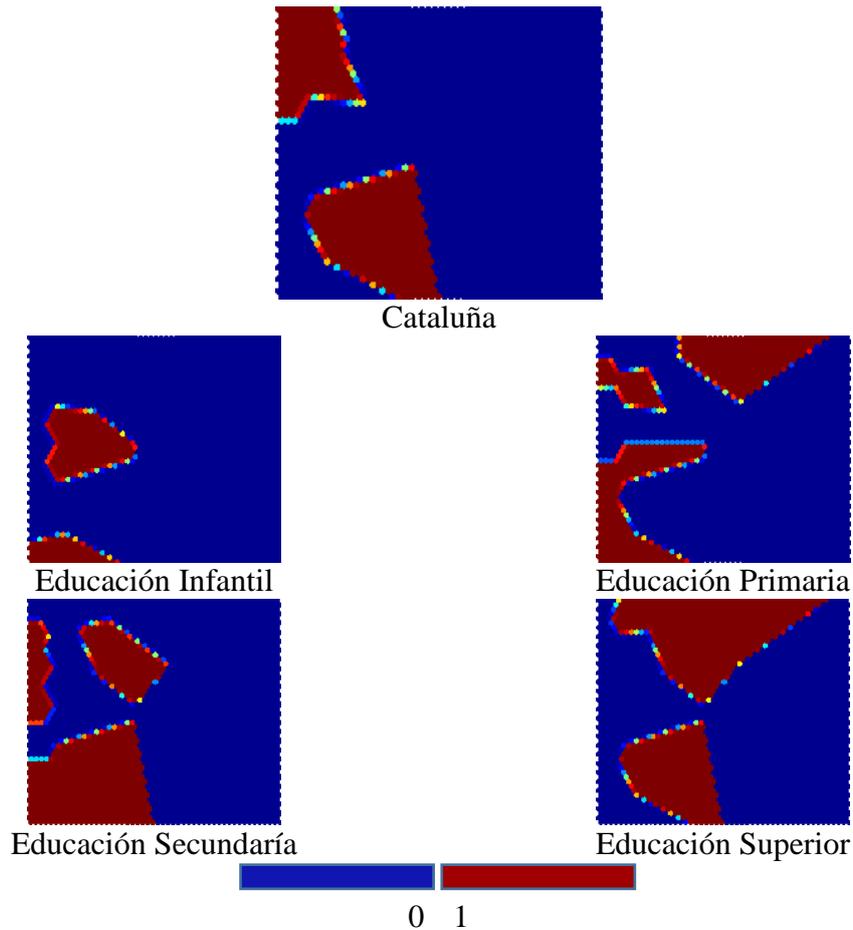


Figura 3. Mapa auto-organizado entrenado para la comunidad autónoma de Cataluña y etapas educativas

En el caso de la comunidad autónoma de Cataluña, podemos observar en la Figura 3 cómo se solapa en gran medida con la investigación relacionada con la Educación

Secundaria y Educación Superior, y un poco con Educación Primaria. La Educación Infantil parece no ser una etapa muy investigada en relación con Cataluña.

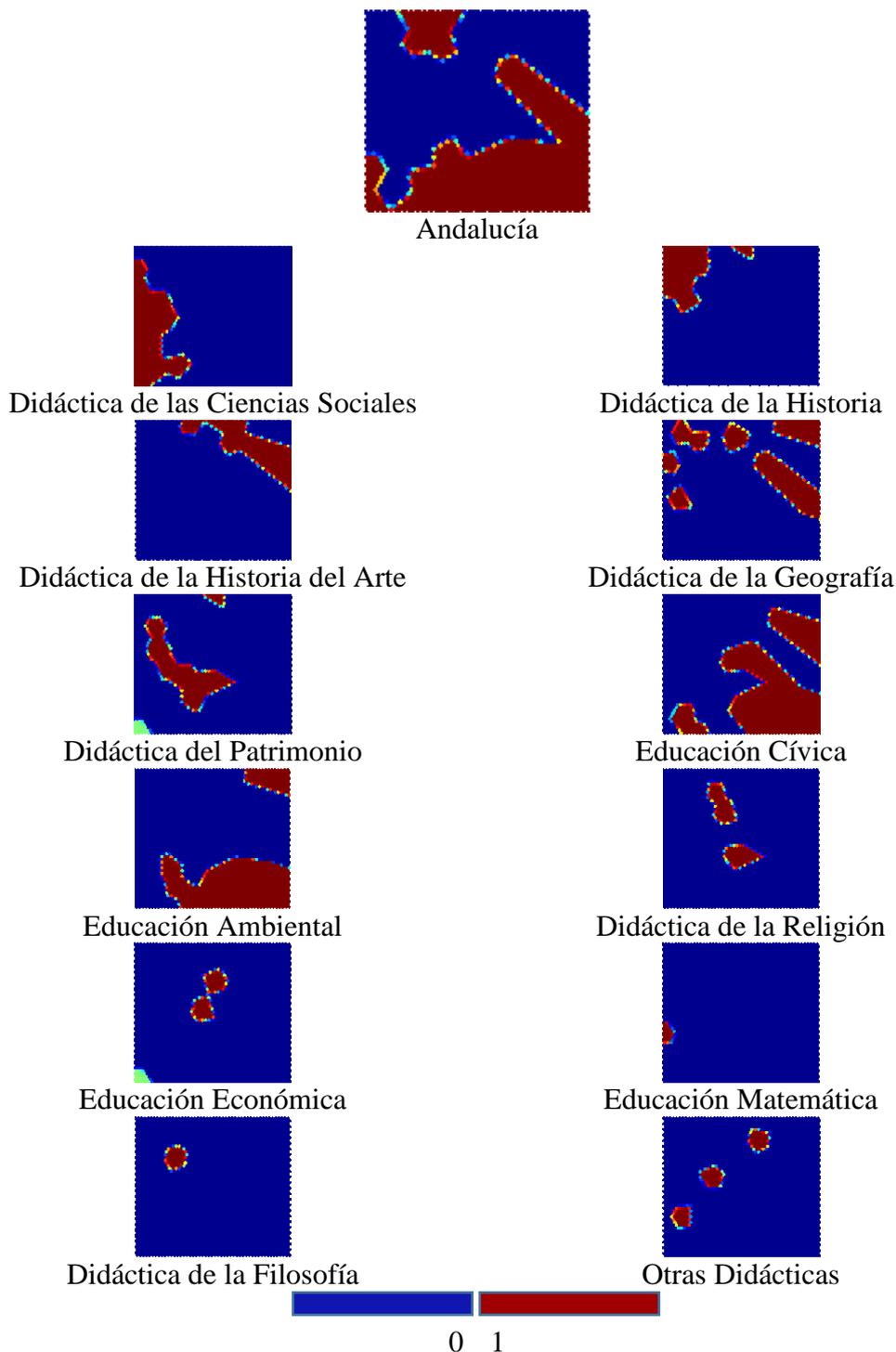


Figura 4. Mapa auto-organizado entrenado para la comunidad autónoma de Andalucía y las variables temáticas

En la Figura 4 advertimos cómo interactúan las investigaciones que relacionan Andalucía con las distintas categorías temáticas en las tesis de Didáctica de las Ciencias Sociales. Se observa claramente

cómo se solapa una parte de la investigación en Didáctica de la Geografía, otra sobre Educación Cívica y Educación Ambiental, y prácticamente todas aquellas relacionadas con Educación Matemática.

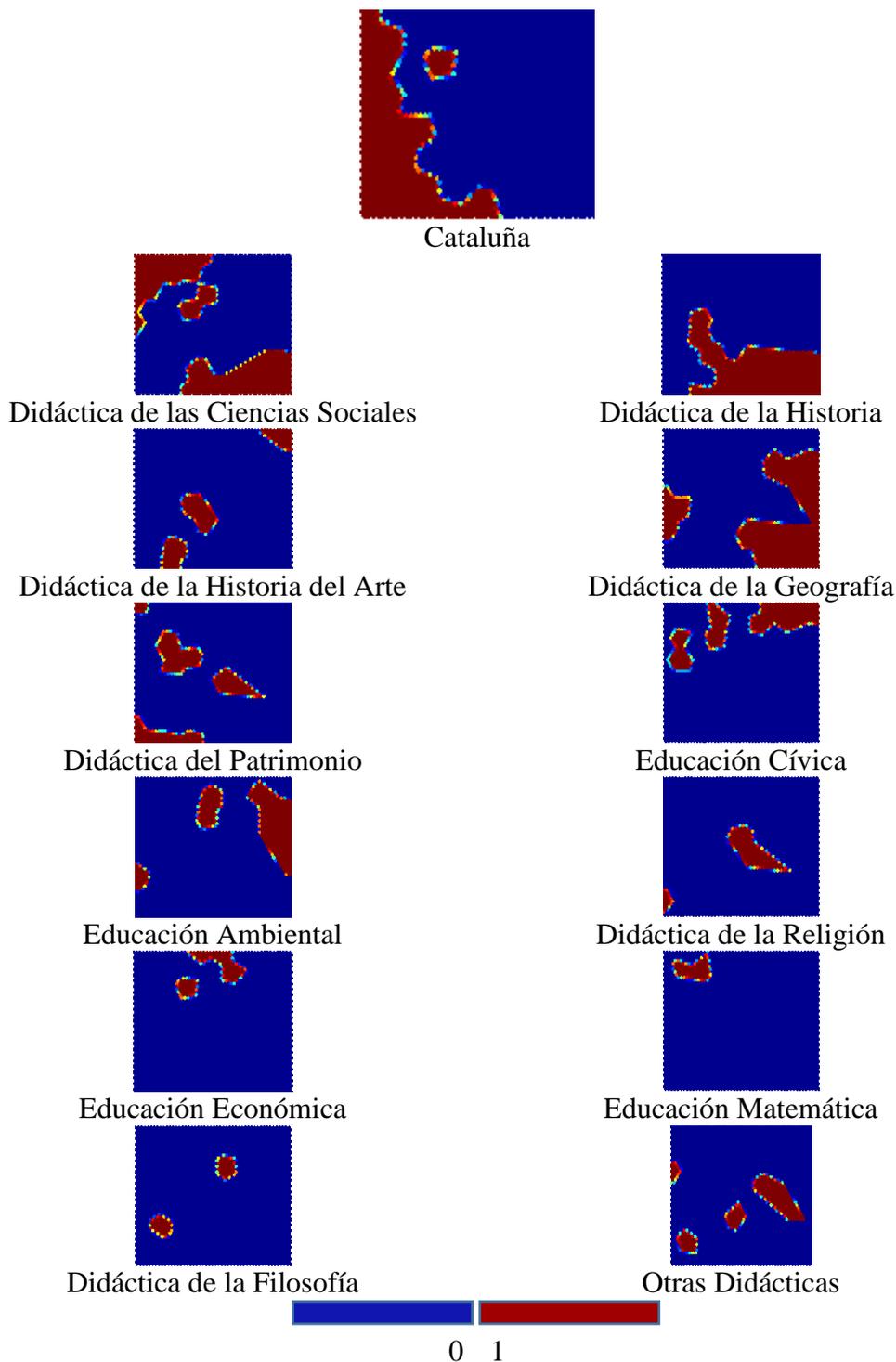


Figura 5. Mapa auto-organizado entrenado para la comunidad autónoma de Cataluña y las variables temáticas

En los mapas auto-organizados para las categorías temáticas en Cataluña, hallamos que los temas de investigación están muy repartidos; coincide en parte con la categoría de Didáctica de las Ciencias Sociales, Didáctica de la Historia, Didáctica de la Historia del Arte, Didáctica de la Geografía,

Didáctica del Patrimonio, Educación Cívica, y en mejor proporción con una pequeña parte de la Educación Ambiental y de la Didáctica de la Religión, así como con la Educación Económica y la Didáctica de la Filosofía (Figura 5).

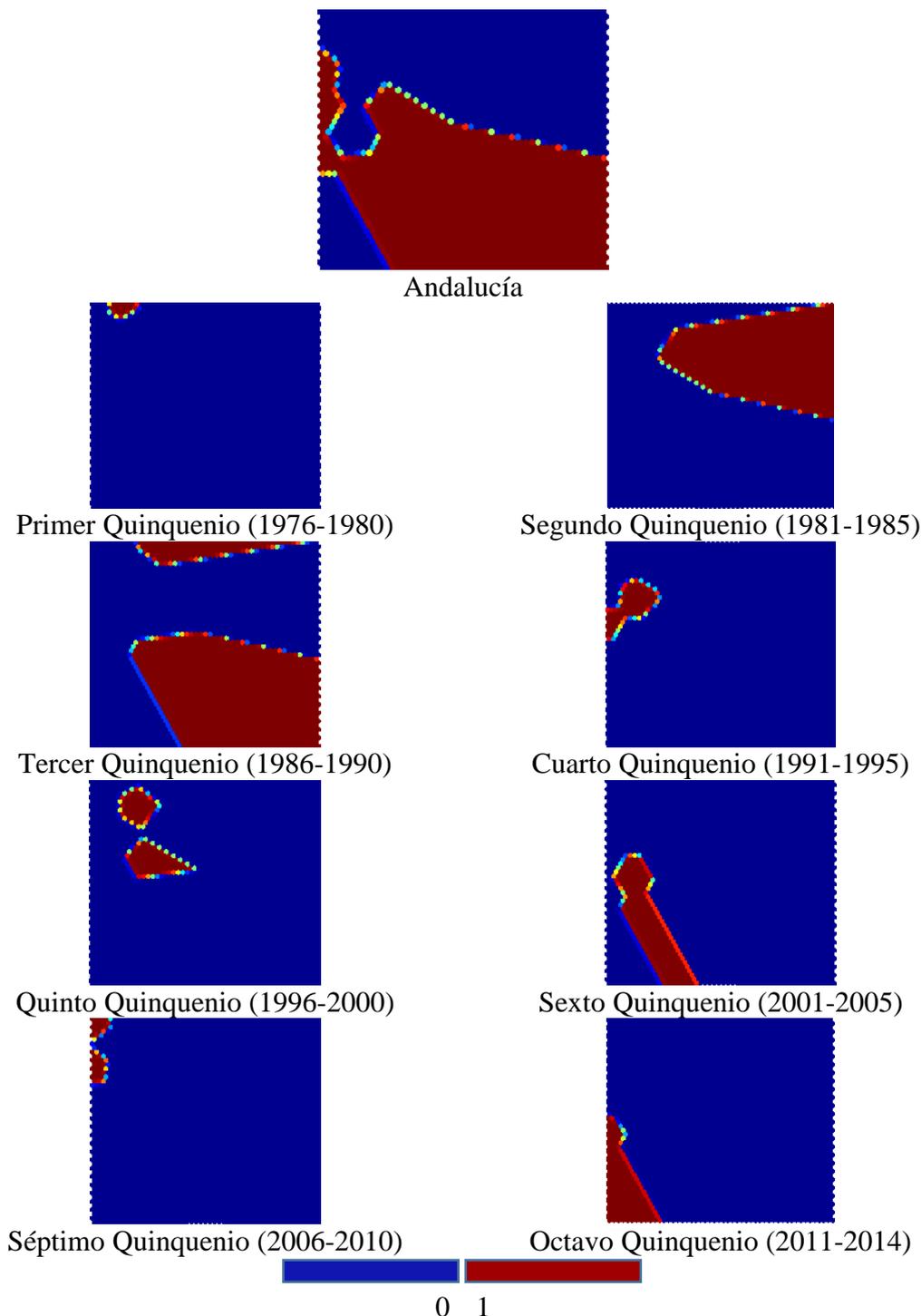


Figura 6. Mapa auto-organizado entrenado para la comunidad autónoma de Andalucía y los quinquenios de producción.

En las figuras 6 y 7 distinguimos la distribución de nuevo para Andalucía y Cataluña, de las investigaciones doctorales recuperadas en función de cuándo se defendieron. En la figura 6 podemos observar cómo la mayor parte de las investigaciones realizadas en Andalucía se realizaron en el

tercer quinquenio (1986-1990) y sexto quinquenio (2001-2005). Para la producción relacionada con Cataluña, la mayor relación la encontramos en el cuarto quinquenio (1991-1995), quinto quinquenio (1996-2000), sexto quinquenio (2001-2005), y séptimo quinquenio (2006-2010) (Figura 7).

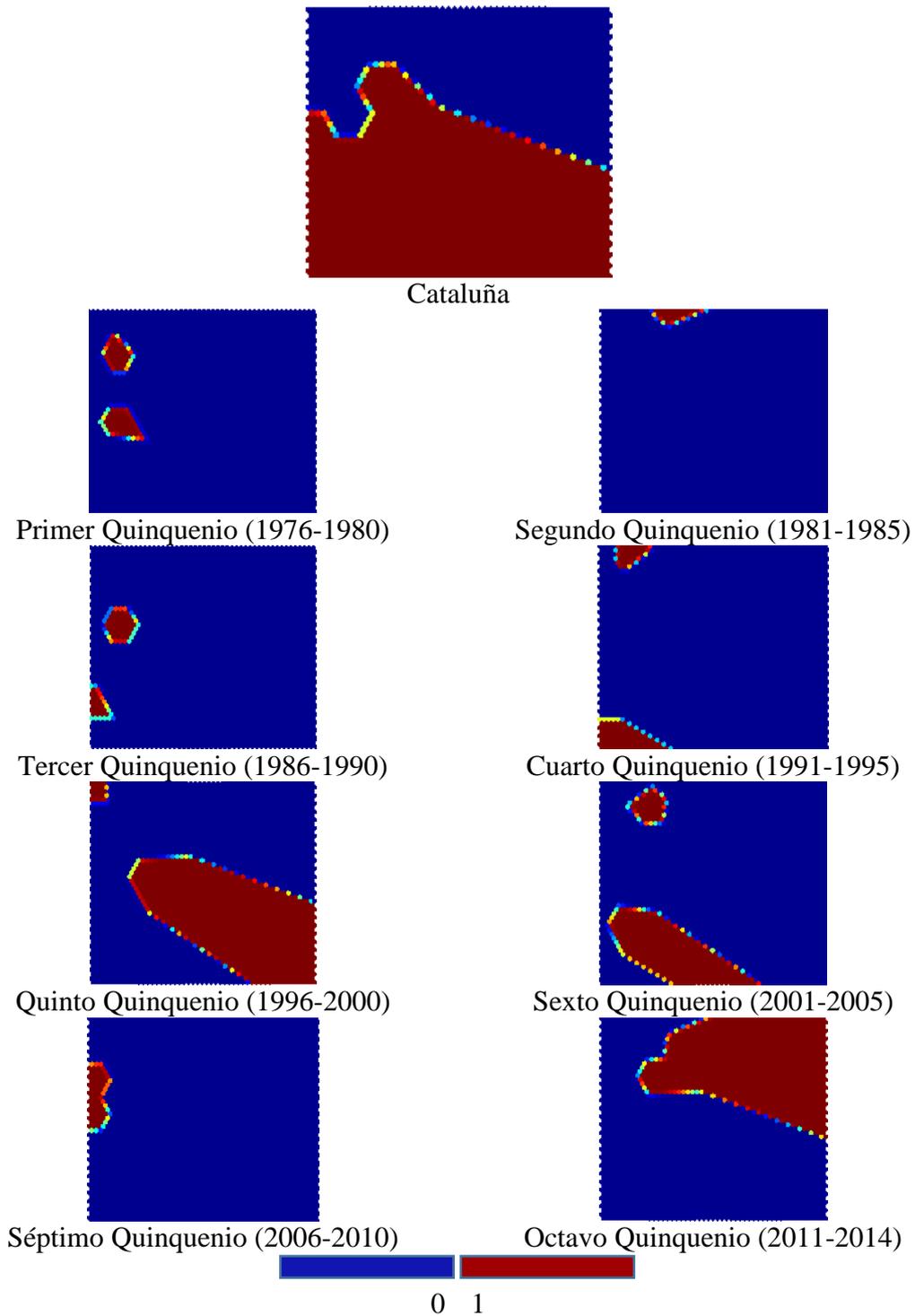


Figura 7. Mapa auto-organizado entrenado para la comunidad autónoma de Cataluña y los quinquenios de producción.

Discusión

El análisis de las tesis doctorales defendidas dentro de un área específica del conocimiento, como es la Didáctica de las Ciencias Sociales en España, permite actualizar la comprensión sobre el estado de la cuestión de la disciplina y conocer la producción por años o categorías temáticas. Este tipo de estudios son especialmente importantes en áreas jóvenes o con poco desarrollo, pues actúan como guía para la agenda de investigación, y pueden dar pautas para su mejora.

La aplicación de las redes neuronales del modo mapas auto-organizados a la visualización de la interacción entre categorías de análisis es una metodología muy novedosa en el área de las Ciencias de la Educación, permitiendo una comprensión practicable y clara de los resultados, especialmente necesaria cuando se manejan grandes cantidades de datos, pues las “metáforas visuales” ayudan a revelar patrones ocultos en los datos (Sotongo-Aguilar, Guzmán-Sánchez & Carrillo, 2002).

Dadas las numerosas posibilidades para entrenar y obtener resultados mediante el uso de los mapas auto-organizados, se decidió mostrar aquellos relacionados con la zona geográfica más productiva en lo relacionado con tesis doctorales españolas en Didáctica de las Ciencias Sociales (Cataluña) (Curiel-Marín, 2017), y con la comunidad autónoma de Andalucía. Para las tesis realizadas en Andalucía, las etapas educativas más estudiadas son Educación Primaria y Educación Secundaria. Para aquellas tesis realizadas en Cataluña, las etapas educativas más estudiadas son Educación Secundaria y Educación Superior. En ambos casos resalta la Educación Infantil como la menos estudiada.

En lo relativo a las investigaciones por categorías temáticas, en Andalucía sobresalen las relacionadas con Didáctica de la Geografía, Educación Cívica, Educación Ambiental y Educación Matemática. En Cataluña, encontramos los temas de investigación muy equilibrados, destacando las categorías de Didáctica de las Ciencias Sociales, Didáctica de la Historia, Didáctica de la Historia del Arte, Didáctica de la Geografía, Didáctica del Patrimonio y Educación Cívica.

Se concluye pues, que tanto la producción general como las categorías más estudiadas coinciden en las investigaciones catalanas. Los quinquenios más fructíferos en Cataluña fueron el cuarto quinquenio (1991-1995), el quinto quinquenio (1996-2000), el sexto quinquenio (1996-2000), y el séptimo quinquenio (2006-2010). Para Andalucía, la mayor productividad tuvo lugar en el tercer quinquenio (1986-1990) y en el sexto quinquenio (2001-2005).

Se hace necesario pues mejorar la investigación doctoral relacionada con la Didáctica de las Ciencias Sociales en Educación Infantil y Educación Superior en Andalucía, así como un incremento en los estudios relacionados con Didáctica de la Historia, Didáctica de la Historia del Arte y Didáctica del Patrimonio. Por otro lado, en Cataluña, para equilibrar aún más los estudios sería interesante ampliar las investigaciones en Educación Infantil y Primaria.

El mayor reto en investigaciones bibliométricas o cienciométricas, especialmente en aquellas relacionadas con tesis doctorales, es la falta de acuerdo en la nomenclatura de las áreas, las palabras clave, y en el caso de la Didáctica de las Ciencias Sociales, el área. La homogeneidad en el uso de las áreas temáticas o palabras clave, al igual que en el caso del nombre y

apellidos cuando se firman publicaciones académicas, es imperativo para hacer posible la recuperación de los documentos.

Otra conclusión crítica de esta investigación es la posibilidad de mejora de la base de datos nacional de tesis doctorales, TESEO, cuya inestabilidad en el uso dificulta la recuperación de la información, llegando a arrojar distintos resultados en las búsquedas realizadas el mismo día, con los mismos criterios; así como la nula transportabilidad de la información generada, que la torna en difícilmente operable.

Referencias

- Agudelo, D., Bretón-López, J., Ortiz-Recio, G., Poveda-Vera, J., Valor-Segura, I. & Vico, C. (2003). Análisis de la productividad científica de la Psicología española a través de las tesis doctorales. *Psicothema*, 15(4), 595-609.
- Albo Hernández, R. O., Sánchez, G., Victoria, M., Álvarez Díaz, I., Bouza Figueroa, J. F. & Calero Ramos, R. (2018). Requerimientos para mejorar la normalización de datos en software de análisis métricos de la información. *Revista Cubana de Información en Ciencias de la Salud*, 29(1), 55-73.
- Andersen, J. P. & Hammarfelt, B. (2011). Price revisited: on the growth of dissertations in eight research fields. *Scientometrics*, 88(2), 71-383. doi: <https://doi.org/10.1007/s11192-011-0408-8>
- Casillas, J. (2016). *Ciencia de Datos, Analítica avanzada y big data* (Material docente), Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada,
- Curiel-Marín, E. (2013). Análisis cuantitativo de tesis doctorales españolas en didáctica de las ciencias sociales (1976-2012). Trabajo Fin de Máster, Universidad de Granada. http://digibug.ugr.es/bitstream/10481/33202/1/CurielMarin_TFM.pdf
- Curiel-Marín, E. (2017) *Análisis cuantitativo de tesis doctorales españolas en Didáctica de las Ciencias Sociales 1976-2014*. Universidad de Granada. [\[http://hdl.handle.net/10481/46522\]](http://hdl.handle.net/10481/46522)
- Curiel-Marín, E. & Fernández-Cano, A. (2015). Análisis cuantitativo de tesis doctorales españolas en didáctica de las ciencias sociales. *Revista Española de Documentación Científica*, 38(4), e110. doi: <https://doi.org/10.3989/redc.2015.4.1282>
- DIALNET (s.d.): <https://dialnet.unirioja.es>
- Fernández-Bautista, A., Torralbo, M. & Fernández-Cano, A. (2014). Análisis longitudinal de tesis doctorales españolas en educación (1841-2012). *RELIEVE*, 20(2). doi: <https://doi.org/10.7203/relieve.20.2.4479>
- Fernández-Cano, A., Torralbo Rodríguez, M. & Vallejo Ruiz, M. (2008). Revisión y prospectiva de la producción española en tesis doctorales de Pedagogía (1979-2006). *Revista de Investigación Educativa*, 26(1), 191-207.
- Fernández-Cano, A., Torralbo, M. & Vallejo, M. (2012). Time series of scientific growth in Spanish doctoral theses (1848–2009). *Scientometrics*, 91(1), 15-36. doi: <https://doi.org/10.1007/s11192-011-0572-x>.
- Ferreira-Villa, C., Pascual-García, L. & Pol, C. (2013). La producción española de Tesis Doctorales sobre orientación en la base de datos TESEO (2001-2012). *RELIEVE*, 19(1). doi: <https://doi.org/10.7203/relieve.10.1.2460>
- Hernández Orallo, J., Ramírez Quintana, M.J. & Ferri Ramírez, C.

- (2004). *Introducción a La Minería de Datos*. Madrid: Pearson Prentice Hall.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429-31. doi: <https://doi.org/10.1038/520429>
- Histodidáctica (s.d.) <http://www.ub.edu/histodidactica/>
- Jiménez-Contreras, E., Ruiz Pérez, R. & Delgado López-Cózar, E. (2014). El análisis de las tesis doctorales como indicador evaluativo: reflexiones y propuestas. *Revista de Investigación Educativa*, 32(2), 295-308. doi: <https://doi.org/10.6018/rie.32.2.197401>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69. doi: <https://doi.org/10.1007/BF00337288>
- Kohonen, T. (1990). The self organizing map. *IEEE Proceedings*, 78(9), 1464-1480. doi: <https://doi.org/10.1109/5.58325>
- Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *IEEE Transactions on Neural Networks*, 6, 895-905. doi: [https://doi.org/10.1016/S0893-6080\(05\)90000-2](https://doi.org/10.1016/S0893-6080(05)90000-2)
- Kohonen, T.S., Kaski, K., Lagus, J., Salojärvi, J., Honkela, V., & Paatero, V. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, 11(3), 574-585. doi: <https://doi.org/10.1109/72.846729>
- Liceras, A. (2004). La investigación sobre formación del profesorado en didáctica de las ciencias sociales. *Profesorado. Revista de Currículum y Formación del Profesorado*, 8 (1). Recuperado de <http://www.ugr.es/~recfpro/rev81ART4.pdf>
- Mathworks (2013). *MATLAB (Software para ordenador)*. Mathworks Inc.: Natick, MA.
- Meireles, M. R. G. & Cendón, B. V. (2017). Citation-based document categorization: an approach using artificial neural networks. *Qualitative and Quantitative Methods in Libraries*, Número especial: 71-79. Recuperado de: http://www.qqml.net/papers/Special_Issue_January_2015_Bibliometrics/4S8QQML_Journal_2015_SpecialIssueBibliometrics_MeirelesandCendon_71-79.pdf
- Nielsen, S. E. & Yeziarski, E. J. (2016). Beyond academic tracking: using cluster analysis and self-organizing maps to investigate secondary students' chemistry self-concept. *Chemistry Education Research and Practice*, 17(4), 711-722. doi: <https://doi.org/10.1039/C6RP00058D>
- Noyons, E. C. & van Raan, A. F. (1998). Monitoring scientific developments from a dynamic perspective: Self organized structuring to map neural network research. *Journal of the Association for Information Science and Technology*, 49(1), 68-81. doi: [https://doi.org/10.1002/\(SICI\)1097-4571\(1998\)49:1<68::AID-ASI9>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-4571(1998)49:1<68::AID-ASI9>3.0.CO;2-1)
- Passoni, L. I. (2003). Las redes neuronales artificiales como elementos del sistema de soporte a las decisiones en la administración universitaria. *FACES*, 9(17), 73-87.
- Passoni, L.I. (2005). *Modelos en Bioingeniería: Caracterización de imágenes estáticas y dinámicas* (Tesis Doctoral). Universidad Nacional de Mar del Plata, Argentina.
- Passoni, L.I., Meschino, G.J., Comas, D.S., Ballarin, V.L. & Scandurra, A.G. (2013). Using SOM as a tool for automated design of clustering systems based on fuzzy predicates. En P. A. Estévez, J. C. Príncipe & P. Zegers (Eds.), *Advances in Self-*

- Organizing Maps*, (pp. 85-94). Heidelberg: Springer. Doi: https://doi.org/10.1007/978-3-642-35230-0_9
- Pellicer, M., García-Massó, X., Morales, J., Serra, P., Solana, M., González, L. M. & Toca-Herrera, J.L. (2015). Physical activity, physical fitness and academic achievement in adolescents: a self-organizing maps approach. *Health Education Research*, 30(3), 436-448. doi: <https://doi.org/10.1093/her/cyv016>
- Prats, J. (1997). *La investigación en Didáctica de las Ciencias Sociales. La formación del profesorado y la didáctica de las Ciencias Sociales*. Sevilla: Díada.
- Prats, J. (2002). La Didáctica de las Ciencias Sociales en la universidad española: estado de la cuestión. *Revista de Educación*, 328, 81-96.
- Prats, J. (2003). Líneas de investigación en Didáctica de las Ciencias Sociales. Apuntes y tendencias. *Història & Ensino*, 9. doi: <https://doi.org/10.5433/2238-3018.2003v9n0p133>
- Robinson-García, N., Leeuwen, T. N. V. & Rafols, I. (2016). SSH & the City. A proposal to map social engagement through social media and web-link analysis. *Open Evaluation Conference*. Recuperado de <http://hdl.handle.net/10261/162532>
- Saavedra Fernández, O., Sotolongo Aguilar, G. & Guzmán Sánchez, M. V. (2002). Contribución al estudio de las revistas de América Latina y el Caribe mediante el mapeo autoorganizado. *ACIMED*, 10(3), 5-6.
- San Francisco Declaration on Research Assessment (DORA): www.ascb.org/dora/
- Sotolongo-Aguilar, G. & Guzmán-Sánchez, M. V. (2001). Aplicaciones de las redes neuronales. El caso de la Bibliometría. *Ciencias de la Información*, 32(1), 27-34.
- Sotongo-Aguilar, G., Guzmán-Sánchez, M. V. & Carrillo, H. (2002). ViBLIOSOM: Visualización de la información bibliométrica mediante el mapeo autoorganizado. *Revista Española de Documentación Científica*, 25(4), 477-484.
- TESEO (s.d.) www.educacion.gob.es/teseo
- Tesis Doctorales en Red (s.d.) <http://www.tesisenred.net/>
- Torralbo, M, Fernández-Cano, A., Rico, L., Maz, A. & Gutiérrez, P. (2003). Tesis doctorales españolas en Educación Matemática. *Enseñanza de las Ciencias*, 21(2), 295-305.
- Torralbo, M. (2002). *Análisis cuantitativo, conceptual y metodológico de las tesis doctorales españolas en Educación Matemática (1976-1998)*. Servicio de Publicaciones de la Universidad de Córdoba.
- Vallejo, M. (2005). *Estudio longitudinal de la producción de tesis doctorales en Educación Matemática (1975-2002)*. (Tesis doctoral). Universidad de Granada.
- Vallejo, M., Torralbo, M. & Fernández-Cano, A. (2005). Análisis diacrónico de la producción española de tesis doctorales en Educación Matemática mediante la metodología ARIMA en datos de diseños longitudinales. En A, Maz, B. Gómez & M. Torralbo (Eds.), *Actas del Noveno Simposio de la Sociedad Española de Educación Matemática SEIEM* (pp. 163-174). Córdoba: Sociedad Española de Investigación en Educación Matemática – SEIEM.
- van Leeuwen, T. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie-Praxis und*

Forschung, 2, 8,1-8,18. doi: <https://doi.org/10.5283/bpf.173>

<https://doi.org/10.1016/j.neucom.2014.02.061>

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T. & Lähdesmäki, H. (2015). Self-organization and missing values in SOM and GTM. *Neurocomputing*, 147(1), 60–70. doi:

White, H. D. (2015). Co-cited author retrieval and relevance theory: examples from the humanities. *Scientometrics*, 102(3), 2275-2299. doi: <https://doi.org/10.1007/s11192-014-1483-4>

Autores / Authors

To know more / Saber más

Curiel-Marín, Elvira (ecuriel@ugr.es).

Profesora del departamento de Métodos de investigación y Diagnóstico en Educación (MIDE) de la Universidad de Granada, y miembro del grupo de investigación HUM-567 de Evaluación de la Investigación y de Programas Educativos Andaluces. Doctora Internacional en Ciencias de la Educación por la Universidad de Granada, es especialista en Evaluación de la Investigación en Educación, centrandó su trabajo actual en el desarrollo de herramientas metodológicas que faciliten la comprensión y transparencia de las evaluaciones cuantitativas, así como la aplicación de métodos mixtos a las mismas. Es la autora de contacto. Dirección Postal: Dpto. MIDE, Facultad de Educación, Economía y Tecnología, Universidad de Granada. C/Cortadura del Valle S/N, Ceuta (España).


[0000-0002-0014-3971](https://orcid.org/0000-0002-0014-3971)



Passoni, Isabel (ipassoni@fi.mdp.edu.ar).

Doctora en Ingeniería, orientación Electrónica y Magister en Gestión Universitaria. Trabaja como catedrática en la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata. Es directora del Laboratorio de Bioingeniería del Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICYTE). Sus áreas de interés se centran en el desarrollo de técnicas de Inteligencia Computacional para sistemas de soporte a las decisiones. Dirección Postal: Facultad de Ingeniería. Departamento de Ingeniería Electrónica y Computación. Avd. Juan B Justo 4302. Cp: 7600. Mar del Plata (Argentina).


[0000-0001-5923-597X](https://orcid.org/0000-0001-5923-597X)



Olmedo-Moreno, Eva M. (emolmedo@ugr.es).

Doctora en Pedagogía por la Universidad de Granada y Profesora Titular en el Departamento MIDE de la Universidad de Granada. Su principal línea de investigación versa sobre “Estrategias de Aprendizaje y el desarrollo de modelos de aprendizaje híbridos en la cultura de la Hiper-globalización-Smart Cities”. Es la directora del Departamento MIDE y coordinadora del Programa de Doctorado de Ciencias de la Educación, además de editora de RELIEVE. Su dirección postal es Facultad de Ciencias de la Educación. Campus Universitario de Cartuja C.P 18071 Granada (España)


[0000-0003-0558-1513](https://orcid.org/0000-0003-0558-1513)



Fernández-Cano, Antonio (afcano@ugr.es).

Doctor en Filosofía y Ciencias de la Educación. Catedrático de en el departamento de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Granada. Sus áreas específicas de interés y especialización son la evaluación de la investigación, la cuantimetría pedagógica, el estudio de las tesis doctorales y las metodologías complementarias para la evaluación de programas educativos. Su dirección postal es Facultad de Ciencias de la Educación. Campus Universitario de Cartuja C.P 18071 Granada (España)


[0000-0003-3991-4443](https://orcid.org/0000-0003-3991-4443)





Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).