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

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

Curiel-Marín, Elvira[1]**; Passoni, Isabel**[2]**; Eva M. Olmedo-Moreno**[1] **& Fernández-Cano, Antonio**[1]

(1) University of Granada (Spain)   (2) National University of 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 cienciométricos, 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

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

Corresponding author / Autor de contacto

Curiel-Marín, Elvira. Faculty of Education, Economics and Technology, University of Granada.
C/ Cortadura del Valle S/N, Ceuta (Spain). ecuriel@ugr.es

| 1

For several decades now, research evaluation has become an area of great interest, and therefore, of great advances within the community of experts in the area, and by allusions and direct interest in their work, of all the members of the scientific community. The rise and generalization of the application of powerful scientometric techniques is the fundamental pillar on which the current evaluation processes of institutions, programs, teachers and researchers are based. Some examples are the application of the quartiles of the *Journal Citation Report* of the *Web of Science* (Clarivate Analytics) or the *Scimago Journal & Country Rank* (Scimago Lab) for the evaluation of research sexenios or accreditation of the Teaching and Research Staff in the Spanish universities and research centers.

There are numerous debates about the methodologies applied in research evaluation results, and about the misuse of metrics and indicators by non-experts, their use without contextualizing or methodological complementarity using mixed methods. However, the scientific community has often attended to an inappropriate use of the interpretation of the indices and indicators used in research evaluation. International declarations such as the Declaration of San Francisco on Evaluation of Research (DORA, 2012) or *Leiden Manifesto* (2015) aim to make visible the problem of the exaggerated and non-specialized use of these metrics.

This debate is especially sensitive among researchers from the Social Sciences, Arts and Humanities, who, with dissimilar traditions of scientific communication than those in Science and Technology, feel that they are inadequately judged on numerous occasions when they are enforced on the same criteria as their peers. Van Leeuwen (2013) states that, given the current coverage of well-known bibliometric databases and the growing focus of English as a lingua franca of modern academic activity, current bibliometric tools are insufficient to support the actual evaluation of academics in Social Sciences, Humanities and Law, being necessary a broader approach with a wider range of types of publications considered, since publications are not always intended for a purely academic audience. Robinson-García, Van Leeuwen and Rafols (2016) recall that the current contexts of research evaluation policies were designed to address life and nature sciences, global research communities and scientific impact, so they do not adapt well to the scientific community of the Social Sciences and the Humanities, to local interests, or to the societal impact from a broad point of view.

One of the challenges faced by experts in scientometrics and research methods is the development of data analysis and visualization techniques that are applicable and understandable to the entire scientific community. Therefore, the continuous updating and application of new data processing techniques arising from collaboration between different areas of knowledge is necessary. This article illustrates how the use of Neural Networks of the Self-Organized type is a valuable tool for research evaluation, exemplifying it in the case of doctoral dissertations in Teaching Social Sciences in Spain.

Both scientometric research with doctoral theses, and the use of self-organized maps as a research methodology in Bibliometrics (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), turn out not to be novel approaches to the study of science, although we consider that due to their potential, both as documents of great value and as a clarifying data visualization methodology, they have not been exploited enough.

SOMs allow the visualization of low-dimensional views of high-dimensional data, similar to multidimensional scaling, evidently multivariate. However, its use in educational research is very scarce.

Corresponding author / Autor de contacto

Curiel-Marín, Elvira. Faculty of Education, Economics and Technology, University of Granada.
C/ Cortadura del Valle S/N, Ceuta (Spain). ecuriel@ugr.es

A search in the Web of Science database with the sequence: TOPIC: "self-organizing maps" and educat * "only yields 25 articles and most of them somewhat tangential to the strictly educational field given by journals in the field of education. Two highly remarkable educational papers stand out (Nielsen & Yezierski, 2016, Pellicer et al., 2015).

Doctoral thesis as a basis document for bibliometric studies

Doctoral theses have been used as the basis of numerous studies in Spain. On numerous occasions, they are considered as the raw material of science. Their functions can be very diverse, although the main ones are to demonstrate the research competence of a future doctor and to generate valuable scientific knowledge (Fernández-Cano, Torralbo & Vallejo, 2008).

The use of the doctoral thesis as a basis for scientometric studies goes back many years, but as Andersen and Hammarfelt (2011) point out, much remains to be investigated, since the application of bibliometric techniques to these has not been generalized. In relation to the use of doctoral theses in Scientometrics:

> One of the historically first units of scientometric research is largely unused today, namely publication counts of dissertations. Derek J. de Solla Price used this unit, alongside other measurements, as an indicator of an addition of a new researcher to a scientific field (p. 372).

Some aspects that can be analyzed to improve the understanding of the importance of doctoral theses are (Jiménez-Contreras, Ruiz Pérez & Delgado López-Cózar, 2014): their value as a source and indicator of trends in research, their condition of training period researchers, their relevance as producers of high-level research results, and their interest for analyzing the organizational structures and their power and influence in academic institutions.

Fernández-Cano, Torralbo and Vallejo (2012), in relation to the importance and consideration of doctoral theses as documents for scientometric studies, underline the relevance they also have for the scientific community to which the doctoral student belongs, representing the three main functions of universities: mentoring and supervision tutoring, research developed by the doctoral student, and management of the resulting information.

Within Spanish educational research, we find a prolific literature related to bibliometric reviews of doctoral theses. Highlights, Agudelo et al. (2003), which analyze the scientific production of doctoral theses of Psychology in the period 1992-2001.

The specific didactics on which more has been investigated related to the production of doctoral theses is the Didactics of Mathematics. Examples of this are the works carried out by Fernández-Cano, Torralbo and Vallejo. In Fernández-Cano, Torralbo and Vallejo (2012) an analysis of the time series of scientific growth is made using data on Spanish doctoral theses in the period 1848 to 2009. In 2013, Ferreira-Villa, Pascual-García and Pol-Asmarats published their study on Spanish doctoral theses on orientation in the TESEO database, 2001-2012 period. Fernández-Bautista, Torralbo and Fernández-Cano (2014) analyzed the longitudinal production of doctoral theses in education, defended in Spain, from 1841 to 2012.

In relation to the area of Teaching Social Sciences, two works stand out that base their study on the analysis of doctoral theses, Prats (2002) explores the state of the art of the area of Teaching Social Sciences at the Spanish University; Liceras (2004) reviews the research related to teacher training in Teaching Social Science.

Potential of the use of neural networks for visualization of bibliometric data

Within the growing trends in the treatment and analysis of data, the use of the resources that Data Science provides to researchers stands out. The huge amount of information, not only bibliometric, but of all kinds (social networks such as Mendeley or Twitter, health records, databases of large multinationals or

governments) demand data management using powerful processors that allow the creation of models for processing that information.

The so-called Data Science, refers to the set of necessary skills for the treatment of large amounts of information, from the management of specific software, through mathematical knowledge and research experience, to the ability to understand, apply and explain the various models they generate. Within Data Science, Data Mining, as explained by Casillas (2016), is the process of extracting information patterns (implicit, non-trivial, unknown and potentially useful) from a large set of data. These patterns are used to explain existing data, make predictions, summarize the massive databases, to facilitate decision making, and to visualize data through structures extraction. Data Mining can be applied to relational, spatial, temporal, documentary, multimedia databases; It can be applied to the World Wide Web, and to large volumes of data, known as

Big Data or Social Big Data. In Figure 1 we see a diagram where Data Mining is located in relation to the process of extracting knowledge from the databases (Figure 1).

Data mining is a useful and understandable knowledge extraction process, previously unknown, from large amounts of data stored in different formats. This process, through the discovery and quantification of predictive relationships in the data, allows the transformation of available information into useful knowledge for the organization. Data mining is an idea based on a simple analogy. The growth of data repositories available in organizations has created mountains of data, which represent a resource of great value to the organization. However, to extract what is valuable from the mountain, it is necessary to dig, or to practice mining, to reach the "nuggets" of precious metal, in our case, knowledge (Hernández Orallo et al 2004).
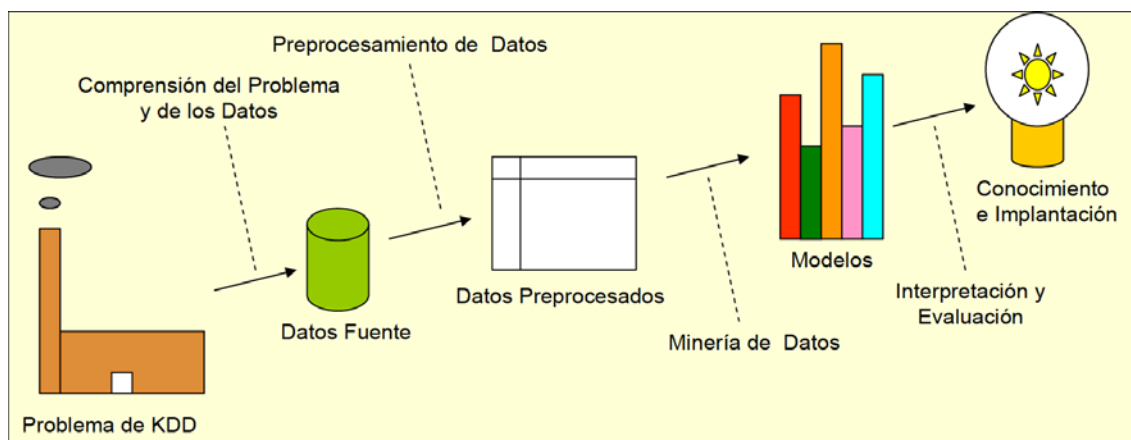

Figure 1. Diagram of *Knowledge Discovery from Databases* process or KDD in which Data Mining is located (Casillas, 2016)

One of the options to approach the problem of representation with models is the use of data, concepts and methods used in the so-called *Machine Learning*. As defined by Passoni (2005):

> A method of automatic learning applied to modeling is an algorithm that estimates or induces a mapping or dependence still unknown among the various system variables (inputs and outputs). The "data" for modeling are

the combinations of available variables that are supposed to "describe" or are presented as relevant to characterize the behavior of the system.

Groups or techniques included in Computational Intelligence, essentially the Artificial Neural Networks, the Fuzzy Systems and the Evolutionary Computation, have demonstrated their efficiency in the generation of models from data, both for the analysis and for

the prediction of the behavior of complex systems. Currently there is a strong tendency to develop hybrid models, which use more than one of the techniques of Computational Intelligence (p.110).

Self-Organized Maps, also known as SOM, are an artificial neural network of an unsupervised type. Self-organized maps of Kohonen are a powerful tool for exploratory data analysis, oriented to the recognition of patterns.

The analysis with Neural Networks carried out in this investigation is based on the work carried out by Passoni (2003, 2013), using the self-organized networks created by Kohonen (1990, 1993).

Teaching Social Sciences in Spain

Within the umbrella of Educational Sciences, specific didactics, such as the case of Teaching Social Sciences, are devoted to the study of teaching and learning of that specific area. Based on this, we could define Teaching Social Sciences as the area in charge of the study of the part related to teaching, learning, innovation, research, methodology, epistemology and evaluation, always related to the teaching and learning process, in any context, of the disciplines that, after the Spanish university reform of 1983, form the curriculum of Teaching Social Sciences in Spain: Didactics of Geography and Didactics of History, which can also be broken down into Didactics of the History of Art. The great vertebrate disciplines that maintain their own entity, for their historical trajectory in the curriculum in our country, in all educational levels from primary to high school, are the Didactics of Geography and the Didactics of History, either grouped in areas as Knowledge of the Environment, or split as in the case of the History of Art (Curiel-Marín and Fernández-Cano, 2015).

Joaquín Prats Cuevas, professor of Teaching Social Sciences at the University of Barcelona, and one of the researchers who has

been most interested in the epistemology of the area, (see Prats 1997, 2003) distinguishes five fundamental research lines in Teaching Social Sciences: Curriculum design and development in its various stages, areas and educational disciplines; Construction of concepts and elements that center the relational and polyvalent content of Teaching Social Sciences; Studies on behavior and development of teaching professionalism in relation to the teaching of History, Geography and Social Sciences; Research linked to the conceptions of History and Geography or other Social Sciences among students, and the evaluation of learning; Research on Heritage Didactics. These research lines have served as a guide for the compilation of the doctoral theses of the area indexed in the TESEO database, as well as a guide for the conformation of the thematic categories for the analysis through Self-Organized Maps.

Prats (2002) in a tour on the state of the art on Teaching Social Sciences at the Spanish University, reviews the profound transformation of the discipline, which since its creation has established itself as a field of knowledge sufficiently differentiated from the rest of educational disciplines.

At first, the constitution of the area came at a difficult time, because shortly after the creation of areas of knowledge in the Spanish university, in mid-1980, the specialty of Social Sciences at Teacher Training Schools was suppressed.

In addition, other peculiarities of the area have hindered its growth, such as the internal distribution of the categories of teaching staff (68% of tenured teachers of university schools in 2002, an outdated position), or the degrees where it teaches, since after the disappearance of the specialty of Teaching Social Sciences in the former teaching diploma, the teaching presence of this discipline was significantly reduced. On the other hand, the effort of the teaching and research group specialized in Social Sciences Didactics has led to the creation of periodic publications and meetings essential for the consolidation of the area, such

as *Enseñanza de las Ciencias Sociales: Revista de Investigación,* the journal *Íber and Asociación Universitaria de Profesorado de Didáctica de las Ciencias Sociales.*

Another notable aspect has been the growth of the number of doctoral theses defended in the area. Although, the area does not have an agreement on the nomenclature of its departments or keywords, not having its own thesaurus, thus hindering the indexing and recovery of these research works, a linear growth has been observed (R = 0' 97) in the production of doctoral theses in the period between 1976 and 2014 (Curiel-Marín, 2017), going from zero or one thesis per year at the beginning of the observed period to 15 or 20 theses defended in recent years.

Method

This research is a census study, since the sample coincides with the population, having recovered all available documents for research, in this case, all the doctoral theses belonging to the area of Teaching Social Sciences, defended in Spain and indexed in the TESEO database, corresponding to the period between 1976 and 2014. To complete this census, four different search sequences were carried out: At first, the doctoral theses were compiled in TESEO by descriptors and keywords. Subsequently, an exhaustive search was carried out in other databases, specifically, in the database of doctoral theses of *Histodidactica*, in *Dialnet* and in *Tesis Doctorales en Red*, always incorporating into the sample those theses that were also indexed in TESEO. In a third moment, another search was carried out in TESEO, investigating each department related or own of the area of Teaching Social Sciences, of all the universities.

To complete the census, a search by authors was carried out, according to the list of authors of the bibliometric index h Index Scholar area of the EC3 Research Group of the University of Granada. This process is described with maximum detail in Curiel-Marín (2017).

After this intentional, non-probabilistic and census sampling, the sample-population was finally made up of 301 doctoral theses.

For the analysis, the self-organized maps of Kohonen, SOM were used (Kohonen 1982). A two-dimensional SOM is a grid of cells or neurons of the same dimension as the input data, connected to all the units of the grid. The purpose of this type of neural network is to achieve a mapping of the data to "discover" easily unobservable characteristics of the same.

An SOM is a type of artificial neural network that "learns" iteratively when a set of data is presented to it, following the learning paradigm called "unsupervised learning". In this type of learning there is no "expected output" with which guide learning, also called "training", but rather the SOM tries to discover similarities in the data by itself.

It is expected that, once a SOM is trained, it will be useful to discover natural groupings of the data, according to their similarity, and allow a detailed analysis of how these groupings occurred and to what types or categories of data they represent. A SOM is an array of neurons, which in this context are also called "cells". Each cell will have "neighboring" cells according to their topographical proximity on the map. It can be thought that each cell "saves" a prototype that represents one or more data of the set, due to its similarity. Thus, each data corresponds to a cell according to which prototype is the most similar, and this cell is called "the representative cell" for each piece of information (Best Matching Unit).

The so-called SOM training is an iterative process in which the "data" vectors are presented consecutively and adjustments are made in the internal state of the map cells, so that a similarity between them and the data is established. Consequently, when the SOM is properly trained it is feasible to verify two characteristics, which contribute to its potential as a method of visualization and analysis:

a. two similar data will have equal or very close representative cells within the map;

b. The prototype of the representative cell will be similar to the data.

Once this process is finished, the results can be visualized in different ways; given a case or a data and a known label that identifies it, it is possible to represent this label on the map according to its most representative cell. The cell is searched and a tag is added to it. Then, the labels that have remained in nearby cells correspond to similar data, which will allow conclusions to be drawn.

An interesting visualization of the results can also be obtained with the map of similarities. In this representation, the similarities between data prototypes of neighboring cells are displayed on a color scale. Thus, high similarities (low differences) indicate that two cells contain similar prototypes and vice versa. Areas of low differences will indicate similar data grouping zones.

Finally, the component maps offer the possibility of knowing the values of each variable (attribute) in each of the cells. It is thus possible to analyze whether a certain group of data (discovered according to its labels) has a certain attribute or not (represented in the variable). For this paper, the analysis using self-organized maps, was carried out using the SOM Toolbox libraries (Vatanen et al) that run under MatLab® 2013 Version

In relation to the data, a table was generated in a spreadsheet with recovered theses, including the various indicators available in TESEO for each indexed thesis.

The following variables of dichotomous type were used, that is, characterized by a "1" if they were present or by a "0" if they were not:

- Autonomous Communities: Andalusia, Aragon, Asturias, Balearic Islands, Canary Islands, Cantabria, Castilla La Mancha, Castilla y León, Catalonia, Extremadura, Galicia, La Rioja, Madrid, Navarra, Basque Country, Murcia, Community of Valencia.

- Quinquenniums: First quinquennium (1976-1980), Second quinquennium (1981-1985), Third quinquennium (1986-1990), Fourth quinquennium (1991-1995), Fifth quinquennium (1996-2000), Sixth quinquennium (2001-2005), Seventh quinquennium (2006-2010) and Eighth quinquennium (2011-2014).

- Thematic categories: Teaching Social Sciences, History Didactics, Art History Didactics, Geography Didactics, Heritage Didactics, Civic Education, Environmental Education, Religion Didactics, Economic Education, Mathematics Education, Didactics of Philosophy, Other didactics.

- Educational Stages: Early Childhood Education, Primary Education, Secondary Education, Higher Education.

Both the instrument for collecting information and the census of theses recovered have been validated by triangulation by three national and international experts, from the area of Research Methods, Teaching Social Sciences and the area of Science and Technology Studies, showing unanimous agreement on the validity of the instrument and the data, and system of categories (variables) for conducting the study. The validity of the doctoral theses recovered for this research is based on the external validity endorsed by the inclusion of the doctoral theses recovered from the TESEO database, determining the authenticity of the document, and internal validity, based on the reliability of the information extracted on the official file that the doctoral students fill out by themselves.

Likewise, following Fraenkel and Wallen (1990, cited in Vallejo, 2005), the following threats to the validity of the design have been controlled (Curiel-Marín, 2017):

- Characteristics of the data collector and expectations: The personal characteristics of the data collector can alter the data collection process. This threat has been controlled by having a validated previous database (Curiel-Marín, 2013), and an experienced collector who has been

systematically testing the data collection, following a standardized process.

- Mortality: The loss or lack of documents is another threat of this type of studies. Being a census study, this threat has been largely controlled, although it remains in the air that due to failure to update, or permission to access the files of the TESEO database, some documents may have been lost.

- Pretest or testing effect: Reference to the variations that occur due to the influence of the pretest on the final results. This bias does not occur in this type of studies, since no pre-test is applied.

- Instrumental deterioration: It does not occur when the record is recovered as it is in TESEO; anyway, sessions of moderate duration were held and replication controls were made to verify the correct collection of the data.

- A last threat to consider in this type of studies is that a poor analysis due to the biases of oversimplification and supergeneralization, but the use of a multivariate methodology such as self-organized maps allows incorporating an enormous amount of data, operable after this mining technique.

Given the large number of possibilities for exploring the different variables through self-organized maps, it has been chosen to train maps for two autonomous communities (Andalusia and Catalonia).

Results

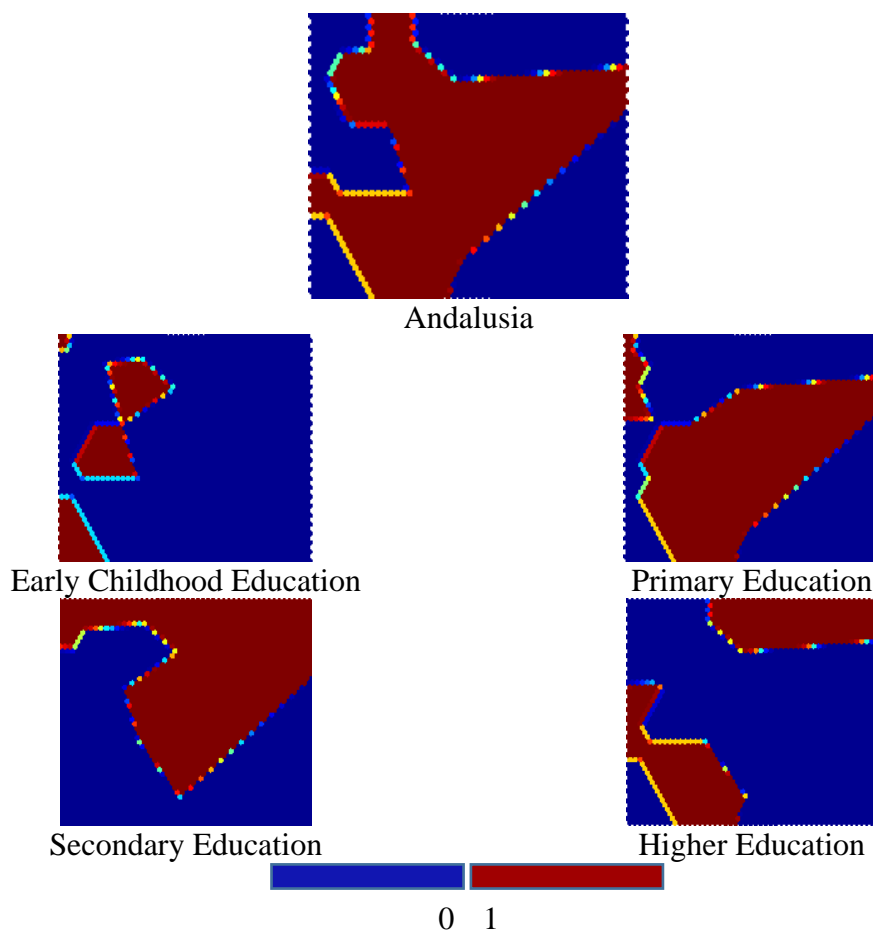Next, the results of the analysis of the variables are presented through self-organized maps.

Figure 2. Self-organized map trained for the autonomous community of Andalusia and educational stages

Figure 2 shows the maps of variables generated from a Self-Organized Map trained with data from the Autonomous Community of Andalusia and the different educational stages. The variable maps show the variable to variable composition of the structure of the trained SOM. In this way, with a simple visualization, one can observe that most of the doctoral theses that treat or consider this autonomous community and the different educational stages, deal with Primary Education and Secondary Education, by coinciding "topologically" in the maps. Early Childhood Education appears, but leaves out part (the lower left corner of the map does not match or the other trapezoidal segment of the map). Higher Education is also to a certain extent excluded in this community (Segment in the upper right corner of the map).
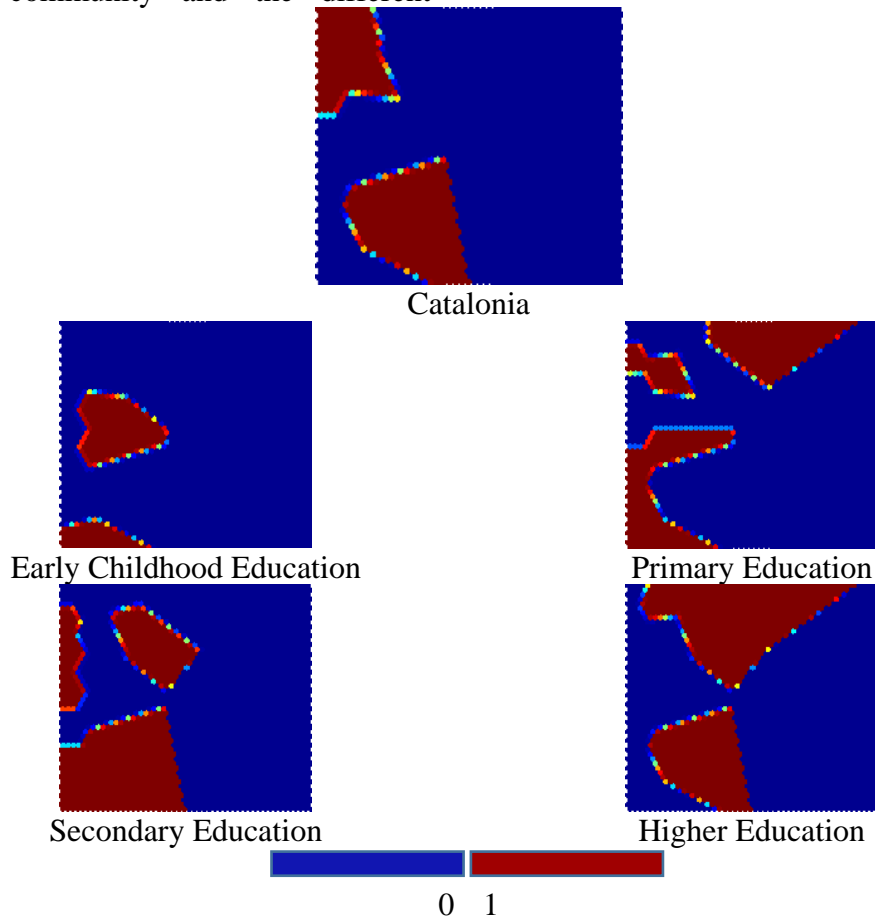

Figure 3. Self-organized map trained for the autonomous community of Catalonia and educational stages

In the case of the autonomous community of Catalonia, we can see in Figure 3 how it overlaps to a large extent with research related to Secondary Education and Higher Education, and a little with Primary Education. Early Childhood Education does not seem to be a highly-researched stage in relation to Catalonia.
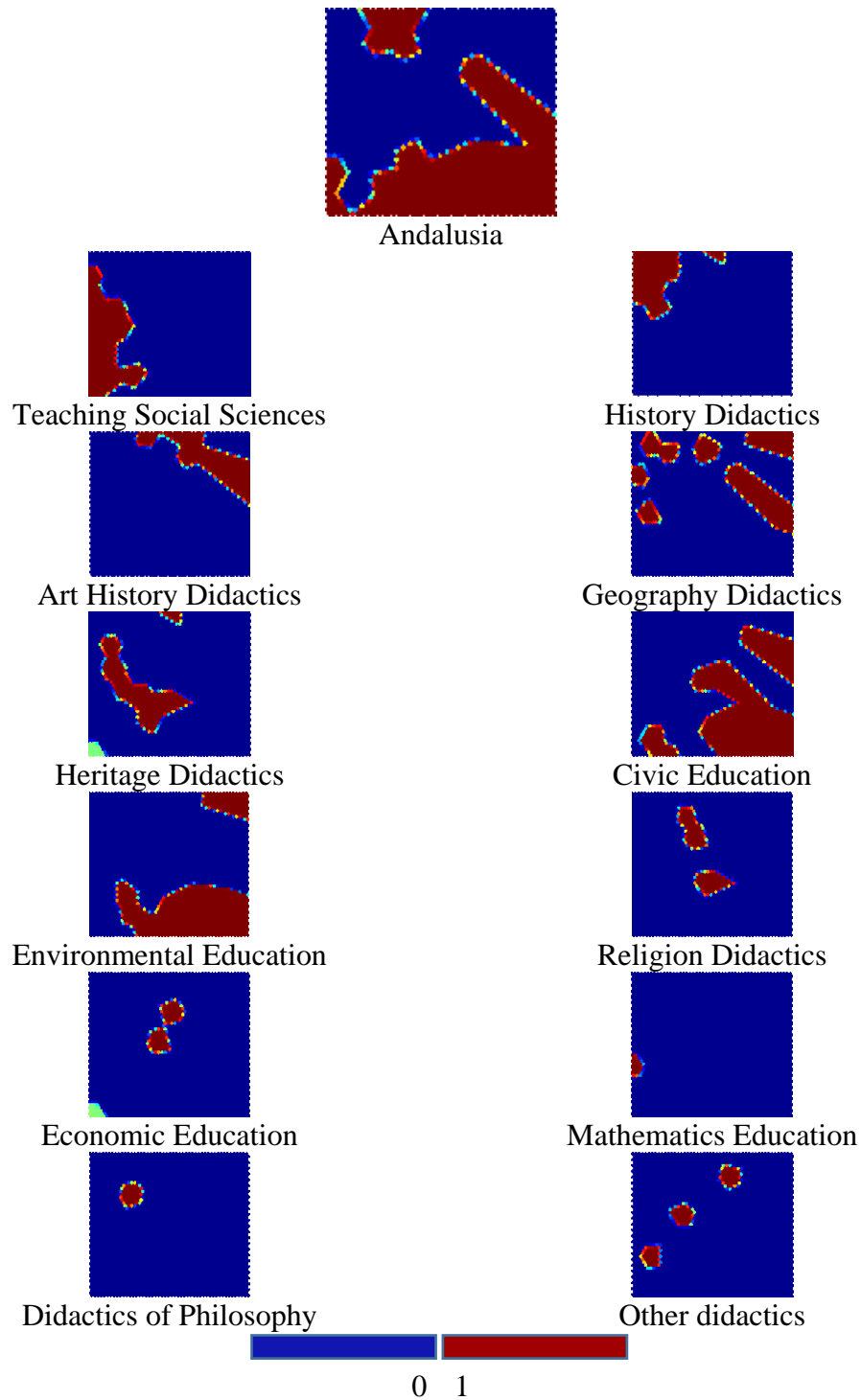
Figure 4. Self-organized map trained for the autonomous community of Andalusia and the thematic variables

In Figure 4 we note how the research that relates Andalusia to the different thematic categories interacts in the Teaching Social Science thesis. It is clearly observed how a part of the research in Geography Didactics overlaps, another on Civic Education and Environmental Education, and practically all those related to Mathematics Education.
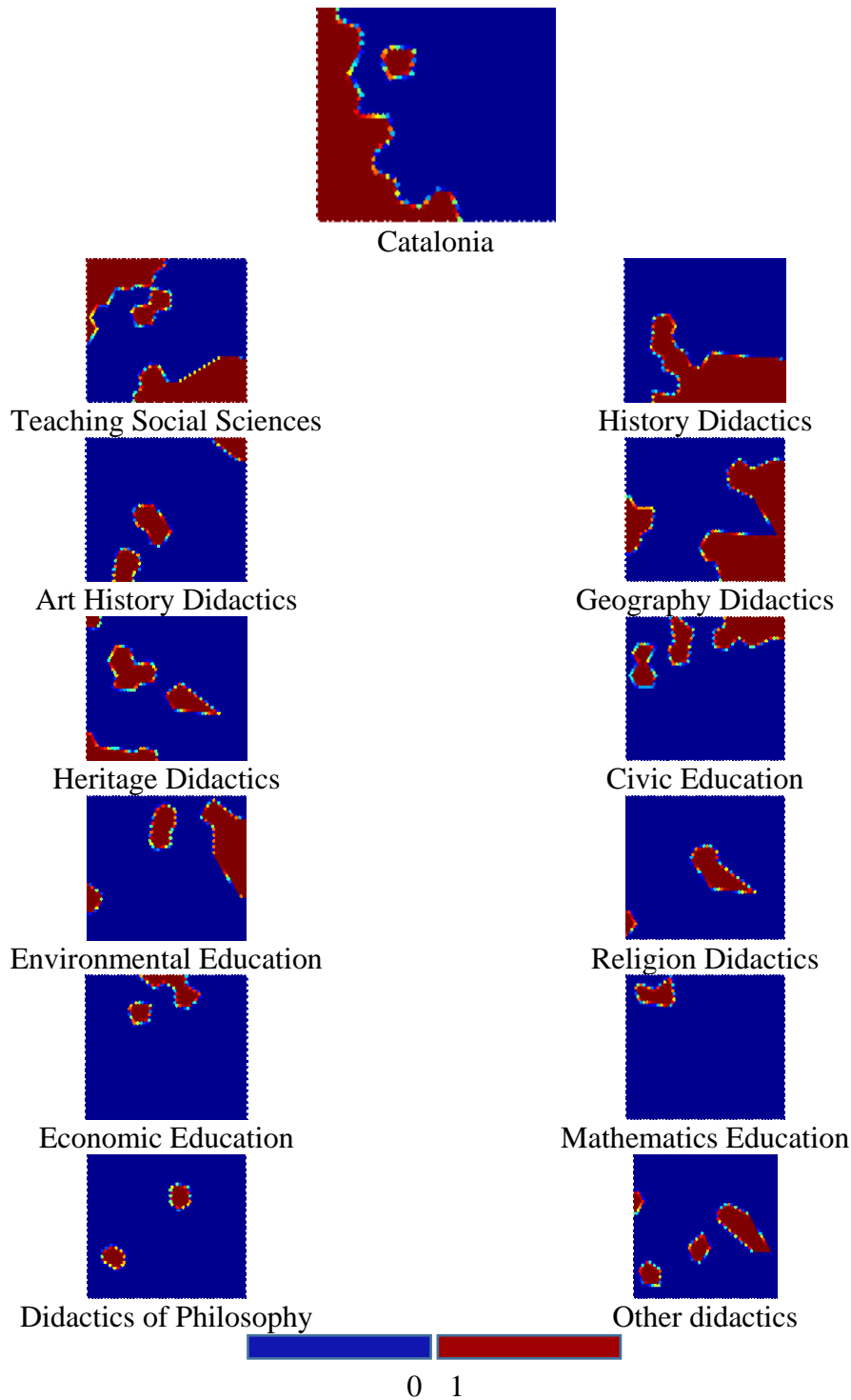
Figure 5. Self-organized map trained for the autonomous community of Catalonia and the thematic variables

In the self-organized maps for the thematic categories in Catalonia, we find that the research topics are widely distributed; it coincides in part with the category of Teaching Social Science, History Didactics, Art History Didactics, Geography Didactics, Heritage Didactics, Civic Education, and in a better proportion with a small part of Environmental Education and of the Didactics of Religion, as well as with the Economic Education and the Didactics of Philosophy (Figure 5).
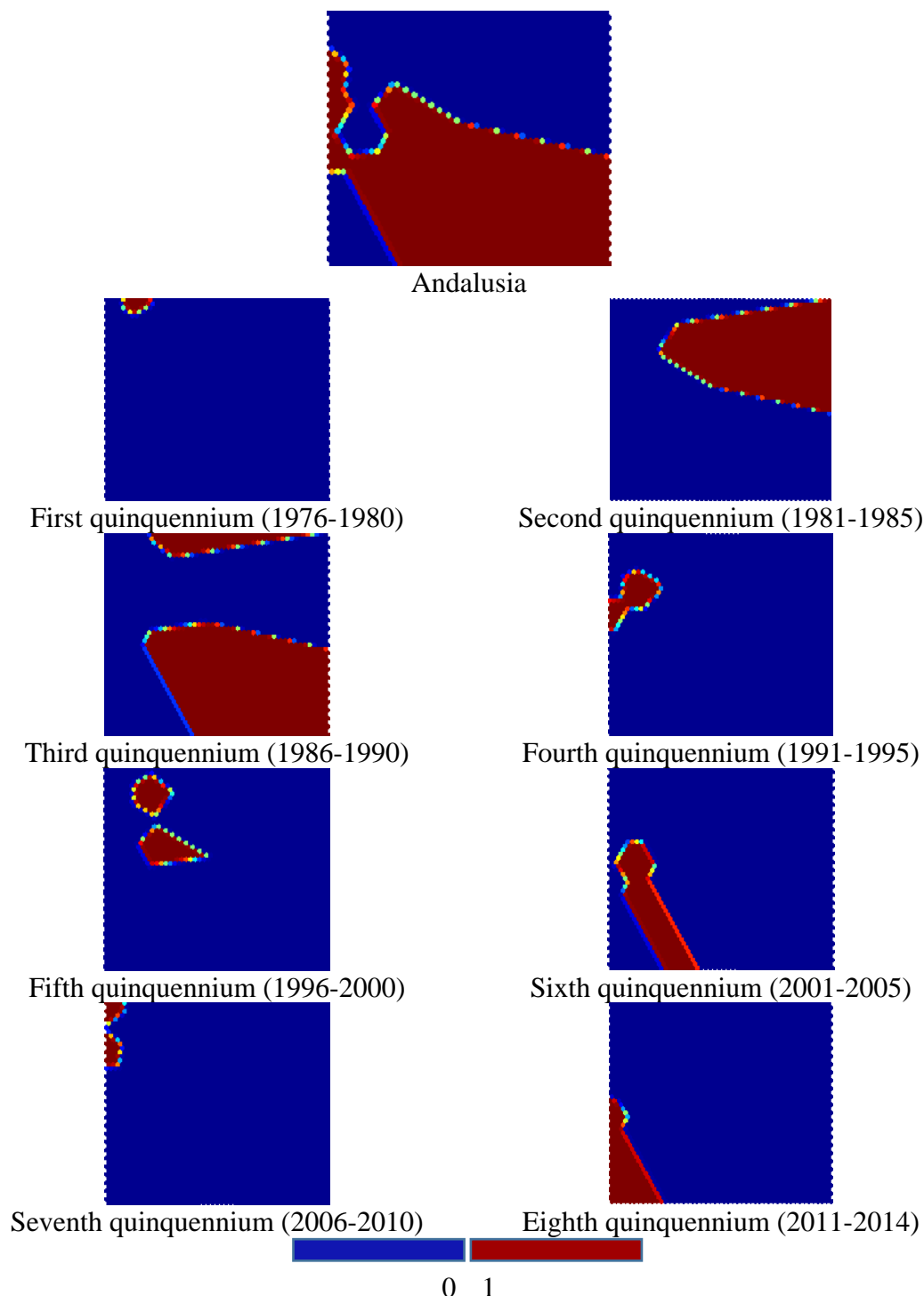
Figure 6. Self-organized map trained for the autonomous community of Andalusia and the five-year periods of production.

In figures 6 and 7 we distinguish the distribution again for Andalusia and Catalonia, from the doctoral research recovered according to when they were defended. In figure 6 we can see how most of the research carried out in Andalusia was carried out in the third quinquennium (1986-1990) and the sixth quinquennium (2001-2005). For production related to Catalonia, the highest ratio is found in the fourth quinquennium (1991-1995), the fifth quinquennium (1996-2000), the sixth quinquennium (2001-2005), and the seventh quinquennium (2006-2010) (Figure 7).
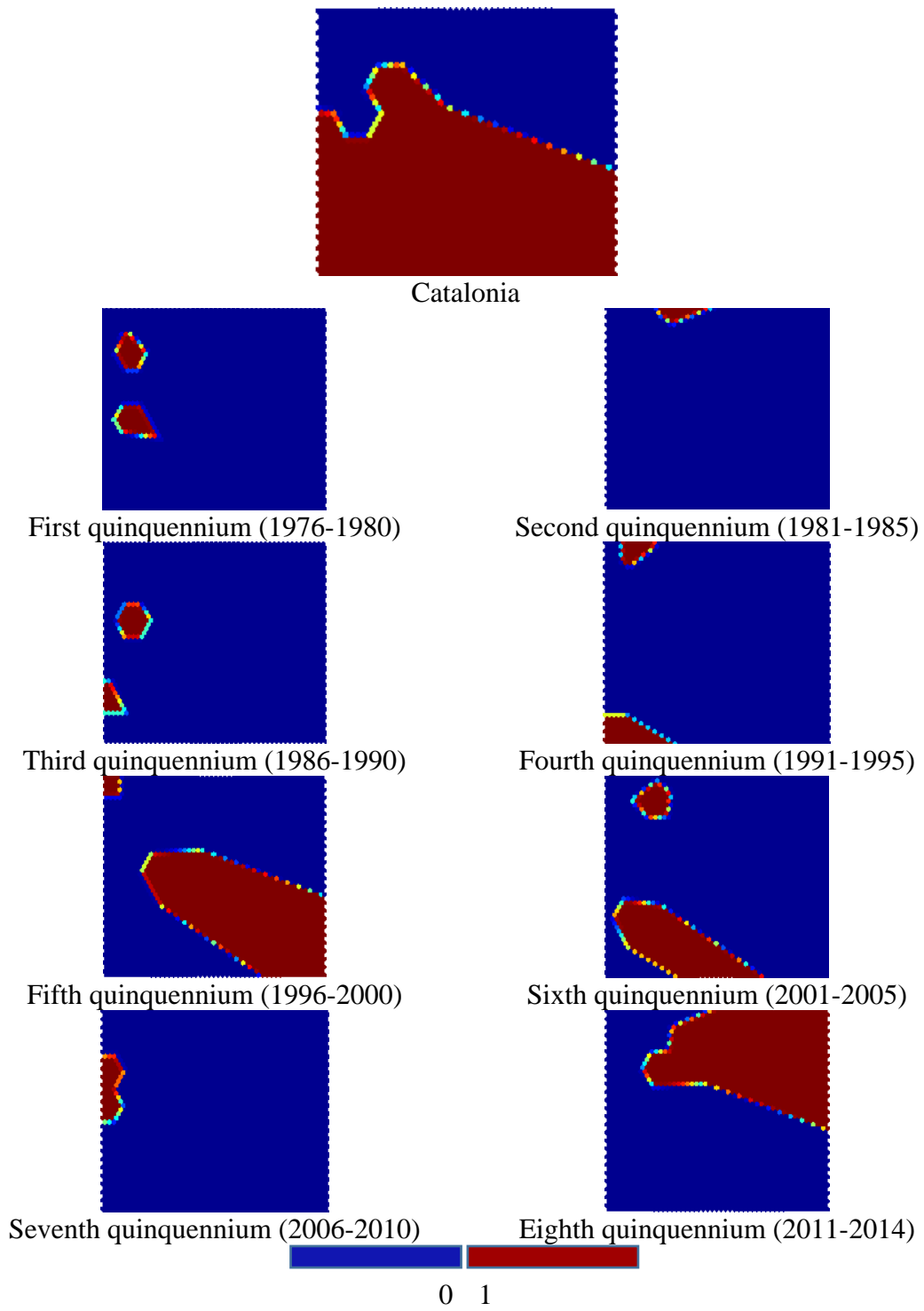
Catalonia

First quinquennium (1976-1980)   Second quinquennium (1981-1985)

Third quinquennium (1986-1990)   Fourth quinquennium (1991-1995)

Fifth quinquennium (1996-2000)   Sixth quinquennium (2001-2005)

Seventh quinquennium (2006-2010)   Eighth quinquennium (2011-2014)

0   1

Figure 7. Self-organized map trained for the autonomous community of Catalonia and the five-year periods of production.

Discussion

The analysis of doctoral theses defended within a specific area of knowledge, such as the Teaching Social Sciences in Spain, allows updating the understanding of the state of the art of the discipline and knowing the production by years or thematic categories. These types of studies are especially important in young or underdeveloped areas, as they act as a guide for the research agenda, and can provide guidelines for improvement.

The application of neural networks of the self-organized maps type to the visualization of the interaction between categories of analysis is a very novel methodology in Educational Sciences, allowing a practical and clear understanding of the results, especially necessary when large amounts of data are handled, since "visual metaphors" help to reveal hidden patterns in the data (Sotongo-Aguilar, Guzmán-Sánchez & Carrillo, 2002).

Given the numerous possibilities to train and obtain results using self-organized maps, it was decided to show those related to the most productive geographical area in relation to Spanish doctoral theses in Teaching Social Sciences (Catalonia) (Curiel-Marín, 2017), and with the autonomous community of Andalusia. For the theses carried out in Andalusia, the most studied educational stages are Primary Education and Secondary Education. For those theses carried out in Catalonia, the most studied educational stages are Secondary Education and Higher Education. In both cases, Early Childhood Education stands out as the least studied.

Regarding research by thematic categories, in Andalusia stand out those related to Didactics of Geography, Civic Education, Environmental Education and Mathematics Education. In Catalonia, we find the research topics very balanced, highlighting the categories of Teaching Social Sciences, Didactics of History, Didactics of the History of Art, Didactics of Geography, Didactics of Heritage and Civic Education.

It is concluded that both the general production and the most studied categories overlap in Catalan research. The most successful quinquennium in Catalonia were the fourth quinquennium (1991-1995), the fifth quinquennium (1996-2000), the sixth quinquennium (1996-2000), and the seventh quinquennium (2006-2010). For Andalusia, the highest productivity took place in the third quinquennium (1986-1990) and in the sixth quinquennium (2001-2005).

It is therefore necessary to improve the doctoral research related to Teaching Social Sciences in Early Childhood Education and Higher Education in Andalusia, as well as an increase in studies related to Didactics of History, Didactics of Art History and Heritage Didactics. On the other hand, in Catalonia, to balance the studies even more, it would be interesting to expand the research in Infant and Primary Education.

The biggest challenge in bibliometric or scientometric research, especially in those related to doctoral theses, is the lack of agreement in the nomenclature of the areas, the key words, and in the case of Teaching Social Sciences, the area. The homogeneity in the use of thematic areas or keywords, as in the case of the name and surnames when signing academic publications, it is imperative to make possible the recovery of documents.

Another critical conclusion of this research is the possibility of improving the national database of doctoral theses, TESEO, whose instability in use hinders the recovery of information, reaching different results in searches performed on the same day, with the same criteria; as well as the null portability of the information generated, which makes it difficult to operate.

References

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, University of Granada.

Curiel-Marín, E. (2013). Análisis cienciométrico de tesis doctorales españolas en didáctica de las ciencias sociales (1976-2012). Master Thesis, University of Granada. http://digibug.ugr.es/bitstream/10481/33202/1/CurielMarin_TFM.pdf

Curiel-Marín, E. (2017) *Análisis cienciométrico de tesis doctorales españolas en Didáctica de las Ciencias Sociales 1976-2014*. Universidad de Granada. [http://hdl.handle.net/10481/46522]

Curiel-Marín, E. & Fernández-Cano, A. (2015). Análisis cienciométrico 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

Kohonen, T.S., Kaski, K., Lagus, J., Salojarvi, J., Honkela, V., & Paatero, V. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks, 11*(3), 574–585. 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). Retrieved from http://www.ugr.es/~recfpro/rev81ART4.pdf

Mathworks (2013). *MATLAB (Computer Software)*. 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*, Special issue: 71-79. Retrieved from: http://www.qqml.net/papers/Special_Issue_January_2015_Bibliometrics/4S8QQML_Journal_2015_SpecialIssueBibliometrics_MeirelesandCendon_71-79.pdf

Nielsen, S. E. & Yezierski, 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

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 Bioingenieria: Caracterización de imágenes estáticas y dinámicas* (Doctoral Thesis). National University of 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. In 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*. Retrieved from 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: Visualizacuón 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 cienciométrico, conceptual y metodológico de las tesis doctorales españolas en Educación Matemática (1976-1998).* Servicio de Publicaciones de la Universidad de Córdoba (Spain).

Vallejo, M. (2005). *Estudio longitudinal de la producción de tesis doctorales en Educación Matemática (1975-2002).* (Doctoral Thesis). University of 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. In A, Maz, B. Gómez & M. Torralbo (Eds.), *Actas del Noveno Simposio de la Sociedad Española de Educación Matemática SEIEM* (pp. 163-174). Cordoba (Spain): 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

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

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

Authors / Autores	To know more / Saber más
Curiel-Marín, Elvira (ecuriel@ugr.es). Assistant Professor of the Department of Research Methods and Diagnosis in Education at University of Granada, member of the research group HUM-567 of Research Evaluation and Andalusian Programs Evaluation. She is PhD in Educational Sciences from University of Granada, focusing her current work on developing methodological tools that facilitate the understanding and transparency of scientometric evaluations, as well as the application of mixed methods to them. She is the contact author. Postal Address: Department of Research and Diagnosis Methods in Education, Faculty of Education, Economics and Technology, University of Granada. C / Cortadura del Valle S/N, Ceuta (Spain).	ORCID 0000-0002-0014-3971 ResearchGate Google académico
Passoni, Isabel (lpassoni@fi.mdp.edu.ar). PhD in Engineering, Electronic orientation and Magister in University Management. She works as a full professor at the National University of Mar del Plata in the Department of Electronics and Computing Engineering of the Faculty of Engineering. Director of the Bioengineering Laboratory of the Institute of Scientific and Technological Research in Electronics (ICYTE). Her areas of interest in R & D are focused on the development of Computational Intelligence techniques for decision support systems. Postal Address: Faculty of Engineering. Department of Electronic Engineering and Computing. Avd. Juan B Justo 4302. Cp: 7600. Mar del Plata (Argentina).	ORCID 0000-0001-5923-597X ResearchGate Google académico
Olmedo-Moreno, Eva M. (emolmedo@ugr.es). PhD in Pedagogy from the University of Granada, Professor of the Department of Research Methods and Diagnosis in Education of the University of Granada, where she teaches the core subject of Educational Research Methods. Her main line of research deals with "Learning Strategies and the development of hybrid learning models in the culture of Hyper-globalization-Smart Cities". She is Head of the MIDE Department and coordinator of the Doctoral Program in Educational Sciences, as well as editor of RELIEVE. Postal Adress: Campus Cartuja S/N Educational Sciences Faculty. Dept. MIDE. 18071 Granada (Spain)	ORCID 0000-0003-0558-1513 ResearchGate Google académico
Fernández-Cano, Antonio (afcano@ugr.es). PhD in Philosophy and Educational Sciences. Full Professor and Chairman of the Department of Research and Diagnosis Methods in Education at University of Granada. His specific areas of interest and specialization are Research Evaluation, pedagogical scientometrics, the study of doctoral theses and complementary methodologies for the evaluation of educational programs. Postal Adress: Campus Cartuja S/N Educational Sciences Faculty. Dept. MIDE. 18071 Granada (Spain)	ORCID 0000-0003-3991-4443 ResearchGate Google académico

RELIEVE

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]