

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL
(FOR EVALUATION OF THE ACT DOCTORAL THESIS)

Año académico (academic year): 2016/17

DOCTORANDO (candidate PHD): **CANTOS MATEOS, GISELA**

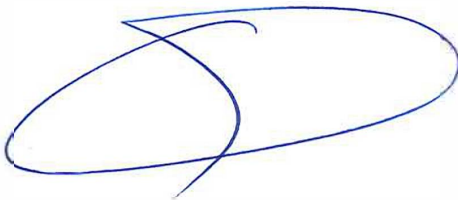
PROGRAMA DE DOCTORADO (Academic Committee of the Programme): **D329 DOCTORADO EN DOCUMENTACIÓN**
DEPARTAMENTO DE (Department): **FILOLOGÍA, COMUNICACIÓN Y DOCUMENTACIÓN**
TITULACIÓN DE DOCTOR EN (Phd title): **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 15/09/17, reunido el tribunal de evaluación, constituido por los miembros que suscriben el presente Acta, el aspirante defendió su Tesis Doctoral **con Mención Internacional** (In today assessment met the court, consisting of the members who signed this Act, the candidate defended his doctoral thesis with mention as International Doctorate), elaborada bajo la dirección de (prepared under the direction of) **FÉLIX DE MOYA ANEGÓN // M. ÁNGELES ZULUETA GARCÍA BENJAMÍN VARGAS QUESADA.**

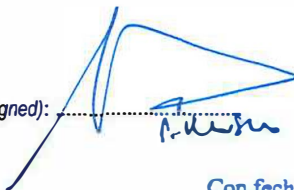
Sobre el siguiente tema (Title of the doctoral thesis): **LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE DOMINIOS CIENTÍFICOS.**

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL¹ de **(no apto, aprobado, notable y sobresaliente)** (After the defense and defense of the thesis, the court agreed to grant the GLOBAL RATING (fail, pass, good and excellent): **SOBRESALIENTE**

Alcalá de Henares, a 15 de Septiembre de 2017



Fdo. (Signed): ELENA GRANDE



Fdo. (Signed): ANTONIO PEÑALVES



FIRMA DEL ALUMNO (candidate's signature),



Fdo. (Signed): GISELA CANTOS MATEOS

Con fecha 4 de octubre de 2017 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

- Conceder la Mención de "Cum Laude"
 No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada



¹ La calificación podrá ser "no apto" "aprobado" "notable" y "sobresaliente". El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad. (The grade may be "fail" "pass" "good" or "excellent". The panel may confer the distinction of "cum laude" if the overall grade is "Excellent" and has been awarded unanimously as such after secret voting.)

INCIDENCIAS / OBSERVACIONES:
(Incidents

/

Comments)



Universidad
de Alcalá

COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 4 de octubre, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *CANTOS MATEOS, GISELA*, el día 15 de septiembre de 2017, titulada *LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE DOMINIOS CIENTÍFICOS.*, para determinar, si a la misma, se le concede la mención “cum laude”, arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado **resuelve otorgar** a dicha tesis la

MENCIÓN “CUM LAUDE”

Alcalá de Henares, 10 de octubre de 2017

EL PRESIDENTE DE LA COMISIÓN DE ESTUDIOS
OFICIALES DE POSGRADO Y DOCTORADO



Juan Ramón Velasco Pérez

Copia por e-mail a:

Doctorando: CANTOS MATEOS, GISELA

Secretario del Tribunal: ELENA CORERA ÁLVAREZ.

Directores de Tesis: FÉLIX DE MOYA ANEGÓN // M. ÁNGELES ZULUETA GARCÍA// BENJAMÍN VARGAS QUESADA



Universidad
de Alcalá

ESCUELA DE DOCTORADO
Servicio de Estudios Oficiales de
Posgrado

DILIGENCIA DE DEPÓSITO DE TESIS.

Comprobado que el expediente académico de D./D^a _____
reúne los requisitos exigidos para la presentación de la Tesis, de acuerdo a la normativa vigente, y habiendo
presentado la misma en formato: soporte electrónico impreso en papel, para el depósito de la
misma, en el Servicio de Estudios Oficiales de Posgrado, con el nº de páginas: _____ se procede, con
fecha de hoy a registrar el depósito de la tesis.

Alcalá de Henares a _____ de _____ de 20 _____



Fdo. El Funcionario



Programa de Doctorado en Documentación (D329)

**LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES
LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE
CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES**

**PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE
DOMINIOS CIENTÍFICOS**

Tesis Doctoral presentada por

GISELA CANTOS MATEOS

Directores:

Dr. Félix de Moya Anegón

Dra. María Ángeles Zulueta García

Dr. Benjamín Vargas Quesada

Alcalá de Henares, 2017

D. Manuel PÉREZ JIMÉNEZ
DIRECTOR DEL DEPARTAMENTO DE FILOLOGÍA, COMUNICACIÓN Y
DOCUMENTACIÓN DE LA UNIVERSIDAD DE ALCALÁ,

H A C E C O N S T A R

Que la Tesis Doctoral presentada por **D^a. Gisela Cantos Mateos**, titulada **“LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE DOMINIOS CIENTÍFICOS”**, bajo la dirección de los Directores D. Félix de Moya Anegón, D^a. M. Ángeles Zulueta García y D. Benjamín Vargas Quesada, reúne las condiciones científicas necesarias para su presentación y defensa en este Departamento de Filología, Comunicación y Documentación de la Universidad de Alcalá.

Y para que conste donde convenga, a los efectos oportunos, firmo el presente en Alcalá de Henares, a diecinueve de abril de dos mil diecisiete.

Fdo.: Manuel Pérez Jiménez

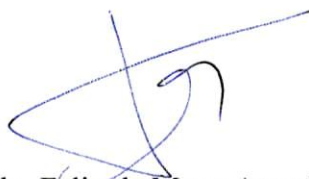


D. Felix de Moya Anegón
PROFESOR DE INVESTIGACIÓN DEL CONSEJO SUPERIOR DE
INVESTIGACIONES CIENTÍFICAS

H A C E C O N S T A R

Como Director de la Tesis Doctoral de **Dña. Gisela Cantos Mateos**, titulada **“LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE DOMINIOS CIENTÍFICOS”**, que este Trabajo de Investigación reúne las condiciones científicas necesarias para su presentación y defensa en el Departamento de Filología, Comunicación y Documentación.

Y para que conste donde convenga, a los efectos oportunos, firmo el presente en Granada, a 6 de abril de dos mil diez y siete



Fdo. Felix de Moya Anegón

Dña. M^a Angeles Zulueta García
Catedrática de la Universidad de Alcalá

H A C E C O N S T A R

Como Director/a de la Tesis Doctoral de **Dña. Gisela Cantos Mateos**, titulada **LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACIÓN TEMÁTICA DE DOMINIOS CIENTÍFICOS**, que este Trabajo de Investigación reúne las condiciones científicas necesarias para su presentación y defensa en el Departamento de Filología, Comunicación y Documentación.

Y para que conste donde convenga, a los efectos oportunos, firmo el presente en Alcalá de Henares, a 4 de abril de dos mil diez y siete



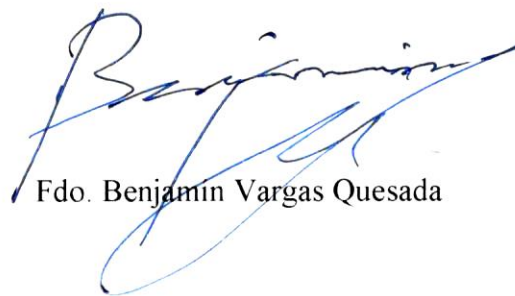
Fdo. M^a Angeles Zulueta

D. Benjamín Vargas Quesada
Profesor Titular de la Universidad de Granada

H A C E C O N S T A R

Como Director de la Tesis Doctoral de **Dña. Gisela Cantos Mateos**, titulada **LOCALIZACIÓN Y VISUALIZACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN A TRAVÉS DEL ANÁLISIS DE CO-PALABRAS Y DEL ANÁLISIS DE REDES SOCIALES. PROPUESTA METODOLÓGICA PARA LA DELIMITACION TEMÁTICA DE DOMINIOS CIENTÍFICOS**", que este Trabajo de Investigación reúne las condiciones científicas necesarias para su presentación y defensa en el Departamento de Filología, Comunicación y Documentación.

Y para que conste donde convenga, a los efectos oportunos, firmo el presente en Granada, a 4 de abril de dos mil diez y siete



Fdo. Benjamín Vargas Quesada

A Dani, por ser lo mejor de mi vida

“Las palabras son como monedas, que una vale por muchas como muchas no valen por una” (Francisco de Quevedo)

ÍNDICE GENERAL

I. PREFACIO	XVII
II. AGRADECIMIENTOS	XIX
III. ÍNDICE DE TABLAS	XXI
IV. RESUMEN	XXIII
V. ABSTRACT.....	XXVII

PARTE I: MARCO GENERAL DE LA INVESTIGACIÓN

1. INTRODUCCIÓN.....	3
2. JUSTIFICACIÓN Y OBJETIVOS.....	7
2.1 Problema de investigación	7
2.2 Objetivo principal del estudio	10
2.3 Objetivos específicos.....	11
3. DELIMITACIÓN DEL ESTUDIO	19
3.1 Estructura de la tesis	19
3.2 Dominios científicos analizados	22
3.2.1. “Dominio 1”	24
3.2.2. “Dominio 2”	25
4. ANTECEDENTES	31
4.1 Análisis de Dominios	33
4.2 Visualización y representación de la información.....	39
4.3 Mapeo de la ciencia.....	43
4.4 Análisis de Redes Sociales	52
4.5 Delimitación temática	55
4.5.1 Emparejamiento bibliográfico.....	56
4.5.2 Co-citación de Autores	56
4.5.3 Co-citación de categorías temáticas	57
4.6 Análisis de co-palabras	58
4.6.1 Evolución histórica	58
4.6.2 Perspectivas de la investigación en estos estudios.....	60
4.6.3 Ventajas e inconvenientes del análisis de co-palabras para la delimitación temática	63
5. LIMITACIONES	67

6. MATERIALES Y MÉTODOS	71
6.1 Materiales.....	71
6.1.1 Fuentes de información	71
6.1.2 Recogida y tratamiento de la información.....	75
6.2 Métodos	83
6.2.1 Unidades de análisis: términos de indización	83
6.2.2 Unidades de medida: análisis de co-palabras	88
6.2.3 Reducción del espacio n-dimensional y visualización de la información.....	89
7. RESULTADOS Y DISCUSIÓN.....	103
7.1. Localización y visualización de las principales líneas de investigación de la producción científica correspondiente al Dominio 1. (<i>Publicación 1</i>).....	103
7.1.1. Detección de las principales líneas de investigación sobre la producción científica sobre salud y mujer (1965-2005) a través de los de los descriptores MeSH: estudio evolutivo (<i>Publicación 1</i>).	103
7.2. Localización y visualización de las principales líneas de investigación de la producción científica correspondiente al Dominio 2.	112
7.2.1. Detección de las principales líneas de investigación sobre la producción científica española en células madre (1997-2007) a través de los KeyWords Plus. (<i>Publicaciones 2 y 3</i>).	112
7.2.2. Detección de las principales líneas de investigación sobre la producción española en células madre (1997-2010) a través de los KeyWords Plus: estudio evolutivo. (<i>Publicaciones 4 y 5</i>).	120
7.2.3. Detección de las principales líneas de investigación sobre la producción española en células madre (1997-2010) a través de los términos de indización del SCI y Medline: estudio comparativo. (<i>Publicación 6</i>).....	126
8. CONCLUSIONES	135
9. CONCLUSIONS	139
10. PERSPECTIVAS DE INVESTIGACIONES FUTURAS	143
11. BIBLIOGRAFÍA.....	147

PARTE II: PUBLICACIONES CIENTÍFICAS

PUBLICACIÓN Nº 1	179
PUBLICACIÓN Nº 2	209
PUBLICACIÓN Nº 3	213
PUBLICACIÓN Nº 4	247
PUBLICACIÓN Nº 5	265
PUBLICACIÓN Nº 6	281

I. PREFACIO

Esta tesis doctoral se ha llevado a cabo gracias a la financiación otorgada por el Consejo Superior de Investigaciones Científicas (CSIC) a través del programa *Junta para la Ampliación de Estudios (JAE)*, destinado a la formación de personal investigador y técnico. Dentro de este programa, el subprograma *JAEPredoctoral* ofrece becas de cuatro años de duración para el desarrollo de tesis doctorales en líneas estratégicas de los centros que integran el CSIC. En concreto, esta tesis se ha desarrollado en el Instituto de Políticas y Bienes Públicos (IPP), del Centro de Ciencias Humanas y Sociales.

II. AGRADECIMIENTOS

Durante el tiempo que ha transcurrido en la elaboración de esta tesis doctoral, son muchas las personas que han contribuido enormemente para que esto fuera posible. En primer lugar, querría expresar mi más profundo agradecimiento a mis directores, el Dr. Félix de Moya Anegón, la Dra. M. Ángeles Zulueta y el Dr. Benjamín Vargas Quesada por su gran apoyo y confianza, además de brindarme la oportunidad de realizar esta tesis. En el terreno académico, sus ideas, recomendaciones, conocimientos y experiencia han mejorado la calidad de esta investigación, pero sin duda, en el terreno personal, han sido capaces de contagiarme su entusiasmo y sus ganas de trabajar más y mejor. Sobre todo, en el caso de la Dra. M. Ángeles Zulueta cuya relación ha traspasado los límites de cordialidad y amistad teniendo la gran satisfacción de haber sido una segunda madre para mí.

Del mismo modo, quiero destacar el privilegio de haber conocido a una gran científica, pero, más aún, grandísima persona como es Zaida Chinchilla-Rodríguez que, gracias a su guía, asesoramiento y su gran humanidad he tenido la oportunidad de aprender muchísimo y poder colaborar con ella en multitud de proyectos.

Deseo agradecer, asimismo, al personal investigador del Instituto de Políticas y Bienes Públicos del CSIC sus recomendaciones y consejos durante el desarrollo de la tesis doctoral, así como, a mis compañeros de despacho por haber compartido tantos cafés, nervios, risas, preocupaciones y muchos otros momentos inolvidables.

También quiero mostrar mi agradecimiento a los miembros que he tenido el privilegio de conocer del grupo SCImago por sus recomendaciones y consejos especialmente, a Elena Corera y a todo su equipo de normalización.

Asimismo, me gustaría mencionar a los profesores de la Universidad de Alcalá, investigadores y estudiantes que he tenido la oportunidad de conocer durante esta etapa, y cuyos consejos y comentarios me han permitido mejorar, pero, sobre todo, aprender. No quisiera olvidarme de mis compañeras de la licenciatura de Documentación (Cecilia, Cristina, Avelina y Mercedes) que, sin duda, supieron comprenderme y animarme de manera incondicional.

En el plano más personal, quiero destacar la labor de mi “gabinete de crisis” particular compuesto por mis amigas Raquel, Cristina, Maru y Aroa, por tantas tardes de terapia colectiva que me hicieron ver el lado positivo de las cosas. Pero, especialmente, quiero agradecer este trabajo a toda mi familia. Especialmente, a mis suegros Julián y M^a Eugenia por sacar en todo momento lo mejor de mí. A mis hermanas Oché y Olivia por todos los ánimos, empujones y, sobre todo, el gran apoyo que, sin duda, han supuesto para mí, no se puede tener mejores referentes. A mis padres, Andrés y María a los que les debo todo lo que soy y de los que he aprendido las lecciones más importantes y más enriquecedoras de mi vida y a los que el agradecimiento se hace infinito. Y, sobre todo, a mi marido Dani que, gracias a su inagotable paciencia, amor, comprensión y un largo etcétera de virtudes de las que cada día me sorprende más, desde luego no hubiera sido posible la culminación de esta tesis.

III. ÍNDICE DE TABLAS

Tabla 1. Resumen de los objetivos fundamentales que originan las publicaciones	13
Tabla 2. Características de los dominios analizados	23
Tabla 3. Publicaciones donde son tratados los dominios	29
Tabla 4. Datos relativos al proceso de normalización de direcciones españolas	78
Tabla 5. Criterios de codificación y normalización de las direcciones españolas	79
Tabla 6. Tratamiento de descriptores MeSH procedentes de Medline	81
Tabla 7. Tratamiento de los términos procedentes del SCI-EXPANDED	82
Tabla 8. Datos relativos a la normalización de los Autores	83
Tabla 9. Resumen de los parámetros metodológicos utilizados en las publicaciones ..	87
Tabla 10. Características de los dominios analizados	101
Tabla 11. Términos de indización analizados	127
Tabla 12. Frecuencias relativas a los términos de indización analizados	129

IV. RESUMEN

La ciencia se caracteriza cada vez más por su multi e interdisciplinariedad. Esto dificulta la demarcación de los distintos ámbitos del conocimiento y complica la delimitación, definición y distribución de las diferentes áreas, categorías o aspectos temáticos involucrados en el desarrollo científico. Desde el punto de vista bibliométrico, este hecho plantea un gran reto metodológico, sobre todo, en la delimitación temática y en la identificación y localización de los diferentes aspectos temáticos que caracterizan cualquier dominio científico o a la ciencia en general. Especialmente, en niveles de agregación muy concretos o en dominios caracterizados por una gran especificidad temática. Más aún, cuando las posibilidades que ofrecen actualmente, la mayoría de las bases de datos bibliográficas para localizar y analizar la producción científica relativa a aspectos temáticos más específicos que el nivel de áreas, categorías o disciplinas científicas es prácticamente inexistente.

El objetivo principal de esta tesis doctoral se centra en el diseño de una propuesta metodológica que permita la localización, identificación y visualización de las principales líneas de investigación de un dominio científico determinado, con independencia de su amplitud o especificidad temática. Dicha propuesta pretende posibilitar una delimitación temática tanto de carácter general como de carácter más específico, pudiendo ser aplicable tanto para estudios de granularidad gruesa como fina. Por este motivo esta propuesta se aplica a dos dominios claramente diferenciados desde el punto de vista de su cobertura temporal, geográfica y cronológica. El “Dominio 1” se basa en la investigación recogida en la base de datos Medline sobre salud y mujer durante el período 1965-2005. El “Dominio 2”, en cambio, se centra en la investigación española con células madre recogida en la base de datos Science Citation Expanded durante los años 1997-2012. De este modo, los mayores retos a los que se enfrenta esta propuesta metodológica son el de descender a niveles de análisis muy específicos y especializados como es la identificación de las subdisciplinas científicas, líneas de investigación, sublíneas de investigación, etc., y el de la identificación de los aspectos temáticos tanto de carácter más consolidado como los de carácter más dinámico.

La propuesta metodológica que se presenta en esta tesis doctoral se centra en la combinación de varias técnicas bibliométricas. En primer lugar, se basa en el empleo del análisis de co-palabras para la identificación de los términos más representativos a partir de su aparición conjunta en los documentos (*Co-words Analysis*). Concretamente, se han utilizado como unidades de análisis los términos de indización que emplean las bases de datos bibliográficas que se han seleccionado en este estudio como fuentes de información. Es decir, los descriptores MeSH en el caso de Medline y los términos clave *KeyWords Plus* y *Author Keywords* en el caso del Science Citation Index Expanded. En segundo lugar, se han combinado técnicas bibliométricas procedentes del Análisis de Redes Sociales y de la representación y visualización de la información, fundamentalmente, mediante el uso de dos programas informáticos especializados como son *Pajek* y *VOSviewer*. En el caso del primero, éste se centra en la generación de mapas basados en grafos que en esta tesis han sido simplificados a través del algoritmo de poda *Pathfinder Networks* (PFNETs) y representados mediante el algoritmo de visualización *Kamada Kawai*. En el caso de *VOSviewer* éste se centra en la generación de mapas basados en la distancia a través de la *visualización de similitudes* (*visualization of similarities – VOS*) y que utiliza diferentes técnicas de *clustering* en la representación gráfica. Dada las ventajas y limitaciones que conllevan el uso de una técnica frente a la otra se han utilizado ambos con la intención de que puedan ser utilizados de manera complementaria.

Los resultados que se han obtenido en esta tesis doctoral han sido presentados en las diferentes publicaciones que la integran y que son consecuencia de las distintas fases de trabajo que fueron planteadas. Concretamente, los resultados muestran que ha sido posible la localización y visualización de los principales aspectos temáticos de los dos dominios científicos analizados. En el caso del “Dominio 1” se obtuvieron cuatro mapas en función de los períodos temporales seleccionados (1965-1974; 1975-1984; 1985-1994 y 1995-2005). En ellos, se pudo observar que la orientación de la investigación en las primeras etapas, se decanta, fundamentalmente, por aquellos aspectos de la salud de las mujeres que tienen que ver con la reproducción. En cambio, a medida que se van sucediendo los años, se aprecia que la investigación bascula hacia otros ámbitos más

heterogéneos y diversos como son los aspectos sociales, laborales y psicológicos que también afectan a la salud de las mujeres. En el caso del “Dominio 2” los resultados permitieron distinguir claramente la investigación clínica y la básica. Además, permitieron la localización e identificación de hasta cuatro líneas de investigación como son las relacionadas con: el uso terapéutico de las células madre en enfermedades hematológicas; las células madre hematopoyéticas; las células madre embrionarias, concretamente, con los procesos de expresión (*Expression*) y diferenciación celular (*Differentiation*), así como, los procesos y las biotecnologías necesarias para localizar, producir, crecer y, sobre todo, analizar in-vitro las células madre embrionarias; y finalmente, las células madre neurales.

Los resultados alcanzados en esta tesis permiten concluir que la propuesta metodológica planteada es adecuada para la localización, identificación y visualización de las principales tendencias o líneas de investigación que caracterizan la producción científica, con independencia de la amplitud o especificidad del dominio utilizado y de la cobertura de las fuentes de información empleadas. Su adecuación metodológica para la delimitación temática la convierten en una herramienta con un gran potencial de cara al desarrollo de nuevos estudios bibliométricos en posibles investigaciones futuras. No obstante, y pese a ello, también esta propuesta metodológica puede ser objeto de mejora. En el futuro, podría ser interesante emplear otras técnicas de análisis como el Análisis de co-citación de Autores (ACA) o el empleo de otros términos como pueden ser los procedentes del título, resumen, palabras clave, etc., como unidades de análisis.

Palabras Clave

Bibliometría, Delimitación temática, Análisis de co-words, Visualización de la información, Análisis de redes, Células madre, Salud y Mujer, Pajek, VOSviewer y PFNET.

V. ABSTRACT

Science is increasingly characterized by its multi and interdisciplinary nature. This complicates the demarcation of the different areas of knowledge and complicates the delimitation, definition and distribution of the different areas, categories or thematic aspects involved in scientific development. From the bibliometric point of view, this fact presents a great methodological challenge, above all, in the thematic delimitation and in the identification and location of the different thematic aspects that characterize any scientific domain or to science in general. Especially, at very specific levels of aggregation or in domains characterized by a great thematic specificity. Moreover, when the possibilities currently offered, most of the bibliographic databases to locate and analyze the scientific production related to thematic aspects more specific than the level of scientific areas, categories or disciplines is practically nonexistent.

The main objective of this doctoral thesis focuses on the design of a methodological proposal that allows the localization, identification and visualization of the main research lines of a given scientific domain, regardless of its amplitude or specificity thematic. This proposal aims to enable a thematic delimitation of both a general and a more specific character, which may be applicable for both gross and fine granularity studies. For this reason, this proposal applies to two domains clearly differentiated from the point of view of their temporal, geographical and chronological coverage. "Domain 1" is based on research from the Medline database on health and women during the period 1965-2005. On the other hand, the "Domain 2" focuses on Spanish stem cell research collected in the Science Citation Expanded database during the years 1997-2012. Thus, the major challenges facing this methodological proposal are that of descending to very specific and specialized levels of analysis, such as the identification of scientific subdisciplines, lines of research, sub-lines of research, etc., and the identification of the thematic aspects, both of a more consolidated nature and those of a more dynamic nature.

The methodological proposal presented in this doctoral thesis focuses on the combination of several bibliometric techniques. First, it is based on the use of Co-words Analysis for the identification of the most representative terms from their joint appearance in the documents. Specifically, the indexing terms used by the bibliographic databases that have been selected in this study as sources of information have been used as units of analysis. That is, the descriptors *MeSH* in the case of Medline and the *KeyWords Plus* and *Author Keywords* in the case of the Science Citation Index Expanded. Secondly, bibliometric techniques have been combined from the analysis of social networks and from the representation and visualization of information, mainly using two specialized computer programs such as *Pajek* and *VOSviewer*. In the case of the first, it focuses on the generation of maps based on graphs that in this thesis have been simplified through the *Pathfinder Networks* pruning algorithm (*PFNETs*) and represented by the visualization algorithm *Kamada Kawai*. In the case of *VOSviewer* this one focuses on the generation of maps based on the distance through *the visualization of similarities (VOS)* and that uses different techniques of clustering in the graphical representation. Given the advantages and limitations involved in the use of one technique over the other, both have been used with the intention that they can be used in a complementary way.

The results obtained in this doctoral thesis have been presented in the different publications that integrate it and are a consequence of the different phases of study that were proposed. Specifically, the results show that it has been possible to locate and visualise the main thematic aspects of the two scientific domains analysed. In the case of “Domain 1”, four maps were obtained according to the selected time periods (1965-1974, 1975-1984, 1985-1994 and 1995-2005). In them, it could be observed that the orientation of research in the early stages, essentially, focuses on those aspects of the health of women that must do with reproduction. However, as the years go by, research is moving towards other more heterogeneous and diverse areas such as the social, labour and psychological aspects that also affect the health of women. In the case of “Domain 2”, the results allowed a clear distinction between clinical and basic research. In addition, they allowed the localization and identification of up to four research lines

such as: the therapeutic use of stem cells in haematological diseases; hematopoietic stem cells; the embryonic stem cells, specifically, the processes of expression and cell differentiation, as well as the processes and biotechnologies necessary to locate, produce, grow and, above all, to analyse in vitro embryonic stem cells; and finally, neural stem cells.

The results obtained in this thesis allow to conclude that the proposed methodological proposal is adequate for locating, identifying and visualizing the main trends or lines of research that characterize scientific production, regardless of the breadth or specificity of the domain used and the coverage of the information sources used. Its methodological suitability for thematic delimitation makes it a tool with great potential for the development of new bibliometric studies in possible future research. However, despite this, this methodological proposal can also be improved. In the future, it may be interesting to use other analytical techniques such as the Author Co-citation Analysis (ACA) or the use of other terms such as titles, abstract, keywords, etc., as units of analysis.

KEYWORDS

Bibliometric analysis, Thematic delimitation, Co-word analysis, Information visualization, Social networks, Stem cells, Health Women, Pajek, VOSviewer and PFNET.

PARTE I: MARCO GENERAL DE LA INVESTIGACIÓN

1. INTRODUCCIÓN

El uso de los estudios bibliométricos para la comprensión y análisis de dominios científicos (Hjørland y Albrechtsen, 1995), junto con el desarrollo y perfeccionamiento de nuevas técnicas y herramientas, facilita la toma de decisiones en materia de política científica y refleja el estado en el que se encuentra la investigación en un momento determinado. Estos procesos son básicos para evaluar la Ciencia, compromiso que ningún país puede eludir (Krauskopf, 2000) dadas las evidentes conexiones entre los avances procedentes de la actividad investigadora y el empleo, el crecimiento y progreso económico y social que mejoran el bienestar de la sociedad, en todos los avances biomédicos en particular y en su conjunto y por extensión, con la necesidad de crear una sociedad más competitiva (Chinchilla- Rodríguez & Moya-Anegón, 2007).

Por otro lado, el gran número de disciplinas científicas que hoy en día la investigación interrelaciona, dota a esta de un carácter muy interesante al tiempo que controvertido. La multidisciplinariedad plantea un gran reto a la hora de delimitar y analizar la composición temática de cualquier dominio científico, exigiendo para ello un análisis muy exhaustivo y riguroso. Precisamente, para superar este reto, la bibliometría cuenta con herramientas complementarias entre las que se han incorporado el Análisis de Redes Sociales (Wasserman y Faust, 1998) como una de las técnicas de análisis y visualización de dominios científicos (Leydesdorff & Rafols, 2009; Vargas-Quesada Chinchilla-Rodríguez, González-Molina & Moya-Anegón, 2010; Rafols, Porter & Leydesdorff, 2010).

Para la elaboración de este tipo de estudios, las fuentes de información que tradicionalmente se vienen utilizando son las bases de datos bibliográficas. Éstas ofrecen información potencialmente útil para la explotación bibliométrica, aun cuando no han sido diseñadas para este fin. Sin embargo, presentan limitaciones a la hora de resolver cuestiones de carácter específico como, por ejemplo, la definición y delimitación concreta de un tema. Para ello, es preciso desvelar la estructura semántica que se

establece entre los documentos a través de la información bibliográfica que recogen las bases de datos.

El *Science Citation Index*, clasifica las revistas agrupándolas por categorías temáticas. Éstas, ofrecen una aproximación muy general a temas que se han establecido en función de la clasificación de dichas revistas y no en la información contenida en los documentos. Además, presenta otras limitaciones como la asignación de varias categorías a una misma revista, la posibilidad de cambio del ámbito de investigación de la revista, o la mutabilidad de las categorías a lo largo del tiempo según la evolución de la propia investigación. En cambio, en el caso de *Medline* la designación de la categoría temática se realiza a partir de la información contenida directamente en los documentos. Esta asignación es de carácter manual y se produce en una base de datos cuya cobertura temática está especializada en el campo de la Medicina y en la que la indización también es manual. A diferencia de la categorización de revistas, este tipo de asignación tiene la ventaja de ser más precisa ya que la clasificación se hace directamente sobre los aspectos temáticos que son tratados en los documentos. No obstante, este sistema de clasificación se enfrenta a los mismos inconvenientes que el anterior, a la hora de realizar delimitaciones temáticas en niveles de agregación mucho más específicos.

Para salvar estas limitaciones se debe recurrir a métodos de análisis que permitan llegar a las ideas y conceptos que sustentan, en última instancia, el discurso científico expresado en la literatura. Para ello, se propone el uso de unidades de análisis más pequeñas que las categorías temáticas o las revistas como son los términos de indización que, según la hipótesis de esta tesis, pueden reflejar mejor la esencia conceptual de cada documento. Éstos, ofrecen la posibilidad de llegar a aquellas ideas y conceptos que sustentan, en última instancia, el discurso científico expresado en la literatura científica. Es decir, permiten descender a las unidades de medida más pequeñas posibles que mejor reflejan la esencia conceptual de cada documento, a partir de las cuales se vertebra la temática de un dominio científico.

El comportamiento de las palabras puede ser analizado a través de las relaciones que se establecen entre ellas a través de los documentos, es decir, a través de sus co-ocurrencias, lo que se conoce como análisis de co-palabras o *Co-words Analysis*. Esta técnica permite el estudio relacional de los principales conceptos implicados en la investigación. Además, si se combina con técnicas procedentes del Análisis de Redes Sociales y con herramientas especializadas en la representación y visualización de la información científica, pueden detectarse agrupaciones de términos relacionados semánticamente y representados a modo de mapa, identificando así los principales aspectos temáticos implicados en un dominio científico dado.

A continuación, se presenta una propuesta metodológica que pretende localizar, visualizar y representar los principales aspectos temáticos que predominan en un dominio científico determinado. Tienen como objetivo principal, conseguir una delimitación temática más específica que la que ofrecen el estudio de las áreas temáticas o disciplinas científicas, con independencia de la amplitud o especificidad del dominio utilizado o el nivel de agregación al que se pretende alcanzar en el estudio. Para ello, la metodología se basa en el uso como unidades de análisis de los términos de indización, en el análisis de co-palabras como técnica de estudio para la delimitación temática y en la combinación de técnicas procedentes del Análisis de Redes Sociales y representación y visualización de la información, a través del uso de programas informáticos especializados como son *Pajek* y *VOSviewer*.

2. JUSTIFICACIÓN Y OBJETIVOS

2.1 Problema de investigación

La ciencia se caracteriza cada vez más por su multidisciplinariedad. Ésta, se presenta como un fenómeno en el que cada vez intervienen mayor número de disciplinas en la generación de nuevo conocimiento. Del mismo modo, no sólo aumenta el número de disciplinas, sino que, la interrelación entre ellas también aumenta. Dado que el mundo se caracteriza por ser multifacético, multidimensional y multi-fenoménico, una de sus consecuencias es la gran diversidad temática que define a los diferentes dominios del conocimiento.

Desde este punto de vista, la visión holística de la ciencia que caracteriza al Análisis de Dominios y que se define más pormenorizadamente en el apartado 4. Antecedentes, tiene en cuenta la interacción y conexión de los dominios científicos. Parte de la premisa de que, para una aproximación o análisis de cualquier dominio es necesario hacerlo desde un enfoque general teniendo en cuenta las prácticas de los distintos campos del conocimiento. Esta interdisciplinariedad plantea dificultades a la hora de delimitar y establecer fronteras entre los diferentes aspectos temáticos involucrados en el desarrollo científico. Las disciplinas científicas no siguen un mismo comportamiento a la hora de desarrollarse. No todas se originan al mismo tiempo ni de la misma manera. La dinamicidad de la ciencia presente en el desarrollo científico, tiene como consecuencia que al mismo tiempo se encuentren disciplinas que estén consolidándose, otras que estén desapareciendo, otras que comienzan a originarse, otras que tienen un desarrollo mayor, etc. Además, la interdisciplinariedad no se produce con la misma intensidad entre todas las disciplinas científicas. En algunas se produce más intensamente y en otras puede estar prácticamente ausente (López Ferrer, Velasco Arroyo, Osca Lluch, & Peñaranda Ortega, 2009). No obstante, la posibilidad de analizar y comprender estos grados de intensidad puede ser muy interesante para evaluar la proximidad de unas disciplinas con otras o para observar vacíos de interdisciplinariedad que pueden plantear nuevos campos de conocimiento.

El aumento de esta interdisciplinariedad se debe en gran medida a la tendencia a la súper especialización científica en el avance hacia la generación de nuevo conocimiento. Aunque parezca paradójico, el desarrollo de disciplinas o líneas de investigación de carácter altamente especializado requieren de la transversalidad de conocimientos procedentes de diferentes disciplinas científicas o de otras áreas o campos científicos.

Del mismo modo, esta interacción interdisciplinar o entre los diferentes aspectos temáticos que caracterizan un dominio científico dado, evidencia un grado de relación variable que puede analizarse de manera estructural. Las interacciones temáticas pueden ser analizadas desde la perspectiva del Análisis de Redes Sociales y representadas a partir de técnicas de visualización y representación de la información descritas más adelante, obteniendo como resultado un mapa bibliométrico.

Teniendo en cuenta todo lo expuesto, la delimitación temática y la localización e identificación de los diferentes aspectos temáticos que caracterizan cualquier dominio científico o a la ciencia en general, ya sea en distintos niveles de agregación o especificidad temática, resulta una tarea compleja desde el ámbito de la Bibliometría que plantea grandes retos metodológicos.

Uno de los principales motivos consiste en las limitaciones que plantean los sistemas de categorización que ofrecen las bases de datos bibliográficas. Generalmente, aquellas bases de datos que tienen un carácter multidisciplinar realizan una clasificación temática basada en la categorización de las revistas en las que son publicados los documentos. Dentro de este tipo de bases de datos se encuentran dos de las más utilizadas por la comunidad científica, como son la Web of Science (WOS)¹ (Thomson Reuters, 2016) y Scopus² (Elsevier, 2016). Ambas, emplean dos niveles de categorización temática a la hora de clasificar las revistas de publicación. Un primer nivel, de carácter más general

¹ Web of Science (2016). Disponible en <http://wokinfo.com/> [Fecha de consulta: 5/11/2016]

² Scopus (2016). Disponible en <https://www.scopus.com/> [Fecha de consulta: 5/11/2016]

definido por las áreas temáticas, y un segundo nivel compuesto por las categorías temáticas. También, existen bases de datos como Medline³ o INSPEC⁴ que, establecen clasificaciones a partir de la información contenida directamente en los documentos. Esta asignación es de carácter manual y, generalmente, se produce en bases de datos con una cobertura temática más especializada y donde la indización también suele ser manual. A diferencia de la categorización de revistas, este tipo de asignación tiene la ventaja de ser más precisa ya que la clasificación se hace directamente sobre los aspectos temáticos que son tratados en los documentos.

No obstante, ambos sistemas de clasificación temática se enfrentan a las mismas limitaciones. Como se ha expuesto, la interdisciplinariedad de la ciencia dificulta la demarcación de los distintos ámbitos del conocimiento. Esto complica la definición, distribución y delimitación de las áreas o categorías temáticas en la elaboración de la clasificación. Además, la continua transformación y evolución de las disciplinas como reflejo de la dinamicidad de la ciencia, choca con un modelo de organización temática que, por su naturaleza, no incorpora con inmediatez los cambios temporales ocasionando, sobre todo, graves problemas en la movilidad y renombramiento de las categorías. Del mismo modo, la múltiple adscripción de categorías a un documento o a una revista, aumenta el solapamiento documental a la hora de aplicar análisis bibliométricos.

Por otro lado, las bases de datos que emplean estos sistemas de clasificación ofrecen una forma aislada, dispersa y fraccionada de buscar información de carácter temático y tienden a ser excluyentes en la fase de recuperación de la información. No satisfacen el enfoque multidimensional de la ciencia. Asimismo, existe el inconveniente de que la mayoría de estos sistemas de clasificación procedentes de las bases de datos bibliográficas no coinciden entre sí.

³ Medline (2016). Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed> [Fecha de consulta: 5/11/2016]

⁴ INSPEC (2016). Disponible en: <http://www.theiet.org/resources/inspec/> [Fecha de consulta: 5/11/2016].

Finalmente, hay que tener en cuenta que todos estos inconvenientes se acentúan aún más a medida que se desciende a niveles de agregación temáticos menores. Si es posible encontrar estas limitaciones cuando se alude a grandes áreas o categorías temáticas, la complejidad aumenta si lo que se pretende localizar son aspectos temáticos más concretos como líneas de investigación, tendencias, sublíneas de investigación, etc. Actualmente, las posibilidades que ofrecen la mayoría de las bases de datos para localizar y analizar la producción científica relativa a aspectos temáticos más específicos que el nivel de áreas, categorías o disciplinas científicas es prácticamente inexistente. La delimitación temática que posibilitan, por tanto, es de carácter general y poco especializada pudiendo suponer un nivel de análisis adecuado, sobre todo, para la realización estudios bibliométricos a nivel macro, pero es insuficiente para análisis que requieran de una información más específica y diversificada como son los estudios enmarcados en un nivel meso o micro.

2.2 Objetivo principal del estudio

El objetivo principal de esta tesis doctoral se centra en el diseño de una propuesta metodológica que permita la localización, identificación y visualización de las principales líneas de investigación de un dominio científico determinado. Dadas las limitaciones que, como se acaba de indicar, presentan las bases de datos bibliográficas para la localización de los principales aspectos temáticos que caracterizan a los documentos, sobre todo, en niveles de agregación específicos, se pretende desarrollar una metodología capaz de llevar a cabo una delimitación temática de carácter específico, permitiendo descender a niveles de análisis mucho más específicos y especializados que el de las áreas o categorías temáticas, como pueden ser las subdisciplinas científicas, líneas de investigación, sublíneas de investigación, etc. De este modo, se pretende obtener una imagen más precisa de los aspectos temáticos que predominan en la producción científica, ya sean los de carácter más consolidado o los de carácter más dinámico, como es la identificación de tendencias en desarrollo debido a la formación o desaparición de nuevos aspectos de la investigación.

El diseño de dicha propuesta se basa, fundamentalmente, en la combinación de diferentes técnicas y herramientas procedentes del campo de la Bibliometría. Concretamente, en la aplicación del Análisis de co-palabras, del Análisis de Redes Sociales y en técnicas relacionadas con la visualización y representación de la información mediante el uso de programas informáticos especializados.

2.3 Objetivos específicos

Los objetivos específicos que se pretenden desarrollar con la propuesta metodológica que plantea esta tesis doctoral, pueden resumirse en los siguientes aspectos:

1. Empleo de la propuesta metodológica en diferentes dominios científicos en función de su cobertura temporal, geográfica y cronológica. Se pretende valorar así, cómo puede afectar la especificidad o la amplitud de los dominios en la delimitación e identificación temática.
2. En relación con el objetivo anterior, obtención de diferentes universos documentales a partir del uso de diferentes fuentes documentales según el dominio a analizar.
3. Análisis estructural de las relaciones que se producen entre los documentos relativos a cada dominio, a través de la co-ocurrencia de palabras, en concreto, a través de los términos de indización que emplean las bases de datos que se han utilizado como fuentes de información.
4. Representación y visualización de la estructura conceptual que vincula a los documentos obtenida del objetivo anterior, a través del empleo de algoritmos y softwares especializados en la visualización de la información y en el Análisis de Redes Sociales.

5. Apreciar la dinamicidad temática de cada uno de los dominios científicos analizados, a través del desarrollo de análisis evolutivos y temporales. De este modo, se podrá apreciar la evolución que han experimentado los aspectos más predominantes de la investigación a lo largo de los períodos de estudio seleccionados.
6. Uso como unidades de análisis de diferentes términos de indización, según su naturaleza documental, para comprobar y valorar si tienen la misma capacidad para identificar, localizar y visualizar líneas de investigación, tomando como objeto de estudio el dominio científico más específico.

Asimismo, todos estos aspectos han sido planteados como objetivos principales que han originado la distribución de la investigación en diferentes fases de trabajo y han dado lugar a la elaboración de las publicaciones científicas. A continuación, se detalla de manera pormenorizada, los objetivos que han motivado el diseño de las preguntas de investigación planteadas en cada una de las publicaciones⁵ y que, además, se responden y discuten en el apartado 6. Resultados y Discusión de esta tesis.

⁵ Adviértase que, las publicaciones correspondientes a los congresos internacionales comparten los mismos objetivos que a los artículos a los que dieron lugar. Es decir, la publicación nº2 dio lugar a la nº3 y la nº4 dio lugar a la nº5. En el caso de la publicación nº 6, no dio tiempo a que se desarrollara en un artículo científico.

Fase de trabajo	Objetivo	Nº de Publicación
1ª	Localización e identificación de las principales líneas de investigación del Dominio 1 a través de los descriptores MeSH, aplicando la propuesta metodológica que se plantea en la tesis.	1
2ª – 4ª	Localización e identificación de las principales líneas de investigación del Dominio 2, aplicando la propuesta metodológica que se plantea en la tesis.	2, 3, 4, 5 y 6.
2ª	Localización e identificación de las principales líneas de investigación del Dominio 2 a través de los <i>KeyWords Plus</i> y aplicando la propuesta metodológica que se plantea en la tesis.	2 y 3
3ª	Localización e identificación de las principales líneas de investigación del Dominio 2 a través de los <i>KeyWords Plus</i> a lo largo del período de estudio para apreciar su evolución temporal.	4 y 5
4ª	Localización e identificación de las principales líneas de investigación del Dominio 2 a través de tipologías diferentes de términos de indización (<i>MeSH</i> , <i>Author Keywords</i> y <i>KeyWords Plus</i>).	6

Tabla 1. Resumen de los objetivos fundamentales que originan las publicaciones

➤ **Publicación nº 1**

• **Objetivos:**

- Localizar y detectar líneas de investigación a partir del uso de una base de datos especializada en medicina (Medline), mediante un tema multidisciplinar (investigación sobre células madre Salud y mujer), con una cobertura temporal amplia y con el empleo de un vocabulario controlado como unidades de análisis (descriptores MeSH).
- Comprobar si el empleo de los descriptores *MeSH* más frecuentes en combinación con *Pajek* y el algoritmo de poda *PfNET* muestran resultados concordantes y complementarios con los que ofrece el Análisis Factorial.
- Valorar si es posible apreciar una evolución temática del dominio a partir de la fragmentación del período de estudio en subperíodos.

- Comprobar si el uso de un vocabulario controlado como son los descriptores *MeSH* permite la identificación de los principales temas de la investigación.
- **Preguntas de investigación:**
 - La propuesta metodológica que se aplica en este estudio ha permitido obtener unos resultados determinados. Concretamente, ¿en qué consisten estos resultados? ¿Éstos permiten obtener una imagen donde se reflejen las principales líneas de investigación que caracterizan el dominio elegido?
 - Existen varios aspectos por lo que la metodología diseñada en este estudio es relevante. Entre ellos, destaca el uso de diferentes técnicas bibliométricas para el análisis de los datos, ¿la combinación del Análisis Factorial y el de redes sociales aplicado al estudio de co-términos ha permitido obtener resultados satisfactorios?
 - Otro aspecto a destacar es el uso de los descriptores MeSH como unidades de análisis, ¿la visualización de las relaciones que se producen entre estos descriptores a partir de su aparición conjunta en los documentos ha permitido representar y definir temáticamente las principales líneas de investigación del dominio? ¿El uso de un vocabulario controlado supone alguna ventaja para este tipo de estudios? ¿Qué inconvenientes presenta? ¿Es posible realizar un análisis estructural de carácter temático llegando a un nivel de agregación tan concreto como es el documental?
 - Desde el punto de vista del análisis que supone el período temporal este estudio se considera de carácter evolutivo-longitudinal. ¿La selección del período temporal abarcando 40 años es suficiente para la identificación temática a partir de un vocabulario controlado? ¿La subdivisión de este período en 4 permite detectar cambios en los temas que reflejen la evolución y dinamismo de la investigación?

- Partiendo de un tema de carácter multidisciplinar como es Salud y Mujer, ¿es posible localizar y representar las tendencias científicas que se han desarrollado en este campo?
- La propuesta metodológica ¿se podría aplicar a otros dominios científicos?, en concreto, ¿a uno más específico? ¿se podrían utilizar otras fuentes documentales para el análisis? ¿Se podrían tener en cuenta como unidades de análisis otra tipología de términos basados en diferentes técnicas de indización?

➤ **Publicaciones nº 2 y 3**

- **Objetivos:**

- Contextualizar la producción científica del dominio de estudio a través de un estudio bibliométrico básico y detectar la posible influencia de ciertos patrones de publicación o colaboración.
- Realizar una delimitación temática a partir del análisis de las categorías que aparecen en los documentos y hallar las áreas y disciplinas que están presentes en el dominio científico.
- Localizar y detectar líneas de investigación a partir del uso de una base de datos multidisciplinar (*Web of Science-EXPANDED*), mediante un tema especializado en medicina (investigación sobre células madre), con una cobertura temporal concreta (1997-2010), circunscrita a un área geográfica determinada (España) y con el empleo de un vocabulario automatizado libre como unidades de análisis (*KeyWords Plus*).

- **Preguntas de investigación:**

- En comparación con el dominio anterior, en esta fase de trabajo se aplica la propuesta metodológica a un dominio de carácter más específico desde el punto de vista temático, cronológico y geográfico. En términos bibliométricos ¿qué indicadores permiten contextualizar mejor este dominio? ¿qué imagen de la investigación ofrecen?
- En esta fase de trabajo se ha utilizado el Análisis de Redes Sociales aplicado al estudio de co-words teniendo en cuenta dos unidades

de análisis diferentes: las categorías y los términos KW+. ¿Qué información aporta el estudio de categorías? ¿se podría aplicar a un estudio de granularidad fina como este? ¿El análisis de los KW+ ofrece más información que el de las categorías? ¿Proporcionan información complementaria? ¿Qué unidades de análisis se podrían considerar más adecuadas para la delimitación temática a nivel de documento?

- ¿Cuál es la estructura general del dominio? ¿De esa estructura se puede extraer alguna conclusión desde el punto de vista temático? ¿Es posible detectar líneas de investigación con la propuesta metodológica sobre un dominio científico tan específico? ¿El uso del SCI-EEXPANDED como fuente de información y de los KW+ como unidades de análisis han permitido la detección de líneas de investigación?
- Actualizando el período de estudio a 1997-2010 ¿se podría realizar un estudio evolutivo con este dominio similar al realizado en el Dominio 1, utilizando los KW+ como unidades de análisis?

➤ **Publicaciones nº 4 y 5**

- **Objetivos:**

- Realizar un análisis temporal para detectar la dinamicidad del campo y apreciar una evolución de los principales temas que predominan la investigación a lo largo del periodo de estudio.
- Aplicar dicho análisis temporal sobre el dominio científico de las publicaciones 2 y 3, para comprobar si es posible localizar, detectar y visualizar líneas de investigación con la metodología que se propone en esta tesis sobre un dominio muy especializado.
- Probar si los softwares especializados en la visualización de la información y del Análisis de Redes Sociales seleccionados para esta tesis son capaces de representar estudios dinámicos de estas características.

- **Preguntas de investigación:**
 - Teniendo en cuenta la especificidad temática del dominio y un período temporal de 14 años ¿sería posible detectar líneas de investigación a través de los KW+ partir de una segmentación temporal?
 - En relación con la imagen que ofrece el período completo ¿sería posible apreciar diferencias o cambios en los temas de investigación en cada uno de los subperíodos? ¿Es posible apreciar la dinámica del dominio a partir de la identificación de líneas de investigación consolidadas y líneas de investigación emergentes?
 - En relación con los softwares de visualización propuestos en esta fase de trabajo ¿han mostrado resultados compatibles? ¿son los que tienen la mayor capacidad para representar estudios temáticos de carácter dinámico o evolutivo?
 - Entre los parámetros metodológicos que se han utilizados en esta fase de trabajo se encuentran el empleo de un dominio de carácter específico, el volumen de años, así como su temporalización, el uso de una fuente documental multidisciplinar y el empleo de los KW+ como unidades de análisis. ¿Alguno de estos parámetros ha supuesto algún inconveniente para la identificación de líneas de investigación a través de la propuesta metodológica?

➤ **Publicación nº 6**

- **Objetivos:**
 - Utilizar como unidades de análisis diferentes términos de indización, según su naturaleza documental, para comprobar y valorar si tienen la misma capacidad para identificar, localizar y visualizar líneas de investigación en un estudio de granularidad fina.
- **Preguntas de investigación:**
 - A partir de la metodología propuesta, ¿sería posible comparar sobre un mismo conjunto documental las representaciones temáticas que se obtendrían a partir del uso de diferentes tipologías de términos desde el punto de vista de la indización? ¿El uso de diferentes unidades de

análisis dentro de esta propuesta metodológica puede repercutir en los resultados de la delimitación temática?

- ¿Se aprecian diferencias en las visualizaciones obtenidas según el tipo de términos de indización utilizado? ¿A qué son debidas?
- Teniendo en cuenta estos resultados, ¿cabe deducir que el empleo de diferentes tipos de términos de indización como unidades de análisis requiere de ajustes metodológicos específicos para cada caso? ¿qué criterios habría que tener en cuenta para desarrollar una propuesta metodológica que contemple dichos ajustes?

3. DELIMITACIÓN DEL ESTUDIO

3.1 Estructura de la tesis

El documento se estructura en dos partes claramente diferenciadas. La primera, denominada Marco general de la investigación, en la cual se recogen los aspectos teóricos y metodológicos, la discusión de los resultados y las conclusiones del estudio. La segunda parte en cambio, recoge las 6 publicaciones que forman del proceso de investigación de esta tesis doctoral.

La primera parte del trabajo comienza con la introducción (capítulo 1) en el que se sitúa el contexto general de la investigación, los motivos por los que se ha desarrollado esta tesis doctoral, así como, el objeto principal de la misma.

En el capítulo 2 se explica, como se puede observar, la estructura general del estudio y se especifica el contexto científico en el que se va aplicar la propuesta metodológica. Se ubica al lector en los dominios que se han tenido en cuenta para su desarrollo, explicando la motivación en cada uno de ellos, como también las consideraciones metodológicas que los diferencian, así como, las publicaciones donde son tratados y analizados.

En el capítulo 3 se expone la justificación y los objetivos de la investigación, uno general y seis específicos. Además, se indican, de manera pormenorizada los objetivos que motivaron cada una de las publicaciones, así como, las preguntas de investigación que se formularon en función de dichos objetivos.

En el capítulo 4 se recoge todo el marco teórico y el estado de la cuestión en relación con el desarrollo de la tesis. Se explican los antecedentes fundamentales que se han desarrollado en el campo y que motivan la elaboración de este estudio. Se especifican

los diferentes puntos de vista e interpretaciones que recoge la literatura sobre el objeto de estudio en el que se enmarca esta tesis doctoral.

El capítulo 5 destaca las limitaciones que condicionan el desarrollo metodológico del estudio y los inconvenientes que han de ser tenidos en cuenta a la hora de hacer una valoración sobre los resultados obtenidos.

El capítulo 6 se titula Material y Métodos. Esta sección se desarrolla ampliamente para explicar todos los elementos que caracterizan la propuesta metodológica, además de justificar el uso de todos los materiales empleados y la argumentación de las decisiones metodológicas tomadas a lo largo de la investigación.

En el capítulo 7 se presentan los resultados y su discusión, mediante la integración de los principales resultados publicados en las seis publicaciones y a través de un formato en el que se da respuesta a las principales preguntas de investigación planteadas en cada publicación.

Finalmente, en el capítulo 8 se exponen las principales conclusiones a las que se llega al final del proceso de investigación y en el capítulo 9 se presentan las principales perspectivas y líneas de trabajo futuras que pueden dar continuidad al trabajo realizado. El trabajo se completa con la lista de referencias bibliográficas utilizadas.

En la segunda parte, aparecen íntegramente las seis publicaciones que integran esta tesis doctoral. En concreto, se trata de tres contribuciones a congresos internacionales que, sirvieron como precedente a dos de los tres artículos publicados en revistas de impacto en el área de la Bibliometría y que también integran el estudio.

- **Publicación nº 1.** Zulueta, M. Ángeles, Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Sánchez, Carmen (2011). Research involving women and health in the Medline database, 1965-2005: Co-term analysis and visualization of main lines of research, *Scientometrics*, 88(3), 679-706.
<http://dx.doi.org/10.1007/s11192-011-0455-1>

- **Publicación nº 2.** Zulueta, M. Ángeles, Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2010). Spanish stem cell research (1997-2007). Investigación española con células madre. Análisis bibliométrico y principales líneas de investigación a través de los *KeyWords Plus*. *I International workshop on scientometric studies related to the biomedical sciences. XV international scientific congress CNIC 2010*. La Habana (Cuba), 28 de junio al 1 de julio 2010, 1-21.

- **Publicación nº 3.** Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida, Zulueta, María-Ángeles (2012). Stem cell research: bibliometric analysis of main research areas through KeyWords Plus, *Aslib Proceedings*, 64 (6), 561-590.
<http://dx.doi.org/10.1108/00012531211281698>

- **Publicación nº 4.** Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2014). Localización y evolución de las principales líneas de investigación a través de los KeyWords Plus. *VI Encontro Ibérico EDICIC 2013 - Globalização, Ciência, Informação*, Oporto (Portugal), 4-6 noviembre 2013, 608-622.

- **Publicación nº 5.** Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2014). Estudio evolutivo de la investigación española con células madre. Visualización e identificación de las principales líneas de investigación. *El profesional de la información*, 23(3), 259-271.
<http://dx.doi.org/10.3145/epi.2014.may.06>

- **Publicación nº 6.** Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2013). Estudio comparativo sobre la visualización de redes de co-words a través de los descriptores del Science citation index y Medline. *I Congresso ISKO Espanha e Portugal / XI Congresso ISKO Espanha*, Oporto (Portugal), 7-9 noviembre 2013, pp. 173-189.
- <http://dx.doi.org/10.1080/00387019808003304>

3.2 Dominios científicos analizados

La propuesta metodológica que se presenta en esta tesis se aplica a dos dominios científicos que tienen grandes diferencias entre sí. Estas diferencias residen fundamentalmente, en la amplitud de la cobertura temática, geográfica y temporal. Para facilitar la comprensión de esta tesis, así como, para agilizar su lectura, se ha decidido utilizar las expresiones “Dominio 1” y “Dominio 2”. El “Dominio 1” constituye el primer dominio sobre el que se aplica la metodología propuesta. Se centra, fundamentalmente, en la investigación recogida en la base de datos Medline sobre salud y mujer durante el período 1965-2005. El diseño metodológico, así como, los resultados de la localización y visualización de las principales líneas de investigación sobre este dominio a través del análisis de co-palabras se recoge en la publicación nº 1. El “Dominio 2”, sin embargo, tiene una cobertura temática, geográfica y temporal más específica que en el caso anterior. Éste se centra en la investigación española con células madre recogida en la base de datos Science Citation Expanded durante los años 1997-2012⁶.

⁶ Este período de estudio ha sufrido varias actualizaciones a lo largo del proceso de creación de las publicaciones que forman parte de esta tesis doctoral. Esto se ha hecho con la intención de utilizar los datos más recientes posibles. Las primeras publicaciones que aplicaron la metodología al “Dominio 2” utilizaron el período temporal 1997-2007 (publicaciones nº 2 y nº 3). Posteriormente, se realizó una actualización de los datos pasando a ser el período de estudio 1997-2010 (publicaciones nº4, nº5 y nº 6).

La elección de aplicar la propuesta metodológica a diferentes dominios tiene como objetivo valorar cómo afecta la especificidad o la amplitud de éstos en la delimitación temática. De este modo, se observa si las características de los dominios pueden condicionar el diseño de la propuesta metodológica, especialmente, en cuanto a la elección de las fuentes y de las unidades de análisis. Por lo que, el uso de diferentes dominios ofrece la posibilidad de detectar cuáles son los ajustes que se deben realizar en la parametrización metodológica en relación con las particularidades del dominio que se analiza. Al mismo tiempo, teniendo presente esto y con los ajustes necesarios, es posible aplicar esta metodología a otros dominios científicos. Es decir, la posibilidad de probar la metodología con dominios diferentes permite valorar qué aspectos se han de tener en cuenta para ajustar la propuesta metodológica a las particularidades de cualquier dominio científico. De este modo, es posible probar una misma metodología, en términos generales, en distintos campos del conocimiento.

Por otro lado, el uso de varios dominios permite comparar los resultados obtenidos y evaluar las posibles limitaciones o inconvenientes de la propuesta. A cuantos más escenarios científicos se aplique la metodología mayor es la exigencia en la aplicabilidad de la misma, ya que aumenta el número de factores metodológicos a tener en cuenta y esto, a su vez, favorece a la robustez o solidez de los resultados que ofrece.

	<i>Cobertura temática</i>	<i>Cobertura temporal</i>	<i>Cobertura geográfica</i>
<i>Dominio 1</i>	Investigación sobre Salud y Mujer	1965-2005	Sin determinar
<i>Dominio 2</i>	Investigación sobre células madre	1997-2012	España

Tabla 2. Características de los dominios analizados

3.2.1. “Dominio 1”

Como se acaba de indicar, este es el primer dominio sobre el que se aplica la metodología propuesta. Se centra en los estudios médicos sobre salud que han tenido a la mujer como sujeto de investigación a lo largo de 40 años. La introducción del concepto del género y el sexo en la salud se ha producido de manera paulatina. Los primeros estudios que hacían referencia a este hecho se limitaban prácticamente a temas relacionados con la reproducción, (Vivian & Pinn, 2005). Su presencia estaba reducida a especialidades médicas como la obstetricia y la ginecología y apenas se la consideraba para estudiar enfermedades que afectaban y afectan indistintamente a hombres y a mujeres (WHO, 1998).

No es hasta los años 70 cuando comienzan a publicarse investigaciones que hablan de los distintos niveles de morbilidad y mortalidad entre ambos sexos, (Nathanson, 1975; Nathanson 1977; Verbrugge 1978; Waldron, 1978) y se llega a afirmar que los hombres tienen un índice de mortalidad mayor pero unas menores tasas de morbilidad que las mujeres, una idea que continuó debatiéndose durante la década de los 80 (Clarke 1983; Verbrugge, 1985, 1989). A pesar de que empieza a reconocerse las diferencias en la salud entre ambos sexos, la presencia de la mujer seguía siendo escasa en la mayoría de las investigaciones y de los estudios clínicos.

A mediados de los años 80 empieza a plantearse la importancia de introducir a las mujeres dentro de la investigación biomédica y las pruebas clínicas. Pero no es hasta los años 90 cuando comienzan a desarrollarse iniciativas concretas al respecto. Por ejemplo, el Food and Drug Administration (FDA) y el National Institutes of Health (NIH) establecieron medidas para incluir a las mujeres como sujetos de análisis en las pruebas clínicas (1995; Merkatz & Junod, 1994), creando la Office of Research on Women’s Health (ORWH) en 1990 para controlar y asegurar la investigación sobre salud y mujeres. Otros organismos e instituciones relacionados con la salud, también empiezan a introducir cambios en su política sanitaria y a elaborar instrumentos para motivar a los países a introducir la perspectiva del género y el sexo dentro de la investigación médica

y los servicios sanitarios (Gender and Health Group, 2000; Women's Health Bureau, 2002; WHO, 2002; Nobelius & Wainer, 2004; Sen, George. & Östlin, 2005).

La introducción del concepto del género y el sexo en la salud, tiene una incidencia directa en el curso de las publicaciones científicas. Se puede apreciar que la mujer comenzó a estar presente en el estudio de enfermedades como la infección por VIH, enfermedades cardíacas, síndrome del colon irritable, cáncer de pulmón, etc. (Vivian & Pinn, 2003; Hamberg, Risberg & Johansson, 2004; Devesa et al. 2005; Blake, Ince & Dean, 2005; Daly et al. 2006; Kerrigan et al. 2008). Las publicaciones de estas investigaciones crecieron considerablemente en muy poco tiempo, generando una gran cantidad de información y de nuevos conceptos. Los propios investigadores y profesionales en temas de la salud han reconocido que tienen dificultades para asimilar y controlar las nuevas líneas de investigación y la información que ha surgido al respecto (Montgomery & Sherif, 2000).

Asimismo, teniendo en cuenta el dinamismo de la investigación científica y su consiguiente producción literaria, resulta difícil tener una imagen clara de la estructura, relaciones y límites de una disciplina o campo de investigación. Una representación simple y asequible de la estructura del dominio, facilitaría la traslación mental de su imagen, así como, mejoraría su asimilación.

3.2.2. "Dominio 2"

Este dominio, de carácter más específico que el anterior, se centra en la investigación española con células madre producida desde 1997-2012. A grandes rasgos, se podría decir que este es uno de los campos del conocimiento que ha tenido un gran impacto en cualquier ámbito de la sociedad mundial. Su estudio y desarrollo no sólo ha despertado un gran interés en el ámbito de la biomedicina, sino que, ha trascendido a

niveles políticos, éticos, culturales, jurídicos, etc., situándolo al frente de la controversia y del paradigma científico.

Desde que en 1949, el científico J. Hammond descubriese el método para mantener los embriones de ratón en cultivo *in vitro*, la investigación con células madre se ha desarrollado de manera incesante ofreciendo resultados muy esperanzadores en el tratamiento de enfermedades, hasta ahora, incurables. Ya en 1981, un grupo de investigadores obtuvieron células pluripotentes, también de ratón (Evans & Kaufman, 1981), y dos años más tarde se desarrolló una nueva técnica para experimentos de transferencia nuclear (McGrath & Solter, 1983) que, permitió el paso a la era de la clonación con manipulación genética que se desarrollaría años más tarde. También, comenzaron a progresar estudios relativos al proceso de diferenciación de células embrionarias *in vitro* (Wobus, Jakel & Schoneich, 1984; Doetschman et al., 1985) y a la descripción de las ventajas de éstas células para la creación de animales transgénicos (Gossler et al., 1986).

Ya en la década de los 90, el salto cualitativo a la clonación por transferencia nuclear, se materializó a través del grupo de investigación de K.H.S. Campbell (1996) del Instituto Roslin de Escocia, donde se intentaron obtener ovejas clónicas. Finalmente, I. Wilmut (1997) lo consiguió dándonos a conocer a la famosa oveja Dolly. Los siguientes pasos de la investigación en esta década se orientaron a la descripción del procedimiento para aislar y crecer en laboratorio células embrionarias obtenidas a partir de la Masa Celular Interna (MCI) del blastocito.

Posteriormente, la dirección de la investigación con células madre se abrió a muchos campos aunque, en última instancia, todos ellos están orientados al desarrollo de nuevas terapias para enfermedades que afectan a órganos vitales, como son las enfermedades cardiovasculares, neurodegenerativas, genéticas, cáncer, diabetes, etc. (Bishop, Buttery & Polak, 2002). Tanto es así que, se están consiguiendo importantes logros médicos como la posibilidad de que pacientes afectados con diabetes de tipo 1

puedan sobrevivir sin recibir insulina mediante un trasplante de células madre (Voltarelli, Martínez & Burt, 2009), o el desarrollo de células cardíacas que podrían ayudar a reparar los músculos cardíacos lesionados por ataques al corazón (Yang et al., 2008) o el trasplante de células hematopoyéticas que restituyan las células neuronales afectadas por la enfermedad de Parkinson (Lu et al., 2001).

Pero al mismo tiempo, la capacidad de estas células para su renovación y la facultad de mantenerse en estado indiferenciado (McKay, 1997; Cordon & Blackett, 1998; Scheffler et al., 1999) especialmente, en el caso de las células embrionarias, ha planteado serios debates políticos, religiosos, éticos, legislativos etc., afectando a la fluidez de los estudios y el desarrollo habitual de la investigación. Es por eso que, muchos trabajos se orientaron hacia la obtención de células madre evitando la destrucción de los embriones. Se desarrollaron técnicas similares al Diagnóstico Genético Preimplantacional (Lu et al., 2007), o técnicas de creación de embriones híbridos, óvulos de procedencia animal a los que se inserta una célula adulta humana (Chung et al., 2006). Pero sin duda, la técnica que tuvo mayor impacto en el curso de la investigación fue la de la reprogramación celular que obtenía las denominadas células *iPS* (*induced Pluripotent Stem*), es decir, células pluripotentes a partir de la reprogramación de células adultas de la piel (Takahashi & Yamanaka 2006; Takahashi et al. 2007).

No obstante, y pese a todas las controversias y polémicas que suscitan el desarrollo de determinadas líneas de investigación con células madre, hemos asistido en los últimos años a una serie de iniciativas políticas, económicas y legislativas que manifiestan una clara intención de adecuar los adelantos científicos a las estructuras sociales de nuestro tiempo. Ciencia y sociedad se baten en un pulso en el que los Estados intentan hallar el equilibrio preciso. Así, encontramos decisiones como la del presidente de los EE.UU., Barack Obama que, determinó levantar el veto de la financiación de proyectos de investigación con células madre embrionarias impuesto por la legislatura anterior (Decreto 09/03/2009), evitando así, quedar rezagado en la carrera científica.

Por su parte, en Europa se pueden destacar importantes esfuerzos para el impulso de la investigación mediante significativas medidas legislativas como son las directivas 2001/20/CE, de 4 de abril; 2003/63/CE, de 25 de junio; 2004/23/CE, de 31 de marzo y la 2006/17/CE, de 8 de febrero, o mediante la incentivación económica a través del VII Programa Marco que destinó una partida de 32.413 millones de euros al área de salud, dentro del apartado cooperación, una de las cuatro secciones en las que se estructura el programa. Además, contempló la creación del Consejo Europeo de Investigación (ERC) para el que se destinaron 7,5 mil millones de euros; o la creación del Registro Europeo de Células Madre Embrionarias con el objetivo de proporcionar información a toda la comunidad científica, sobre las líneas de células disponibles en Europa.

A su vez, Reino Unido, Australia y Asia entre los grandes productores científicos han ido incentivando y legislando al respecto, aunque a lo largo de los años estas regulaciones han sido sometidas a una continua revisión (EuroStemCell, iCeMS & Elsevier, 2013).

Del mismo modo, en lo referente a España, también se han sucedido importantes pasos para el desarrollo de la investigación científica en el campo de las células madre. A nivel legislativo, se derogó la Ley de Reproducción Asistida 45/2003, de 21 de noviembre que, a su vez, derogaba la antigua Ley 35/1988, de 22 de noviembre, por la actual Ley 14/2006, de 26 de mayo, ofreciendo la posibilidad del trasplante autólogo y, por lo tanto, la autorización para la creación de bancos privados de sangre de cordón umbilical en nuestro país, siempre y cuando, se cumplan las previsiones que han sido desarrolladas por el Real Decreto 2132/2004, de 29 de octubre. Esta Ley es fruto de la adaptación de las directivas europeas 2004/23/CE y 2006/17/CE. También, se aprobó la Ley 17/2007, de 3 de julio, sobre Investigación Biomédica, que supuso la actualización del ordenamiento jurídico español en materia de células madre posibilitando, entre otras cosas, la clonación terapéutica y la investigación con embriones. A nivel ejecutivo, se han tomado decisiones significativas en materia de financiación como puede ser el gasto en la inversión en I+D+i para la legislatura del 2008, ascendiendo a 417 millones

de euros (en 2004 fue de 220 millones de euros), o el apoyo financiero desde el Fondo de Investigaciones Sanitarias, o las acciones para el retorno de la inversión como consecuencia de la adaptación de España al VII Programa Marco de la UE, creándose para ello por un lado, el Fondo Euroingenio, en colaboración con las Comunidades Autónomas y gestionado por la Oficina de Proyectos Europeos en Biomedicina y Ciencias de la Salud, y por otro, el Programa Eurociencia, gestionado por la Oficina MEC-Europa.

Todos estos hitos han contribuido a generar un importante pool de resultados de esta actividad científica que son susceptibles de analizar desde la perspectiva bibliométrica. Sobre todo, teniendo en cuenta que esta investigación con células madre ha trascendido a otros campos científicos, siendo un tema con un fuerte carácter multidisciplinar. Esto supone un reto a la hora de definir temáticamente el dominio y, sobre todo, a la hora de localizar las principales líneas de investigación que se están desarrollando.

<i>Publicaciones en las que son tratados</i>	
<i>Dominio 1</i>	Nº 1
<i>Dominio 2</i>	Nº 2, 3, 4, 5 y 6

Tabla 3. Publicaciones donde son tratados los dominios

4. ANTECEDENTES

La propuesta metodológica que se presenta en esta tesis, tiene como objetivo final la localización e identificación de los principales temas que son abordados en las publicaciones científicas dentro de un dominio científico determinado.

En un sentido amplio, la delimitación temática sirve como herramienta para la organización del conocimiento que, en definitiva, es generado a través de la práctica científica. Localizar, estructurar y organizar los diferentes aspectos temáticos del conocimiento ayuda a comprender las características y la evolución del mismo. En última instancia, permite apreciar la distribución de los fundamentos, principios, métodos y técnicas científicas que están presentes en un paradigma científico dado.

No obstante, para abordar este tipo de análisis es necesario tener un enfoque o perspectiva desde la que comprender cómo se produce el conocimiento científico y, además, un método, herramientas y técnicas que permitan analizarlo de manera objetiva y cuantificable. En este sentido, los estudios bibliométricos son fundamentales porque son los encargados del análisis, evaluación y seguimiento de la actividad científica. Tienen como objetivo *“la aplicación de los métodos estadísticos y matemáticos dispuestos para definir los procesos de la comunicación escrita y la naturaleza y el desarrollo de las disciplinas científicas mediante técnicas de recuento y análisis de dicha comunicación”* (Pritchard, 1969).

En cuanto al enfoque o perspectiva epistemológica que sustenta, dirige y fundamenta el desarrollo de los análisis bibliométricos, actualmente está basada en una interpretación holística de la ciencia en la que la evaluación de la misma debe realizarse a partir del conocimiento de las prácticas sociales de los científicos. En la que la mejor forma de comprender la información, consiste en estudiar los dominios de conocimiento como parte del discurso de las comunidades de las que proceden, que son el reflejo de la división social y laboral de la sociedad (Vargas-Quesada, 2005). Esta perspectiva se

fundamenta en los principios que definen los Análisis de Dominios, expuestos por primera vez por Hjørland y Albrechtsen en 1995.

Desde este punto de vista, cobra especial importancia el contexto social y el valor de las comunidades científicas como creadoras del discurso científico. Las interacciones que se establecen dentro de estas comunidades tienden a formar un entramado de relaciones donde se desarrollan los flujos de conocimiento e información científica. De este modo, se forman estructuras relacionales que pueden ser objeto de análisis ofreciendo una visión muy relevante sobre el desarrollo científico y para la identificación de patrones de comportamiento en el proceso de generación de nuevo conocimiento. Así, técnicas procedentes del campo de la Visualización de la Información y del Análisis de Redes Sociales, se convierten en elementos indispensables para el ámbito de la Bibliometría, ya que, ofrecen métodos tanto visuales como matemáticos para el estudio de dichas relaciones estructurales. La aplicación de estas herramientas en los Análisis de Dominios, facilita una mejor comprensión y percepción del desarrollo científico.

Asimismo, el valor del discurso científico adquiere un valor significativo para esta interpretación holística de la ciencia porque, es considerado como el resultado del conocimiento social, cultural e histórico de la comunidad que lo elabora. Las palabras se convierten, por tanto, en unidades de expresión seleccionadas por los actores principales del desarrollo científico encargadas de vehicular las ideas y los conceptos presentes en el diálogo científico. De este modo, debido a este rastro que dejan los signos lingüísticos, es posible detectar, localizar y analizar los aspectos temáticos que estructuran los diferentes ámbitos del conocimiento y favorecer estudios relacionados con la delimitación temática.

4.1 Análisis de Dominios

El Análisis de Dominios (AD) proporciona una perspectiva teórica y contextual desde la que abordar el análisis de un ámbito del conocimiento científico dado dentro del campo de la *Information Science* (IS). Constituyen un punto de partida de fundamentos teóricos, principios, métodos y herramientas desde el que emprender el estudio, evaluación, interpretación o comprensión de los resultados de la actividad científica.

En el caso de esta tesis es necesario comprender, brevemente, en que consiste el AD ya que la propuesta metodológica que se presenta se aplica a dos dominios científicos claramente diferenciados por su cobertura temática, geográfica y temporal. De este modo, se comprenderá mejor el marco teórico sobre el que se desarrolla esta tesis en particular y al mismo tiempo, la perspectiva epistemológica desde la que se realizan, actualmente, la mayor parte de los estudios bibliométricos en general.

Aunque el AD en el campo de la IS ha sido abordado en la literatura científica por diferentes autores, fueron Hjørland y Albrechtsen (1995) los que lo definieron por primera vez. Para ellos, está basado en un paradigma dominio-analítico que obedece a tres naturalezas, fundamentalmente. En primer lugar, como paradigma social, ya que entienden que la IS es una ciencia social que emplea fundamentos procedentes del campo de la sociología, psicología y sociolingüística. En segundo lugar, como una aproximación funcionalista, pues intenta comprender las funciones implícitas y explícitas de la ciencia, además de resaltar o hacer visibles los mecanismos subyacentes de comunicación. Y, en tercer lugar, como una aproximación filosófico-realista, pues intenta establecer la base científica de un dominio, a través de factores externos a la percepción subjetiva e individualista de los sujetos, en contraposición, por ejemplo, a los paradigmas cognitivo y conductivista.

Asimismo, Hjørland y Albrechtsen (1995) afirman que la mejor forma de comprender la información es la que consiste en estudiar los dominios de conocimiento como parte del discurso de las comunidades de las que proceden, las cuales son el reflejo de la división social y laboral de la sociedad. Sostienen que *“la organización del conocimiento, su estructuración, los patrones de cooperación, las formas y el lenguaje de comunicación, los sistemas de información y los criterios de relevancia, son un reflejo del trabajo de estas comunidades y del papel que juegan en la sociedad”*. No obstante, a pesar del peso que tiene el entorno social en la actividad científica estos autores también destacan el valor de la psicología individual de cada sujeto, su conocimiento, sus necesidades informativas, así como, sus criterios subjetivos de relevancia a la hora de interactuar en el diálogo científico.

Para Hjørland y Albrechtsen las comunidades que construyen el discurso científico se definen como un conjunto de actores sociales que comparten una determinada visión del mundo y participan, de manera consensuada, en la creación de soluciones universales ante los problemas existentes. Es decir, en el AD la dimensión histórica, cultural, social y objetiva tiene una especial incidencia en la construcción de un paradigma científico dado. Esta idea está relacionada con el modelo científico desarrollado por Thomas S. Kuhn que supuso un gran cambio en el debate filosófico de la segunda mitad del siglo XX y que propició una ruptura con el modelo formalista que imperaba en ese momento. En su obra *La estructura de las revoluciones científicas* (1962), Kuhn configura una teoría de la ciencia como un proceso cíclico, dando importancia a las características sociológicas de la comunidad, donde el conocimiento científico se daría como tal cuando la actividad de los científicos se agrupase en torno a un conjunto de afirmaciones y métodos consensuados o admitidos por la comunidad. A este autor se debe el concepto de paradigma científico que definió como *“realizaciones científicas universalmente reconocidas que, durante cierto tiempo, proporcionan modelos de problemas y soluciones a una comunidad científica”* (Kuhn, 1962). Además, Kuhn describió que la ciencia se desarrolla siguiendo un patrón cíclico compuesto por una serie de estadios que conducen al cambio de paradigma. Este proceso se inicia con lo que él denominó “ciencia normal” pasando por “crisis” y “revoluciones científicas”

hasta llegar al establecimiento de un nuevo paradigma. Este enfoque sigue estando presente en la actualidad en el campo de la IS. En concreto, en el área de la Bibliometría Marx & Bornmann (2013), desarrollan aún más este modelo definiendo cuáles son exactamente los requisitos que deben cumplirse para que se produzca un cambio de paradigma a través de lo que ellos denominan el Principio Ana Karenina (Anna Karenina Principle).

Esta metateoría kuhniana está basada en el constructivismo social que responde a una corriente de pensamiento que surge como contrapunto a las interpretaciones fundamentalistas y empíricas desarrolladas en la primera mitad del siglo XX. Éstas últimas, se deben en gran medida a las teorías desarrolladas por el Círculo de Viena. Este movimiento se caracterizó por la combinación de principios procedentes del empirismo y el positivismo clásico con la lógica formal. Sus postulados se basaban en el ideal de una ciencia unificada fundamentada en la observación empírica y sustentada en el análisis lógico desarrollado por Whitehead y Russefl (1962), bajo el consecuente rechazo a toda metafísica. Su objetivo se basaba, por un lado, en la construcción de un lenguaje perfecto que dispensara al científico de tomar decisiones, para realizar razonamientos libres de error y contradicción entre los datos y la derivación de las teorías. Y, por otro lado, en la aplicación de la teoría verificacionista procedente del positivismo lógico inspirado en gran medida en Wittgenstein (1921), según la cual una proposición tiene sentido únicamente si se refiere a un conjunto que puede verificarse.

Estas corrientes de pensamiento parten de supuestos muy diferentes a los del AD, donde el peso de la dimensión histórica, cultural, social y objetiva cobra un papel fundamental. En este caso, la realidad no puede ser captada ni comprendida inocentemente por un individuo aislado y sin preparación. Es el propio conocimiento del individuo construido a partir de la historia y la cultura, incluido aquel que se haya podido desarrollar sobre un determinado dominio de conocimiento, el que ofrece la posibilidad de percibir la realidad. En definitiva, el individualismo metodológico que considera al conocimiento como un estado mental del individuo, es reemplazado por un colectivismo

metodológico u holístico, que entiende el conocimiento como un proceso social o como un producto cultural (Hjørland & Albrechtsen, 1995).

Del mismo modo, la interpretación de la realidad a través del lenguaje juega un papel fundamental en el AD. Se abandona la idea de que el lenguaje se reduce a cumplir únicamente con una función nominalista cuya principal utilidad consiste en etiquetar el conocimiento percibido mediante los sentidos o adquirido por la razón. Las corrientes positivistas y racionalistas excluían claramente al lenguaje en el proceso de percepción de la realidad. Consideraban que éste actuaba, simplemente, como vehículo de comunicación entre los individuos, donde la percepción individual del conocimiento estaba libre de las tradiciones culturales. Por el contrario, la interpretación holística del AD otorga un gran valor al lenguaje como herramienta para la elaboración del discurso científico. Éste último, se considera como el resultado del conocimiento social, cultural e histórico de la comunidad que lo elabora. Está compuesto por el intercambio de pareceres que se produce entre los miembros que constituyen la comunidad, la cual es el reflejo de la división social y laboral de la sociedad. De este modo, el lenguaje y los signos semánticos, actúan como herramientas de comunicación dependientes de la Historia y de una *sociocultura* específica (Ma, 2006). En concreto, en el campo de la IS autores como Y. Wang (2013) consideran que la información codificada en las bases de datos bibliográficas está completamente condicionada por el entorno y la actividad social de la comunidad que la genera. En definitiva, para el AD la actividad científica y la generación de nuevo conocimiento son considerados como procesos sociales donde adquiere especial relevancia el lenguaje y las palabras que construyen el discurso científico.

En cuanto a lo que se refiere a las técnicas, métodos, herramientas, etc, más adecuadas para llevar a cabo el AD desde el campo de la IS, Hjørland (2002) propone once métodos, algunos de corte tradicional y otros más innovadores. En lo referente a los estudios métricos, Hjørland (2002) considera que la Bibliometría puede emplearse tanto como herramienta como método. Como herramienta, puede ser utilizada para la

generación, por ejemplo, de mapas bibliométricos basados en el análisis de co-citación, que facilitan la visualización de una determinada disciplina (White & McCain, 1998). Como método, la bibliometría puede mostrar las relaciones reales entre los documentos individuales y revelar el reconocimiento explícito que unos autores hacen a otros, al mismo tiempo que, refleja las relaciones entre los distintos campos científicos (Garfield, 1976). De este modo, puede mostrar y describir tendencias en distintas áreas de conocimiento. De acuerdo con esta dualidad metodológica de la Bibliometría se muestran trabajos más recientes como el de Vega Almeida et al. (2011) que proponen que esta disciplina puede emplearse como herramienta y método para la identificación de paradigmas científicos en los dominios de conocimiento. Sin embargo, Hjørland y Albrechtsen (1995) indican que a la hora de obtener una interpretación sociocultural de los datos y de las visualizaciones bibliométricas ha de recurrirse a otras disciplinas más amplias, haciendo así necesario el enfoque holístico del análisis de dominios (Garfield, 1992).

Del mismo modo, este enfoque holístico del AD que destaca la idea de que el conocimiento de las prácticas de los distintos campos científicos es esencial en la IS tiene como consecuencia que las fronteras, en el caso de la Bibliometría con respecto a otras disciplinas, como por ejemplo la Organización del Conocimiento o la Recuperación de la Información, son cada vez más difusas (Smiraglia, 2015).

Asimismo, dicho enfoque también afecta a la hora de definir y delimitar un dominio determinado. Tennis (2003), considera que es necesario precisar cuáles son los procedimientos operacionales que ayudaban a establecer los límites de un dominio científico. Su objetivo se centra en proporcionar una definición clara y estandarizada comprensible para otros analistas de dominios y transferible a otros campos de conocimiento. Para ello, propone dos ejes de análisis. El primero denominado *Áreas de Modulación*, que se refiere al conjunto de parámetros que definen nominalmente al dominio y establecen su extensión. El segundo corresponde al *Grado de Especialización* que consiste en calificar y establecer la intención de un dominio frente a otros. Por su

parte, Christensen (2014) detecta la dificultad que tienen los investigadores y profesionales a la hora de realizar búsquedas sobre dominios específicos. Proponen el modelo UNISIST que procura facilitar el entendimiento de los procesos de comunicación en un marco determinado o localizar un dominio específico en relación con el mejor proceso de búsqueda de información. Este análisis se aborda desde la perspectiva de la Recuperación de la Información, en concreto, con la mejora de las búsquedas en dominios científicos especializados.

Finalmente, ha quedado de manifiesto la importancia que para el AD tiene el contexto social y el valor de las comunidades científicas como creadoras del discurso científico. Esto favorece que cobren especial relevancia las interacciones que se dan en el sistema de actores, su implicación y repercusión en la sociedad. De este modo, se crea un entramado de relaciones donde se desarrollan los flujos de conocimiento e información científica, formándose así estructuras relacionales susceptibles de ser objeto de análisis. En el campo de la IS la información proviene, fundamentalmente, de la literatura científica recogida en las bases de datos bibliográficas. En este caso, los documentos se relacionan entre sí a través de los elementos bibliográficos. El estudio de estas relaciones proporciona una gran variedad de información sobre el dominio científico seleccionado, ya sea mediante el empleo de indicadores relacionales o mediante técnicas como la co-citación, la co-ocurrencia de palabras y el análisis de colaboración, en varios niveles de agregación, sean autores, revistas, categorías, palabras claves, resúmenes, etc. (Chinchilla- Rodríguez & Moya-Anegón, 2007).

Para llevar a cabo el estudio de las relaciones estructurales en el área de la IS se emplea una técnica que procede del ámbito de la Sociología conocida como *Análisis de Redes Sociales (ARS)*. Basada, fundamentalmente, en la teoría de grafos que proporciona herramientas tanto visuales como matemáticas para el estudio de las relaciones humanas. Actualmente, es una técnica muy utilizada en el campo de la Bibliometría, tanto para la visualización y representación de la información como para el análisis de las estructuras relaciones que se producen en un dominio científico, especialmente,

para la obtención de lo que se conoce como mapas de la ciencia o cienciogramas (Vargas-Quesada, 2005).

De este modo, se podría resumir lo que supone el AD para el campo de la bibliometría y marco conceptual sobre el que se desarrolla esta tesis. Cabe decir, que sigue siendo un aspecto sobre el que se sigue reflexionando y sobre el que, actualmente, se centran algunos estudios. Además de las referencias indicadas anteriormente, se pueden destacar los trabajos de Richard P. Smiraglia (2012; 2015) que aborda el AD desde la perspectiva de la Organización del Conocimiento y reconoce el valor de la Bibliometría como método y herramienta. También destaca el trabajo de González Guitián & Zayas Pérez (2012), que recoge un estudio bibliométrico sobre Auditorias del Conocimiento en las bases de datos del Science Citation Index y Scopus como ejemplo práctico de un AD. O el artículo de Martens & Van Fleet (2012) que también tratan de ejemplificar un AD, pero en esta ocasión enfocado al campo de las instituciones documentales como parte de la infraestructura del conocimiento. Por último, señalar el estudio de Gutierrez Castanha & Cabrini Gracio (2014), en el que desarrollan un metateoría sobre el AD aplicado al campo de la Información y bajo la concepción holística de la Ciencia.

4.2 Visualización y representación de la información

Dentro del proceso científico, el tratamiento y transformación de los datos en información para su posterior análisis resulta un paso fundamental para la consecución de unos resultados óptimos. Una de las maneras más accesibles para comprender la realidad que subyace detrás de los datos científicos y de poder concretar los conceptos e ideas que pertenecen al ámbito de lo abstracto, es mediante la generación de representaciones visuales. La acción de visualizar la información facilita su comprensión y asimilación, contribuyendo a la generación de nuevo conocimiento (Vargas-Quesada, 2005). Si se acude al origen etimológico del término información es posible comprobar que guarda cierta relación con el concepto de representación ya que, proviene del

sustantivo latino *informatio (-nis)* (que, a su vez, proviene del verbo *informare*) que, entre otras acepciones, significa "dar forma a la mente".

En términos generales, la visualización puede ayudar a hacer visible al ojo humano aquello que es muy pequeño y/o complejo de ver o, por el contrario, aquello que es excesivamente grande para comprenderlo, también aquello que sólo existe en la mente, o a poner de manifiesto fenómenos que no son visibles por sí mismos, como es el caso de la visualización de la información, así como, las relaciones entre los elementos que la componen (Araya, 2003). Según Card, Mackinlay & Shneiderman *"la visualización puede aumentar la capacidad de procesamiento, recurriendo directamente a los recursos del sistema visual. O puede actuar indirectamente descargando de trabajo al proceso cognitivo, o reducir las tareas memorísticas haciendo que parte de estas sean externas y visuales [...]. La visualización permite realizar inferencia de una forma fácil, que de otra forma no sería posible"* (Card, Mackinlay & Shneiderman, 1999). En el caso de la visualización de la información puede permitir el reconocimiento de patrones, de estructuras, tendencias y relaciones que ayudan a la comprensión de los datos y del fenómeno que representan que, de otro modo, serían incomprensibles por su complejidad (Olmeda, 2014).

De este modo, para Card, Mackinlay y Shneiderman (1999), la visualización de la información puede definirse como *"el uso de representaciones visuales interactivas de datos abstractos, no físicos, para ampliar el conocimiento"*. Es decir, se centra en la extracción de estructuras de grandes conjuntos de datos que puedan ser representadas de forma visual. En consecuencia, en las visualizaciones de información intervienen distintos procesos. Hay una transformación de datos brutos en abstracciones analíticas que, a continuación, se transforman en un modelo espacial-visual abstracto, para que finalmente, mediante procesos de diseño, el modelo visual se presente de forma gráfica y visible (Olmeda, 2014).

El gran desarrollo de la informática en las últimas décadas ha tenido una especial incidencia en la visualización de la información. Por un lado, gracias a las posibilidades que ofrece de almacenamiento, tratamiento y difusión, se han diversificado notablemente las fuentes de información suministrando enormes cantidades de información dificultando cada vez más su procesamiento e interpretación. Tanto es así, que I. Alcalde (2012) indica que la cantidad de datos que se producen en el mundo crece exponencialmente, siendo la expectativa de generación de datos para el 2020 de 35 Zettabytes/año, (es decir, un 1 seguido de 21 ceros y multiplicado por 35). Además, según M. Hilbert y P. López (2011) desde el 2007, el 99% de la información generada en el mundo es en formato digital y según el informe *Business, Analytics and Optimization for the Intelligent Enterprise* de IBM (2009), se produce un promedio de unos 2,5 trillones de bytes de datos cada día, habiéndose generado el 90% de los datos de todo el mundo en los últimos dos años. Esto es lo que se ha denominado recientemente, como el fenómeno *Big Data* o era de los grandes volúmenes de información. Por otro lado, el desarrollo de la informática también ha impulsado el incremento de herramientas y técnicas que tratan de favorecer la comunicación de resultados científicos y de promover una mayor interacción entre los investigadores y el tratamiento informático de los datos.

Una de las grandes posibilidades que brinda la informática al campo de la visualización de la información es la oportunidad de interactuar con las imágenes. El observador de una representación visual está obligado a interpretarla a partir de la forma física externa de la estructura y de los atributos de los datos que se usan para definirlos de forma visual como objetos y además tiene la posibilidad de interactuar pudiendo manipularla, tratarla y gestionarla de la misma forma que lo hace con otros objetos de la realidad. Según C. Olmeda (2014), *“las visualizaciones de información son procesos dinámicos, donde se permite al usuario en su grado máximo, el control virtual de todas las fases de los procedimientos, desde la selección y control de los datos que se desean ver, hasta la posible selección del color o el modo de afinar las vistas”*. No obstante, en este proceso de interacción con las imágenes hay que tener en cuenta las capacidades cognitivas relacionadas con las habilidades y las limitaciones propias del

sistema visual de cada observador. Es esencial tener en consideración las habilidades de percepción de cada individuo y comprender cómo se ve, ya que, puede ayudar a generar visualizaciones más efectivas y eficaces (Costa, 1998).

Del mismo modo, también cabe tener en cuenta que, el marco en el que se desarrolla la visualización de la información es en un mundo multifacético, multidimensional y multi-fenoménico que se presenta como un *continuum* donde la intersección de la imagen, la palabra, el número y el arte ayudan a transformar la información en conocimiento (Vargas-Quesada, 2005). La complejidad de este mundo que presenta tantas dimensiones, plantea un reto a la hora de visualizar la información ya que las representaciones gráficas están limitadas a dos o tres dimensiones como máximo y a un formato comprimido y reducido (Tufte, 1994).

Dentro del campo de la visualización de la información es posible distinguir diferentes tipologías en función del tipo de información que se esté representando. A continuación, se indican las tipologías que C. Olmeda define en su estudio (Olmeda, 2014):

- **Visualización de datos geoespaciales:** describen objetos o fenómenos con una localización específica en el mundo real. La forma más habitual de representar objetos de información geolocalizable es en mapas temáticos. Una vez ubicados en el espacio, los datos geolocalizables pueden revelar patrones y relaciones difíciles de descubrir o de ser apreciados de otro modo.
- **Visualización de redes:** se centra en las interrelaciones de la información relacional. La manera más habitual de representar este tipo de información es mediante la generación de grafos. Éstos, mediante el empleo de nodos y enlaces permiten reflejar estructuras relacionales y patrones que reflejan la conexión de los elementos de análisis.

- **Visualización de textos:** consiste en transformar los objetos contenidos en los corpora o colecciones de documentos, en un formato visual. Estos objetos pueden ser palabras, sentencias, párrafos, atributos y metadatos, a nivel individual o en colecciones. Si se identifican entidades importantes en el seno de los textos, se puede analizar su contenido en tres niveles de representación: en un nivel léxico, sintáctico o semántico. Por ejemplo, mediante la obtención de frecuencias o distribuciones de palabras; el contenido semántico y repeticiones; el tema o los temas que definen conjuntos de documentos o los temas centrales de los documentos. También pueden detectarse semejanzas mediante conexiones, cambios o fenómenos de difusión de la información.

Finalmente, indicar que la representación gráfica de la información para su posterior visualización, es una actividad común en la mayoría de las disciplinas científicas (Klovdhal, 1981; Crosby, 1997) aunque tiene una especial incidencia en el campo de la Bibliometría, ya que, tiene una gran utilidad para poner al descubierto y divulgar la esencia y estructura de la ciencia. Pese a sus años de existencia, actualmente, se encuentra en fase de evolución y desarrollo teniendo especial interés para el Análisis de Dominios, el mapeo de la ciencia y el Análisis de Redes.

4.3 Mapeo de la ciencia

La posibilidad de capturar el proceso científico en un momento determinado para poder analizarlo, evaluarlo y comprenderlo mejor en todas sus posibilidades y en toda su complejidad, ha sido un anhelo perseguido por los investigadores durante mucho tiempo. Poder captar la realidad en su máxima expresión mediante la obtención de una imagen a modo de fotografía y así tratar de expresar la mayor cantidad de información posible, ha sido un objetivo fundamental para el AD.

La visualización de un gran dominio científico o *The Big Picture*, a modo de representación gráfica o mapa bibliométrico, ofrece una visión general de la estructura de la literatura científica mostrándose útil para poder identificar las principales áreas de investigación dentro de un campo científico o para tener una idea del tamaño de las diferentes áreas o, simplemente, para ver cómo las áreas se relacionan entre sí. De este modo, las agrupaciones de documentos muestran los intereses intelectuales comunes que comparten los autores de la comunidad científica reflejando que la ciencia es una red de especialidades interconectadas entre sí, que se pueden contemplar y analizar utilizando como base el análisis cuantitativo de la producción escrita.

Los orígenes de los primeros intentos de mapear la ciencia se remontan a la década de los años sesenta. En 1963 Price publica su libro *Little Science, Big Science*, en el que analiza el sistema de comunicación de la ciencia, advierte sobre la ausencia de una metodología objetiva y, además, presenta las bases conceptuales para representar la estructura de la ciencia a partir de sus propios datos. Se le considera el padre de la representación de dominios científicos porque fue el primero en demostrar que los patrones de citas utilizados por los autores de artículos científicos definían los frentes de investigación, y que éstos podían ser aprovechados para delinear una topología que reflejase la estructura de la producción científica de un dominio (Price, 1965). Al mismo tiempo, tiene lugar otro hito fundamental como es el nacimiento de la Teoría de la Citación de manos de Eugene Garfield. En 1963, este autor comenzó a interesarse por la construcción de mapas históricos mediante el uso de citas (Garfield, 1963). Interés que, junto a I. Sher en 1964, se plasmó en la realización de un mapa histórico en el que se mostraba el desarrollo del ADN desde Mendel hasta Niremberg. Ese mismo año, Garfield junto con Sher y Torpie, generan manualmente distintos mapas histórico-topológicos de la ciencia, a partir de las citas de la producción científica sobre la investigación en el ADN, utilizando como variable el emparejamiento bibliográfico (*bibliographic coupling*), (Garfield, Sher, & Torpie, 1964).

Durante la década de los setenta se produce un gran avance en el diseño de mapas sobre un dominio científico gracias, en gran medida, a las aportaciones de H. Small. Este autor propuso la co-citación de documentos como variable de estudio en los análisis de citas de la producción científica, utilizando como fuente de información el Science Citation Index (Small, 1973; Small & Griffith, 1974; Griffith et al., 1974). Esto supuso una revolución en el desarrollo de la representación de dominios científicos. A continuación, el trabajo de Aaronson también supone una gran aportación al campo. Sus representaciones gráficas o mapas bibliométricos (Aaronson, 1975), destacan no sólo por reflejar la evolución del mismo a través del tiempo, o por mostrar como las distintas disciplinas interactúan entre sí, también lo son por la información que suministran, pues los documentos aparecen agrupados conformando un mapa de clusters, en donde cada cluster tiene su propia denominación y número de documentos que lo componen, así como el grado de conexión entre ellos, representado por líneas acompañadas del valor de co-citación que los conectan. Además, permiten detectar los clusters más importantes, pues son aquellos que tienen un mayor número de relaciones con el resto.

Durante los años ochenta, se publican los primeros resultados por parte del Institute for Scientific Information (ISI) de su planteamiento de crear un gran Atlas de la Ciencia, hecho que fue anunciado años antes por Garfield (1975). Concretamente, en 1981 aparecen los dos primeros volúmenes referentes a la bioquímica y a la biología molecular (Garfield, 1981). Las técnicas utilizadas para la generación de los mapas no difieren mucho de las utilizadas por Small y Griffith en 1974 aunque incorporan la novedad de un nuevo posicionamiento espacial de los clusters, mediante el uso de técnicas de escalamiento multidimensional (MDS). En el transcurso de los años se publicaron otros volúmenes del Atlas de la Ciencia: Biotecnología y Genética Molecular (Garfield, 1984), Bioquímica, Inmunología y Biología Animal y Vegetal (Garfield, 1988), Farmacología (Seiden & Swanson, 1989), siendo este el último volumen publicado. Mientras tanto, Small, siguió trabajando prácticamente hasta la actualidad en el diseño de mapas de dominios científicos, depurando y mejorando las técnicas utilizadas en sus primeros mapas, convirtiéndose en el principal especialista del ISI en la investigación y elaboración de mapas de la ciencia (Small, 1981; Small & Garfield, 1985; Small &

Sweeney, 1985; Small, Sweeney & Greenlee, 1985; Small, 1993; Small, 1994; Small, 1999; Small, 2000; Small, 2003; Small & Upham, 2009; Small, 2010; Small, 2011; Boyack, Small, & Klavans, 2013).

A partir de los años 90, se produce un gran desarrollo de nuevos métodos y técnicas, tanto para la recopilación de la información, como para el análisis, visualización y posicionamiento espacial de la información (Börner, Chen, & Boyack, 2003). Esto favorece la proliferación de estudios basados en técnicas de visualización de pequeños dominios científicos, abandonando por el momento, la elaboración de mapas de grandes dominios científicos o *The Big Picture*. Entre los trabajos más destacados de este período se encuentran los de White y McCain que, proponen la representación gráfica o visualización, como modelo para el análisis y recuperación de la información que el usuario estime oportuna (White & McCain, 1997; White & McCain, 1998). También, se producen abundantes propuestas relacionadas con el uso combinado de la co-citación y el análisis de palabras (*co-word analysis*), para la generación de mapas de la ciencia, haciendo hincapié en su estructura y aspectos dinámicos (Braam, Moed & van Raan, 1991a; Braam, Moed & van Raan, 1991b; Brandenburg, Himsolt & Rohrer, 1996). Por otro lado, destaca el trabajo de Lin, Soergel y Marchionini que, desarrollan un mapa auto-organizativo (*SOM*) para representar de forma gráfica y esquemática las relaciones semánticas entre los documentos representados (Lin, Soergel & Marchionini, 1991). Por su parte, White, Lin, y McCain, realizan un estudio comparativo confrontando esta técnica de visualización con la que ofrece el Escalamiento Multidimensional (*MDS*), llegando a la conclusión de que los mapas auto-organizativos permiten una más fácil integración entre información bibliográfica y recuperación de la misma (White, Lin & McCain, 1998). Otra gran aportación de esta década es la de Chen con la incorporación de las redes *Pathfinder (PfNET)* al campo de la Documentación, como técnica de visualización de la información (Chen, 1998a; Chen, 1998b). Del mismo modo, Garfield (1998) considera que gracias a las nuevas técnicas de visualización que se están desarrollando, es posible la generación de mapas globales de la ciencia que, a través de su representación en distintos períodos temporales, o mediante el uso del zoom sobre ellos, permiten la identificación de frentes de investigación emergentes. Y finalmente,

destaca la aportación de Ding, Chowdhury & Foo (1999), que valoran que las técnicas bibliométricas permiten la simplificación de un área de conocimiento en sus elementos esenciales posibilitando la representación gráfica de dichas áreas y de sus correspondientes subáreas lo que supone una mejor comprensión del dominio.

En cuanto a la década perteneciente al cambio del milenio, la generación de mapas bibliométricos y la visualización de dominios se caracterizan, fundamentalmente, por tratar de mostrar relaciones entre documentos, detectar los autores más importantes de una determinada disciplina o analizar la estructura de un área de conocimiento y su evolución mediante su representación en sucesivos espacios temporales. Para ello, las técnicas que habitualmente se emplean son el *Clustering*, el *MDS*, el *Análisis Factorial* o el Análisis de Redes Sociales (ARS) basado en modelos de grafos, ya sea de manera independiente o de manera combinada. Para Merton (2000), el empleo de estas técnicas son una magnífica herramienta para el estudio de la Sociología de la ciencia. Incluso, para muchos autores de este período el uso de los mapas como metáforas de una disciplina científica tiene un gran potencial como interfaz para la recuperación de la información (Noyons, Moed & Luwel, 1999; Buter & Noyons, 2001; Noyons, 2001). Durante este período, se desarrollan multitud de estudios encuadrados en esta línea de trabajo que evidencian su consolidación en el campo de la Bibliometría. Dentro de los estudios más significativos se encuentran los de White, que propone la elaboración de redes centradas en el sujeto a partir del nombre de un autor que suministraría el usuario (CAMEO) (White, 2000). Posteriormente, se implementaron en un sistema dinámico de visualización (*Authorlink*), basado en la co-citación de autores, facilitando la recuperación de información en tiempo real (White, Buzydlowski & Lin, 2000; Buzydlowski, White & Lin, 2002; Lin, White & Buzydlowski, 2003). También destacan las aportaciones de Chen, Paul y O'keefe (2001), que ponen de manifiesto que la proliferación de las técnicas de visualización de la información juega un papel fundamental en el proceso de modelización y representación de la estructura o mapa intelectual de un determinado dominio, ya sea geográfico, temático o intelectual. Éstos autores presentan una propuesta metodológica para la creación de mapas del conocimiento a través de los patrones estructurales de la literatura científica que

pueden ser representados en mapas 3D. En la línea de la exploración de nuevas técnicas de visualización, Ingwersen y Larsen (2001) investigan las ventajas que presenta la representación gráfica de la producción científica de un área geográfica determinada, a través de un mapa de la producción científica en nueve áreas de las ciencias sociales de diecisiete países europeos. Del mismo modo, destacan trabajos que tratan de desarrollar mapas sobre la producción científica a través de una evolución temporal como el de Chen y Kuljis (2003), que estudian la aparición y evolución de nuevos frentes de investigación en el campo de la Física a lo largo del tiempo, o como el de Morris, Yen, Wu y Asnake (2003), que trabajan en la misma temática mediante la visualización, detección e identificación de los cambios temporales en los frentes de investigación o como el de Ding, Chowdhury y Foo (2001), que se proponen reflejar la estructura intelectual del campo de la Recuperación de la Información durante un período de 10 años (1987-1997). Sobre este mismo campo, también destaca el trabajo de Guerrero Bote, Moya Anegón y Herrero Solana (2002), que describen un método en el que los documentos de una base de datos pueden ser clasificados de forma automática mediante un mapa SOM. Por su parte, Boyack y Börner (2003), generan mapas con fines evaluativos, haciendo visible la relación entre financiación gubernamental y número de citas recibidas. Por otra parte, White (2003), adopta Pathfinder Networks (PFNeT) como técnica para la construcción de mapas de co-citación de autores (ACA), detallando sus ventajas. Del mismo modo, Glenisson, Glanzel y Persson, realizaron un estudio aplicable tanto al campo de la evaluación de la investigación como a la recuperación de la información, en el que combinaron el análisis a texto completo con métodos bibliométricos tradicionales con el objetivo de mejorar la eficiencia de los métodos individuales en el mapeo de la ciencia. Estos autores aplicaron esta metodología, en un primer momento, para cartografiar la producción científica de un número especial de *Scientometrics* (Glenisson, Glanzel & Persson, 2005) que, posteriormente, ampliaron a todas las publicaciones de la misma revista del año 2003 (Glenisson et al., 2005). Los resultados obtenidos posibilitaban que esta metodología tuviera aplicación tanto al campo de la evaluación de la investigación como al de la recuperación de la información. El estudio de Marshakova-Shaikevich (2005), proponía la elaboración de mapas bibliométricos mediante la combinación del análisis de co-citación de revistas y el de análisis de co-words, a través de las palabras clave obtenidas de los títulos y textos,

aportando una interesante propuesta para la normalización estadística de los datos. En cambio, Calero et al. (2006) proponía la aplicación de técnicas de mapeo bibliométrico combinadas con el análisis de redes para la identificación de grupos de investigación en áreas científicas concretas.

También destaca en este período el trabajo de Boyack, Klavans y Börner (2005) que, mapearon la estructura de toda la ciencia a través de 7.121 revistas que abarcaban más de un millón de documentos relacionados, tanto con el ámbito de las ciencias naturales como de las ciencias sociales. En sus mapas se podían identificar visualmente las principales áreas de la ciencia, su tamaño, similitud e interconexión de manera local como global. Asimismo, posteriormente, Klavans y Boyack (2006) analizaron cuantitativamente diez medidas de similitud que habitualmente se utilizaban en la generación de mapas bibliométricos en cuanto a la relación de los elementos de análisis, concluyendo que para datos con baja cobertura la correlación de Pearson es la medida más adecuada y para una cobertura mayor el Índice del Coseno.

Otro trabajo significativo es el de Moya-Anegón et al. (2006) que, visualizaron la estructura de la producción científica española (1990-2005) a través del análisis de la co-citación de las *ISI Subject Category* en combinación con el análisis de redes con la aplicación del algoritmo de poda *PFNeT* y el algoritmo de visualización *Kamada-Kawai*. Esta metodología fue desarrollada y ampliada en estudios posteriores donde fue posible representar la estructura de la producción científica a nivel mundial (Moya-Anegón et al., 2007) y detectar y visualizar estructuras esenciales tanto de grandes dominios como de dominios más reducidos incluso apreciar tendencias y patrones a través de análisis temporales (Vargas-Quesadas et al., 2010).

Por su parte, Porter y Rafols (2009), propusieron un índice para medir el grado de interdisciplinariedad de la ciencia mediante el uso de indicadores bibliométricos y en combinación con la generación de mapas científicos aplicado a un amplio periodo temporal (1975-2005) y a seis dominios de investigación diferentes. En esta línea

también destaca el trabajo de Rafols y Meyer (2010) que analizan el campo de la Bionanociencia, empleando también diferentes indicadores bibliométricos en combinación con la representación de redes para reflejar la novedad temática que supone en la integración del conocimiento.

En cuanto a los trabajos más recientes, es imprescindible destacar los estudios de Van Eck & Waltman que dieron un nuevo impulso a la investigación referida al área del mapeo de la ciencia gracias, en gran medida, al desarrollo del software *VOSviewer* (2010). Esta herramienta, se centra en la visualización de grandes redes de información donde se incorporan y combinan técnicas de clustering y la representación de redes basadas en la distancia entre los elementos de análisis (Waltman, van Eck & Noyons, 2010; van Eck et al., 2010). Como se verá más adelante en el apartado 6. Materiales y Métodos, este software tiene un gran potencial en el campo de la visualización y actualmente es muy utilizado, ya que, ofrece diferentes opciones de análisis y representación que aportan informaciones complementarias sobre las unidades de estudio. Son muchos los autores que incorporan este software como herramienta de análisis y mapeo de la producción científica en sus estudios (Tijssen, 2010; Leydesdorff, Hammarfelt & Akdag Salah, 2011; Rafols et al., 2012; Leydesdorff, Carley & Rafols, 2013; Romo-Fernández, Guerrero-Bote & Moya-Anegón, 2013; Leydesdorff, Park & Wagner, 2014; Zhao, & Chen, 2014; Leydesdorff, Khan & Bornmann, 2014; Gómez-Núñez et al., 2014; Kozak, Bornmann & Leydesdorff, 2015; Sedighi, 2016). Del mismo modo, en los últimos años se han desarrollado otros programas informáticos que combinan e integran técnicas de análisis y visualización de dominios para la generación de mapas bibliométricos que han contribuido a dar impulso a los estudios orientados al mapeo de la ciencia. En este sentido, destaca el trabajo de Chaomei Chen que desarrolló CiteSpace⁷ como herramienta para la obtención de mapas bibliométricos. Este software, permite visualizar y analizar la literatura de un dominio científico y obtener patrones y tendencias de la investigación a través de la información bibliográfica, especialmente, información sobre las citas de la Web of Science generando visualizaciones interactivas

⁷ CiteSpace (2003). Disponible en <http://cluster.cis.drexel.edu/~cchen/citespace/> [Fecha de consulta 10/12/2016].

(Chen, 2004; Chen, 2006; Chen et al., 2009; Chen, Dubin & Kim, 2014; Kim & Chen, 2015; Niazi, 2016). En la misma línea de trabajo, destaca el trabajo de Katy Börner que, junto a sus correspondientes equipos, ha desarrollado varias herramientas encaminadas a la obtención de mapas bibliométricos, como Network Workbench (NWB)⁸ que, consiste en un conjunto de herramientas de análisis, modelado y visualización de redes a gran escala orientada a la investigación biomédica, física y a las Ciencias Sociales (Börner et al., 2010b), o como el programa Science of Science (SCI²)⁹ que, también se trata de un conjunto de herramientas diseñado en este caso para la realización de análisis temporales, geoespaciales, temáticos y visualización de redes tanto a nivel micro, meso y macro (Börner et al., 2010a). En el caso de España, también destaca el trabajo de M.J. Cobo y su equipo de la Universidad de Granada que han desarrollado el Science Mapping Analysis software Tool (SciMAT)¹⁰ que, se trata de un software libre de cartografía científica que incorpora métodos, algoritmos y medidas implicadas en el proceso que se desarrolla en la generación de los mapas bibliométricos, desde el pre-procesamiento hasta la visualización de los resultados. Se basa en diferentes medidas de normalización y similitud de los datos para la obtención de diferentes tipos de redes, también utiliza diferentes algoritmos de clustering y permite obtener tres tipos de representaciones diferentes: diagramas estratégicos, redes de clúster y áreas de evolución (Cobo et al., 2012).

Finalmente, la tendencia en el campo del mapeo de la ciencia en los últimos años se encamina al desarrollo de nuevos métodos y técnicas que permiten obtener mayor información a la hora de generar las “instantáneas” que mejor expresen la realidad científica. En este sentido, es posible resaltar estudios dirigidos hacia diferentes aspectos del campo de la representación de la información. Por un lado, destacan los estudios basados en los denominados *Overlays Maps* que, consisten en la superposición de mapas en una sola representación donde es posible la comparación de una serie de

⁸ NWB Team. (2006). Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan. Disponible en: <http://nwb.slis.indiana.edu> [Fecha de consulta 10/12/2016].

⁹ Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies. Disponible en: <https://sci2.cns.iu.edu> [Fecha de la consulta 10/12/2016].

¹⁰ SciMAT (2012). Disponible en <http://sci2s.ugr.es/scimat/index.html> [Fecha de consulta 10/12/2016].

datos relacionados entre sí en diferentes niveles de agregación analítica (Rafols, Porter & Leydesdorff, 2010; Leydesdorff & Rafols, 2012; Chen & Leydesdorff, 2014; Leydesdorff, Kushnir & Rafols, 2014; Leydesdorff, Moya-Anegón & Guerrero-Bote, 2015; Leydesdorff, Moya-Anegón & Nooy, 2016; Bornmann & Haunschild, 2016). Por otro lado, también destacan los estudios centrados en la obtención de mapas bibliométricos que incorporan técnicas procedentes de la visualización cartográfica conocidas como *Geographic Information Systems (GIS)* (Wei, Grubestic & Bishop, 2015; Liu et al., 2015). Esto permite obtener mapas más enriquecidos, en los que se cruza la información bibliométrica con la cartográfica.

4.4 Análisis de Redes Sociales

El Análisis de Redes Sociales (ARS) es considerado un elemento fundamental en la representación y análisis de la estructura de todo tipo de dominios científicos, ya sean geográficos, temáticos o institucionales. Aunque no existe un consenso claro en la literatura científica sobre su definición, si es posible convenir que su objetivo principal se centra en la identificación y análisis de estructuras relacionales dentro de un conjunto finito de elementos (actores) vinculados entre sí.

El origen del ARS se remonta a mediados de los años treinta del s. XX y se debe, fundamentalmente, a la confluencia de diferentes teorías psicológicas, sociológicas y antropológicas, al igual que a modelos matemáticos en los que se sustenta la formalización de la misma como sucede con la teoría de los grafos. Para una revisión más profunda y detallada sobre el origen y la historia del ARS se remite al trabajo de John Scott (1991) y Carlos Lozares (1996). No obstante, tras los incipientes comienzos del ARS donde la Psiquiatría y la Antropología Social empezaban a interesarse por el estudio de pequeños grupos de individuos a través de sus relaciones, es posible reseñar como grandes precursores de este campo a J.L. Moreno (1934) y a J.A. Barnes (1954). El primero, proviene del campo de la Psicología y consiguió representar mediante lo que

denominó el *Sociograma*, las conexiones entre los actores y aplicar medidas con las que obtener patrones de comportamiento mediante el uso de matrices de coincidencia de datos. El *Sociograma* tenía la gran ventaja de transformar la información matemática contenida en matrices numéricas (datos numéricos), en información visual (grafos). Este método dio lugar a la Sociometría y precedió a los avances en la teoría matemática de los grafos, hacia finales de los 40 y principios de los 50, con el fin de dar explicación de la estructura social de pequeños grupos, intentando evidenciar el impacto de los mismos sobre la conducta individual (Galaskiewicz & Wasserman, 1993). En cuanto a J.A. Barnes, es considerado el creador del concepto de red social (*social network*). Para este autor, la red social consiste en un conjunto de lazos que vinculan a los miembros del sistema social a través y más allá, de las categorías sociales y de los grupos cerrados (Vargas-Quesada, 2005).

El desarrollo de la informática durante las décadas de los años 70 y 80, también tuvo consecuencias para este campo. Entre otras, facilitó la gestión de grandes volúmenes de datos, posibilitó la representación y análisis de estructuras complejas, la obtención de gráficos en 2 ó 3 dimensiones y la aplicación a las redes sociales de técnicas de análisis multivariante para el posicionamiento y reducción espacial de las representaciones (Freeman, 2000). Por otro lado, esto supuso un gran impulso para las matemáticas discretas, especialmente, en el desarrollo de la teoría de grafos que comenzó a servir como marco conceptual para el ARS. De este modo, el desarrollo del *Sociograma* supuso la representación gráfica de una matriz de datos a través de un grafo y la teoría de grafos proporcionó los conceptos y teoremas desde el que emprender el Análisis de Redes Sociales. Para una revisión profunda y detallada de lo que supone el Análisis de Redes Sociales y la teoría de grafos, se remite al estudio de Wasserman y Faust (1998).

La evolución de esta disciplina desde el punto de los estudios bibliométricos se ha orientado, fundamentalmente, hacia el análisis de las relaciones estructurales que se producen entre los actores de la red, desprendiéndose de un enfoque individualista centrado, únicamente, en las características de los actores. Concretamente, esta

perspectiva se sustenta bajo la premisa de que la producción científica está organizada intelectualmente en términos de disciplinas y especialidades formando una serie de estructuras que surgen y se conectan de manera cognitiva mediante los autores, los conceptos y los textos embebidos en dicha producción científica. Este concepto de estructura cognitiva, fue introducido en el ARS por D. Krackhardt en 1987 donde hasta el momento, la cognición era considerada como un atributo psicológico sólo propio de la naturaleza humana. De este modo, para los estudios bibliométricos las estructuras socio-cognitivas se refieren a la organización intelectual a un nivel supraindividual. Según L. Leydesdorff (2011) *“esta organización intelectual emerge y es reproducida por conjuntos de autores que están organizados no sólo en términos de relaciones interpersonales, sino también, de manera más abstracta, en términos de códigos de comunicación que son específicos del campo”*.

De este modo, el comportamiento relacional de las disciplinas científicas que integran la producción científica puede ser detectado, analizado y estudiado desde la perspectiva del ARS. Para comprender mejor esta información compleja por su carácter abstracto y en el que se contemplan grandes volúmenes de datos, los métodos y herramientas que brinda el campo de la visualización de la información se muestran especialmente útiles como se ha indicado en el apartado 4.2. Visualización y representación de la información. Más concretamente, la visualización y representación de las disciplinas científicas, en este caso, detectando y analizando su comportamiento relacional, constituye el objetivo fundamental en la construcción y generación de los mapas que representan el conocimiento científico, es decir, en lo que el apartado anterior se ha definido como mapeo de la ciencia. De este modo, los mapas de la ciencia se convierten en una herramienta muy adecuada para comprender la interacción de las disciplinas científicas y permitir navegar por la literatura científica por medio de la representación de sus relaciones espaciales (Garfield, 1986). En consecuencia, la obtención de estos mapas favorece y facilita el análisis de dominios al mostrar de forma gráfica la estructura y relaciones de los elementos que se representan en el dominio. El dominio se recrea a partir del discurso de la comunidad en que se forma y es un método objetivo, porque posibilita el análisis de la estructura no subjetiva, por medio del consenso intelectual de

las relaciones existentes entre los elementos que lo representan (Vargas-Quesada, 2005). En síntesis, el ARS en combinación con las técnicas de visualización de la información que, permiten la obtención de mapas bibliométricos científicos permite la representación de dominios no como grupos de elementos aislados en los que sólo se estudian sus características, sino como conjuntos completos en los que además se estudian sus relaciones y se establecen conclusiones sobre ellas.

4.5 Delimitación temática

En relación con la metodología que, desde el campo de la Bibliometría se puede aplicar a la hora de delimitar y visualizar un campo científico, existe una gran variedad de técnicas que se pueden emplear con el fin de detectar las principales líneas o temas de investigación. La mayoría están basados en los principios de la co-citación. Básicamente, ésta puede ser definida como la frecuencia con la que dos elementos bibliográficos (autores, documentos, revistas, categorías, etc.), son citados por otros documentos que han sido publicados con posterioridad. Fue formulada por primera vez de forma independiente y simultánea por Small (1973) y Marshakova (1973), proponiendo el análisis de la co-citación como un modelo objetivo para revelar la estructura intelectual de las especialidades científicas. En un primer momento, fue desarrollada tomando como unidad de análisis a los documentos, con el objetivo de medir el grado de relación que se produce cuando dos ítems de la literatura existente son citados juntos en la literatura posterior. Con el paso del tiempo, fueron desarrollándose otras medidas de co-citación mediante el empleo de otras unidades de análisis, siendo las más representativas para la delimitación temática los autores (Análisis de Co-citación de Autores), las categorías (Análisis de Co-citación de Categorías), las palabras clave (Análisis de co-palabras o co-words), etc. A continuación, se explica, brevemente, en qué consiste cada una de ellas.

4.5.1 Emparejamiento bibliográfico

También conocido como análisis de co-citación de documentos o de trabajos. Como se ha indicado, es el precursor del resto de análisis basados en la co-citación y se define como la frecuencia con la que dos documentos son citados por otros documentos que han sido posteriormente publicados.

El emparejamiento bibliográfico parte de la asunción de que entre dos o más documentos que son co-citados (citados juntos) en un tercer y posterior trabajo existe, al menos desde la perspectiva del autor citante, una similitud temática, y que cuanto mayor es la frecuencia de co-citación, mayor será la afinidad entre ellos. La intensidad de esta relación está dada, entonces, por la cantidad de documentos citantes que tienen el mismo par de documentos en sus referencias. Asimismo, si dos documentos son frecuentemente co-citados, ellos serán también, necesariamente, frecuentemente citados. Por tanto, las relaciones entre documentos pueden ser utilizadas para representar la estructura científica de un dominio o incluso para la detección de sus frentes de investigación. Entre los trabajos que han utilizado este tipo de análisis se pueden destacar los siguientes: Kessler, 1963; Small & Griffith, 1974; Griffith et al., 1974; Garfield, 1981; Small & Garfield, 1985; Small, 1997; Garfield, 1998; Boyack & Klavans, 2010; Chen et al., 2011; Glänzel & Thijs, 2011.

4.5.2 Co-citación de Autores

El Análisis de Co-citacitación de Autores (ACA) fue propuesta por primera vez en 1981 por White y Griffith como una nueva técnica para contribuir al conocimiento de la estructura intelectual de las disciplinas científicas, entendiendo por “autor” el conjunto de trabajos escritos por una persona. En este contexto, la co-citación resulta cuando alguien cita cualquier trabajo de cualquier autor, junto con cualquier trabajo de cualquier otro autor en un nuevo documento que le es propio (White and Griffith, 1981).

Se puede considerar el ACA como una herramienta de gran potencial para la visualización de la estructura intelectual de diferentes ámbitos temáticos ya que, muestra y valida dicha estructura por medio del consenso de sus principales autores, que son los que, al fin y al cabo, componen el dominio científico. Existe gran cantidad y variedad de trabajos que utilizan esta metodología como herramienta para la delimitación temática. A pesar de ello, se destacan a continuación los más representativos: White & McCain, 1998; Chen & Paul, 2001; Chen, Paul & O'keefe, 2001; White, Buzydlowski, & Lin, 2000; White, 2000; White, 2003; Small & Upham, 2009; Chen, Ibekwe-SanJuan & Hou, 2010; Zhao & Strotmann, 2011; entre otros.

4.5.3 Co-citación de categorías temáticas

El análisis de co-citación a través de las categorías temáticas se define de manera similar a la de los casos anteriores, es decir, como la frecuencia con la que dos categorías son citadas por otras categorías. En concreto, se refiere al número de veces que dos categorías co-ocurren conjuntamente en los documentos. Resulta ser una metodología muy interesante, sobre todo, para el análisis de grandes dominios temáticos, ya que, permite detectar la estructura relacional de grandes áreas o disciplinas. Su empleo y desarrollo corresponde a años más recientes que en los casos anteriores. Uno de los precursores fue F. de Moya Anegón que, en 2004 junto con su equipo, propusieron la co-citación de categorías ISI como una nueva técnica para la construcción de mapas de grandes dominios científicos. Posteriormente, Vargas-Quesada (2005), planteó la aplicación de PfNET (Schvaneveldt, R. W., 1990) como algoritmo de poda con el que mostrar la estructura más “saliente” de un dominio, así como, utilizar valores de co-citación sin ningún tipo de normalización, para provocar la agrupación de las categorías ISI en grandes áreas temáticas a modo de racimos. A continuación, aplicaron esta misma metodología, para la visualización y análisis de la estructura científica tanto del dominio español (Moya Anegón et al., 2006), como del mundo (Moya Anegón et al., 2007), poniendo de manifiesto la macro-estructura, micro-estructura y columna vertebral de la investigación. Posteriormente, otros autores siguieron esta opción de análisis para detectar y analizar estructuras temáticas. Entre los trabajos más relevantes destacan:

Leydesdorff & Rafols, 2009; Rafols & Meyer, 2010; Rafols, Porter & Leydesdorff, 2010; Wolfram & Zhao, 2014; Wang & Wolfram, 2015; entre otros.

4.6 Análisis de co-palabras

Los análisis de co-citación basados en las palabras como unidades de análisis también son conocidos como análisis de co-palabras o *Co-Words Analysis*. Se pueden definir como la frecuencia de aparición de los términos que aparecen de manera conjunta en los documentos, es decir, en el estudio de su co-ocurrencia. En virtud de la fuerza de las asociaciones entre los términos más representativos de la literatura científica, es posible obtener información sobre la estructura y las tendencias de una disciplina científica (Ding, Chowdhury & Foo, 2001). De este modo, si se combinan los análisis de co-palabras con el de ARS y con las técnicas de visualización de la información, es posible obtener mapas bibliométricos que representen los principales aspectos temáticos que son tratados en los documentos.

4.6.1 Evolución histórica

Los principios del análisis de co-palabras vienen fraguándose desde la década de los años 80 (Callon et al., 1983) cuando se plantea revelar la estructura de la ciencia y su evolución a través de la co-ocurrencia de palabras. El fundamento principal se basa en que la aparición conjunta de dos términos en un documento indica una relación entre los conceptos que representan (Cambrosio et al., 1993; Courtial, Cahlik & Callon, 1994) y pone de manifiesto la estructura y tendencias de una disciplina científica, por medio de la fuerza de las asociaciones entre los términos representativos de la literatura publicada (Ding, Chowdhury & Foo, 2001). Por su parte, Van Raan y Tijssen (1993), aplicaron este tipo de análisis a la investigación realizada sobre redes neuronales, llegando a la conclusión de que posee un valor epistemológico que permite descubrir relaciones insospechadas entre los conceptos de una disciplina y revelando, además, problemas que de otra forma pasarían inadvertidos. Ding, Chowdhury y Foo (2001) lo utilizaron para mapear la estructura intelectual del campo de la Recuperación de la

Información, mostrando patrones y tendencias en dicho campo. Baldwin et al. (2003) lo emplearon para mapear la investigación en ética y demencia. Onyancha y Ocholla (2005), recurrieron a él para ayudar a los investigadores y a los gestores de la ciencia a identificar nuevas líneas de investigación sobre el VIH/SIDA, así como, los vínculos que se establecen para planificar la investigación y formular las políticas científicas adecuadas. Bessalar y Heimeriks (2006) lo emplearon para estudiar las publicaciones de Ciencias de la Información producidas entre 1986 y 2002 a través de las palabras más significativas de los títulos y de las referencias. Chalík y Jirina (2006) lo relacionaron con la ley acumulativa de la ciencia de Price aplicado al campo de la economía. Por su parte, Lee (2008) utilizó este modelo de análisis para identificar las tendencias y aspectos subyacentes dentro del área de la Seguridad de la Información. Tian, Wen, & Hong (2008) mediante el uso de palabras clave, lo emplearon para medir la producción científica en el campo de Sistemas de Información Geográfica (SIG) de la producción científica recogida en la base de datos del Instituto de Institute for Scientific Information (ISI). En cambio, Neff y Corley (2009) recurrieron a él, para estudiar las publicaciones producidas en Ecología durante el periodo comprendido entre 1970 y 2005. Leydesdorff & Welbers (2011), se centraron en el análisis de co-palabras para analizar, comparativamente, las medidas semánticas de los patrones de similitud (correlaciones) y las variables latentes (factor de análisis) mediante técnicas de computación y el uso de estadísticos.

Asimismo, se han realizado otros análisis de co-palabras similares, para mapear la evolución del conocimiento en diversos campos tales como la ingeniería de software (Coulter, Monarch & Konda, 1998), la química (Callon, Courtial & Lavoie, 1991), la investigación en redes neuronales (Noyons & Van Raan, 1998), la optometría (Noyons & Van Raan, 1994), la bioelectrónica (Hinze, 1994), las reacciones adversas a los fármacos (Rikken, Kiers & Vos, 1995; Clarke et al., 2007), la biotecnología (Rip & Courtial, 1984; Looze & Lemarie, 1997), ciencias ambientales (Ho, 2007), la física de la materia condensada (Bhattacharya & Basu, 1998), el síndrome respiratorio agudo severo (Chiu, Huang & Ho, 2004), el fenómeno del tsunami (Chiu & Ho, 2007), la enfermedad de

Parkinson (Li, Ho & Li, 2008), el grafeno (Lv, 2011) y la seguridad del agua (Huai & Chai, 2016).

4.6.2 Perspectivas de la investigación en estos estudios

La tendencia de los últimos años en los análisis de co-palabras como herramienta para la delimitación temática, refleja un gran desarrollo de estos estudios, encaminándose hacia varias direcciones. Una de ellas, se centra en la elección adecuada de las palabras como un aspecto fundamental a la hora de posibilitar la delimitación temática en cualquier campo de investigación. Predomina la preocupación por conocer cuáles son los términos que tienen mayor capacidad de expresar el contenido conceptual de los documentos, así como, su pontencial relacional a la hora de reflejar estructuras que permitan la definición de aspectos temáticos, sobre todo, en niveles de agregación muy específicos. Entre otros, este es uno de los objetivos principales que propician el desarrollo de esta tesis doctoral y que ha estado muy presente en todas las publicaciones que la integran. El empleo de diferentes términos, según su naturaleza documental, a la hora de representar sus relaciones estructurales y analizar su potencial para la localización e identifcación de líneas de investigación y otros aspectos temáticos de carácter específico.

Existen una gran variedad de criterios a la hora de seleccionar los términos que se consideran más adecuados como unidades de análisis. Uno de ellos se basa en la extracción de palabras clave a través de los títulos o resúmenes de las publicaciones. En esta línea es posible localizar trabajos precedentes como los de Stapley & Benoit (2000), Chaussabel & Sher (2002), Wren & Garner (2004) y Chen & Sharp (2004). También existen trabajos más recientes como el de Milojevic et al. (2011), que utilizan las palabras procedentes de los títulos de los artículos con el objetivo de analizar la estructura cognitiva de la categoría *Library Information Science (LIS)* durante el período 1988-2007 y, así, detectar y representar las tendencias más representativas del campo. En este sentido, también destaca el trabajo de de Yao et al. (2012), en el que, no sólo utilizan las palabras clave procedentes de los títulos, sino que también, utilizan las

palabras clave empleadas por los autores (*Author Keywords*) de las publicaciones para sintetizar su contenido y que son recogidas por las bases de datos como elementos de indización documental. De este modo, son cada vez más los autores que utilizan los términos de indización empleados por las bases de datos, como unidades de análisis en los estudios de co-palabras. Este es el caso del trabajo de Su & Lee (2010) que, visualizaron la estructura de conocimiento del área *Technology Foresight* mediante la generación de mapas bibliométricos de términos obtenidos a partir de los *Author Keywords* de los documentos. También, en el trabajo de Romo-Fernández, Guerrero-Bote & Moya-Anegón (2013), se utilizan las palabras clave que la base de datos Scopus emplea para indizar los documentos, con el objetivo de analizar la producción científica sobre Energías Renovables. En este caso, se utilizan tanto términos procedentes del lenguaje natural, como son los *Author Keywords*, como términos procedentes de un vocabulario controlado, como son los *Index Terms*. El empleo de éstos últimos como unidades de análisis en los estudios de co-palabras también ha sido utilizado por otros autores como Kestin, Rindfleisch & Hristovski (2014) que, emplearon los descriptores *MeSH*, términos que proceden de un Tesauro y que son indizados de manera manual por la base de datos *Medline*, con la intención de estudiar las relaciones estructurales que se producen entre términos procedentes de un vocabulario controlado. Estos autores, obtuvieron dos redes de descriptores que compararon entre sí, una con todos los términos y otra sólo con *Major MeSH* (términos que aluden a la temática principal del texto). En cambio, autores como Zhang et al. (2016), se han centrado en los términos de indización que emplea la base de datos *Web of Science (WOS)*, donde además de emplear los *Author Keywords* se utilizan los *Keywords Plus*, que son palabras clave extraídas, automáticamente, de las referencias de los documentos. Precisamente, la extracción automática de las palabras clave es otro aspecto de la investigación que es objeto de estudio para muchos autores. Es posible encontrar trabajos donde se ha recurrido a la extracción de texto libre para obtener las unidades de análisis, (Kostoff et al., 2005; Janssens et al., 2006). También, existen trabajos donde se obtienen palabras clave con la ayuda de un software especializado, como ocurre en el estudio de Zahng et al. (2012) en el que analizan, de manera temporal, la producción científica sobre el tratamiento de la adherencia en pacientes durante los años 2000 al 2011, a través de la extracción de palabras clave mediante el software

Biblexcel. También, el estudio de María Pinto (2015), en el que se analizan los aspectos temáticos más relevantes de la producción científica sobre Evaluación de la Alfabetización Informacional en la Educación Superior (ILAHE), a través del análisis de términos extraídos automáticamente con el software *Atlas.ti*, pero en este caso se aplica únicamente a determinados campos bibliográficos (título, resumen, *Author Keywords* y *Keywords Plus*). En el trabajo de Xie (2015), también se realiza una extracción automática de palabras clave centrada en los mismos campos que en el estudio anterior.

No obstante, otro de los aspectos que caracteriza a las últimas tendencias de investigación en el análisis de co-palabras, se orienta al desarrollo de metodologías que, combinan diferentes análisis estructurales basados en el empleo de diferentes medidas de co-citación. La idea fundamental, reside en que la combinación de diferentes análisis genera un estudio integral enriquecido por los diferentes enfoques metodológicos. De este modo, destacan trabajos en los que se han empleado tanto análisis de citas como análisis de co-words. Como ocurre con el estudio de Kajikawa et al. (2008) que utilizaron un análisis de las citas recogidas del *Science Citation Index (SCI)* y las palabras clave de los documentos para analizar los cambios estructurales que se han producido en el campo de las energías sostenibles o en biomasa y biocombustibles (Kajikawa & Takeda, 2008), o en los diodos orgánicos de emisión de luz (Kajikawa & Takeda, 2009). También, en el estudio de Uddin & Khan (2016), donde emplearon los *Author Keywords* recogidos en los documentos en relación con el recuento de citas, generando y analizando redes de co-ocurrencia a través de medidas de centralidad procedentes del *ARS*.

Por otro lado, también existen abundantes estudios que basados en la combinación de Análisis de Co-citación de Autores (ACA) y análisis de co-palabras. En este sentido, es posible localizar los trabajos de Zitt & Bassecouard, 1996; Zitt & Bassecouard, 2006; Zitt, Lelu & Bassecouard, 2008; Zitt, Lelu & Bassecouard, 2011; Jung, 2013.

Por último, también se están desarrollando nuevos modelos de análisis en los que están presentes nuevos elementos de análisis. Este es el caso del trabajo de Yang et al.

(2016), donde se obtienen mapas bibliométricos creados a partir de la combinación de dos modelos de análisis: uno basado en el *Author Bibliographic Coupling Analysis (ABCA)* y el otro basado en el *Author Keyword Coupling Analysis (AKCA)*, desarrollado a partir del análisis de co-palabras.

4.6.3 Ventajas e inconvenientes del análisis de co-palabras para la delimitación temática

Como ha quedado de manifiesto, desde la perspectiva del AD el lenguaje y las palabras cobran una gran relevancia ya que son las encargadas de sostener el discurso científico elaborado por una comunidad encargada de generar nuevo conocimiento y que es el reflejo de la división social y laboral de la sociedad. Según Ma (2006), *“las relaciones semánticas de los signos dependen de una sociocultura específica y de la Historia”* y para Wang (2013), *“es necesario estudiar los contextos culturales e históricos inherentes en los signos”*. De este modo, las palabras cobran un gran protagonismo para el AD y para la delimitación temática, ya que aluden al contenido de los documentos permitiendo identificar y analizar los temas que despiertan mayor interés en los científicos (He, 1999).

Asimismo, las palabras pueden definirse como las unidades mínimas de expresión en las que se puede resumir el contenido de un documento. Esto todavía se hace más evidente si se trata de términos de indización, ya que, por su naturaleza como lenguaje documental para el Análisis de Contenido (AC), tienen mayor vocación de sintetizar los temas o aspectos conceptuales más relevantes que son tratados en los documentos. Esta característica es muy importante desde el punto de vista de la delimitación temática, porque, permite analizar la producción científica a nivel documental, hecho que no es posible si se toman unidades de análisis mayores, como es el caso de las categorías temáticas. Al mismo tiempo, al ser las unidades mínimas de expresión conceptual también permiten realizar análisis temáticos de granularidad más fina que, una vez más, en el caso de las categorías temáticas. Su versatilidad conceptual otorga a las palabras y especialmente, a los términos de indización, de una gran capacidad de

identificar aspectos temáticos más específicos que las áreas o categorías temáticas, permitiendo descender a niveles de análisis mucho más específicos y especializados como pueden ser las subdisciplinas científicas, líneas de investigación, sublíneas de investigación, etc., profundizando tanto como los resultados que se pretendan obtener.

Como consecuencia de esta capacidad para identificar aspectos temáticos de carácter más específicos, los análisis de co-palabras resultan ser muy adecuados, no sólo para estudios a nivel macro, sino que, también permiten análisis a nivel meso o micro. Una delimitación temática con mayor profundidad de análisis permite la caracterización de dominios más concretos. Por lo tanto, los análisis de co-palabras, podrían ser empleados para identificar, por ejemplo, líneas de investigación tanto de un país, como de instituciones o centros, o sobre determinados grupos de investigación.

Por otro lado, según Michel Zitt y colaboradores (2011), las palabras, a diferencia de otras unidades de análisis como por ejemplo las citas, parecen adaptarse mejor a la “ciencia viva” porque son universales y son coetáneas al propio documento (Soos, Kampis & Gulyás, 2013). Para estos autores las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes. Esto todavía se aprecia con mayor claridad si se trata de términos de indización como las palabras clave, ya sean procedentes de los autores o extraídas de manera automática. Éstas, favorecen la aparición de una terminología más actualizada y específica en comparación con otros términos de indización, como son los procedentes de un vocabulario controlado (Braam; Moed, & Van-Raan, 1991). Las palabras clave son unidades terminológicas muy permeables a los cambios de tendencias, proceden del lenguaje natural por lo que tienen un carácter muy actualizado, permitiendo detectar los cambios de la Ciencia con gran inmediatez. Como demuestran estudios recientes, estas cualidades contribuyen a reflejar mejor los aspectos temáticos más dinámicos de un dominio. Facilitan, sobre todo, tanto la detección de nuevas tendencias de investigación como de aquellas que

comienzan a desaparecer, pese a los problemas de dispersión terminológica que plantean.

No obstante, seleccionar las palabras para el análisis temático de un dominio puede ser objeto de crítica. Leydesdorff, (1997) considera que las *“palabras y las co-palabras no pueden mapear el desarrollo de las ciencias”*. Su argumento se basa en la idea de que las palabras que contienen los documentos a veces no son lo suficientemente específicas y pueden variar de un contexto a otro. De ahí que, Leydesdorff haya desarrollado estudios posteriores encaminados a salvar este problema y poder realizar estudios de co-words teniendo en cuenta los contextos de las palabras (Leydesdorff & Hellsten, 2006; Leydesdorff & Welbers, 2011).

5. LIMITACIONES

Como ocurre en cualquier trabajo exhaustivo existen una serie de inconvenientes o salvedades que forman parte de la propia investigación y que han de ser tenidas en cuenta para la comprensión del estudio. En este caso, las limitaciones que afectan a esta tesis se resumen a continuación.

En primer lugar, las bases datos que se utilizan como fuentes de información en este estudio tienen por definición una serie de sesgos que, inevitablemente, afectan en la recuperación de la producción científica que se pretender estudiar. Principalmente, se pueden indicar los siguientes (Gómez Nuñez, 2014):

- Sesgo temático reflejado, por ejemplo, en una extensa representación de áreas como la Ingeniería y, en especial, las Ciencias de la Vida y de la Salud.
- Sesgo temporal ocasionado como consecuencia de indexar únicamente aquellas publicaciones originadas en el periodo temporal transcurrido desde el año 1900 en el caso de la *Web of Science* y desde 1950 en el caso de *Medline*, hasta la actualidad.
- Sesgo geográfico hacia ciertos países anglosajones como UK y, principalmente, USA.
- Sesgo idiomático claro en favor del inglés y escasa cobertura de otros idiomas.
- Sesgo editorial derivado principalmente de la inclusión en el sistema de amplios paquetes de revistas pertenecientes a grandes grupos editoriales, como Springer, Taylor & Francis, Blackwell y Elsevier.

En segundo lugar, y también en relación con las bases de datos bibliográficas, éstas ocasionan una serie de inconvenientes debido, principalmente, a la ausencia de normalización de la mayoría de sus campos, dado que no están diseñadas originalmente, para la explotación bibliométrica. Éstos problemas afectan, sobre todo, a tres aspectos fundamentales como son la afiliación institucional y a las indicaciones geográficas (países, ciudades, centros, departamentos, etc.), los nombres de los autores y a los términos de indización en el caso de que no se trate de un vocabulario controlado como ocurre en Medline.

Asimismo, las estrategias de búsqueda diseñadas en esta tesis para la creación de la colección documental, no han sido elaboradas bajo la pretensión de obtener toda la producción científica de la manera más exhaustiva y precisa posible en relación con los temas de estudio. El objetivo de estas estrategias está más orientado a contar con una cantidad de publicaciones lo suficientemente representativa, con un pool de estudios más o menos relacionados por una temática determinada sobre el que realizar una delimitación temática mediante el desarrollo metodológico que se plantea en esta tesis.

Finalmente, también se ha de tener en consideración las limitaciones que se derivan del empleo de los programas informáticos seleccionados en esta propuesta metodológica para la generación de las redes de información científica y la elaboración de los mapas bibliométricos. En el caso del empleo de *Pajek* para la visualización de las relaciones entre los términos, es necesario seleccionar un umbral máximo de términos a representar dado que no es posible representar redes muy amplias que superen los 100 – 150 términos. A pesar de que se utilice el algoritmo de poda *PfNET* que esquematiza y reduce el número de relaciones, sigue siendo necesario el establecimiento de un umbral máximo de términos para que la visualización de las estructuras relacionales sea clara y comprensible. Esa es la razón por la que se seleccionan los términos más frecuentes. En cambio, en el caso de *VOSviewer*, no es necesario reducir el número de ítems para ser representados ya que este programa

permite la representación de redes más amplias, aunque tiende a mostrar en primer lugar en las visualizaciones los nodos que tienen mayor peso en la red debido al solapamiento que se produce con los nombres de los mismos.

6. MATERIALES Y MÉTODOS

La propuesta metodológica que se presenta en esta tesis doctoral se centra en la combinación de varias técnicas bibliométricas. En primer lugar, se basa en el empleo del análisis de co-palabras para la identificación de los términos más representativos a partir de su aparición conjunta en los documentos. En este caso, se considera que los términos de indización son las unidades de análisis más adecuadas para la representación conceptual de los documentos. De este modo, se produce un conjunto de relaciones entre los términos que puede ser analizada y representada a partir de programas informáticos especializados en el Análisis de Redes Sociales y técnicas de representación y visualización de la información, obteniendo como resultado un mapa bibliométrico en el que se plasma la estructura relacional de los principales aspectos temáticos del dominio. A continuación, se detalla en profundidad los parámetros metodológicos que forman parte de la propuesta de esta tesis doctoral.

6.1 Materiales

6.1.1 Fuentes de información

La elección de las fuentes de información en cualquier estudio bibliométrico resulta fundamental para la consecución de los objetivos que se pretenden alcanzar. Las bases de datos con información bibliográfica son las principales fuentes para los análisis bibliométricos. De esta realidad son cada vez más conscientes los productores de las principales bases de datos que incluyen con mayor frecuencia mejoras que enriquecen la explotación bibliométrica. No obstante, para una interpretación adecuada de los resultados es necesario conocer bien la cobertura temporal, geográfica y temática; la estructuración y exportación de los campos de información; el uso de determinados índices y clasificaciones temáticas o las peculiaridades de los sistemas de indización que utilizan, entre otras características (Costas, Moreno & Bordons, 2008).

En este caso, la elección es muy importante porque de ello depende, en gran medida, la definición de los dominios a analizar. Los criterios de selección están muy condicionados por la dualidad de dominios que se estudian en esta tesis en relación con la propuesta metodológica. Por un lado, se ha seleccionado la base de datos Medline¹¹ para la identificación de las principales líneas de investigación sobre un tema de carácter general dentro del campo de la Biomedicina, como son los estudios sobre Salud y Mujer, con una amplia cobertura temporal (1965-2005) y sin delimitación geográfica. Dada la amplitud del dominio se ha considerado adecuado seleccionar una base de datos cuya especialización temática favorezca una discriminación documental inicial. Por otro lado, se ha utilizado la base de datos Science Citation Index Expanded (SCI-EXPANDED)¹² para la identificación de las principales líneas de investigación sobre un aspecto biomédico de carácter más específico como es la investigación con células madre, durante un período temporal más reducido, de 14 años (1997-2012) y circunscrito al ámbito español. Ya que se trata de un dominio más específico no se ha considerado inapropiado utilizar una fuente de carácter multidisciplinar.

6.1.1.1 Medline

Se trata de una base de datos bibliográfica creada por la *National Library of Medicine* (NLM)¹³ de Maryland, Estados Unidos. Contiene más de 22 millones de referencias desde 1946 e indexa cerca de 5.600 revistas. Está especializada en biomedicina y salud, abarcando también ciencias de la vida, del comportamiento, ciencias químicas y bioingeniería. Este hecho, es muy significativo para la primera fase de análisis de esta tesis. Los estudios sobre salud desde una perspectiva de género pueden suponer un gran reto a la hora de su recuperación en una base de datos de carácter multidisciplinar. A pesar de utilizar una estrategia de búsqueda muy definida se corre el riesgo de recuperar estudios que no estén relacionados con el campo de la biomedicina. Más aún, cuando el período de estudio es tan amplio, 40 años. Al contar con una base de datos

¹¹ Medline (2016). Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed> [Fecha de consulta: 5/11/2016]

¹² Web of Science (2016). Disponible en <http://wokinfo.com/> [Fecha de consulta: 5/11/2016]

¹³ National Library of Medicine (2016). Disponible en: <https://www.nlm.nih.gov/> [Fecha de consulta: 6/11/2016]

especializada en biomedicina se elimina el riesgo de obtener un exceso de “ruido documental” en la recuperación de la información.

Además, Medline posee otra característica que fue determinante para su elección como fuente de información de dicha fase de análisis de esta tesis. Se trata del sistema de indización que emplea para facilitar las búsquedas y recuperación de la información. Es una indización de carácter manual que se basa en el empleo de términos procedentes de un vocabulario controlado que forman parte del tesoro *Medical Subject Headings* (MeSH)¹⁴ proporcionado también por la NLM. Éste, está compuesto por 25.000 descriptores estructurados jerárquicamente en 16 categorías que cubren todos los aspectos relacionados con la biomedicina. Dado que las principales unidades para el análisis del estudio son las palabras es necesario tener en cuenta la terminología que caracteriza el dominio. En este caso, la mayor parte del léxico está compuesto por términos que tienden a la ambigüedad y son poco específicos. La posibilidad de acceder a un conjunto documental indizado manualmente con un vocabulario controlado, disminuye la ambigüedad y la escasa especificidad de los términos procedentes del lenguaje natural que se expresa en los documentos del dominio.

Por último, otro de los aspectos que se tuvieron en cuenta en la elección de Medline como fuente de información fue la ausencia de una delimitación geográfica para el dominio. Como es sabido, Medline no tiene normalizado el campo relativo a las afiliaciones de los autores que firman los documentos, ni siquiera los datos referidos al país, a diferencia de otras bases de datos. Este hecho, suele suponer un inconveniente a la hora de hacer estudios bibliométricos circunscritos a países. En este caso, esta circunstancia no plantea problemas porque el dominio de estudio no está determinado por ningún país o región geográfica.

¹⁴ Medical Subject Headings (2016). Disponible en : <http://www.ncbi.nlm.nih.gov/mesh> [Fecha de consulta: 6/11/2016]

6.1.1.2 *Science Citation Index Expanded*

Esta base de datos está elaborada por la empresa estadounidense Thomson Reuters (anteriormente, por el *Institute for Scientific Information*, ISI) es una de las bases de datos que integran el portal electrónico *Web of Science* (WoS). Indexa más de 12.000 revistas de gran prestigio internacional siendo una de sus principales características su cobertura temática de carácter multidisciplinar. No se consideró que este rasgo plantease ningún inconveniente para la elección de esta base de datos como fuente para la localización y visualización de las principales líneas de investigación de la investigación española con células madre desde 1997 al 2012. Al tratarse de un dominio concreto cuya definición es clara tanto por su cobertura temporal como geográfica o temática, está dotado de identidad suficiente como para que el uso de una base de datos multidisciplinar no suponga un inconveniente a la hora de la recuperación documental.

Además, a diferencia del dominio de salud y mujer (1965-2005), este dominio tiene una terminología más específica. El discurso científico está compuesto por un léxico más definido, cuyos términos mayoritariamente, describen de manera unívoca los conceptos biomédicos, reduciendo los posibles problemas de ambigüedad terminológica como la homonimia, sinonimia o polisemia. Esto contribuye a la posibilidad de utilizar otro tipo de términos que no correspondan a un vocabulario controlado como unidades de análisis a la hora de aplicar la metodología que se plantea en esta tesis doctoral. El sistema de indización utilizado por el SCI-EXPANDED es lo que motivó su elección. Los documentos son indizados con dos tipos de palabras clave de diferentes naturalezas. Por un lado, las palabras clave que utilizan los autores para expresar el contenido conceptual de los documentos, *Author Keywords* (AKW). Estas palabras clave responden a una indización de carácter manual a partir de la selección de términos procedentes del lenguaje natural. Por otro lado, el SCI-EXPANDED utiliza una indización automatizada mediante el uso de las palabras clave *KeyWords Plus* (KW+). Éstas, se obtienen a partir de un algoritmo de búsqueda que extrae los términos de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Gierfield & Sher, 1993). Es un sistema de indización propio y exclusivo de esta base de datos ideado por Garfield en los años 90.

Finalmente, otras de las razones por las que se seleccionó esta base de datos es porque, a diferencia de otras, aparecen todas las afiliaciones de los autores que firman los documentos y además el campo país está normalizado. Esto posibilita la elaboración de estudios bibliométricos delimitados geográficamente por países. En el caso del dominio de la investigación española con células madre esta característica resulta fundamental.

6.1.2 Recogida y tratamiento de la información

6.1.2.1 Estrategias de búsquedas

Teniendo en cuenta que la propuesta metodológica que se presenta en esta tesis se aplica a dos dominios distintos, se han diseñado diferentes estrategias de búsqueda en función de cada uno de los dominios y de los objetivos a alcanzar en cada fase de estudio.

➤ Dominio 1

Para este dominio se diseñó una estrategia de búsqueda que se dividió en dos fases y que tuvo lugar el 18 de abril del 2006:

- **Primera fase:** tiene como fin obtener los términos más representativos de las investigaciones sobre salud y mujer. Es decir, aquellos que representan los conceptos sobre los que más se ha investigado en esta área. Para ello, se realizó una primera búsqueda con los 9 descriptores Desc (Descriptores en Ciencias de la Salud) (BIREME, 1967) que indican cuales son los descriptores más apropiados para la búsqueda y recuperación de información, a través del MeSH, que hacen referencia al concepto “mujeres”. Estos términos son, Women, Women’s health, Women’s Health Services, Women’s rights, Battered women, Women, Working, Dentists women, Physicians women y Pregnant women. Con el fin de recuperar los documentos más relevantes, los descriptores se

buscaron en el campo MajorMesh, y se utilizaron las etiquetas Human y Female. En esta primera búsqueda se recuperaron 19.524 documentos (journal articles, review, historical article, editorial, letter, comment, news, case reports, clinical trials).

- **Segunda fase:** tiene como fin extraer y delimitar claramente el objeto de estudio, así como su corolario o “family words” (Van Raan y Tijssen, 1993). Para ello, se tomaron los sesenta descriptores más ocurrentes de los 19.524 documentos recuperados en la fase anterior y se volvieron a lanzar nuevamente contra la base de datos Medline. El resultado es un total de 365.714 documentos representados por 1.300.881 descriptores recogidos en el campo *MajorMesh*.

➤ Dominio 2

Para este dominio se diseñó una estrategia de búsqueda cuyo periodo temporal fue revisado en dos ocasiones con el fin de obtener los datos más recientes.

La primera búsqueda tuvo lugar el 15 de abril del 2008 y se utilizaron los términos “stem* cell*” en el campo *Topic*. Este campo localiza los términos o frases en partes concretas de los documentos, como el título, el resumen los *Author Keywords* (AKW) y los *KeyWords Plus* (KW+). Además, se delimitó la búsqueda por “Spain” en el campo *Address* y por el período temporal [1997-2007]. El número total de documentos que se recuperaron fue de 2.467. Estos datos han sido utilizados en las publicaciones nº 2 y 3.

Posteriormente, se realizó una de las actualizaciones ampliando el período de estudio al 2010. Para ello, se sustituyó el periodo temporal por [2008-2010]. La descarga de estos datos tuvo lugar el 7 de noviembre del 2011 y se añadieron a los del período

anterior. Los datos del período completo (1997-2010) fueron utilizados en las publicaciones nº 4, 5 y 6¹⁵ y el total de documentos analizados fue de 4.148.

6.1.2.2 Organización de la información. Estructura relacional

La información recuperada de las bases de datos bibliográficas, generalmente, requiere de un tratamiento previo antes de su análisis bibliométrico. Es preciso estructurarla en unidades de información que permitan analizarlas de manera individualizada pero siempre relacionadas con el registro bibliográfico del que proceden. De este modo, se desarrolló una base de datos relacional en Access que posibilitó la ordenación y estructuración de los campos, así como, el estudio y análisis de los datos.

6.1.2.3 Tratamiento de la información. Normalización

Uno de los principales problemas del uso de las bases de datos bibliográficas para la explotación bibliométrica es la falta de normalización de algunos de sus campos. En el caso de esta tesis, este problema afecta principalmente a tres niveles: geográfico, temático y autoría.

➤ Nivel geográfico e institucional

Este nivel estaría relacionado con toda aquella información que tiene que ver con las referencias geográficas e institucionales del documento, tales como el país, la ciudad, la institución, el centro, el departamento, etc., que se obtienen a través de la afiliación institucional de los autores. En esta tesis, este nivel ha sido de interés para las publicaciones relacionadas con el Domino 2, que está circunscrito al ámbito español.

Así, esta información de carácter geográfico e institucional resulta fundamental para poder analizar la distribución de la producción. En el caso del SCI-EXPANDED ésta aparece reflejada en el campo *Address*. Como ya se ha indicado, esta base de datos sólo

¹⁵ La publicación nº 6 se centra en el análisis comparativo de los diferentes tipos de descriptores que utilizan las bases datos SCI-EXPANDED y Medline. La recuperación de los documentos parte de una búsqueda inicial en el SCI-EXPANDED con la estrategia de búsqueda que se ha descrito. Pero cabe decir, que hay una segunda fase en la que se localizan uno a uno estos documentos en Medline, para poder hacer un análisis sobre el mismo conjunto documental. Debido a que Medline no indiza algunas revistas y tampoco los documentos tipo *Meeting Abstract*, el número de documentos que se analizaron en este estudio se redujo a 3.078.

normaliza de este campo el nombre del país, no así el resto de indicaciones geográficas e institucionales. Esto significa, que para analizar la distribución de la producción científica a un nivel de agregación más pormenorizado que el del país, es necesario realizar un proceso previo de normalización de las direcciones. En las publicaciones que componen esta tesis, este proceso se ha llevado a cabo en dos ocasiones de manera manual y sólo referido a las direcciones españolas.

Publicaciones donde se han analizado	Período de estudio	Direcciones Españolas Sin Normalizar	Direcciones Españolas Normalizadas
Publicaciones 2 y 3	1997-2007	5.121	4.935 ¹⁶
Publicación 7	1997-2012	13.159	13.612 ¹⁷

Tabla 4. Datos relativos al proceso de normalización de direcciones españolas

Para este proceso de normalización de las direcciones se codificaron cada una de ellas teniendo en cuenta hasta 5 niveles de descripción. A continuación, se detallan los criterios de codificación que se llevaron a cabo.

¹⁶ Durante esta fase del proceso de normalización se advirtió que existían documentos que no correspondían con el período de estudio o que estaban duplicados o que estaban asignados equivocadamente a España. Por lo tanto, se eliminaron un total de 98 documentos. Así que, el universo de documentos sobre el que se hizo este estudio durante esta etapa pasó de 2.467 a 2.369. La cifra resultante de la normalización de direcciones españolas es sobre los 2.369 documentos.

¹⁷ La cifra aumenta tras la normalización debido a la adscripción múltiple de una dirección a varias instituciones diferentes. El criterio que se ha seguido para estos casos es replicar la dirección tantas veces como instituciones intervienen.

1 ^{er} Nivel	2 ^o Nivel		3er Nivel	4 ^o Nivel	5 ^o Nivel
Código Postal	Código de Sector		Sector Institucional	Centro	Nombre de la Institución
	H	H+Acrónimo	Sector Sanitario		
	1	1+Acrónimo	Universidades	1 Facultad 2 Escuela Técnica Sup. 3 Escuela Universitaria 4 Colegios Universitarios 5 Centros de Postgrado 6 Inst. Invest.	
	2	2	CSIC		
	3	3	Admón. Central		
	4	4	Admón. Autonómica		
	5	5	Admón. Municipal o Diputación		
	6	6	Empresas públicas		
	7	7	Fundaciones/ Asociaciones sin ánimo de lucro		
	8	8	Empresas Privadas		
	21	21	CSIC + Universidad		
	24	24	CSIC + Admón. Autonómica		
	124	124	Universidad + CSIC + Admón. Autonómica		
	H124	H124	Sector Sanitario.+ Universidad+ CSIC + Admón. Autonómica	CENTROS MIXTOS MULTIDEPENDIENTES	
	H14	H14	Sector Sanitario+ Universidad+ Admón. Autonómica		
	14	14	Universidad + Admón. Autonómica		
	148	148	Universidad + Admón. Autonómica + Empresas Privadas		
	34	34	Admón. Central + Admón. Autonómica		

Tabla 5. Criterios de codificación y normalización de las direcciones españolas

➤ Nivel temático

Este nivel está relacionado con aquella información referida al contenido temático del documento. Son varios los campos que aluden a este tipo de información en una base de datos bibliográfica. Por un lado, se podría aludir a las categorías temáticas y por otro a los descriptores o términos de indización.

En el caso de las categorías el SCI-EXPANDED emplea la clasificación que el *Journal Citation Reports (JCR)*¹⁸ utiliza para clasificar las revistas que son indexadas. Estas categorías son recogidas en el campo *Web of Science Categories*. Uno de los inconvenientes referido a su uso es que se trata de una clasificación dinámica y cambiante. El número de revistas recogidas por el JCR varía ligeramente a lo largo del tiempo, debido a cambios de títulos, fusiones o escisiones de las revistas, así como, a cambios en la cobertura que, incorpora o excluye revistas. Este fenómeno repercute de forma negativa tanto en los recuentos como en la asignación temática (Maltrás, 2003). En el caso de esta tesis, todos los análisis referidos a este campo se ha mantenido la información original procedente del documento, sin ningún tipo de tratamiento al respecto (Publicaciones 2, 3 y 7).

En relación con los descriptores o términos de indización, ofrecen información sobre los contenidos conceptuales que son tratados en los documentos. Tanto el SCI-EXPANDED como Medline los utilizan para indizar sus documentos, aunque obedecen a naturalezas documentales diferentes.

En el caso de Medline, son el resultado de una indización manual basada en el uso de un vocabulario controlado como son los descriptores MeSH. Cabe señalar, que estos descriptores poseen una estructura muy concreta. Por un lado, se encuentran los términos *Major*, que aluden a los conceptos relacionados con la temática principal del documento y por otro lado, se encuentran los términos *Minor*, que identifican conceptos que también son discutidos en el documento, pero no aluden a la temática principal¹⁹.

En lo referido a esta tesis, en concreto en las Publicaciones nº 1 y 6, se han utilizado únicamente los descriptores *Major*. En la publicación nº 1 de los 365.714 documentos recuperados en Medline se descargaron únicamente los términos *Major*, obteniéndose un total de 1.300.881 descriptores. En cambio, en la publicación nº 6 la descarga de los

¹⁸ Journal Citation Reports (2017). Disponible en: <https://jcr.incites.thomsonreuters.com> [Fecha de consulta: 20/02/2017]

¹⁹ National Library of Medicine (2016). Disponible en: <https://www.nlm.nih.gov/bsd/mms/medlineelements.html> [Fecha de consulta: 01/12/2016].

datos se hizo sin tener en cuenta esta discriminación en los *MeSH* por lo que el proceso de selección de los términos *Major* se realizó posteriormente. De este modo, de los 3.078 documentos recuperados en Medline en esta fase, se obtuvieron un total de 43.053 descriptores que, tras el proceso de selección de los términos *Major*, se redujeron a 11.940 descriptores totales.

Publicación	Dominio de estudio	Período de estudio	Nº de Documentos	Descriptores MeSH	Descriptores Major MeSH Totales	Descriptores Major MeSH Únicos
1	Dominio 1	1965-2005	365.714	-	1.300.881	13.865
6	Dominio 2	1997-2010	3.078	43.053	11.940	2.868

Tabla 6. Tratamiento de descriptores MeSH procedentes de Medline

En el caso del SCI-EXPANDED, se emplean dos tipologías diferentes de palabras clave desde el punto de vista documental. Como ya se ha indicado en el apartado 6.1.1. Fuentes de información, por un lado, se utiliza una indización manual como son los *AKW* y por otro, los términos *KW+*, procedentes de una indización automatizada. Una característica que afecta a ambas tipologías es la falta de normalización de los términos. Este hecho repercute en el análisis de contenido de los documentos y en una elevada dispersión terminológica a causa de los problemas semánticos propios del uso de un vocabulario libre como son la homonimia, polisemia y sinonimia. No obstante, a pesar de esta posible limitación, el criterio que se ha seguido en los análisis realizados sobre estos términos ha sido el de respetar dicha ausencia de normalización. Aun así, lo que se ha realizado es un leve tratamiento de estandarización de los términos buscando la uniformidad gramatical (plurales y singulares) y ortográfica (abreviaturas, guiones, mayúsculas y minúsculas).

Publicación	Período de estudio	Nº de Documentos	Términos Totales		Términos Únicos Sin Normalizar		Términos Únicos normalizados	
4 y 5	1997-2010	4.148	KW+	28.733	KW+	9.833	KW+	9.465
6	1997-2010	3.078	KW+	26.956	KW+	9.107	KW+	8.764
			AKW	10.031	AKW	5.548	AKW	5.212
7	1997-2012	5.838	KW+	41.600	KW+	12.960	KW+	12.417
			AKW	15.480	AKW	8.050	AKW	7.452

Tabla 7. Tratamiento de los términos procedentes del SCI-EXPANDED

➤ Nivel de autoría

Se trata de la información relacionada con el nombre de los autores que firman los documentos. Esta información es recogida en las bases de datos en el campo *Author*.

Recientemente, la Web of Science ha comenzado a proporcionar herramientas que permiten buscar autores de una forma más precisa, sin embargo, estas herramientas todavía no han sido evaluadas formalmente para conocer su utilidad en los análisis a nivel individual. Los principales problemas referidos a la falta de normalización de este campo se deben fundamentalmente:

- A las variaciones del nombre que un autor utiliza en su producción científica para firmar sus trabajos.
- A una falta de rigor y homogeneidad en los procesos de tratamiento de la información por las bases de datos.

Esta falta de normalización afecta a los recuentos e indicadores bibliométricos relacionados con la autoría. Para solucionar este problema, en esta tesis se ha recurrido a una normalización de carácter manual ayudada fundamentalmente, por la existencia en algunos registros del nombre completo del autor y por la frecuencia de colaboración del autor con otros autores. Únicamente, se ha trabajado este campo para el Dominio 2.

Publicación	Período de Estudio	Nº de documentos	Autores Sin Normalizar	Autores Normalizados
2 y 3	1997-2007	2.369	7.809	7.397
7	1997-2012	5.838	23.676	19.632

Tabla 8. Datos relativos a la normalización de los Autores

6.2 Métodos

La propuesta metodológica que se presenta en esta tesis se basa en la combinación de diferentes técnicas y herramientas bibliométricas que se desarrollarán en los siguientes apartados. Esta estructura está inspirada en el modelo que propusieron Börner, Chen y Boyack (2003) para la visualización de dominios científicos.

6.2.1 Unidades de análisis: términos de indización

Las unidades de análisis son los elementos informativos procedentes de la colección documental que se utilizan como objeto de estudio principal para el análisis bibliométrico. Como ya se ha dicho anteriormente, en el campo de la bibliometría hay una gran variedad de elementos bibliográficos que pueden ser objeto de análisis desde los propios documentos hasta las revistas, las categorías, los autores, las citas, las palabras, etc. Además, en el caso de la delimitación temática existe una gran diversidad de técnicas y métodos que permiten la identificación de áreas y subáreas. En función del grado de profundidad que se quiera alcanzar en el análisis del dominio y acorde con los objetivos planteados, se utilizarán unas unidades de análisis u otras. Como se ha

indicado en el apartado 4. Antecedentes, existen estudios bibliométricos²⁰ que se inclinan por el Análisis de Co-citación de Autores (ACA), otros por el emparejamiento bibliográfico, otros por el Análisis de Co-palabras o co-words o estudios híbridos que combinan el ACA y Análisis de Co-palabras. En definitiva, no existe un consenso claro en la literatura científica que determine cuáles son las unidades de análisis más adecuadas en la delimitación temática de un dominio.

A pesar de las múltiples posibilidades, en este caso se ha decidido utilizar las palabras como unidades principales de análisis para la delimitación temática. Uno de los motivos es la ventaja que supone su uso en relación, por ejemplo, con las citas. Según Michel Zitt y colaboradores (2011), las palabras, parecen adaptarse mejor a la "ciencia viva" porque son universales. Además, las citas podrían considerarse que aluden a temas menos actuales, ya que conforman el trasfondo del documento, mientras que las palabras son coetáneas al propio documento (Soos et al. 2013). Para estos autores, las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes. Asimismo, la procedencia directa de las palabras de cada registro bibliográfico sitúa al análisis bibliométrico al nivel documental, es decir, el análisis desciende a un nivel de agregación mucho más concreto y de granularidad más fina que el que supondría el uso de otras unidades de análisis mayores como pueden ser las categorías temáticas.

Dentro de las posibilidades que ofrece el uso de las palabras para el análisis bibliométrico en esta tesis se ha decidido utilizar concretamente, los términos de indización que emplean las bases de datos para la búsqueda y recuperación documental. Una de las principales razones reside en la ventaja que supone la obtención de estos términos, ya que no requiere de minería textual ni de ningún pre-procesamiento lingüístico. Estos términos son incluidos en cada registro bibliográfico y son el resultado

²⁰ Se recomienda acudir a los apartados 4.5. *Delimitación temática* y 4.6. *Análisis de co-palabras* para acceder a los estudios más representativos que han empleado cada una de las técnicas que se aluden a continuación.

de un tratamiento documental como es el análisis de contenido, en el que se seleccionan los términos que reflejan la esencia conceptual de los documentos.

Como se ha indicado en el apartado 6.1.1. Fuentes de Información, se han utilizado los términos de indización que utilizan las dos bases de datos empleadas en esta tesis. Por un lado, los descriptores *MeSH* utilizados por Medline y por otro, las palabras clave *AKW* y *KW+* empleadas por el SCI-EXPANDED.

- **Descriptores *MeSH*.** Se trata de un vocabulario controlado especializado en el campo de la biomedicina desarrollado por la *National Library of Medicine*. Este tesoro es el resultado de una indización de carácter manual y está compuesto por 25.000 descriptores estructurados jerárquicamente en 16 categorías. Desde el punto de vista documental, este sistema de indización se caracteriza por un riguroso control semántico de los términos. Esto favorece al análisis de contenido donde los descriptores representan de manera unívoca los conceptos fundamentales tratados en los documentos. De este modo, se reduce la ambigüedad semántica eliminando los problemas de homonimia, sinonimia y polisemia propios del lenguaje natural y se mejora la calidad de la indización. Al mismo tiempo, esto repercute en la búsqueda y recuperación de la información obteniéndose mayor precisión y exhaustividad en los resultados. No obstante, este vocabulario también presenta algunas desventajas. Por un lado, la indización manual requiere de un coste mayor en cuanto a recursos humanos y de mayor plazo temporal en el proceso de indización. Y, por otro lado, y a consecuencia de esto último, este vocabulario controlado tiene menor capacidad para la incorporación de nueva terminología. Existe un margen de tiempo para que un término se considere que forma parte del léxico científico y adaptarlo a los criterios conceptuales y documentales que rigen dicho vocabulario controlado. Como consecuencia, se produce una rigidez terminológica que puede afectar a la detección y visualización de áreas y subáreas en la delimitación temática.

- **Author Keywords (AKW).** Se trata de términos procedentes del lenguaje natural que son seleccionados por los propios autores del documento. Como en el caso de los MeSH, corresponden a una indización de carácter manual, pero en este caso no es controlada sino libre. Esto supone una ventaja en el análisis de contenido ya que los autores son los mayores expertos en su campo de conocimiento y pueden ser más precisos en cuanto a la elección de los términos que mejor expresen el contenido de sus documentos. A pesar de ello, no existe un criterio de uniformidad semántica de los términos y esto favorece la aparición de los problemas de homonimia, sinonimia y polisemia propios del lenguaje natural, repercutiendo en la calidad de la indización y disminuyendo la precisión y exhaustividad en la búsqueda y recuperación de la información. Del mismo modo, esto supone un aumento de la dispersión terminológica en relación con los descriptores MeSH. Aunque, por otro lado, este tipo de indización requiere de un coste menor y su procesamiento en la base de datos es inmediato. No obstante, cabe decir que el SCI-EXPANDED incorporó estos términos como campo en el registro bibliográfico de manera paulatina a partir del año 1991²¹, por lo que, la presencia de estos términos en los documentos no es tan alta como en el caso de los otros dos términos de indización (véase publicaciones 4, 5 y 6).

- **KeyWords Plus (KW+).** Son términos procedentes de una indización automatizada que, a partir del empleo de un algoritmo de búsqueda, se extraen las palabras clave de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Gierfield & Sher, 1993). Este sistema de indización es exclusivo del SCI-EXPANDED y fue ideado por Garfield en los años 90. Al tratarse de términos procedentes de los títulos de los artículos que aparecen en las referencias, son los propios autores, en definitiva, los que seleccionan y comunican cuáles son los términos esenciales para expresar el

²¹ Web of Science (2016). Disponible en: https://images.webofknowledge.com/WOK46/help/WOS/h_fullrec.html#keywords_fr [Fecha de consulta: 16/12/2016]

contenido más importante de sus trabajos de investigación. Esta característica favorece la aparición de una terminología más actualizada y específica en comparación con otros tipos de términos de indización como pueden ser los procedentes de un vocabulario controlado (Braam et al., 1991). Esta cualidad puede contribuir a reflejar mejor los aspectos temáticos más dinámicos de un dominio. Por el contrario, estos términos tienen la desventaja de presentar una amplia dispersión terminológica debida a la falta de normalización y de control semántico de los términos.

Después de todo lo expuesto, a continuación, se incluye una tabla en la que trata de resumir brevemente, los criterios metodológicos que se han seguido en las publicaciones realizadas para realizar el análisis de cada dominio científico, teniendo en cuenta los elementos que los caracterizan:

<i>Dominio</i>	<i>Publicación</i>	<i>Cobertura temática</i>	<i>Cobertura temporal</i>	<i>Cobertura geográfica</i>	<i>Fuente de información</i>	<i>Términos de indización</i>	<i>Universo documental</i>
<i>Dominio 1</i>	Publicación 1	Investigación sobre Salud y Mujer	1965-2005	Internacional	Medline	Desc. MeSH	365.714
<i>Dominio 2</i>	Publicación 2 y 3	Investigación sobre células madre	1997-2007	España	SCI-EXPANDED	KW+ AKW	2.369
	Publicación 4 y 5	Investigación sobre células madre	1997-2010	España	SCI-EXPANDED	KW+	4.148
	Publicación 5 y 6	Investigación sobre células madre	1997-2010	España	SCI-EXPANDED Medline	KW+ AKW Desc. MeSH	3.078
	Publicación 7	Investigación sobre células madre	1997-2012	España	SCI-EXPANDED	KW+ AKW	5.838

Tabla 9. Resumen de los parámetros metodológicos utilizados en las publicaciones

6.2.2 Unidades de medida: análisis de co-palabras

El objeto de las unidades de medida consiste en cuantificar las relaciones de cada uno de los elementos que componen las unidades de análisis seleccionadas, en este caso los términos de indización, con el resto de elementos. Existe una gran variedad de medidas que pueden ser utilizadas en la visualización de dominios. Las más comunes son citación, emparejamiento bibliográfico, co-citación, co-ocurrencias, entre otras. Dado que el objetivo fundamental de la tesis consiste en la localización e identificación de líneas de investigación a través de los términos de indización, se ha optado por la co-ocurrencia como unidad de medida. Es decir, se han calculado las relaciones que se producen entre los términos de indización a partir de su aparición conjunta en los documentos. El resultado es una matriz cuadrada de $N \times N$ elementos, donde N es el término a representar, a partir de las veces que ocurre en los documentos. Este tipo de análisis basado en la co-ocurrencia de elementos utilizando, concretamente, las palabras como unidades de medida con el objetivo de detectar la estructura relacional de un dominio se conoce con el nombre de *Análisis de Co-Palabras* o *Co-Words Analysis*. En este caso, no se utilizan palabras procedentes del lenguaje natural extraídas del discurso científico de los documentos como suele ser lo más habitual, sino que, se utilizan términos procedentes de un lenguaje de indización postcoordinado.

6.2.2.1 Normalización de los datos de co-ocurrencia

Una vez obtenidos los datos de co-ocurrencia en relación con el mapeo y la visualización de los dominios de estudio en esta tesis, cabe decir que, en el campo de la bibliometría existe un amplio debate en la comunidad científica sobre la normalización de los datos de co-ocurrencia. No existe un consenso claro sobre en qué casos es necesaria la normalización de los datos y cuáles son las medidas más adecuadas. Trabajos como los de Jones & Furnas (1987), Hamers et al. (1989), Luukkonen et al. (1993), Peters & Van Raan (1993), Rorvig (1999), Chung & Lee (2001), Gmür (2003), Boyack et al. (2005), Klavans & Boyack (2006), Egghe & Rousseau (2006), Schneider & Borlund (2007a, 2007b), Leydesdorff & Vaughan (2006), Leydesdorff (2008), Egghe (2009), Van Eck & Waltman (2009), etc., muestran resultados empíricos sobre la

comparación de diferentes medidas. Se trata, generalmente, de medidas de similitud que cuantifican el grado de semejanza entre dos elementos de cara a su representación dimensional a modo de mapa. Para ello, se relacionan el número de co-ocurrencias con el número de veces que ocurren los elementos. Las más utilizadas son la Fuerza de Asociación, el Índice de Inclusión, el Índice de Jaccard, coeficiente del coseno de Salton, entre otras.

En este caso, siguiendo las recomendaciones de Leydesdorff & Vaughan (2006), no se han normalizado los datos de co-ocurrencia procedentes de la matriz simétrica obtenida de $N \times N$ elementos. Leydesdorff & Vaughan (2006) indican que no es necesario la normalización de los datos de co-ocurrencia cuando se utilizan matrices simétricas. Por lo que, consideran que el uso de los datos en bruto o *raw data* es más adecuado de cara a representaciones dimensionales. Del mismo modo, White (2003) considera que las representaciones obtenidas a través de datos de co-ocurrencia en bruto y tras el uso de un algoritmo de poda como puede ser *Pathfinder Networks* ofrecen estructuras más claras que las versiones normalizadas. Teniendo en cuenta esto, en este estudio no se ha aplicado ninguna medida de similitud sobre los datos de co-ocurrencia. Además, esta decisión está condicionada por la elección de las técnicas y herramientas de visualización que a continuación se detallan y que definen la metodología utilizada para el posterior análisis y representación de los datos.

6.2.3 Reducción del espacio n-dimensional y visualización de la información

Tras la obtención de la matriz de co-ocurrencia es necesario representar estos datos de tal manera que puedan transmitir la misma información, pero cuya interpretación sea visualmente más comprensible y clara. El objetivo es representar la interacción de las unidades de análisis a través de los documentos de manera que sea posible reflejar su estructura relacional (red) y que permita la identificación de agrupaciones bien definidas (agrupamiento o *clustering*). Para ello, uno de los métodos más habituales en

el ámbito de la bibliometría es la representación gráfica conocida como técnica de mapeo. Esto conlleva una transformación del espacio n-dimensional en uno bidimensional o tridimensional.

Según Van Eck y Waltman (2010), fundamentalmente, se utilizan dos tipos de visualizaciones basadas en diferentes técnicas de mapeo y *clustering* en la representación de redes. Por un lado, se encuentran los mapas basados en la distancia (*distance-based maps*). Éstos se caracterizan porque la distancia entre dos elementos refleja la fuerza de la relación entre ellos, una distancia pequeña indica, generalmente, una fuerte relación. Esta técnica tiene la ventaja de que facilita más claramente la identificación de grupos o *clusters* pero tiene el inconveniente de que dificulta la identificación de los nombres de los elementos debido a un fuerte solapamiento entre ellos. Dentro de esta técnica destacan el *Escalamiento Multidimensional* (*Multidimensional Scaling – MDS*), *VxOrd* (Davidson et al., 2001) y *VOS* (Van Eck et al., 2006; Van Eck & Waltman 2007a, b). Por otro lado, se encuentran los mapas basados en grafos (*graph-based maps*). Éstos, se caracterizan porque la fuerza de relación entre los elementos no viene dada por la distancia entre los elementos sino por los enlaces que los unen. Los elementos suelen estar distribuidos en el espacio de una manera más uniforme. Esta técnica tiene la ventaja de que el solapamiento de los nombres es mucho menor, pero en cambio, tiene la desventaja de que es más difícil detectar la fuerza de relación entre los elementos por lo que la identificación de los clústeres es más difusa. Generalmente, para desarrollar este tipo de mapas basados en grafos, se utilizan los algoritmos de tipo *springs embedders* como *Kamada Kawai* (1989) y *Fruchterman and Reingold* (1991) que están implementados en el software *Pajek* (Batagelj & Mrvar, 2010). También, los programas de ordenador *CiteSpace* (Chen, 2006) y *Network Workbench Tool* (NWB Team, 2006) entre otros, utilizan esta técnica de mapeo.

En esta tesis, se han utilizado ambas técnicas de cara a la visualización y representación de los datos para la detección de líneas o áreas de investigación. Para la aplicación de cada una de ellas, se ha recurrido al uso de dos de los softwares más

utilizados en la visualización de redes sociales y que constituyen un buen ejemplo de ambas técnicas como son *Pajek*²² en la generación de mapas basados en grafos y *VOSviewer*²³ en la generación de mapas basados en la distancia. Dada las ventajas y limitaciones que conllevan el uso de una técnica frente a la otra se han utilizado ambas con la intención de que puedan ser utilizadas de manera complementaria. Estas propuestas comportan grandes diferencias metodológicas que se describen a continuación.

6.2.3.1 Mapas basados en grafos (*Pajek*): Términos más frecuentes, *PFNETs* y *Kamada Kawai*

El uso de esta técnica de mapeo tiene como propósito representar el comportamiento de las unidades de análisis de manera relacional. El objetivo fundamental consiste en obtener una representación en forma de red que consta de dos elementos esenciales, por un lado, las propias unidades de análisis que son objeto de estudio y que reciben el nombre de *nodos*, *actores*, *vértices*, etc., y, por otro lado, la relación en sí que vincula a estos elementos entre sí y que recibe el nombre de *enlaces*, *aristas*, *líneas*, *lados* o *conexiones*. De este modo, es posible, representar y analizar la estructura relacional que existe entre las unidades de análisis a través de los documentos. Mediante este tipo de técnica de mapeo se obtiene una imagen estructural de la interacción de las unidades de análisis que es cuantificable y evaluable mediante el denominado análisis de redes dentro del ámbito de la teoría de grafos.

Generalmente, en redes donde existen un gran número de elementos que se relacionan frecuentemente entre sí y que su interacción es elevada, se obtiene una representación poco inteligible formando un amasijo o maraña de enlaces que dificulta la visualización de la estructura relacional principal. Este es el caso de esta tesis, ya que

²² Pajek (2016). Disponible en: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> [Fecha de consulta: 5/02/2016].

²³ VOSviewer (2016). Disponible en: <http://www.vosviewer.com/> [Fecha de consulta: 5/02/2016].

todas las redes que se han obtenido se han realizado a partir de un número muy elevado de *nodos* y *enlaces*. Por eso, ha sido necesario desarrollar una serie de técnicas y herramientas que han permitido seleccionar las relaciones principales y descartar las relaciones menos significativas o menos importantes. Es decir, ha sido necesario un proceso de simplificación de la red que permita representar la estructura y esencia de la red original. De este modo, en esta tesis se han utilizado dos criterios para la simplificación de los elementos esenciales de la red. Por un lado, se han seleccionado únicamente las unidades de análisis más frecuentes, es decir, los términos de indización más frecuentes. No existe un número de elementos que *a priori* pueda constituir un umbral óptimo para las representaciones, pero si se ha demostrado que las redes compuestas en torno a 100 nodos ofrecen resultados más satisfactorios (Van Eck & Waltman, 2010). En todas las publicaciones que integran esta tesis se han generado y analizado redes con los términos de indización que oscilan desde los 60 a los 150 nodos, en función de la estructura relacional de cada conjunto de análisis y según los objetivos planteados en cada fase de estudio. Por otro lado, se ha utilizado como criterio de simplificación de la red un algoritmo denominado de poda como es *Pathfinder Networks* (PFNETs) (Schvaneveldt et al., 1988 & Schvaneveldt, 1990), que permite detectar cuáles son los enlaces más relevantes de una red. Su objetivo fundamental es la extracción de la estructura principal de una red por medio del análisis de proximidad entre sus variables. Se basa, principalmente, en la aplicación de dos parámetros: r , asociado a la distancia de *Minkowski*, que se utiliza para calcular la distancia entre puntos a través de varios enlaces y se define mediante una ecuación paramétrica que subsume a la distancia euclidiana para $r = 2$. Esta distancia admite que se haga tender r hasta el infinito, lo que equivaldría a hallar el máximo de las distancias intermedias. Y el segundo parámetro es q , que está basado en el principio de desigualdad del triángulo. Este principio está relacionado con la longitud, en número de enlaces, de los caminos que se comparan, es decir, elimina aquellos enlaces que tienen asociada una distancia menor que otro camino que una los mismos nodos pasando por otros intermedios. De este modo, prevalecen los enlaces con mayor distancia y, por consiguiente, los enlaces con mayor peso e importancia de la red. El mayor valor que podría alcanzar q sería $n-1$, donde n es el número de nodos. Por lo tanto, en esta tesis se ha utilizado *PFNETs* bajo los parámetros $r = \infty$ y $q = n - 1$, como así lo recomiendan otros autores (Vargas-

Quesada, 2005; Moya-Anegón, et al., 2007; Quirin et al., 2008 & Vargas-Quesada et al., 2010).

PFNETs viene utilizándose en el campo de la Documentación desde 1990 aunque fue Chen el primero en aplicarlo a la citación (Chen, 1998a; Chen, 1998b & Chen, 1999). Desde entonces, son muchos los trabajos que han utilizado este algoritmo de poda para el estudio, visualización y detección de comunidades científicas a través del análisis de redes (Buzydlowski, 2003; White, 2003; Vargas-Quesada, 2005; Moya-Anegón, et al., 2007; Quirin et al., 2008; Espinosa-Calvo et al., 2009; Vargas-Quesada et al., 2010; Quirin et al., 2010; Yoo et al., 2013; White, 2015).

Una vez detectada la estructura principal de la red, otro aspecto a tener en cuenta en la metodología que se propone es la distribución espacial de los elementos que componen dicha red. Es decir, definir qué método se va a utilizar para la generación automática del grafo. Existen multitud de métodos, pero sin duda los más utilizados en el campo de la Bibliometría son los algoritmos de tipo *spring embedders* (insertadores de muelles). Se trata de programas cuyo objetivo principal es el de realizar grafos atractivos siguiendo una serie de principios estéticos como puede ser utilizar el máximo espacio disponible, forzar la posición de los nodos y reducir el número de enlaces cruzados, entre otros. En definitiva, estos algoritmos consisten en asignar coordenadas a los nodos de tal modo que, el grafo final sea estéticamente agradable al ojo humano (Vargas-Quesada, 2005). De todos ellos, el que se ha empleado en este trabajo y el que resulta más utilizado en el campo de la Bibliometría es *Kamada Kawai* (1989). Este algoritmo parte del fundamento de que la posición de los vértices no está restringida y que los enlaces son dibujados como líneas rectas, su propósito final es determinar únicamente la posición de los nodos o vértices. La idea es que se considera que la distancia deseable entre dos nodos es la misma que la representada en el gráfico basándose en el algoritmo de Floyd-Warshall (Floyd, 1963). Se introduce un sistema dinámico virtual formado por anillos (nodos) y por muelles (enlaces) haciéndolo evolucionar hasta que disminuye la energía acumulada por los muelles.

Algorítmicamente, se calcula la evolución de cada nodo por separado, fijando todos los nodos menos aquel que más energía acumula, dejándolo evolucionar hasta que la energía acumulada sea menor que un determinado límite y se vuelve a fijar. Este proceso se repite, continuamente hasta que, ninguno de los nodos acumule una energía superior al límite.

La combinación del algoritmo de poda *PFNETs* y el de visualización *Kamada Kawai* ha demostrado ser una propuesta interesante que ofrece buenos resultados en el análisis de dominios y que ha sido ampliamente utilizada en el campo de la Bibliometría (Chen, 1999; White, 2003; Moya-Anegón et al., 2007; Quirin et al., 2008; Vargas-Quesada et al., 2008; Vargas-Quesada et al., 2010).

Finalmente, se ha desarrollado una última fase dentro de esta técnica de análisis basada en el mapeo del dominio mediante grafos. Se trata de la validación gráfica. Una vez obtenidas las representaciones finales, se ha recurrido a dos técnicas para la interpretación de la información obtenida en los mapas:

- **Análisis Factorial.** Es una técnica estadística de reducción de datos que se utiliza para explicar las correlaciones entre las variables observadas en términos de un número menor de variables no observadas llamadas factores. Es una técnica exploratoria multivariante cuya aplicación práctica principal consiste en reducir el número de variables, detectar la estructura por medio de sus relaciones, así como clasificarlas. En definitiva, se ha aplicado esta técnica con el fin de agrupar los términos de indización e identificar cada una de las líneas de investigación representadas. Esta técnica ha sido empleada en el Dominio 1 y está recogido en la Publicación nº 1.
- **Consulta de expertos.** Se ha recurrido a la opinión de expertos especializados para la interpretación de las líneas de investigación representadas en los

mapas a través de los términos de indización. En este caso, se ha contado con la colaboración de especialistas relacionados con la investigación española en células madre, ya que, este método se ha aplicado, exclusivamente, al Dominio 2, presente en las publicaciones nº 2, 3, 4, 5 y 6.

6.2.3.2 *Mapas basados en la distancia (VOSviewer): todos los términos, técnica de mapeo VOS y técnicas de clustering.*

Esta técnica de mapeo basada en la distancia comparte el mismo propósito que en el caso de la técnica anterior, representar el comportamiento de las unidades de análisis de manera relacional. La diferencia fundamental reside en que las relaciones y el valor de las mismas, no viene representado por enlaces que conectan los ítems entre sí, sino, por la distancia existente entre ellos, cuanto menor sea la distancia entre los ítems más fuerte es la relación. A diferencia de la técnica basada en grafos, en esta técnica resulta fundamental la disposición de los elementos en la representación para interpretar la relación existente entre ellos.

En el campo de la bibliometría existen muchas técnicas de mapeo basadas en la distancia para la elaboración de mapas bibliométricos. Una de las más utilizadas por la comunidad científica ha sido el *Escalamiento Multidimensional (Multidimensional Scaling – MDS)*. Esta técnica comprende una amplia familia de procedimientos multivariados, algunos de los cuales hacen posible representar las proximidades entre los elementos como distancias en un espacio de un número reducido de dimensiones. La idea central es que las distancias que median entre los puntos se corresponden con las proximidades entre los objetos por medio de una función de ajuste resultante de un proceso iterativo de optimización, pudiéndose describir las relaciones entre los objetos sobre la base de las proximidades observadas. De este modo, se establece una analogía entre el concepto “psicológico” de proximidad (semejanza o desemejanza) y el concepto geométrico de distancia (Arce, 1996), quedando reflejadas las diferencias entre ambos en una función denominada stress (Kruskal, 1964a y b).

Tomando esta técnica como referencia, Ludo Waltman y Nees Jan van Eck desarrollaron otra técnica alternativa de mapeo basada en la distancia conocida como técnica de *visualización de similitudes* (*visualization of similarities – VOS*) (Van Eck & Waltman, 2007a). Ésta, se fundamenta en gran medida, en los principios metodológicos en los que se asienta el *MDS*, pero incorpora a su vez, una serie de parámetros y ajustes algorítmicos que ofrecen mejores resultados desde el punto de vista de la visualización de la información. Para profundizar más sobre los aspectos que asemejan o diferencian a estas dos técnicas de mapeo se recomienda la lectura de las siguientes publicaciones (Van Eck & Waltman, 2007a; Van Eck et al., 2008; Van Eck et al., 2010). No obstante, tanto *VOS* como *MDS* parten de un mismo objetivo esencial que consiste en situar a los ítems en un espacio de reducidas dimensiones, de tal manera que, la distancia entre dos ítems refleje, tanto como sea posible, la similitud o grado de relación entre ellos (Van Eck et al., 2010).

Para medir dicha similitud la técnica de mapeo *VOS* requiere de la matriz de datos de co-ocurrencias. Para Van Eck y Waltman el empleo de estos datos no es suficiente para calcular la proximidad entre los ítems por lo que, consideran indispensable la aplicación de una medida de similitud a los datos de co-ocurrencia. En concreto, estos autores proponen dentro de las que se consideran medidas de similitud directas la denominada fuerza de asociación (*association strength*) (Van Eck et al., 2006; Van Eck & Waltman, 2007b). También es conocida esta medida con el nombre de índice de proximidad (*proximity index*) (Rip & Courtail, 1984; Peters & Van Raan, 1993) o índice de afinidad probabilística (*probabilistic affinity index*) (Zitt et al., 2000). Se formula de la siguiente manera:

$$AS_{ij} = \frac{C_{ij}}{C_i C_j}$$

Esta fórmula se interpreta como la fuerza de asociación (AS_{ij}) de los ítems i y j es proporcional a la relación entre el número observado de co-ocurrencias de i y j por un

lado, y al número esperado de co-ocurrencias de i y j por otro, asumiendo que las co-ocurrencias de i y j son estadísticamente independientes (Van Eck & Waltman, 2009).

Calculada la fuerza de asociación como medida de similitud entre los ítems, la técnica de mapeo *VOS* tiene una forma predeterminada de ubicar los elementos en un mapa bidimensional. Ésta consiste en minimizar la suma ponderada de las distancias euclidianas al cuadrado entre todos los pares de ítems. Cuanta mayor sea la similitud entre dos elementos, mayor será el peso de la suma al cuadrado de su distancia. Con el fin de evitar mapas triviales en los que todos los elementos puedan tener la misma ubicación, *VOS* impone la restricción que la distancia media entre dos elementos debe ser igual a 1. Además de esta restricción, se aplica el denominado *algoritmo de mayorización (majorization algorithm)* que es una variante del algoritmo *SMACOF (Scaling by MAyorizing a COmplicated Function)* cuyo desarrollo inicial se debe, fundamentalmente, a De Leeuw (De Leeuw, 1977; De Leeuw & Heiser, 1977) y que ha sido ampliamente utilizado en la literatura del *MDS* (Van Eck & Waltman, 2010; Van Eck et al., 2010b). La característica fundamental de este algoritmo consiste en que, si se aplica de manera iterativa, se generan secuencias no crecientes de valores de la función que termina, generalmente, en un mínimo local. La idea general radica en reemplazar la función a minimizar $f(x)$ por una función auxiliar $g(x, z)$ donde z es un valor fijo. De este modo, no se trabaja con cada punto de manera independiente considerando fijos los demás, sino que se hace modificando las posiciones de todos los puntos a la vez. *VOS* tiende a localizar los objetos cerca de lo que considera su ideal de coordenadas.

Asimismo, esta técnica de mapeo está integrada en un software diseñado para la visualización y generación de mapas bibliométricos denominado *VOSviewer*. Sus creadores, Van Eck y Waltman, han implementado en una misma herramienta informática la técnica de mapeo *VOS* en combinación con técnicas bibliométricas de *clustering*. Éstas últimas se basan, fundamentalmente, en una variante ponderada y parametrizada de la conocida función de similitud de Newman and Girvan (2004).

VOSviewer está diseñado con una configuración en la que, por defecto, se aplican y combinan todas las técnicas y variantes algorítmicas descritas sin necesidad de que los usuarios tengan que intervenir metodológicamente, en la construcción o visualización de los mapas bibliométricos. Consta de una serie de parámetros que obedecen a una metodología previamente definida y basada en los criterios que se acaban de describir. Esto supone una ventaja para los usuarios que en una sola herramienta tienen agrupadas un conjunto de técnicas bibliométricas que simplifican y agilizan la creación o visualización de mapas bibliométricos. No obstante, ofrece la posibilidad de modificar algunos parámetros por otros que prefieran los usuarios. En el caso de los análisis orientados a la delimitación temática, *VOSviewer* ofrece la opción de extraer directamente los términos más relevantes de cada documento y mapearlos. En este caso, se ha descartado esta opción porque el objetivo del estudio se basa en el uso exclusivo de los términos de indización como unidades de medida para la identificación temática. Del mismo modo, *VOSviewer* ofrece la posibilidad de generar mapas bibliométricos a partir de datos de co-ocurrencia en bruto, es decir, sin emplear la medida de similitud basada en la fuerza de asociación que aplica por defecto. El software permite a los usuarios desactivar la normalización de los datos de co-ocurrencia que, por defecto, aplica a los datos de entrada. En este caso, cabe decir que, se han utilizado los datos en bruto de la matriz de co-ocurrencias como datos de entrada, pero al mantener los parámetros iniciales de *VOSviewer*, el programa, finalmente, ha normalizado los datos mediante la medida de similitud basada en la fuerza de asociación entre los ítems para la construcción de los mapas bibliométricos.

También cabe destacar que, a diferencia de *Pajek*, este software permite trabajar con redes más amplias, donde intervienen mayor número de actores o nodos. Dependiendo de la distribución de los datos de co-ocurrencia es posible representar mayor o menor número de elementos. En la publicación 3, por ejemplo, se han representado todos los elementos. En principio, la matriz de co-ocurrencias que se facilita al programa contempla a todos los elementos, pero es posible modificar su número estableciendo algún umbral mínimo relacionado con los valores del índice de proximidad. Todo ello, en función del criterio de visualización que se quiera adoptar en relación con la claridad

de las representaciones de cara a su interpretación y la información que aporte del dominio. Este ha sido el caso de los términos de indización representados en las publicaciones número 4, 5 y 6, donde se han utilizado diferentes umbrales acordes con los criterios de visualización.

Asimismo, una vez generados los mapas bibliométricos, *VOSviewer* ofrece diferentes opciones para visualizarlos. Estas opciones varían en función de las versiones que se utilicen del software. Actualmente, está disponible la versión 1.6.5²⁴ en la que sólo se ofrecen dos opciones:

1.- Visualización de la Red (*Network Visualization*): en esta opción los ítems se muestran etiquetados por su nombre y, por defecto, representados por un círculo (también existe la opción de utilizar rectángulos). Para cada ítem, tanto el tamaño de la fuente del nombre con el que es etiquetado como el tamaño del círculo dependen del peso de cada ítem. Si los ítems han sido asignados a clústeres, el círculo de cada ítem se muestra en el color del clúster al que pertenece. Opcionalmente, los colores de los clústeres también pueden ser asignados de manera manual. De forma predeterminada, para evitar la superposición de los nombres etiquetados, sólo se muestra un subconjunto de todos ellos. Además, por defecto, no se muestran las líneas entre los ítems. Sin embargo, esto también es posible cambiarlo a través de las opciones que proporciona *VOSviewer*. Esta visualización es particularmente útil para examinar detalladamente el mapa.

2.- Visualización de Densidad (*Density Visualization*): esta opción se caracteriza porque cada elemento en el mapa es representado con un color que va del rojo al azul, reflejando la densidad de las relaciones entre los ítems. Cuanto mayor es la densidad, es decir, la co-ocurrencia entre los términos, más se aproximarán a la tonalidad roja, en

²⁴ La versión 1.6.5 de *VOSviewer* está disponible desde el 28 de Septiembre del 2016. Para más información véase <http://www.vosviewer.com/> [Fecha de consulta: 20/02/2017].

cambio sí se aproximan más al color azul, esto indica una mayor dispersión y, por lo tanto, menos co-ocurrencia. Esta visualización resulta especialmente útil porque permite detectar los principales núcleos de interacción de un dominio.

Además de estas opciones, en versiones anteriores de *VOSviewer*²⁵ estaban disponibles otras dos opciones más:

3.- Vista de Densidad de Agrupaciones (*Cluster Density View*): esta opción sólo está disponible si los ítems han sido asignados a clústeres. Esta visualización es similar a la de Densidad, pero en este caso, la densidad de los ítems se muestra para cada clúster por separado. Así, el color de un punto en un mapa es similar al color de un determinado clúster, si hay un gran número de elementos que pertenecen a ese clúster alrededor de dicho punto. Al igual que en la vista de Densidad, los ítems con pesos altos se destacan más que los ítems con pesos bajos. Esta vista es especialmente útil para obtener una visualización rápida y general de cada clúster por separado y, por tanto, de las áreas importantes de un mapa. Para una descripción detallada de la ejecución técnica de esta opción de visualización véase Van Eck & Waltman, 2009.

4.- Vista de Dispersión (*Scatter View*): en esta opción los ítems son representados mediante un círculo pequeño. Si el ítem ha sido asignado a un clúster, el círculo de cada elemento aparecerá del color del clúster al que pertenece. Opcionalmente, los colores de los clústeres también pueden ser asignados de manera manual. En esta visualización no se muestran los nombres etiquetados de cada ítem. Por defecto, tampoco aparecen las líneas entre los ítems, pero esto se puede cambiar de manera opcional. Esta visualización puede ser especialmente útil para obtener una visión general de la estructura de un mapa.

²⁵ En versiones anteriores a la actual (v. 1.6.5) la opción *Network Visualization* era denominada como *Label View* y la opción *Density Visualization* como *Density View*.

En esta tesis, fundamentalmente, se ha trabajado con versiones antiguas de *VOSviewer*, donde constaban las cuatro opciones de visualización. Se han generado todos los mapas bibliométricos utilizando todas las opciones de visualización que el software ofrece. No obstante, en las publicaciones, por motivos de extensión se han incluido sólo algunas de ellas de manera conjunta.

VERSIÓN	PUBLICACIÓN
V. 1.3.0	Publicación nº 3
V. 1.5.3	Publicación nº 4 y 5
V. 1.5.4	Publicación nº 6

Tabla 10. Características de los dominios analizados

7. RESULTADOS Y DISCUSIÓN

En esta sección se presenta un compendio general de los resultados más significativos de cada una de las publicaciones, así como, una breve discusión de los mismos de acuerdo con los objetivos planteados en cada fase de estudio. Todo ello, se aborda desde una perspectiva global ya que, todos estos aspectos están desarrollados, pormenorizadamente, en cada una de las publicaciones. El objetivo de este apartado consiste, fundamentalmente, en ofrecer una visión de conjunto que contemple la discusión de los resultados más destacados en relación con el marco general de la tesis.

Para ello, se ha decidido mostrar y discutir aquellos resultados que dan respuesta a las principales preguntas de investigación planteadas en cada una de las publicaciones y que corresponden, en esencia, a los objetivos principales planteados en la tesis. Por lo tanto, a continuación, se muestran una relación de preguntas estructuradas en función del dominio de estudio que se aborda. Cabe recordar que, como se ha indicado en otras secciones, cada dominio está condicionado por el uso de diferentes fuentes de información, unidades de análisis y periodos temporales.

7.1. Localización y visualización de las principales líneas de investigación de la producción científica correspondiente al Dominio 1. (Publicación 1)

7.1.1. Detección de las principales líneas de investigación sobre la producción científica sobre salud y mujer (1965-2005) a través de los de los descriptores *MeSH*: estudio evolutivo (Publicación 1).

- 1. La propuesta metodológica que se aplica en este estudio ha permitido obtener unos resultados determinados. Concretamente, ¿en qué consisten estos resultados? ¿Éstos permiten obtener una imagen donde se reflejen las principales líneas de investigación que caracterizan el dominio elegido?**

El resultado de aplicar la propuesta metodológica al dominio seleccionado se resume, básicamente, en un conjunto de tablas y gráficos que son mostrados en la Publicación 1. Fundamentalmente, los resultados se muestran en función de los análisis aplicados y de cada una de las divisiones temporales en las que se segmenta el período de estudio (1965-1974; 1975-1984; 1985-1994 y 1995-2005).

La combinación del Análisis Factorial (Análisis de Componentes Principales con rotación VARIMAX) y el de redes aplicado al estudio de co-términos, generaron unos resultados que fueron sintetizados, esencialmente, en cuatro mapas. Cada uno de ellos está constituido por los 60 descriptores *MeSH* más frecuentes de cada período relacionados entre sí, en función de su aparición conjunta en los documentos. Al mismo tiempo, cada descriptor está representado mediante un símbolo que indica el factor al que pertenece. Para más detalle, en el Anexo I de la Publicación 1 se muestran todos los descriptores *MeSH* correspondientes a cada uno de los factores detectados en cada período. De este modo, se visualiza la estructura relacional de los descriptores más predominantes de cada período y, además, debido a la factorización, es posible agruparlos por temas, lo que permite la identificación y localización de los principales aspectos tratados en los documentos. Además, gracias a las visualizaciones generadas a partir del análisis de redes es posible detectar y examinar cómo interactúan estos aspectos en cada período analizado.

Asimismo, los resultados obtenidos del Análisis Factorial (AF) han permitido localizar en cada uno de los períodos los factores predominantes en función del porcentaje de la varianza que ocupan. En definitiva, se han identificado los aspectos que reflejan líneas de investigación y subcampos de disciplinas científicas. Éstos resultados se muestran por cada período en la Publicación 1 en forma de tabla.

La aplicación de la metodología propuesta por sucesivos períodos temporales, que son el resultado de la subdivisión de un período de estudio amplio, permite comparar los resultados entre sí. Esto posibilita detectar si hay una continuidad o no tanto en los factores como en los descriptores que los integran. Por eso, en cada período se ha incluido una tabla donde se indican los descriptores que aparecen o desaparecen en

cada mapa. De este modo, se aprecia una evolución temática de la investigación lo que evidencia el dinamismo de la ciencia y la progresión de las tendencias del dominio.

2. Existen varios aspectos por lo que la metodología diseñada en este estudio es relevante. Entre ellos, destaca el uso de diferentes técnicas bibliométricas para el análisis de los datos, ¿la combinación del Análisis Factorial y el de redes sociales aplicado al estudio de co-términos ha permitido obtener resultados satisfactorios?

La combinación del AF y el de redes sociales aplicado al estudio de co-términos ha permitido, la obtención de resultados concordantes entre sí y al mismo tiempo complementarios, en la mayoría de los casos. El AF al tener en cuenta el grado de correlación entre los descriptores ha favorecido la detección de las principales líneas de investigación. En cambio, el estudio de redes sociales aplicado al estudio de co-términos ayudado con el sistema de poda de *Pathfinder Networks* (PFNETs), ha representado las relaciones entre los descriptores más frecuentes que caracterizan a cada una de esas líneas de investigación permitiendo ver la interacción entre los factores. En cada uno de los mapas, se ha podido apreciar cómo, generalmente, ambos análisis coinciden. El hecho de que algunos descriptores aparezcan conectados con otros, indica que esos descriptores están relacionados en investigaciones (documentos) comunes. Esto no implica que pertenezcan a la misma línea de investigación (factores). Algunos de los descriptores incluidos en los mapas pueden ser de nuevo cuño o de reciente incorporación a la investigación y por su “juventud”, no correlacionar lo suficiente con el resto como para ser “factorizados”. También ha podido ocurrir lo contrario, que aparezcan descriptores que están cayendo en desuso en la investigación, y que por tanto, están dejando de correlacionar con el resto hasta el punto de no ser “factorizados”. Como el objetivo del estudio es realizar el análisis de los descriptores más frecuentes, se han tenido en cuenta tanto los “factorizados” como los no “factorizados” mostrando sus relaciones. Por eso, se identifican como zonas “calientes” a aquellos descriptores que pertenecen a más de un factor, indicando así la interacción entre distintas líneas de investigación, y como zonas “frías” a aquellos descriptores que,

aun siendo frecuentes en la investigación, no interactúan con el resto de los mostrados en el mapa.

- 3. Otro aspecto a destacar es el uso de los descriptores *MeSH* como unidades de análisis, ¿la visualización de las relaciones que se producen entre estos descriptores a partir de su aparición conjunta en los documentos ha permitido representar y definir temáticamente las principales líneas de investigación del dominio? ¿El uso de un vocabulario controlado supone alguna ventaja para este tipo de estudios? ¿Qué inconvenientes presenta? ¿Es posible realizar un análisis estructural de carácter temático llegando a un nivel de agregación tan concreto como es el documental?**

Los resultados reflejan que el empleo de descriptores *MeSH*, han contribuido a crear una imagen clara y limpia de la investigación. El hecho de que se trate de un vocabulario controlado, sintetiza y simplifica los conceptos o ideas de los documentos de una manera más sistemática y coherente que si se tratara de un vocabulario libre o del lenguaje natural. Para estudios de granularidad gruesa como este, donde se persigue localizar y representar los temas fundamentales y esenciales que mejor resumen la investigación de un dominio científico y donde se utilizan volúmenes documentales muy amplios, resulta muy adecuado el uso de este tipo de sistemas de indización. No obstante, hay que tener en cuenta que a veces, el uso de términos controlados en estudios de granularidad gruesa como este, impiden estudiar en profundidad la variedad del campo. Por lo que, la selección de información esencial a través del AF puede estar sesgada por esta opción metodológica.

En general, el uso de las palabras y en definitiva de los análisis de *co-words* permiten utilizar unidades de análisis muy pequeñas que identifican, a la mínima expresión, ideas o conceptos. En el caso de los descriptores, estas palabras, además, tienen la obligación de corresponder con las ideas o conceptos que mejor resumen la información conceptual de un documento. De este modo, es posible definir temas o líneas de investigación en función de la información obtenida directamente de cada documento,

favoreciendo una caracterización temática más precisa que si se utilizan otras unidades de análisis mayores como pueden ser las categorías de revistas.

4. Desde el punto de vista del análisis que supone el período temporal este estudio se considera de carácter evolutivo-longitudinal. ¿La selección del período temporal abarcando 40 años es suficiente para la identificación temática a partir de un vocabulario controlado? ¿La subdivisión de este período en 4 permite detectar cambios en los temas que reflejen la evolución y dinamismo de la investigación?

La decisión de la temporalización en períodos de 10 años en este estudio ha venido motivada, fundamentalmente, por la producción documental. Si se observa la tabla 1 de la Publicación 1, se observa que el último período es, especialmente, influyente puesto que su producción supone más de la mitad de la producción total (193.659 documentos de 365.714 documentos). Por lo que, si se quisieran representar las líneas de investigación de todo el período en su conjunto se apreciaría un claro sesgo por el peso de los estudios del último período, que tendrían una mayor influencia y serán más visibles que los concernientes a las primeras décadas. Además, el análisis del período completo ofrecería una imagen muy estática y superficial de los temas de investigación que han caracterizado los estudios sobre Salud y Mujer. No sería posible reflejar una tendencia evolutiva en la que se apreciara los cambios de orientación en la investigación. Por eso, los resultados obtenidos por décadas si revelan dicha tendencia y muestran una serie de líneas de investigación predominantes en cada etapa.

5. Partiendo de un tema de carácter multidisciplinar como es Salud y Mujer, ¿es posible localizar y representar las tendencias científicas que se han desarrollado en este campo?

En este estudio se ha podido demostrar que, a través de la coocurrencia de los 60 descriptores más frecuentes de un período, se puede resumir en una imagen la

investigación más importante. Los mapas realizados para cada década, han servido para mostrar cómo han evolucionado los principales frentes de investigación referentes a este ámbito desde los años 60 hasta nuestros días. No obstante, tratándose de un tema multidisciplinar y de una base de datos especializada en medicina se ha apreciado, por parte de esta última, un esfuerzo por ampliar su cobertura temática a lo largo de los años. A medida que el concepto de salud se ha ido ampliando, yendo más allá de los aspectos estrictamente biológicos, los resultados reflejan como la presencia del contexto social y del entorno cada vez son más importantes.

El estudio longitudinal a través de mapas de diez años muestra las líneas de investigación predominantes en cada periodo, así como, su evolución en el tiempo. Mediante su análisis se puede ver que, en el primer período (1965-1974) los aspectos relativos al embarazo y al parto, así como, sus posibles consecuencias mentales y psicológicas, constituyen el eje central sobre el que se desarrolla la investigación. Incluso la presencia de estudios de temática legal se justifica por las consecuencias médicas, legales, éticas y sociales que durante estos años planteaba la despenalización del aborto en la mayoría de los países desarrollados. Al parecer, sólo las patologías que se generan a consecuencia de los procesos biológicos que son exclusivos de las mujeres, son las únicas en las que la mujer es objeto de estudio para la investigación.

No obstante, al contemplar todos los períodos en su conjunto, se puede observar que esta característica va cambiando a lo largo de los años. En la década de 1975-1984 se puede apreciar como el peso y preponderancia de estos estudios disminuye considerablemente. El factor Embarazo experimenta un protagonismo menor a pesar de que sigue siendo el elemento que mayor vinculación y proximidad tiene con el resto de los factores. Parto y Enfermedades Mentales ya ni siquiera son representados. Los Estudios Médico-Legales que siguen aludiendo, fundamentalmente, a los problemas legales, éticos y jurídicos que subyacen bajo la regulación del aborto inducido, se ven claramente mermados. En cambio, parece despertar un interés nuevo por aspectos de carácter más social que tienen que ver más con el contexto y entorno en el que gana

presencia la mujer. Aparecen factores como Estudios de Género y conceptos como la actitud psicológica del paciente, la educación sanitaria, la actitud del personal sanitario, etc. que denotan un cambio de orientación en la investigación.

Este cambio, aún es más claro en la siguiente década. Entre 1985 y 1994 los estudios relacionados con Entorno Socio-Familiar. Psicología comienzan a ser el núcleo a partir del cual se vinculan el resto de factores, desbancando de esta posición, definitivamente, a los estudios relacionados con la reproducción femenina. Tanto es así, que la línea de investigación relativa al Embarazo desaparece. Se observa que, los aspectos sociales impregnan y son un importante trasfondo para todos los estudios de este período, en el que, además, se detecta el mayor número de factores de todas las décadas. Probablemente, esto haya sido motivado en gran medida, por dos importantes fenómenos que tienen lugar en este momento. Por un lado, irrumpe un importantísimo fenómeno biomédico cuya trascendencia es notabilísima no sólo en el campo de la medicina sino de la sociedad en general, como es la aparición del SIDA. El estudio de esta enfermedad junto con el descubrimiento del VIH en 1986, tuvo una importante repercusión para la producción científica. Hasta el punto de que, como se aprecia en los resultados, el factor que recoge estos estudios se convierte en una de las líneas de investigación más importantes del período. Además, propicia e impulsa el interés hacia otros estudios de temática más social, como son los que pueden tener una relación causal con la enfermedad (Drogadicción) o los que pueden suponer una respuesta preventiva hacia la misma (Educación Sanitaria). De este modo, los estudios sobre SIDA e Infecciones por HIV contribuyen a fortalecer las relaciones entre los aspectos sociales que condicionan la salud de las mujeres presentes en este período.

Por otro lado, la transformación sociológica que tiene lugar en este momento con la incorporación de la mujer al mundo laboral, también añade valor a la investigación sobre la importancia del contexto y el entorno social. La presencia de la mujer en un escenario nuevo más allá del ámbito familiar y doméstico introduce un campo desconocido para la investigación que lo hace muy atractivo. Las consecuencias legales que se derivan de

este fenómeno, también introducen un cambio de perspectiva para los Estudios Médico-Legales, que ya no sólo se limitan a la regulación del aborto.

Aunque indiscutiblemente los estudios sobre Neoplasias Ginecológicas suponen una constante para la investigación y una línea de trabajo consolidada. En este período la presencia de la mujer en los estudios médicos contempla nuevos determinantes para la salud. No sólo es considerada para los aspectos que establecen una diferencia respecto a los hombres a nivel sexual (sexo), sino que también lo es a nivel social, en función del rol que desempeña (género). Por lo que ya podemos advertir una serie de cambios en la presencia de una perspectiva de género en temas de salud.

Finalmente, en la última década, se puede apreciar cómo estos aspectos sociales obtienen aún mayor protagonismo. Tanto el número de descriptores como la variedad temática de los mismos enriquecen las líneas de investigación que tienen que ver con estos aspectos. El fuerte retroceso que sufren los estudios sobre el SIDA e Infecciones por HIV con respecto al periodo anterior, confirma, por un lado, que, el auge de esta línea fue debido a un impulso coyuntural y por otro que el interés de la investigación se desplaza hacia otros aspectos que no son los estrictamente biológicos.

En esencia, los resultados obtenidos en esta fase de estudio muestran las principales líneas de investigación relativas a Salud y Mujer a partir de las publicaciones recogidas en la base de datos Medline durante un período de 40 años (1965-2005).

- 6. La propuesta metodológica ¿se podría aplicar a otros dominios científicos?, en concreto, ¿a uno más específico? ¿se podrían utilizar otras fuentes documentales para el análisis? ¿Se podrían tener en cuenta como unidades de análisis otra tipología de términos basados en diferentes técnicas de indización?**

Estas cuestiones se plantean después de realizar esta fase de trabajo formando parte de la tesis. En concreto, serían los objetivos a alcanzar en la siguiente fase. Probar la metodología en dominios de carácter más específico desde el punto de vista temático, cronológico y geográfico es interesante para valorar su capacidad de adaptación a otros contextos científicos. De este modo, se comprobaría si esta metodología se puede aplicar a cualquier dominio científico o si se producen sesgos o problemas condicionados por las características del dominio.

También, observar si el uso de otras fuentes documentales condiciona en alguna medida los resultados que se obtienen al aplicar la propuesta metodológica. La cobertura, criterios de indexación y el alcance de la fuente documental pueden interferir en la imagen de la investigación científica que se pretende analizar. La naturaleza de la colección documental puede ser muy influyente en la propuesta metodológica.

Del mismo modo, el sistema de indización que utiliza la fuente documental juega un papel muy importante en la metodología diseñada. La naturaleza y la tipología del sistema de indización es clave para esta propuesta puesto que son las unidades de análisis. Valorar de qué manera la metodología propuesta se adapta o no a diferentes sistemas de indización puede ser muy interesante para comprobar su robustez y adaptabilidad.

7.2. Localización y visualización de las principales líneas de investigación de la producción científica correspondiente al Dominio 2.

7.2.1. Detección de las principales líneas de investigación sobre la producción científica española en células madre (1997-2007) a través de los *KeyWords Plus*. (Publicaciones 2 y 3).

- 1. En comparación con el dominio anterior, en esta fase de trabajo se aplica la propuesta metodológica a un dominio de carácter más específico desde el punto de vista temático, cronológico y geográfico. En términos bibliométricos ¿qué indicadores permiten contextualizar mejor este dominio? ¿qué imagen de la investigación ofrecen?**

En esta fase de estudio se ha analizado un conjunto documental referido a un dominio más específico que en la fase anterior, la investigación española sobre células madre durante el período 1997-2007. Se recuperaron 2.369 documentos a partir de una estrategia de búsqueda general. Se consideró apropiado realizar un breve análisis bibliométrico compuesto por indicadores de carácter general para obtener una imagen que ayudase a contextualizar mejor el estudio y a valorar el estado en el que se encuentra la investigación científica desde el punto de vista temático. De este modo, además de probar la metodología en un dominio más concreto se obtiene una referencia de lo que puede aportar la propuesta metodológica relacionada con la delimitación temática dentro de los estudios bibliométricos tradicionales.

Cabe decir que, los resultados de esta fase de estudio fueron presentados, de manera preliminar, en un congreso internacional (Publicación 2) y, posteriormente, fueron revisados, mejorados y ampliados dando lugar a un artículo científico (Publicación 3). Para explicarlos de una manera más clara y precisa todas las alusiones o referencias que se hacen a continuación provienen de la Publicación 3.

Los resultados del análisis bibliométrico mostraron que la tendencia en la producción de los 2.369 documentos fue ascendente a lo largo del período 1997-2007 (Fig. 1, p. 568). No obstante, también se detectaron pequeñas fluctuaciones en la producción que pudieron obedecer a diversos factores. Por un lado, a la posible interferencia de celebraciones de congresos que elevaría el número de publicaciones en un momento determinado. En este sentido, el ligero incremento que se aprecia en la evolución temporal de la producción en *Meeting Abstract* (Fig. 6, p. 574) explicaría el pico registrado en los últimos años de estudio. Por otro, la influencia de los documentos hechos en colaboración. En el año 2004 se aprecia un aumento destacado de los documentos sin colaboración entre instituciones nacionales y/o internacionales (Fig. 6, p. 572), hecho que puede estar relacionado con el acusado incremento de autores que también se detecta en ese mismo año (Fig. 3, p. 571). Del mismo modo, se puede tener en consideración si las reformas legislativas han podido repercutir e impulsar la investigación hacia nuevos campos, pudiendo diversificar el perfil temático del dominio científico.

Por otro lado, los datos del análisis bibliométrico indicaron que el sector institucional más predominante fue el sanitario (Tab. 3, p. 570). Precisamente, los centros más productivos correspondieron a hospitales, en concreto, al de Santa Cruz y San Pablo, al Clínico de Barcelona y al Clínico de Salamanca. Además, las afiliaciones de los autores más productivos correspondieron a estas instituciones. J. Sierra fue el más destacado que, junto a R. Martino pertenecen al Hospital Santa Cruz y San Pablo. J.F. San Miguel, junto con M. D. Caballero, proceden del Hospital Clínico de Salamanca. Del mismo modo, la producción a nivel geográfico se concentró, fundamentalmente, en Cataluña y Madrid. Comunidades con una producción más modesta como Galicia y el Principado de Asturias presentaron una tasa de variación notablemente elevada (Fig. 2, p. 569). En cuanto a los indicadores de coautoría y productividad reflejaron una relación coherente y una tendencia estable durante el período de estudio, siguiéndose una evolución lógica hacia el aumento de documentos coautorados en detrimento de los realizados por un único autor (fig.3, p. 571). En relación a las colaboraciones a nivel institucional predominan las de carácter bilateral frente a las trilaterales o cuatrilaterales, igualmente ocurren con la colaboración internacional. Los principales países con los que se ha

colaborado son Estados Unidos, Alemania, Inglaterra, Francia e Italia, seguidos de productores más modestos pero importantes en el campo de las células madre (Tab. 5, p. 573).

Los datos relacionados con la distribución temática se centraron fundamentalmente, en el estudio de las categorías y de los *KW+*. En el caso de las primeras las más productivas del período fueron: Hematología, Oncología y Biofísica (Tab. 7, p. 576). Cabe decir que, gran parte de esa producción se debe al volumen de publicación de las revistas *Bone Marrow Trasplantation* (Hematología, Oncología y Biofísica) y *Blood* (Hematología) que, concentran el mayor número de documentos monopolizando así la investigación (Tab. 6, p. 575). Además, estas categorías reflejaron un desarrollo y comportamiento similares a los de la producción total, a excepción de los últimos años donde el crecimiento no es tan acusado (Fig. 7, p. 576). Dejando al margen a estas categorías se observa que la productividad del resto aumenta notablemente, sobre todo, a partir del año 2004 (Fig. 8. p. 577).

En el caso de los *KW+*, cabe señalar su elevada presencia en los documentos ya que el 78% de las publicaciones analizadas (1.846 documentos) los contienen (Tab. 8, p. 578). El *KW+* con mayor presencia en los documentos fue *Bone-Marrow-Transplantation*, con casi el doble de apariciones que el siguiente *KW+* que fue *Stem Cell* (Tab. 10, p. 579). A pesar de que este término fue el empleado en la estrategia de búsqueda, al haber sido buscado en el campo *topic* no necesariamente tiene que ser el término más frecuente en los documentos, ya que dicho campo no sólo recupera los documentos que tengan ese término en los *KW+* sino que también busca en los campos *AKW*, título y resumen.

En líneas generales, estos resultados procedentes del análisis bibliométrico ofrecen una imagen general del dominio y desde el punto de vista temático aportan una información descriptiva y esquemática, insuficiente para identificar de manera consistente los focos, los temas o las tendencias de investigación.

2. En esta fase de trabajo se ha utilizado el Análisis de Redes Sociales aplicado al estudio de *co-palabras* teniendo en cuenta dos unidades de análisis diferentes: las categorías y los términos *KW+*. ¿Qué información aporta el estudio de categorías? ¿se podría aplicar a un estudio de granularidad fina como este? ¿El análisis de los *KW+* ofrece más información que el de las categorías? ¿Proporcionan información complementaria? ¿Qué unidades de análisis se podrían considerar más adecuadas para la delimitación temática a nivel de documento?

La propuesta metodológica planteada en esta tesis se aplica en esta fase de estudio bajo la premisa de poder identificar las principales áreas o líneas de investigación en un dominio especializado que, como ya se ha podido comprobar en la anterior pregunta, no sería posible identificar en un estudio bibliométrico de corte tradicional. En concreto, se aplica utilizando como unidades de análisis tanto los términos de indización que emplea el *SCI*, en este caso los *KW+*, como las categorías temáticas que el *Journal Citation Reports (JCR)* utiliza para clasificar temáticamente a las revistas. De este modo, los resultados obtenidos pueden compararse entre sí y valorar qué unidades permiten una identificación temática más clara, completa y precisa.

Los resultados obtenidos a partir del análisis de las categorías más frecuentes se muestran a modo de mapa. Éstos permitieron distinguir claramente dos tipos diferenciados de investigación: la clínica y la básica (Fig. 9, p. 577). Las categorías relacionadas con la clínica se localizaron en la parte central e inferior del mapa y se corresponden con Hematología, Oncología y Biofísica, siendo esta agrupación la más destacada del mapa por el peso de sus relaciones. A su vez, estas categorías se relacionan con otras agrupaciones que también forman parte de la investigación clínica como son Medicina, Investigación y Experimentación; Enfermedad Vascular Periférica y Trasplante. La investigación básica, en cambio, se encuentra ubicada en la zona superior del mapa agrupando un conjunto de categorías relacionadas con Biología Celular que, a su vez, da lugar a otras subagrupaciones.

Los resultados obtenidos a partir del análisis de los *KW+* se muestran en dos representaciones diferentes. Una obtenida a través del software *VOSviewer* (Fig. 10, p. 580) que, fundamentalmente, obedece a un mapa de calor donde la interacción de los *KW+* a través de los documentos, destaca los principales focos o temas predominantes en la investigación. En este mapa, nuevamente, se puede distinguir la investigación clínica de la básica. En concreto, en la zona derecha del mapa se pudo localizar la investigación de carácter más clínico compuesta por los documentos relacionados con el uso terapéutico del trasplante de progenitores hematopoyéticos, como indica el peso de *Therapy* en este conjunto de *KW+*. Por el contrario, en la posición central se concentran los documentos que abordan la investigación de carácter más básico, reflejada por el peso del término *In-vitro*, percibiéndose, además, la investigación basada en experimentación animal. Por último, a la izquierda se perfilan los trabajos con células madre relacionados con las neurociencias.

La otra representación que ha generado el análisis de los *KW+* se ha obtenido a través del software *Pajek* en combinación con *PFNETs* (Fig. 11, p. 582). En este caso, la imagen obtenida ofrece una información más clara sobre la estructura relacional del dominio destacando agrupaciones de términos relacionados entre sí a partir de una vinculación temática. La representatividad de los *KW+* más frecuentes que aparecen en el mapa es bastante elevada, ya que, están presentes en un 68,80% de los documentos. Este dato confirma la validez e idoneidad de este método para localizar y representar áreas de investigación, pese a que el grado de dispersión de los *KW+* puede ser mayor que el de otros sistemas de indización.

Las relaciones de estos *KW+* han permitido la localización e identificación de hasta cuatro agrupaciones bien diferenciadas que definen las principales líneas de investigación (L.I.) del dominio. Por un lado, la relacionada con la investigación clínica representada por los términos agrupados bajo *Bone-Marrow-Transplantation* y que, están relacionados con los estudios ligados al uso terapéutico de las células madre en enfermedades hematológicas (L.I.1). Por otra parte, la investigación básica representada por las tres agrupaciones restantes. En primer lugar, es posible localizar los *KW+* que están relacionados con los estudios encaminados a la investigación de células madre

hematopoyéticas (L.I.2), es decir, las células madre procedentes de la médula ósea (*Stem Cells*). En segundo lugar, los *KW+* relacionados con el estudio de las células madre embrionarias (L.I.3). Concretamente, con los términos relacionados con los procesos de expresión (*Expression*) y diferenciación celular (*Differentiation*), así como, los procesos y las biotecnologías necesarias para localizar, producir, crecer y, sobre todo, analizar in-vitro las células madre embrionarias. Finalmente, la última agrupación identificada y que conecta con la línea de investigación de las células madre embrionarias, está relacionada con la investigación con células madre neurales (L.I.4) (*Central-Nervous-System*). Estos términos están muy relacionados con los estudios centrados en el desarrollo e identificación de las células madre neurales que se diferencian activamente en el tejido cerebral.

Pese a que las visualizaciones obtenidas de los *KW+* se han realizado con metodologías distintas (véase el apartado 6. Materiales y Métodos), ambas coinciden en identificar las dos grandes naturalezas en las que se divide la investigación como son la clínica y la básica. Además de esto, las visualizaciones aportan más información sobre la temática del dominio localizando las principales áreas o líneas de investigación sobre células madre. Ambas representaciones coinciden en detectar el mismo número y los mismos aspectos temáticos.

Al entrar a valorar comparativamente los resultados obtenidos del estudio de categorías con los de los *KW+* hay que tener presente que son unidades de análisis que obedecen a diferentes niveles de agregación temática. Las categorías sólo aportan información relacionada con la clasificación temática de las revistas donde son publicados los documentos, en cambio, los *KW+* aportan información procedente de cada documento. Por tanto, el análisis de los *KW+* permite detectar las relaciones temáticas interdocumentales, por medio de unidades de medida más pequeñas que tratan de sintetizar la esencia conceptual de cada documento.

A pesar de ello, en líneas generales, los resultados obtenidos del análisis de categorías coinciden con el de los *KW+*, indicando que el mayor peso de la investigación con células madre reside en los estudios hematológicos y oncológicos. No obstante, el análisis de

categorías ofrece unos resultados de carácter general y superficial ya que no permite definir líneas de investigación, más allá del conjunto de disciplinas que están implicadas en el dominio científico. En cambio, el análisis de *KW+* ofrece unos resultados más exhaustivos y precisos mostrando, con un mayor nivel de desagregación, la distribución de términos que integran cada perfil temático de manera específica. Además, posibilita la identificación de tendencias en la investigación tanto de las que son más productivas como de las que no lo son tanto, bien por estar en declive o bien porque se están desarrollando de manera incipiente.

3. ¿Cuál es la estructura general del dominio? ¿De esa estructura se puede extraer alguna conclusión desde el punto de vista temático? ¿Es posible detectar líneas de investigación con la propuesta metodológica sobre un dominio científico tan específico? ¿El uso del *SCI-EEXPANDED* como fuente de información y de los *KW+* como unidades de análisis han permitido la detección de líneas de investigación?

Los resultados de esta fase de trabajo referidos a la delimitación temática de un dominio más específico a través de los términos de indización *KW+*, se han reducido, esencialmente, a la información ofrecida por las visualizaciones generadas en forma de mapa. En concreto, las figuras 10 y 11 (p. 580 y 582) permitieron detectar diferentes agrupaciones de términos que interactúan desde el punto de vista semántico y estructural, posibilitando la identificación de tendencias o áreas temáticas. Principalmente, se detectaron hasta un total de cuatro líneas de investigación, agrupadas, a su vez, por las dos grandes naturalezas que caracterizan a este dominio, la investigación de carácter clínica y básica. La primera de ellas definida, fundamentalmente, por la L.I.1 y la segunda por el resto de líneas (L.I.2; L.I.3 y L.I.4).

A partir de la obtención de estos resultados, es posible valorar que el empleo de un dominio más específico en comparación con el dominio de la fase de trabajo anterior, no supuso una limitación ni una desventaja a la hora de localizar e identificar las principales líneas de investigación a partir de la aplicación de la propuesta metodológica

que se presenta en esta tesis. Se podría decir que, hasta el momento, el dominio no ha condicionado a la propuesta metodológica.

Del mismo modo, el uso de una fuente documental de carácter multidisciplinar no especializada en medicina como es el SCI-EXPANDED, no ha supuesto una limitación para localizar estudios especializados en aspectos biomédicos, permitiendo incluso detectar diferentes tendencias temáticas dentro del dominio de estudio. En la fase anterior de trabajo se utilizaba una fuente documental especializada para un dominio más amplio y en esta fase se ha utilizado una fuente documental multidisciplinar para un dominio de estudio más específico. En ambos casos, la propuesta metodológica ha posibilitado la localización de las principales líneas de investigación, lo que avala la potencialidad y solidez de la misma.

Finalmente, los resultados han mostrado que el uso de los *KW+* como unidades de análisis ha permitido localizar e identificar los aspectos más representativos de la investigación. Dado que son términos extraídos automatizadamente del lenguaje natural, posibilita la existencia de una terminología más actualizada lo que favorece la detección e identificación de las tendencias en el discurso científico. Además, su incorporación a la base de datos como términos de indización por su procesamiento automatizado es más rápido en comparación con otros sistemas de indización como puede ser un vocabulario controlado. No obstante, el uso de estos términos también presenta una clara desventaja debido a su gran dispersión terminológica.

4. Actualizando el período de estudio a 1997-2010 ¿se podría realizar un estudio evolutivo con este dominio similar al realizado en el Dominio 1, utilizando los *KW+* como unidades de análisis?

La actualización de los datos es un aspecto que se ha tenido en consideración a lo largo de las fases de trabajo de esta tesis. En cada una de ellas, se ha creído apropiado trabajar con los períodos temporales más actualizados, siempre teniendo en cuenta el

proceso de indización de las bases de datos. De cara a la siguiente fase, se consideró apropiado actualizar el período de estudio en tres años (1997-2010).

Por otro lado, esta fase de trabajo se ha caracterizado, fundamentalmente, por utilizar un dominio de carácter más específico con la propuesta metodológica que se propone. Como se ha indicado en la anterior pregunta, los resultados muestran que la naturaleza más o menos específica del dominio no ha afectado en la identificación y localización de las principales líneas de investigación. No obstante, sería interesante comprobar si es posible detectar una evolución de estas líneas a lo largo del período. Es decir, comprobar si la metodología que se propone tiene la capacidad de reflejar la dinámica y evolución temática de un dominio muy reducido y concreto sujeto a grandes cambios y a una gran actividad científica.

7.2.2. Detección de las principales líneas de investigación sobre la producción española en células madre (1997-2010) a través de los *KeyWords Plus*: estudio evolutivo. (Publicaciones 4 y 5).

1. Teniendo en cuenta la especificidad temática del dominio y un período temporal de 14 años ¿sería posible detectar líneas de investigación a través de los *KW+* partir de una segmentación temporal?

En la anterior fase de trabajo, se aplicó la propuesta metodológica a un dominio de carácter más específico con respecto al Dominio 1, donde se obtuvieron unos resultados que posibilitaban la identificación y localización de hasta cuatro líneas de investigación. La caracterización temática que se obtuvo ofreció una imagen fija y estática del dominio. En este caso, el objetivo de esta fase de trabajo se centró, fundamentalmente, en obtener distintas visualizaciones de diferentes tramos temporales en los que se dividió el período de estudio. Si los resultados que se obtienen permiten la identificación de las principales líneas de investigación de cada subperíodo, al compararlos entre sí, tal y como ocurría en el análisis del Dominio 1, se podría observar cuál ha sido la evolución de estas líneas y como se han ido configurando a lo largo del período. El reto

metodológico en esta fase de trabajo residió en comprobar si las características del Dominio 2 podrían impedir la localización de líneas de investigación con una producción documental tan concreta.

Cabe decir que, al igual que ocurría en el apartado anterior, los resultados de esta fase de estudio fueron presentados, de manera preliminar, en un congreso internacional (Publicación 4) y, posteriormente, fueron revisados, mejorados y ampliados dando lugar a un artículo científico (Publicación 5). Para explicarlos de una manera más clara y precisa todas las alusiones o referencias que se hacen a continuación provienen de la Publicación 5.

2. En relación con la imagen que ofrece el período completo ¿sería posible apreciar diferencias o cambios en los temas de investigación en cada uno de los subperíodos? ¿Es posible apreciar la dinámica del dominio a partir de la identificación de líneas de investigación consolidadas y líneas de investigación emergentes?

El período completo de estudio para esta fase de trabajo se compone de 14 años (1997-2010). Se obtuvieron un total de 4.148 documentos. Este periodo se subdividió en tres tramos temporales (1997-2001; 2002-2006; 2007-2010). Si se observa la tendencia de la productividad documental (Tab. 2, p. 262), se aprecia un importante crecimiento anual, sobre todo, en los últimos años. La producción de cada período casi se duplica con respecto al anterior. Se observa que, más de la mitad de los documentos (2.174) corresponden al último período. Esto es importante tenerlo en cuenta de cara a las interpretaciones de los resultados relativos al período completo, donde la identificación de las tendencias puede verse afectada por el peso de los documentos pertenecientes al último período.

Las visualizaciones obtenidas del período completo tanto a través de *Pajek* como de *VOSviewer* permiten identificar las mismas líneas de investigación que en el caso de la fase de estudio anterior: el uso terapéutico de las células madre en enfermedades onco-

hematológicas (L.I.1); las células madre hematopoyéticas (L.I.2), las células madre embrionarias (L.I.3) y finalmente, las células madre neurales (L.I.4). Además, nuevamente se distingue estructuralmente dos naturalezas temáticas de la investigación claramente diferenciadas: la clínica y la básica.

En cuanto a los resultados obtenidos de cada subperíodo, se pudieron apreciar claras diferencias en la estructura y configuración de las líneas de investigación detectadas en relación al período completo. A pesar de ello, la distinción entre la investigación clínica y básica siempre ha sido clara en todos los tramos temporales, aunque varíe el peso de una respecto a la otra.

En líneas generales, se puede observar que la presencia y las relaciones de determinados *KW+* cambia en función del tramo temporal. Existen agrupaciones muy cohesionadas que se definen claramente. Sus relaciones tienen un peso importante en comparación con el resto lo que denota que, las publicaciones tienen una gran consistencia y están temáticamente muy interrelacionadas. Esto, permite identificarlas con líneas de investigación de carácter consolidado para ese período de estudio. En cambio, existen otras agrupaciones de *KW+* menos cohesionadas y poco definidas compuestas por términos que, aunque puedan estar presentes en otros subperíodos, no tienen la identidad estructural y relacional suficiente en ese momento. Esto conduce a pensar que obedecen a publicaciones con una definición temática más imprecisa, propias de líneas de investigación en formación, bien por tratarse de tendencias que pierden interés científico o bien, por tratarse de nuevas tendencias que empiezan a desarrollarse en este campo. De este modo, a partir de las visualizaciones que genera cada subperíodo es posible detectar una evolución de los principales temas.

Concretamente, en los resultados referidos al primer subperíodo (1997-2001), se observó la numerosa presencia de *KW+* relacionados con la L.I.1. Esta agrupación que está claramente relacionada con la investigación clínica, es la más definida y la más predominante del período. Esto refleja que estos estudios están muy consolidados. El resto de líneas que integraban la investigación de carácter más básico (L.I.2, L.I.3 y L.I.4) no parecen estar tan definidas. Fue posible detectar (fig. 4, 5 y 6, p. 264 y 265) que la

mayor parte de los *KW+* estuvieron presentes en el período completo, incluso los que ejercían mayor influencia dentro de cada línea de investigación. A pesar de ello, el comportamiento relacional entre ellos es distinto y no es posible distinguir agrupaciones de términos que tengan una relación temática clara. En líneas generales, la investigación básica en este período aparece poco definida y poco cohesionada. Posiblemente, estas débiles agrupaciones formen parte de las denominadas líneas de investigación emergentes, como se puede observar en los siguientes periodos.

En cuanto al segundo subperíodo (2002-2006) se detectó que, al igual que ocurría en el primer subperíodo la agrupación más numerosa y definida fue la L.I.1. La presencia de esta línea nuevamente, confirma la consolidación de estos estudios. Respecto a la investigación de carácter más básico, estructuralmente, se muestran mejor definidos que en el subperíodo anterior. Los estudios relacionados con la L.I.2 identificados mediante los descriptores conectados con *Stem-Cells*, ya no concentran la mayor parte de las relaciones, si no que han pasado a formar una agrupación con mayor independencia relacional y temática. Al definirse mejor esta línea también se han definido los términos relacionados con los estudios sobre las células madre embrionarias (L.I.3). Si en el primer subperíodo estaban muy entremezclados con los estudios de las células madre hematopoyéticas (L.I.2), en este subperíodo se definen mejor. Finalmente, los estudios relacionados con las células madre neurales (L.I.4), también están mejor definidos tanto temática como estructuralmente. Por lo tanto, es posible considerar la consolidación de esta línea de investigación durante este período.

En el tercer subperíodo (2007-2010) se detectan cambios. La distinción entre la investigación clínica y la básica sigue siendo clara, aunque esta última aparece mucho más desarrollada y con mayor peso en la investigación. La presencia de *KW+* de carácter clínico (L.I.1) es más proporcionada al resto de las líneas de investigación, tanto en número como en el peso de sus relaciones. Lo que llama la atención en estos años, es la distinción (Fig. 10) de los estudios relacionados con los procesos oncológicos, identificados por el descriptor *Bone-Marrow-Transplantation* y, por otro lado, los que están más relacionados con el trasplante de progenitores hematopoyéticos, identificados con el descriptor *Stem-Cell Transplantation*. Esta diferenciación estructural

puede deberse a la formación de una nueva tendencia en la investigación que con el tiempo pueda dar lugar a una nueva línea de investigación. Asimismo, se identifican cinco agrupaciones de carácter más básico. La primera, mucho mejor definida que en el período anterior, está constituida por los estudios relacionados con las células madre hematopoyéticas (L.I.2), lo que significa su consolidación para este período. Además, esta línea está conectada tanto con los estudios relacionados con el uso terapéutico de las células madre en enfermedades hemato-oncológicas (L.I.1), como con la investigación relacionada con las células madre embrionarias (L.I.3). Concretamente, en esta última línea, aparecen mejor definidos los estudios relacionados con los procesos de proliferación y diferenciación celular y aquellos relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y analizar in-vitro las células madre embrionarias. Los *KW+* que constituyen la última línea de investigación (L.I.4) han perdido definición en este período y ya no forman un grupo tan homogéneo. Incluso se ha desagregado en dos aspectos de la investigación distintos: el desarrollo e identificación de las células madre neurales y los estudios sobre el infarto de miocardio mediante el empleo de progenitores celulares. Es posible que esta pérdida de definición se deba a una nueva línea de investigación emergente que pueda consolidarse en los próximos años.

3. En relación con los softwares de visualización propuestos en esta fase de trabajo ¿han mostrado resultados compatibles? ¿son los que tienen la mayor capacidad para representar estudios temáticos de carácter dinámico o evolutivo?

En esta fase de estudio se han obtenido resultados complementarios a través de las imágenes ofrecidas por *Pajek* y *VOSviewer*. Por su parte, *Pajek* en combinación con *PFNETs*, ha permitido sintetizar y resaltar la estructura relacional, mientras que *VOSviewer* ha posibilitado la identificación de los grupos y focos de investigación de una manera más plástica. Además, ambos coinciden en la localización de las principales líneas de investigación, así como, en la identificación de los términos más influyentes de la red. No obstante, presentan algunas limitaciones ya que las visualizaciones obtenidas

en esta fase de trabajo presentan visualizaciones estáticas e independientes de cada periodo. Sería interesante de cara a futuros trabajos, explorar otras funcionalidades más recientes de estos softwares que están encaminadas a mejorar estos aspectos, como es la posibilidad de crear mapas superpuestos (*overlay maps*), que permiten fijar la posición de los nodos para hacer estudios evolutivos o comparativos (Rafols, Porter & Leydesdorff 2010; Leydesdorff, Rafols & Chen, 2013).

- 4. Entre los parámetros metodológicos que se han utilizados en esta fase de trabajo se encuentran el empleo de un dominio de carácter específico, el volumen de años, así como su temporalización, el uso de una fuente documental multidisciplinar y el empleo de los KW+ como unidades de análisis. ¿Alguno de estos parámetros ha supuesto algún inconveniente para la identificación de líneas de investigación a través de la propuesta metodológica?**

Como se ha visto en la fase de trabajo anterior la fuente documental, la especificidad del dominio o las unidades de análisis no supusieron ningún inconveniente para la identificación de los principales temas de investigación. En esta fase tampoco lo debían ser, al menos para el estudio del período completo, ya que, se han tenido en cuenta los mismos parámetros metodológicos, a excepción del período temporal, que se ha aumentado en 3 años lo que incrementó el número de publicaciones analizadas, pasando de 2.369 a 4.148. A pesar de este incremento, la fragmentación temporal para realizar el análisis por períodos hace que el volumen documental a analizar en cada tramo disminuya notoriamente además de no ser proporcional, debido al crecimiento exponencial de la Ciencia. Esto, sin duda, afecta a la frecuencia de aparición y a las relaciones que se establecen entre los KW+, pudiendo repercutir en la visualización e identificación de líneas de investigación. Más aún cuando se trata de un campo muy activo cuyas líneas están en continua evolución. La identificación temática se hace más exigente pudiendo comprometer la capacidad de la propuesta metodológica para adaptarse a dominios muy reducidos, específicos y activos.

Como se ha observado en las anteriores preguntas, los resultados han mostrado que es posible la identificación de líneas de investigación, tanto las consolidadas como las que están en formación. Gran parte de ello es posible gracias al uso de los términos *KW+*. La capacidad que tiene este sistema de indización para incorporar una terminología actualizada resulta una ventaja para estudios de granularidad tan fina como éste, especialmente, en la localización de tendencias de manera evolutiva porque favorece la identificación de estudios con menor representatividad en términos de producción y es más sensible a la identificación de líneas en formación.

No obstante, una vez observada la capacidad de los descriptores *MeSH* y los *KW+* para localizar e identificar líneas de investigación sería interesante comparar ambos sistemas de indización sobre un mismo conjunto documental y valorar comparativamente qué sistema lo representa y lo caracteriza mejor desde el punto de vista temático.

7.2.3. Detección de las principales líneas de investigación sobre la producción española en células madre (1997-2010) a través de los términos de indización del SCI y Medline: estudio comparativo. (Publicación 6).

- 1. A partir de la metodología propuesta, ¿sería posible comparar sobre un mismo conjunto documental las representaciones temáticas que se obtendrían a partir del uso de diferentes tipologías de términos desde el punto de vista de la indización? ¿El uso de diferentes unidades de análisis dentro de esta propuesta metodológica puede repercutir en los resultados de la delimitación temática?**

El objetivo principal de esta fase de trabajo ha consistido en analizar y comparar las visualizaciones que se obtienen del empleo de diferentes términos de indización como unidades de análisis en la propuesta metodológica que se plantea en esta tesis.

En las anteriores fases de trabajo se ha podido comprobar que el uso de diferentes unidades de análisis en función de su naturaleza documental está relacionado con la especificidad del dominio científico y con la granularidad del análisis que se pretende realizar. Estos factores han de ser tenidos en cuenta ya que pueden condicionar la precisión y profundidad de los resultados en la delimitación temática. En el estudio del Dominio 1 se utilizaron los descriptores *MeSH* como unidades de análisis y se observó que el uso de un vocabulario controlado especializado en medicina sobre un dominio amplio desde su cobertura temporal, geográfica y temática permitió detectar las principales tendencias del campo e incluso observar su evolución a través del tiempo. Del mismo modo, los estudios sobre el Dominio 2 también han conseguido la identificación de las principales líneas de investigación y observar su evolución a través del tiempo. En este caso, se utilizaron como unidades de análisis los términos *KW+* procedentes de indización libre automatizada y se empleó una base de datos multidisciplinar. Desde el punto de vista metodológico, el análisis de este dominio es más exigente que la del Dominio 1, ya que su cobertura temporal, geográfica y temática es más especializada, convirtiéndolo en un estudio de granularidad fina.

Por lo tanto, se ha considerado interesante valorar en esta fase de trabajo si la naturaleza documental en el uso de diferentes unidades de análisis afecta a los resultados de la propuesta metodológica en estudios de granularidad fina como este. Para ello, se utilizaron tres tipos de unidades de análisis que obedecen a tres tipos de indización distintas procedentes de las dos bases documentales empleadas en los estudios anteriores:

Término de Indización	Base de datos	Tipo de indización	
		Desde el punto de vista de la asignación de términos	Desde el punto de vista del vocabulario
KW+	SCI-EXPANDED	Automatizada	Libre
AKW	SCI-EXPANDED	Manual	Libre
MeSH	Medline	Manual	Controlado

Tabla 11. Términos de indización analizados

El universo documental que se ha empleado en esta fase de trabajo ha partido del conjunto documental utilizado en la fase anterior (4.148 documentos). Esto se debe a que se ha tenido en cuenta el dominio de estudio más exigente desde el punto de vista metodológico, como es el Dominio 2 y en el que ha sido posible la identificación de las principales líneas de investigación. Se ha considerado fundamental para esta fase el empleo del mismo conjunto documental para la comparación de las unidades de análisis. De este modo, se puede valorar con mayor precisión la capacidad que tiene cada unidad de análisis para representar temáticamente el mismo dominio compuesto por las mismas publicaciones. Para ello, los 4.148 documentos obtenidos del SCI-EXPANDED fueron buscados en Medline. No se pudieron localizar todos los documentos porque en algunos casos pertenecían a revistas que no están indexadas en Medline, generalmente, relacionadas con el ámbito de la Zoología y de la Botánica y en otros casos, porque correspondían a la tipología documental de *Meeting Abstract* que no recoge Medline. Por lo tanto, el conjunto documental quedó reducido a 3.078 documentos.

2. ¿Se aprecian diferencias en las visualizaciones obtenidas según el tipo de términos de indización utilizado? ¿A qué son debidas?

Los resultados más significativos de esta fase de trabajo quedan resumidos en la visualización de las relaciones que se producen entre los términos de indización a través de su aparición conjunta en los documentos. En total, se han obtenido nueve mapas que corresponden a dos representaciones obtenidas a través de *VOSviewer* y una obtenida a través de *Pajek* por cada tipología de términos de indización utilizada. Como se ha visto en las fases de trabajo anteriores, la combinación de ambos softwares ofrece resultados complementarios que enriquecen la información que se obtiene del dominio.

Antes de describir y discutir la información que ofrece cada mapa, cabe decir que, la primera diferencia que llama la atención está relacionada con la presencia de los términos de indización en los documentos (Tab. 1, p. 177). Si en el caso de los *MeSH* y *KW+* el volumen de documentos en los que están presentes es muy similar, aproximándose al 99% de los documentos, en el caso de los *AKW* es muy destacado que tan sólo están presentes en el 65,82% de los documentos. Esta escasa presencia en los documentos afecta a los niveles de frecuencia y a las relaciones que se producen entre los *AKW*, influyendo en las visualizaciones obtenidas a través de los mapas, dificultando la formación de agrupaciones temáticas que posibiliten la identificación y localización de líneas de investigación.

Términos de Indización (T.I.)	T.I. Totales	T.I. Únicos	Frecuencia Media de Aparición de los T.I.	Número de T.I. por documento	Número de relaciones
KW+	26.956	9.107	2,96	8,91	93.341 (el 10,35% se producen más de una vez)
AKW	10.031	5.548	1,81	4,95	20.883 (el 4,93% se producen más de una vez)
MeSH	11.940	2.884	4,14	3,91	16.134 (el 14,11% se producen más de una vez)

Tabla 12. Frecuencias relativas a los términos de indización analizados

Otras diferencias importantes pueden apreciarse en la tabla 12. En ella, se puede observar los diversos volúmenes de términos que utiliza cada sistema de indización. En el caso de los *KW+* se aprecia que es el vocabulario que utiliza mayor variedad de términos (9.107), duplicando a los *AKW* (5.548) y triplicando a los *MeSH* (2.884). Esto conduce a que los *KW+* también son los términos que más veces coinciden con otros términos en el mismo documento y es la tipología que más términos utiliza por documento. Del mismo modo, es de esperar que los *KW+* al tratarse de un vocabulario libre, sin ningún tipo de control terminológico la dispersión de términos sea mayor que la que se produce en un vocabulario controlado como es el caso de los *MeSH*. A pesar de ello, como puede apreciarse en la tabla 12, la dispersión terminológica que presentan

los *KW+* es menor que la que muestran los *AKW*. En cuanto a estos últimos, los datos reflejan que los *AKW* son los términos menos numerosos, los que tienen las frecuencias de aparición más bajas y se evidencia una gran dispersión terminológica. Además, sus relaciones son más escasas que las de los *KW+* pero más numerosas que las de los *MeSH*, aunque la mayor parte de ellas sólo se producen una vez. En el caso de los *MeSH* son los términos menos numerosos y que menos cantidad se utilizan para indizar cada documento. Esto puede deberse, en gran medida, a que en el análisis sólo se han tenido en cuenta los términos *Major* reduciendo, sensiblemente, el número de descriptores por documento. No obstante, son los términos que menos dispersión terminológica presentan y que, a pesar de que ofrecen el número más bajo de relaciones, son los que están más conectados en proporción al volumen de términos.

Estas diferencias que presentan cada tipología de términos en las distribuciones de sus frecuencias de aparición y de sus relaciones interdocumentales influyen en la delimitación temática a través de las visualizaciones. El grado de cohesión entre los términos condiciona la discriminación de las relaciones que son superfluas de las que no, comprometiendo la claridad y nitidez de las visualizaciones en la formación de agrupaciones temáticas que permitan identificar líneas o tendencias en la investigación.

Teniendo esto presente, se observa importantes diferencias en la información que ofrecen los mapas generados según cada tipología de términos de indización. En el caso de los *KW+* se aprecia que la identificación temática es muy similar a la obtenida en las fases de trabajo anteriores (publicaciones 3, 4 y 5). Aunque haya podido variar la presencia de algunos términos, la estructura temática no ha cambiado. Nuevamente, es posible distinguir claramente hasta un total de cuatro líneas de investigación, agrupadas, a su vez, por las dos grandes naturalezas que se han visto que caracterizan a este dominio, la investigación de carácter clínica y básica. La primera de ellas definida, fundamentalmente, por la L.I.1 y la segunda por el resto de líneas (L.I.2; L.I.3 y L.I.4).

En cuanto a los *AKW*, los mapas evidencian que no existe una delimitación temática clara y comprensiva del dominio que permita detectar líneas de investigación. Se observa que la presencia de términos relacionados con la investigación de carácter clínico es mucho más reducida que en el caso de los *KW+*. En cambio, la investigación de carácter más básico está mejor representada, aunque está poco definida y no es posible la identificación de líneas de investigación. En términos generales, el mapa se reduce a una red en la que, a partir del nodo central *Stem Cell*, se desarrollan una serie de relaciones entretejidas donde es muy difícil identificar algún tipo de conexión temática.

En cuanto a los *MeSH*, lo primero que llama la atención es una gran interrelación entre ellos. Las agrupaciones se conectan por la presencia de numerosos “nodos puente” que dificultan considerablemente la identificación de la estructura general. En este caso, tampoco es posible una identificación clara de líneas de investigación, aunque las agrupaciones de términos están temáticamente mejor relacionadas que en el caso de los *AKW*. A pesar de ello, lo que sí se puede apreciar es que en este caso la investigación de carácter más clínico está mejor representada que la básica, como se puede apreciar en la fig. 9 (p. 184), donde el foco de la investigación básica aparece más disperso que el de la clínica.

Por lo tanto, se puede considerar que los *AKW* utilizan una terminología más relacionada con la investigación de carácter básico, y se muestra como un vocabulario más permeable a líneas de investigación emergentes. En cambio, en el caso de los *MeSH* al utilizar una terminología más relacionada con la investigación de carácter clínico, se evidencia como un vocabulario más permeable a líneas de investigación de carácter más consolidado. Por su parte los *KW+*, han permitido obtener las visualizaciones más claras donde ha sido posible la localización e identificación de las principales líneas de investigación. Además, la representación de la investigación clínica y básica guarda un mayor equilibrio, siendo un vocabulario muy permeable tanto a líneas de investigación emergentes como a las consolidadas.

3. Teniendo en cuenta estos resultados, ¿cabe deducir que el empleo de diferentes tipos de términos de indización como unidades de análisis requiere de ajustes metodológicos específicos para cada caso? ¿qué criterios habría que tener en cuenta para desarrollar una propuesta metodológica que contemple dichos ajustes?

Los resultados obtenidos a través de las diferentes visualizaciones ofrecen informaciones distintas en función del término de indización utilizado como unidad de análisis. La razón principal se debe a la distribución de las frecuencias de aparición y el número de relaciones que se producen entre los diferentes términos. La cantidad de términos utilizados, el número de veces que aparecen en los documentos, así como, el número de relaciones que mantienen con otros términos, influyen en los parámetros de visualización que han caracterizado la propuesta metodológica de esta tesis.

Las diferencias más destacables pueden resumirse de la siguiente manera:

- **AKW:** de las tres tipologías es la que menor número de términos utiliza y la que mayor dispersión terminológica presenta. Esto afecta al número y a la distribución de las relaciones, donde los valores de la co-ocurrencias tienden a ser bajos y su variedad es la más escasa de las tres tipologías. Estas características afectan a la propuesta metodológica de esta tesis. En el caso de la construcción de mapas basados en grafos desarrollados con Pajek, se utilizan valores de co-ocurrencia sin normalizar trabajando con cifras absolutas que, en el caso de los AKW, alcanzan valores muy similares entre sí. Este hecho dificulta profundamente, la discriminación de las relaciones superfluas con el algoritmo de poda PFNeT, esquematizando excesivamente la estructura relacional de estos términos. El resultado que se obtiene es una red donde es difícil apreciar agrupaciones claras y definidas como se ha visto en la fig. 4 (Publicación 6, p. 181). En el caso de la construcción de mapas basados en la distancia desarrollados con *VOSviewer*, a pesar de que los valores de co-ocurrencia se normalizan en función de la frecuencia de aparición del término, ofreciendo la posibilidad de obtener

mayor diversidad de valores que en el caso de los datos en bruto o *raw data*, al contar con frecuencias que alcanzan cifras bajas, el peso de las relaciones obtienen valores muy similares entre sí, dificultando un agrupamiento claro de los elementos de la red, como se observa en la fig. 5 (Publicación 6, p. 181), generando un elevado número de clústeres.

- **MeSH:** en relación con los AKW, esta tipología tampoco utiliza una gran variedad de términos, pero en cambio, es la que menor dispersión terminológica presenta. La naturaleza del vocabulario condiciona esta realidad, el número medio de apariciones por documento es más escaso y son términos más específicos y consolidados. Esto se ve reflejado en el número y distribución de las relaciones donde los valores de la co-ocurrencias tienden a ser elevados y cuya variedad tiende a ser la más uniforme de las tres tipologías. Es decir, estas unidades de análisis se caracterizan por el empleo de pocos términos, pero están muy relacionados entre sí, su presencia en los documentos es muy recurrente. Esto también tiene consecuencias a la hora de la elaboración de los mapas bibliométricos. En el caso de las representaciones obtenidas mediante *Pajek*, no es posible obtener una representación estructural clara debido a que *PFNet* no puede discriminar adecuadamente las relaciones superfluas debido a que los valores de co-ocurrencia son muy similares. El resultado es una red donde los elementos están muy interrelacionados y es difícil apreciar agrupaciones claras y definidas como ocurre en la fig. 7 (Publicación 6, p. 183). En cambio, las representaciones obtenidas con *VOSviewer*, muestran agrupaciones algo más claras y un poco mejor definidas, aunque el elevado número de clústeres impide la identificación clara de líneas de investigación como se puede observar en la fig. 8 (Publicación 6, p. 183). Esto, conduce a pensar que la normalización empleada por *VOSviewer* permite eliminar más número de relaciones superfluas que si se emplean valores *raw data*. No obstante, también hay que decir que, en el caso del análisis desarrollado en el Dominio 1, si fue posible localizar e identificar las principales líneas de investigación ya que se trataba de un estudio de granularidad gruesa donde la distribución de las relaciones si se ajustaban a los parámetros de la propuesta metodológica desarrollada para ese análisis.

- **KW+**: es la tipología que mayor número de términos utiliza a pesar de que presenta una gran dispersión terminológica. En cuanto al número y a la distribución de las relaciones, es posible apreciar que existe una gran heterogeneidad entre los valores de co-ocurrencia, donde se distingue un gran número de términos que se relacionan frecuentemente y otro gran número de términos que se relacionan escasamente. De este modo, es posible discriminar con mayor facilidad cuáles son las relaciones que tienen mayor peso de las que no. Esto favorece a que haya una distribución de ocurrencias y co-ocurrencias lo suficientemente amplia y variada que se ajusta correctamente a la propuesta metodológica de esta tesis. Tanto las representaciones obtenidas mediante *Pajek* como las obtenidas mediante *VOSviewer*, salvando las diferencias metodológicas que conlleva el empleo de cada software, han permitido la identificación y localización más clara de las líneas de investigación correspondientes al dominio analizado, fig. 1 y 2 (Publicación 6, p. 179).

Todas estas diferencias metodológicas detectadas en el desarrollo de esta publicación, sería recomendable que fueran estudiadas y tenidas en cuenta para futuros análisis, si lo que se pretende es obtener visualizaciones óptimas que permitan la mejor identificación posible de las principales tendencias o líneas de investigación.

8. CONCLUSIONES

A continuación, se exponen las conclusiones generales que se obtienen tras los resultados mencionados en el capítulo anterior y que han sido obtenidos en las distintas fases de trabajo planteadas y en las sucesivas publicaciones desarrolladas a lo largo de esta tesis doctoral.

- **Conclusión 1**

La propuesta metodológica planteada en esta tesis ha permitido localizar, identificar y visualizar las principales tendencias o líneas de investigación que caracterizan la producción científica de los dominios de estudio seleccionados. Ha sido posible detectar los aspectos temáticos de carácter específico más allá de disciplinas científicas o áreas temáticas, con independencia de la amplitud o especificidad del dominio utilizado. De este modo, dicha metodología ofrece la posibilidad de adaptarse a diferentes niveles de agregación temática y bibliométrica permitiendo realizar estudios tanto de granularidad fina como gruesa.

- **Conclusión 2**

Las fuentes de información seleccionadas han permitido reunir el universo documental sobre el que se han realizado los diversos análisis realizados. La propuesta metodológica se ha adaptado adecuadamente, a las diferencias que existen entre ambas fuentes utilizadas, Medline y SCI-EXPANDED. La cobertura temática que en un caso es especializada y en el otro multidisciplinar, no ha supuesto un inconveniente a la hora de localizar tendencias de investigación o aspectos temáticos con mayor o menor grado de especificidad. Del mismo modo, las diferencias en cuanto a los sistemas de indexación que emplean, tampoco han supuesto una limitación para utilizar los términos de indexación como elementos representativos del contenido conceptual de los documentos y detectar con ellos las principales líneas de investigación.

- **Conclusión 3**

El empleo del análisis de co-palabras como parte de la propuesta metodológica planteada, ha demostrado ser una herramienta muy adecuada para la delimitación temática. Ha hecho posible la identificación de las principales tendencias de la investigación gracias a la aparición conjunta de los términos de indización a través de los documentos y a las relaciones que se establecen entre ellos. Como se ha observado en los resultados obtenidos en las publicaciones que integran esta tesis, el uso de los análisis de co-palabras ha demostrado captar los aspectos temáticos más dinámicos de la investigación ya que se trata de unas unidades de análisis muy permeables a los cambios de tendencias y a reflejar con gran inmediatez la incidencia de los contextos científicos, sociales y políticos propios de los dominios más controvertidos y áreas emergentes.

- **Conclusión 4**

La combinación de los programas informáticos especializados en el análisis de grandes redes y visualización de la información científica ha demostrado que, tanto *Pajek* como *VOSviewer* pueden ser utilizados como herramientas complementarias según la propuesta metodológica planteada, siempre y cuando se tenga en cuenta las diferencias metodológicas que conlleva el empleo de cada uno de ellos. *Pajek* en combinación con *PfNET* ha permitido sintetizar y resaltar la estructura relacional de los términos de indización en función de los dominios de análisis, mientras que, *VOSviewer* ha posibilitado la identificación de las agrupaciones temáticas y los focos de investigación de una manera más plástica.

- **Conclusión 5**

El desarrollo de estudios temporales con el fin de observar la dinamicidad temática de los dominios analizados, ha permitido la obtención de una sucesión de imágenes que posibilitan la localización e identificación tanto de las líneas de investigación claramente

consolidadas como de las emergentes. Este hecho destaca la capacidad de predicción de este tipo de estudios en el desarrollo de un dominio temático a lo largo del tiempo. Una vez más, estos análisis han sido posibles con independencia de la cobertura temporal, temática o geográfica del dominio empleado, lo que refuerza la capacidad de la propuesta metodológica para la delimitación temática en diferentes niveles de agregación.

- **Conclusión 6**

El empleo como unidades de análisis de diferentes términos de indización procedentes de las bases de datos bibliográficas utilizadas en la tesis, sobre un mismo dominio científico, más concretamente, sobre la misma colección documental, ha permitido evaluar la capacidad de cada uno de ellos para localizar e identificar las principales tendencias de investigación. Los resultados han demostrado que es necesario tener en cuenta las diferencias que se producen entre los términos de indización debidas, fundamentalmente, a su naturaleza documental, sobre todo, en dominios científicos muy específicos y en niveles de agregación más concretos en los análisis bibliométricos.

Es posible llegar a las siguientes conclusiones según las características de cada término de indización:

- *MeSH*: han demostrado ser un lenguaje más lento a la hora de incorporar nueva terminología debido a la rigidez que le otorga ser un vocabulario controlado. Este rasgo afecta, sobre todo, a la localización e identificación de líneas de investigación de carácter emergente o en formación. En cambio, no es un inconveniente para la identificación de tendencias de investigación de carácter más consolidado, como ocurre en el caso de la Publicación nº1. Por lo tanto, este tipo de términos se adaptan mejor a estudios de granularidad gruesa.

- *Author Keywords*: a diferencia de los anteriores, estos términos tienen mayor capacidad para detectar líneas de investigación emergentes, ya que, como se ha comprobado, se trata de una tipología más permeable hacia términos más actualizados. Aunque, por otro lado, presentan el inconveniente de experimentar una gran dispersión terminológica por falta de una estandarización entre los términos y además, todavía tienen una escasa presencia en las bases de datos lo que dificulta la identificación clara de líneas de investigación.

- *KeyWords Plus*: son términos que, como en el caso anterior, también tienen una gran capacidad para incorporar terminología actualizada a pesar de que también sufren una gran dispersión terminológica por su falta de normalización. Pero a pesar de ello, debido a su abundancia y al elevado número de relaciones que mantienen con otros términos a través de los documentos, favorece la identificación temática y permite localización de líneas de investigación, tanto de carácter consolidado como emergentes. Por lo tanto, de las tres tipologías es la que mejor se adecúa a la propuesta metodológica y a través de la cual se ha obtenido la identificación, localización y visualización más clara de las principales líneas de investigación.

9. CONCLUSIONS

Next, we present the general conclusions that are obtained after the results mentioned in the previous chapter and that we have obtained in the different phases of study proposed and in the successive publications developed throughout this doctoral thesis.

- **Conclusion #1**

The methodology proposed in this thesis has allowed us to locate, identify and visualize the main trends or lines of research that characterize the scientific production of the selected study domains. We have been able to detect the thematic aspects of specific character beyond scientific disciplines or thematic areas, regardless of the amplitude or specificity of the domain used. Thus, this methodology offers the possibility of adapting to different levels of thematic and bibliometric aggregation, allowing studies of both fine and coarse granularity.

- **Conclusion #2**

The sources of information selected have allowed us to gather the documentary universe on which we have performed the various analyzes. The methodological proposal has been adequately adapted to the differences between the two sources used, Medline and SCI-EXPANDED. Thematic coverage that in one case is specialized and in the other multidisciplinary, has not been an inconvenience when it comes to locating research trends or thematic aspects with a greater or lesser degree of specificity. In the same way, the differences in the indexing systems used do not have a limitation to use indexing terms as some representative elements of the conceptual content of the documents and to detect the main lines of research with them.

- **Conclusion #3**

The use of the co-words analysis as part of the raised methodological proposal has proved to be a very adequate tool for thematic delimitation. It has made possible the identification of the main trends research thanks to the joint appearance of the indexing terms through the documents and to the relations established between them. As it has been observed in the results obtained in the publications that integrate this thesis, the use of the co-words analysis has been shown to capture the most dynamic thematic aspects of the research since these are units of analysis very permeable to the changes of trends and to reflect with great immediacy the incidence of the scientific, social and political contexts typical of the most controversial domains and emerging areas.

- **Conclusion #4**

The combination of specialized software in the analysis of large networks and the visualization of scientific information has shown that both Pajek and VOSviewer can be used as complementary tools according to the approached methodological, considering the methodological differences that entails the use of each one of them. Pajek in combination with PfNET has allowed to synthesize and highlight the relational structure of the indexing terms according to of the domains of analysis, whereas, VOSviewer has made it possible to identify thematic clusters and research focus in a more plastic way.

- **Conclusion #5**

The development of temporal studies to observe the thematic dynamicity of the analyzed domains has allowed the obtaining of a succession of images that make it possible to locate and identify both the clearly consolidated and emerging research lines. This highlights the predictive capacity of this type of studies in the development of a thematic domain over time. Once again, these analyzes have been possible regardless of the temporal, thematic or geographical coverage of the domain used,

which reinforces the capacity of the methodological proposal for thematic delimitation at different levels of aggregation.

- **Conclusion #6**

The use, as units of analysis, of different indexing terms from the bibliographic databases used in the thesis, on the same scientific domain, more specifically, on the same documentary collection, has allowed to evaluate the capacity of each of them to locate and identify the main research trends. The results have shown that it is necessary to consider the differences that occur between indexing terms due to their documentary nature, especially in very specific scientific domains and in more specific levels of aggregation in the bibliometric analyzes.

It is possible to conclude according to the characteristics of each indexing term the follow:

- MeSH: they have shown be a slower language when incorporating new terminology due to the rigidity that gives to be a controlled vocabulary. This trait affects, above all, the location and identification of lines of research of emergent or in formation character. On the other hand, it is not a problem for the identification of research trends of a more consolidated nature, as happened in the case of Publication #1. Therefore, these types of terms are best suited to studies of coarse granularity.
- Author Keywords: unlike the previous ones, these terms have greater capacity to detect research lines of research emerging, since, as has been proven, it is a more permeable typology towards more updated terms. Although, on the other hand, they have the disadvantage of experiencing a large terminological dispersion due to a lack of standardization between the terms and, in addition, they still have a scarce presence in the databases, which makes it difficult to clearly identify lines of research.

- KeyWords Plus: they are terms that, as in the previous case, also have a great capacity to incorporate updated terminology although they also suffer a great terminological dispersion by them in the absence of normalization. But despite this, due to its abundance and the high number of relationships they maintain with other terms through the documents, it favors the thematic identification and allows the localization of research lines, both of a consolidated and emerging nature. Therefore, of the three typologies is the one that best suits the methodological proposal and through which the identification, location and clear visualization of the main research lines has been obtained.

10. PERSPECTIVAS DE INVESTIGACIONES FUTURAS

La propuesta metodológica planteada en esta tesis ha permitido conseguir el objetivo principal de la misma como es la localización y visualización de las principales líneas de investigación. Además, se ha comprobado su validez aplicándola a dos dominios científicos con grandes diferencias en cuanto a cobertura temática, temporal y geográfica y empleando diferentes términos de indización como unidades de análisis. Por todo ello, ha demostrado su adecuación como metodología para la delimitación temática, lo que la convierte en una herramienta con un gran potencial para ser aplicada a diferentes estudios bibliométricos que pueden ser desarrollados como líneas de trabajo futuras.

Una de esas posibles aplicaciones podría consistir en emplear la propuesta metodológica para la identificación de cualquier dominio científico, con independencia, como se ha comprobado, de su cobertura o especificidad temática. Dada su demostrada capacidad para detectar los aspectos temáticos más dinámicos y las líneas de investigación emergentes, sería idóneo para aquellos dominios que son más controvertidos, más innovadores o en continua transformación. Aunque, igualmente podría utilizarse para otros dominios más estables, donde puede ser interesante caracterizar la investigación más consolidada.

En cuanto a la línea de trabajo que ha iniciado esta tesis, otra aplicación interesante podría consistir en comparar las líneas de investigación detectadas en el dominio más específico como es el Dominio 2 (análisis de la producción científica española con células madre, 1997-2010), con las líneas que se obtendrían de otras regiones o países del mundo, con el fin de valorar las diferencias o similitudes temáticas que se producen en el campo. Del mismo modo, también sería interesante seleccionar las líneas de investigación obtenidas en dicho dominio y caracterizar la producción científica de cada línea a través de un análisis bibliométrico a nivel micro. Asimismo, se podrían seleccionar los términos que integran cada línea de investigación detectada y relanzarlos como

términos de búsqueda en las bases datos, con el fin de recoger una colección documental más precisa y exhaustiva sobre el campo y poder realizar un estudio bibliométrico más delimitado. Como también, podría ser objeto de estudio en el futuro, rehacer nuevos análisis temporales o evolutivos con programas informáticos especializados en la representación de redes y en la visualización de la información teniendo en cuenta la perspectiva temporal como *Visione*, o generar mapas que permitan la representación fija de los elementos en la red con sucesivas imágenes a lo largo del período, como pueden ser las técnicas de superposición de los denominados *Overlays Maps*.

A pesar de que la propuesta metodológica ha permitido alcanzar los objetivos que se perseguían, hay algunos aspectos que pueden ser mejorados. Como se ha comprobado, el empleo de diferentes tipos de términos de indización ha impedido que se puedan identificar las mismas líneas de investigación con cada uno de ellos. Podría ser interesante de cara al futuro, detectar cuáles son los parámetros metodológicos que hay que tener en cuenta con cada tipología para así obtener imágenes más nítidas sobre el dominio, a pesar de las inevitables diferencias que por su naturaleza documental han de producirse.

En relación con los términos de indización, también podría ser interesante probar esta metodología con otros tipos de términos. Éstos podrían proceder de los campos del título, resumen, palabras clave, etc., o incluso del texto completo y ser extraídos de manera automática con algunos de los softwares especializados en este fin.

No obstante, aunque los términos de indización han contribuido en esta tesis a localizar las principales líneas de investigación a través del análisis de co-palabras, es posible que si se combinara esta técnica con otras que han demostrado ser válidas para la delimitación temática como el Análisis de Co-citación de Autores por ejemplo, es posible que se obtuviera una mayor información temática sobre el dominio analizado.

Por último, indicar que de igual modo podría ser un campo de trabajo a desarrollar con el tiempo, el uso de otras bases de datos bibliográficas como fuentes de información. Sería interesante, por ejemplo, el uso de la base de datos *Scopus*, ya que cuenta con una cobertura más amplia que la Web of Science o Medline. Esto permitiría contar con una producción científica más variada.

11. BIBLIOGRAFÍA

- Aaronson, S. (1975). The footnotes of science. *Mosaic*, 6(2), 22-27.
- Alcalde, I. (2012). Visualización de la información: ideas, conocimiento e innovación. En *ignasialcalde from data to knowledge*. Recuperado de <http://www.ignasialcalde.es/visualizacion-de-la-informacion-ideas-conocimiento-e-innovacion/>
- Araya, A.A. (2003). The hidden side of visualization. *Techné: Research in Philosophy and Technology*, 7(2), 74-119.
- Arce, C. (1996). Escalamiento multidimensional. En J. Arnau, *Métodos y técnicas avanzadas de análisis de datos en Ciencias del Comportamiento* (p. 23-46). Barcelona: Universitat de Barcelona.
- Batagelj, V. & Mrvar, A. (2010). *Pajek 1.28: Package for large network analysis*. Cambridge, (New York): CUP.
- Baldwin, C., Hughes, J., Hope, T., Jacoby, R. & Ziebland, S. (2003). Ethics and dementia: mapping the literature by bibliometric analysis. *International Journal of Geriatric Psychiatry*, 18(1), 41-54.
- Barnes, J.A. (1954). Class and Committees in a Norwegian Island Parish. *Human Relations*, 7(1), 3-58.
- Bessalar, P. & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: a method and an exploratory case study. *Scientometrics*, 68(3), 377-93.
- Bhattacharya, S. & Basu, P.K. (1998). Mapping a research area at the micro level using co-word analysis. *Scientometrics*, 43(3), 359- 372.
- Biblioteca Regional de Medicina (BIREME). (1967). *Descriptores en ciencias de la salud, DeCS*; Brasil. Recuperado de <http://decs.bvs.br/E/decswebe.htm>
- Bishop, A.E., Buttery, L.D.K. & Polak, J.M. (2002). Embryonic stem cells. *Journal of Pathology*, 197(4), 424-429.

- Blake, M.B., Ince, M. & Dean, C.J. (2005). Inclusion of women in cardiac research: Current trends and need for reassessment. *Gender Medicine*, 2(2), 71-75.
- Bornmann, L. & Haunschild, R. (2016). Overlay maps based on Mendeley data: The use of altmetrics for readership networks. *Journal of the American Society for Information Science and Technology (JASIST)*, 67(12), 3064-3072.
- Boyack, K.W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.
- Boyack, K.W. & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology (JASIST)*, 61(12), 2389-2404.
- Braam, R.R., Moed, H.F. & van Raan, A.F.J. (1991a). Mapping of Science by combined co-citation and word analysis. I: structural aspects. *Journal of the American Society for Information Science (JASIS)*, 42(4), 233-251.
- Braam, R.R., Moed, H.F. & van Raan, A.F.J. (1991b). Mapping of Science by combined co-citation and word analysis. II: dynamic aspects. *Journal of the American Society for Information Science (JASIS)*, 42(4), 252-266.
- Brandenburg F.J., Himsolt M. & Rohrer C. (1996) An experimental comparison of force-directed and randomized graph drawing algorithms. En Brandenburg F.J. (Eds.) *Graph Drawing*. GD 1995. Lecture Notes in Computer Science, v. 1027. Heidelberg (Berlin): Springer.
- Börner, K., Chen, C. & Boyack, K.W. (2003). Visualizing knowledge domains. *Annual Review of Information Science & Technology*, 37, 179-255.
- Börner, K., Contractor, N.S., Falk-Krzesinski, H.J., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W. & Uzzi, B. (2010a). A Multi-Level Systems Perspective for the Science of Team Science. *Science Translational Medicine*. 2(49), 49(cm)24.

- Börner, K., Huang, W., Linnemeier, M., Duhon, R.J., Phillips, P., Ma, N., Zoss, A., Guo, H. & Price, M. (2010b). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Network Workbench Tool. *Scientometrics*, 83(3), 863-876.
- Boyack, K.W. & Börner, K. (2003). Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5) 447-461.
- Boyack, K.W., Klavans, R. & Borner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351-374.
- Boyack, K.W., Small, H. & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(9), 1759-1767.
- Buter, R.K. & Noyons, E.C.M. (2001). Improving the functionality of interactive bibliometric science maps. *Scientometrics*, 51(1), 55-67.
- Buzydowski, L. (2003). *A comparison of self-organizing maps and Pathfinder Networks for the mapping of co-cited Authors*. (Tesis Doctoral. Universidad de Drexel, Drexel). Recuperado de: <https://idea.library.drexel.edu/islandora/object/idea%3A150>
- Buzydowski, J., White, H.D. & Lin, X. (2002). Term co-occurrence analysis as an interface for digital libraries. *Lecture Notes in Computer Science Series*, 2539, 133-144.
- Calero, C., Buter, R., Valdes, C.C. & Noyons, E. (2006). How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, 66(2), 365-376.
- Callon, M., Courtial, J.P. & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1), 155-205.
- Callon, M., Courtial, J.P., Turner, W.A. & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191-235.

- Campbell, K.H.S, McWhir, J., Ritchie, W.A. & Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature*, 380(6569), 64-66.
- Cambrosio, A., Limoges, C., Courtial, J.P., Laville, F., (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27(2), 119-143.
- Card, S.K., Mackinlay, J.D. & Shneiderman, B. (1999). Information Visualization. En S.K. Card, J.D. Mackinlay & B. Shneiderman (Eds.), *Readings in information visualization, using vision to think* (p. 1-34). San Francisco: Morgan Kaufmann.
- Chalik, T. & Jirina, M. (2006). Law of cumulative advantages in the evolution of scientific fields. *Scientometrics*, 66(3), 441-9.
- Chaussabel, D. & Sher, A. (2002). Mining microarray expression data by literature profiling. *Genome Biology*, 3(10), research0055.1–research0055.16.
- Chen, C. (1998a). Generalised Similarity Analysis and Pathfinder Network Scaling. *Interacting with computers*, 10(2), 107-128.
- Chen, C. (1998b). Bridging the gap: the use of pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 9(3), 267-286.
- Chen, C. (1999). Visualising semantic spaces and author cocitation networks in digital libraries. *Information Processing & Management*, 35(3), 401–420.
- Chen, C. (2004) Searching for intellectual turning points: Progressive Knowledge Domain Visualization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 101(Suppl. 1), 5303-5310.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literatur. *Journal of the American Society for Information Science and Technology (JASIST)*, 57(3), 359-377.
- Chen, C., Chen Y., Hou J. & Liang Y. (2009). CiteSpace I: detecting and visualizing emerging trends and transient patterns in scientific literatur. *Journal of the China Society for Scientific and Technical Information*, 28(3), 401-21.

- Chen, C., Dubin, R. & Kim, M. C. (2014). Emerging trends and new developments in regenerative medicine: a scientometric update (2000 - 2014). *Expert Opinion on Biological Therapy*, 14(9), 1295-1317.
- Chen, C., Ibekwe-SanJuan, F. & Hou, J. (2010). The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(7), 1386-409.
- Chen, C. & Kuljis, J. (2003). The rising landscape: a visual exploration of superstring revolutions in physics. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5), 435-446.
- Chen, C & Leydesdorff, L. (2014). Patterns of Connections and Movements in Dual-Map Overlays: A New Method of Publication Portfolio Analysis. *Journal of the American Society for Information Science and Technology (JASIST)*, 65(2), 334-351.
- Chen, C. y Paul, R.J. (2001). Visualizing a knowledge domain's intellectual structure. *Computer*, 34(3), 65-71.
- Chen, C., Paul, R. J. & O'keefe, B. (2001). Fitting the jigsaw of citation: information visualization in domain analysis. *Journal of the American Society for Information Science and Technology (JASIST)*, 52(4), 315-330.
- Chen, D.Z., Huang, M.H., Hsieh, H.C., & Lin, C.P., (2011). Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Infometrics*, 5(3), 400-412.
- Chen, H. & Sharp, B.M. (2004). Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics*, 5, 147.
- Chinchilla-Rodríguez, Z., & Moya-Anegón, F. (2007). *La Investigación Científica Española: Una Aproximación Métrica*. Granada: Universidad de Granada.
- Chiu, W.T. & Ho, Y.S. (2007). Bibliometric analysis of tsunami research. *Scientometrics*, 73(1), 3-17.

- Chiu, W.T., Huang, J.S. & Ho, Y.S. (2004). Bibliometric analysis of Severe Acute Respiratory Syndrome-related research in the beginning stage. *Scientometrics*, 61(1), 69-77.
- Christensen, H.D. (2014). The framing of scientific domains: about UNISIST, domain analysis and art history. *Journal of Documentation*, 70(2), 261-281.
- Chung, Y., Klimanskaya, I., Becker, S., Marh, J., Lu, S.J., Johnson, J., Meisner, L. & Lanza, R. (2006). Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature*, 439(7073), 216-219.
- Chung, Y.M. & Lee, J.Y. (2001). A corpus-based approach to comparative evaluation of statistical term association measures. *Journal of the American Society for Information Science and Technology*, 52(4), 283-296.
- Clarke, J.N., (1983). Sexism, feminism and medicalism: a decade of literature on gender and illness. *Sociology of Health and Illness*, 5(1), 62-82.
- Clarke, A., Gatineau, M., Thorogood, M. & Wyn-Roberts, N. (2007). Health promotion research literature in Europe 1995-+2005. *The European Journal of Public Health*, 17(suppl_1), 24-28.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2012). SciMAT: A new Science Mapping Analysis Software Tool. *Journal of the American Society for Information Science and Technology (JASIST)*, 63(8), 1609-1630.
- Cordon, M.Y. & Blackett, N.M. (1998). Reconstruction of the hematopoietic system after stem cell trasplantation. *Cell Transplant*, 7(4), 339-44.
- Costa, J. (1998). *La esquemática. Visualizar la información*. Barcelona: Paidós.
- Costas, R., Moreno, L., & Bordons, M. (2008). Solapamiento y singularidad de MEDLINE, WoS e IME para el análisis de la actividad científica de una región en Ciencias de la Salud. *Revista Española de Documentación Científica*, 31(3), 327-343.
- Courtial, P., Cahlik, T. & Callon., M. (1994). A model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics*, 31(2), 173-192.

- Coulter, N., Monarch, I. & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science (JASIS)*, 49(13), 1206–1223.
- Crosby, A.W. *The Measure of Reality: Quantification and Western Society 1250-1600*. 97. London: Cambridge University Press.
- Daly, C., Clemens, F., Lopes Sendon, J.L., Tavazzi, L., Boersma, E., Danchin, N., Delahaye, F., Gitt, A., Julian, D., Mulcahy, D., Ruzylo, W., Thygessen, K., Berheugt, F. & Fox, K.M. (2006). Gender differences in the management and clinical outcome of stable angina. *Circulation*, 113(4) 490-498.
- Davidson, G. S., Wylie, B. N. & Boyack, K. W. (2001). Cluster stability and the use of noise in interpretation of clustering. En *Proceedings of the IEEE symposium on information visualization*, (pp. 23–30).
- De Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. En J.R. Barra, F. Brodeau, G. Romier & B. van Cutsem (Eds.), *Recent developments in statistics* (p. 133-145). Amsterdam: Nort-Holland.
- De Leeuw, J. y Heiser, W. J. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. En J.C. Lingoes, E.E. Roskam e I. Borg (Eds.). *Geometric representations of relational data* (p. 735-752). Ann Arbor, MI: Mathesis Press.
- Devesa, S.S., Bray, F., Vizcaino, A.P. & Parkin, D.M. (2005). International lung cancer trends by histologic type: Male: female differences diminishing and adenocarcinoma rates rising. *International Journal of Cancer*, 117(2), 294-299.
- Ding, Y., Chowdhury, G.G. & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997. *Journal of Information Science*, 25(1), 67-78.
- Ding, Y., Chowdhury, G.G. & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, 37(6), 801-817.

- Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem-cell lines: formation of visceral Yolk-Sac, blood islands and myocardium. *Journal of Embryology and Experimental Morphology*, 87(87), 27-45.
- Egghe, L. (2009). New relations between similarity measures for vectors based on vector norms. *Journal of the American Society for Information Science and Technology*, 60(2), 232–239.
- Egghe, L., & Rousseau, R. (2006). Classical retrieval and overlap measures satisfy the requirements for rankings based on a Lorenz curve. *Information Processing and Management*, 42(1), 106–120.
- Elsevier (2002). Scopus [Base de datos]. Recuperado de <http://www.scopus.com/home.url>
- Espinosa-Calvo, M.E., Vargas-Quesada, B., Guerrero-Bote, V. & Moya-Anegón, F. (2009). Comparative study of six National Scientific Dominions. *Revista Española de Documentación Científica*, 32(3), 9-28.
- EuroStemCell, iCeMS, Elsevier (2013). *Stem cell research. Trends and perspectives on the evolving international landscape*. Recuperado de http://www.eurostemcell.org/files/Stem-Cell-Report-Trends-and-Perspectives-on-the-Evolving-International-Landscape_Dec2013.pdf
- Evans, M.S. & Kaufman, M.H. (1981). Establishment in culture of pluripotential stem from mouse embryos. *Nature*, 292(5819), 154-156.
- Fernández, M.T., Cabrero, A., Zulueta, M.A. and Gómez I. (1993). Constructing a relational database for bibliometric analysis. *Research Evaluation*, 3(1), 55-62.
- Floyd, R.W. (1962). Algorithm 97: Shortest Path. *Communications of the ACM*, 5(6), 345.
- Freeman, L. C. (2000). Visualizing social networks. *Journal of Social Structure*, 1, 1-15.
- Fruchterman, T. M. J. & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164.

- Galaskiewicz, J. & Wasserman, S. (1993). Social Network Analysis. Concepts, Methodology, and Directions for the 1990s. *Sociological Methods & Research*, 22(1), 3-22.
- Garfield, E. (1963). Citation indexes in sociological and historical research. *American Documentation*, 14(4), 289-291.
- Garfield, E. (1975). ISI's Atlas of Science may help students in choice of career in science. *Current Contents*, (29), 5-8.
- Garfield, E. (1976). Social Sciences Citation Index Clusters. *Current Contents*, 27, 5-11.
- Garfield, E. (1981). Introducing the ISI Atlas of Science: Biochemistry and molecular biology, 1978-80. *Current Contents*, (42), 5-13.
- Garfield, E. (1984). Introducing the ISI Atlas of Science: Biotechnology and molecular genetics, 1981/82 and bibliographic update for 1983/84. *Current Contents*, (41), 3-15.
- Garfield, E. (1986). *Towards scientography*. Essays of an Information Scientist, 9, 324.
- Garfield, E. (1988). The encyclopedic ISI-Atlas of Science launches 3 new sections: Biochemistry, Immunology, and Animal & Plant Sciences. *Current Contents*, (7), 3-8.
- Garfield, E. (1990). Keywords Plus-ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current contents on diskette. *Current Contents*, 1(32), 5-9.
- Garfield, E. (1992). Psychology research, 1986-1990: a citationist perspective on the highest impact papers, institutions and authors. *Current Contents*, 41, 5-13.
- Garfield, E. (1998). Mapping the world of science. En *Presentation - Topical paper presented at the 150th Anniversary Meeting of the AAAS, Philadelphia, February 14, (Unpublished)*, (pp. 1-19). Recuperado de <http://www.garfield.library.upenn.edu/papers/mapsciworld.html>
- Garfield, E., Sher, I.H. (1964). *The Use of Citation in Writings the History of Science*. Philadelphia: Institute for Scientific Information.

- Garfield, E., Sher, I.H., Torpie, R.J. (1964). *The use of citation data in writing the history of science*. Philadelphia: Institute for Scientific Information.
- Garfield, E., Sher, I. H. (1993). Keywords Plus™– Algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44(5), 298-299.
- Gender and Health Group. (2000). Guidelines for the analysis of Gender and Health Group. Recuperado el 2 de marzo de 2010, de <http://www.lstmed.ac.uk/sites/default/files/pictures/Guidelines%20for%20the%20Analysis%20of%20Gender%20and%20Health.pdf>
- Glänzel, W. & Thijs, B. (2011), Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88(1), 297-309.
- Glenisson, P., Glanzel, W., Janssens, F. & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management*, 41(61), 1548-1572.
- Glenisson, P., Glanzel, W. & Persson, O. (2005). Combining full-text analysis and bibliometric indicators. A pilot study. *Scientometrics*, 63(1), 163-180.
- Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, 57(1), 27–57.
- Gómez Núñez, A.J. (2014). *Una aproximación multimetodológica para la clasificación de las revistas de Scimago Journal & Country Rank (SJR)*. (Tesis doctoral. Universidad de Granada, Granada). Recuperado de <http://digibug.ugr.es/bitstream/10481/39955/1/24589457.pdf>
- Gómez-Núñez, A.J., Batagelj, V., Vargas-Quesada, B., Moya-Anegón, F. de, Chinchilla-Rodríguez, Z. (2014). Optimizing SCImago Journal & Country Rank classification by community detection. *Journal of Informetrics*. 8(2), 369-383.
- González Guitián, M. V. & Zayas Pérez, M. R. (2012). Auditorías de conocimiento. Análisis de dominio en las bases de datos Scopus y WoK. *Revista Interamericana de Bibliotecología*, 35(1), 17-25.

- Gossler, A., Doetschman, T., Korn, R., Serflingt, E. & Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *PNAS- Proceedings of the National Academy of Science of the United States of America*, 83(23), 9065-9069.
- Griffith, B.C., Small, H., Stonehill, J. A. & Dey, S. (1974). The structure of scientific literature, II: toward a macro and microstructure for science. *Science Studies*, 4(4), 339-365.
- Guerrero Bote, V.P., Moya Anegón, F. & Herrero Solana, V. (2002). Document organization using Kohonen's algorithm. *Information Processing & Management*, 38(1), 79-89.
- Gutierrez Castanha, R.C., Cabrini Gracio, M.C. (2014). Bibliometrics Contribution to the Metatheoretical and Domain Analysis Studies. *Knowledge Organization*, 41(2), 171-174.
- Hamberg, K., Risberg, G. & Johansson, E.E. (2004). Male and female physicians show different patterns of gender bias. A paper-case study of management of irritable bowel syndrome. *Scandinavian Journal of Public Health*, 32(2) 144-152.
- Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R. & Vanhoutte, A. (1989). Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing and Management*, 25(3), 315–318.
- Hammond, J. (1949). Recovery and culture of tubal mouse ova. *Nature*, 163(4131) 28-29.
- He, Q. (1999). Knowledge Discovery through Co-Word Analysis. *Library Trends*, 48(1), 133–159.
- Hinze, S. (1994). Bibliographical cartography of an emerging interdisciplinary discipline: the case of bioelectronics. *Scientometrics*, 29(3), 353-376.
- Hjørland, B. (2002). Domain analysis in information science. Eleven approaches – traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.

- Hjørland, B. & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Hilbert, M. & López P (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60-65.
- Ho, Y.S. (2007). Bibliometric Analysis of Adsorption Technology in Environmental Science. *Journal of Environmental Protection Science*, 1, 1-11.
- Huai, C. & Chai, L. (2016). A bibliometric analysis on the performance and underlying dynamic patterns of water security research. *Scientometrics*, 108(3), 1531-1551.
- IBM Institute for Business Value (2009). *Business, Analytics and Optimization for the Intelligent Enterprise* [Informe]. Recuperado de <http://www-05.ibm.com/de/services/bao/pdf/gbe03211-usen-00.pdf>
- Inclusion of Women and Minorities in Clinical-Trials and the NIH-Revitalization-Act of 1993. The perspective of NIH Clinical Trialists. *Controlled Clinical Trials*, 16(5), 277-312.
- Ingwersen, P. & Larsen, B. (2001). Mapping national research profiles in social science disciplines. *Journal of Documentation*, 57(6), 715-740.
- Janssens, F., Leta, J., Glänzel, W., & De Moor, B., (2006). Towards mapping library and Information science. *Information Processing and Management*, 42(6), 1614-1642.
- Jones, W.P., & Furnas, G.W. (1987). Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science*, 38(6), 420-442.
- Jung, M. (2013). Bibliometric investigation on preventive medicine in North Korea: a coauthor and keyword network analysis. *The Health Care Manager*, 32(3), 253-9.
- Kajikawa, Y. & Takeda, Y. (2008). Structure of research on biomass and bio-fuels: a citation-based approach. *Technological Forecasting and Social Change*, 75(9), 1349-1359.

- Kajikawa, Y. & Takeda, Y. (2009). Citation Network Analysis of Organic LEDs, *Technological Forecasting and Social Change*, 76(8), 1115–1123.
- Kajikawa, Y., Yoshikawa, J. & Takeda, Y. & Matsushima K. (2008). Tracking emerging technologies in energy research: toward a roadmap for sustainable energy. *Technological Forecasting and Social Change*, 75(6), 771-782.
- Kamada, T. & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31, 7–15.
- Kastin, A., Rindflesch, T.C & Hristovski, D. (2014). Large-Scale Structure of a Network of Co-Occurring MeSH Terms: Statistical Analysis of Macroscopic Properties. *PLoS ONE*, 9(7), e102188.
- Kerrigan, D., Andrinopoulos, K., Chung, S., Glass, B. & Ellen, J. (2008). Gender Ideologies, Socioeconomic Opportunities, and HIV/STI-related Vulnerability among Female, African-American Adolescents. *Journal of Urban Health*, 85(5) 717-726.
- Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10-25.
- Kim, M.C. & Chen, C.A. (2015). A scientometric review of emerging trends and new developments in recommendation systems. *Scientometrics*, 104(1), 239-263.
- Klavans, R. & Boyack, K.W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology (JASIST)*, 57(2), 251-263.
- Klovdhal, A.S. (1981). A note of images of social networks. *Social Networks*, 3, 197-214.
- Kostoff, R.N., Butchtel, H.A., Andrews, J. & Pfiel, K.M. (2005). The hidden structure of neuripsychology: Text mining of the journal Cortex, 1991-2001. *Cortex*, 41(2), 103-15.
- Kozak, M., Bornmann, L., & Leydesdorff, L. (2015) How have the Eastern European countries of the former Warsaw Pact developed since 1990? A bibliometric study. *Scientometrics*, 102(2), 1101–1117.

- Krauskopf, D. (2000). La construcción social de la Ciencia y la Tecnología. *Boletín SEBBM*, 130, 12-16.
- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27.
- Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2), 115-129
- Kuhn T.S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Lee, W.H. (2008). How to identify emerging research fields using scientometrics: an example in the field of information security. *Scientometrics*, 76(3), 503-25.
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science (JASIST)*, 48(5), 418-427.
- Leydesdorff, L. (2008). On the normalization and visualization of author Cocitation data: Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77-85.
- Leydesdorff, L. (2011). Bibliometrics/Citation networks. En G.A. Barnett (Ed.), *The Encyclopedia of Social Networks* (p. 72-74), Universidad de California, Davis, (EE.UU.): SAGE Publications, 1.
- Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589-593.
- Leydesdorff, L., Hammarfelt, B., & Akdag Salah, A.A. (2011). The structure of the Arts & Humanities Citation Index: A mapping on the basis of aggregated citations among 1,157 journals. *Journal of the American Society for Information Science and Technology (JASIST)*, 62(12), 2414-2426.
- Leydesdorff, L. & Hellsten, L. (2006). Measuring the Meaning of Words in Contexts: An automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells'. *Scientometrics*, 67(2), 231-258.

- Leydesdorff, L., Khan, G.-F., & Bornmann, L. (2014). The generation of large networks from Web of Science data. *El Profesional de la Informacion (EPI)*, 23(6), 589-593.
- Leydesdorff, L., Kushnir, D. & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, 98(3), 1583-1599.
- Leydesdorff, L., de Moya-Anegón, F. & Guerrero-Bote, V.P. (2015). Journal Maps, Interactive Overlays, and the Measurement of Interdisciplinarity on the Basis of Scopus Data (1996-2012). *Journal of the Association for Information Science and Technology (JASIST)*, 66(5), 1001-1016.
- Leydesdorff, L., de Moya-Anegón, F. & de Nooy, W. (2016). Aggregated journal-journal citation relations in scopus and web of science matched and compared in terms of networks, maps, and interactive overlays. *Journal of the Association for Information Science and Technology (JASIST)*, 67(9), 2194-2211.
- Leydesdorff, L., Park, H.W., & Wagner, C. (2014). International coauthorship relations in the Social Sciences Citation Index: Is internationalization leading the network? *Journal of the Association for Information Science and Technology (JASIST)*, 65(10), 2111-2126.
- Leydesdorff, L. & Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology (JASIST)*, 60(2), 348-362.
- Leydesdorff, L. & Rafols, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Infometrics*, 6(2), 318-332.
- Leydesdorff, L., Rafols, I., Chen, C. (2013). Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal–journal citations. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(12), 2573–2586.

- Leydesdorff, L. & Vaughan, L. W. (2006). Co-occurrence matrices and their applications in information science: extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, 57(12), 1616-1628.
- Leydesdorff, L. & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5(3), 469-75.
- Li, T., Ho, Y.S. & Li, C.Y. (2008). Bibliometric analysis on global Parkinson's disease research trends during 1991–2006. *Neuroscience Letters*, 441(3), 248-252.
- Lin, X., Soergel, D. & Marchionini, G. (1991). A self-organizing semantic map for information retrieval. En *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, (pp. 262-269), Chicago.
- Lin, X., White, H.D. & Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. *Information Processing & Management*, 39(5), 689-706.
- Liu, F.L., Lin, A.W., Wang, H.H., Peng, Y.L. & Hong, S. (2015). Global research trends in landslides during 1991-2014: a bibliometric analysis. *Landslides*, 12(6), 1215-1226.
- Looze, M.A. De & Lemarié, J. (1997). Corpus relevance through co-word analysis: An application to plant proteints. *Scientometrics*, 39(3), 267-280.
- López Ferrer, M., Velasco Arroyo, E., Osla Lluçh, M.J. & Peñaranda Ortega, M. (2009). Aplicación de las redes sociales a la delimitación de áreas temáticas en bases de datos multidisciplinares. En *Nuevas perspectivas para la difusión y organización del conocimiento: actas del congreso / coord. por Nuria Lloret Romero, Vol. 2*, (pp. 803-817).
- Lozares, C. (1996). La teoría de redes sociales. *Papers*, 48, 103-126.
- Lu, S.J., Feng, Q., Caballero, S., Chen, Y., Moore, M.A.S., Subvención, M.B. & Lanza, R. (2007). Generation of functional hemangioblasts from human embryonic stem cells. *Nature Methods*, 4(6), 501-509.

- Lu, D., Mahmood, A., Wang, L., Li, Y., Lu, M. & Chopp, M. (2001). Adult bone marrow stromal cells administered intravenously to rats after traumatic brain injury migrate into brain and improve neurological outcome. *Neuroreport*, 12(3), 559-563.
- Luukkonen, T., Tijssen, R.J.W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36.
- Lv, P.H., Wang, G., Wan, Y., Liu, J., Liu, Q. & Ma, F.C. (2011). Bibliometric trend analysis on global graphene research. *Scientometrics*, 88(2), 399-419.
- Ma, Y.K. (2006) The inheritance and development of social constructionism psychology to Vygotsky's ideas. *Advances in Psychological Science*, 14(1), 154-160.
- Maltrás, B. (2003). Los indicadores bibliométricos: fundamentos y aplicación al análisis de la ciencia. Asturias: Trea.
- Marshakova, I.V. (1973). System of document connection based on references. *Nauchno-Tekhnicheskaya Informatsiya*, Series II, (6), 3-8.
- Marshakova-Shaikevich, I. (2005). Bibliometric maps of field of science. *Information Processing & Management*, 41(6), 1534-1547.
- Martens, B.V. & Van Fleet, C. (2012). Opening the black box of "relevance work": A domain analysis. *Journal of the American Society for Information Science and Technology (JASIST)*. 63(5), 936-947.
- Marx, W. & Bornmann, L. (2013). The emergence of plate tectonics and the Kuhnian model of paradigm shift: a bibliometric case study based on the Anna Karenina principle. *Scientometrics*, 94(2), 595-614.
- McGrath, J. & Solter, D. (1983). Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. *Science*, 220(4603), 1300-1302.
- McKay, R. (1997). Stem cells in the central nervous system. *Science*, 276(5309), 66-71.
- Merkatz, R. B. & Junod, S. W. (1994). Historical background of changes in FDA policy on the study an evaluation of drugs in women. *Academic Medicine*, 69(9), 703-707.

- Merton, R.K. (2000). On the Garfield input to the sociology of science: a retrospective collage. En B. Cronin & H.B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield* (p. 435-448). New Jersey: Information Today.
- Milojevic, S., Sugimoto, C., Yan, E. & Ding, Y. (2011). The Cognitive Structure of Library and Information Science: Analysis of Article Title Words. *Journal of the American Society for Information Science and Technology (JASIST)*, 62(10), 1933-1953.
- Montgomery, C.H., Sherif, K. (2000). The Information Problem in Women's Health: a Piece of the Solution. *Journal of Womens Health & Gender-Based Medicine*. 9(5), 529-36.
- Moreno, J.L. (1934). *Who shall survive?* Washington, DC: Nervous and Mental Disease Publishing Company.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z. Corera-Alvarez, E., Gonzalez-Molina, A., Muñoz-Fernández, F.J. & Herrero-Solana, V. (2006). Visualizing and analyzing the Spanish science structure: ISI Web of science 1990-2005. *El Profesional de la Información (EPI)*, 15(4), 258-269.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z. Corera-Alvarez, E., Gonzalez-Molina, A., Muñoz-Fernández, F.J. & Herrero-Solana, V. (2007). Visualizing the Marrow of Science. *Journal of The American Society for Information Science and Technology (JASIST)*, 58(14), 2167-2179.
- Moya Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E. & Munoz-Fernández, F.J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61(1), 129-145.
- Morris, S.A., Yen, G., Wu, Z. & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5), 413-422.
- Nathanson, C.A., (1975). Illness and the feminine role: a theoretical review. *Social Science and Medicine*, 9(2), 57-62.

- Nathanson, C.A., (1977). Sex, illness and medical care: a review of data, theory and method. *Social Science and Medicine*. 11(1), 13–25.
- Neff, M.W. & Corley, E.A. (2009). 35 years and 160.000 articles: a bibliometric exploration of the evolution of ecology. *Scientometrics*, 80(3), 657-82.
- Newman, M.E.J. & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.
- Niazi, M.A. (2016). CiteSpace: A Practical Guide for Mapping Scientific Literature. *Complex Adaptive Systems Modeling*, 4(23).
- Nobelius, A. M. & Wainer, J. (2004). Gender and medicine: A conceptual guide for medical educators. Victoria: Monash University School of Rural Health. Recuperado de <http://www.med.monash.edu.au/gendermed/docs/A4TutorManual120504screen.pdf>
- Noyons, E.C.M. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, 50(1), 83-98.
- Noyons, E.C.M., Moed, H.F. & Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study. *Journal of the American Society for Information Science (JASIS)*, 50(2), 115-131.
- Noyons, E.C.M. & Van Raan, A.F.J. (1998a). Bibliometric cartography of scientific and technological developments of an R & D field. *Scientometrics*, 30(1), 157-173.
- Noyons, E.C.M. & Van Raan, A.F.J. (1998b). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science (JASIS)*, 49(1), 68–81.
- NWB Team. (2006). Network Workbench Tool. Indiana : Indiana University, Northeastern University, and University of Michigan. Recuperado de <http://nwb.slis.indiana.edu>
- Olmeda-Gómez, C. (2014). Visualización de información. *El Profesional de la Información (EPI)*, 23(3), 213-219.

- Onyancha, O.B. & Ocholla, D. (2005). Informetric investigation of the relatedness of opportunistic infections to HIV/AIDS. *Information Processing & Management*, 41(6), 1573-88.
- Peters, H.P.F., & Van Raan, A.F.J. (1993). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy*, 22(1), 23–45.
- Pinto, M. (2015). Viewing and exploring the subject area of information literacy assessment in higher education (2000-2011), *Scientometrics*, 102(1), 227-245.
- Porter, A.L. & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3), 719-745.
- Price, D. d. S. (1963). *Little Science, Big Science*. New York: Columbia University Press.
- Price, D. d. S. (1965). Networks of Scientific Papers. *Science*, 149(3683), 510–515.
- Pritchard, A. (1969). Statistical bibliography or Bibliometrics. *Journal of Documentation*, 25(4), 348-369.
- Quirin, A., Cordon, O., Guerrero-Bote, V.P., Vargas-Quesada, B. & Moya-Anegón, F. (2008). A quick MST-based algorithm to obtain pathfinder networks (∞ , $n - 1$). *Journal of the American Society for Information Science & Technology*, 59(12), 1912-1924.
- Quirin, A., Cordon, O., Vargas-Quesada, B. & Moya-Anegón, F. (2010). Graph-based data mining: A new tool for the analysis and comparison of scientific domains represented as scientograms. *Journal of Informetrics*, 4(3), 291-312.
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, 41(7), 1262-1282.
- Rafols, I. & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.

- Rafols, I., Porter, A.L. & Leydesdorff, L. (2010). Science Overlay Maps: A New Tool for Research Policy and Library Management. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(9), 1871-1887.
- Rikken, F., Kiers H.A.L. & Vos, R. (1995). Mapping the dynamics of adverse drug reactions in subsequent time periods using INDSCAL. *Scientometrics*, 33(3), 367-380.
- Rip, A. & Courtial, J.P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381-400.
- Romo-Fernández, L.M., Guerrero-Bote, V.P., & Moya-Anegón, F. (2013). Co-word based thematic analysis of renewable energy (1990-2010). *Scientometrics*, 97(3), 743-765.
- Rorvig, M. (1999). Images of similarity: A visual exploration of optimal similarity metrics and scaling properties of TREC topic-document sets. *Journal of the American Society for Information Science*, 50(8), 639–651.
- Scott, J. (1991). *Social Network Analysis*. Newbury Park, Londres: Sage, 1-38.
- Scheffler, B., Horn, M., Blumcke, I., Lywell, E.E., Coones, D., Kukekov, V.G. & Steindler, D.A. (1999). Marrow- mindedness: A perspective on neurogenesis. *Trends Neurosci*, 22(8), 348-56.
- Schneider, J.W. & Borlund, P. (2007a). Matrix comparison, part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. *Journal of the American Society for Information Science and Technology*, 58(11), 1586–1595.
- Schneider, J.W. & Borlund, P. (2007b). Matrix comparison, part 2: Measuring the resemblance between proximity measures or ordination results by use of the Mantel and Procrustes statistics. *Journal of the American Society for Information Science and Technology*, 58(11), 1596–1609.
- Schvaneveldt, R.W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood: ALEX.

- Schvaneveldt, R.W., Dearholt, D.W. & Durso, F.T. (1988). Graph theoretic foundations of Pathfinder Networks. *Computers and Mathematics with Applications*, 15(4), 337-345.
- Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of Informetrics). *Library Review*, 65(1/2), 52–64.
- Seiden, L.S. & Swanson, D.R. (1989). ISI Atlas of Science: Pharmacology 1987, Vol 1 -Inst-SCI-Informat. *Library Quarterly*, 59(1), 72-73.
- Sen, G., George, A. & Östlin, P. (2005). *Incorporación de la perspectiva de género en la investigación y en las políticas en salud: una perspectiva de género*. Harvard: Harvard Center for Population and Development Studies.
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science (JASIS)*, 24(4), 265-269.
- Small, H. (1981). The relationship of information science to the social sciences: a co-citation analysis. *Information Processing & Management*, 17(1), 39-50.
- Small, H. (1993). Macrolevel changes in the structure of cocitation clusters: 1983-1989. *Scientometrics*, 26(1), 5-20.
- Small, H. (1994). A SCI-MAP case-study: building a map of AIDS research. *Scientometrics*, 30(1), 229-241.
- Small, H. (1997). Update on science mapping: creating large document spaces. *Scientometrics*, 38(2), 275-293.
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science (JASIS)*, 50(9), 799-813.
- Small, H. (2000). Charting pathways through science: exploring Garfield's visión of a unified index to science. En B. Cronin & H.B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield* (p. 449-473). New Jersey: Information Today.

- Small, H. (2003). Paradigms, citations, and maps of science: A personal history. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5), 394-399.
- Small, H. (2010). Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy. *Scientometrics*, 83(3), 835-849.
- Small, H. (2011). Interpreting maps of science using citation context sentiments: a preliminary investigation. *Scientometrics*, 87(2), 373-388.
- Small, H. & Garfield, E. (1985). The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11(4), 147-159.
- Small, H. & Griffith, B.C. (1974). The structure of scientific literature, I: identifying and graphing specialities. *Science Studies*, 4(1), 17-40.
- Small, H. & Sweeney, E. (1985). Clustering the science citation index using co-citations. 1. A comparison of methods. *Scientometrics*, 7(3-6), 391-409.
- Small, H., Sweeney, E. & Greenlee, E. (1985). Clustering the science citation index using co-citations. 2. Mapping science. *Scientometrics*, 8(5-6), 321-340.
- Small, H. & Upham, P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79(2), 365-375.
- Smiraglia, R. (2012). Knowledge organization: some trends in an emergent domain. *El Profesional de la Informacion*, 21(3), 225-227.
- Smiraglia, R. (2015). *Domain Analysis for Knowledge Organization, Tools for Ontology Extraction*. Kidlington (Reino Unido): Chandos Publishing.
- Soos, S. Kamps, G. & Gulyás, L. (2013). Large-scale temporal analysis of computer and information science. *The European physical journal special topics*, 222(6), 1441-1465.
- Stapley, B.J. & Benoit, G. (2000). Bibliometrics information retrieval and visualization from co-occurrences of gene names in Medline abstracts. *Pacific Symp Biocomput*, 5, 526-537.

- Su, H.N. & Lee P.C. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, 85(1), 65-79
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5), 861-872.
- Takahashi, K. & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), 663-676.
- Tian, Y., Wen, C. & Hong, S. (2008). Global scientific production on GIS research by bibliometric analysis from 1997 to 2006. *Journal of Informetrics*, 2(1), 65-74.
- Tennis, J.T. (2003). Two Axes of Domain Analysis. *Knowledge Organization*, 30(3/4), 191-195.
- Thomson Reuters (2010). Web of Science. Recuperado de <http://wokinfo.com/>
- Tijssen, R.J.W. (2010). Discarding the 'basic science/applied science' dichotomy: A knowledge utilization triangle classification system of research journals. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(9), 1842-1852.
- Tufte, E.R. (1990). *Envisioning information*. Cheshire: Graphic Press.
- Uddin, S. & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics*, 10(4), 1166-1177.
- Van Eck, N. J. & Waltman, L. (2007a). VOS: A new method for visualizing similarities between objects. In H.-J. Lenz & R. Decker (Eds.), *Advances in data analysis: Proceedings of the 30th annual conference of the German Classification Society* (pp. 299–306). Heidelberg: Springer.
- Van Eck, N. J. & Waltman, L. (2007b). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5), 625–645.

- Van Eck, N.J. & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635-1651.
- Van Eck, N. J. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.
- Van Eck, N.J., Waltman, L., Dekker, R., & Van den Berg, J. (2008). *An experimental comparison of bibliometric mapping techniques*. Paper presented at the 10th International Conference on Science and Technology Indications, Vienna.
- Van Eck, N. J., Waltman, L., Dekker, R. & van den Berg, J. (2010). A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS. *Journal of the American Society for Information Science and Technology (JASIST)*, 61(12), 2405-2416.
- Van Eck, N.J., Waltman, L & Noyons, E.C.M. (2010). A unified approach to mapping of bibliometric networks. *Journal of Infometrics*, 4(4), 629-635.
- Van Eck, N. J., Waltman, L., Van den Berg, J. & Kaymak, U. (2006). Visualizing the computational intelligence field. *IEEE Computational Intelligence Magazine*, 1(4), 6–10.
- Van Raan, A.F.J., Tijssen, R.J.W. (1993). The Neuronal net of neuronal network research. An exercise in bibliometric mapping. *Scientometrics*, 26(1), 169-192.
- Vargas-Quesada, B. (2005). *Visualización y análisis de grandes dominios científicos mediante redes pathfinder (pfnet)*. (Tesis Doctoral. Universidad de Granada, Granada). Recuperado de <http://www.ugr.es/~benjamin/Tesis-Benjamin.pdf>
- Vargas-Quesada, B., Chinchilla-Rodríguez, Z., González-Molina, A. & Moya-Anegón, F. (2010). Showing the Essential Science Structure of a Scientific Domain and its Evolution. *Information Visualization*, 9(4), 288-300.
- Vargas-Quesada, B., Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E. & Guerrero-Bote, V. (2008). Evolución de la estructura científica española: ISI Web of Science 1990-2005. *El Profesional de la Información*, 7(1), 22-37.

- Vega Almeida, R.L., Fernández Molina, J.C., Moya Anegón, F. (2011). The bibliometric approach to identify paradigms in knowledge domains. *Acimed*, 22(3), 251-261.
- Verbrugge, L.M., (1978). Females and illness: recent trends in sex differences in the United States. *Journal of Health and Social Behavior* 17(4), 387–403.
- Verbrugge, L.M., (1985). Gender and health: an update of hypotheses. *Journal of Health and Social Behavior*. 26(3), 156–182.
- Verbrugge, L.M., (1989). The twain meet: empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior* 30(3), 282–304.
- Vivian, W. Pinn, MD. (2003). Sex and Gender Factors in Medical Studies. Implications for Health and Clinical Practice. *Journal of the American Medical Association*, 289(4), 397-400.
- Vivian, W., Pinn, MD. (2005). Research on Women`s Health. Progress and Opportunities. *Journal of the American Medical Association*, 294(11), 1407-1410.
- Voltarelli, J.C., Martinez E. & Burt R.K. (2009). Autologous Nonmyeloablative Hematopoietic Stem Cell Transplantation in Newly Diagnosed Type 1 Diabetes Mellitus Reply. *JAMA- Journal of the American Medical Association*, 302(6), 624-625.
- Waldron, I., (1978). Why do women live longer than men? *Social Science and Medicine*, 10(7-8), 349–362.
- Wang, L. (2013). Cultural-historical activity theory and domain analysis: metatheoretical implications for information science. Proceedings of the Eighth International Conference on Conceptions of Library and Information Science, Copenhagen, Denmark, 19-22 August, 2013. *Information Research*, 18(3), suppl Sep 2013.
- Wang, F. & Wolfram, D. (2015). Assessment of Journal Similarity Based on Citing Discipline Analysis. *Journal of the Association for Information Science and Technology (JASIST)*, 66(6), 1189-1198.
- Waltman, L., van Eck, N.J. & Noyons, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Infometrics*, 4(4), 629-635.

- Wasserman, S. & Faust, K. (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.
- Wei, F., Grubestic, T.H. & Bishop, B.W. (2015). Exploring the GIS Knowledge Domain Using CiteSpace. *El Profesional de la Información (EPI)*, 67(3), 374-384.
- White, H.D. (2000). Toward ego-centered citation analysis. En B. Cronin & H.B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield*. New Jersey: Information Today.
- White, H.D. (2003). Pathfinder networks and author cocitation analysis: a remapping of paradigmatic information scientist. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5), 423-434.
- White, H. (2015). Co-cited author retrieval and relevance theory: examples from the humanities. *Scientometrics*, 102(3), 2275-2299.
- White, H.D., Buzydlowski, J. & Lin, X. (2000). Co-cited author maps as interfaces to digital libraries: designing Pathfinder Networks in the humanities. En *IEEE International Conference on information visualization* (p. 25-30), London.
- White, H.D. & Griffith B.C. (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163-171.
- White, H.D., Lin, X. & McCain, K.W. (1998). Two modes of automated domain analysis: multidimensional scaling vs. Kohonen feature mapping of information science authors. En *Proceedings of the Fifth International ISKO Conference* (pp. 57-61). Würzburg: Ergon Verlag.
- White, H.D. & McCain, K.W. (1981). Author cocitation: a literature measure of intellectual structure. *Journal of the American Society for Information Science (JASIS)*, 32(3), 163-171.
- White, H.D. & McCain, K.W. (1997). Visualization of literatures. *Annual Review of Information Systems and Technology (ARIST)*, 32, 99-168.

- White, H.D. & McCain, K.W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science (JASIS)*, 49(4), 327-355.
- Whithead, A. N., & Russell, B. (1962). *Principia Mathematica*. Cambridge: University Press.
- Wilmot I., Schnieke, A.E., McWhir, J., Kind, A.J. & Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature*, 395(6619), 810-813.
- Wittgenstein, L. (1921): *Tractatus Logico-Philosophicus*. Routledge & Regan Paul, 1922 (ed. bilingüe inglés-alemán); primera aparición en alemán: «Logisch-Philosophische Abhandlung», en *Annalen der Naturphilosophie*. Traducción al castellano en Alianza Universidad, Madrid, 1987.
- Wobus, A.M., H. Jakel, P. & Schoneich, J. (1984). Characterization of a pluripotent stem line derived from a mouse embryo. *Experimental Cell Research*, 152(1), 212-219.
- Wolfram, D. & Zhao, Y. (2014). A comparison of journal similarity across six disciplines using citing discipline analysis. *Journal of Informetrics*, 8(4), 840-853.
- Women's Health Bureau. (2002). *Exploring concepts of gender and health*. Ottawa: Health Canada.
- World Health Organization (WHO). (1998). *Gender and Health: A Technical Paper*. Ginebra.
- World Health Organization (WHO). (2002). *Integrating gender perspectives in the work of WHO: WHO Gender Policy*. Recuperado de <http://www.who.int/gender/documents/sppolicy.pdf>
- Wren, J.D. & Garner, H.R. (2004). Shared relationship analysis: ranking set cohesion and commonalities within a literature-derived relationship network. *Bioinformatics*, 20(2), 191-198.
- Xie, P. (2015). Study of international anticancer research trends via co-word and document co-citation visualization analysis. *Scientometrics*, 105(1), 611-622.

- Yang, S., Han, R., Wolfram, D. & Zhao, Y.H. (2016). Visualizing the intellectual structure of information science (2006-2015): Introducing author keyword coupling analysis. *Journal of Informetrics*, 10(1), 132-150.
- Yang, L., Soonpaa, M.H., Adler, E.D., Roepke, T.K., Kattman, S.J., Steven, J., Kennedy, M., Henckaerts, E., Bonham, K., Abbott, G.W., Linden, R.M., Field, L.J. & Keller, G.M. (2008). Human cardiovascular progenitor cells develop from a KDR plus embryonic-stem-cell-derived population. *Nature*, 453(7194), 524-U6.
- Yao, Q., Chen, J., Lyu, P., Zhang, S.J., Ma, F.C. & Fang, J.G. (2012). Knowledge map of artemisinin research in SCI and Medline database. *Journal of Vector Borne Diseases*, 49(4), 205-216.
- Yoo, Y.J.; Lee, J.Y. & Choi, S. (2013). Intellectual structure of Korean theology 2000-2008: Presbyterian theological journals. *Journal of Information Science.*, 39(3), 307-318.
- Zhao, D. & Strotmann, A. (2011). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, 87(1), 115-131.
- Zhang, J., Xie, J., Hou, W., Tu, X.C., Xu, J., Song, F.J., Wang, Z.H. & Lu, Z.X. (2012). Mapping the Knowledge Structure of Research on Patient Adherence: Knowledge Domain Visualization Based Co-Word Analysis and Social Network Analysis. *PLoS ONE*, 7(4), e34497.
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu Z. & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology (JASIST)*, 67(4), 967-972.
- Zhao, R., & Chen, B. (2014). Applying author co-citation analysis to user interaction analysis: A case study on instant messaging groups. *Scientometrics*, 101(2), 985-997.
- Zhao, D. & Strotmann, A. (2011). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, 87(1), 115-31.

- Zitt, M. & Bassecouard, E. (1996). Reassessment of co-citation methods for science indicators: effect of methods improving recall rates. *Scientometrics*, 37(2), 223-44.
- Zitt, M., Bassecouard, E. & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627-657.
- Zitt, M. & Bassecouard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: an application to nanosciences. *Information Processing & Management*. 42(6), 1513-31.
- Zitt, M., Lelu, A. & Bassecouard, E. (2008). Hybrid maps of scientific fields (terms and citations): an application to nanosciences. Paper presented at the Tenth *International Conference on Science & Technology Indicators: Excellence and Emergence – A New Challenge for the Combination of Quantitative and Qualitative Approaches*, Vienna.
- Zitt, M., Lelu, A. & Bassecouard, E. (2011). Hybrid citation-word representations in science mapping: Portolan charts or research fields? *Journal of the American Society for Information Science and Technology (JASIST)*, 62(1), 19-39.

PARTE II: PUBLICACIONES CIENTÍFICAS

PUBLICACIÓN Nº 1

Zulueta, M. Ángeles, Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Sánchez, Carmen (2011). Research involving women and health in the Medline database, 1965-2005: Co-term analysis and visualization of main lines of research, *Scientometrics*, 88(3), 679-706. <http://dx.doi.org/10.1007/s11192-011-0455-1>

Research involving women and health in the Medline database, 1965–2005: co-term analysis and visualization of main lines of research

M. Angeles Zulueta · Gisela Cantos-Mateos · Benjamín Vargas-Quesada · Carmen Sánchez

Received: 4 February 2010
© Akadémiai Kiadó, Budapest, Hungary 2011

Abstract This paper shows the main lines of research concerning health and women, as registered in the Medline database, broken down into four 10-year periods: 1965–1974, 1975–1984, 1985–1994, and 1995–2005. The units of analysis used were the Medline “MeSH” major terms, processed by means of co-term analysis. For graphic representation, the social network approach was used, with pruning performed by Pathfinder Networks (PFNET), so as to concentrate the displays. Factor analysis was used to group the descriptors and identify the main lines of research involving health and women. The results show that research on *Health and Women* has increased and undergone significant changes over the past 40 years, yet such studies are not given due importance.

Keywords Medline · Co-term · Health · Women · Research · Social networks · PFNET · Maps

Introduction

Research on health and women has undergone a vast transformation over the past few decades. It was not until the 1960s when fundamental social and clinical differences between men and women were first acknowledged and became the focus of study. Since then, interest in how gender plays a role in health has grown to the point where it is accepted as a determining factor in the well-being of citizens.

M. A. Zulueta · G. Cantos-Mateos · C. Sánchez
Information Science Faculty, University of Alcalá, C/ San Cirilo s/n, 28804 Alcalá de Henares, Madrid, Spain

M. A. Zulueta · B. Vargas-Quesada
CSIC, Unidad Asociada Grupo SCImago, Madrid, Spain

B. Vargas-Quesada (✉)
Communication and Information Science Faculty, University of Granada,
Colegio Máximo de Cartuja s/n, 18071 Granada, Spain
e-mail: benjamin@ugr.es

Linked with this transformation we have witnessed a substantial change in the perception of health per se. Whereas a healthy state used to be equated to the absence of illness or disease, health is now related with an overall state of physical, mental and social well-being (WHO 2006). In addition to strictly biological factors, the impact of environment, lifestyle and the organization of health care are now acknowledged as well (Lalonde 1974). This new perspective gives rise to the consideration of aspects such as sex or gender within medical research.

It was not always so. The first studies to focus on woman as the topic of medical research were largely related to reproduction (Vivian and Pinn 2005, 2006). That is, their presence was limited to medical specialties such as obstetrics and gynecology, and gender was hardly considered when dealing with diseases that affect men and women indistinctly (WHO 1998; Sen et al. 2005). Clinical trials excluded pregnant or fertile women for fear of possible damage to the fetus or the reproductive system (Berlin and Ellenberg 2009).

In the 1970s, however, research began to publish studies of the different levels of mortality and morbidity between the two sexes (Nathanson 1975, 1977; Verbrugge 1978; Waldron 1978), determining that men have a higher mortality but lower morbidity than women (Singh-Manoux et al. 2008). This notion was still under debate throughout the 80s (Wingard 1984; Verbrugge 1985, 1989; Clarke 1983), and despite acknowledgement of minor differences regarding health (Leviatan and Cohen 1985), focus on women continued to be scarce in most research efforts and clinical studies.

Then, in the mid 80s, the need to consider women within biomedical research and clinical trials became a more central issue. This came largely as a result of the development of HIV/AIDS. Many women were not allowed access to new medications due to the lack of clinical trials assessing potential risks in the context of child-bearing (Gowen 2009). In the 90s, specific initiatives were finally undertaken. For instance, the Food and Drug Administration (FDA) and the National Institutes of Health (NIH) set forth measures to include women in clinical trials (1995; Merkatz and Junod 1994), creating the Office of Research on Women's Health (ORWH) in 1990 to control and ensure research on health and women. Other health-related institutions followed the lead, introducing changes in their health policy and elaborating instruments to urge countries to introduce the gender perspective in medical research and public health services (Gender and Health Group 2000; Women's Health Bureau 2002; WHO 2002; Nobelius and Wainer 2004; Sen et al. 2005).

In short, the perspective of gender or sex in health over the decades has led to advances in the consideration of women with HIV infection, irritable bowel syndrome, lung cancer, heart disease, and other realms of study (Vivian and Pinn 2003; Hamberg et al. 2004; Devesa et al. 2005; Blake et al. 2005; Daly et al. 2006; Kerrigan et al. 2008). The use of the terms "gender" and "sex" in medicine is more and more generalized; yet despite efforts such as that of the Institute of Medicine (Institute of Medicine 2001) to foment their unified use by scientists, there is still considerable confusion regarding how each term is to be used. *Sex* is commonly understood to represent the physical, anatomical and physiological differences between men and women. *Gender*, in turn, refers to the social construct based on cultural conventions, attitudes and relations between men and women. Both have implications for health, as personal identity is a matter of both gender and sex at the same time (Borrell and Artazcoz 2008).

Related works

The publication of studies that consider the female perspective has evolved considerably in a very short time, generating a great deal of information and new concepts. Researchers

and health professionals in general acknowledge the fact that they have difficulty assimilating the new lines of research and information.

Studies of co-words may prove very adequate in resolving this problem. This type of study, forged since the 1980s (Callon et al. 1983), stems from the notion that the co-occurrence of two terms in a document signals a key relation between the concepts that they represent (Cambrosio et al. 1993; Courtial et al. 1994), providing insight as to the structure and trends of a scientific discipline, by virtue of the strength of associations among the representative terms in the literature (Ding et al. 2001).

Van Raan and Tijssen (1993), applied this notion to research in neural networks, concluding that the epistemological value of co-words allows one to discover unsuspected relationships among the concepts of a discipline. Moreover, it can reveal problems that would have otherwise gone undetected. Braam et al. (1991a, b) combine co-citation and word analysis to map the structure and dynamics of scientific research to improve the capability of quantitative techniques. Onyancha and Ocholla (2005), resorted to co-word analysis to help researchers and policy-makers identify new lines of research surrounding HIV/AIDS, as well as the links established to plan research and formulate adequate policies. The most recent works have focused on the identification and visualization of research fronts (Morris et al. 2003).

The objective of the present paper resides in the schematic identification and representation of the main lines of research that have been the focus of studies surrounding women and health over the past 40 years. Bibliometric, statistical and visualization techniques are used to recreate the underlying thematic structure by decades. This allows us to corroborate whether the main lines of research detected correspond with the actual changes in the literature described in the Introduction. A coarse-grained study such as this, whose thematic and temporal coverage are so broad, calls for a methodology capable of representing, in a clear, concise and systematic way, all the essential information gathered up in scientific output. We put this methodology to the test to see if it adequately detects and represents domains with a longitudinal trend, where multidisciplinary can be considered a strong point rather than an obstacle.

Materials and methods

Data source

We chose Medline as the informational source. This database contains some 16 million bibliographic references and the abstracts of over 5,200 biomedical journals. One of its main features, making it the ideal source for our approach, is that it utilizes an indexing system based on MeSH, which lends it a high level of precision and comprehensiveness. It consists of some 25,000 descriptors structured hierarchically into 16 categories, covering all health-related subjects. This is crucial for our work, as we shall see below.

Extraction of data

This process involved the following stages:

Stage one: In order to obtain the most representative terms in research about health and women, we carried out an initial search of the term “women” in DeCS (*Descriptores en Ciencias de la Salud*) (1967), which led us to the most appropriate MeSH Terms.

The descriptors obtained were *Women*; *Women's health*; *Women's Health Services*; *Women's rights*; *Battered Women*; *Women, Working*; *Dentists, Women*; *Physicians, Women* and *Pregnant Women*. Then, to retrieve the most relevant documents, the descriptors were sought in the *MeSH Major Topic* field, in the Medline database by means of WebSpirs 5.03 (2005), using the check tags *Human* and *Female*. This initial search yielded 19,524 papers. *Stage two*: The aim was to extract and clearly delimit the object of study, as well as any corollary or “family words” (Van Raan and Tijssen 1993). The sixty most common descriptors from the aforementioned 19,524 documents retrieved in stage one were taken, and then launched once again against the Medline database. The result was a total of 365,714 documents represented by 13,865 descriptors. By extending the search beyond the 9 initial DeCS descriptors, it was possible to determine the main lines of research meriting the interest of scientists, as expressed in their articles and synthesized by the Medline indexers through descriptors.

Data processing

Unit of analysis

Obviously, careful choice of terms is a crucial point in work based on word analysis. The terms obtained through the title or abstract or from the free text are often not specific enough and vary from one context to another (Leydesdorff 1997). The use of a controlled language lends the advantage of avoiding the traps of natural language processing, while it has the shortcoming of offering a coarse-grained view with little insight. In our case, due to the nature, amplitude and type of maps used in this study, we resorted to descriptors to carry out the analysis. The visualization of the relations existing among them (Chen 2002) stands as a very adequate point of departure from which to arrive at the contents of the main and/or different lines of research within a subject area. To achieve the greatest precision possible, only major descriptors were used, eliminating qualifiers and geographic descriptors.

Units of measure

Different similarity measures applied to word analysis over the past 30 years in order to construct maps would include raw co-occurrence counts, Pearson r correlations, cosine, Jaccard, etc. We apply co-occurrence values between pairs of descriptors, and then form a raw count matrix as the measure of similarity between them. Though not well suited in studies of the intellectual structure, raw co-occurrence can be combined with dimension reduction techniques (Pathfinder) and layout algorithms of the spring embedder type to identify and highlight the main lines of research, as authors White (2003) or Moya-Anegón et al. (2007) have demonstrated.

Dimensionality reduction

A raw co-occurrence count reflects how many times any two terms used in a specialized text appear together in posterior works. The Pathfinder Networks (Schvaneveldt 1990) deals with these recounts as weights on a path drawn between nodes that represent any two terms in the input set. If all possible term-to-term paths are drawn, the sums of the weights on them will vary. The Pathfinder Networks (PFNET) algorithm seeks minimum weight paths between nodes. Two parameters must be set: r defines the Minkowski metrics used to

compute the distance of a path, while q limits the scope of the minimum-weight paths that will ultimately be drawn (White 2003). In our opinion, PFNET, with pruning parameters $r = \infty$, and $q = n - 1$ (Vargas-Quesada et al. 2010), serves to eliminate the “superfluous” relations between descriptors and preserves the salient ones, thereby providing a new matrix for schematic displays that are easy to interpret.

Scalar

There are many different methods for the automatic generation of graphs. The spring embedder type is widely used in the area of Documentation. Spring embedders assign coordinates to the nodes in such a way that the final graph will be pleasing to the eye. One noteworthy example is the Kamada and Kawai (1989) algorithm, featuring the capacity to minimize differences with respect to theoretical distances in the entire graph, good computation times, and the fact that it subsumes multidimensional scaling. As Krempel (1999) indicates, the Kamada and Kawai algorithm relies on an energy similar to the *stress* of multidimensional scaling as the measure for adaptation to theoretical distances.

A term must have a relatively high occurrence overall if it is to be highly *co-occurred* with others and perhaps linked to more than one other term. The terms with many links to others tend to situate themselves toward the center of the display, as a result of the greater number of links, and they group themselves into bunches that will later be identified through factor analysis (FA) as the main lines of research. The great advantage of raw-count PFNET and Kamada and Kawai is that, in one step, the main lines of research of a discipline are drawn into bunches and connected by the dominant terms. This facilitates analysis and interpretation.

Graph validation

A method based on a statistical process—Factor Analysis, or FA—was finally used to validate our findings. FA is conducted on the basis of raw data co-occurrence. The number of factors identified is extracted, each factor is tagged, and the factors identified are transferred to a graph. This multivariate exploratory technique can be used to examine a wide array of data sets. Its main practical application is to reduce the number of variables, and detect a structure by means of relationships and classification (Benzecri 1976; Benzecri et al. 1981; White and McCain 1998).

In the context of our study, FA (Principal Component Analysis with a varimax rotation) was conducted on each one of the co-occurrence matrixes to identify salient dimensions that would explain data variance. In Information Science, these dimensions correspond to groupings of research lines and subfields of scientific disciplines, i.e., specialities (Chen et al. 2001). The criterion adopted to stop extracting factors was the detection of a brusque change in the decline of the factors' eigenvalue. This simple criterion has proved very adequate in previous research (Leydesdorff and Rafols 2009). To capture the nature of each factor so as to tag it, we followed the methodology proposed by Moya-Anegón et al. (2007), that is: the descriptors contained in each factor by their weighting index *factor loading* were put in descending order, and the cut-off for being included in a factor was established as a value equal to or greater than 0.5. Then, in tagging each factor, we took into account only those descriptors of each factor having a value greater than or equal to 0.7. In this way, the tagging of each factor is consistent with the more representative descriptors.

The results obtained using FA were transferred to the maps, and the descriptors belonging to a specific research line were designated with pre-established symbols and

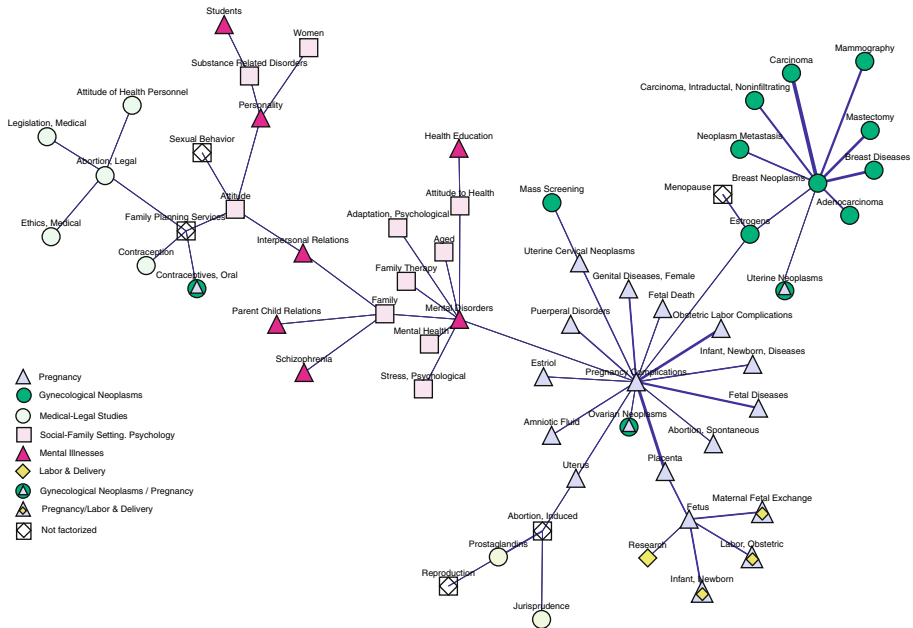


Fig. 1 Map of the 60 most frequent descriptors for the period 1965–1974

colours. Thus, for instance in Fig. 1, the descriptors identified in Factor 1 (*Pregnancy*), appear as blue triangles, while those of Factor 2 (*Gynecological Neoplasms*) are green circles, etc. When descriptors belong to more than one line of research, the symbols and colours are integrated and can be perceived as “hot” points of interaction among descriptors or research lines. Finally, a light grey diamond shows the “cold” descriptors that went unidentified by FA and could not be assigned to any line of research.

Representation and visualization of data

In order to express and display the results in a clear, simple and precise manner, we opted to use graphs and tables as is usual in FA. This entailed the application of Pajek software (Batagelj and Mrvar 2010) for network analysis in the context of co-term study.

The final result is a series of four maps pertaining to each one of the temporal divisions made. Each map contains 60 descriptors that are interconnected and evoke the structure of a human neuron. We worked with maps of 200, 150, 100 and 50 descriptors, and came to the conclusion that the best choice—a compromise solution offering good visualization and informational structure—was the map of 60 descriptors: it identifies the same lines of research (factors) as does that of 200, but its terminological information is better restricted. Figures under 60 imply a loss of structural information.

To help visually establish the relationship between the symbols and lines of research, the equivalence between them is shown in the left corner of the graphs. The lines that connect the different symbols show the salient relations of co-occurrence among descriptors. These associations are thicker or thinner depending on the intensity of co-occurrence: the greater the intensity, the greater the thickness of the link.

Results

A breakdown into 10-year periods is shown (the last period spans 11 years). For each period, we can find parallels between the results obtained using FA and the results using Network analysis as applied to the study of co-terms.

The data shown above attest to the growth in studies of health and women throughout the period analyzed, with increases near 100% from one period to the next. Table 1 reflects the high level of representativity of the descriptors selected for our study. In the Annex of complementary information (II) we include a list of the frequency of appearance of the main descriptors detected in the full period.

First period (1965–1974)

Six lines of research stand out, corresponding to six specialized areas. The first three factors explained 76.5% of the variance, whereas the top six factors accounted for practically 90% of the variance (Table 2). This means that the results can be explained to a great extent on the basis of these six factors. For more detailed information see Annex I, Table 6. Of all, the field sparking most interest is that made up of studies related with *Pregnancy* (43.1% of variance). As seen in the map, these studies are closely linked with the rest of the factors, excepting *Socio-familiar Setting. Psychology*, linked by means of *Mental Illnesses*. A glance at the thematic profile of the descriptors shows that the greatest weight resides in the studies relative to the complications that may arise during gestation (*Fetus; Estriol; Amniotic Fluid; Placenta...*). Indeed, nearly all the descriptors corresponding to this factor are related through the actor *Pregnancy Complications*, which is the

Table 1 Level of representativity of the study

	Total no. of documents	No. of documents represented	%
1st period (1965–1974)	26,438	25,939	98.11
2nd period (1975–1984)	49,001	47,443	96.82
3rd period (1985–1994)	96,616	92,551	95.79
4th period (1995–2005)	193,659	179,621	92.75

Table 2 Factor analysis for the period 1965–1974

Factor	Name	Eigenvalue	%Variance	%Accumulated variance
1	<i>Pregnancy</i>	25.866	43.1	43.1
2	<i>Gynecological Neoplasms</i>	12.330	20.5	63.7
3	<i>Medical-Legal Studies</i>	7.272	12.9	76.5
4	<i>Social-Family Setting. Psychology</i>	3.277	5.5	82.0
5	<i>Mental Illnesses</i>	2.562	4.3	86.3
6	<i>Labor and Delivery</i>	1.581	2.6	88.9

most frequent descriptor (6,177 appearances). However, we also note interest in research concerned with the fetus or neonate (*Fetal Death; Fetal Diseases; Infant, Newborn, Diseases; Abortion, Spontaneous*), possible birth complications (*Puerperal Disorders*) and illnesses affecting the female reproductive organs (*Genital Diseases, Female; Uterus*), particularly when cancer is involved (*Uterine Cervical Neoplasms and Ovarian Neoplasms*).

Whereas complications during pregnancy stand out as topics, the lowest value of variance (2.6%) can be attributed to the studies factorized as *Delivery*, which nonetheless has sufficient relief to constitute a research line independent of *Pregnancy*. In the lower right section of the map we see the group of descriptors that integrate it (*Maternal Fetal Exchange; Labor Obstetric, Infant, Newborn and Research*).

Another set of predominant studies during this period in that of *Gynecological Neoplasms* (20.5% of variance), whose descriptors have a very compact and homogeneous presence in the network. This group is largely concerned with the study of tumors (*Carcinoma; Carcinoma, Intraductal, Noninfiltrating; Adenocarcinoma; Neoplasms Metastasis*), especially when affecting female reproductive organs (*Uterine Neoplasms; Breast Diseases*), and the techniques applied to their detection (*Mammography; Mastectomy; Mass Screening*). Not surprisingly, breast cancer (*Breast Neoplasms*, appearing 5,540 times) is seen to be the focal aspect of this factor.

The third most important line of research seen here is *Medical-Legal Studies* (12.9% of variance). The map reveals that these studies address ethical, legal or judicial problems surrounding the legalization of abortion, a highly relevant issue during this period: the descriptors *Abortion, Legal* (the most frequent one for this factor, 922 appearances) and *Abortion, Induced* are the actors integrating this line of study (*Ethics, Medical; Legislation, Medical; Attitude of Health Personnel; Jurisprudence...*).

The other two lines, *Social-Family Setting. Psychology* and *Mental Illnesses* have a discrete presence reflected in their respective FA variance values of 5.5 and 4.3%. We also note that these two lines are closely interrelated, their actors somewhat intertwined in the network. This suggests that research regarding the family environment is tied to mental disorders or illnesses that affect women's health, and which also may bear some relation with *Pregnancy*.

Second period (1975–1984)

Caption 1 Descriptors entering/exiting the map of the period 1975–1984

<i>Gynecological Neoplasms</i>	<i>Labor and Delivery</i>
Entering the map are (5): <i>Receptors, Estrogens; Mothers; Antineoplastic Agents; Breast; Estradiol; Tamoxifen</i>	Gone are (2): <i>Maternal Exchange; Research Gender Studies</i>
Gone are (3): <i>Neoplasm Metastasis; Ovarian Neoplasms; Contraceptives, Oral</i>	Entering are (4): <i>Identification Psychology; Gender Identity; Employment; Physicians, Women</i>
<i>Social-Family Setting. Psychology</i>	<i>Medical-Legal Studies</i>
Entering are (4): <i>Mothers; Violence, Neoplasms; Parents.</i>	Entering : (1) <i>Women's Rights</i>
Gone is (1): <i>Mental Health</i>	Gone (3): <i>Prostaglandins; Contraception; Legislation, Medical</i>
<i>Pregnancy</i>	<i>Mental Illnesses</i>
Entering are (2): <i>Ultrasonography; Smoking</i>	Gone (4): <i>Schizophrenia; Students; Parent-Child Relations; Interpersonal Relations</i>
Gone are (11): <i>Fetal Diseases; Maternal Fetal Exchange; Amniotic Fluid; Genital Diseases, Female; Uterus; Fetal Death; Puerperal Disorders; Estriol; Abortion, Spontaneous; Ovarian Neoplasms; Contraceptives, Oral</i>	

Table 3 Factor analysis for the period 1975–1984

Factor	Name	Eigenvalue	%Variance	%Accumulated variance
1	<i>Gynecological Neoplasms</i>	24.536	40.9	40.9
2	<i>Pregnancy</i>	12.335	20.6	61.5
3	<i>Social-Family Setting, Psychology</i>	6.204	10.3	71.8
4	<i>Medical-Legal Studies</i>	4.163	6.9	78.7
5	<i>Gender Studies</i>	2.364	3.9	82.7

In this decade, some important changes in the orientation of research become apparent. For one, the studies surrounding female reproduction lose their protagonism; the factor *Pregnancy* goes from 43.1% of variance to just 20.6%. We also observe a sharp reduction in the number of descriptors that represent it (from 19 to 9), and a simplification of the topics of interest in the area. Thus, complications in pregnancy (*Pregnancy Complications; Obstetric Labor Complications; Infant, Newborn, Diseases*) and the possible anomalies affecting the fetus or neonate (*Infant Newborn; Labor Obstetric; Placenta; Fetus*) are the main focuses of research during this period. This may have to do with the development of new techniques for monitoring gestation (*Ultrasonography*). Notwithstanding, *Pregnancy* is still the factor most interrelated with the rest, linking *Gynecological Neoplasms, Medical-Legal Studies* and *Social-Family Setting, Psychology* (Table 3).

Noteworthy is a certain line near that of *Pregnancy* involving a number of descriptors related with unhealthy habits that may harm the fetus (*Alcoholism; Substance Related Disorders*). Although they are clearly linked to pregnancy, they can not be ascribed to any specific factor.

Also important in this decade is the presence of the descriptor *Acquired Immunodeficiency Syndrome*. It marks the earliest research efforts regarding AIDS.

Meanwhile, the line of research into *Delivery* has disappeared. These studies no longer constitute a factor, and largely overlap with studies about *Pregnancy*.

Social-Family Setting, Psychology incorporates two descriptors that previously pertained to *Mental Illnesses*, which is no longer a factor. This allows them to connect directly with *Pregnancy* and gives them a clearer and more independent position in the network. We see that work surrounding the family environment constitutes an important aspect of this factor, as it did in the previous period. The psychological profile of the patient coping with illness (*Attitude to Health*), public health awareness (*Health Education*) and the profile of health professionals (*Attitude of Health personnel*) are new facets attracting the interest of researchers.

Another topic on the decline in this period is *Medical-Legal Studies* (going from 12.9% of variance to 6.9%), whose position on the map becomes less disperse. One exception is *Reproduction*, which belongs to this factor but appears connected to *Fetus*, in turn corresponding to the factor *Pregnancy*. This points to a significant phenomenon: the groupings made by FA do not always coincide with the pruning done by PFNET. The weight of these studies still resides in legal/ethical problems surrounding abortion. However, the appearance of two new descriptors suggests an incipient interest in other aspects of a legal nature involving women and health, such as *Family Planning Services* and *Women's Rights*.

In contrast, the studies of greatest interest in this period are those related with *Gynecological Neoplasms*. It may be that this factor benefits from the declining interest in studies of female reproduction. At any rate, its variance doubles with respect to the previous period (from 20.5 to 40.9%) and the number of descriptors represented increases considerably (from 13 to 22). Most of these descriptors are related to *Breast Neoplasms*, which is the most frequent descriptor in this period, having 13,049 appearances. It is still a very compact group, showing the strongest associations in the network. In it, new therapies or treatments for women with cancer (*Tamoxifen*; *Antineoplastic, Agents*; *Receptors, Estrogens...*) are seen to spark interest.

Perhaps the most remarkable aspect of the period, however, is the appearance of *Gender Studies*. While the variance is very low (3.9%), its presence as a factor reveals a novel perspective within studies involving women and health. Its position near the areas of *Social-Family Setting, Psychology* and *Gynecological Neoplasms* evidences a fundamental concern with the female role overall (*Gender Identity*; *Identification Psychology*) or more specifically in the working world (*Employment*) and within the medical profession (*Physicians, Women*).

Third period (1985–1994)

Caption 2 Descriptors entering/exiting the map of the period 1985–1994

<i>Gynecological Neoplasms</i>	<i>Gender Studies</i>
Entering the map are (3): <i>Antineoplastic Combined Chemotherapy Protocols</i> ; <i>Estrogen Replacement Therapy</i> ; <i>Patient Acceptance of Health Care</i>	Gone (1): <i>Physicians, Women</i>
Gone are (7): <i>Mastectomy</i> ; <i>Estrogens</i> ; <i>Antineoplastic Agents</i> ; <i>Adenocarcinoma</i> ; <i>Breast Diseases</i> ; <i>Estradiol</i> ; <i>Uterine Neoplasms</i>	<i>Drug Addiction</i>
<i>Social-Family Setting, Psychology</i>	Entering (1): <i>Cocaine</i>
Entering are (3): <i>Social Environment</i> ; <i>Social Support</i> ; <i>Sick Role</i>	<i>Work-related Aspects</i>
Gone are (2): <i>Family Therapy</i> ; <i>Parents</i>	Entering (2): <i>Women, Working</i> ; <i>Women's Health</i>
<i>Pregnancy</i>	<i>Health Education</i>
Gone (6): <i>Ultrasonography</i> ; <i>Infant, Newborn, Diseases</i> ; <i>Obstetric Labor Complications</i> ; <i>Labor, Obstetric</i> ; <i>Infant, Newborn</i> ; <i>Placenta</i>	Entering (4): <i>Health Promotion</i> ; <i>Health Behavior</i> ; <i>Patient Acceptance of Health Care</i> ; <i>African Americans</i>
	<i>AIDS and HIV Infection</i>
	Entering (6): <i>HIV Infections</i> ; <i>Health Knowledge, Attitudes, Practice</i> ; <i>HIV 1</i> ; <i>Pregnancy Complications, Infectious</i> ; <i>HIV Seropositivity, Women's Health</i>

In this period, we see that the research topics are more diversified, giving rise to a total of eight factors. Deserving special mention among them is *AIDS and HIV Infection*. FA gives it a high value for variance (20.7%), making it the second most predominant factor of the decade. The map illustrates its close relationship with two other research lines that also appear for the first time in this ten-year period: *Drug Addiction* and *Health Education*. Interestingly, the former develops as an offshoot of *Mental Disorders*. In other words, the map shows how research itself relates drug use with AIDS transmission. At the same time, the descriptor (*Acquired Immunodeficiency Syndrome*) acts as the nucleus from which the rest of the descriptors pertaining to this factor stem; it is the second most prevalent of the period, with 8,771 appearances. The most significant relations are oriented towards knowledge of HIV infection (*HIV 1*; *HIV Infections*; *HIV Seropositivity*). However, we also detect some interest in researching other possible causes of contagion or propagation of HIV during gestation (*Pregnancy Complications, Infectious*), or through sexual conduct (*Sexual Behavior*). Measures surrounding HIV/AIDS prevention are also present in this

Table 4 Factor analysis for the period 1985–1994

Factor	Name	Eigenvalue	%Variance	%Accumulated variante
1	<i>Gynecological Neoplasms</i>	15.802	26.3	26.3
2	<i>AIDS and HIV Infection</i>	12.226	20.7	46.7
3	<i>Social-Family Setting, Psychology</i>	10.868	18.1	64.5
4	<i>Drug Addiction</i>	5.807	9.7	74.5
5	<i>Work-related Aspects</i>	3.393	5.7	80.2
6	<i>Medical-Legal Studies</i>	2.569	4.3	84.4
7	<i>Gender Studies</i>	2.127	3.5	88.0
8	<i>Health Education</i>	1.676	2.8	90.8

line, and public health initiatives (*Health Promotion; Health Behavior; Attitude to Health*), become noteworthy focuses of study at this time, especially regarding populations at a higher risk of transmission (*African Americans*). Therefore, the descriptors pertaining to *Health Education* are closely linked to *AIDS and HIV Infection* (Table 4).

In parallel, studies of *Gynecological Neoplasms* constitute another salient line. While it is the factor with the highest percentage of variance (26.3%), its decline since the previous period (40.9%) is considerable. The diversification of factors may be the cause of this decline. Even so, we can consider it a consolidated research line, in which the strongest associations in the network come from the descriptor *Breast Neoplasm* (the most frequent one in this period, with 23,188 appearances). In this case, research efforts are seen to lean toward the development of new treatments (*Estrogen Replacement Therapy; Antineoplastic Combined Chemotherapy Protocols*).

A panoramic look at this map reveals that the factor most interrelated with the rest is *Social-Family Setting, Psychology* (linked with four of six possible factors). This is also a new development, as the core factor in previous depictions was *Pregnancy*. During this period, however, researchers have lost interest in pregnancy and have turned to studying broader, more social aspects of health and the environment of women, not just the context of reproduction. In addition to a substantial increase in the percentage of variance, the map indicates a consolidated prevalence of this interrelated factor. Its descriptors tend to form more homogeneous groups that occupy more or less central positions. We might underline the case of *Aged* which, despite belonging to this factor, is most closely linked to the descriptor *Health Status*, belonging to *Medical-Legal Studies*. Again, *the groupings made by FA do not always coincide with the pruning done by PFNET*.

Other interesting developments have to do with the thematic profile of these descriptors. Aside from the possible psychological or social implications of the family environment (*Mothers; Family*) and the psychological consequences of female cancer patients with *Gynecological Neoplasms (Neoplasms)*, interest has grown in key elements for the overall well-being of women (*Social Environment*) and patient assistance (*Social Support*).

Much more visible in this period is the link clearly existing between women's health and the working world. Whereas in the previous decade we saw descriptors of this realm factorized under *Gender Studies*, they are now seen to forge a research line of their own

(Work-related Aspects). Variance is substantial (5.7%), as is the number of descriptors (7). The basis of such studies is made up of health and employment (*Women Working; Employment; Women; Women's Health; Health*), maternity (*Reproduction*) and related legal regulations (*Women's Rights*). In view of the social character of the workplace, and the rights of working women, the position occupied by this factor on the map acts as the intermediary between factors *Social-Family Setting. Psychology* and *Medical-Legal Studies*.

In other words, as the process of legalizing abortion was becoming widespread in Europe during this decade, *Medical-Legal Studies* moved on to a concern with other problems, such as the legal implications of woman in the working world.

Finally, *Gender Studies* shows a limited presence in terms of the number of descriptors (dropping from 6 to just 2) while it remains relatively stable in terms of variance (from 3.9 to 3.5%). It can therefore be considered a consolidated line of research in this period.

Fourth period (1995–2004)

Caption 3 Descriptors entering/exiting the map of the period 1995–2005

<i>Gynecological Neoplasms</i>	<i>Social-Family Setting. Psychology</i>
Entering are (9): <i>Carcinoma, Ductal, Breast; Decision Making; Patient Selection; Adolescent Behavior; Antineoplastic Agents; Antineoplastic Agents, Hormonal; Tumor Markers, Biological; Patient Education; Ovarian Neoplasms</i>	Entering are (4): <i>Quality of Life; Questionnaires; Caregivers; Depression</i>
Gone are (3): <i>Receptors, Estrogen; Carcinoma; Carcinoma, Intraductal, Noninfiltrating</i>	Gone are (2): <i>Social Environment; Sick Role</i>
<i>AIDS and HIV Infection</i>	<i>Health Education</i>
Entering are (2): <i>Anti HIV Agents; Poverty</i>	Entering are (5): <i>Health Services Accessibility; Decision Making; Adolescent Behavior; Poverty; Patient Education</i>
Gone is (1): <i>HIV Seropositivity</i>	

The scientific subject matter of this decade is much less diverse. The number of factors detected drops to four, and none represent new developments. Again, *Gynecological Neoplasms* constitutes the outstanding focus of research (36.4% of variance). Attention continues to revolve around possible therapies (*Antineoplastic Agents; Antineoplastic Agents, Hormonal*) or studies of other tumors (*Ovarian Neoplasms; Carcinoma Ductal, Breast*). This is, moreover, the factor that has the most robust connections in the network, most stemming from *Breast Neoplasms* (48,899 hits), again the most frequent descriptor. Thus, we can consider it as the most consolidated line of research. However, despite its compact and autonomous presence as a group, here it is linked to a number of descriptors belonging to the factor *Social-Family Setting. Psychology*. These descriptors (*Quality of Life; Health Status; Employment*) are closely tied with *Gynecological Neoplasms* in that their concern is with the quality of life of cancer patients (Table 5).

Another line apparently consolidated in this period is *Social-Family Setting. Psychology*. Its variance has shown a steady rise over the four decades studied (5.5; 10.3; 18.1, and now 25.6%), as has the number of descriptors representing it (10; 14; 11 and 17). In addition to the aforementioned aspects related with *Gynecological Neoplasms* we see a persistent interest in family-related problems or difficulties in psychological adaptation for women with cancer or chronic illness.

Table 5 Factor analysis for the period 1995–2005

Factor	Name	Eigenvalue	%Variance	%Accumulated variance
1	<i>Gynecological Neoplasms</i>	21.836	36.4	36.4
2	<i>Social-Family Setting, Psychology</i>	15.379	25.6	62.0
3	<i>Health Education</i>	9.411	15.7	77.7
4	<i>AIDS and HIV Infection</i>	4.049	6.7	84.5

A careful look at the map reveals that one descriptor connects three of the four factors of this decade: *Attitude to Health* (9,177 appearances). Aside from belonging to the factors *Health Education* and *Social-Family Setting, Psychology*, it is related with *Gynecological Neoplasms*. This leads us to grasp the importance of attitude towards health at the social level, through programs or initiatives aimed to ensure effective Health Education as well as the personal or individual attitude of patients coping with cancer or aiming to prevent adverse effects for the family etc.

Studies in *Health Education* increase considerably, from 2.8% of variance to 15.7%. They are geared mostly toward public health initiatives (*Health Promotion; Health Behavior*), or increasing awareness of public health (*Health Knowledge, Attitudes, Practice; Health Education*), whether for health professionals (*Attitude of Health Personnel*) or, more importantly, for populations at risk (*Poverty; Africans Americans*).

Finally, it is apparent that studies relative to *AIDS and HIV Infection* are on the decline, dropping sharply with respect to the previous period. The number of descriptors goes from 19 to just 9, and the value of variance drops from 20.7 to 6.7%. One might surmise that the normalization of research in this area has led to a more moderate interest in the topic. The results of the previous decade would accordingly reflect how this research topic had been a landmark development in the world of biomedicine.

Discussion

In the first place, we should underline that the results obtained in this study may be influenced by factors that are not necessarily a consequence of the methodology used. The very nature of the data, the coverage of the database, and the natural behaviour of science itself can affect the findings.

The use of a controlled language favors the synthesis and simplification of the concepts or ideas represented by the contents of documents in a more systematic and coherent way than does natural language. While it is true that controlled terms have a tendency to lead one towards coarse-grained views, providing little insight into the variety of a given field, in our case they actually made it easier to produce schematic depictions and simplify analysis, which was one of our main objectives.

Straight co-occurrences may induce, as any other good technique, an interesting but truncated view of reality. Also, it reinforces a possible shortcoming of the choices made (high-frequency controlled terms), an increase of noise which make it risky to interpret some time changes.

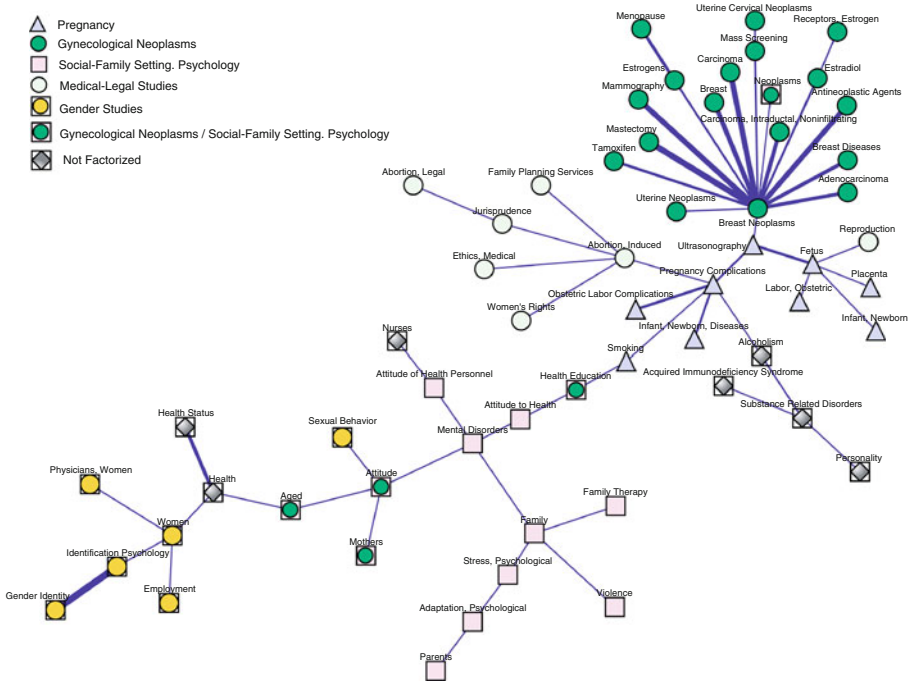


Fig. 2 Map of the 60 most frequent descriptors for the period 1975–1984

Once these have been mentioned, we have shown how raw co-occurrence in combination with dimension reduction techniques (PFNET) and layout algorithms type spring embedders, can identify and highlight salient lines of research. It is also shown how FA facilitates their delimitation. The combination of the two techniques allows us to re-build bridges (through FA groupings) not apparent when using the “no-cycle” option of PFNET. Though the two techniques are used in complementary fashion, there are exceptions, such as with *Reproduction* (Fig. 2) and *Aged* (Fig. 3). The fact that some descriptors appear connected with others indicates that these descriptors are related in common research articles (papers); but it does not imply that they pertain to the same line of research (factors). Some of the descriptors included in the maps may be recently coined terms, or recently incorporated into the research panorama, and therefore they are not yet correlated with the rest enough to be “factorized”. The opposite may also be true, when descriptors grow obsolete or become less used (and less correlated) in the context of scientific research, to the point of not being “factorized” (Fig. 4).

The decision to undertake a longitudinal study in ten-year periods is fundamentally based on documental production. Table 1 shows how the final period is especially influential. Its output stands as over half of the total documents studied (193.659 documents out of 365.714). If we were to represent the research lines of the entire forty-year period, we would see a clear bias due to the weight of documents in the last period, which would have a greater influence and would be more visible.

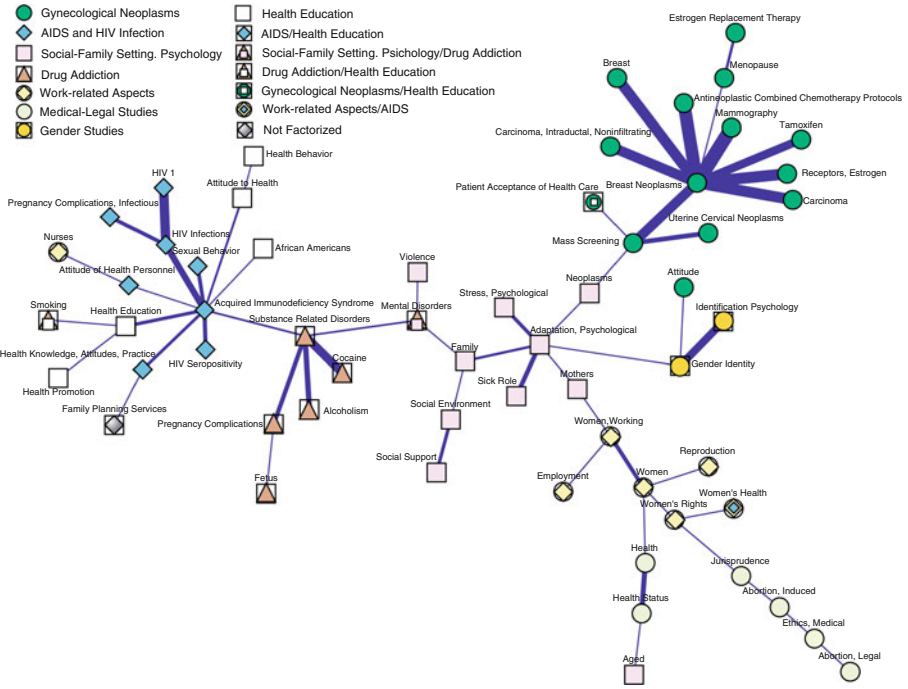


Fig. 3 Map of the 60 most frequent descriptors 1985–1994

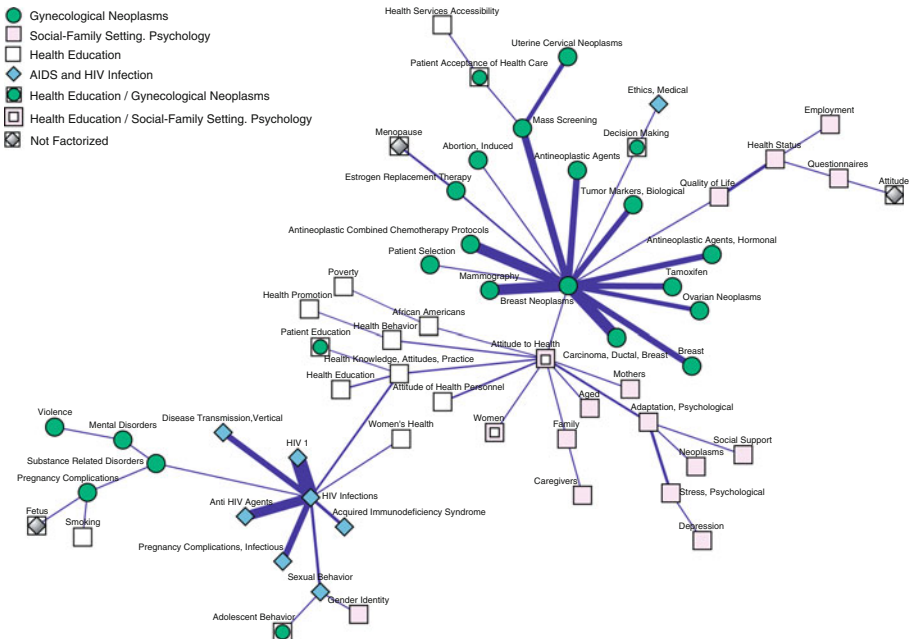


Fig. 4 Map of the 60 most frequent descriptors 1995–2005

The longitudinal study by means of ten-year maps shows the predominating research lines in each period plus their evolution over time. Through their analysis we can see that the aspects relative to Pregnancy and Delivery, as well as mental or psychological states possibly associated with them, constitute the central axis around which research took place during the first period. Even the studies about legal issues have to do with Pregnancy in the sense that legalizing abortion was a key issue at that particular time in most developed nations. It would seem that only those pathologies generated as a consequence of biological processes exclusive to women (i.e., reproduction) merited research with a female focus.

However, when we look at the entire period, we note that this characteristic changes over the years. In the decade 1975–1984, the weight and preponderance of these studies diminishes considerably. The factor *Pregnancy* is not as key or central, although it continues to be the element with more links and proximity to the rest of the factors. *Delivery* and *Mental Illnesses* are no longer even represented. *Medical-Legal Studies* are still largely concerned with legal, ethical and judicial problems surrounding abortion, but are on the decline. In turn, there is new interest in topics having to do with social setting of women. This interest gives rise to factors such as *Gender Studies* and concepts such as the psychological attitude of female patients, health education of women, or female health care professionals and their perspectives, all evidencing a change in the orientation of research.

This change is even clearer in the following decade. Between 1985 and 1994, studies related with *Social-Family Setting*. *Psychology* begin to be the nucleus from which the rest of the factors stem, finally displacing studies related with female reproduction. Thus, the line of research into *Pregnancy* disappears. We notice how the social aspects become prevalent, and are an important background for all the studies of this period, which features a greater number of factors than any other decade. This is most likely due to two important phenomena taking place at this point in time. For one, there is an important biomedical phenomenon whose transcendence is remarkable not only in the field of medicine but in the society in general—the appearance of AIDS. The study of this disease, together with the discovery of the HIV virus in 1986, had a great impact on scientific output. Indeed, as seen in our results, the factor corresponding to these studies becomes a predominant line of research of the period. Moreover, it gives rise to an interest in studies of a more social orientation, for instance bearing a causal relation with HIV transmission (*Drug addiction*) or having a focus on preventive or therapeutic measures (*Health Education*). In this way, studies of *AIDS and HIV Infection* contribute to fortifying relations between the social aspects conditioning the health of women as viewed in this ten-year period.

The sociological transformation that takes place at this time, with women joining the work force, lends added value to research about the importance of the social setting and workplace. The presence of women on a new stage beyond the family or domestic setting introduces an unexplored (and therefore attractive) field for research. The legal implications of this development also serve to modify the perspective of *Medical-Legal Studies*, no longer limited to abortion-related studies.

Whereas studies of *Gynecological Neoplasms* are indisputably a stronghold for research and a consolidated line of work, in this period the presence of the women in medical studies branches out. The female is not only considered when differences are established with regard to the male (sex), but woman is also perceived on a social level, in terms of the role played by women (gender). For this reason, we see a growing presence of the gender perspective within health issues.

Finally, in the last decade studied, such social aspects are even more significant. The number of descriptors and the thematic variety of studies enrich the corresponding lines of research. A sharp drop in studies of *AIDS and HIV Infection* with respect to the previous period confirms that its peak was conditioned by (then) current events, and that research shifts towards other aspects that are not strictly biological.

In sum, we can affirm that research involving Women and Health has undergone meaningful changes over the forty years studied here. Notwithstanding, concern with pathologies that affect both men and women, such as lung cancer or colon cancer, heart disease and others, were not detected. This may be because they do not have sufficient representativity, or, as pointed out by Fuster (2010) that woman is not lent due importance in these subject areas.

Conclusion

To conclude, we present a methodology that allows for the detection, identification and representation of the main lines of research that focus on Women and Health, looking at a forty-year period broken down into decades, and using the Medline database and MeSH descriptors. Medline provided us with the intellectual foundation used to discover the relevant research topics. The MeSH descriptors were the vehicle used to achieve a clear, synthesized and representative image of the main lines of research into Women and Health. The identification of these lines was possible because, despite the multidisciplinary nature of the information, there are always concrete limits to be found (Noyons and Calero-Medina 2008), and in our case these limits are established by PFNET pruning and drawn by FA.

Furthermore, we reflect the evolution of these research lines by means of a longitudinal study, which facilitates the interpretation of results over time. Research surrounding Health and Women has undergone important changes in this period on the whole. However, there are aspects that are not detected, possibly because as a research focus, woman is not yet given proper importance.

Finally, the scope of this study opens a door to future action. Aside from constituting a valid tool for the thematic identification of a specific domain, the methodology presented here may stand as an early stage for further bibliometric studies on the micro or middle level of analysis aimed at complex realms of knowledge.

Acknowledgments This study was made possible through an agreement with the *Observatorio de Salud de la Mujer*, under Spain's Ministerio de Sanidad y Consumo. The authors wish to express their gratitude to the anonymous reviewers for their many useful comments, suggestions and help.

Annex I

See Tables 6, 7, 8, 9 and 10.

Table 6 Factor membership

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4	Descriptors	F5	Descriptors	F6
<i>Genital Diseases, Female</i>	0.985	<i>Neoplasm Metastasis</i>	0.968	<i>Jurisprudence</i>	0.952	<i>Stress, Psychological</i>	0.886	<i>Mental Disorders</i>	0.751	<i>Research</i>	0.710
<i>Puerperal Disorders</i>	0.981	<i>Carcinoma</i>	0.964	<i>Ethics, Medical</i>	0.930	<i>Family Therapy</i>	0.842	<i>Parent Child Relations</i>	0.712	<i>Labor, Obstetric</i>	0.648
<i>Infant, Newborn, Diseases</i>	0.977	<i>Breast Diseases</i>	0.963	<i>Legislation, Medical</i>	0.918	<i>Family</i>	0.827	<i>Schizophrenia</i>	0.703	<i>Infant, Newborn</i>	0.636
<i>Obstetric Labor Complications</i>	0.976	<i>Adenocarcinoma</i>	0.962	<i>Abortion, Legal</i>	0.860	<i>Attitude to Health</i>	0.813	<i>Personality</i>	0.693	<i>Maternal Fetal Exchange</i>	0.624
<i>Fetal Death</i>	0.970	<i>Mammography</i>	0.962	<i>Prostaglandins</i>	0.713	<i>Adaptation, Psychological</i>	0.775	<i>Interpersonal Relations</i>	0.678		
<i>Fetal Diseases</i>	0.966	<i>Carcinoma, Intraductal, Noninfiltrating</i>	0.956	<i>Attitude of Health Personnel</i>	0.674	<i>Mental Health</i>	0.753	<i>Students</i>	0.663		
<i>Estriol</i>	0.966	<i>Mastectomy</i>	0.955	<i>Contraception</i>	0.501	<i>Aged</i>	0.738	<i>Health Education</i>	0.653		
<i>Uterus</i>	0.964	<i>Breast Neoplasms</i>	0.838			<i>Attitude</i>	0.717				
<i>Placenta</i>	0.947	<i>Estrogens</i>	0.756			<i>Substance Related Disorders</i>	0.694				
<i>Abortion, Spontaneous</i>	0.944	<i>Uterine Neoplasms</i>	0.734			<i>Women</i>	0.576				
<i>Amniotic Fluid</i>	0.925	<i>Mass Screening</i>	0.665								
<i>Uterine Cervical Neoplasms</i>	0.858	<i>Contraceptives, Oral</i>	0.599								
<i>Ovarian Neoplasms</i>	0.827	<i>Ovarian Neoplasms</i>	0.509								
<i>Pregnancy Complications</i>	0.656										
<i>Infant, Newborn</i>	0.644										

Table 6 continued

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4	Descriptors	F5	Descriptors	F6
<i>Fetus</i>	0.631										
<i>Labor, Obstetric</i>	0.624										
<i>Maternal Fetal Exchange</i>	0.618										
<i>Uterine Neoplasms</i>	0.600										
<i>Contraceptives, Oral</i>	0.521										

Descriptors for the period 1965–1974 with a higher value than 0.5 ranked by factor loadings

Table 7 Factor membership

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4	Descriptors	F5
<i>Mammography</i>	0.968	<i>Placenta</i>	0.936	<i>Stress, Psychological</i>	0.867	<i>Jurisprudence</i>	0.919	<i>Identification Psychology</i>	0.776
<i>Antineoplastic Agents</i>	0.967	<i>Infant, Newborn</i>	0.932	<i>Adaptation, Psychological</i>	0.854	<i>Ethics, Medical</i>	0.897	<i>Gender Identity</i>	0.775
<i>Mastectomy</i>	0.965	<i>Labor, Obstetric</i>	0.911	<i>Attitude to Health</i>	0.834	<i>Abortion, Legal</i>	0.890	<i>Physicians, Women</i>	0.703
<i>Receptors, Estrogen</i>	0.965	<i>Ultrasonography</i>	0.881	<i>Family Therapy</i>	0.830	<i>Family Planning Services</i>	0.774	<i>Women</i>	0.635
<i>Adenocarcinoma</i>	0.962	<i>Infant, Newborn, Diseases</i>	0.879	<i>Family</i>	0.824	<i>Women's Rights</i>	0.711	<i>Employment</i>	0.555
<i>Breast Diseases</i>	0.959	<i>Obstetric Labor Complications</i>	0.851	<i>Violence</i>	0.798	<i>Reproduction</i>	0.703	<i>Sexual Behavior</i>	0.513
<i>Estradiol</i>	0.958	<i>Fetus</i>	0.781	<i>Attitude of Health Personnel</i>	0.644	<i>Abortion, Induced</i>	0.696		
<i>Breast</i>	0.957	<i>Pregnancy Complications</i>	0.713	<i>Parents</i>	0.628				
<i>Tamoxifen</i>	0.957	<i>Smoking</i>	0.625	<i>Mental Disorders</i>	0.621				
<i>Carcinoma, Intraductal, Noninfiltrating</i>	0.954			<i>Aged</i>	0.578				
<i>Carcinoma</i>	0.953			<i>Attitude</i>	0.553				
<i>Mass Screening</i>	0.946			<i>Mothers</i>	0.531				
<i>Estrogens</i>	0.879			<i>Neoplasms</i>	0.520				
<i>Uterine Neoplasms</i>	0.843			<i>Health Education</i>	0.516				
<i>Uterine Cervical Neoplasms</i>	0.821								
<i>Breast Neoplasms</i>	0.776								
<i>Health Education</i>	0.625								
<i>Neoplasms</i>	0.563								
<i>Menopause</i>	0.551								
<i>Mothers</i>	0.518								
<i>Attitude</i>	0.508								
<i>Aged</i>	0.503								

Descriptors for the period 1975–1984 with a higher value than 0.5 ranked by factor loadings

Table 8 Factor membership

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4
Antineoplastic Combined Chemotherapy, Protocols	0.960	HIV Seropositivity	0.952	Stress, Psychological	0.950	Cocaine	0.959
Tamoxifen	0.960	Pregnancy Complications, Infectious	0.950	Family	0.950	Pregnancy Complications	0.956
Receptors, Estrogen	0.957	HIV 1	0.950	Adaptation, Psychological	0.932	Alcoholism	0.925
Breast	0.957	Sexual Behavior	0.941	Sick Role	0.927	Substance Related Disorders	0.864
Mammography	0.957	HIV Infections	0.922	Social Support	0.913	Mental Disorders	0.683
Carcinoma, Intraductal, Noninfiltrating	0.956	Health Knowledge, Attitudes, Practice	0.889	Social Environment	0.893	Fetus	0.518
Carcinoma	0.955	Acquired Immunodeficiency Syndrome	0.871	Mothers	0.769	Smoking	0.514
Mass Screening	0.930	Health Education	0.832	Neoplasms	0.759		
Breast Neoplasms	0.854	Attitude of Health Personnel	0.757	Aged	0.711		
Menopause	0.836	Attitude to Health	0.657	Violence	0.543		
Estrogen Replacement Therapy	0.781	Women's Health	0.577	Mental Disorders	0.538		
Uterine Cervical Neoplasms	0.689						
Patient Acceptance of Health Care	0.678						
Attitude	0.515						

Descriptors	F5	Descriptors	F6	Descriptors	F7	Descriptors	F8
Women, Working	0.886	Ethics, Medical	0.733	Identification Psychology	0.929	Health Promotion	0.869
Employment	0.863	Health Status	0.730	Gender Identity	0.905	Health Behavior	0.804
Women	0.845	Health	0.707			Smoking	0.760
Reproduction	0.731	Abortion, Legal	0.670			African Americans	0.737
Nurses	0.705	Abortion, Induced	0.635			Attitude to Health	0.636
Women's Rights	0.696	Jurisprudence	0.601			Patient Acceptance of Health Care	0.532
Women's Health	0.614					Health Education	0.505

Descriptors for the period 1985–1994 with a higher value than 0.5 ranked by factor loadings

Table 9 Factor membership

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4
Ovarian Neoplasms	0.972	Social Support	0.956	Health Education	0.962	Pregnancy Complications, Infectious	0.909
Mammography	0.968	Adaptation, Psychological	0.950	Health Promotion	0.942	Disease Transmission, Vertical	0.907
Antineoplastic Agents	0.964	Stress, Psychological	0.930	Health Behavior	0.912	HIV 1	0.906
Antineoplastic Combined Chemotherapy Protocols	0.962	Depression	0.912	Health Knowledge, Attitudes, Practice	0.892	Anti HIV Agents	0.904
Breast	0.958	Caregivers	0.902	African Americans	0.890	Acquired Immunodeficiency Syndrome	0.863
Carcinoma, Ductal, Breast	0.954	Family	0.871	Women's Health	0.863	HIV Infections	0.796
Tumor Markers, Biological	0.953	Employment	0.819	Health Services Accessibility	0.794	Ethics, Medical	0.771
Tamoxifen	0.951	Aged	0.812	Attitude of Health Personnel	0.792	Sexual Behavior	0.624
Antineoplastic Agents, Hormonal	0.951	Quality of Life	0.784	Smoking	0.778		
Mass Screening	0.933	Neoplasms	0.781	Women	0.771		
Patient Selection	0.915	Attitude to Health	0.745	Patient Education	0.764		
Estrogen Replacement Therapy	0.820	Health Status	0.721	Poverty	0.731		
Breast Neoplasms	0.751	Questionnaires	0.720	Patient Acceptance of Health Care	0.729		
Abortion, Induced	0.679	Mothers	0.704	Adolescent Behavior	0.596		
Substance Related Disorders	0.645	Gender Identity	0.689	Attitude to Health	0.571		
Uterine Cervical Neoplasms	0.640	Women	0.554	Decision Making	0.549		
Violence	0.578						
Patient Acceptance of Health Care	0.574						
Mental Disorders	0.562						
Adolescent Behavior	0.552						
Decision Making	0.542						
Patient Education	0.539						
Pregnancy Complications	0.514						

Descriptors for the period 1995–2005 with a higher value than 0.5 ranked by factor loadings

Table 10 Factor membership

Descriptors	F1	F2	F3	F4	F5	F6
<i>Mammography</i>	0.979	0.969	0.940	0.918	0.898	0.938
	<i>Social Support</i>	<i>Health Promotion</i>		<i>Abortion, Induced</i>	<i>Pregnancy Complications</i>	<i>Anti HIV Agents</i>
<i>Antineoplastic Agents</i>	0.971	0.958	0.938	0.890	0.873	0.938
	<i>Adaptation, Psychological</i>	<i>Health Education</i>		<i>Abortion, Legal</i>	<i>Alcoholism</i>	<i>HIV 1</i>
<i>Antineoplastic Combined Chemotherapy Protocols</i>	0.971	0.939	0.919	0.886	0.736	0.930
	<i>Stress, Psychological</i>	<i>Health Behavior</i>		<i>Ethics, Medical</i>	<i>Substance Related Disorders</i>	<i>Pregnancy Complications, Infectious</i>
<i>Tamoxifen</i>	0.967	0.881	0.850	0.593	0.601	0.875
	<i>Family</i>	<i>Health Knowledge, Attitudes, Practice</i>		<i>Decision Making</i>	<i>Fetus</i>	<i>HIV Infections</i>
<i>Receptors, Estrogen</i>	0.966	0.820	0.842		0.593	0.840
	<i>Aged</i>	<i>African Americans</i>			<i>Mental Disorders</i>	<i>Acquired Immunodeficiency Syndrome</i>
<i>Carcinoma Breast</i>	0.964	0.778	0.808		0.543	0.685
	<i>Neoplasms</i>	<i>Women's Health</i>			<i>Smoking</i>	<i>Sexual Behavior</i>
	0.964	0.766	0.743			
	<i>Employment</i>	<i>Family Planning Services</i>				
<i>Carcinoma, Intraductal, Noninfiltrating</i>	0.963	0.740	0.726			
	<i>Quality of Life</i>	<i>Health Services Accessibility</i>				
<i>Carcinoma, Ductal, Breast</i>	0.961	0.733	0.712			
	<i>Health Status</i>	<i>Attitude of Health Personnel</i>				
<i>Mass Screening Estrogens</i>	0.951	0.726	0.692			
	<i>Mothers</i>	<i>Smoking</i>				
	0.866	0.717	0.692			
	<i>Attitude to Health</i>	<i>Poverty</i>				
<i>Breast Neoplasms</i>	0.826	0.647	0.671			
	<i>Women</i>	<i>Patient Acceptance of Health Care</i>				
<i>Estrogen Replacement Therapy</i>	0.747	0.618	0.649			
	<i>Gender Identity</i>	<i>Women</i>				
<i>Uterine Cervical Neoplasms</i>	0.696	0.576	0.595			
	<i>Nurses</i>	<i>Attitude to Health</i>				

Table 10 continued

Descriptors	F1	Descriptors	F2	Descriptors	F3	Descriptors	F4	Descriptors	F5	Descriptors	F6
<i>Patient Acceptance of Health Care</i>	0.605	<i>Mental Disorders</i>	0.534	<i>Adolescent Behavior</i>	0.583						
<i>Menopause</i>	0.580			<i>Sexual Behavior</i>	0.501						
<i>Gender Identity</i>	0.531										
<i>Violence</i>	0.525										
<i>Adolescent Behavior</i>	0.502										

Descriptors for the period 1995–2005 with a higher value than 0.5 ranked by factor loadings

Annex II

Descriptors (Ranked by number of axis and factor loadings)	Frequency Full period 1965–2005	Frequency 1st period 1965–1974	Frequency 2nd period 1975–1984	Frequency 3rd period 1985–1994	Frequency 4th period 1995–2005
<i>Mammography</i>	5,842	256	577	1,630	3,379
<i>Antineoplastic Combined Chemotherapy Protocols</i>	4,405			1,404	3,001
<i>Antineoplastic Agents</i>	2,532		591		1,941
<i>Tamoxifen</i>	2,976		299	918	1,759
<i>Receptors, Estrogen</i>	2,044		822	1,222	
<i>Carcinoma</i>	2,261	455	590	1,216	
<i>Breast</i>	3,598		543	1,227	1,828
<i>Carcinoma, Intraductal, Noninfiltrating</i>	1,744	229	432	1,083	
<i>Carcinoma, Ductal, Breast</i>	3,108				3,108
<i>Mass Screening</i>	15,023	712	1,183	3,998	9,130
<i>Estrogens</i>	1,081	443	638		
<i>Breast Neoplasms</i>	90,676	5,540	13,049	23,188	48,899
<i>Estrogen Replacement Therapy</i>	6,733			1,251	5,482
<i>Uterine Cervical Neoplasms</i>	3,528	383	414	819	1,912
<i>Patient Acceptance of Health Care</i>	3,611			919	2,692
<i>Menopause</i>	7,460	834	1,472	2,422	2,732
<i>Gender Identity</i>	4,377		732	1,862	1,783
<i>Violence</i>	4,028		355	1,096	2,577
<i>Adolescent Behavior</i>	2,008				2,008
<i>Social Support</i>	2,658			815	1,843
<i>Adaptation, Psychological</i>	11,305	455	950	3,529	6,371
<i>Stress, Psychological</i>	12,245	630	1,740	3,187	6,688
<i>Family</i>	9,529	724	1,594	3,322	3,889
<i>Aged</i>	6,809	653	1,765	1,979	2,412
<i>Neoplasms</i>	3,229		354	856	2,019
<i>Employment</i>	3,101		486	798	1,817
<i>Quality of Life</i>	2,986				2,986
<i>Health Status</i>	8,120		388	1,677	6,055
<i>Mothers</i>	7,030		651	1,949	4,430
<i>Attitude to Health</i>	13,507	492	1,143	2,755	9,117
<i>Women</i>	5,860	372	1,316	2,004	2,168
<i>Gender Identity</i>	4,377		732	1,862	1,783
<i>Nurses</i>	1,774		560	1,214	
<i>Mental Disorders</i>	4,609	447	555	1,146	2,461
<i>Health Promotion</i>	5,045			1,256	3,789
<i>Health Education</i>	7,012	387	1,084	2,125	3,416
<i>Health Behavior</i>	4,508			931	3,577
<i>Health Knowledge, Attitudes, Practice</i>	8,906			1,580	7,326

Annex II continued

Descriptors (Ranked by number of axis and factor loadings)	Frequency	Frequency	Frequency	Frequency	Frequency
	Full period 1965–2005	1st period 1965–1974	2nd period 1975–1984	3rd period 1985–1994	4th period 1995–2005
<i>African Americans</i>	2,953			691	2,262
<i>Women's Health</i>	6,230			934	5,296
<i>Family Planning Services</i>	3,300	1,108	1,189	1,003	
<i>Health Services Accessibility</i>	2,842				2,842
<i>Attitude of Health Personnel</i>	8,681	253	667	1,868	5,893
<i>Poverty</i>	1,906				1,906
<i>Smoking</i>	3,571		417	955	2,199
<i>Patient Acceptance of Health Care</i>	3,611			919	2,692
<i>Women</i>	5,860	372	1,316	2,004	2,168
<i>Attitude to Health</i>	13,507	492	1,143	2,755	9,117
<i>Adolescent Behavior</i>	2,008				2,008
<i>Sexual Behavior</i>	8,418	570	1,242	2,490	4,116
<i>Abortion, Induced</i>	7,064	958	2,010	1,796	2,300
<i>Abortion, Legal</i>	2,708	922	878	908	
<i>Ethics, Medical</i>	4,019	239	584	1,414	1,782
<i>Decision Making</i>	2,340				2,340
<i>Pregnancy Complications</i>	31,321	6,177	6,570	8,017	10,557
<i>Alcoholism</i>	1,703		541	1,162	
<i>Substance Related Disorders</i>	16,657	1,656	2,444	5,053	7,504
<i>Fetus</i>	10,884	2,227	3,138	2,492	3,027
<i>Mental Disorders</i>	4,609	447	555	1,146	2,461
<i>Smoking</i>	3,571		417	955	2,199
<i>Anti HIV Agents</i>	3,019				3,019
<i>HIV 1</i>	6,345			1,523	4,822
<i>Pregnancy Complications, Infectious</i>	3,097			891	2,206
<i>HIV Infections</i>	25,671			4,771	20,900
<i>Acquired Immunodeficiency Syndrome</i>	14,441		417	8,771	5,253
<i>Sexual Behavior</i>	8,418	570	1,242	2,490	4,116

References

- Batagelj, V., & Mrvar, A. (2010). *Pajek 1.28: Package for large network analysis*. Cambridge, NY: CUP.
- Benzecri, J. P. (1976). *L'Analyse des Données, Tome II: L'Analyse des Correspondances*. Paris: Dunod.
- Benzecri, J. P., et al. (1981). *Pratique de l'analyse des données, Linguistique et lexicologie*. Paris: Dunod.
- Berlin, A. J., & Ellenberg, S. S. (2009). Inclusion of women in clinical trials. *BMC Medicine*, 7(56). <http://www.biomedcentral.com/1741-7015/7/56>.
- Blake, M. B., Ince, M., & Dean, C. J. (2005). Inclusion of women in cardiac research: Current trends and need for reassessment. *Gender Medicine*, 2(2), 71–75.
- Borrell, C., & Artazcoz, L. (2008). The inequalities of gender in health: Challenges for the future. *Revista Española de Salud Pública*, 82(3), 245–249.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991a). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science*, 42(4), 233–251.

- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991b). Mapping of science by combined co-citation and word analysis. II. Dynamical aspects. *Journal of the American Society for Information Science*, 42(4), 252–266.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information (Sur Les Sciences Sociales)*, 22, 191–235.
- Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27(2), 119–243.
- Chen, C. H. (2002). *Mapping scientific frontiers*. London: Springer.
- Chen, C., Paul, R. J., & O’Keefe, B. (2001). Fitting the jigsaw of citation: Information visualization in domain analysis. *Journal of the American Society for Information Science and Technology (JASIST)*, 52(4), 315–330.
- Clarke, J. N. (1983). Sexism, feminism and medicalism: A decade of literature on gender and illness. *Sociology of Health & Illness*, 5(1), 62–82.
- Courtial, P., Cahlik, T., & Callon, M. (1994). A model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics*, 31(2), 173–192.
- Daly, C., Clemens, F., Lopes Sendon, J. L., Tavazzi, L., Boersma, E., Danchin, N., et al. (2006). Gender differences in the management and clinical outcome of stable angina. *Circulation*, 113(4), 490–498.
- Biblioteca Regional de Medicina (BIREME). (1967). Descriptores en ciencias de la salud, DeCS, Brasil. Accessed February 20, 2009, from <http://decs.bvs.br/E/decswebe.htm>.
- Devesa, S. S., Bray, F., Vizcaino, A. P., & Parkin, D. M. (2005). International lung cancer trends by histologic type: Male:female differences diminishing and adenocarcinoma rates rising. *International Journal of Cancer*, 117(2), 294–299.
- Ding, Y., Chowdhury, G., & Foot, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing Management*, 37(6), 67–78.
- Fuster, V. (2010). No se ha dado a la mujer la importancia que se debería en temas de salud cardiovascular, Entrevista en el “El País Digital”. Accessed July 21, 2010, from http://www.elpais.com/articulo/sociedad/ha/dado/mujer/importancia/deberia/temas/salud/cardiovascular/elpepusoc/20100721elpepusoc_7/Tes.
- Gender and Health Group. (2000). Guidelines for the analysis of Gender and Health Group. Accessed March 8, 2010, from <http://www.lstmliverpool.ac.uk/research/academic-groups/international-health/gender-and-health-group/guidelines>.
- Gowen, A. (2009). Pregnant women in D.C. Area cautious about Flu Shot. Washington Post.
- Hamberg, K., Risberg, G., & Johansson, E. E. (2004). Male and female physicians show different patterns of gender bias. A paper-case study of management of irritable bowel syndrome. *Scandinavian Journal of Public Health*, 32(2), 144–152.
- Inclusion of Women and Minorities in Clinical-Trials and the NIH-Revitalization-Act of 1993. (1995). The perspective of NIH Clinical Trialists. *Controlled Clinical Trials*, 16(5), 277–312.
- Institute of Medicine. (2001). Committee on understanding the biology of sex and gender differences. In T. M. Wizemann & M.-L. Pardu (Eds.), *Exploring the biological contributions to human health: Does sex matter?*. Washington: National Academy Press.
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31, 7–15.
- Kerrigan, D., Andrinopoulos, K., Chung, S., Glass, B., & Ellen, J. (2008). Gender ideologies, socioeconomic opportunities, and HIV/STI-related vulnerability among female, African-American adolescents. *Journal of Urban Health*, 85(5), 717–726.
- Krempel, L. (1999). Visualizing networks with spring embedders: Two-mode and valued graphs. In *International sunbelt social network conference*, Charleston, SC.
- Lalonde, M. (1974). Social values and public hygiene. *Canadian Journal of Public Health (Revue Canadienne de Sante Publique)*, 65(4), 260–268.
- Leviatan, U., & Cohen, J. (1985). Gender differences in life expectancy among kibbutz members. *Social Science Medicine*, 21(5), 545–551.
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, 48(5), 418–427.
- Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348–462.
- Merkatz, R. B., & Junod, S. W. (1994). Historical background of changes in FDA policy on the study an evaluation of drugs in women. *Academic Medicine*, 69(9), 703–707.
- Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Timeline visualization of research fronts. *Journal of the American Society for Information Science and Technology*, 54(5), 413–422.

- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., González-Molina, A., Muñoz-Fernández, F. J., et al. (2007). Visualizing the Marrow of Science. *Journal of the American Society for Information Science and Technology (JASIST)*, 58(14), 2167–2179.
- Nathanson, C. A. (1975). Illness and the feminine role: A theoretical review. *Social Science and Medicine*, 9, 57–62.
- Nathanson, C. A. (1977). Sex, illness and medical care: A review of data, theory and method. *Social Science and Medicine*, 11, 13–25.
- Nobelius, A. M., & Wainer, J. (2004). *Gender and medicine: A conceptual guide for medical educators*. Victoria: Monash University School of Rural Health. <http://www.med.monash.edu.au/gendermed/docs/A4TutorManual120504screen.pdf>.
- Noyons, E. C. M., & Calero-Medina, C. (2008). Applying bibliometric mapping in a high level science policy context. *Scientometrics*, 79(2), 261–275.
- Onyancha, O. B., & Ocholla, D. N. (2005). Informetric investigation of the relatedness of opportunistic infections to HIV/AIDS. *Information Processing and Management*, 41(6), 1573–1588.
- Ovid Technologies. (2005). <http://www.ovid.com>.
- Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood: ABLEX.
- Sen, G., George, A., & Östlin, P. (2005a). *Incorporación de la perspectiva de género en la investigación y en las políticas en salud: una perspectiva de género*. Harvard: Harvard Center for Population and Development Studies.
- Sen, G., George, A., & Östlin, P. (2005b). *Incorporar la perspectiva de género en la equidad en salud: un análisis de la investigación y las políticas*. Washington: Organización Panamericana de la Salud.
- Singh-Manoux, A., Gueguen, A., Ferrie, J., Shipley, M., Martikainen, P., Bonenfant, S., et al. (2008). Gender differences in the association between morbidity and mortality among middle-aged men and women. *American Journal of Public Health*, 98(12), 2251–2257.
- Van Raan, A. F. J., & Tijssen, R. J. W. (1993). The neuronal net of neuronal network research. An exercise in bibliometric mapping. *Scientometrics*, 26(1), 169–192.
- Vargas-Quesada, B., Chinchilla-Rodríguez, Z., González-Molina, A., & Moya-Anegón, F. (2010). Showing the essential science structure of a scientific domain and its evolution. *Information Visualization*, 9, 288–300.
- Verbrugge, L. M. (1978). Females and illness: Recent trends in sex differences in the United States. *Journal of Health and Social Behavior*, 17(4), 387–403.
- Verbrugge, L. M. (1985). Gender and health: An update of hypotheses. *Journal of Health and Social Behavior*, 26(3), 156–182.
- Verbrugge, L. M. (1989). The Twain meet: Empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior*, 30(3), 282–304.
- Vivian, W., & Pinn, M. D. (2003). Sex and gender factors in medical studies. Implications for health and clinical practice. *Journal of the American Medical Association*, 289(4), 397–400.
- Vivian, W., & Pinn, M. D. (2005). Research on women's health. Progress and opportunities. *Journal of the American Medical Association*, 294(11), 1407–1410.
- Vivian, W., & Pinn, M. D. (2006). Women's health research and health leadership: Benchmarks of the continuum. *Journal of Dental Education*, 70(11 Supplement), 27–34.
- Waldron, I. (1978). Why do women live longer than men? *Social Science and Medicine*, 10(7–8), 349–362.
- White, H. D. (2003). Pathfinder networks and author co-citation analysis: A remapping of paradigmatic information scientist. *Journal of the American Society for Information Science and Technology (JASIST)*, 54(5), 423–434.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science (JASIS)*, 49(4), 327–355.
- Wingard, D. (1984). The sex differential in morbidity in morbidity, mortality and lifestyle. *Annual Review of Public Health*, 5, 433–458.
- Women's Health Bureau. (2002). *Exploring concepts of gender and health*. Ottawa: Health Canada.
- World Health Organization (WHO). (1998). Gender and health: A technical paper. Geneva.
- World Health Organization (WHO). (2002). Integrating gender perspectives in the work of WHO: WHO Gender Policy. <http://www.who.int/gender/documents/engpolicy.pdf>.
- World Health Organization (WHO). (2006). Constitution of the World Health Organization, Basic documents, Forty-fifth edition, Supplement, October 2006. http://www.who.int/governance/eb/who_constitution_en.pdf.

PUBLICACIÓN Nº 2

Zulueta, M. Ángeles, Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2010). Spanish stem cell research (1997-2007). Investigación española con células madre. Análisis bibliométrico y principales líneas de investigación a través de los *KeyWords Plus*. *I International workshop on scientometric studies related to the biomedical sciences. XV international scientific congress CNIC 2010*. La Habana (Cuba), 28 de junio al 1 de julio 2010, 1-21.

Presentación disponible en:

https://www.researchgate.net/publication/260900828_Investigacion_espanola_con_celulas_madre_Analisis_bibliometrico_y_principales_lineas_de_investigacion_a_traves_de_los_KeyWords_Plus

Para más información sobre la comunicación:

<http://redec-mes.blogspot.com/2010/10/1st-international-workshop-of.html>

Para más información sobre el congreso:

<http://ferbor.blogspot.com.es/2009/02/international-workshop-on-scientometric.html>

Investigación española en células madre (1997-2007). Análisis bibliométrico y principales líneas de investigación a través de los KeyWords Plus

M. A. Zulueta¹, G. Cantos-Mateos¹, B. Vargas-Quesada² y Z. Chinchilla-Rodríguez³

1ma.zulueta@uah.es

Information Science Faculty. University of Alcalá., C/ San Cirilo s/n, 28804, Alcalá de Henares, Madrid (Spain)

1gisela.cantos@gmail.com

Information Science Faculty. University of Alcalá., C/ San Cirilo s/n, 28804, Alcalá de Henares, Madrid (Spain)

2benjamin@ugr.es

Communication and Information Science Faculty, University of Granada, Colegio Máximo de Cartuja s/n, 18071, Granada (Spain)

(CSIC), Unidad Asociada Grupo SCImago, Madrid (Spain)

[3 zaida.chinchilla@scimago.es](mailto:3zaida.chinchilla@scimago.es)

(CSIC), Unidad Asociada Grupo SCImago, Madrid (Spain)

Abstract

Introduction: La investigación española con células madre constituye uno de los aspectos biomédicos de mayor repercusión científica. La utilidad de los estudios bibliométricos para el análisis de dominios científicos está ampliamente demostrada. El objetivo de este trabajo consiste en la realización de un estudio de la investigación sobre este tema durante el período 1997-2007 y la aplicación de una metodología válida para la detección de las principales líneas de investigación. **Material and Methods:** Los documentos proceden de la base de datos del SCI mediante la búsqueda de los términos *stem* cell** en el campo *topic*. Los análisis realizados se han dirigido, por un lado, a la obtención de indicadores bibliométricos y por otro, al estudio de las co-ocurrencias de los KeyWords Plus contenidos en los documentos. **Results and Discussion:** Se han obtenido 2.369 documentos procedentes mayoritariamente de las Comunidades Autónomas de Cataluña, Madrid y Valencia y circunscrita al sector sanitario, fundamentalmente. Los centros más productivos han sido los hospitales de Santa Cruz y San Pablo (Barcelona), el Clínico de Barcelona y el Clínico de Salamanca. Los autores que han destacado han sido J. Sierra, R. Martino y J. F. San Miguel. *Bone Marrow Transplantation*, *Blood* y *Haematologica* han sido las revistas más productivas y las categorías más frecuentes: Hematología, Oncología y Neurociencias. El análisis de las relaciones de los KeyWords Plus ha permitido detectar cuatro líneas prioritarias de investigación: *Trasplante de progenitores hematopoyéticos*, *Proliferación y diferenciación de líneas de células madre*, *Caracterización y diferenciación de células progenitoras* y *Células progenitoras neuronales*. **Conclusion:** la metodología planteada permite una delimitación temática más precisa que la que ofrecen las propias bases de datos y aplicable a cualquier dominio científico.

Keywords: Bibliometry, Co-word analysis, Stem cells

Spanish stem cell research (1997-2007). Bibliometric analysis and main lines of research through KeyWords Plus

M. A. Zulueta¹, G. Cantos-Mateos¹, B. Vargas-Quesada² y Z. Chinchilla-Rodríguez³

1ma.zulueta@uah.es

Information Science Faculty. University of Alcalá., C/ San Cirilo s/n, 28804, Alcalá de Henares, Madrid (Spain)

1gisela.cantos@gmail.com

Information Science Faculty. University of Alcalá., C/ San Cirilo s/n, 28804, Alcalá de Henares, Madrid (Spain)

2benjamin@ugr.es

Communication and Information Science Faculty, University of Granada, Colegio Máximo de Cartuja s/n, 18071, Granada (Spain)

(CSIC), Unidad Asociada Grupo SCImago, Madrid (Spain)

³ zaida.chinchilla@scimago.es

(CSIC), Unidad Asociada Grupo SCImago, Madrid (Spain)

Abstract

Introduction: The Spanish stem cell research is one of the most influential biomedical sciences. The usefulness of bibliometric studies to analyze scientific domains is well established. The aim of this paper is conducting a research study on this subject during the period 1997-2007 and the implementation of a adequate methodology for the detection of the main lines of research. **Material and Methods:** The documents come from the SCI database by searching for the terms stem * cell * in the topic field. The analysis performed has been directed, first to obtain bibliometric indicators and secondly, to study the co-occurrence of KeyWords Plus contained in documents. **Results and Discussion:** We had obtained 2369 documents mainly from the Autonomous Communities of Cataluña, Madrid and Valencia and limited to the health sector, principally. The most productive institutions were the hospital Santa Cruz and San Pablo (Barcelona), the Clinic of Barcelona and the Clinic of Salamanca. The authors are highlighted J.Sierra, R. Martino and J. F. San Miguel. Bone Marrow Transplantation, Blood and Haematologica were the most productive journals and Hematology, Oncology and Neurosciences the most frequent categories. The analysis of the relations of KeyWords Plus identified four main lines of research: Transplantation of hematopoietic progenitors, proliferation and differentiation of stem cell lines, characterization and differentiation of progenitor cells and neural progenitor cells. **Conclusion:** The proposed methodology allows a more precise delimitation issue than is offered their own databases and apply any scientific domain.

Keywords: Bibliometry, Co-word analysis, Stem cells

PUBLICACIÓN Nº 3

Cantos-Mateos, Gisela, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida, Zulueta, María-Ángeles (2012). Stem cell research: bibliometric analysis of main research areas through KeyWords Plus, *Aslib Proceedings*, 64 (6), 561-590.
<http://dx.doi.org/10.1108/00012531211281698>



Aslib Proceedings

Emerald Article: Stem cell research: bibliometric analysis of main research areas through KeyWords Plus

G. Cantos-Mateos, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, M.A. Zulueta

Article information:

To cite this document: G. Cantos-Mateos, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, M.A. Zulueta, (2012), "Stem cell research: bibliometric analysis of main research areas through KeyWords Plus", Aslib Proceedings, Vol. 64 Iss: 6 pp. 561 - 590

Permanent link to this document:

<http://dx.doi.org/10.1108/00012531211281698>

Downloaded on: 27-11-2012

References: This document contains references to 86 other documents

To copy this document: permissions@emeraldinsight.com

Access to this document was granted through an Emerald subscription provided by UNIVERSIDAD DE ALCALA

For Authors:

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service. Information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

With over forty years' experience, Emerald Group Publishing is a leading independent publisher of global research with impact in business, society, public policy and education. In total, Emerald publishes over 275 journals and more than 130 book series, as well as an extensive range of online products and services. Emerald is both COUNTER 3 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.



Stem cell research: bibliometric analysis of main research areas through KeyWords Plus

Stem cell
research

561

G. Cantos-Mateos, B. Vargas-Quesada, Z. Chinchilla-Rodríguez and
M.A. Zulueta

(Author affiliations can be found at the end of the article.)

Received 22 July 2011
Revised 6 October 2011
Accepted 16 October 2011

Abstract

Purpose – Research with stem cells is a biomedical venture with great scientific impact, and whose development flows over into many other areas. This article aims to present a dual analysis of Spain's scientific output in this field during the period 1997-2007.

Design/methodology/approach – The authors used bibliometric indicators of a basic nature as well as techniques for the visualization and analysis of networks of scientific information based on a study of KeyWords Plus.

Findings – The output is mainly concentrated in Cataluña and Madrid, and hospitals are the most productive centres (followed by health institutes), where the main authors are affiliated. Main categories are hematology, oncology and biophysics. The outstanding areas of study revolve around the therapeutic use of transplant of hematopoietic progenitors, the processes of generation, proliferation and differentiation of lines of cells, and the study of neurosciences.

Originality/value – This study provides an overview of Spanish research involving stem cells, detecting and representing the main areas of research. The article considers the potential of KeyWords Plus in combination with the proposed methodology as particularly useful for the analysis and delimitation of a scientific domain.

Keywords Stem cells, Spain, Co-word analysis, Bibliometric analysis, Information visualization, Research work, Physical sciences

Paper type Research paper

1. Introduction

Research into stem cells is a biomedical field of great expectations. It was in 1949 when scientist J. Hammond (1949) discovered the method to maintain mouse embryos in culture *in vitro*. Since, research with stem cells has developed to the point where it holds very hopeful perspectives for the treatment of thus far incurable diseases. At present, research is mainly oriented toward developing new therapies for hematological, cardiovascular, neurodegenerative and genetic diseases, as well as cancer and diabetes, among others (Martínez Serrano and Bjorklund, 1996; Bishop *et al.*, 2002; Cao *et al.*, 2002; Di Giorgio *et al.*, 2007). Its intrinsic nature makes stem cell research transcend to other fields as diverse as politics, ethics, culture, and law, placing it in the arena of social controversy.

The authors thank Jean Sanders for translating and editing the text. This study was made possible by funding from the project CAM-UAH CCG07-UAH/HUM-1906. G. Cantos-Mateos received a grant for Introduction to Research under this project. At present, she has a pre-doctoral grant (JAE-Predoc of Consejo Superior de Investigaciones Científicas of Ministerio de Ciencia e Innovación), cofinanced by the Fondo Social Europeo.

Aslib Proceedings: New Information
Perspectives
Vol. 64 No. 6, 2012
pp. 561-590
© Emerald Group Publishing Limited
0001-253X
DOI 10.1108/00012531211281698



The use of bibliometric studies to comprehend and analyse scientific domains (Hjørland and Albrechtsen, 1995), together with the development and fine-tuning of new techniques and tools, facilitates decision-making in areas of scientific policy and reflects the “state of the art” of research at a given time. These processes, necessary for the evaluation of science (Camí *et al.*, 1997; Bordons and Zulueta, 1999) are a responsibility that no country can elude (Krauskopf, 2000) given the evident connections between advancement through research activity, economic growth and progress, and the enhanced well-being of society (Chinchilla-Rodríguez and Moya-Anegón, 2007).

The number of scientific disciplines interrelated by stem cell research lends it an interesting yet complicated character (Zhao and Strotmann, 2011). Its interdisciplinarity presents a great challenge when delimiting and analyzing its thematic composition, demanding a very precise analysis. Precisely to face this challenge, bibliometrics has complementary tools that more recently include social network analysis (Perianes-Rodríguez *et al.*, 2011) and the visualization of scientific domains (Wasserman and Faust, 1998; Boyack *et al.*, 2009; Leydesdorff and Rafols, 2009; Rafols *et al.*, 2010; Aharony, 2010; Chinchilla-Rodríguez *et al.*, 2010; Vargas-Quesada *et al.*, 2010).

When searching for reliable information, studies of this nature have traditionally used bibliographic databases as the most usual source (Fernández *et al.*, 1993). Though not designed for this purpose, the information they offer has demonstrated its bibliometric potential. However, they present limitations when resolving specific matters, for example the precise definition and delimitation of a topic. It is therefore essential to reveal the semantic structure established among documents by means of the bibliographic information contained in a database.

Overcoming such limitations calls for analytical methods that will allow us to arrive at the ideas and concepts that ultimately sustain the scientific discourse expressed in the literature. To this end, units of analysis smaller than thematic categories or journals may be used: namely, descriptors. We opted to use KeyWords Plus (KW +) based on the hypothesis that they could better reflect the conceptual essence of each document. These descriptors are automatically generated by the SCI from an algorithm that extracts key words from all the titles referenced or cited in the documents (Garfield, 1990; Garfield and Sher, 1993).

2. Literature review

There is a great diversity of methods and techniques that may be used for delimiting and visualizing a scientific domain in order to detect research areas or subareas. The array includes author cocitation (ACA) (White and Griffith, 1981; White and McCain, 1998; Small and Upham, 2009; Chen *et al.*, 2010; Zhao and Strotmann, 2011), bibliographic coupling analysis (Kessler, 1963; Boyack and Klavans, 2010; Chen *et al.*, 2011; Glänzel and Thijs, 2011), co-word studies (see the following) or hybrid studies combining ACA and co-words (Zitt and Bassecoulard, 1996; Zitt and Bassecoulard, 2006, Zitt *et al.*, 2008, 2011). Still other authors have demonstrated how the results of structural analysis and research performance assessment of a given research field can be enriched by combining elements of both into one integrated analysis (Noyons *et al.*, 1999; Cobo *et al.*, 2011). In view of the advantages and drawbacks that condition the choice of one methodology over another, for the present study we opted to focus on co-word analysis. According to

Michel Zitt and collaborators (Zitt *et al.*, 2011), if words seem better adapted to “science alive”, it is not because citations cannot describe recent science, but rather because words are universal. Not confined to scientific and kindred databases, linguistic markers are more able to reflect the interweaving of scientific, social, and political contexts, typical of emerging and controversial areas – at the expense of some “translation” effort. Furthermore, they are considered as a content analysis technique that is effective in mapping the strength of association between information items in textual data. It deals directly with sets of terms shared by documents, mapping the pertinent literature directly from the interactions of key terms (Cobo *et al.*, 2011).

The application and development of this type of analysis is nothing new. In the 1980s there were attempts to reveal the structure of science and its evolution through word co-occurrence. This marked a relationship between the concepts that documents represent (Cambrosio *et al.*, 1993; Courtial *et al.*, 1994) and made manifest the structure and trends of a scientific discipline in view of the strength of associations among the representative terms in the published literature (Ding *et al.*, 2001).

Authors such as Van Raan and Tijssen (1993) applied it to research on neural networks, concluding that there is an epistemological value that allows us to discover unsuspected relations among the concepts of a discipline besides revealing problems that otherwise might go undetected. Ding *et al.* (2001) used it to map the intellectual structure of the field of information retrieval, showing patterns and trends within. Onyancha and Ocholla, 2005, used it to help researchers and project managers identify new research lines in HIV/AIDS, as well as study the links established, to better plan research and formulate adequate scientific policies. Bessalar and Heimeriks (2006) used it to study the publications in *Information Science* that came out between 1986 and 2002 by means of the most relevant words in titles and references. Chalik and Jirina (2006) related it with Price’s cumulative law of science as applied to Economics. In turn, Lee (2008) used this model to identify trends and underlying aspects in the area of Information Security, and Neff and Corley (2009) applied it to the study of publications in the area of Ecology from 1970 to 2005. More recently Leydesdorff and Welbers (2011) focuses on co-word analysis in relation with the semantic measures of similarity patterns (correlations) and latent variables (analysis factor) using computational and statistical techniques. And Zulueta *et al.* (2011) use it to study publications involving health and women as recorded in Medline from 1965 to 2005, combining network visualization techniques and factor analysis.

Despite such a proliferation of studies, and such a broad area of application, there is very little evidence of preceding studies that resort to KW + and that systematically undertakes analysis of the topic at hand, stem cell research. Maybe for the limitations pointed out by Zhao and Strotmann (2011) who considered that the traditional core journal or keyword search based methods for collecting data using existing citation databases, especially the ISI data source, do not work well. This is particularly complicated in the case of journals, because papers on stem cell research are scattered over numerous journals rooted in very different research areas, from biology to social sciences and law. However, we encountered an analytical endeavor using keywords that could be related with our study and serve as a baseline for it. Qin (2000) uses KW + to compare differences among the descriptors of the SCI obtained by automated indexing and the MeSH terms obtained by manual indexing using a controlled language. In later works (Chiau and Ho, 2007; Ho, 2007), a brief recount of

the Author KeyWords (AKW) is incorporated as a complement to bibliometric analysis, overlooking KW + . Li *et al.* (2009) complement their bibliometric analysis with AKW as well as KW + , whereas Su and Lee (2010) analyze documents pertaining to the field of “Technological Prospective” registered in the Web of Science through a topographic representation of AKWs alone. From the bibliometric standpoint in the stem cell research field, Ho *et al.* (2003) tried to gauge the scientific output of Asia, focusing on Hong Kong, Singapore, South Korea and Taiwan. Li *et al.* (2009) carried out a more recent and comprehensive analysis, appraising world output from 1991 to 2006. The methodological approach of Leydesdorff and Hellsten (2005, 2006) involved using the term “stem cell” in a co-word study as an element of analysis to show differences in its use and measure its significance in diverse contexts.

3. Research questions

This paper aims to answer a number of questions about Spanish research surrounding Stem Cells:

- RQ1. What is the overall structure of stem cell research field? Does this structure allow us to characterize the field or extract some sort of conclusion from a thematic standpoint?
- RQ2. Do co-word studies applied to categories supply complementary information? Are they suitable for fine-grained studies such as this? Could they be considered adequate units of analysis for thematic delimitation at the document level?
- RQ3. Are Kw + better measures of analysis than subject categories? Do the two provide the same information? Can they be considered adequate units of analysis for the thematic delimitation at the document level?

The present contribution is a general research performance assessment of a research field enriched by combining elements of structural analysis into one integrated analysis of Spain's research with stem cells. Our aim is to detect and depict the main lines of research and actors as institution and authors as well as the main patterns of publication. To this end, we compare category analysis and the analysis of KW + , weighing the information and the potential that each offer. This study is part of a broader and deeper research effort and the subject matter of a doctoral thesis. It constitutes the first stage in which we propose and develop a methodology for thematic delimitation that allows one to detect the conceptual relations established among documents. Future efforts would be dedicated to testing this model using other bibliographic information sources and alternative indexing languages.

In the next section we describe the material and methodology used. Section 4 shows the results in terms of the type of analysis applied: bibliometric or thematic delimitation. Finally, section 5 provides discussion of the results obtained.

4. Research design

4.1 Data collection

4.1.1 *Sources.* The original information sources were the Science Citation Index (SCI) database and the Journal Citation Report (JCR) of Thomson Reuters, where we extracted information about scientific output and the impact factor of the journals

where Spanish researchers had published their findings in the respective field of study. The data, available thanks to the access license contracted through the Fundación Española de Ciencia y Tecnología (Fecyt) as a public service for all public research institutions, was gathered from January 2004 onwards.

These databases have come to play a crucial role in many countries in the sense that they lend added value to particular basic parameters for the calculation of bibliometric indicators that are key for evaluative and comparative analyses at an international or a multidisciplinary level (Chinchilla-Rodríguez, 2005; Chinchilla-Rodríguez and Moya-Anegón, 2007; Moya-Anegón *et al.*, 2007a, b).

The fact that they have a different structure for recording the institutional affiliation of all the undersigning authors of a study makes them most appropriate for discerning scientific collaboration (not all databases collected institutional affiliation of all authors). Although there are certain drawbacks owing to a lack of standardization, for the purposes of this study we undertook a thorough effort to reduce any possible errors in author-institution ascription. At the same time, the value placed on the reference or bibliographic citation facilitates the development of indicators that may be representative of visibility and or the underlying scientific structure.

These databases have traditionally been criticized for a coverage bias reportedly involving disciplinarity and nationality. Yet studies that compare the coverage of the SCI with that of Ulrich's International Periodicals Directory (U-S&T) demonstrate that, on the whole, the SCI-JCR journals present a good balance with respect to the U-S&T when dealing with countries and disciplines. In contrast with widespread claims, there is no bias in favor of countries such as the USA, or areas such as Biomedicine; in some cases, under-representation is even seen (Braun *et al.*, 2000). Since the aim was to ensure global coverage of a discipline and have a very homogeneous point of reference for comparison, this source was considered optimal. Moreover, because analysis was to be done at a national level (comparing the output of autonomous communities within Spain), without excessive detail in terms of specialized fields, the source was held to be fully appropriate for the study at hand.

In peripheral countries, the utilization of these sources to evaluate scientific activity is not widely supported by the scientific communities. Some voices say that the current system penalizes researchers who publish in Spanish-language journals or journals published in other (non-English) languages; and that research lines of regional interest hardly find a niche in the journals of high impact (Sanz-Casado *et al.*, 1995; García-Guinea and Ruis, 1998; Kyvik, 2003), which may have led to a wave of "migration" from domestic journals toward the English-language journals harbored in the ISI (Jiménez-Contreras *et al.*, 2001).

In this specific case, focusing on Spanish research studies, the source selection follows current regulations[1] established by the Comisión Nacional Evaluadora (CNEAI) regarding criteria for the system of research incentives, in which the journals registered by the former ISI are consulted as a reference in the evaluative processes for assigning research funds or incentives (Jiménez Contreras *et al.*, 2003).

In the end, and despite the possible drawbacks just mentioned, these sources are the *de facto* reference used by very numerous governmental institutions. This is certainly true of Spain (Chinchilla-Rodríguez and Moya-Anegón, 2007). We therefore consider that the data source used here, and the raw data supplied, provide a realistic reflection of Spain's internationally visible research over the ten-year period of analysis.

4.1.2 *Search strategy.* The search strategy consisted of using the term “stem* cell*” in the field topic, delimited by “Spain” in the field address and parameterized by the time period (1997-2007). The search in the field topic allowed us to locate terms or phrases in specific parts of the document such as title, abstract, AKW and KW + . In April 2008 we finished downloading the records into an *ad hoc* relational database. The total number of documents retrieved was 2,467.

4.1.3 *Depuration and standardization of data.* The ISI database presents certain disadvantages for bibliometric applications. For this reason it was necessary to standardize and purge the data, in particular the authors’ and institutional names, to suitably analyze and mine the data. The address field usually comprises four levels: main organization, department within the organization, city and country. In many cases, only three levels are listed, excluding the department or institutional level. The country is generally highly standardized, and the city may be standardized using postal codes. Many variations can be found at all these levels. The procedure followed involved locating the variations in each institution’s name, choosing an acceptable denomination that encompassed them all, and attributing it to the respective region semi-automatically. Standardization was then carried out manually for the authors as well as the Spanish addresses (see Table I).

In total, 98 documents were eliminated for one reason or another – the date of publication did not match our period of study, documents were mistakenly attributed to the domain of Spain by homonymy with the name of the institution, there were cases of record duplication, etc. In the end, the number of documents used for this study was 2,369.

4.2 Data analysis

4.2.1 *Bibliometric indicators.* In order to analyze the study about stem cells from different perspectives, we adopted two separate approaches. First we carried out a basic bibliometric study of a descriptive nature, through which we could generally appraise the state of research in the area and characterize the main actors at meso and micro levels. The indicators are: N Documents is the number of papers. Ncol is the total number of co-authored papers. Co-authorship index is the mean authorship per document. Productivity is the number of documents per author. Naut is the number of authors per year. Impact Factor is a measure reflecting the average number of citations to articles published in science and social science journals and is calculated yearly for those journals that are indexed in Thomson Reuters Journal Citation Reports. International collaboration is the number of papers co-authored with foreign countries; and finally, Institutional Collaboration is the number of papers in which two or more Spanish institutions participated.

4.2.2 *Representation of information and thematic analysis.* The generation of the maps requires the sequential application of several steps. We follow the sequence proposed by Börner *et al.* (2003), explained in the subsections which follow.

Table I.
Standardization of data

	Not standardized	Standardized
Authors	7,809	7,397
Spanish address	5,121	4,935

4.2.2.1 Units of analysis. The most common units for mapping literatures are categories, journals, documents, authors, and descriptive terms or words. Each presents different facets of a domain and enables different types of analysis. For instance, a map of categories can be used to obtain a macro view of science, showing the relative positions and relationships between disciplines. Journal maps are used to show fine distinctions within a discipline. Documents are used to visualize a knowledge domain for a variety of purposes, including document retrieval or domain analysis. Authors are also relatively common as units of analysis and occur in two main forms: author co-citation, typically used to infer the intellectual structure of a field, and co-authorship, used to show the social network of a discipline, institution, etc. or to reveal international collaboration. Words are the last link in the chain, used to shed light on the cognitive structure of a field through semantic maps often referred to as co-word analysis.

We undertook a thematic delimitation by means of distributive and network analysis, using as units of analysis the Journal Citation Report categories and the KW + assigned to each one of the documents retrieved.

4.2.2.2 Unit of measure. A number of measures may be used to quantify the way that units of analysis are mutually related, the most common ones being citation, coupling, co-citation or co-occurrence. Since one of our objectives was to detect main research areas, we adopted co-occurrence as the unit of measure. It is widely accepted for obtaining relational information from the documents within a particular domain. Indeed, in the case of KW + , there is no another viable option.

4.2.2.3 Dimensionality reduction. We have take into account the fact that the networks resulting from co-occurrence analysis are usually dense. Hence, the Pathfinder algorithm (Moya-Anegón *et al.*, 2007a, b) is applied to the co-occurrence matrix to prune the network, parameterized to $r = \infty$ and $q = n - 1$. In conjunction with raw data and algorithm type spring embedders, it helps to preserve and highlight the salient relationships between categories or KW + (White, 2003; Vargas-Quesada *et al.*, 2008; Moya-Anegón *et al.*, 2009), capturing, identifying and delimiting the main research areas of a domain.

The final result was a square matrix of $N \times N$ elements, where N stands for the subject categories or KW + to be represented.

4.2.2.4 Information visualization. For visualization of the data, two programs were used: Pajek (Batagelj and Mrvar, 2010) and VOSviewer (Van Eck and Waltman, 2010).

The first one is free, widely available software for the visualization and analysis of large networks. It has been used to create maps of categories (the 29 most frequent ones) and KW + (the 102 most frequent KW +) via the Kamada-Kawai algorithm (Kamada and Kawai, 1989), which is included in Pajek and is one of the most broadly used methods to perform this task.

The second one is another example of recent, free software for the representation and analysis of information, which appears as an alternative to the traditional techniques of multidimensional representation and network display. Vosviewer combines visualization and clustering techniques, favoring analysis while bypassing unnecessary technical complications. Of the four types of representations that it offers, we chose to use the so-called “heat maps” (or density maps) to display the total KW + (6.199 KW +). This type of map is especially useful for detecting the main nuclei of interaction of a domain. In it, each descriptor is represented by a tag, whose

size is proportional to its weight. Each point on the map has a color that reflects the density of its descriptors, from red to blue. The greater the density, or co-occurrence among the KW + , the redder the display; whereas being closer to blue indicates a greater dispersion and therefore less co-occurrence. VOSviewer is based on multidimensional scaling for the spatial distribution of information, for which reason the positions of the descriptors – but not the foci – differ from other programs, such as Pajek, which uses the Kamada-Kawai algorithm for layout.

5. Results and discussion

In the following we present our results according to the type of study applied: the general data obtained by bibliometric analysis, and the results related with the thematic delimitation obtained by analyzing the categories and KW + . In either case, the body of documents we worked with was the same, numbering 2,369.

5.1 Bibliometric analysis

5.1.1 Output trends. There is a clear rise in scientific output involving stem cell research over the period studied (see Figure 1). The thresholds of production we encountered are 94 documents for the year 1997, and 395 for 2007, meaning a gross increase of over 300 per cent in the decade. Noteworthy is the period from the year 2000 to 2003, during which output stabilized. The year 1999 shows the highest rate of production, and 2003 would be the lowest. Such fluctuations in national output could be affected by a number of factors such as the amount of congresses held, technological advances, levels of international collaboration, and political policy or specific measures applied to the area of stem cell research. While true that we do not have all the information needed to identify the causes for this output trend, some data and observations could shed light on the aforementioned factors and therefore on the evolution of scientific production related to stem cells.

5.1.2 Distribution by autonomous communities of Spain. Out of the 17 regional “Autonomous Communities” (AC) into which Spain is divided, there are five that contributed most substantially to output, according to author affiliation (see Figure 2);

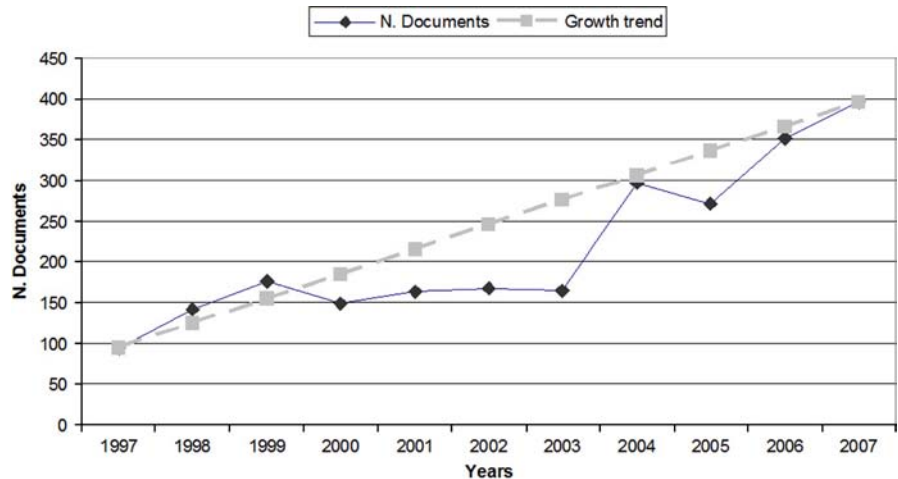


Figure 1.
Evolution of output trend

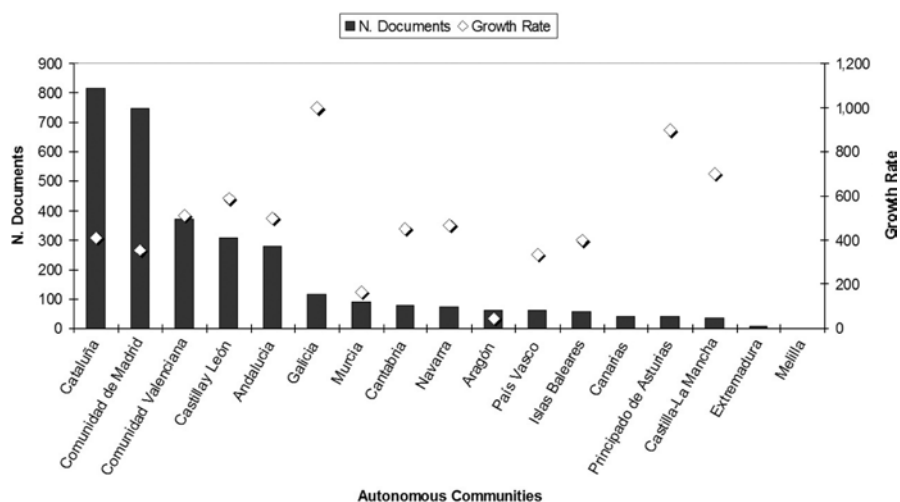


Figure 2.
Distribution of output and
growth rate by AC

they are Cataluña, Madrid, Comunidad Valenciana, Castilla y León and Andalucía. The first two stand out way above the rest, with well over half of total production (67.2 per cent). Comunidad Valenciana has a share of 16 per cent, while Castilla y León and Andalucía both contribute roughly 13 per cent to total output, at somewhat of a distance from the rest of the AC.

Of the 644 institutions where national output originated, we see in Table II that the top 20 produce 50 publications in the period of study. Cataluña has two main sources of publications, the Hospital de Santa Cruz y San Pablo, with nearly 12 per cent of output,

Institutions	No of documents	Per cent documents	Growth rate
Hosp. Sta. Cruz y San Pablo (Barcelona)	283	11.95	192
Hosp. Clin. De Barcelona.	197	8.32	900
Hosp. Clin. (Salamanca)	194	8.19	1,000
Hosp. La Fe (Valencia)	130	5.49	767
IDIBAPS (Barcelona)	115	4.85	
Hosp. Univ. La Princesa (Madrid)	95	4.01	400
Hosp. Gregorio Marañón (Madrid)	87	3.67	
Hosp. La Paz (Madrid)	86	3.63	600
Hosp. Vall del Hebrón (Barcelona)	84	3.55	600
Hosp. Ramón y Cajal (Madrid)	77	3.25	600
Hosp. 12 Octubre (Madrid)	76	3.21	533
Hosp. Clin. (Valencia)	74	3.12	
Ctr. Inv. Energet. Medioamb. Tecnol. (Madrid)	68	2.87	100
Hosp. Marques de Valdecilla (Cantabria)	67	2.83	1,000
Hosp. Niño Jesús (Madrid)	65	2.74	200
Inst. Catalán de Oncología (Barcelona)	57	2.41	0
Clin. Univ. Navarra (Navarra)	56	2.36	333
Ctr. Inv. del Cáncer (Salamanca)	54	2.28	350
Hosp. Reina Sofía (Córdoba)	54	2.28	
Hosp. Germans Triás Pujol (Barcelona)	51	2.15	1,400

Table II.
Distribution of
institutional output

and the Hospital Clínico de Barcelona and the Hospital Clínico de Salamanca, with over 8 per cent. Following them, the most productive institution would be the Hospital la Fe de Valencia, with its roughly 5.5 per cent contribution to output.

Also deserving mention is the output from the Instituto de Investigaciones Biomédicas August Pi i Sunyer of Barcelona (IDIBAPS). This “mixed institution” of administrative multi-dependence is dependent on the Hospital Clínico de Barcelona, Universidad de Barcelona, Inst. de Invest. Biomédicas de Barcelona del CSIC (IIBB-CSIC) and the Generalitat de Cataluña (Department of Economía y Conocimiento (DEC), and is physically located in the Hospital Clínico de Barcelona.

Table III shows how most of Spain’s research with stem cells is carried out in the Health Services. Approximately 64 per cent of the publications can be traced to health institutions, which is twice as much as the output from the university sector, second in production. Other productive institutions are those that depend on the Administration of the Autonomous Communities (7.34 per cent), the mixed institutions of the CSIC-Higher Education (6.37 per cent) and the Consejo Superior de Investigaciones Científicas (CSIC) itself (5.61 per cent). Documents from the Central Government or Administración Central (3.63 per cent) correspond almost exclusively to the Instituto de Salud Carlos III, which at that time was governed under the Ministerio de Sanidad (Ministry of Health).

5.1.3 Most productive authors. Table IV shows, in descending order, the percentage of documents that the authors participated in, along with the rate of growth over the period on the whole. The most productive authors, J. Sierra, J.F. San Miguel and R. Martino, have output that surpasses 120 documents.

The total sum of all authors signing documents comes to 18,571. This figure breaks down to a total of 7,397 actual individual authors, since many authors sign more than one paper. As some studies have reported (Newman, 2001), this field presents a very high collaboration rate. The mean number of authors per paper in the biomedical fields focused in Medline database was 3.75 and concretely in cardiovascular subfield 6.18 (Bordons *et al.*, 1996). However, our data show that during the period of study (see Figure 3), the mean authorship per document (co-authorship index) was higher with a rate of 7.8; this figure was quite stable over time (line with squares). This represents a

Institutional sector	No. of documents	Per cent documents	Growth rate
Health	1,515	63.95	282
Higher education	692	29.21	337
Public administration	174	7.34	1,700
CSIC-higher education	151	6.37	540
Consejo Superior de Investigaciones Científicas CSIC	133	5.61	833
Multi-dependent mixed institutions	121	5.11	
Central administration	86	3.63	600
Private	29	1.22	
CSIC-higher education (local)	15	0.63	
State-owned companies	8	0.34	
Foundations	2	0.08	
Local administration	1	0.04	
CSIC-autonomous communities	1	0.04	

Table III.
Distribution and growth of output by institutional sector

Author	No. of documents	Per cent documents	Growth rate
Sierra, J.	155	6.54	80
San Miguel, J.F.	143	6.04	1,150
Martino, R.	140	5.91	67
Caballero, M.D.	126	5.32	33
Sureda, A.	120	5.07	137
Carreras, E.	112	4.73	150
Brunet, S.	100	4.22	-11
Montserrat, E.	98	4.14	120
Urbano-Ispizua, A.	94	3.97	150
Rovira, M.	83	3.50	125
Perez-Simon, J.A.	72	3.04	1,000
De La Rubia, J.	64	2.70	1,300
Ribera, J.M.	60	2.53	
Del Canizo, M.C.	58	2.45	600
Lahuerta, J.J.	56	2.36	250

Table IV.
Most productive authors

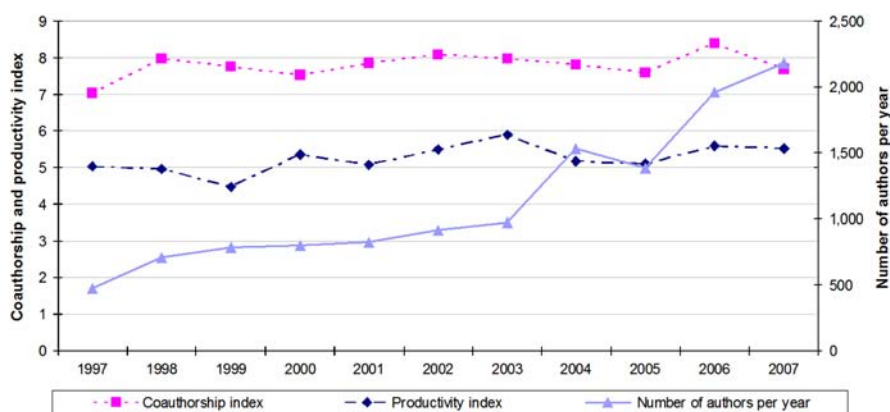


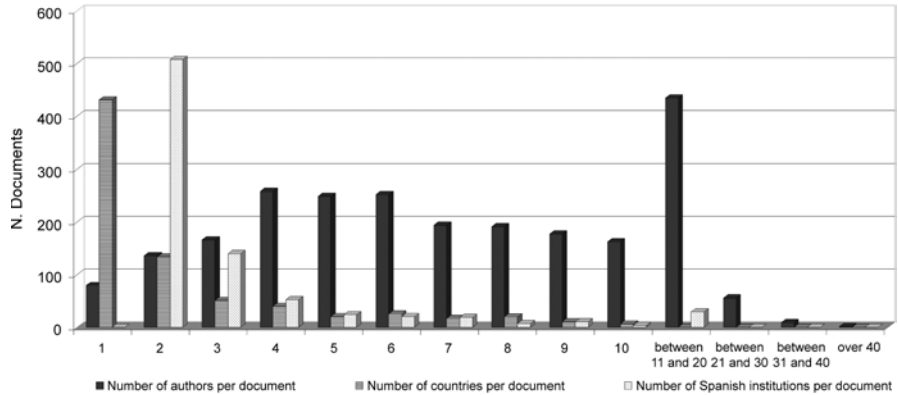
Figure 3.
Index of co-authorship,
productivity and evolution
of the number of authors
by year

strong pattern of coauthorship, even higher than the global of the biomedical subfields as Cronin pointed out at the beginning of 2000 (Cronin, 2001), and this rate is similar to that found in the study of Zhao and Strotmann (2011).

Another indicator associated with productivity is the number of documents per author (dotted line with diamonds), which is seen to also be fairly stable over time, with a mean of five documents per author (dotted line with squares), though there is a slight increase in the middle of the study period. In contrast, the trend of authors per year (line with triangles) is on the rise, with a noteworthy surge in the year 2004.

The co-authorship and productivity indicators reflect a coherent relation and a stable trend within the period of study, and trace a logical evolution towards an increase in co-authored documents at the expense of those put out by a single author. In fact, the transitory index, or authors with just one document, is 69 per cent. In Figure 4 we find the distribution of the number of documents at each level of aggregation for the authors, Spanish institutions and foreign countries. At the micro level, the documents signed by four, five or six authors constitute the bulk of output, which is coherent with

Figure 4.
Percentage of documents signed in collaboration, according to number of participating actors



the average number of authors per document that we saw in the previous figure. Therefore, we may affirm that the number of co-authored documents is considerably greater than the number of documents with a single author. Multiple authorship (more than two authors) is particularly prevalent, with a great proportion of documents signed by 11 to 20 authors.

With respect to the number of institutions participating, we see that over 20 per cent of output is signed by two institutions. Similarly, international collaboration amounts to roughly 18 per cent of output signed by another country in conjunction with Spain. Trilateral or multilateral collaboration on the institutional or international level is very scarce. Therefore, when there is institutional and international collaboration, studies of a bilateral origin prevail over those of a trilateral or quadrilateral nature.

5.1.4 Patterns of collaboration. Analysis of the different types of collaboration (see Figure 5) reveals patterns that tend to match those observed at the national or international level. As the period advances, the documents involving no collaboration descend, while the documents undersigned by more than one institution (national or international) are on the rise. These rates are higher than the ones reported by Zhao

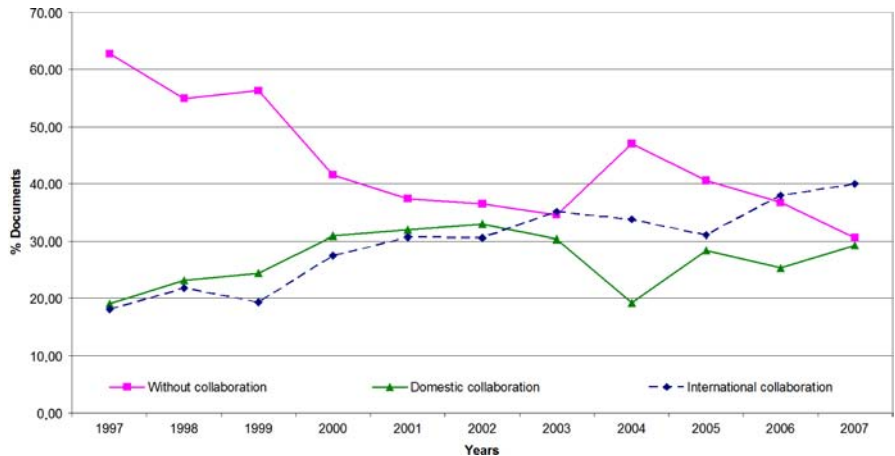


Figure 5.
Types of collaboration

and Strotmann for the period 2004-2009. This might be due to differences between periods and sources. In any case, it would be desirable correct this tendency to avoid fluctuations over time and attain reasonable rates in accordance with international standards and patterns (less than 10 per cent), bearing in mind the positive correlation between institutional and international collaboration and the high number of citations of collaborating authors as compared to single authorship (Narin *et al.*, 1991, Chinchilla-Rodriguez *et al.*, 2010)

There is a remarkable tendency in the year 2004 for documents in collaboration with Spanish institutions to drop. Likewise, international collaboration decreases considerably the following year, 2005. Yet if we place these findings in the context of more authors per year, as reflected in Figure 3, one may infer greater interdepartmental collaboration is underway. Further study, at the level of institutional desegregation, would be necessary to confirm this inferred behaviour.

Therefore, if we relate these collaboration indicators with the overall output trend depicted in Figure 1, it is difficult to attribute the increase in publications detected in 2004 to such causes.

In the period overall, 31.95 per cent of the documents analysed correspond to international collaborative efforts. The main countries participating are the USA, Germany, Great Britain, France, and Italy, followed by production involving modest but important countries in the area of stem cell research as can be seen in Table V. There are noteworthy rates of growth in the two main countries of association, followed by Switzerland and Canada.

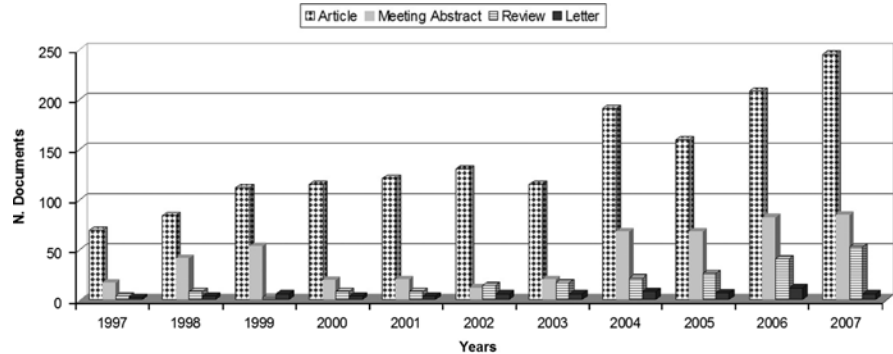
5.1.5 Document type. The predominating document type is the research article (1,555 papers), followed by congress proceedings, as seen in Figure 6. Deserving mention is the high number of reviews on the subject of stem cell research. In the period studied, 95 per cent of output was in the form of a scientific article (over 65 per cent), congress proceedings (20 per cent) and reviews (8.82 per cent). Over the years of study, however, there are some variations in the production by type of document: the growing number of congress proceedings and the reviews contribute to a relative decrease in the percentage of original research articles. An increase in the latter is seen over the period of study, especially after 2004. Regardless, this finding cannot be singled out as the cause of the fluctuations in total output observed in Figure 1, as the number of publications in the article format also adopts a growing trend.

As the patterns of collaboration and the document type do not, in themselves, explain the increase in total output seen in the year 2004, we also must consider that

Country	No. of documents	Per cent documents	Growth rate
The USA	304	40.43	1,220
Germany	183	24.34	1,450
The UK	178	23.67	900
France	174	23.14	420
Italy	156	20.74	833
Netherlands	109	14.49	800
Sweden	69	9.18	800
Switzerland	69	9.18	1,000
Belgium	52	6.91	175
Canada	48	6.38	1,200

Table V.
Main countries associated
in Spain's international
collaboration and output
in stem cell research

Figure 6.
Document type



legislative reforms undertaken by the Spanish institutions could have implications for research, propelling it toward new and diversified fields, enriching the thematic profile of this scientific domain. This explanation is quite probable, as in recent years stem cell research has generated much debate in nearly all the sectors involved in biomedical research and development. For instance, at the legislative level, key reforms came about such as the Law of Assisted Reproduction on 21 November 2003, by virtue of which the “Commission for the Follow-up and Control of the Donation of Human Cells and Tissues” was created, which in turn authorized, in 2005, the very first research projects with embryonic stem cells. This law was derogated by the one in force, from 26 May 2006; among other things, it authorized the performance of autologous transplants from umbilical cord blood, allowing for the creation of private blood and cell banks. Another recent example is Law 14/2007 (3 July 2007) on Biomedical Research, which authorizes therapeutic cloning and research with embryos.

5.1.6 Journals of publication. The total number of journals where the documents were published came to 610. Outstanding among them is the output in Bone Marrow Transplantation and Blood, where practically 25 per cent of the total appears. The rest are at a great distance in terms of the volume of production, and 97.7 per cent of the journals show fewer than 20 publications involving stem cell research and Spain. Table VI shows the journals that put out more than ten documents during the study period, accounting for just over half of production.

We add a column with the Impact Factor for the year 2006 to orient the reader, as it is impossible to compare journals on an equal basis, and so to highlight the amount of output in view of each journal’s visibility in terms of the IF. Thus we see that one of the most visible journals in the area surrounding stem cell research is second in production.

5.2 Thematic analysis

5.2.1 Analysis of subject categories. Altogether, the documents studied were published in journals classified under 110 different categories. The category Hematology stands out with its 997 documents, which account for 42 per cent of the total. Next in line would be: Oncology, Biophysics, Neurosciences, Cell Biology, Biochemistry and Molecular Biology and Immunology. We would add that a major part of this production comes from the documents appearing in the journals Bone Marrow Transplantation (Hematology, Oncology and Biophysics) and Blood (Hematology) where the majority of these documents are concentrated and as shown in Table VII.

Journal	No of. documents	Per cent documents	Growth rate	Impact factor 2006
<i>Bone Marrow Transplantation</i>	328	13.85	80	2.62
<i>Blood</i>	249	10.51	50	10.37
<i>Haematologica</i>	75	3.17	- 100	
<i>British Journal of Haematology</i>	56	2.36	- 33	4.50
<i>Medicina Clinica</i>	44	1.86	- 50	1.33
<i>Haematologica-the Hematology Journal</i>	40	1.69	380	5.03
<i>Leukemia</i>	33	1.39	800	6.15
<i>Annals of Oncology</i>	31	1.31	0	5.18
<i>Experimental Hematology</i>	31	1.31	200	3.41
<i>Biology of Blood and Marrow Transplantation</i>	29	1.22	400	3.46
<i>Journal of Neuroscience</i>	26	1.10	200	7.45
<i>European Journal of Cancer</i>	24	1.01	- 86	4.17
<i>Journal of Clinical Oncology</i>	21	0.89	100	13.60
<i>European Journal of Neuroscience</i>	20	0.84	0	3.71
<i>Journal of Comparative Neurology</i>	18	0.76	0	3.83
<i>Leukemia & Lymphoma</i>	18	0.76	100	1.56
<i>Transfusion</i>	18	0.76	- 50	3.28
<i>Brain Research</i>	15	0.63	100	2.34
<i>Histology and Histopathology</i>	15	0.63	100	2.18
<i>Revista De Neurologia</i>	15	0.63	0	0.53
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	13	0.55	200	9.64
<i>Cytotherapy</i>	12	0.51	300	2.16
<i>Journal of Biological Chemistry</i>	12	0.51	100	5.81
<i>Stem Cells</i>	12	0.51	200	7.92
<i>Transplantation Proceedings</i>	12	0.51	- 33	0.96
<i>Annals of Hematology</i>	11	0.46	- 67	2.25
<i>Brain Research Bulletin</i>	11	0.46	- 100	1.68
<i>European Journal of Haematology</i>	11	0.46	200	1.86
<i>International Journal of Developmental Biology</i>	11	0.46	300	3.58
<i>Physiologia Plantarum</i>	11	0.46	- 33	2.17
<i>Revista Espanola De Cardiologia</i>	11	0.46	200	2.18

Table VI.
Most productive journals
(more than ten
documents)

A look at the evolution of these categories with regard to total productions reveals the most productive ones to be responsible for the overall trend in production (see Figure 1). Neurosciences constitutes a very productive category throughout the period of study; yet from 2004 onward, its output rises even more dramatically. Similarly, Cell Biology, Biochemistry and Molecular Biology, Biotechnology and Applied Microbiology and Immunology undergo sharp rises. In this comparative framework, certain parallels can be drawn with Hematology, Oncology and Biophysics (see Figure 7). Thus, in Figure 8 we remove these three categories to emphasise the weight and scientific influence of the rest.

Only the most productive ones are represented (over 50 documents).

Up to now, the most solid trend derived from our findings is that most output comes from Cataluña and Madrid, and is fundamentally produced in the health sector. More specifically, the outstanding centers would be: Hospital Sta. Cruz and S. Pablo de Barcelona, Hospital Clínico de Barcelona and the Hospital Clínico de Salamanca; and

AP
64,6

576

Subject categories	No of documents	Per cent documents	Growth rate
Hematology	997	42.09	209
Oncology	570	24.06	161
Biophysics	351	14.82	120
Neurosciences	215	9.08	367
Cell biology	181	7.64	1,175
Biochemistry and molecular biology	142	5.99	617
Immunology	117	4.94	2,600
Biotechnology and applied microbiology	99	4.18	2,300
Medicine, research and experimental	91	3.84	950
Plant sciences	77	3.25	40
Transplantation	68	2.87	
Medicine, general and internal	62	2.62	-50
Genetics and heredity	60	2.53	
Clinical neurology	59	2.49	
Developmental biology	55	2.32	333
Pathology	52	2.20	433
Pharmacology and pharmacy	49	2.07	200
Cardiac and cardiovascular systems	48	2.03	
Surgery	39	1.65	
Microbiology	37	1.56	100
Infectious diseases	33	1.39	
Endocrinology and metabolism	30	1.27	
Anatomy and morphology	23	0.97	400
Multidisciplinary sciences	23	0.97	500
Biochemical research methods	22	0.93	
Dermatology	21	0.89	200
Radiology, nuclear medicine and medical imaging	21	0.89	100
Zoology	21	0.89	0
Biology	18	0.76	300
Peripheral vascular disease	18	0.76	

Table VII.
Most productive subject categories

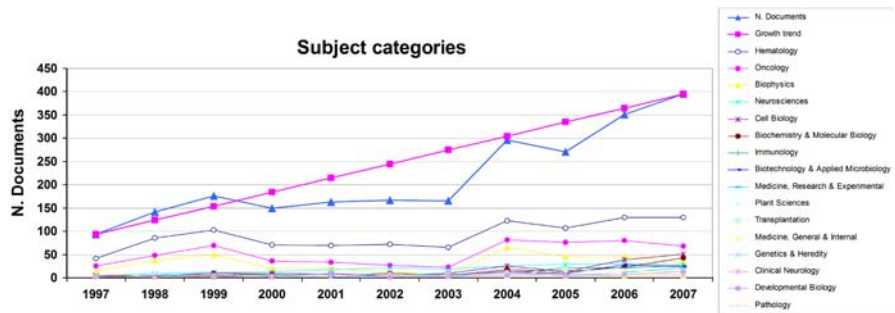


Figure 7.
Evolution of the most frequent categories

that the most productive authors belong to these centres, namely J. Sierra, and R. Martino, and J.F. San Miguel. We also corroborated that the most productive journals correspond precisely with the categories of greatest output. All this leads us to designate one single research area in Spain as the one mainly related with stem cell

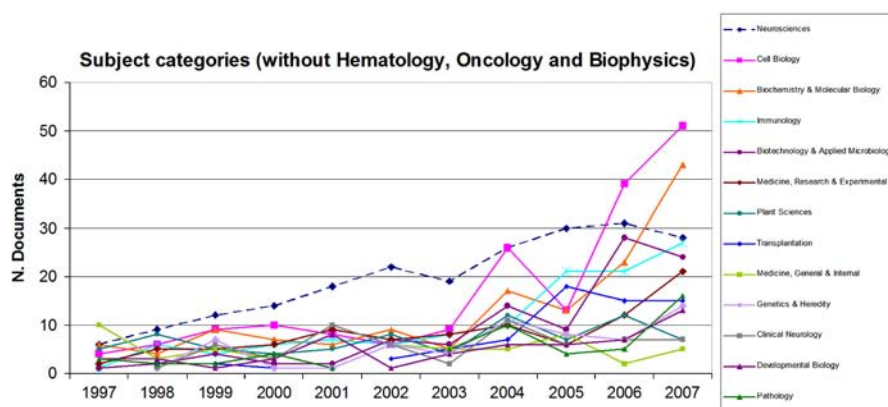


Figure 8. Evolution of the most frequent categories without Hematology, Oncology and Biophysics

research – Hematology and Oncology. That is, studies involving clinical research are foremost in the field and in the literature. Its weight is so substantial that, when we carried out the bibliometric study using indicators, it was not possible to detect the implications of other less productive disciplines or research areas. We shall proceed to see if, through the representation of the relations established among the most productive categories, we might detect disciplines other than Oncology and Hematology involved in the bulk of stem cell research in Spain.

Visual subject category analysis. Figure 9 displays the network of the 29 most frequent categories in the journals publishing the 2,369 documents retrieved with reference to stem cell research.

The map of categories clearly shows a single main nucleus comprising Hematology, Oncology and Biophysics, where the most solid (red) relations are seen. Indeed, the

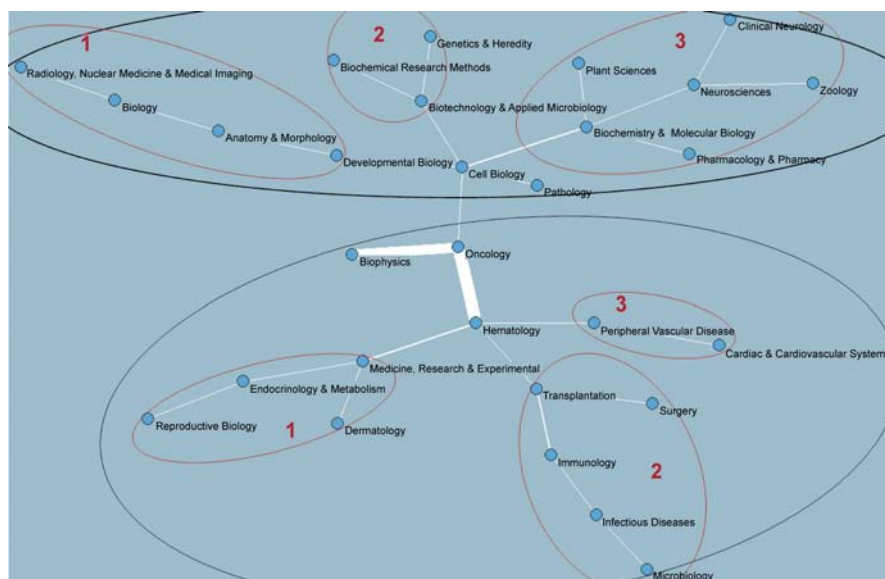


Figure 9. PFNET network of the most frequent categories related with stem cells

most productive journal, Bone Marrow Transplantation (328 documents), is classified in these three categories.

Aside from this main nucleus, we see that Oncology gives rise to two rather weak groupings. The one toward the top of the map contains Cell Biology, which in turn embraces four subgroups, namely: Developmental Biology, Pathology, Biochemistry and Molecular Biology and Biotechnology and Applied Microbiology. In the lower part of the map, Hematology is a category that shows three other subgroupings, in this case: Medicine, Research and Experimentation, Peripheral Vascular Disease and Transplant.

Unlike what we have seen so far, this map allows us to distinguish two types of research, clinical and basic. The categories related with the former are situated in the central and lower part of the map and they correspond with the thematic groupings detected on the one hand, such as Hematology, Oncology and Biophysics, where the most solid (red) relations are seen and on the other, Medicine, Research and Experimentation, or Peripheral Vascular Disease and Transplant. The categories related with basic research, in contrast, are found in the upper area of the map, grouped under Cell Biology, which in turn harbors a number of subgroups.

However, this category analysis provides information more closely related with the subject classification of the journals where the documents are published (Moya-Anegón *et al.*, 2004) than with the information contained in the documents *per se*. We must bear in mind, then, that the data and conclusions to be deduced from this type of analysis serve only for journal characterization, not for any thematic breakdown. The type of information it offers is of a general character, and does not allow us to descend to a lower level of aggregation that might enable us to detect groupings with sufficient identity to reflect more specific areas or lines of research. Therefore, we shall use the KW + as minor units of analysis that try to capture the intellectual content of documents so that one may visualize the subject relations existing among them.

5.2.2 *KW + analysis*. Of the 2,369 documents obtained, 2,077 were indexed in the database using some sort of descriptor (1,846 documents had KW + and/or AKW, and 231 had only the AKW). This means that 78 per cent of the publications are indexed under KW + descriptors – a remarkably high documental presence. The total number of KW + was 16,024, of which 6,199 KW + were unique. The mean number of the descriptors per document is 8.68 when considered in terms of the documents with some type of KeyWords (1,846) (see Tables VIII and IX).

Table VIII.
Documents indexed

	<i>n</i>
Total	2,369
Total with descriptors	2,077
With KW +	1,846
Only with AKW	231

Table IX.
Number of KW +

	No. of KW +
Total	16,024
Unique	6,199
No. of KW+ per document (<i>n</i> = 1,846)	8.68

Regarding the prevalence of appearance of the KW + , the most frequent descriptor is Bone-Marrow-Transplantation with 390 hits, nearly twice the yield of stem cell, with 207. Although the term used in our search strategy was “stem cell” in the field “topic”, it retrieved documents that included this term in the KW Author, Title, Abstract, and of course, KW + as well. For this reason, the term appearing the most times in the KW + is not necessarily the search term. Expression, Stem-Cell Transplantation, Therapy and Progenitor Cells are found over 120 times, while Central Nervous-System, Colony-Stimulating Factor, Bone-Marrow and Differentiation close the group of descriptors with frequencies of appearance over 50 (see Table X).

Co-word network. Figure 10 offers a heat map where the main lines of research stand out. This visualization was obtained using VOSviewer, taking as the matrix of co-occurrence of the 6,199 unique KW + s, standardized by the measure of strength of association (Van Eck and Waltman, 2009).

Although the KW + , which represent the domain of stem cells, show a highly interrelated distribution, the structure can be clearly seen. To the right of the map the

Descriptor	Frequency	Per cent documents with KW + *
Bone-marrow-transplantation	390	21.13
Stem-cells	207	11.21
Expression	160	8.67
Stem-cell transplantation	142	7.69
Therapy	124	6.72
Progenitor cells	123	6.66
Central-nervous-system	113	6.12
Colony-stimulating factor	106	5.74
Bone-marrow	103	5.58
Differentiation	103	5.58
Transplantation	99	5.36
In-vitro	97	5.25
Graft-versus-host-disease	94	5.09
Chemotherapy	80	4.33
Gene-expression	78	4.23
Hematopoietic stem-cells	73	3.95
Leukemia	69	3.74
High-dose chemotherapy	66	3.58
Engraftment	65	3.52
Randomized-trial	65	3.52
Peripheral-blood	65	3.52
Survival	64	3.47
In-vivo	61	3.30
Blood	60	3.25
Brain-stem	59	3.20
Disease	58	3.14
Embryonic stem-cells	55	2.98
Mice	54	2.93
Recipients	53	2.87
Acute Myeloid-Leukemia	52	2.82
Non-Hodgkins-Lymphoma	51	2.76

Note: *In view of docs that have KW+ (1.846)

Table X.
Most frequent descriptors

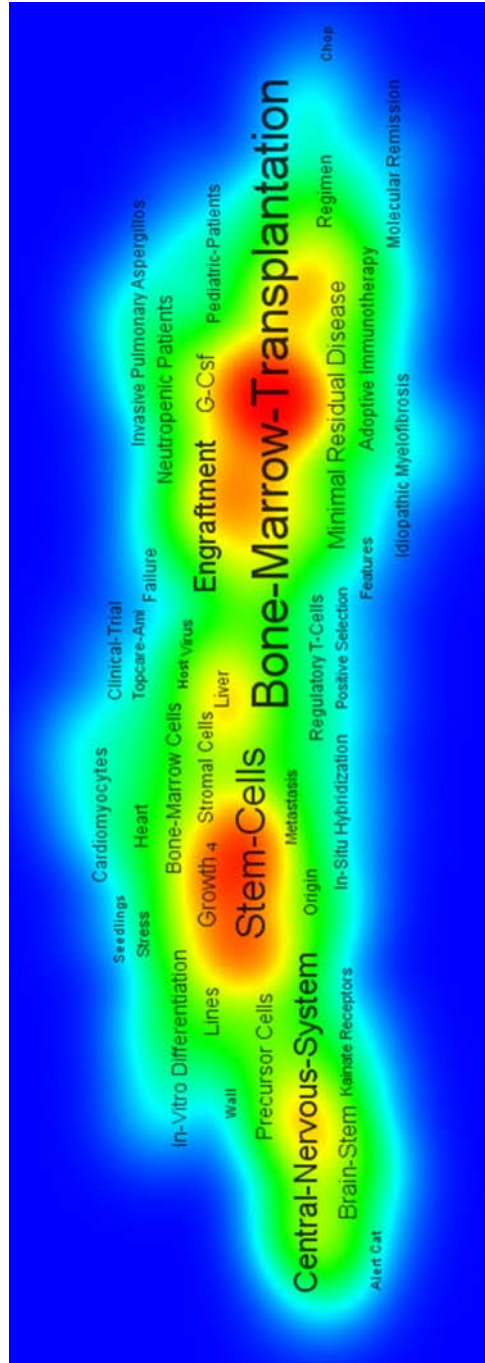


Figure 10.
Visualization of the main
lines of research involving
stem cells (VOSviewer 1.3)

documents related with the therapeutic use of hematopoietic progenitor stem cells are grouped, as indicated by the weight of Therapy in this KW + set. As expected, this core appears linked to terms related to different clinical aspects of hematopoietic stem cell transplantation such as diseases (i.e. leukemia, lymphoma, breast cancer [...]), technical aspects (i.e. allogeneic, leukapheresis, immunosuppression [...]), outcome and complications. In the central position we see a concentration of documents referring to more basic research, reflected by the weight of the term *in-vitro* and animal experimental research (i.e. mice). Peripheral to this core we find terms related to early applications of stem cell research for regenerative medicine (i.e. cardiomyocytes and acute myocardial infarction) as well as the term cancer, presumably reflecting studies on cancer stem cell biology. Finally, to the left works involving stem cells in the context of neurosciences appear clustered and well defined.

As we stated earlier, the visualization obtained with VOSviewer permits quick and easy identification of the main areas of research in this domain. However, if we also wish to view the most relevant descriptions within each line of research, we will need to resort to other types of visualization, based on alternative techniques and methods.

Figure 11 shows the PFNET network of the 102 KW + that appear most frequently in the 2,369 documents retrieved for this study. The representativeness of these descriptors is quite high, as they are present in 68.80 per cent of the documents. This finding confirms the validity and adequacy of the method in locating and representing major areas of research, despite the fact that the degree of dispersion of the KW + can be greater than that of other indexing systems.

Analysis of the relations among the most frequent descriptors reveals four well defined groups, but with a variable degree of dispersion. In the upper left part, around the node Bone Marrow Transplantation, we find the most compact group of the network. It includes descriptors that represent documents in the areas of Hematology and Oncology, related fundamentally with the application and clinical research into the transplant of hematopoietic progenitor cells begun in the 1980s. From the descriptor Colony-Stimulating Factor, three other groupings appear. To the left, the KW + stem cells connects with a well defined cluster of descriptors related with hematopoietic stem cell research (Bone Marrow, Placental Blood, Hematopoietic stem cells, Transplantation, Cancer [...]) Meanwhile, the term Progenitor Cells branches out into two other lines. To the right, we have the KW + associated with basic research related with the processes of generation, proliferation and the differentiation of lines of stem cells (Embryonic stem-cells, Mesenchymal stem-cells, Gene-Expression, Cell, Protein, Growth, Culture, Lines, Self-Renewal, Proliferation, Activation, Apoptosis) Finally, in the lower zone and around the descriptor Central-Nervous-System we can clearly discern a line grouping descriptors specifically related with neural progenitor cells (Neural Stem-Cells, Neurogenesis, Spinal-Cord, Brain-Stem, Neurons [...]), which reveals the special attention that they are given from the realm of the neurosciences.

Even though the depictions of Figures 10 and 11 were made using different programs, data, techniques, methods and layout algorithms, the two coincide in identifying the number and name of the main areas of research surrounding stem cells in Spain.

In addition, the results obtained from category analysis in the previous section coincide with those derived from the KW + , indicating that the most weight in stem cell research in Spain resides in hematological and oncological studies. This is an

interesting finding because, despite the theoretically highly interdisciplinary nature of the field, this area is reportedly dominated by a few central medical research areas (Zhao and Strotmann, 2011). Category analysis does not allow us to define research lines beyond the set of disciplines directly involved in the domain of study. For this reason, the information obtained through KW + analysis is much more comprehensive and precise. Even though the two largely coincide or overlap, that of subject categories (see Figure 9) is more general or superficial than the information based on KW + (Figures 10 and 11). The latter show, with a greater level of desegregation, the distribution of the descriptors that specifically configure each thematic profile, plus the less productive or incipient research that would otherwise remain hidden.

6. Conclusions and further research

The results we harvested and present here give substantial information about the state of research involving stem cells in Spain to date, and allow us to draw profiles from different viewpoints.

First, bibliometric analysis, which relies on a general search strategy, reflects a series of data that serve to characterize the scientific domain and lend it an identity of its own to better contextualize the study. This provides the reader or the decision-maker with a point of departure for grasping the domain. This study has allowed us to depict a general panorama of Spain's research efforts surrounding stem cells by means of a bibliometric analysis, and evidences the possible influence of certain patterns of publication and collaboration. It also makes it possible to detect and represent the research trends and areas that characterize this particular scientific domain.

As could be expected, the thematic analysis based on the study of categories offers much more general, superficial and imprecise information than the KW + . The choice of categories as the units of measure for representing thematic relations among documents is very promising indeed, as it can have a dual function. It can be used to locate the main lines of research under a given scientific domain, and can moreover serve as a means to locate the documents pertaining to specific research lines, allowing for their direct application to carry out bibliometric studies at the micro level.

We are, of course, aware of the fact that this methodology has its limitations, which should be brought to light here for the benefit of future work. Granted, using one or another data source may condition the results of the study. Database coverage is determinant for obtaining more or less comprehensive results. Future studies may therefore lead us to apply this same methodology to alternative databases, so as to arrive at comparative results.

The use of different information sources would furthermore serve to analyze the indexing systems from diverse perspectives. First, using the same systems and the same documental or disciplinary set (for instance, AKW) could lead us to uncover differences in data representation. Second, using different indexing systems might help us perceive the different foci of interest and the emergence or consolidation of research lines, depending on the use of controlled or non-controlled vocabularies.

We should moreover mention that the search strategy applied here was not chosen for the purpose of retrieving all the Spanish research papers on stem cells publisher, but rather to retrieve a volume of documents that would be heterogeneous enough to

serve as a pool of studies from which to discern research lines through thematic analysis. In the near future, it would indeed prove interesting to take the most frequent KW + from each research line detected, and then relaunch them onto the database. In this way, the retrieval of data on this specific subject would be more exhaustive, and the results could be compared with those obtained using other search (categories, keywords, topics [...]). Similarly, we could focus on the study of one or more lines detected, in order to develop bibliometric studies at the micro level.

Finally, we should acknowledge that the methodology applied and described here may entail limitations that could condition the results expounded. There are other analytical models (aside from co-words) that could reflect the research lines detected and offer additional, complementary information. Alternative or future techniques for visualization could likewise enhance the display of results.

In short, the results put forth here can be viewed as a composite still-life photograph of the period of study within the diffuse intellectual framework of Spain. It would prove very interesting to apply this methodology over a series of time periods to view the evolution of specific specialized areas and the degree of participation of the institutions, authors and geographic regions involved in biomedical research. The very nature of scientific collaboration makes it desirable to further explore these networks of co-authorship and author citation. We therefore stress that the notion of chronological studies incorporating new analytical and display techniques are a source of motivation for our research group.

For the time being, we conclude that the proposed methodology may prove very helpful for the thematic delimitation of stem cell research or any other scientific domain, and it can also be used as a tentative approach for the conceptual dissection of a domain, making manifest the main foci of interest.

Note

1. Resolución de 28 de agosto de 1989, modificada y completada por el Real Decreto 1325/2002.

References

- Aharony, N. (2010), "Information literacy in the professional literature: an exploratory analysis", *Aslib Proceedings*, Vol. 62 No. 3, pp. 261-82.
- Batagelj, V. and Mrvar, A. (2010), "Pajek 2.0: package for large network analysis", available: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> (accessed 8 March 2010).
- Bessalar, P. and Heimeriks, G. (2006), "Mapping research topics using word-reference co-occurrences: a method and an exploratory case study", *Scientometrics*, Vol. 68 No. 3, pp. 377-93.
- Bishop, A.E., Buttery, L.D.K. and Polak, J.M. (2002), "Embryonic stem cells", *Journal of Pathology*, Vol. 197 No. 4, pp. 424-9.
- Bordons, M. and Zulueta, M.A. (1999), "Evaluation of the scientific activity through bibliometric indices", *Revista Española Cardiología*, Vol. 52 No. 10, pp. 780-800.
- Bordons, M., Gómez, I., Fernández, M.T., Zulueta, M.A. and Mendez, A. (1996), "Local, domestic and international scientific collaboration in biomedical research", *Scientometrics*, Vol. 37 No. 2, pp. 279-95.
- Börner, K., Chen, C. and Boyack, K.W. (2003), "Visualizing knowledge domains", *Annual Review of Information Science & Technology*, Vol. 37, pp. 179-255.

-
- Boyack, K.W. and Klavans, R. (2010), "Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 12, pp. 2389-404.
- Boyack, K.W., Börner, K. and Klavans, R. (2009), "Mapping the structure and evolution of chemistry research", *Scientometrics*, Vol. 79 No. 1, pp. 45-60.
- Braun, T., Glänzel, W. and Schubert, A. (2000), "How balanced is the Science Citation Index's journal coverage? – A preliminary overview of macrolevel statistical data", in Cronin, B. and Atkins, A.B. (Eds), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Asist Monograph Series, pp. 251-77.
- Cambrosio, A., Limoges, C., Courtial, J.P. and Laville, F. (1993), "Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis", *Scientometrics*, Vol. 27 No. 2, pp. 119-43.
- Camí, J., Zulueta, M.A., Fernández, M.I., Bordons, M. and Gómez, I. (1997), "Spanish scientific production in biomedicine and health sciences during the period 1990-1993 (Science Citation Index and Social Science Citation Index) and comparison to period 1986-1989", *Medicina Clínica*, Vol. 109 No. 13, pp. 481-96.
- Cao, Q.L., Benton, R.L. and Whitemore, S.R. (2002), "Stem cell repair of central nervous system injury", *Journal of Neuroscience Research*, Vol. 68 No. 5, pp. 501-10.
- Chalik, T. and Jirina, M. (2006), "Law of cumulative advantages in the evolution of scientific fields", *Scientometrics*, Vol. 66 No. 3, pp. 441-9.
- Chen, C., Ibekwe-SanJuan, F. and Hou, J. (2010), "The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 7, pp. 1386-409.
- Chen, D.Z., Huang, M.H., Hsieh, H.C. and Lin, C.P. (2011), "Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology", *Journal of Informetrics*, Vol. 5 No. 3, pp. 400-12.
- Chiau, W.T. and Ho, Y.S. (2007), "Bibliometric analysis of tsunami research", *Scientometrics*, Vol. 73 No. 1, pp. 3-17.
- Chinchilla-Rodríguez, Z. (2005), "Análisis del dominio científico español: 1995-2002 (ISI, Web of Science)", PhD dissertation, Universidad de Granada, Granada.
- Chinchilla-Rodríguez, Z. and Moya-Anegón, F. (2007), "La investigación científica española (1995-2002): una aproximación métrica", Universidad de Granada, Granada.
- Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Hassan-Montero, Y., González-Molina, A. and Moya-Anegón, F. (2010), "New approach to the visualization of international scientific collaboration", *Information Visualization*, Vol. 9 No. 4, pp. 277-87.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2011), "An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field", *Journal of Informetrics*, Vol. 5 No. 1, pp. 146-66.
- Courtial, P., Cahlik, T. and Callon, M. (1994), "A model for social interaction between cognition and action through a key-word simulation of knowledge growth", *Scientometrics*, Vol. 3 No. 2, pp. 173-92.
- Cronin, B. (2001), "Hyperauthorship: a postmodern perversion or evidence of a structural shift in scholarly communication practices", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 7, pp. 558-69.

- Di Giorgio, F.P., Carrasco, M.A., Siao, M.C., Maniatis, T. and Eggan, K. (2007), "Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model", *Nature Neuroscience*, Vol. 10 No. 5, pp. 608-14.
- Ding, Y., Chowdhury, G. and Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis", *Information Processing & Management*, Vol. 37 No. 6, pp. 67-78.
- Fernández, M.T., Cabrero, A., Zulueta, M.A. and Gómez, I. (1993), "Constructing a relational database for bibliometric analysis", *Research Evaluation*, Vol. 3 No. 1, pp. 55-62.
- García-Guinea, J. and Ruis, J.D. (1998), "The consequences of publishing in journals written in Spanish in Spain", *Interciencia*, Vol. 23 No. 3, pp. 185-7.
- Garfield, E. (1990), "Keywords plus: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on *Current Contents on Diskette*", *Current Contents*, Vol. 1 No. 32, pp. 5-9.
- Garfield, E. and Sher, I.H. (1993), "Keywords PlusM – algorithmic derivative indexing", *Journal of the American Society for Information Science*, Vol. 44 No. 5, pp. 298-9.
- Glänzel, W. and Thijs, B. (2011), "Using 'core documents' for the representation of clusters and topics", *Scientometrics*, Vol. 88 No. 1, pp. 297-309.
- Hammond, J. (1949), "Recovery and culture of tubal mouse ova", *Nature*, Vol. 163 No. 4131, pp. 28-9.
- Hjørland, B. and Albrechtsen, H. (1995), "Toward a new horizon in information science: domain analysis", *Journal of the American Society Information Science*, Vol. 46, pp. 400-25.
- Ho, Y.S. (2007), "Bibliometric analysis of adsorption technology in environmental science", *Journal of Environmental Protection Science*, Vol. 1 No. 1, pp. 1-11.
- Ho, Y.S., Chiu, C.H., Tseng, T.M. and Chiu, W.T. (2003), "Assessing stem cell research productivity", *Scientometrics*, Vol. 57 No. 3, pp. 369-76.
- Jiménez-Contreras, E., Faba-Pérez, C. and Moya-Anegón, F. (2001), "El destino de la revistas científicas nacionales. El caso español a través de una muestra (1950-90)", *Revista Española de Documentación Científica*, Vol. 24 No. 2, pp. 147-61.
- Jiménez Contreras, E., Moya Anegón, F. and Delgado López-Cózar, E. (2003), "The evolution of research activity in Spain. The impact of the National Commission for the Evaluation of Research Activity (CNEAD)", *Research Policy*, Vol. 32 No. 1, pp. 123-42.
- Kamada, T. and Kawai, S. (1989), "An algorithm for drawing general undirected graphs", *Information Processing Letters*, Vol. 31 No. 1, pp. 7-15.
- Kessler, M.M. (1963), "Bibliographic coupling between scientific papers", *American Documentation*, Vol. 14 No. 1, pp. 10-25.
- Krauskopf, D. (2000), "La construcción social de la Ciencia y la Tecnología", *Boletín SEBBM*, Vol. 130, pp. 12-16.
- Kyvik, S. (2003), "Changing trends in publishing behaviour among university faculty", *1980-2000*, *Scientometrics*, Vol. 58 No. 1, pp. 35-48.
- Lee, W.H. (2008), "How to identify emerging research fields using scientometrics: an example in the field of information security", *Scientometrics*, Vol. 76 No. 3, pp. 503-25.
- Leydesdorff, L. and Hellsten, I. (2005), "Metaphors and diaphors in science communication", *Science Communication*, Vol. 27 No. 1, pp. 64-99.
- Leydesdorff, L. and Hellsten, I. (2006), "Measuring the meaning of words in contexts: an automated analysis of controversies about 'Monarch butterflies', 'Frankenfoods' and 'stem cells'", *Scientometrics*, Vol. 67 No. 2, pp. 231-58.

-
- Leydesdorff, L. and Rafols, I. (2009), "A global map of science based on the ISI subject categories", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 2, pp. 348-62.
- Leydesdorff, L. and Welbers, K. (2011), "The semantic mapping of words and co-words in contexts", *Journal of Informetrics*, Vol. 5 No. 3, pp. 469-75.
- Li, L.L., Ding, G., Feng, N., Wang, M.H. and Ho, Y.S. (2009), "Global stem cell research trend: bibliometric analysis as a tool for mapping of trends from 1991 to 2006", *Scientometrics*, Vol. 80 No. 1, pp. 39-58.
- Martínez Serrano, A. and Bjorklund, A. (1996), "Protection of the neostriatum against excitotoxic damage by neurotrophin-producing, genetically modified neural stem cells", *Journal of Neuroscience*, Vol. 16 No. 15, pp. 4604-16.
- Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solan, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E. and Muñoz-Fernández, F.J. (2004), "A new technique for building maps of large scientific domains based on the cocitation of classes and categories", *Scientometrics*, Vol. 61 No. 1, pp. 29-145.
- Moya-Anegón, F., Chinchilla Rodríguez, Z., Corera-Álvarez, E., Gómez-Crisóstomo, R., González-Molina, A., Muñoz-Fernández, F. and Vargas-Quesada, B. (2007a), "Indicadores bibliométricos de la actividad científica española: 1990-2004", Fundación Española para la Ciencia y la Tecnología, Madrid.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F.J. and Herrero-Solana, V. (2007b), "Visualizing the marrow of science", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 14, pp. 2167-79.
- Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Gómez-Crisóstomo, R., Hassan-Montero, Y. and Vargas-Quesada, B. (2009), "Indicadores Bibliométricos de la Actividad Científica Española: 2007", Fundación Española para la Ciencia y la Tecnología (Fecyt), Madrid.
- Neff, M.W. and Corley, E.A. (2009), "35 years and 160.000 articles: a bibliometric exploration of the evolution of ecology", *Scientometrics*, Vol. 80 No. 3, pp. 657-82.
- Narin, F., Stevens, K. and Whitlow, E.S. (1991), "Scientific cooperation in Europe and the citation of multinationally authored papers", *Scientometrics*, Vol. 21 No. 3, pp. 313-23.
- Newman, M.E.J. (2001), "The structure of scientific collaboration networks", *Proceedings of the National Academy of Science of the USA*, Vol. 98 No. 2, pp. 404-9.
- Noyons, E.C.M., Moed, H.F. and Luwel, M. (1999), "Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study", *Journal of the American Society for Information Science*, Vol. 50 No. 2, pp. 115-31.
- Onyancha, O.B. and Ocholla, D. (2005), "Informetric investigation of the relatedness of opportunistic infections to HIV/AIDS", *Information Processing & Management*, Vol. 41 No. 6, pp. 1573-88.
- Perianes-Rodríguez, A., Olmeda-Gómez, C., Ovalle-Perandones, M.A., Chinchilla-Rodríguez, Z. and Moya-Anegón, F. (2011), "R&D collaboration in 50 major Spanish companies", *Aslib Proceedings*, Vol. 63 No. 1, pp. 5-27.
- Qin, J. (2000), "Semantic similarities between a keyword database and a controlled vocabulary database: an investigation in the antibiotic resistance literature", *Journal of the American Society for Information Science*, Vol. 51 No. 2, pp. 166-80.

- Rafols, I., Porter, A. and Leydesdorff, L. (2010), "Science overlay maps: a new tool for research policy and library management", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 9, pp. 1871-87.
- Sanz-Casado, E., Aragon, I. and Mendez, A. (1995), "The function of national journals in disseminating applied science", *Journal of Information Science*, Vol. 21 No. 4, pp. 319-23.
- Small, H. and Upham, S.P. (2009), "Citation structure of an emerging research area on the verge of application", *Scientometrics*, Vol. 79 No. 2, pp. 365-75.
- Su, H.N. and Lee, P.C. (2010), "Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in technology foresight", *Scientometrics*, Vol. 85 No. 1, pp. 65-79.
- Van Eck, N.J. and Waltman, L. (2009), "How to normalize cooccurrence data? An analysis of some well-known similarity measures", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 8, pp. 1635-51.
- Van Eck, N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84 No. 2, pp. 523-38.
- Van Raan, A.F.J. and Tijssen, R.J.W. (1993), "The neural net of neural network research. An exercise in bibliometric mapping", *Scientometrics*, Vol. 26 No. 1, pp. 169-92.
- Vargas-Quesada, B., Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E. and Guerrero-Bote, V. (2008), "Development of the Spanish scientific landscape: ISI web of science 1990-2005", *Profesional de la Información*, Vol. 17 No. 1, pp. 22-37.
- Vargas-Quesada, B., Chinchilla-Rodríguez, Z., González-Molina, A. and Moya-Anegón, F. (2010), "Showing the essential science structure of a scientific domain and its evolution", *Information Visualization*, Vol. 9, pp. 288-300.
- Wasserman, S. and Faust, K. (1998), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
- White, H.D. (2003), "Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 5, pp. 423-34.
- White, H.D. and Griffith, B.C. (1981), "Author co-citation: a literature measure of intellectual structure", *Journal of the American Society for Information Science*, Vol. 32, pp. 163-71.
- White, H. and McCain, K. (1998), "Visualizing a discipline: an author co-citation analysis of information science, 1972-1995", *Journal of the American Society for Information Science*, Vol. 49 No. 4, pp. 327-75.
- Zhao, D. and Strotmann, A. (2011), "Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field", *Scientometrics*, Vol. 87 No. 1, pp. 115-31.
- Zitt, M. and Bassecouard, E. (1996), "Reassessment of co-citation methods for science indicators: effect of methods improving recall rates", *Scientometrics*, Vol. 37 No. 2, pp. 223-44.
- Zitt, M. and Bassecouard, E. (2006), "Delineating complex scientific fields by an hybrid lexical-citation method: an application to nanosciences", *Information Processing & Management*, Vol. 42 No. 6, pp. 1513-31.
- Zitt, M., Lelu, A. and Bassecouard, E. (2008), "Hybrid maps of scientific fields (terms and citations): an application to nanosciences", paper presented at the Tenth International Conference on Science & Technology Indicators: Excellence and Emergence – A New Challenge for the Combination of Quantitative and Qualitative Approaches, Vienna.
- Zitt, M., Lelu, A. and Bassecouard, E. (2011), "Hybrid citation-word representations in science mapping: portolan charts or research fields?", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 1, pp. 19-39.

Zulueta, M.A., Cantos-Mateos, G., Sánchez, C. and Vargas-Quesada, B. (2011), "Research involving women and health in the Medline database, 1965-2005, co-term analysis and visualization of main lines of research", *Scientometrics*, Vol. 88 No. 3, pp. 679-706.

Further reading

- Bellavista, J., Guardiola, E., Méndez, A. and Bordons, M. (1997), "Evaluación de la investigación", Centro de Investigaciones Sociológicas (Cuadernos Metodológicos), Madrid.
- Bordons, M. and Gómez Caridad, I. (1997), "La actividad científica española a través de indicadores bibliométricos en el período 1990-93", *Revista General de Información y Documentación*, Vol. 7 No. 2, pp. 69-86.
- España (2003), "Ley 45/2003, de 21 de noviembre, por la que se modifica la Ley 35/1988", de 22 de noviembre, sobre Técnicas de Reproducción Asistida. Boletín Oficial del Estado, 22 noviembre de 2003, núm. 280, 41.458-41.463 [Law on Techniques of Assisted Reproduction].
- España (2006), "Ley 14/2006, de 26 de mayo, sobre técnicas de reproducción humana asistida. Boletín Oficial del Estado", 27 de mayo de 2006, núm. 126, 19947-19956 [Law on Techniques of Assisted Reproduction].
- España (2007), "Ley 14/2007, de 3 de julio, de Investigación biomédica. Boletín Oficial del Estado", 4 de julio de 2007, núm. 159, 28.826-28.848 [Law on Biomedical Research].
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., González-Molina, A., Muñoz-Fernández, F.J. and Herrero-Solana, V. (2006), "Visualizing and analyzing the Spanish science structure: ISI Web of Science 1990-2005", *Profesional de la Información*, Vol. 15 No. 4, pp. 258-69.
- Quirin, A., Cordon, O., Guerrero-Bote, V.P., Vargas-Quesada, B. and Moya-Anegón, F. (2008), "A quick MST-based algorithm to obtain pathfinder networks ($\infty, n - 1$)", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 12, pp. 1912-24.
- Resolución de 28 de agosto de 1989 (1989), Resolución de 28 de agosto de 1989, modificada y completada por el Real Decreto 1325/2002 [Law on Performance Research].
- Thomson Reuters ISI Web of Science (2010), *Thomson Reuters ISI Web of Science*, available at: <http://science.thomsonreuters.com/es/productos/wos> (accessed 11 April 2008).

Author affiliations

G. Cantos-Mateos is based at the Information Science Faculty, University of Alcalá, Madrid, Spain and Consejo Superior de Investigaciones Científicas, Centro de Ciencias Humanas y Sociales, Instituto de Políticas y Bienes Públicos (CSIC-IPP), SCImago Research Group Madrid, Madrid, Spain.

B. Vargas-Quesada is based at the Communication and Information Science Faculty, University of Granada, Granada, Spain and Consejo Superior de Investigaciones Científicas, Centro de Ciencias Humanas y Sociales, Instituto de Políticas y Bienes Públicos (CSIC-IPP), SCImago Research Group Madrid, Madrid, Spain.

Z. Chinchilla-Rodríguez is based at Consejo Superior de Investigaciones Científicas, Centro de Ciencias Humanas y Sociales, Instituto de Políticas y Bienes Públicos (CSIC-IPP), SCImago Research Group Madrid, Madrid, Spain

M.A. Zulueta is based at Information Science Faculty, University of Alcalá, Madrid, Spain and Consejo Superior de Investigaciones Científicas, Centro de Ciencias Humanas y Sociales, Instituto de Políticas y Bienes Públicos (CSIC-IPP), SCImago Research Group Madrid, Madrid, Spain.

About the authors

G. Cantos-Mateos obtained a pre-doctoral grant (JAE-Predoc) from the Consejo Superior de Investigaciones Científicas (CSIC), under Spain's Ministerio de Ciencia e Innovación, in 2010. Previous research grants include Introduction to Research under a project of the Community of Madrid and the University of Alcalá, and Initiation to Research from the University of Alcalá. At present, she is completing a Master's in Documentation (Library and Information Science) while collaborating with the SCImago Research Group. Her future objective is a doctoral thesis related to scientometric domain analysis and the representation and visualization of scientific information.

B. Vargas-Quesada obtained his PhD in Library and Information Science in 2005 at the University of Granada (Spain). He is currently an Associate Professor at that university and a member of the SCImago Research Group since its inception. His areas of expertise are information visualization and scientometrics, with special emphasis on domain analysis of scientific information.

Z. Chinchilla-Rodríguez has a tenure position as Researcher at the Institute of Public Goods and Policies (IPP) of the Consejo Superior de Investigaciones Científicas (CSIC) in Madrid. She obtained her PhD in Library and Information Science at the University of Granada in 2005. A member of the SCImago Research Group since its inception, her areas of expertise are scientometric domain analysis and the representation and visualization of scientific information, with special emphasis on scientific collaboration networks, as well as the methodological design of scientific information systems.

M.A. Zulueta is an MD graduate of the University of Valencia, and obtained her PhD at the Autonomous University of Madrid. Having worked at the Spanish National Research Council in Madrid (CSIC), she is currently an Associate Professor at the University of Alcalá de Henares. Her research areas are framed in the field of bibliometrics. She has directed and collaborated on research projects funded with public funds in this area, authoring a number of publications in the field. M.A. Zulueta is the corresponding author and can be contacted at: ma.zulueta@uah.es

PUBLICACIÓN Nº 4

Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2014). Localización y evolución de las principales líneas de investigación a través de los KeyWords Plus. *VI Encontro Ibérico EDICIC 2013 - Globalização, Ciência, Informação*, Oporto (Portugal), 4-6 noviembre 2013, 608-622.

Investigación española en células madre (1997-2010).

Localización y evolución de las principales líneas de investigación a través de los KeyWords Plus

Tema 2

Gisela Cantos-Mateos

Consejo Superior de Investigaciones Científicas / Instituto de Políticas y Bienes Públicos (CSIC-IPP) / SClmago Research Group (Espanha)
gisela.cantos@csic.es

Maria Angeles Zulueta

Universidad de Alcalá, Facultad de Documentación / SClmago Research Group (Espanha)
ma.zulueta@uah.es

Benjamin Vargas-Quesada

Universidad de Granada, Facultad de Comunicación y Documentación / SClmago Research Group (Espanha)
benjamin@ugr.es

Zaida Chinchilla-Rodríguez

Consejo Superior de Investigaciones Científicas / Instituto de Políticas y Bienes Públicos (CSIC-IPP) / SClmago Research Group (Espanha)
zaida.chinchilla@csic.es

RESUMEN

- **Objetivos:** analizar y representar la evolución y la dinámica de las principales líneas de investigación en la producción científica española sobre células madre entre los años 1997 y 2010.
- **Material y Métodos:** los documentos recuperados proceden de la base de datos *Science Citation Index (SCI)*. Las unidades de análisis han sido los descriptores *KeyWords Plus (KW+)*. Las herramientas utilizadas para la visualización han sido el software *Pajek*, en combinación con el algoritmo *PathfinderNetwork (PfNET)* y el software *VOSviewer*. Metodológicamente, el estudio analiza el periodo completo (1997-2007) y la evolución de las líneas de investigación en tres subperiodos: 1997-2001, 2002-2006, 2007-2010.
- **Resultados y Discusión:** Los resultados del periodo completo permiten localizar hasta seis líneas de investigación. En el primer periodo, se observa una fuerte presencia de los descriptores que representan documentos relacionados con la hematología y oncología, en cambio, el resto de líneas no comienzan a detectarse de manera más clara hasta el segundo periodo. Aquí, es posible localizar agrupaciones de descriptores relacionados con la investigación sobre células madre hematopoyéticas, otros referidos a los estudios sobre la generación, proliferación y diferenciación de células madre y otra agrupación emergente relacionada con las células progenitoras neurales. Finalmente, los resultados muestran que el peso de estas últimas agrupaciones tienen una presencia mucho más evidente en el último periodo.
- **Conclusión:** La visualización de las relaciones que se establecen entre los KW+ ha permitido obtener dos imágenes complementarias de la situación y evolución de la investigación española sobre células madre. La metodología ha permitido identificar tanto líneas de investigación claramente consolidadas como las emergentes y se intuye la capacidad de predicción del desarrollo de un dominio temático a lo largo del tiempo. Los programas de visualización se complementan muy bien y coinciden tanto en las líneas de investigación como en la localización de los descriptores más influyentes de la red.

Palabras clave: Células Madre, Análisis de Co-Words, Visualización de la Información, Análisis de Dominios Científicos

ABSTRACT

- **Objective:** To render and analyze the evolution and dynamics of the main research lines in the Spanish scientific output on stem cells between 1997 and 2010.
- **Material and Methods:** We retrieved the documents from the Science Citation Index (SCI). The units of analysis were the descriptors KeyWords Plus (KW +). The tools used for visualization have been software Pajek combined with Pathfinder Network (PFnet), and VOSviewer software. This study analyzes the period includes between 1997-2007, and the evolution of the research into three sub-periods: 1997-2001, 2002-2006 and 2007-2010.
- **Results and Discussion:** The results of the full period locate up to six main research lines. In the first period, there is a strong presence of the descriptors that represent documents related to hematology and oncology. However, the other research lines do not begin to be clearly detected until the second period. Here, we can locate clusters of descriptors related to hematopoietic stem cell research, others related with the generation, proliferation and differentiation of stem cells and finally, alternate associated to an emerging cluster of neural progenitor cells. The results show that the weight of these latter groups make them more evident in the last period.
- **Conclusion:** The visualization of the relationships between the KW + has yielded two complementary images of the situation and evolution of Spanish research on stem cells. The methodology has identified areas of research both consolidated and emerging, intuiting the development of a thematic domain over time. The visualization software complement each other quite well, matching in the identification of the main research lines and in the location of the most influential descriptors.

Keywords: Stem Cells, Co-Words Analysis, Information Visualization, Scientific Field's Analysis

1. INTRODUCCIÓN

La investigación española con células madre es un tema biomédico de gran importancia y repercusión científica por las expectativas que genera. Desde que en 1949, el científico J. Hammond (1949) descubriese el método para mantener los embriones de ratón en cultivo *in vitro*, la investigación con células madre se ha desarrollado gradualmente ofreciendo en los últimos años perspectivas muy esperanzadoras para el tratamiento de enfermedades, hasta ahora, incurables. Actualmente, la dirección de la investigación está orientada, fundamentalmente, al desarrollo de nuevas terapias para enfermedades hematológicas, cardiovasculares, neurodegenerativas, genéticas, cáncer, diabetes, etc. (Martínez Serrano y Bjorklund, 1996; Bishop et al., 2002; Cao et al., 2002; Di Giorgio et al., 2007). Por su naturaleza, la investigación con células madre trasciende inevitablemente a otros campos de carácter tan diverso como el político, ético, cultural, jurídico, etc., situándose en el escenario de la controversia social.

Por otro lado, resulta cada vez más evidente, la utilidad que tienen los estudios bibliométricos para reflejar el estado en el que se encuentra la investigación en un momento determinado, predecir tendencias y facilitar la toma de decisiones en materia de política científica. Esto impulsa el desarrollo y perfeccionamiento de nuevas técnicas y herramientas bibliométricas que tratan de mejorar la comprensión y análisis de los dominios científicos. Una muestra es la incorporación de las redes sociales (Wasserman y Faust, 1998) como técnica de análisis y visualización de dominios científicos (Boyack et al., 2009; Leydesdorff y Rafols, 2009; Vargas-Quesada et al., 2010; Rafols et al., 2010).

Existe una gran diversidad de métodos y técnicas a la hora de delimitar y visualizar un campo científico con el fin de detectar líneas de investigación. Hay estudios que utilizan la

cocitación de autores (ACA) (White y Griffith, 1981; White y McCain, 1998; Small and Upham, 2009; Chen et al., 2010; Zhao y Strotmann, 2011), el análisis de emparejamiento bibliográfico (Kessler, 1963; Boyack y Klavans, 2010; Chen et al., 2011; Glänzel y Thijs, 2011), estudios de co-words (Callon et al., 1983; Rip y Courtial, 1984, Cambrosio et al., 1993; Van Raan y Tijssen, 1993; Ding et al., 2001; Lee, 2008; Leydesdorff y Welbers, 2011; Zulueta et al., 2011) o los estudios híbridos en los que se combinan tanto ACA como co-words (Braam et al., 1991a, b; Zitt y Bassecouard, 1996; Zitt et al., 2011). Según Zitt y colaboradores (2011), las palabras, a diferencia de las citas, parecen adaptarse mejor a la «ciencia viva» porque son universales. Para estos autores, las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes. Además, el análisis de co-words es considerado como una técnica de análisis de contenido muy eficaz en el mapeo de la fuerza de asociación entre los ítems procedentes de datos textuales (Cobo et al., 2011).

En vista de las ventajas e inconvenientes que condicionan la elección de una metodología sobre otra, para el presente estudio se ha optado por el análisis de co-words (Zitt et al., 2011), continuando así estudios precedentes sobre la investigación española sobre células madres. En un estudio anterior (Cantos-Mateos et al., 2012), se analizó la producción científica desarrollada entre 1997 y 2007 sobre esta temática. Se analizó mediante indicadores bibliométricos la producción científica relativa al dominio y se generó una representación topográfica de los descriptores *KeyWords Plus* más frecuentes. Este trabajo puede considerarse una continuidad, ya que, se detectaron las principales líneas de investigación del período comprendido entre 1997 y 2007. En este caso particular, el objetivo fundamental se centra en la localización y representación de las principales líneas de investigación que tienen lugar en las distintas fases o periodos temporales en los que se ha dividido el estudio con el fin de apreciar la evolución y la dinámica, tanto de las líneas de investigación consolidadas como de los emergentes.

2. MATERIAL Y MÉTODOS

2.1. FUENTES, UNIDADES DE ANÁLISIS Y UNIDADES DE MEDIDA

La base de datos utilizada fue Science Citation Index (SCI). Entre las razones para su elección destaca que el campo *Address* está normalizado, característica fundamental para estudios que pretenden centrarse en un dominio geográfico determinado. Además, emplea los descriptores *KeyWords Plus* (KW+) para indizar los documentos. Estos descriptores son obtenidos mediante un algoritmo de búsqueda que extrae las palabras clave de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Garfield y Sher, 1993).

La elección de los KW+ como unidades de análisis se debe a su elevada representatividad, ya que están presentes en el 78,25% de los documentos (Tabla 2) Los *Keywords Author* se descartaron por su baja representatividad, ya sólo están presentes en el 52,96% (Cantos-Mateos, et al, 2013). Además, al tratarse de términos procedentes de los títulos de las referencias, son los propios autores, los que seleccionan los términos esenciales para expresar el contenido más importante de sus trabajos de investigación. Esta característica favorece la aparición de una terminología más actualizada y específica, en comparación con otros tipos de descriptores como son los procedentes de un vocabulario controlado (Braam et al., 1991; Cantos-Mateos, et al. 2013). Estas cualidades, pueden contribuir mejor a reflejar los aspectos temáticos más dinámicos de un dominio. Entre los inconvenientes hay que destacar que presentan una amplia dispersión terminológica. La falta de normalización de los

términos y de una selección estandarizada de los mismos contribuye a que se generen muchos descriptores para expresar un mismo concepto o idea (Quin, 2000). Para paliar en cierto modo esta limitación, se ha realizado una normalización unificando criterios gramaticales (plurales y singulares) y ortográficos (guiones, mayúsculas y minúsculas).

La estrategia de búsqueda empleada para la recuperación de los documentos consistió en utilizar los términos *stem* cell** en el campo *Topic*, *Spain* en el campo *Address* y limitada al periodo temporal de 1997-2010. El criterio de selección de la ventana temporal está determinado por la disponibilidad de los *Journal Citation Report* a partir del año 1997, con el objetivo de analizar el impacto de las principales revistas del área. De cara al estudio evolutivo, además de analizar el periodo completo en su conjunto (1997-2010), éste se ha subdividido en otros tres: 1997-2001; 2002-2006 y 2007-2010. Esta división temporal estuvo motivada precisamente por el volumen documental, teniendo en cuenta el crecimiento exponencial de la producción científica (De Solla Price, 1963). Quizá, el haber considerado más subdivisiones hubiera fragmentado excesivamente la información impidiendo ver diferencias importantes entre un periodo y el siguiente. Y por el contrario, realizar menos divisiones hubiera agrupado demasiada información ofreciendo una imagen más estática y corriendo el riesgo de detectar sólo líneas muy consolidadas. El número de años por periodo es de 5, a excepción de último que contiene 4. Al tratarse de un periodo de 14 años, se ha preferido que los dos primeros tramos temporales sean más amplios ya que, el volumen de producción de los primeros años es mucho menor en comparación con los últimos.

Para la identificación y visualización de las principales líneas de investigación, se han calculado las relaciones que se producen entre los KW+ a partir de su aparición conjunta en los documentos, lo que se conoce como *Co-Words Analysis*. Dicho análisis consiste, en generar una matriz cuadrada de $N \times N$ elementos, donde N es el descriptor KW+ a representar, a partir de las veces que ocurre en los documentos. El resultado es una matriz de co-ocurrencias que refleja el número de veces que un par de descriptores aparecen conjuntamente en dos documentos.

2.2. VISUALIZACIÓN DE LA INFORMACIÓN

Para la visualización de la estructura relacional entre los KW+, se han utilizado dos software libres especializados en la generación de redes. Por un lado, Pajek (Batagelj and Mrvar, 2010) y en concreto, el algoritmo *Kamada-Kawai*, (Kamada y Kawai, 1989). Aunque Pajek permite representar grandes redes, se decidió simplificarlas con el fin de que las visualizaciones fueran lo más claras posibles. Por un lado, se decidió representar únicamente los descriptores más frecuentes. Tras varios ensayos con distintos umbrales de frecuencias, se observó que las visualizaciones más claras eran las compuestas en torno a 100 nodos. Para la reducción del espacio dimensional se aplicó el algoritmo de poda *Pathfinder* (Moya-Aneón et al., 2007a, b), basado en el principio de desigualdad del triángulo bajo los parámetros $r = \infty$ y $q = n - 1$. Éste ayuda a preservar y poner de relieve las relaciones más destacadas entre los descriptores (White, 2003; Vargas-Quesada et al., 2008).

Para la representación del periodo completo, se utilizó el software VOSviewer v.1.5.4 (Van Eck y Waltman, 2010), que se presenta como una alternativa a las técnicas tradicionales de representación multidimensional y visualización de redes. VOSviewer combina técnicas de visualización y clustering, lo que favorece el análisis. En este estudio se han utilizado los mapas etiquetados con nombres, donde cada descriptor está representado por una etiqueta y cuyo tamaño es proporcional a su peso, junto con los denominados «mapas de calor» (o mapas de densidad). Estos últimos se caracterizan porque cada nodo se representa con un color que va del rojo al azul, reflejando la densidad de las relaciones entre los

descriptores. Cuanto mayor es la densidad (co-ocurrencia entre los descriptores), más se aproximarán a la tonalidad roja y cuando hay más dispersión por la menor co-ocurrencia, se aproximan más al color azul.

3. RESULTADOS Y DISCUSIÓN

Se recuperaron un total de 4.148 documentos para el período 1997-2010. La tabla 1 muestra el volumen de producción por tramos temporales. 724 documentos en el primero, 1.725 documentos en el segundo (2002-2006) por y el tercero (2007-2010) por 2.174 documentos.

TABLA 1. Número de documentos por años.

Período	Primer Período					Segundo Período					Tercer Período			
Años	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nº docs.	94	142	176	149	163	167	165	296	271	351	395	494	608	677
Total docs.	724					1.250					2.174			

3.1. PERÍODO TEMPORAL COMPLETO (1997-2010)

De los 4.148 documentos recuperados, el 78,25% están indizados con KW+, contabilizándose un total de 28.733 descriptores totales y 9.465 únicos (Tabla 2). En la visualización que ofrece Pajek (Figura 1), se pueden apreciar hasta 6 agrupaciones que, por su significado y relación con los demás, podemos considerar que definen las principales líneas de investigación. Como primera aproximación, es posible distinguir estructuralmente dos naturalezas temáticas claramente diferenciadas. Por un lado, la relacionada con la investigación clínica representada por los descriptores agrupados bajo *Bone-Marrow-Transplantation* y por otra parte, la investigación básica representada por el resto de agrupaciones. Ambas zonas quedan conectadas por dos KW+ que actúan como puentes: *T-Cells* y *Colony-Stimulating Factor*. Esta distinción resulta aún más evidente en la Fig. 3, donde se distinguen claramente estos dos focos de la investigación.

TABLA 2. Datos relativos al Período Completo (1997-2010)

Período Completo (1997-2010)			
Documentos	Totales		4.148
	Indizados con KW+ (%)		3.246 (78,25)
KW+	Totales		28.733
	Únicos	Sin Normalización	9.833
		Con Normalización	9.465
	Por documento		8,85
	Frec. Media de aparición		3,04
Visualización	Pajek	Nº Nodos	99
		Frec \geq	32
		Parejas	1.703
	VOSviewer	Nº Nodos	668
		Fuerza	5

En particular, los descriptores que conectan con *Bone-Marrow-Transplantation* (Fig. 1 y 3) están relacionados con los estudios ligados al uso terapéutico de las células madre en enfermedades hematológicas. Concretamente, muchos de los descriptores están relacionados con el tratamiento de los procesos oncológicos como la Leucemia (*Chronic Lymphocytic-Leukemia, Acute Myeloid-Leukemia; Leukemia, etc.*) a partir de trasplantes de progenitores hematopoyéticos (*Marrow-Transplantation; Stem-Cell Transplantation*).

La investigación básica está representada por las cinco agrupaciones restantes de (Figura 1) Existen dos agrupaciones que conectan directamente con la investigación clínica: en la parte central derecha del mapa, los descriptores vinculados con *Stem Cells*, relacionados con los estudios encaminados a la investigación de células madre hematopoyéticas que son las células madre procedentes de la médula ósea. Por otro lado, con aquellos descriptores que forman parte de la línea de investigación relacionada con el estudio de las células madre embrionarias. Concretamente, con los descriptores relacionados con los procesos de expresión (*Expression*) y diferenciación celular (*Differentiation*). En realidad, esta línea se encuentra desagregada en tres agrupaciones: las dos que se acaban de describir (*Expression* y *Differentiation*) y la que se detecta en torno al descriptor *In-Vitro*. En esta última, se pueden localizar descriptores relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y sobre todo, analizar in-vitro las células madre embrionarias.

La última agrupación que se detecta y que conecta con la línea de investigación de las células madre embrionarias, está relacionada con la investigación con células madre neurales (*Central-Nervous-System*). Estos descriptores están muy relacionados con los estudios centrados en el desarrollo e identificación de las células madre neuronales que se diferencian activamente en el tejido cerebral.

FIG. 1 Período Completo (1997-2010). Representación en Pajek de los 99 KW+ más frecuentes. (Frecuencia ≥ 32 ; coocurrencias = 1.703)

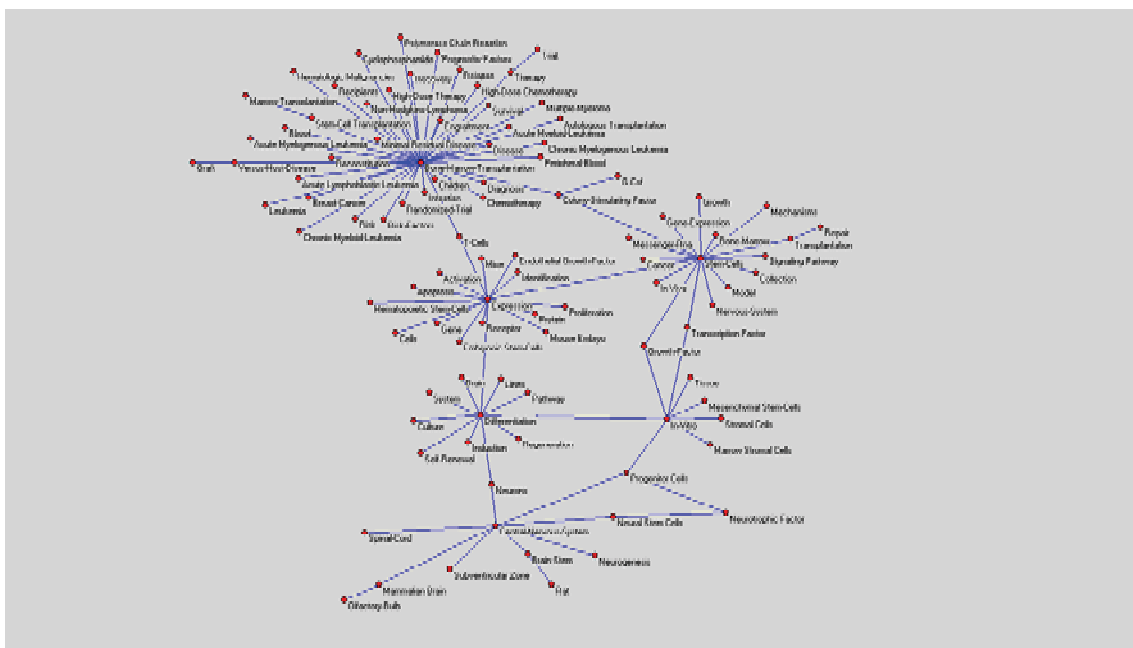
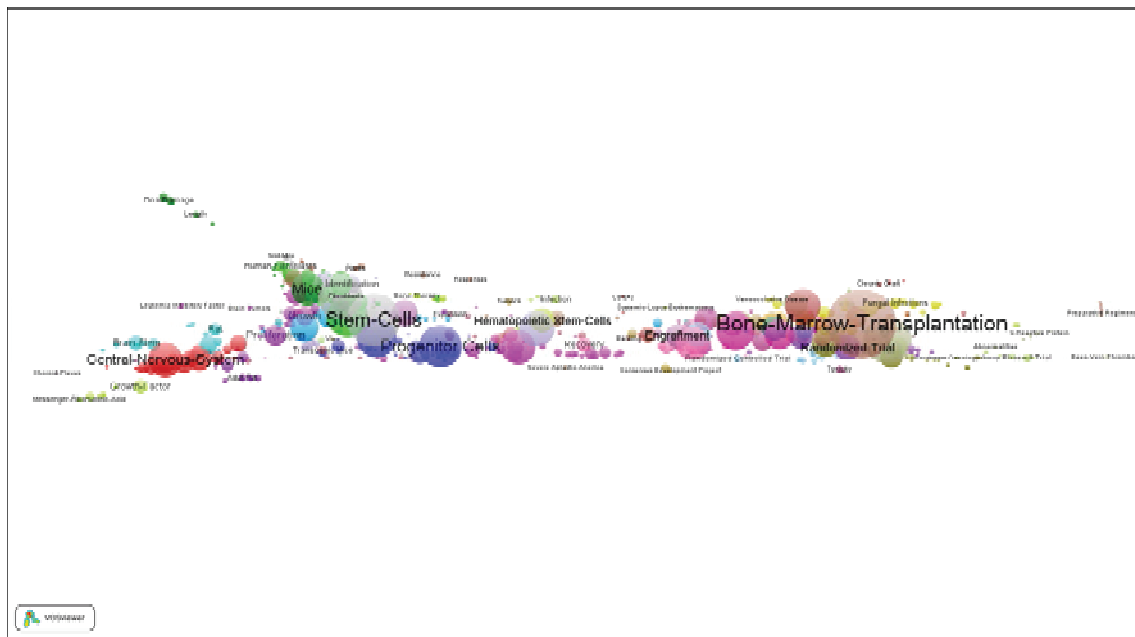
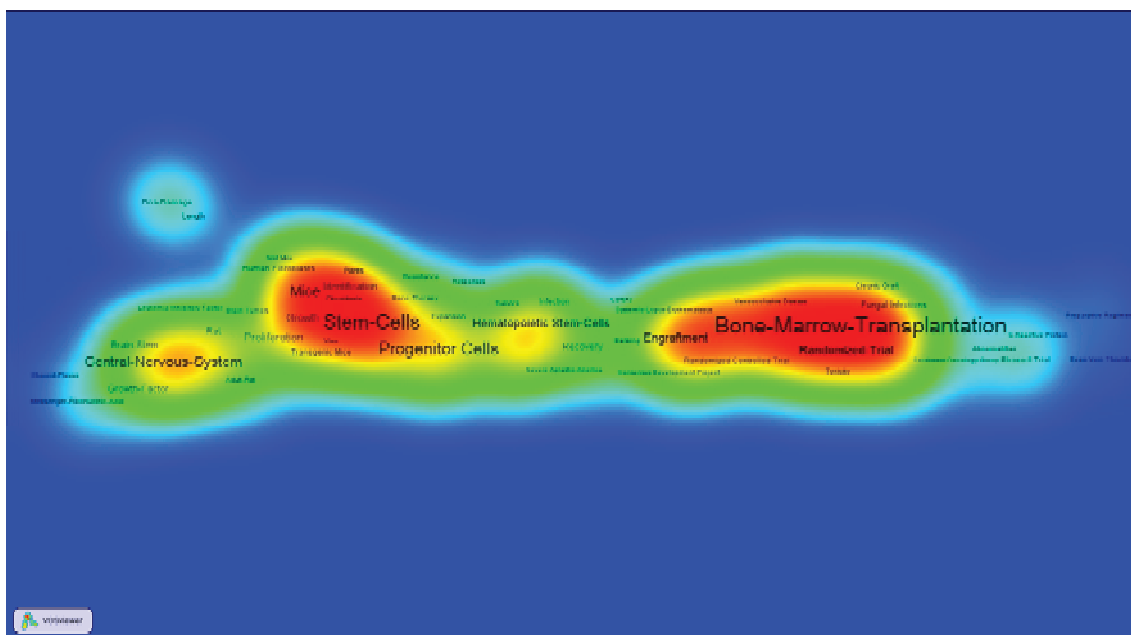


FIG. 2 Período completo (1997-2010). Mapa de clúster representado con VOSviewer.
(Nº de nodos = 668; Fuerza de Asociación = 5)



En la representación de VOSviewer (Fig. 2), no se identifica de una manera tan nítida, el mismo número de agrupaciones que en Pajek. Sólo se identifican claramente 3 líneas de investigación. La representada por el clúster de color marrón encabezada por el descriptor *Bone-Marrow-Transplantation*, la situada en la parte central bajo el descriptor *Stem-Cells*, y la situada en la zona derecha del mapa, representada por el clúster de color rojo encabezada por el descriptor *Central-Nervous-System*. El resto de nodos, aparecen muy entremezclados y son difícilmente detectables.

FIG. 3 Período Completo (1997-2010). Mapa de calor representado con VOSviewer.
(Nº de nodos = 668; Fuerza de Asociación = 5)



3.2. PRIMER PERÍODO (1997-2001)

El volumen de documentos para este periodo es de 724 y el número de descriptores sobre el que se realiza el análisis relacional, de 2.080 KW+ únicos (Tabla 3).

Tabla 3. Datos relativos al Primer Período (1997-2001)

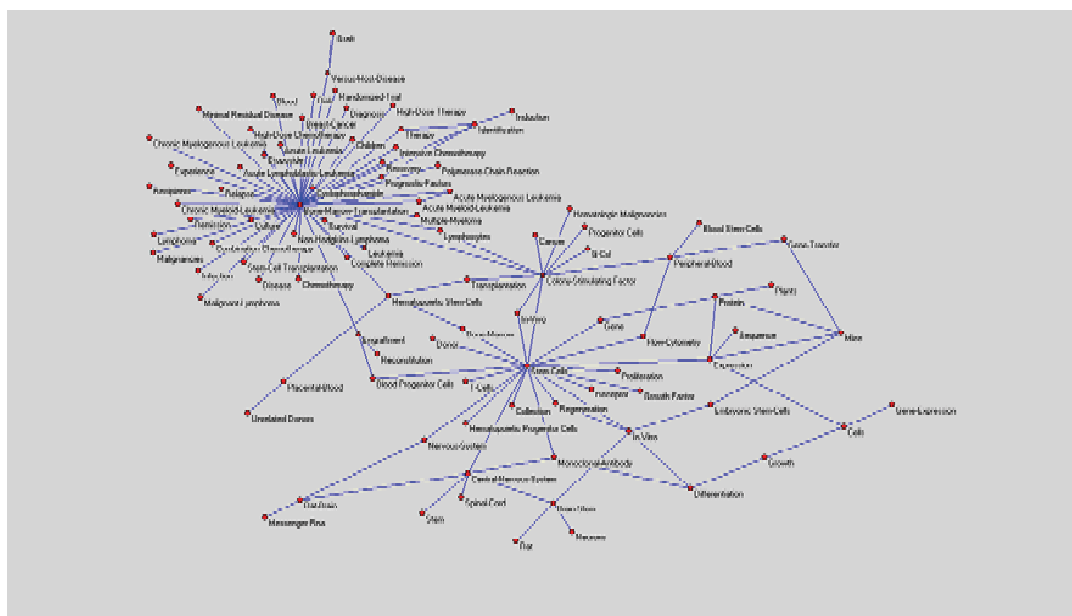
Primer Período (1997-2001)			
Documentos	Totales		724
	Indizados con KW+ (%)		560 (77,35)
KW+	Totales		4.631
	Únicos	Sin Normalización	2.134
		Con Normalización	2.080
	Por documento		8,27
	Frec. Media de aparición		2,23
Visualización	Pajek	Nº Nodos	95
		Frec \geq	8
		Parejas	678
	VOSviewer	Nº Nodos	543
		Fuerza	15

La visualización del dominio ha cambiado sustancialmente, respecto a la imagen del período completo. A pesar de apreciar importantes cambios estructurales en las visualizaciones, la distinción entre los estudios relacionados con la investigación clínica y la básica, es incluso más clara que en el período completo.

La investigación clínica se agrupa claramente en la zona superior izquierda, con un grupo muy numeroso de KW+ relacionados con el uso terapéutico de las células madre en enfermedades hematológicas. La composición de los descriptores de esta línea de investigación es muy similar a la descrita en el período completo. Debido a la clara definición que esta agrupación refleja, es posible considerar que está bien consolidada desde los primeros años del periodo.

El resto de líneas de carácter más básico, no parecen estar tan definidas. Aunque es posible detectar (Figura 4) la mayor parte de los descriptores que estaban presentes en el período anterior, incluso los que ejercían mayor influencia dentro de cada línea de investigación (*Stem Cells, Expression, Differentiation, In-Vitro, Central-Nervous-System*), el comportamiento relacional entre ellos es distinto. La mayoría de las relaciones parten del descriptor *Stem Cells*, nodo principal de la investigación de carácter básico.

FIG. 4 Primer período (1997-2001). Representación en Pajek de los 95 KW+ más frecuentes.
(Frecuencia ≥ 8 ; coocurrencias = 678)



Las visualizaciones de este período sólo permiten identificar 4 agrupaciones. Por un lado, la investigación clínica, y por otro, la básica. En esta última es posible localizar otras dos agrupaciones más, de carácter más emergente y muy poco definidas pero que pueden tener cierta identidad respecto al resto de descriptores. Éstas muestran estudios relacionados con la investigación de células madre neurales y con algunos de los procesos que intervienen en la proliferación de líneas celulares. Es posible que estas débiles agrupaciones se desarrollen de manera más consistente en el futuro, pudiendo ser consideradas como líneas de investigación emergentes.

2.3. SEGUNDO PERÍODO (2002-2006)

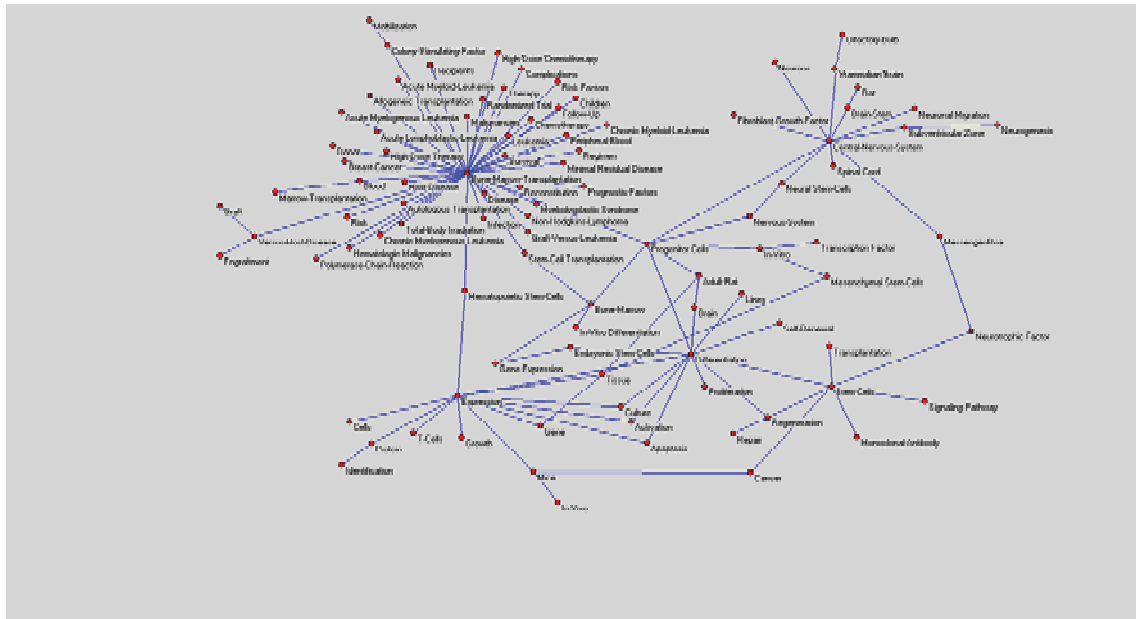
En este período el número de documentos ascendiendo a 1.250 documentos (Tabla 4) y también incrementa el número de KW+ únicos a 3.852 descriptores. En cuanto a las relaciones que se producen entre los KW+, se repite una clara distinción entre los descriptores relacionados con la investigación clínica y la básica con mucho mayor densidad en el caso de la clínica.

TABLA 4. Datos relativos al Segundo Período (2002-2006)

Segundo Período (2002-2006)			
Documentos	Totales		1.250
	Índizados con KW+ (%)		979 (78,32)
KW+	Totales		8.581
	Únicos	Sin Normalización	3.966
		Con Normalización	3.852
	Por documento		8,76
	Frec. Media de aparición		2,23
Visualización	Pajek	Nº Nodos	96
		Frec \geq	11
		Parejas	820
	VOSviewer	Nº Nodos	363
		Fuerza	30

La agrupación más numerosa y definida es la que se localiza en torno al descriptor *Bone-Marrow-Transplantation* (figura 5), que se identifica temáticamente con los estudios relacionados con el uso terapéutico de las células madre en enfermedades hematológicas. La presencia de esta línea de investigación, detectada en el primer período confirma la consolidación de estos estudios a lo largo del tiempo.

FIG. 5 Segundo período (2002-2006). Representación en Pajek de los 96 KW+ más frecuentes. (Frecuencia \geq 11; coocurrencias = 820)



Respecto a la investigación de carácter más básico, estructuralmente se muestran mejor definidos. Los estudios relacionados con las células madre hematopoyéticas identificados mediante los descriptores conectados con *Stem-Cells*, ya no concentran la mayor parte de las relaciones como ocurría en el primer período, sino que han pasado a formar una agrupación con mayor independencia relacional y temática. Al definirse mejor esta línea de investigación también se han definido con mayor claridad los descriptores relacionados con los estudios sobre las células madre embrionarias. Si en el primer período estaban muy entremezclados con los estudios de las células madre hematológicas y tan sólo los procesos de proliferación de líneas celulares podían estar algo más definidos, en este período además de definirse mejor estos estudios (zona inferior del mapa), es posible detectar la presencia de aquellos que se relacionan con los procesos de generación de líneas celulares en torno al descriptor *Expression*. A pesar de la presencia del descriptor *In-Vitro*, aún no se pueden identificar de manera definida los estudios relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y analizar in-vitro las células madre embrionarias como ocurría en el período completo. Es muy posible que esta línea se desarrolle en el siguiente período.

Finalmente, los estudios relacionados con las células madre neurales que aparecían en el período anterior bajo el descriptor *Central-Nervous-System*, no sólo siguen estando presentes para estos años sino que, además, están mejor definidos temáticamente y estructuralmente. Por lo tanto, es posible considerar la consolidación de esta línea de investigación para este período de estudio.

2.4. TERCER PERÍODO (2007-2010)

Este es el periodo con el mayor número de documentos (2.174 documentos) y el mayor número de descriptores (6.178 KW+ únicos). Esta elevada producción puede afectar a la visualización del período completo por la presencia de determinados KW+, y sus frecuencias de aparición que alcanzan valores mayores que otros periodos

TABLA 5. Datos relativos al Tercer Período (2007-2010)

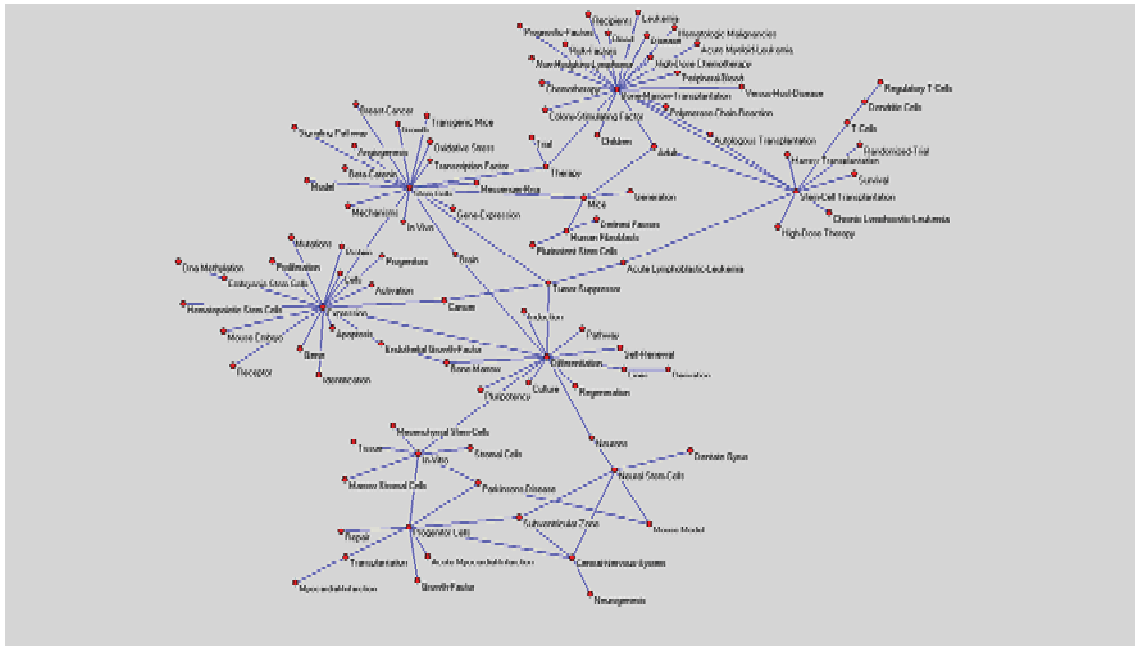
Tercer Período (2007-2010)			
Documentos	Totales		2.174
	Indizados con KW+ (%)		1.707 (78,52)
KW+	Totales		15.521
	Únicos	Sin Normalización	6.393
		Con Normalización	6.178
	Por documento		9,1
	Frec. Media de aparición		2,51
Visualización	Pajek	Nº Nodos	98
		Frec \geq	18
		Parejas	1.103
	VOSviewer	Nº Nodos	390
		Fuerza	50

Las visualizaciones permiten detectar cambios importantes en la evolución de la estructura y composición temática de las líneas de investigación localizadas en los períodos anteriores. La distinción entre los descriptores relacionados con la investigación clínica y la básica sigue siendo clara, pero éstos últimos aparecen mucho más desarrollados.

La presencia de descriptores de carácter clínico es más proporcionada al resto de las líneas de investigación, tanto en número de descriptores como en el peso de sus relaciones. Lo que llama la atención en estos años, es la distinción clara que se aprecia (fig. 6) entre los estudios relacionados con los procesos oncológicos, identificados por el descriptor *Bone-Marrow-Transplantation* y que conectan con la investigación básica a través de *Therapy*. Y por otro lado, los que están más relacionados con el trasplante de progenitores hematopoyéticos, identificados con el descriptor *Stem-Cel Transplantation* y que conectan con la investigación básica a través de *Acute-Lymphoblastic-Leukemia*.

Se identifican cinco agrupaciones de carácter más básico. La primera estaría constituida por los estudios relacionados con las células madre hematopoyéticas. Esta línea de investigación, identificada por los descriptores relacionados con *Stem-Cells*, aparece mucho mejor definida que en el período anterior (figura 6, zona superior izquierda del mapa). Esto significa, que estos estudios se han ido desarrollando hasta consolidarse a lo largo de tiempo.

FIG. 6. Tercer período (2006-2010). Representación en Pajek de los 98 KW+ más frecuentes. (Frecuencia \geq 18; cocurrencias = 1.103)



Esta línea de investigación está conectada con los estudios relacionados con el uso terapéutico de las células madre en enfermedades oncológicas y con la investigación relacionada con las células madre embrionarias. Si en el período anterior era posible localizar estudios relacionados con los procesos de proliferación y diferenciación celular, en este período aparecen mejor definidos. Por lo que se puede decir que esta línea de investigación se consolida. Pero lo hace incorporando una novedad con respecto a periodos anteriores. Se trata de la localización e identificación de manera clara de los estudios relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y sobre todo, analizar *in-vitro* las células madre embrionarias (figura 6, zona inferior). Si bien es cierto, que en el período anterior no era posible identificar de manera definida estos estudios, en este período si es posible.

La última línea de investigación la conforman los estudios relacionados con las células madre neurales. A diferencia del período anterior, los KW+ que identifican esta línea, han perdido definición y ya no constituyen un grupo tan homogéneo. Ahora, es posible localizar dos aspectos de la investigación en este ámbito: la consolidación de los estudios relacionados con el desarrollo e identificación de las células madre neurales mediante los descriptores relacionados con *Neural-Stem-Cell*, y por otro, los relacionados con el infarto de miocardio mediante el empleo de progenitores celulares. Es posible que la pérdida de definición de esta agrupación en este período se deba a la gestación de una nueva línea de investigación que es posible que se consolide en los próximos años como ha ocurrido en el caso de los estudios relacionados con los procesos de proliferación y diferenciación celular.

4. CONCLUSIONES

La visualización de las relaciones que se establecen entre los KW+ ha permitido obtener dos imágenes complementarias de la investigación española sobre células madre, proyec-

tando tanto líneas de investigación claramente consolidadas como las emergentes, lo que nos permite intuir metodológicamente la capacidad de predicción de este tipo de estudios en el desarrollo de un dominio temático a lo largo del tiempo.

El estudio del período completo ha mostrado la estructura estática del dominio con la identificación de las 4 líneas de investigación más consolidadas. Por un lado la investigación relacionada con el uso terapéutico de las células madre en enfermedades hematológicas, por otro, la investigación relacionada con las células madre hematopoyéticas, también la investigación de las células madre embrionarias desagregada en tres aspectos fundamentales como son la proliferación, generación y diferenciación de las líneas celulares, y finalmente, la investigación relacionada con las células madre neuronales.

En cambio, a través del análisis y comparación de la evolución de los resultados por subperíodos, se ha podido reflejar la consolidación de las líneas de investigación emergentes en unos casos y la pérdida de protagonismo en otros. En el primer periodo (1997-2001), se observó una fuerte presencia de los descriptores que representaban los documentos relacionados con la hematología y oncología. En cambio, el resto de líneas no comenzaron a detectarse de manera más clara hasta el segundo período (2002-2006). Fue aquí, donde fue posible localizar claramente agrupaciones de descriptores relacionados con la investigación sobre células madre hematopoyéticas, los descriptores referidos a los estudios sobre la generación, proliferación y diferenciación de células madre y los descriptores relacionados con las células progenitoras neurales. Finalmente, los resultados han mostrado que el peso de estas últimas agrupaciones tuvo una presencia mucho más evidente en el último periodo (2007-2010), en detrimento de las líneas de investigación relacionadas con la aplicación e investigación clínica sobre el trasplante de progenitores hematopoyéticos, a diferencia de lo que ocurría en el primer y segundo período.

En cuanto a los programas de visualización utilizados, el estudio ha mostrado que tanto las visualizaciones en Pajek como en VOSviewer, coinciden tanto en las líneas de investigación como en la localización de los descriptores más influyentes de la red. Pero además de coincidir, se complementan muy bien, ya que Pajek permite detectar de manera más clara la estructura relacional, mientras que VOSviewer posibilita la detección de los grupos y los focos de investigación de una manera más plástica. Sin embargo, el empleo de este tipo de software presenta la limitación de generar sólo visualizaciones correspondientes a un momento determinado de la investigación, sin poder reflejar la evolución en sí misma del dominio científico.

REFERENCIAS

- Batagelj, V., y Mrvar, A. (2010). Pajek 2.0: package for large network analysis, disponible en: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> (accedido 19/01/2012).
- Bishop, A.E., Buttery, L.D.K., y Polak, J.M. (2002). Embryonic stem cells. *Journal of Pathology*, 197 (4), 424-429.
- Boyack, K.W., Börner, K., y Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, 79 (1), 45-60.
- Boyack, K.W., y Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? *Journal of the American Society for Information Science and Technology*, 61 (12), 2389-2404.
- Braam, R.R., Moed, H.F., y Van Raan, F.J. (1991a). Mapping of science by combined co-citation and word analysis. I: Structural Aspects. *Journal of the American Society for Information Science*, 42 (4), 233-251.

- Braam, R.R., Moed, H.F., y Van Raan, F.J. (1991b). Mapping of science by combined co-citation and word analysis. II: Dynamical Aspects. *Journal of the American Society for Information Science*, 42 (4), 252-266.
- Callon, M., Courtail, J.P., Turner, W.A., y Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information Sur Les Sciences Sociales*, 22, 191-235.
- Cambrosio, A., Limoges, C., Courtial, J.P., y Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27 (2), 119-143.
- Cantos-Mateos, G., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., y Zulueta, M.A. (2012). Stem cell research: bibliometric analysis of main research areas through KeyWords Plus. *Aslib Proceedings*, 64 (6), 561-590.
- Cantos-Mateos, G., Zulueta, M.A., Vargas-Quesada, B., y Chinchilla-Rodríguez, Z. (2013). *Estudio comparativo sobre la visualización de redes de co-words a través de los descriptores del Science Citation Index y Medline*. I Congreso ISKO España y Portugal / XI Congreso ISKO España. Fundamentos, métodos y modelos para la organización y representación de la información. Oporto, noviembre (paper).
- Cao, Q.L, Benton, R.L., y Whittemore, S.R. (2002). Stem cell repair of central nervous system injury, *Journal of Neuroscience Research*, 68 (5), 501-510.
- Chen, C., Ibekwe-SanJuan, F., y Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis, *Journal of the American Society for Information Science and Technology*. 61 (7), 1386-1409.
- Chen, D.Z., Huang, M.H., Hsieh, H.C., y Lin, C.P., (2011). Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Informetrics*, 5 (3), 400-412.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., y Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Infometrics*, 5 (1), 146-166.
- De Solla Price, D.J. (1963) *Little Science, Big Science*, New York, USA: Columbia University Press.
- Di Giorgio, F.P., Carrasco, M.A., Siao, M.C., Maniatis T., y Eggan, K. (2007). Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. *Nature Neuroscience*, 10 (5), 608-614.
- Ding, &, Chowdhury, G., y Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, 37 (6), 67-78.
- Garfield, E. (1990). Keywords plus-ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette. *Current Contents*, 1, (32), 5-9.
- Garfield, E., y Sher, I.H. (1993). Keywords Plus™ – Algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44 (5), 298-299.
- Glänzel, W., y Thijs, B. (2011). Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88 (1), 297-309.
- Hammond, J. (1949). Recovery and culture of tubal mouse ova. *Nature*, 163, (4131), 28-29.
- Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14 (1), 10-25.
- Lee, W.H. (2008). How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics*, 76 (3), 503-525.
- Leydesdorff, L., y Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60 (2), 348-362.
- Leydesdorff, L., y Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5 (3), 469-475.

- Martínez Serrano, A., y Bjorklund, A. (1996). Protection of the neostriatum against excitotoxic damage by neurotrophin-producing, genetically modified neural stem cells, *Journal of Neuroscience*, 16 (15), 4604-4616.
- Moya-Anegón, F., Chinchilla Rodríguez, Z., Corera-Álvarez, E., Gómez-Crisóstomo, R., González-Molina, A., Muñoz-Fernández, F., y Vargas-Quesada, B. (2007a). *Indicadores bibliométricos de la actividad científica española: 1990-2004*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F.J., y Herrero-Solana, V. (2007b). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology*, 58 (14), 2167-79.
- Quin, J. (2000). Semantic similarities between a keyword database and a controlled Vocabulary Database: an investigation in the antibiotic resistance Literature. *Journal of the American for Information Science (JASIS)*, 51 (3), 166-180.
- Rafols, I., Porter, A., y Leydesdorff, L. (2010). Science overlay maps: a new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61 (9), 1871-1887.
- Rip, A., y Courtial, P. (1984) Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6 (6), 381-400.
- Small, H., y Upham, S. P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79 (2), 365-375.
- Van Eck, N.J., y Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84 (2), 523-38.
- Van Raan, A.F.J., y Tijssen, R.J.W. (1993). The neural net of neural network research. An exercise in bibliometric mapping. *Scientometrics*, 26 (1), 169-192.
- Vargas-Quesada, B., Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., y Guerrero-Bote, V. (2008). Development of the Spanish scientific landscape: ISI web of science 1990-2005. *Profesional de la Información*, 17 (1), 22-37.
- Vargas-Quesada, B., Chinchilla-Rodríguez, Z., González-Molina, A., y Moya-Anegón, F. (2010). Showing the Essential Science Structure of a Scientific Domain and its Evolution. *Information Visualization*, 9, 288-300.
- Wasserman, S., y Faust, K. (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.
- White, H.D., y Griffith, B.C. (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163-171.
- White, H., y McCain, K. (1998). Visualizing a Discipline: an Author Co-Citation Analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science*, 49 (4), 327-375.
- White, H.D. (2003). Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54 (5), 423-34.
- Zhao, D., y Strotmann, A. (2011). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, 87, 115-131.
- Zitt, M., y Bassecouard, E. (1996). Reassessment of co-citation methods for science indicators: Effect of methods improving recall rates. *Scientometrics*, 37 (2), 223-244.
- Zitt, M., Lelu A., y Bassecouard, E. (2011). Hybrid citation-word representations in Science mapping: portolan charts or research fields? *Journal of the American society for information Science and Technology*, 62 (1), 19-39.
- Zulueta, M.A., Cantos-Mateos, G., Sánchez, C., y Vargas-Quesada, B. (2011). Research involving women and health in the Medline database, 1965-2005, Co-term analysis and visualization of main lines of research. *Scientometrics*, 88 (3), 679-706.

PUBLICACIÓN Nº 5

Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2014). Estudio evolutivo de la investigación española con células madre. Visualización e identificación de las principales líneas de investigación. *El profesional de la información*, 23(3), 259-271.

<http://dx.doi.org/10.3145/epi.2014.may.06>

ESTUDIO EVOLUTIVO DE LA INVESTIGACIÓN ESPAÑOLA CON CÉLULAS MADRE. VISUALIZACIÓN E IDENTIFICACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN

Gisela Cantos-Mateos, María-Ángeles Zulueta, Benjamín Vargas-Quesada y Zaida Chinchilla-Rodríguez



Gisela Cantos-Mateos es licenciada en historia y en documentación y máster en documentación. Becaria predoctoral (programa JAE-Predoc) del *Instituto de Políticas y Bienes Públicos (IPP)* del *Consejo Superior de Investigaciones Científicas (CSIC)*. Realiza su tesis doctoral relacionada con análisis bibliométricos y con la representación y visualización de la información científica. Colabora con el *Grupo SCImago*.

<http://orcid.org/0000-0003-2690-1790>

Consejo Superior de Investigaciones Científicas. Instituto de Políticas y Bienes Públicos (CSIC-IPP)
Albasanz, 26-28. 28037 Madrid, España
gisela.cantos@csic.es



María-Ángeles Zulueta es licenciada en medicina y doctora en ciencias por la *Universidad Autónoma de Madrid*. Profesora titular del *Departamento de Filología, Comunicación y Documentación* de la *Universidad de Alcalá* e investigadora del *Grupo SCImago*. Sus áreas de investigación de interés se enmarcan en el campo de la bibliometría y análisis de dominios científicos. Ha dirigido y colaborado en varios proyectos de investigación financiados con fondos públicos en esta área y es autora de numerosas publicaciones en este campo.

<http://orcid.org/0000-0002-3983-258X>

Universidad de Alcalá. Facultad de Documentación
San Cirilo, s/n. 28804, Alcalá de Henares, Madrid, España
ma.zulueta@uah.es



Benjamín Vargas-Quesada es profesor de técnicas avanzadas de recuperación de información en la *Facultad de Comunicación y Documentación* de la *Universidad de Granada* y miembro del *Grupo SCImago*. Sus líneas de investigación se centran en la visualización de la información científica y en la evaluación de la ciencia y de la comunicación científica.

<http://orcid.org/0000-0001-5115-7460>

Universidad de Granada. Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja, s/n. 18071, Granada, España
benjamin@ugr.es



Zaida Chinchilla-Rodríguez es doctora en documentación e información científica y científica titular del *Consejo Superior de Investigaciones Científicas (CSIC)*, en el *Instituto de Políticas y Bienes Públicos (IPP)* de Madrid. Miembro del *Grupo SCImago*, desarrolla su investigación en el análisis cuantitativo, la representación y visualización de información, redes de colaboración científica y propuestas metodológicas para el diseño de sistemas de información científica.

<http://orcid.org/0000-0002-1608-4478>

Consejo Superior de Investigaciones Científicas. Instituto de Políticas y Bienes Públicos (CSIC-IPP)
Albasanz, 26-28. 28037 Madrid, España
zaida.chinchilla@csic.es

Resumen

Se analiza la investigación española en células madre en el *Science Citation Index (SCI)* entre 1997 y 2010 con una división temporal en tres sub-períodos, mediante técnicas de visualización basadas en redes sociales. Se han utilizado los descriptores *KeyWords Plus (KW+)* como unidad de análisis, su coocurrencia como unidad de medida, y *Pajek* y *VOSviewer* como herramientas para la generación y visualización de redes sociales. Los resultados muestran dos imágenes complementarias de la investigación: la estructura estática, donde se distingue investigación clínica y básica, y la dinámica del análisis evolutivo, mostrando las líneas más consolidadas y las emergentes. Se plantea una propuesta metodológica para la visualización y detección de las principales líneas de investigación a lo largo del tiempo, su aplicabilidad y capacidad de predicción en dominios científicos y geográficos.

Artículo recibido el 30-01-2014

Aceptación definitiva: 05-06-2014

Palabras clave

Análisis de *co-words*, Visualización de la información, Análisis de redes, Células madre, Tendencias de investigación, Evolución de la ciencia, España.

Title: Development of Spanish research on stem cells. Visualization and identification of the main research fronts

Abstract

Using visualization techniques based on social networks, this study aims to analyze stem cell research in Spain, as reflected in the *Science Citation Index (SCI)* database between 1997 and 2010, divided into three sub-periods. The selected unit of analysis was the *KeyWords Plus* descriptors (*KW+*), the unit of measurement was their co-occurrence, and the *Pajek* and *VOSviewer* tools were used to generate and display the social networks. The results show two complementary images of research: the static structure, distinguishing between clinical and basic research, and the evolutionary dynamic, analysing both the most established and the emerging lines. The main contribution of this work is to present a methodology for the visualization and detection of the main research lines over time, demonstrating its applicability and its predictability in scientific and geographic domains.

Keywords

Co-word analysis, Network analysis, Information visualization, Stem cells, Science trends, Research trends, Spain.

Cantos-Mateos, Gisela; Zulueta, María-Ángeles; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida (2014). "Estudio evolutivo de la investigación española con células madre. Visualización e identificación de las principales líneas de investigación". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 259-271.

<http://dx.doi.org/10.3145/epi.2014.may.06>

1. Introducción

Los estudios con células madre son una de las áreas de investigación que más ha evolucionado. Desde que en 1949 el científico **John Hammond** (1949) descubriese el método para mantener los embriones de ratón en cultivo *in vitro*, el progreso en este campo ha ido en aumento, ofreciendo perspectivas muy esperanzadoras en las últimas décadas para el tratamiento de enfermedades que hasta ahora parecían incurables. La investigación en este campo está teniendo grandes repercusiones científicas y sociales, desde los experimentos como la conocida clonación de la oveja Dolly (**Wilmut et al.**, 1997) que tuvieron un especial impacto en el campo de la transferencia nuclear y la diferenciación y expresión celular en 1997, hasta la aparición entre 2006 y 2007 de las células de Yamanaka (**Takahashi; Yamanaka**, 2006; **Takahashi et al.**, 2007) que ofrecen una alternativa mediante la posibilidad de obtener células pluripotentes a partir de la reprogramación de células adultas. Al mismo tiempo se ha generado un importante debate bioético que afecta a la regulación legislativa y la financiación de la investigación con células madre embrionarias, reglamentando así su desarrollo.

Aunque es difícil establecer una relación directa entre las reformas legislativas y la evolución de la investigación, los hitos científicos podrían estar influenciados por las regulaciones legislativas. Pese a las controversias que suscita el desarrollo de determinadas líneas de investigación con células madre, en los últimos años ha habido iniciativas políticas, económicas y legislativas que manifiestan una clara intención de adecuar los hitos científicos a las estructuras sociales de nuestro tiempo. En Europa se pueden destacar importantes esfuerzos para el impulso de la investigación mediante medidas legislativas adoptadas desde 2001: la in-

centivación económica a través del *VII Programa marco* y la creación del *Registro europeo de células madre embrionarias* con el objetivo de proporcionar información a toda la comunidad científica, sobre las líneas de células disponibles en Europa. España ha ido modificando su legislación adoptando las directivas europeas hasta 2006. Estados Unidos, Reino Unido, Australia y Asia han ido legislando al respecto, y a lo largo de los años las regulaciones han sido sometidas a una continua revisión (*EuroStemCell et al.*, 2013). En cualquier caso, las normativas y la financiación de proyectos de investigación con células madre podrían afectar a los resultados de la producción científica tanto a nivel internacional como nacional. Aunque no es el objeto de este estudio, parece adecuado plantear el marco en el que se desarrolla este campo científico.

“ La idea es presentar un análisis aplicable no sólo a la investigación española con células madre, sino a otros dominios temáticos, geográficos y temporales ”

Todos estos hitos han contribuido a generar importantes resultados de esta actividad científica, susceptibles de ser analizados desde la perspectiva bibliométrica, especialmente porque esta investigación con células madre ha trascendido a otros campos científicos, siendo un tema con un importante carácter multidisciplinar. Por ello es un reto definir temáticamente el dominio y las principales líneas de investigación en las que se está trabajando.

Los estudios bibliométricos ayudan en esta tarea ya que el estudio de nuevas técnicas y herramientas ha mejorado

notablemente la comprensión y análisis de dominios científicos (Leydesdorff; Rafols, 2009; Vargas-Quesada *et al.*, 2010; Rafols; Porter; Leydesdorff, 2010; Boyack; Börner; Klavans, 2009), y especialmente en el campo de la visualización de la información (Börner; Chen; Boyack, 2003; De-Moya-Anegón *et al.*, 2006; Vargas-Quesada *et al.*, 2008). La incorporación del análisis estructural de redes sociales (Wasserman; Faust, 1998) lo ha convertido en una herramienta fundamental para visualizar y analizar las relaciones estructurales que se producen en un contexto científico determinado. En el caso de la delimitación temática ayuda a visualizar la interacción de los aspectos más destacados en el dominio, haciendo posible la identificación de áreas o líneas de investigación.

La obtención de los descriptores *KeyWords Plus* no requiere de minería textual ni de ningún pre-procesamiento lingüístico

Existe gran diversidad de métodos y técnicas a la hora de delimitar y visualizar un campo científico. Hay estudios que utilizan la cocitación de autores (*author co-citation analysis* o *ACA*) (White; Griffith, 1981; White; McCain, 1998; Small; Upham, 2009; Chen; Ibekwe-SanJuan; Hou, 2010; Zhao; Strotmann, 2011), el análisis de emparejamiento bibliográfico (*bibliographic coupling*) (Kessler, 1963; Boyack; Klavans, 2010; Chen *et al.*, 2011; Glänzel; Thijs, 2011), estudios de *co-words* (Callon *et al.*, 1983; Rip; Courtial, 1984, Cambrosio *et al.*, 1993; Van-Raan; Tijssen, 1993; Ding; Chowdhury; Foo, 2001; Lee, 2008; Leydesdorff; Welbers, 2011; Zulueta *et al.*, 2011; Cantos-Mateos *et al.*, 2013; Romo-Fernández; Guerrero-Bote; De-Moya-Anegón, 2013) o los estudios híbridos en los que se combinan tanto *ACA* como *co-words* (Zitt; Bassecouard, 1996; Zitt; Lelu; Bassecouard, 2011).

En vista de las ventajas e inconvenientes que condicionan la elección de una metodología sobre otra, para el presente estudio se ha optado por el análisis de *co-words* (Zitt; Lelu; Bassecouard, 2011). Según Michel Zitt y colaboradores (2011), las palabras, a diferencia de las citas, parecen adaptarse mejor a la “ciencia viva” porque son universales. Además las citas podría considerarse que aluden a temas menos actuales, ya que conforman el trasfondo del documento, mientras que las palabras son coetáneas al propio documento (Soos; Kampis; Gulyás, 2013). Para estos autores las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes.

Se puede apreciar con mayor claridad la dinámica del campo si se aplica una perspectiva temporal. Este tipo de análisis permite detectar la evolución de los principales temas de investigación a lo largo del tiempo (Braam; Moed; Van-Raan, 1991; Gábor, 2006; Leydesdorff; Schank, 2008; Boyack; Börner; Klavans, 2009; Chen; Ibekwe-SanJuan; Hou, 2010; Soos; Kampis; Gulyás, 2013). En este caso las visualizacio-

nes de redes de *co-words* de distintos sub-períodos permiten localizar qué temas persisten, desaparecen o emergen en el campo (Braam; Moed; Van-Raan, 1991).

A partir de este marco, los objetivos de este estudio son de carácter fundamentalmente metodológicos. La idea es presentar un análisis que sea aplicable no sólo a la investigación española con células madre, sino a otros dominios temáticos, geográficos y temporales, distintos al que se aborda en este trabajo. Por tanto la presente propuesta metodológica pretende por un lado identificar las principales líneas de investigación a partir de la localización y visualización de la estructura temática del dominio. Y por otro lado reflejar la dinámica del campo a través de la evolución que experimentan las principales líneas de investigación a lo largo del período de estudio. Todo ello a través de la visualización de redes de *co-words*¹.

2. Material y métodos

2.1. Fuente de información

Los documentos de este estudio proceden de la base de datos *Science Citation Index (SCI)*. La estrategia de búsqueda empleada consistió en utilizar los términos *stem* cell** en el campo *Topic*, *Spain* en el campo *Address* y limitada al período temporal de 1997-2010. Se recuperó un total de 4.148 documentos.

2.2. Unidades de análisis

Se han seleccionado los descriptores *KeyWords Plus (KW+)*² como unidades de análisis. Se trata de un vocabulario automatizado que utiliza el *SCI* para indizar los documentos, que se obtiene mediante un algoritmo de búsqueda que extrae las palabras clave de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Garfield; Sher, 1993). En la investigación con células madre estos descriptores tienen una elevada representatividad (78,25% de los documentos) en comparación con los *Author keywords*³ (52,96%), como ha quedado de manifiesto en trabajos anteriores (Cantos-Mateos *et al.*, 2013).

El uso de descriptores *KeyWords Plus* permite realizar un análisis temático de granularidad fina y contiene terminología más actualizada y específica, en comparación con otros descriptores

El uso de estos descriptores tiene la ventaja de que su obtención no requiere de minería textual ni de ningún pre-procesamiento lingüístico. Permite además realizar un análisis temático de granularidad más fina que otro tipo de unidades como pueden ser las categorías, ya que aluden directamente al contenido conceptual de los documentos (Soos; Kampis; Gulyás, 2013). Por otro lado, al tratarse de términos procedentes de los títulos de los artículos que aparecen en las referencias, son los propios autores quienes seleccionan y comunican cuáles son los términos esenciales para expresar el contenido más importante de sus trabajos de investigación. Esta característica favorece la aparición de una

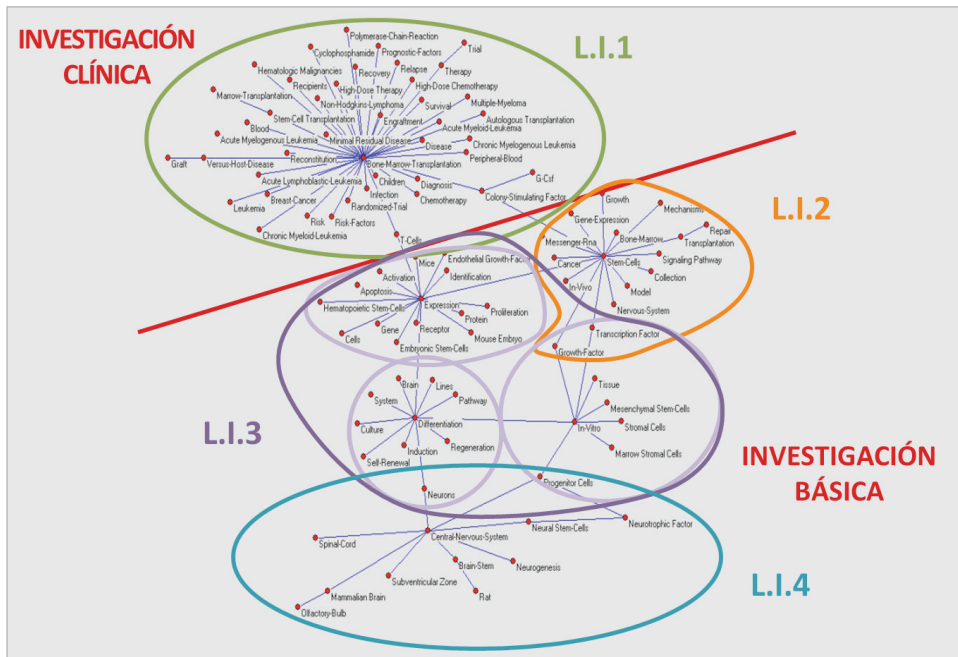


Figura 1. Período completo (1997-2010). Representación en Pajek de los 99 KW+ más frecuentes.

terminología más actualizada y específica en comparación con otros tipos de descriptores, como son los procedentes de un vocabulario controlado (Braam; Moed; Van-Raan, 1991). Como demuestran estudios recientes, estas cualidades contribuyen a reflejar mejor los aspectos temáticos más dinámicos de un dominio.

Los KW+ tienen la desventaja sin embargo de presentar una amplia dispersión terminológica debida a la falta de normalización. Este aspecto hay que resaltarlo como una de las posibles limitaciones de este trabajo. Para paliarla en cierto modo se ha realizado una normalización moderada en la que se han tratado de unificar criterios gramaticales (plurales y singulares) y ortográficos (guiones, mayúsculas y minúsculas).

2.3. Unidades de medida y análisis temporal

El análisis bibliométrico para la detección y visualización de las principales líneas de investigación y su evolución a lo largo del tiempo, ha consistido en calcular las relaciones que se producen entre los KW+ a partir de su aparición conjunta en los documentos, lo que se conoce como *co-words analysis*. Dicho análisis consiste en generar una matriz cuadrada de NxN elementos, donde N es el descriptor KW+ a representar, a partir de las veces que ocurre en los documentos. El

Tabla 1. Períodos temporales analizados

Períodos	Años
Período completo	1997-2010
Primer período	1997-2001
Segundo período	2002-2006
Tercer período	2007-2010

Tabla 2. Número de documentos por años

Período	Primer período					Segundo período					Tercer período			
Años	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nº docs.	94	142	176	149	163	167	165	296	271	351	395	494	608	677
Total docs.	724					1.250					2.174			

resultado es una matriz de co-ocurrencias que refleja el número de veces que un par de descriptores aparecen conjuntamente en dos documentos.

Para el estudio evolutivo se ha analizado el período completo en su conjunto (1997-2010), y subdivido en otros tres períodos, como muestra la tabla 1. La segmentación temporal ha sido establecida considerando dos criterios complementarios.

El primero son los principales hitos científicos. Partiendo de 1997, cuando tuvieron lugar los exitosos ensayos del *Roslin Institute* de Edimburgo, se desencadenó un tiempo de incertidumbre en la regulación legislativa y las políticas de financiación pública que condicionaron el avance de la investigación en células madre embrionarias.

El proceso comenzó a normalizarse fundamentalmente entre los años 2002 y 2006. A partir de 2007 esta investigación recibió un importante impulso gracias al descubrimiento de las células madre de pluripotencia inducida (células iPS) a partir de células humanas adultas que planteaban una alternativa a los problemas éticos-jurídicos sobre la obtención de células madre embrionarias obtenidas a partir de la masa celular interna de un embrión, lo que suponía la destrucción del mismo (*EuroStemCell et al., 2013*).

El segundo criterio ha sido que el volumen documental de cada período sea lo suficientemente representativo como para permitir la detección de líneas de investigación.

2.4. Visualización de la información

La visualización de las relaciones entre los KW+ se ha hecho siguiendo dos métodos en base al uso de dos programas especializados en visualización y análisis de grandes redes, que permiten obtener unas representaciones que favorecen la identificación de las principales líneas de investigación y su evolución.

Por un lado se ha empleado el programa *Pajek* (Batagelj; Mrvar, 2010). Se utilizó una matriz de co-ocurrencias de descriptores con datos en bruto en combinación con el algoritmo de poda *Pathfinder* (De-Moya-Aneón et al., 2007a, 2007b; Zulueta et al., 2011) para la reducción del espacio dimensional. Este algoritmo está basado en el principio de desigualdad del triángulo bajo los parámetros $r=\infty$ y $q=n-1$. Éste ayuda a preservar y poner de relieve las relaciones más

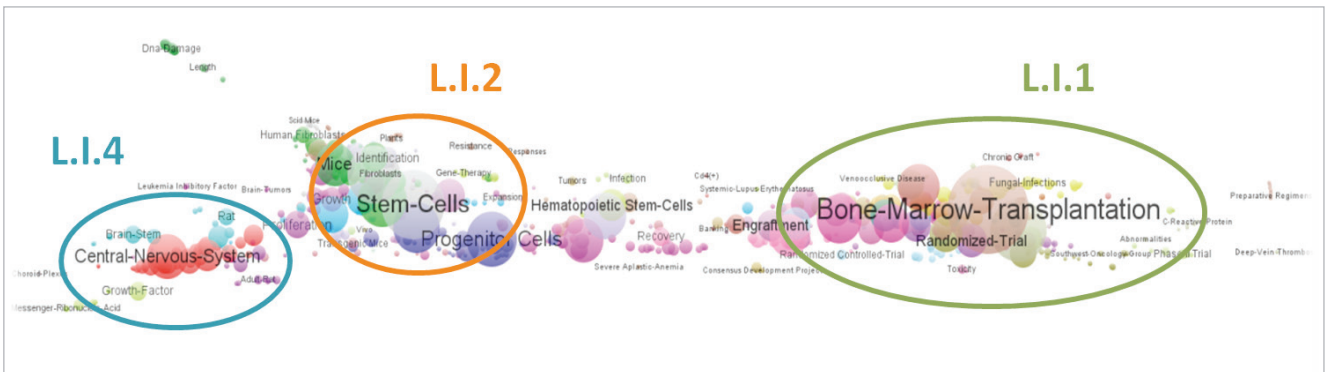


Figura 2. Período completo (1997-2010). Mapa de clusters representado con VOSviewer (número de nodos: 668).

destacadas entre los descriptores (White, 2003; Vargas-Quesada *et al.*, 2008; De-Moya-Anegón *et al.*, 2009). Como algoritmo de visualización se utilizó Kamada-Kawai (Kamada; Kawai, 1989) incluido en Pajek. Aunque este programa permite representar grandes redes, se decidió incorporar únicamente los descriptores más frecuentes. Tras varios ensayos donde se probaron distintos umbrales de frecuencias para la generación de los mapas, se observó que el mejor compromiso entre visualización y análisis, teniendo en cuenta que se trabaja con las limitaciones propias de la pantalla de un ordenador, es el de redes compuestas en torno a los 100 nodos (Zulueta *et al.*, 2011).

Por otro lado se ha utilizado el programa VOSviewer v. 1.5.3 (Van-Eck; Waltman, 2010), que es una alternativa a las representaciones multidimensionales combinando técnicas de visualización y clustering. En este caso se ha construido una matriz de co-ocurrencia normalizada a través de la medida de similitud que utiliza VOSviewer, basada en la fuerza de asociación de los ítems. El umbral de dicha co-ocurrencia ha sido fijo para todas las representaciones. De los cuatro tipos de visualizaciones que ofrece, se han utilizado los mapas etiquetados con nombres, donde cada descriptor está representado por una etiqueta y cuyo tamaño es proporcional a su peso, junto con los denominados “mapas de calor” (o mapas de densidad). Estos últimos se caracterizan porque cada nodo en el mapa se representa con un color que va del rojo al azul, reflejando la densidad de las relaciones entre los descriptores. Cuanto mayor es la densidad, es decir la co-ocurrencia entre descriptores, más se aproximarán a la tonalidad roja; en cambio la mayor aproximación al azul indica una mayor dispersión y por lo tanto menos co-ocurrencia.

Tabla 3. Datos relativos al período completo (1997-2010)

Período completo (1997-2010)			
Documentos	Totales	4.148	
	Indizados con KW+	3.246 (78,25%)	
KW+	Totales	28.733	
	Únicos	Sin normalización	9.833
		Con normalización	9.465
	Por documento	8,85	
	Frecuencia media de aparición	3,04	

3. Resultados y discusión

El número de documentos recuperados fue de 4.148 para el período 1997-2010. La tabla 2 muestra el volumen de producción por tramos temporales.

3.1. Período temporal completo (1997-2010)

De los 4.148 documentos recuperados para todo el período, el 78,25% están indizados con KW+, contabilizándose 28.733 descriptores, de los cuales 9.465 son únicos (tabla 3).

En la figura 1 se aprecian hasta 6 agrupaciones de descriptores KW+ que, por su significado y relación con los demás, podemos considerar que definen líneas de investigación. Se distinguen estructuralmente dos naturalezas temáticas de la investigación claramente diferenciadas. Por un lado la investigación clínica representada por los descriptores agrupados bajo Bone-Marrow-Transplantation (figuras 1 y 3), relacionados con los estudios sobre el uso terapéutico de

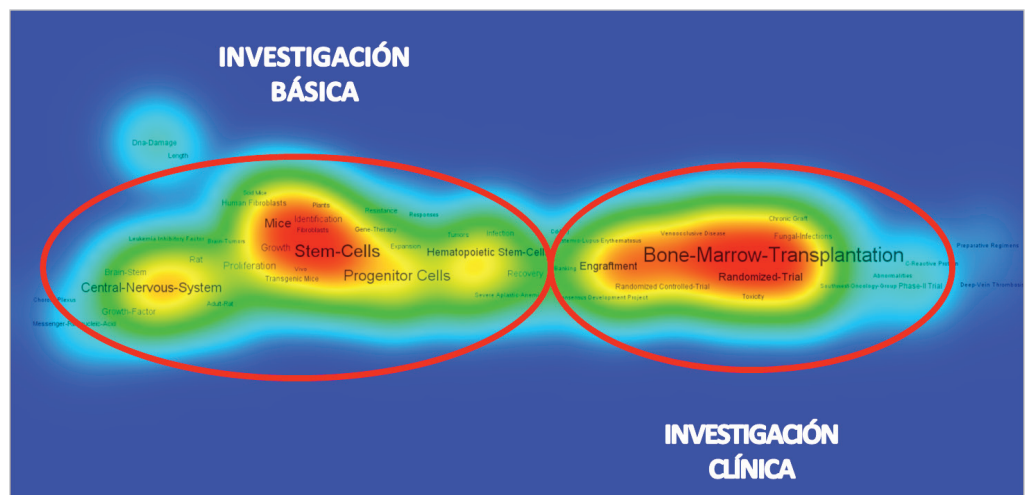


Figura 3. Período completo (1997-2010). Mapa de calor representado con VOSviewer.

Tabla 4. Datos del primer período (1997-2001)

Primer período (1997-2001)			
Documentos	Totales	724	
	Indizados con KW+	560 (77,35%)	
KW+	Totales	4.631	
	Únicos	Sin normalización	2.134
		Con normalización	2.080
	Por documento	8,27	
Frecuencia media de aparición	2,23		

las células madre en enfermedades hematológicas (L.I.1). Por otro la investigación básica representada por el resto de agrupaciones.

En la investigación básica (figura 1) existen dos agrupaciones que conectan directamente con la clínica: los descriptores vinculados con Stem Cells, relacionados con los estudios sobre células madre hematopoyéticas (L.I.2) y los que tie-

nen relación con las células madre embrionarias (L.I.3). Esta línea se encuentra desagregada en tres agrupaciones: procesos de expresión, diferenciación celular y procesos y biotecnologías necesarias para localizar, producir, crecer y analizar in-vitro las células madre embrionarias.

La última agrupación se relaciona con la investigación con células madre neurales (L.I.4). Estos descriptores están muy ligados con los estudios sobre desarrollo e identificación de las células madre neuronales que se diferencian activamente en el tejido cerebral.

En la representación de *VOSviewer* (figura 2) no se identifica de una manera tan nítida el mismo número de agrupaciones. Sólo se identifican claramente 3:

- agrupación del clúster marrón encabezada por el descriptor Bone-Marrow-Transplantation (L.I.1);
- la situada en la parte central bajo el descriptor Stem-Cells (L.I.2);
- la situada en la zona de la derecha, representada por el clúster de color rojo encabezada por el descriptor Central-Nervous-System (L.I.4).

El resto de los nodos aparecen muy entremezclados y son difícilmente detectables.

En las zonas con mayor densidad de la red (figura 3), se distinguen claramente dos zonas que concentran una importante actividad de relaciones, una de investigación básica y otra de clínica.

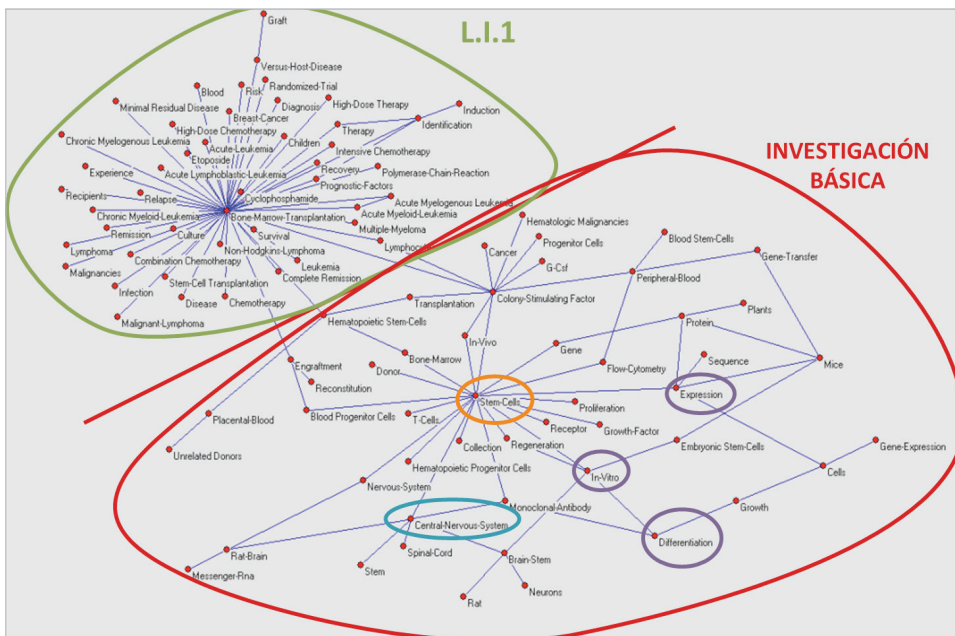


Figura 4. Primer período (1997-2001). Representación en *Pajek* de los 95 KW+ más frecuentes.

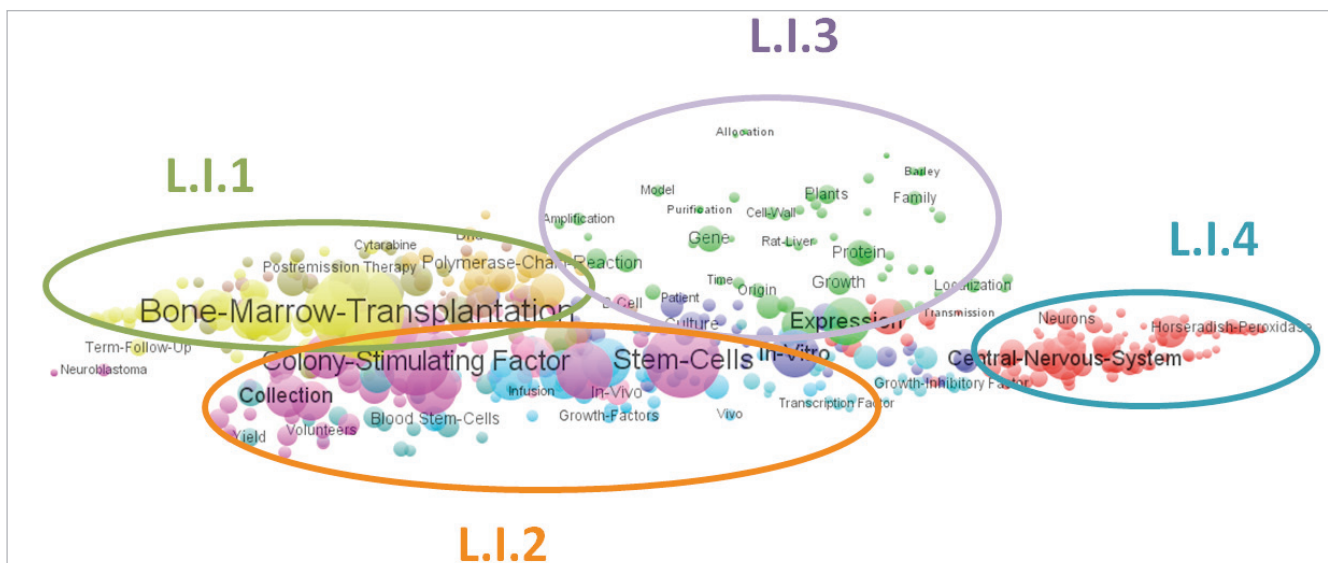


Figura 5. Primer período (1997-2001). Mapa de clusters representado con *VOSviewer* (número de nodos: 543)

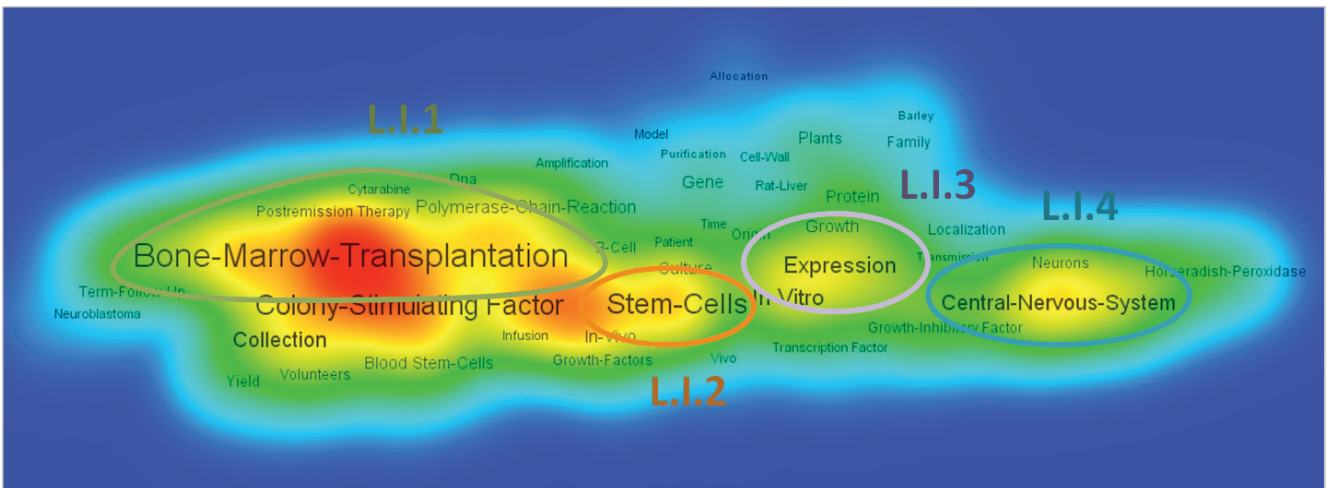


Figura 6. Primer período (1997-2001). Mapa de calor representado con VOSviewer.

3.2. Primer período (1997-2001)

El número de documentos es de 724 y el de descriptores de 2.080 KW+ únicos (tabla 4).

La visualización del dominio ha cambiado sustancialmente. A pesar de ello, la distinción entre la investigación clínica y la básica sigue siendo muy clara, incluso más, que con respecto al período completo.

En la figura 4 se aprecia un grupo muy numeroso de KW+ relacionados con la investigación clínica. La composición de los descriptores de esta línea (L.I.1) es muy similar a la que se ha descrito en el período completo. Su clara definición refleja que estos estudios están muy consolidados.

El resto de las líneas que integraban la investigación de carácter más básico (L.I.2, L.I.3 y L.I.4) no parecen estar tan definidas. Según la figura 4, es posible detectar la mayor parte de los descriptores presentes en el período completo, incluso los que ejercían mayor influencia dentro de cada línea de investigación. A pesar de ello, el comportamiento relacional entre ellos es distinto en este primer período. La mayoría de relaciones parten del descriptor Stem Cells, constituyendo el nodo principal. También se aprecia en la figura 5, donde constituye uno de los cuatro clusters que mejor se diferencian en el mapa. De manera más específica y poco definida, se detectan dos agrupaciones que mantienen cierta identidad:

- procesos que intervienen en la proliferación de líneas celulares (figura 5, clúster verde);
- relacionados con células madre neurales (figura 5, clúster rojo).

Lo mismo se puede localizar

Tabla 5. Datos del segundo período (2002-2006)

Segundo período (2002-2006)			
Documentos	Totales	1.250	
	Indizados con KW+	979 (78,32%)	
KW+	Totales	8.581	
	Únicos	Sin normalización	3.966
		Con normalización	3.852
	Por documento	8,76	
Frecuencia media de aparición	2,23		

en el mapa de densidades de la figura 6. Estas débiles agrupaciones forman líneas de investigación emergentes, como se muestra en los siguientes períodos.

3.3. Segundo período (2002-2006)

El número de documentos asciende a 1.250. El número de KW+ únicos también se incrementa, hasta 3.852 descriptores.

Se mantiene la distinción entre investigación clínica y básica. En la figura 9 se aprecia que esos son los dos focos más importantes, aunque la densidad es mucho mayor en la investigación clínica.

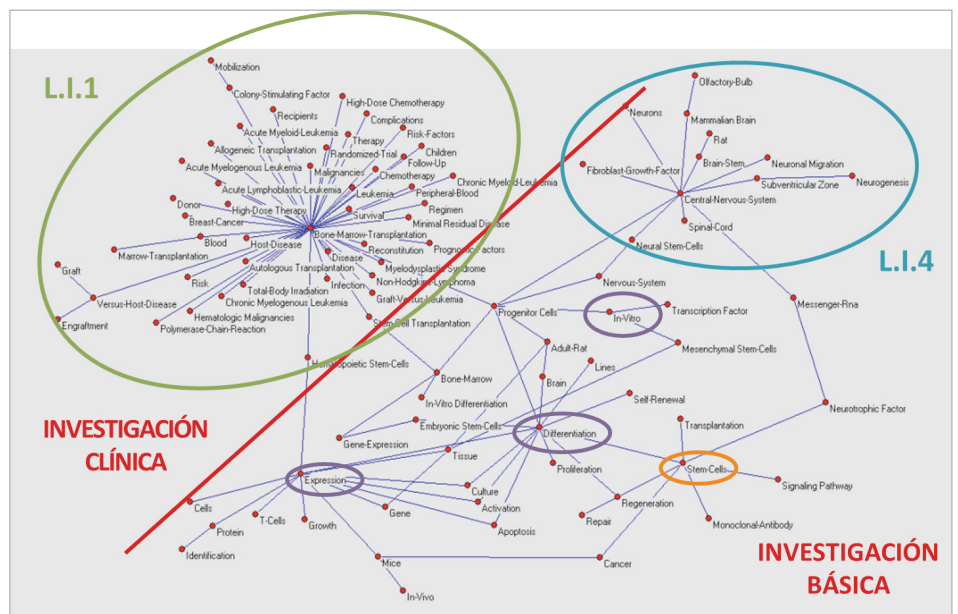


Figura 7. Segundo período (2002-2006). Representación en Pajek de los 96 KW+ más frecuentes.

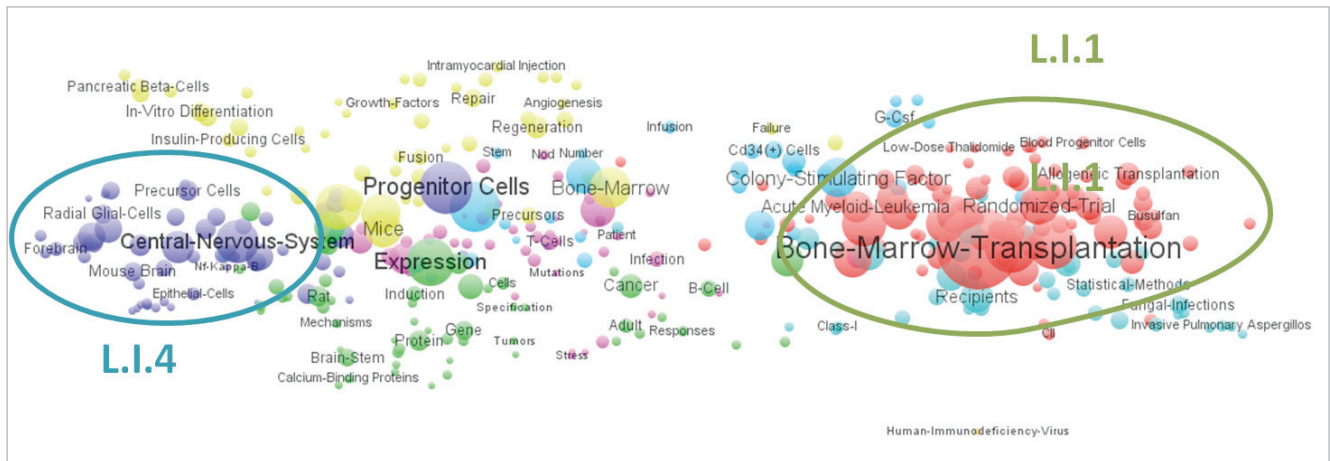


Figura 8. Segundo período (2002-2006). Mapa de clusters representado con VOSviewer (número de nodos: 363).

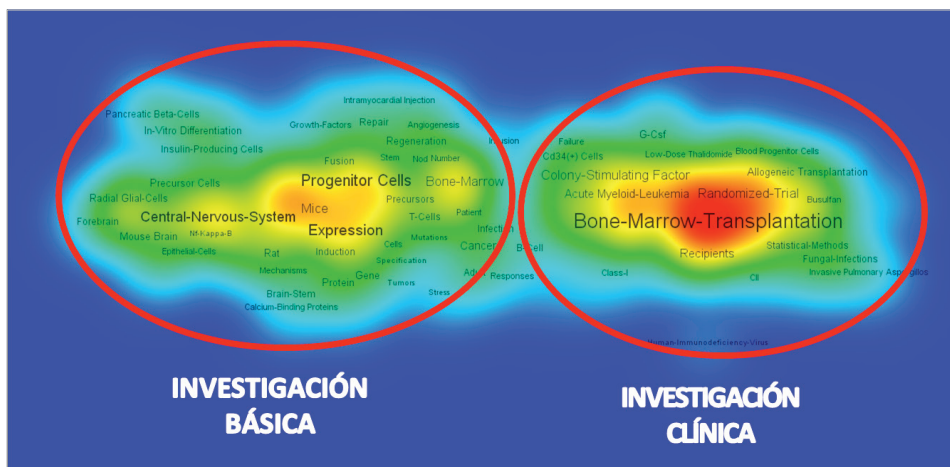


Figura 9. Segundo período (2002-2006). Mapa de calor representado con VOSviewer.

La agrupación más numerosa y definida (figura 7) es la del uso terapéutico de las células madre en enfermedades hematológicas (L.I.1). La presencia de esta línea confirma la consolidación de estos estudios para este período.

La investigación de carácter más básico se muestra estructuralmente mejor definida. Los estudios relacionados con las células madre hematopoyéticas (L.I.2) identificados mediante los descriptores conectados con Stem-Cells, ya no concentran la mayor parte de las relaciones como ocurría en el primer período, sino que han pasado a formar una agrupación con mayor independencia relacional (figura 7) y temática (figura 8, clúster amarillo). Al definirse mejor esta línea de investigación también se han definido los descriptores relacionados con los estudios sobre las células madre embrionarias (figura 7, zona inferior; figura 8, clúster verde; L.I.3). Si en el primer período estaban muy entremezclados con los estudios de las células madre hematopoyéticas (L.I.2), en este período se definen mejor y además es posible detectar los descriptores relacionados con los procesos de generación de líneas celulares en torno al descriptor Expression (figura 7).

Finalmente los estudios sobre células madre neurales (L.I.4) están mejor definidos tanto temática como estructuralmente. Por tanto se puede señalar la consolidación de esta línea

de investigación durante este período.

3.4. Tercer período (2007-2010)

Es el período con mayor número de documentos (2.174) y descriptores (6.178 KW+ únicos).

Las visualizaciones permiten detectar cambios importantes en la evolución de la estructura y composición temática de las líneas de investigación. La distinción entre investigación clínica y básica sigue siendo clara (figura 12),

aunque esta última aparece mucho más desarrollada.

La presencia de descriptores de carácter clínico (L.I.1) es más proporcionada al resto de las líneas de investigación, tanto en número como en el peso de sus relaciones. Llama la atención en estos años la distinción (figura 10) de los estudios sobre procesos oncológicos, identificados por el descriptor Bone-Marrow-Transplantation y de los de trasplante de progenitores hematopoyéticos, identificados con el descriptor Stem-Cell Transplantation.

Se identifican 5 agrupaciones de carácter más básico. La primera, mucho mejor definida que en el período anterior, son los estudios sobre células madre hematopoyéticas (figura 10, zona superior izquierda; figura 11, clúster verde; L.I.2), lo que significa su consolidación. Asimismo está conectada con los estudios de uso terapéutico de las células madre

Tabla 6. Datos relativos al tercer período (2007-2010)

Tercer período (2007-2010)			
Documentos	Totales	2.174	
	Indizados con KW+	1.707 (78,52%)	
KW+	Totales	15.521	
	Únicos	Sin normalización	6.393
		Con normalización	6.178
	Por documento	9,1	
	Frecuencia media de aparición	2,51	

en enfermedades hematológicas (L.I.1) y con la investigación en células madre embrionarias (L.I.3). En este período aparecen mejor definidos los estudios sobre procesos de proliferación y diferenciación celular y los relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y sobre todo, analizar in-vitro las células madre embrionarias (figura 10, zona inferior; figura 11, clúster azul oscuro).

La última línea de investigación (L.I.4) la conforman los estudios relacionados con

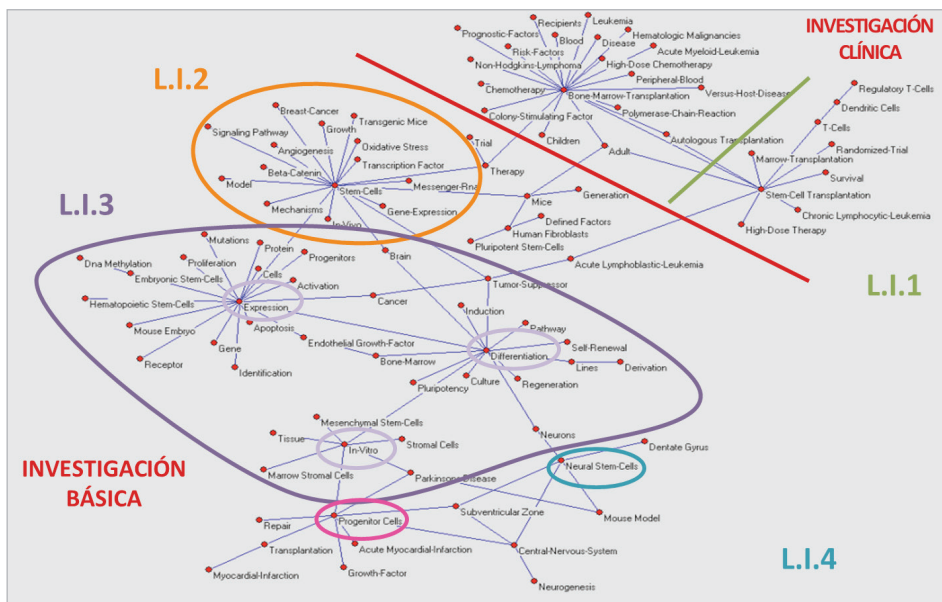


Figura 10. Tercer período (2007-2010). Representación en Pajek de los 98 KW+ más frecuentes.

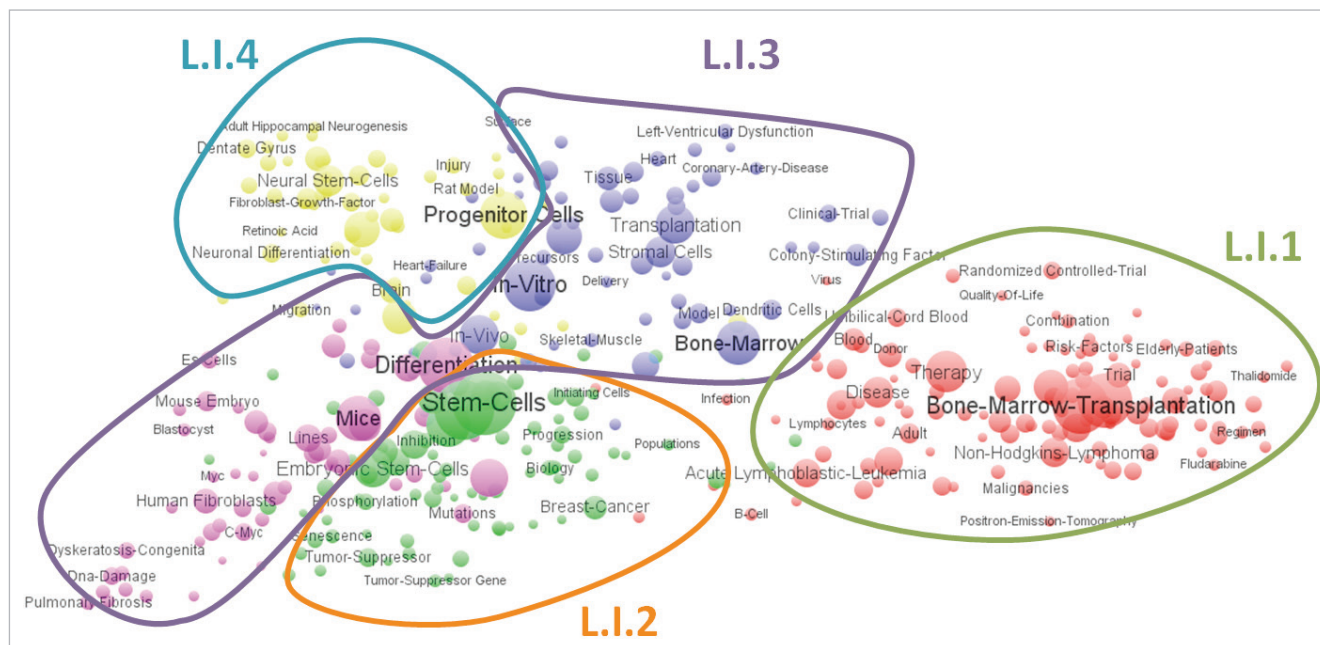


Figura 11. Tercer período (2007-2010). Mapa de clusters representado con VOSviewer (número de nodos: 390).

las células madre neurales (figura 11, clúster amarillo). En este período los KW+ han perdido definición y ya no constituyen un grupo tan homogéneo. Incluso se han desagregado en dos aspectos de la investigación: el desarrollo e identificación de las células madre neurales y el infarto de miocardio empleando progenitores celulares (figura 10). Es posible que esta pérdida de definición se deba a



Figura 12. Tercer período (2007-2010). Mapa de calor representado con VOSviewer.

una nueva línea de investigación emergente que pueda consolidarse con los años.

El estudio del período completo identifica las 4 líneas de investigación más consolidadas: uso terapéutico de células madre en enfermedades hematológicas; células madre hematopoyéticas; embrionarias; y neurales

4. Conclusiones

La evolución de la investigación española y su crecimiento están en consonancia con la que se ha llevado a cabo a nivel internacional, tanto en el crecimiento de la literatura a lo largo del período, como su fuerte expansión en el último período analizado. También las líneas de investigación detectadas en las visualizaciones obtenidas en este estudio están relacionadas con la evolución científica internacional, especialmente en la investigación en células madre embrionarias (*EuroStemCell et al.*, 2013).

El estudio del período completo muestra la estructura estática del dominio con la identificación de las 4 líneas de investigación más consolidadas:

- uso terapéutico de células madre en enfermedades oncohematológicas (L.I.1);
- células madre hematopoyéticas (L.I.2);
- células madre embrionarias (L.I.3);
- células madre neuronales (L.I.4).

A través del análisis evolutivo se aprecia cómo se han configurado estas líneas a partir de la consolidación de algunas emergentes y la pérdida de protagonismo de otras. En el primer período hay una fuerte presencia de la hematología y la oncología. El resto de las líneas no comenzaron a detectarse hasta el segundo período con la investigación sobre células madre hematopoyéticas. Finalmente los resultados han mostrado que el peso de estas últimas agrupaciones tuvo una presencia mucho más evidente en el último período, en detrimento de las líneas de investigación relacionadas con la aplicación y la clínica, a diferencia de lo que ocurría en el primer y segundo períodos.

Con estos resultados no es posible concluir que exista una relación directa entre los hitos científicos y normativos acontecidos en España entre los años 1997 y 2010. Sin embargo, parece que las tendencias de investigación están en sintonía con la evolución del marco jurídico-legislativo español y con las principales acciones de incentiación científica en el campo de las células madre. Aunque no es el objeto de estudio, sí parece interesante plantear como estudios futuros un análisis que examine si hay una relación causa-efecto entre los hitos normativos y los científicos.

La propuesta metodológica basada en la visualización de las relaciones entre los KW+ ha permitido obtener dos imágenes complementarias de la investigación española sobre cé-

lulas madre, proyectando tanto líneas de investigación claramente consolidadas como emergentes. Esto permite intuir metodológicamente la capacidad de predicción de este tipo de estudios en el desarrollo de un dominio temático a lo largo del tiempo. Su aplicabilidad trasciende a otros dominios científicos. La rigidez de las clasificaciones documentales utilizadas por las bases de datos, fundamentalmente de revistas científicas, que proporcionan un nivel de agregación de los dominios científicos excesivamente amplios y poco definidos se vería complementada por este tipo de análisis al permitir detectar un mayor grado de profundidad de los diferentes dominios científicos.

En cuanto a los programas de visualización utilizados, las imágenes ofrecidas por *Pajek* y *VOSviewer* coinciden en las líneas de investigación y en la identificación de los descriptores más influyentes de la red. Además se han complementado bien: *Pajek* en combinación con *PfNET* ha permitido sintetizar y resaltar la estructura relacional, mientras que *VOSviewer* ha hecho posible la identificación de los grupos y focos de investigación de una manera más plástica. Las metodologías de visualización empleadas en este estudio presentan visualizaciones estáticas e independientes de cada período.

Tanto en las visualizaciones del período completo como en la de los sub-períodos se distingue claramente la investigación básica y la clínica

De cara a estudios futuros sería interesante explorar las últimas funciones incorporadas a los dos programas: mapas superpuestos (*overlay maps*) que permiten fijar la posición de los nodos para hacer estudios evolutivos o comparativos (**Rafols; Porter; Leydesdorff, 2010; Leydesdorff; Rafols; Chen, 2013**), así como el uso de otros softwares de visualización como *Vison* utilizado en estudios similares (**Leydesdorff et al., 2008; Leydesdorff; Schank, 2008; Baur; Schank, 2008**). También sería interesante contrastar esta propuesta metodológica empleando otras unidades de análisis distintas a los KW+ con el objetivo de identificar el mejor sistema para definir las líneas de investigación presentes en cada dominio.

5. Notas

1. Un avance de los resultados de este trabajo se presentó en el *VI Encontro Ibérico Edicic 2013; Globalização, ciência, informação*, celebrado en Oporto (Portugal), 4-6 nov. 2013.

2. *KeyWords Plus (KW+)*
<http://interest.science.thomsonreuters.com/content/WOKUserTips-201010-SEA>
<http://wokinfor.com/media/pdf/qrc/wosqrc.pdf>

3. Los *Author keywords* son el otro tipo que utiliza *SCI* para indizar los documentos. Éstos se descartaron como unidades de análisis para el estudio, por su baja representatividad a partir de los resultados del estudio que se menciona.

6. Agradecimientos

Este trabajo ha sido posible gracias a la financiación pública obtenida mediante la beca pre doctoral del programa JAE-Predoc concedida por el Consejo Superior de Investigaciones Científicas (CSIC) y desarrollada en el Instituto de Políticas y Bienes Públicos (IPP).

7. Bibliografía

Batagelj, Vladimir; Mrvar, Andrej (2010). *Pajek 2.0: package for large network analysis*.

<http://vlado.fmf.uni-lj.si/pub/networks/pajek>

Baur, Michael; Schank, Thomas (2008). *Dynamic graph drawing in visone*. Karlsruhe: Technical University Karlsruhe

<http://i11www.itl.uni-karlsruhe.de/extra/publications/bs-dgdv-08.pdf>

Börner, Katy; Chen, Chaomei; Boyack, Kevin W. (2003). "Visualizing knowledge domains". *Annual review of information science and technology*, v. 37, n. 1, pp. 179–255.

<http://nwb.cns.iu.edu/papers/arist02.pdf>

<http://dx.doi.org/10.1002/aris.1440370106>

Boyack, Kevin W.; Börner, Katy; Klavans, Richard (2009). "Mapping the structure and evolution of chemistry research". *Scientometrics*, v. 79, n. 1, pp. 45-60.

<http://cns.iu.edu/images/pub/2007-boyack-mapchem.pdf>

<http://dx.doi.org/10.1007/s11192-009-0403-5>

Boyack, Kevin W.; Klavans, Richard (2010). "Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?". *Journal of the American Society for Information Science and Technology*, v. 61, n. 12, pp. 2389-2404.

<http://dx.doi.org/10.1002/asi.21419>

Braam, Robert R.; Moed, Henk F.; Van-Raan, Anthony F. J. (1991). "Mapping of science by combined co-citation and word analysis. II: Dynamical aspects". *Journal of the American Society for Information Science*, v. 42, n. 4, pp. 252-266.

<http://www.cwts.nl/TvR/documents/AvR-CoCit-Word-II.pdf>

[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4%3C252::AID-ASIZ3E3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1097-4571(199105)42:4%3C252::AID-ASIZ3E3.0.CO;2-G)

Callon, Michell; Courtail, Jean-Pierre; Turner, William A.; Bauin, Serge (1983). "From translations to problematic networks: an introduction to co-word analysis". *Social science information sur les sciences sociales*, v. 22, n. 2, pp. 191-235.

<http://dx.doi.org/10.1177/053901883022002003>

Cambrosio, Alberto; Limoges, Camille; Courtail, Jean-Pierre; Laville, Françoise (1993). "Historical scientometrics? Mapping over 70 years of biological safety research with cword analysis". *Scientometrics*, v. 27, n. 2, pp. 119-143.

<http://dx.doi.org/10.1007/BF02016546>

Cantos-Mateos, Gisela; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Zulueta, María-Ángeles (2012)

"Stem cell research: bibliometric analysis of main research areas through KeyWords plus". *Aslib proceedings*, v. 64, n. 6, pp. 561-590.

<http://digital.csic.es/handle/10261/62888?locale=es>

<http://dx.doi.org/10.1108/00012531211281698>

Cantos-Mateos, Gisela; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Zulueta, María-Ángeles (2013).

"Estudio comparativo sobre la visualización de redes de co-words a través de los descriptores del Science citation index y Medline". En: *I Congresso ISKO Espanha e Portugal / XI Congresso ISKO Espanha*, Oporto (Portugal), 7-9 nov. 2013, pp. 173-189.

<http://digital.csic.es/handle/10261/89290>

<http://dx.doi.org/10.1080/00387019808003304>

Chen, Chaomei; Ibekwe-SanJuan, Fidelia; Hou, Jianhua (2010). "The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis". *Journal of the American Society for Information Science and Technology*, v. 61, n. 7, pp. 1386-1409.

<http://arxiv.org/pdf/1002.1985.pdf>

<http://dx.doi.org/10.1002/asi.21309>

Chen, Dar-Zen; Huang, Mu-Hsuan; Hsieh, Hui-Chen; Lin, Chang-Pin (2011). "Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology". *Journal of informetrics*, v. 5, n. 3, pp. 400-412.

<http://dx.doi.org/10.1016/j.joi.2011.02.005>

De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Gómez-Crisóstomo, Rocío; González-Molina, Antonio; Muñoz-Fernández, Francisco-José; Vargas-Quesada, Benjamín (2007a).

Indicadores bibliométricos de la actividad científica española: 1990-2004. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt). ISBN: 84 690 5059 0

<http://eprints.rclis.org/16646>

De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Core-ra-Álvarez, Elena; Gómez-Crisóstomo, Rocío; Hassan-Montero, Yusef; Vargas-Quesada, Benjamín (2009).

Indicadores bibliométricos de la actividad científica española: 2007. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).

http://icono.fecyt.es/informesypublicaciones/Documents/indicadores_bibliometricos_2007.pdf

De-Moya-Anegón, Félix; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco José; Herrero-Solana, Víctor (2007b).

"Visualizing the marrow of science". *Journal of the American Society for Information Science and Technology*, v. 58, n. 14, pp. 2167-79.

<http://digital.csic.es/handle/10261/78646>

<http://dx.doi.org/10.1002/asi.20683>

De-Moya-Anegón, Félix; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco José; Herrero-Solana, Víctor (2006).

"Visualización y análisis de la estructura científica española: ISI Web of Science 1990-2005". *El profesional de la información*, v. 15, n. 4, pp. 258-269.

<http://eprints.rclis.org/8632>

Ding, Ying; Chowdhury, Gobinda G.; Foo, Schubert (2001). "Bibliometric cartography of information retrieval research by using co-word analysis". *Information processing & management*, v. 37, n. 6, pp. 67-78.

<http://core.kmi.open.ac.uk/display/9015701>

[http://dx.doi.org/10.1016/S0306-4573\(00\)00051-0](http://dx.doi.org/10.1016/S0306-4573(00)00051-0)

EuroStemCell, iCeMS, Elsevier (2013). "Stem cell research. Trends and perspectives on the evolving international landscape". http://www.eurostemcell.org/files/Stem-Cell-Report-Trends-and-Perspectives-on-the-Evolving-International-Landscape_Dec2013.pdf

Gábor, Csárdi (2006). "Dynamics of citation networks". En: *Artificial neural networks - Icnan 2006, 16th Intl conf*, Athens, Greece, Sept. 10-14, pp. 698-709.

Garfield, Eugene (1990). "Keywords Plus-ISI's breakthrough retrieval method. Part 1. Expanding your searching power on *Current contents* on diskette". *Current contents*, v. I, n. 32, pp. 5-9. <http://www.garfield.library.upenn.edu/essays/v13p295y1990.pdf> <http://dx.doi.org/10.1007/s00338-008-0431-2>

Garfield, Eugene; Sher, Irving H. (1993). "Keywords Plus™ – Algorithmic derivative indexing". *Journal of the American Society for Information Science*, v. 44, n. 5, pp. 298-299. [http://www.garfield.library.upenn.edu/papers/jasis44\(5\)p298y1993.html](http://www.garfield.library.upenn.edu/papers/jasis44(5)p298y1993.html) [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199306\)44:5%3C298::AID-ASI5%3E3.0.CO;2-A](http://dx.doi.org/10.1002/(SICI)1097-4571(199306)44:5%3C298::AID-ASI5%3E3.0.CO;2-A)

Glänzel, Wolfgang; Thijs, Bart (2011). "Using 'core documents' for the representation of clusters and topics". *Scientometrics*, July, v. 88, n. 1, pp. 297-309. <http://dx.doi.org/10.1007/s11192-011-0347-4>

Hammond, John (1949). "Recovery and culture of tubal mouse ova". *Nature*, v. 163, n. 4131, pp. 28-29. <http://dx.doi.org/10.1038/163028b0>

Kamada, Tomihisa; Kawai, Satoru (1989). "An algorithm for drawing general undirected graphs". *Information processing letters*, v. 31, n. 1, April, pp. 7-15. [http://dx.doi.org/10.1016/0020-0190\(89\)90102-6](http://dx.doi.org/10.1016/0020-0190(89)90102-6)

Kessler, Meyer-Mike (1963). "Bibliographic coupling between scientific papers". *American documentation*, v. 14, n. 1, pp. 10-25. <http://dx.doi.org/10.1002/asi.5090140103>

Lee, Woo-Hyoung (2008). "How to identify emerging research fields using scientometrics: an example in the field of information security". *Scientometrics*, v. 76, n. 3, pp. 503-525. <http://dx.doi.org/10.1007/s11192-007-1898-2>

Leydesdorff, Loet; Rafols, Ismael (2009). "A global map of science based on the *ISI* subject categories". *Journal of the American Society for Information Science and Technology*, v. 60, n. 2, pp. 348-362. <http://arxiv.org/pdf/0911.1057.pdf> <http://dx.doi.org/10.1002/asi.20967>

Leydesdorff, Loet; Rafols, Ismael; Chen, Chaomei (2013). "Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations". *Journal of the American Society for Information Science and Technology*, v. 64, n. 12, pp. 2573-2586. <http://arxiv.org/pdf/1301.1013.pdf> <http://dx.doi.org/10.1002/asi.22946>

Leydesdorff, Loet; Schank, Thomas (2008). "Dynamic animations of journal maps: indicators of structural change and interdisciplinary developments". *Journal of the American Society for Information Science and Technology*, v. 59, n. 11, pp. 1810-1818. <http://arxiv.org/pdf/0911.1437v1.pdf>

Leydesdorff, Loet; Schank, Thomas; Scharnhorst, Andrea; De Nooy, Wouter (2008). "Animating the development of social networks over time using a dynamic extension of multidimensional scaling". *El profesional de la información*, nov.-dic., v. 17, n. 6, pp. 611-626. <http://www.elprofesionalde lainformacion.com/contenidos/2008/noviembre/04.pdf> <http://dx.doi.org/10.3145/epi.2008.nov.04>

Leydesdorff, Loet; Welbers, Kasper (2011). "The semantic mapping of words and co-words in contexts". *Journal of informetrics*, July, v. 5, n. 3, pp. 469-475. <http://arxiv.org/pdf/1011.5209v2.pdf> <http://dx.doi.org/10.1016/j.joi.2011.01.008>

Rafols, Ismael; Porter, Alan L.; Leydesdorff, Loet (2010). "Science overlay maps: a new tool for research policy and library management". *Journal of the American Society for Information Science and Technology*, v. 61, n. 9, pp. 1871-1887. <http://www.leydesdorff.net/overlaytoolkit/overlaytoolkit.pdf> <http://dx.doi.org/10.1002/asi.21368>

Rip, Arie; Courtial, Jean-Pierre (1984). "Co-word maps of biotechnology: An example of cognitive scientometrics". *Scientometrics*, v. 6, n. 6, pp. 381-400. <http://doc.utwente.nl/57102/1/co-word.pdf> <http://dx.doi.org/10.1007/BF02025827>

Romo-Fernández, Luz M.; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix (2013). "Co-word based thematic analysis of renewable energy (1990-2010)". *Scientometrics*, v. 97, n. 3, pp. 743-765. <http://dx.doi.org/10.1007/s11192-013-1009-5>

Small, Henry; Upham, Phineas (2009). "Citation structure of an emerging research area on the verge of application". *Scientometrics*, v. 79, n. 2, pp. 365-375. <http://dx.doi.org/10.1007/s11192-009-0424-0>

Soos, Sandor; Kampis, George; Gulyás, László (2013). "Large-scale temporal analysis of computer and information science". *The European physical journal special topics*, v. 222, n. 6, pp. 1441-1465. <http://dx.doi.org/10.1140/epjst/e2013-01936-6>

Takahashi, Kazutoshi; Tanabe, Koji; Ohnuki, Mari; Narita, Megumi; Ichisaka, Tomoko; Tomoda, Kiichiro; Yamanaka, Shinya (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors". *Cell*, v. 131 n. 5, pp. 861-872. <http://images.cell.com/images/Edimages/Cell/IEPs/3661.pdf> <http://dx.doi.org/10.1016/j.cell.2007.11.019>

Takahashi, Kazutoshi; Yamanaka, Shinya (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors". *Cell*, v. 126, n. 4, pp. 663-676. [http://www.cell.com/cell/fulltext/S0092-8674\(06\)00976-7](http://www.cell.com/cell/fulltext/S0092-8674(06)00976-7)

<http://dx.doi.org/10.1016/j.cell.2006.07.024>

Van-Eck, Nees-Jan; Waltman, Ludo (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, v. 84, n. 2, pp. 523-38.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2883932>
<http://dx.doi.org/10.1007/s11192-009-0146-3>

Van-Raan, Anthony F. J.; Tijssen, Robert J. W. (1993). "The neural net of neural network research. An exercise in bibliometric mapping". *Scientometrics*, v. 26, n. 1, pp. 169-192.
<http://dx.doi.org/10.1007/BF02016799>

Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; González-Molina, Antonio; De-Moya-Anegón, Félix (2010). "Showing the essential science structure of a scientific domain and its evolution". *Information visualization*, v. 9, n. 4, pp. 288-300.
<http://www.ugr.es/~benjamin/showing.pdf>
<http://dx.doi.org/10.1057/ivs.2009.33>

Vargas-Quesada, Benjamín; De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Guerrero-Bote, Vicente (2008). "Development of the Spanish scientific landscape: ISI Web of science 1990-2005". *El profesional de la información*, v. 17, n. 1, pp. 22-37.
http://eprints.rclis.org/16683/1/evolucion_estructura_cientifica_espa%C3%B1ola.pdf
<http://dx.doi.org/10.3145/epi.2008.ene.03>

Wasserman, Stanley; Faust, Katherine (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press. ISBN: 97 805 1181 5478.
<http://dx.doi.org/10.1017/CBO9780511815478>

White, Howard D. (2003). "Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists". *Journal of the American Society for Information Science and Technology*, v. 54, n. 5, pp. 423-34.
<http://dx.doi.org/10.1002/asi.10228>

White, Howard D.; Griffith, Belder C. (1981). "Author co-citation: a literature measure of intellectual structure". *Journal of the American Society for Information Science*, mayo,

v. 32, n.3, pp. 163-171.

<http://dx.doi.org/10.1002/asi.4630320302>

White, Howard D.; McCain, Katherine W. (1998). "Visualizing a discipline: an author co-citation analysis of information science, 1972-1995". *Journal of the American Society for Information Science*, v. 49, n. 4, pp. 327-375.
<http://comminfo.rutgers.edu/~kantor/601/Readings2004/Week2/w2r2.PDF>
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(19980401\)49:4%3C327::AID-ASIA4%3E3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1097-4571(19980401)49:4%3C327::AID-ASIA4%3E3.0.CO;2-4)

Wilmut, Ian; Schnieke, Angelika E.; McWhir, Jim; Kind, Alexander J.; Campbell, Keith H. S. (1997). "Viable offspring derived from fetal and adult mammalian cells". *Nature*, v. 385, n. 6619, pp. 810-813.
<http://dx.doi.org/10.1038/385810a0>

Zhao, Dangzhi; Strotmann, Andreas (2011). "Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field". *Scientometrics*, v. 87, n. 1, pp. 15-131.
<http://dx.doi.org/10.1007/s11192-010-0317-2>

Zitt, Michel; Bassecouard, Elise (1996). "Reassessment of co-citation methods for science indicators: effect of methods improving recall rates". *Scientometrics*, v. 37, n. 2, pp. 223-244.
<http://dx.doi.org/10.1007/BF02093622>

Zitt, Michel; Lelu, Alain; Bassecouard, Elise (2011). "Hybrid citation-word representations in science mapping: Portolan charts or research fields?" *Journal of the American Society for Information Science and Technology*, v. 62, n. 1, pp. 19-39.
<http://dx.doi.org/10.1002/asi.21440>

Zulueta, María-Ángeles; Cantos-Mateos, Gisela; Sánchez, Carmen; Vargas-Quesada, Benjamín (2011). "Research involving women and health in the Medline database, 1965-2005. Co-term analysis and visualization of main lines of research". *Scientometrics*, v. 88, n. 3, pp. 679-706.
<http://dx.doi.org/10.1007/s11192-011-0455-1>

Máster Universitario en
Gestión de Información en las Organizaciones
Facultad de Comunicación y Documentación de la Universidad de Murcia
Grupo semipresencial en la Universidad Autónoma Metropolitana de México Unidad Xochimilco



Preinscripción 2014/2015:
• 1ª fase: 18 de febrero - 2 de mayo de 2014
• 2ª fase: 1 - 18 de julio de 2014
• 3ª fase: 15 - 26 de septiembre de 2014

Más información:
<http://www.um.es/web/comunicacion>

UNIVERSIDAD DE MURCIA



PUBLICACIÓN Nº 6

Cantos-Mateos, Gisela, Zulueta, María-Ángeles, Vargas-Quesada, Benjamín, Chinchilla-Rodríguez, Zaida (2013). Estudio comparativo sobre la visualización de redes de co-words a través de los descriptores del Science citation index y Medline. / *Congresso ISKO Espanha e Portugal / XI Congresso ISKO Espanha*, Oporto (Portugal), 7-9 noviembre 2013, pp. 173-189.

<http://dx.doi.org/10.1080/00387019808003304>

ESTUDIO COMPARATIVO SOBRE LA VISUALIZACIÓN DE REDES DE CO-WORDS A TRAVÉS DE LOS
DESCRIPTORES DEL SCIENCE CITATION INDEX Y DE MEDLINE

G. CANTOS MATEOS
Consejo Superior de Investigaciones Científicas - Instituto de Políticas y Bienes Públicos / SCImago
Research Group
gisela.cantos@csic.es

M. A. ZULUETA
Universidad de Alcalá / SCImago Research Group
ma.zulueta@uah.es

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

Z. CHINCHILLA-RODRÍGUEZ
Consejo Superior de Investigaciones Científicas - Instituto de Políticas y Bienes Públicos / SCImago
Research Group
zaida.chinchilla@csic.es

Resumen

Objetivos: El presente estudio se centra en la investigación desarrollada en España sobre células madre comprendida entre los años 1997 y 2010. El objetivo fundamental consiste en la comparación de las líneas de investigación que ofrece el análisis de distintos tipos de descriptores, según su naturaleza documental, a partir de su aparición conjunta en los documentos.

Material y Métodos: Las fuentes utilizadas han sido las bases de datos del Science Citation Index Expanded (SCI-E) y Medline, empleando para el estudio el mismo conjunto documental. El análisis aplicado ha consistido en la representación y visualización de las relaciones que se establecen entre los términos de indización. De un lado, se han empleado los descriptores utilizados por el SCI para indizar sus documentos: *KeywordsPlus* (KW+) y *Keywords Author* (KWA) y de otro, los descriptores *MeSH* utilizados por Medline. Las herramientas utilizadas para la visualización han sido el software Pajek, en combinación con el algoritmo PathfinderNetwork (PfNET) para la simplificación de las relaciones y el software VOSviewer.

Resultados y Discusión: se han recuperado 3.078 documentos. A partir de ellos, en función del tipo de descriptor seleccionado, se han obtenido distintas imágenes sobre la investigación española en células madre entre 1997 y 2010. La visualización más clara y completa es la que ofrecen los descriptores KW+, permitiendo detectar hasta un total de seis líneas de investigación. La visualización de los KWA, por su parte, ofrece una imagen más diluida de las líneas de investigación, reflejando, sobre todo, la investigación de carácter más básico. Finalmente, la representación de las relaciones de los descriptores MeSH también se aproxima, sobre todo, a los estudios de carácter más básico.

Conclusión: la comparación de las visualizaciones ha permitido determinar que los descriptores KW+ son la unidad de análisis más adecuada a la hora de realizar un análisis temático sobre un dominio científico.

Palabras-clave Análisis de Co-words. Células Madre. Visualización de la Información.

Abstract

Objectives: The present study is focused on the research conducted on stem cells in Spain between 1997 and 2010. Its main objective is the comparison of the main lines of research that provides the analysis of different types of descriptors through their co-occurrence in the documents.

Material and Methods: We used Science Citation Index Expanded (SCI-E) and Medline in order to study the same set of documents. Applied analysis involved the representation and visualization of relationships established between indexing terms of the study documents. On the one hand, we have employed the descriptors used by the SCI-E to index the documents: KeywordsPlus (KW+) and Keywords Author (KWA), on the other hand, MeSH descriptors used by Medline. The softwares handled for visualization are Pajek, in combination with the algorithm PathfinderNetwork (PfnET), and Vosviewer.

Results and Discussion: 3,078 documents were recovered. On the basis of them, depending on the type of descriptor selected, different visualizations were obtained from the Spanish stem cell research between 1997 and 2010. The better display is that offered by KW+ descriptors, allowing detection up to a total of six lines of research. The visualization of the KWA, meanwhile, is more diffuse, showing an overview of the more basic research. Finally, the representation of relations using MeSH descriptors also approaches especially the more basic studies.

Conclusion: The comparison of the displays has revealed that KW+ descriptors are the most appropriate units to analyse a scientific domain.

Keywords Co-words Analysis. Stem Cell. Information Visualization.

Introducción

La investigación española con células madre es un tema biomédico de gran importancia y repercusión científica por las expectativas que genera. Desde que en 1949, el científico J. Hammond (1949) descubriese el método para mantener los embriones de ratón en cultivo in vitro, la investigación con células madre se ha desarrollado gradualmente ofreciendo en los últimos años perspectivas muy esperanzadoras para el tratamiento de enfermedades, hasta ahora, incurables. Actualmente, la dirección de la investigación está orientada, fundamentalmente, al desarrollo de nuevas terapias para enfermedades hematológicas, cardiovasculares, neurodegenerativas, genéticas, cáncer, diabetes, etc. (Martínez Serrano y Bjorklund 1996; Bishop et al. 2002; Cao et al. 2002; Di Giorgio et al. 2007). Por su naturaleza, la investigación con células madre trasciende inevitablemente a otros campos de carácter tan diverso como el político, ético, cultural, jurídico, etc., situándose en el escenario de la controversia social.

Resulta cada vez más evidente, la utilidad que tienen los estudios bibliométricos para reflejar el estado en el que se encuentra la investigación en un momento determinado y facilitar la toma de decisiones en materia de política científica. Esto impulsa el desarrollo y perfeccionamiento de nuevas técnicas y herramientas bibliométricas que tratan de mejorar la comprensión y análisis de los dominios científicos. Una muestra es la incorporación de las redes sociales (Wasserman y Faust 1998) como técnica de análisis y visualización de dominios científicos (Boyack et al. 2009; Leydesdorff y Rafols 2009; Vargas-Quesada et al. 2010; Rafols et al. 2010).

Existe una gran diversidad de métodos y técnicas a la hora de delimitar y visualizar un campo científico con el fin de detectar subcampos, áreas o líneas de investigación. Hay estudios que utilizan la cocitación de autores (ACA) (White y Griffith, 1981; White y McCain, 1998; Small and Upham, 2009; Chen et al., 2010; Zhao y Strotmann, 2011), el análisis de emparejamiento bibliográfico (Kessler, 1963; Boyack y Klavans, 2010; Chen et al., 2011; Glänzel y Thijs, 2011), estudios de co-words (Callon et al., 1983; Rip y Courtial, 1984, Cambrosio et al., 1993; Van Raan y Tijssen, 1993; Ding et al., 2001; Lee, 2008; Leydesdorff y Welbers, 2011; Zulueta et al., 2011) o los estudios híbridos en los que se combinan tanto ACA como co-words (Zitt y Bassecoulard, 1996; Zitt et al., 2011). Sin embargo, otros autores han demostrado que los resultados del análisis estructural y la evaluación de los

resultados de investigación de un campo determinado pueden ser enriquecidos mediante la combinación de ambos elementos en un análisis integrado (Noyons et al, 1999; Cobo et al, 2011). En vista de las ventajas e inconvenientes que condicionan la elección de una metodología sobre otra, para el presente estudio se ha optado por el análisis de co-word. (Zitt et al., 2011). Según Michel Zitt y colaboradores (Zitt et al., 2011), las palabras, a diferencia de las citas, parecen adaptarse mejor a la "ciencia viva" porque son universales. Para estos autores, las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes. Además, el análisis de co-words es considerado como una técnica de análisis de contenido muy eficaz en el mapeo de la fuerza de asociación entre los ítems procedentes de datos textuales (Cobo et al., 2011).

En un estudio anterior (Cantos-Mateos et al., 2012), se analizó la producción científica sobre la investigación española con células madre desarrollada entre 1997 y 2007. En él, se analizaba mediante indicadores bibliométricos la producción científica relativa al dominio y se generó una representación topográfica de los descriptores *KeyWords Plus* más frecuentes. Este trabajo puede considerarse un estudio preliminar a este ya que se realizó una delimitación temática a través de una tipología concreta de descriptor.

En este caso, el objetivo fundamental de este trabajo consistirá en la comparación de las visualizaciones que ofrece el análisis de tres tipos distintos de descriptores, según su naturaleza documental, a partir de su aparición conjunta en los documentos, con el fin de detectar líneas o áreas de investigación en cada caso.

Existen otros trabajos en los que se han comparado descriptores como el de Quin (2000) desde la perspectiva del análisis de las similitudes semánticas entre los descriptores KW+ y MeSH. También hay muestras de estudios que han mapeado un dominio científico a través de los descriptores de los autores (Su y Lee, 2010). Así como trabajos en los que se ha realizado un análisis temático sobre la producción científica con células madre aplicando otro tipo de análisis como el de co-citación de autores (Zhao y Strotmann 2011).

1 Material y Métodos

1.1 Fuentes

Se han utilizado como fuentes de información las bases de datos bibliográficas del Science Citation Index Expanded (SCI-E) y Medline, ambas accesibles a través de la Web of Science.

Estas bases de datos se diferencian en muchos aspectos, uno de ellos es el tipo de indización que utilizan. El SCI-E indiza los documentos mediante el empleo de dos tipos de descriptores. Por un lado, los *Keywords de Autor* (KWA), que son las palabras clave que los propios autores utilizan para definir el contenido de sus documentos. Y por otro, los *KeyWords Plus* (KW+), que son las palabras clave extraídas de manera automatizada de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Garfield and Sher, 1993). Por su parte, Medline emplea una indización manual basada en el uso de los *Medical Subject Headings* (MeSH). Se trata de un vocabulario controlado especializado en términos médicos proporcionado por la National Library of Medicine (NLM).

Con el empleo de estas bases de datos se obtendrían los términos procedentes de tres sistemas de indización distintos. Uno de carácter automatizado a través de los KW+ del SCI y otros dos de carácter manual, uno de tipo libre mediante los KWA del SCI y otro de tipo controlado a través de los descriptores MeSH utilizados por Medline.

1.2 Recuperación y tratamiento de la información

En primer lugar, se inició la búsqueda documental en el SCI-E porque esta base de datos, a diferencia de Medline, normaliza el nombre de todos los países a través del campo *Address*. Esto es una característica fundamental para estudios como este que pretenden centrarse en un dominio geográfico determinado.

De este modo, la estrategia de búsqueda empleada para la recuperación de los documentos en el SCI fue utilizando los términos *stem* cell** en el campo *Topic*, *Spain* en el campo *Address* y limitada al periodo temporal de 1997-2010.

Una vez recuperados los documentos se extrajeron los KWA y los KW+ y se buscaron los mismos documentos en la base de datos Medline. Al tratarse de la misma colección documental es posible comparar con mayor precisión la imagen que ofrece la visualización de las relaciones que se establecen entre los documentos a través de los descriptores.

Finalmente, fueron extraídos los descriptores MeSH pasando por un proceso posterior de normalización. Se seleccionaron únicamente los términos *Major*. Éstos aluden únicamente a los conceptos más importantes de los documentos. Descartándose así, los términos *Minor* que, aunque identifican conceptos que también se han discutido en el documento, no aluden a la temática principal.

1.3 Análisis y visualización de los datos

El análisis aplicado a cada tipo de descriptor ha consistido en calcular las relaciones que se producen entre ellos a partir de su aparición conjunta en los documentos, lo que se conoce como *Co-words Analysis*. Este análisis consiste, en generar una matriz cuadrada de $N \times N$ elementos, donde N es el tipo de descriptor a representar, a partir de las veces que ocurren los descriptores en los documentos. El resultado es una matriz de co-ocurrencias que refleja el número de veces que un par de descriptores aparecen conjuntamente en dos documentos.

Para la visualización de dichas relaciones, se han utilizado dos software libres especializados en la generación de redes, que permiten representar la estructura relacional de cada tipo de descriptor. Por un lado, Pajek (Batagelj and Mrvar, 2010) y en concreto, el algoritmo de tipo *spring embedders* como es *Kamada-Kawai*, (Kamada y Kawai, 1989). Aunque Pajek permite representar grandes redes, se decidió simplificarlas con el fin de que las visualizaciones fueran las más claras posibles. Por un lado, se decidió representar únicamente los descriptores más frecuentes. Tras varios ensayos donde se probaron distintos umbrales de frecuencias para la generación de los mapas, se observó que los resultados más claros, en cuanto a criterios de visualización se refiere, eran aquellas redes compuestas entre 100 y 150 nodos. Además, para la reducción del espacio dimensional se aplicó el algoritmo de poda *Pathfinder* (Moya-Anegón et al., 2007a, b). Este algoritmo está basado en el principio de desigualdad del triángulo bajo los parámetros $r = \infty$ y $q = n - 1$. Éste ayuda a preservar y poner de relieve las relaciones más destacadas entre los descriptores (White, 2003; Vargas-Quesada et al., 2008; Moya-Anegón et al., 2009).

Por otro lado, también se utilizó el software WOSviewer v. 1.5.4 (Van Eck y Waltman, 2010). Es otro ejemplo reciente de software libre para la representación y análisis de la información. Se presenta como una alternativa a las técnicas tradicionales de representación multidimensional y visualización de redes. WOSviewer combina técnicas de visualización y clustering, lo que favorece el análisis. De los cuatro tipos de representaciones que ofrece, se optó por utilizar los mapas etiquetados con nombres, donde cada descriptor está representado por una etiqueta y cuyo tamaño es proporcional a su peso, junto con los denominados “mapas de calor” (o mapas de densidad). Estos últimos se caracterizan porque cada nodo en el mapa es representado con un color que va del rojo al azul, reflejando la densidad de las relaciones entre los descriptores. Cuanto mayor es la

densidad, es decir, la co-ocurrencia entre los descriptores, más se aproximarán a la tonalidad roja, en cambio si se aproximan más al color azul, esto indica una mayor dispersión y por lo tanto, menos co-ocurrencia.

Para la distribución espacial de la información, VOSviewer se puede considerar como una especie de escalamiento multidimensional ponderado, por lo que las posiciones de los descriptores, aunque no los frentes de investigación, pueden diferir de los detectados mediante Pajek y el algoritmo de layout con fines estéticos, como es Kamada-Kawai.

2 Resultados y discusión

El número de documentos recuperados en la base de datos del SCI-E ascendió a 4.148. De estos documentos, se localizaron y recuperaron en Medline 3.078. Los 1.070 documentos restantes no se recuperaron porque correspondían, en su mayor parte, a la tipología documental de *Meeting Abstract* o porque pertenecían a revistas que no están indexadas en Medline, y que están relacionadas, generalmente, con el ámbito de la Zoología y de la Botánica.

De los 3.078 documentos recuperados en ambas bases de datos (tabla 1), el 98,34% están indizados con KW+, el 99,22% con descriptores MeSH y tan, sólo el 65,82% de los documentos contienen KWA. Esta escasa representatividad de los KWA ha de tenerse en cuenta a la hora de interpretar los resultados, ya que, puede afectar a la visualización temática del dominio.

(N=3.078 docs.)					
SCI				Medline	
Nº docs. con KW+		Nº de docs. con KWA		Nº de docs. con MeSH	
Total	%	Total	%	Total	%
3.027	98,34	2.026	65,82	3.054	99,22

Tabla 1. Nº de documentos indizados según el tipo de vocabulario utilizado por las bases de datos

La representación en forma de mapa de las relaciones que se producen entre los descriptores más frecuentes según cada sistema de indización ofrece diferencias importantes entre las temáticas. Para su análisis se han generado un total de 9 mapas (tabla 2) a partir del uso de los dos software de visualización de redes indicados en la metodología.

	Software de visualización	
	Pajek	VOSviewer
Nº de desc. KW+	153 desc. más frec.	1.186 (Fuerza de Asociación = 30)
Nº de desc. KWA	103 desc. más frec.	492 (Fuerza de Asociación = 15)
Nº de desc. MeSH	151 desc. más frec.	760 (Fuerza de Asociación = 10)

Tabla 2. Descriptores representados

2.1 Descriptores KeyWords Plus (KW+)

En el caso de los KW+, es posible distinguir claramente la estructura temática del dominio a partir de los 153 términos más frecuentes (Fig. 1). Las dos zonas que se diferencian a primera vista en el mapa, se corresponden con las dos grandes tendencias de la investigación: la relacionada con la investigación clínica que estaría representada por los descriptores agrupados bajo *Bone-Marrow-Transplantation* y por otra parte, la investigación básica representada por el resto de agrupaciones. Ambas zonas quedan conectadas gracias a la presencia de dos KW+ que actúan a modo de puentes, como son *Tumor-Suppressor* y *Colony-Stimulating Factor*. Esta distinción resulta aún más clara en la Fig. 3, donde se distinguen claramente estos dos focos de la investigación.

En particular, los descriptores que conectan con *Bone-Marrow-Transplantation* (Fig. 1 y 3) están relacionados con los estudios ligados al uso terapéutico de las células madre en enfermedades hematológicas. Concretamente, muchos de los descriptores están relacionados con el tratamiento de los procesos oncológicos como la Leucemia (*Chronic Lymphocytic-Leukemia*, *Acute Myeloid-Leukemia*; *Leukemia*, etc.) a partir de trasplantes de progenitores hematopoyéticos (*Marrow-Transplantation*; *Stem-Cell Transplantation*).

En cuanto a la investigación básica representada por los KW+, se detectaron en los resultados hasta cuatro agrupaciones. En primer lugar, los descriptores vinculados con *Stem Cells* que están relacionados con los estudios encaminados a la investigación de células madre hematopoyéticas, es decir, las células madre procedentes de la médula ósea. Estos KW+ conectan claramente con la investigación clínica y con aquellos descriptores que forman parte de la siguiente línea de investigación, como es la relacionada con el estudio de las células madre embrionarias. En realidad, esta línea de investigación se encuentra desagregada en tres agrupaciones. Por un lado, en torno al descriptor *In-Vitro* se pueden localizar descriptores relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y sobre todo, analizar in-vitro las células madre embrionarias. Por otro lado, este grupo conecta con los descriptores relacionados con los procesos de diferenciación (*Differentiation*) y expresión celular (*Expression*). En el caso de la diferenciación se pueden localizar estudios relacionados con las células madre pluripotentes, aquellas que tienen la cualidad de poder diferenciarse en cualquier tipo de célula, ya sea, a partir de células madre embrionarias o inducidas. Estos estudios estarían muy relacionados con la Biología Molecular aplicada al interior de la célula encargada de regular la pluripotencia celular (*Pluripotency*; *Induction*) y con la diferenciación de las células madre pluripotentes en células madre tisulares (*Regeneration*; *Tissue*). Finalmente, la última agrupación de KW+ estaría relacionada con la investigación con células madre neurales (*Central-Nervous-System*). Estos descriptores están muy relacionados con los estudios centrados en el desarrollo e identificación de las células madre neuronales que se diferencian activamente en el tejido cerebral.

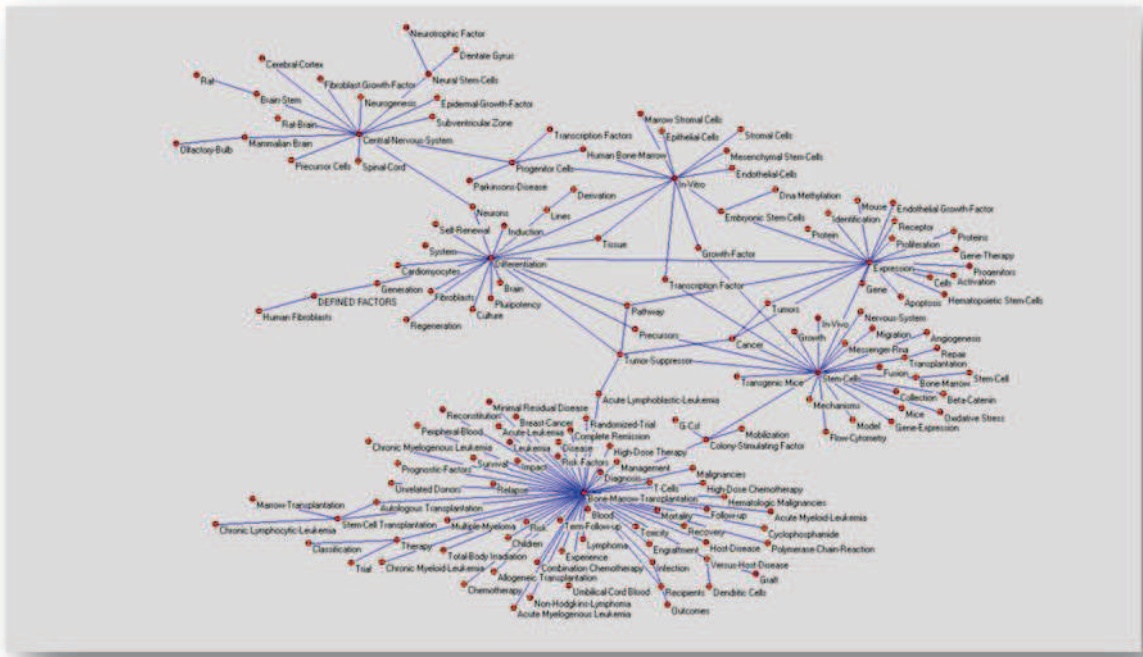


Fig. 1 Visualización con Pajek de las relaciones de co-ocurrencia de los 153 KW+ más frecuentes

VOSviewer (fig. 2), identifican el mismo número de agrupaciones (6) y además se puede apreciar el peso que tiene cada descriptor dentro de la red por la dimensión del nodo que lo representa.

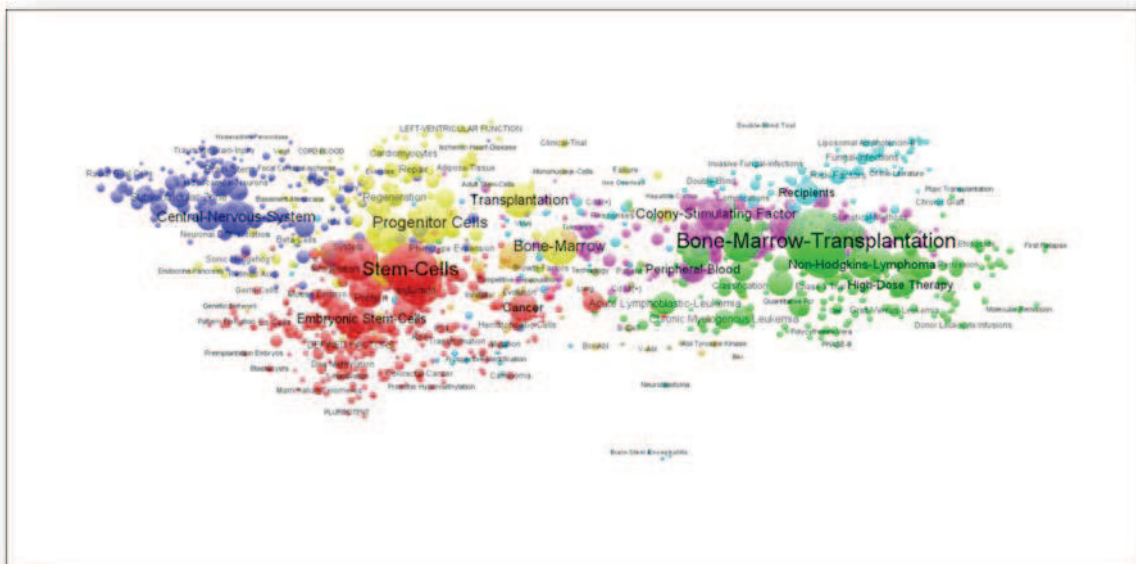


Fig. 2 Visualización con VOSviewer de 1.186 KW+ con fuerza de asociación = 30

En cuanto a las zonas con mayor densidad de la red (fig. 3), es posible distinguir claramente, dos zonas que concentran una importante actividad de relaciones. Una en torno al KW+ *Stem-Cells* y otra en torno al descriptor *Bone-Marrow-Trasplantation*. En menor medida, con una intensidad más moderada, también destacan los KW+ relacionados con *Central-Nervous-System*.

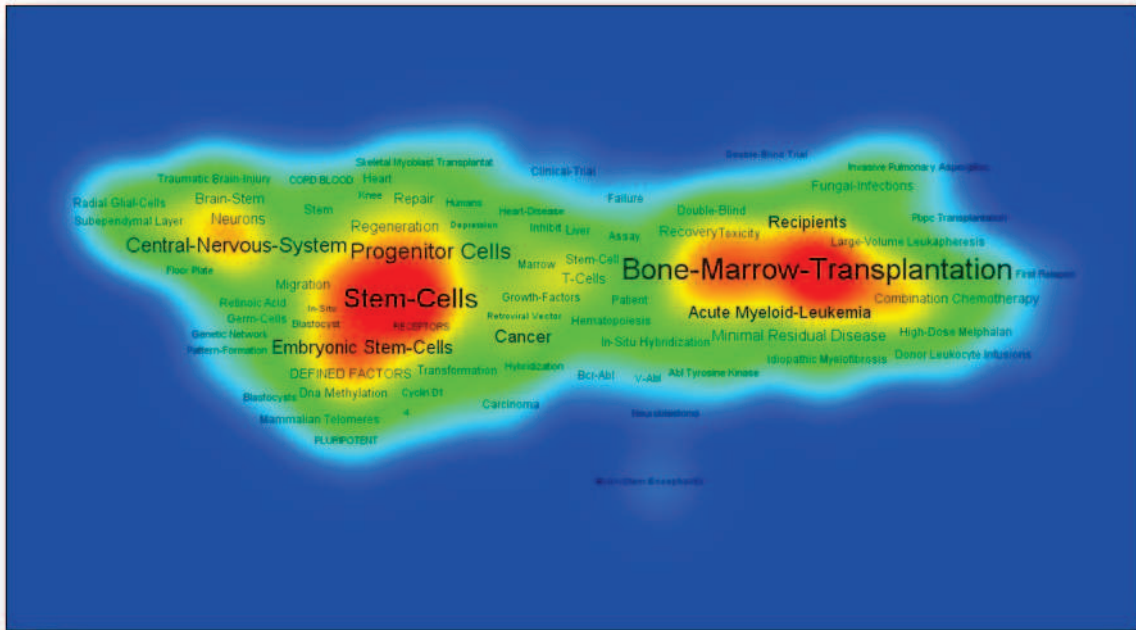


Fig. 3 Visualización con VOSviewer de la densidad de los 1.186 KW+ con fuerza de asociación = 30

2.2 Descriptores Key Words de Autor (KWA)

Tanto la fig. 4 como la fig. 5 ponen de manifiesto que no existe una delimitación temática clara y comprensiva del dominio que permita detectar líneas de investigación. Lo que se observa es que la presencia de los descriptores relacionados con la investigación de carácter clínico es mucho más reducida que en el caso de los KW+. Las visualizaciones de los KWA, por tanto, proporcionan una imagen de la investigación de carácter más básico y muy poco definida.

El resultado es una red en la que, a partir del nodo central *Stem Cell*, se desarrollan una serie de relaciones entretejidas donde es muy difícil identificar algún tipo de conexión temática. Por destacar algunas formaciones, aunque no se manifiesten de manera evidente, se podría señalar la localizada en la parte superior izquierda del mapa en torno al KWA *Neurogenesis*. También, en la parte superior derecha, se podría destacar la agrupación vinculada al descriptor *Multiple Myeloma* o la que se genera a partir del descriptor *Transplantation*, situada en la parte inferior izquierda del mapa.

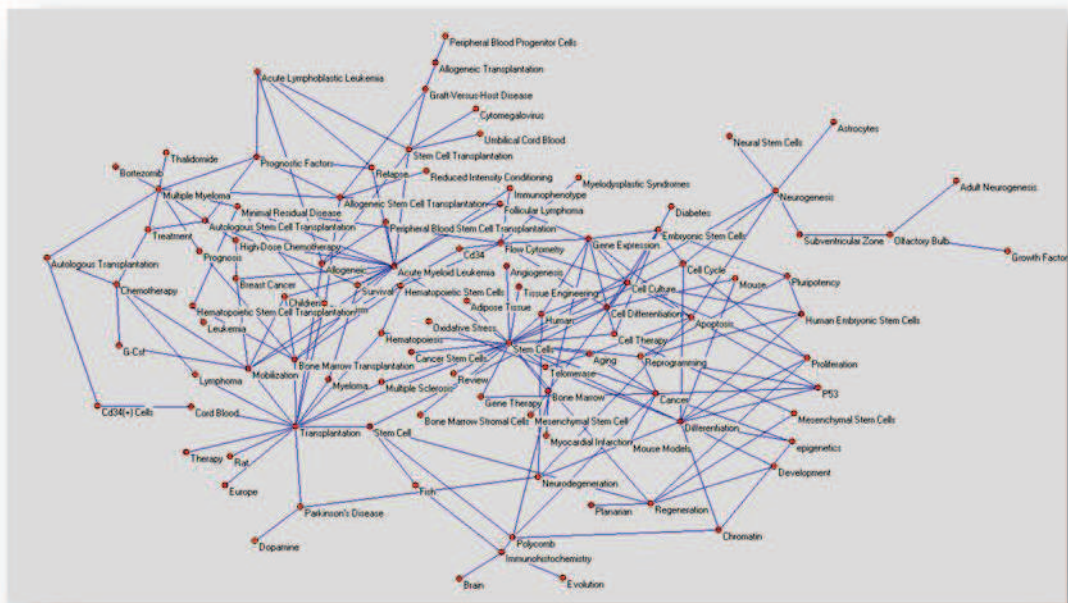


Fig. 4 Visualización con Pajek de las relaciones de co-ocurrencia de los 103 KWA más frecuentes

De igual modo, ocurre en la visualización que ofrece VOSviewer (fig. 5) que dibuja 15 agrupaciones, lo que significa que la identificación de temáticas se produce de una manera más diluida donde es más difícil identificar la vinculación de unos descriptores con otros.

Tan sólo se podrían destacar algunos clúster que parecen ser más predominantes y que coinciden a grandes rasgos, con los identificados por Pajek. Es el caso del clúster de color rojo que agrupa a los descriptores relacionados con *Multiple Myeloma*, el clúster de color azul oscuro donde destaca el KWA *Neurogenesis* y el clúster de color rosa al que pertenece el nodo más importante de la red: *Stem Cells*.

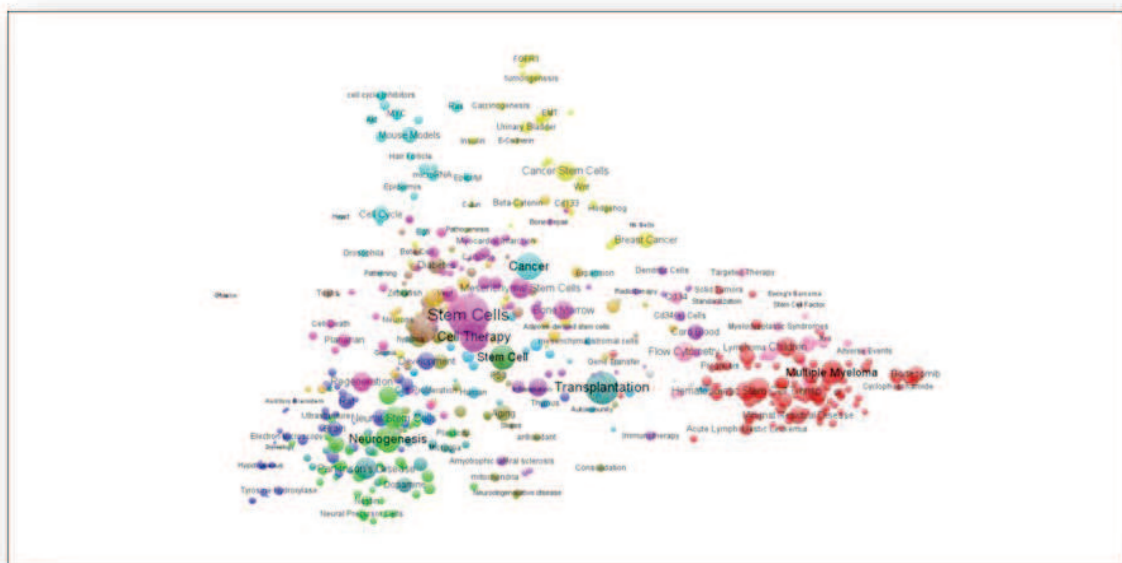


Fig. 5 Visualización con VOSviewer de 492 KWA con fuerza de asociación = 15

Así mismo, esta descripción de los clúster coincide con la imagen que ofrece la fig. 6. Las zonas con mayor densidad son las mismas que los clúster que mejor se identifican en la fig. 5.

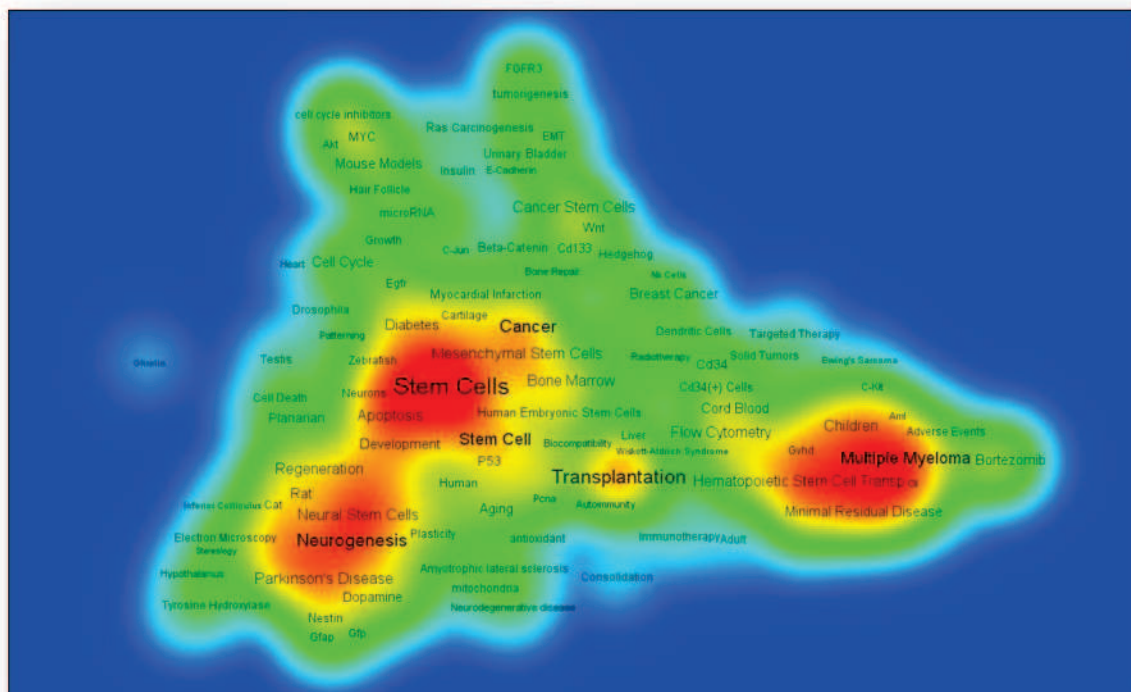


Fig. 6 Visualización con VOSviewer la densidad de los 492 KWA con fuerza de asociación = 15

2.3 Descriptores MeSH

La primera impresión al visualizar las relaciones entre descriptores MeSH, es una gran interrelación entre ellos. Las agrupaciones se conectan por la presencia de numerosos nodos-puente que dificultan considerablemente la identificación de la estructura general. Aun así, se pueden identificar 6 agrupaciones de términos que guardan una cierta vinculación temática. En la parte superior izquierda, se localiza la más amplia, cuyos elementos están vinculados al descriptor *Hematopoietic Stem Cell Transplantation* más representativa de la investigación clínica que de la básica. Otra formación más pequeña se sitúa en la parte superior derecha en torno al descriptor *Stem Cells*. En la zona central izquierda se ubica la investigación relacionada con las células madre hematopoyéticas a partir de un conjunto de descriptores relacionados con *Hematopoietic Stem Cells* (Figs. 7 y 8). En la parte inferior del mapa se localiza un grupo vinculado con el descriptor *Neurons*. Y un poco más a la derecha otro relacionado mediante el descriptor *Mesenchymal Stem Cells*. Finalmente, es posible detectar un pequeño grupo generado a partir de *Embryonic Stem Cells*.

Lo que sí se detecta claramente es una mejor representación de la investigación clínica que de la básica, esto se aprecia muy claramente en la fig. 9 donde el foco de la investigación básica aparece más disperso que el de la clínica.

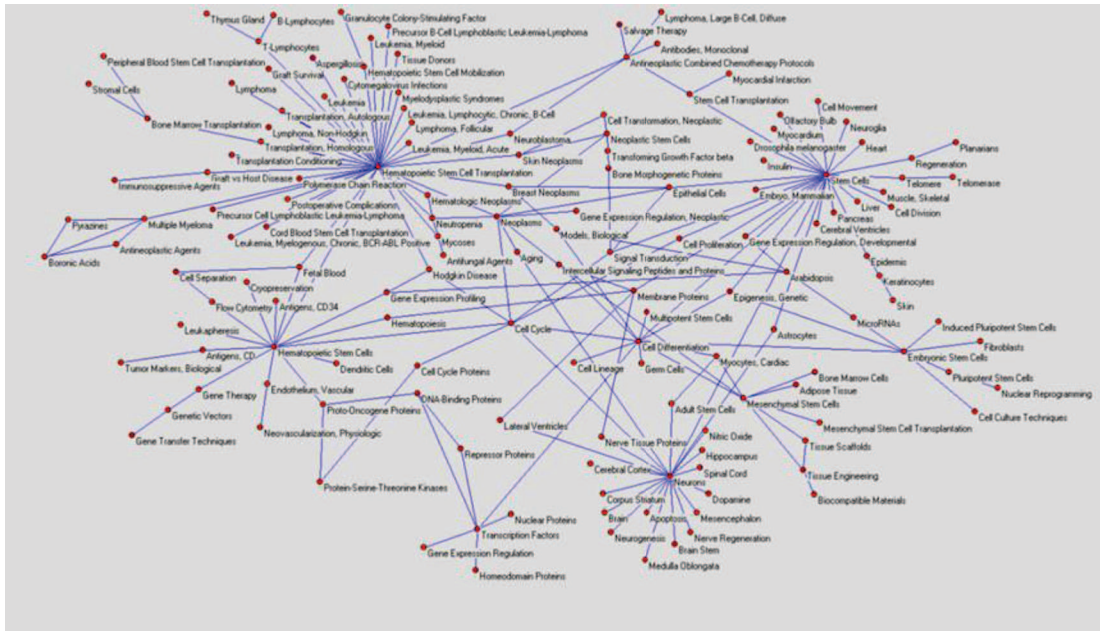


Fig. 7 Visualización con Pajek de las relaciones de co-ocurrencia de los 151 MeSH más frecuentes.

Todas esas agrupaciones pueden ser identificadas también en la fig. 8. A excepción de los descriptores *Haematopoietic Stem Cells* y *Embryonic Stem Cells* que no aparecen en el mapa. Esto se debe a los parámetros de visualización. A través de ellos se trata de evitar que se produzca un posible solapamiento de los nombres de los nodos cuando estos están muy próximos unos de otros.

A pesar de que se pueden identificar algunos clúster con claridad, el alto número de éstos (VOSviewer ha detectado 19 clúster), indica que hay muchas agrupaciones integradas con un escaso número de nodos que fraccionan excesivamente las relaciones temáticas del dominio.

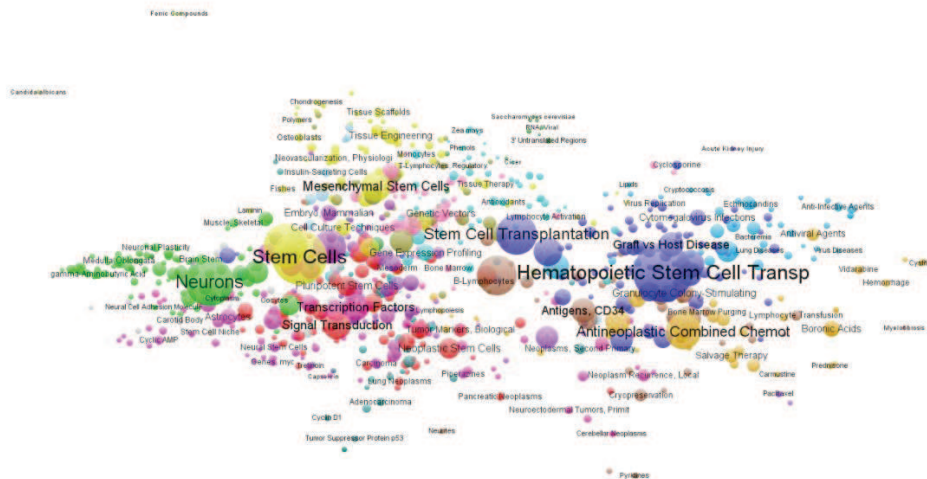


Fig. 8 Visualización con VOSviewer de 760 MeSH con fuerza de asociación = 10

En cuanto a localización de las zonas con mayor densidad (fig. 9), se aprecia que coinciden con los grupos identificados en los dos mapas anteriores. Además, esta representación indica que dichas zonas ocupan un gran tamaño y están muy próximas entre sí.

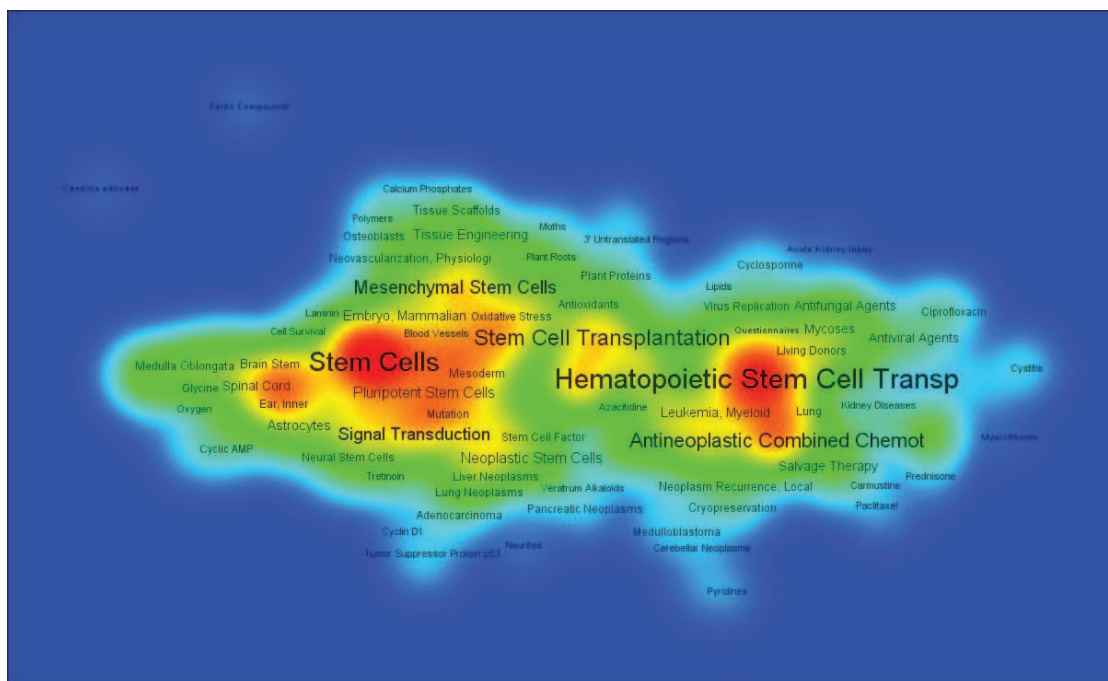


Fig. 9 Visualización con VOSviewer de la densidad de los 760 MeSH con fuerza de asociación = 10

Conclusiones

Los resultados de este estudio ofrecen una sucesión de datos e imágenes que, desde una perspectiva general, caracterizan temáticamente la investigación española con células madre entre 1997 y 2010. La elección de un periodo de 14 años, parece ser lo suficientemente amplio como para detectar aquellas líneas de investigación que tienen un carácter más consolidado. Esto facilita la comparación de los resultados en cuanto a la representación temática que ofrece cada tipo de descriptor. Por otra parte, la selección del mismo conjunto documental, ha hecho posible comparar las visualizaciones que se establecen entre los distintos sistemas de indización que utilizan las bases de datos SCI-E y Medline, cuyas coberturas temáticas son distintas. El estudio de diferentes tipos de descriptores, que sintetizan el contenido de un mismo conjunto documental, facilita el análisis comparativo del comportamiento que desarrolla cada descriptor y su capacidad de expresar los aspectos fundamentales que componen los principales temas de una investigación.

La visualización más clara y completa ha correspondido a los descriptores KW+, detectando hasta un total de seis líneas de investigación. Los KWA han ofrecido una imagen más diluida de las líneas de investigación, reflejando, sobre todo, la investigación de carácter más básico. Y la representación de las relaciones de los descriptores MeSH también se ha aproximado, sobre todo, a los estudios de carácter más básico. Luego podemos decir los KW+ son las unidades de análisis más adecuadas de las tres tipologías de descriptores, según la metodología aplicada, mostrando una estructura temática más definida donde las líneas de investigación se localizan claramente y expresan mayor información sobre el campo.

La explicación de estas diferencias temáticas puede verse afectada por la propia naturaleza documental de cada tipo de descriptor. Por definición, los KW+ tienen la ventaja de incorporar con mayor inmediatez los términos nuevos que van surgiendo en el desarrollo de las investigaciones de un determinado dominio científico. Al tratarse de un sistema de indización automatizado basado en la extracción de términos del lenguaje natural, la incorporación de términos que van configurando el discurso científico se incorporan de una manera más inmediata. Del mismo modo, ocurre con los KWA, ya que son los propios autores los que seleccionan los términos que consideran más precisos para definir su investigación. En cambio, este proceso de incorporación de términos nuevos es más

lento en el caso de los descriptores MeSH. Al tratarse de un vocabulario controlado, existe un margen de tiempo hasta que cada término se considera que forma parte del léxico científico y tiene que ser adaptado a los criterios conceptuales y documentales que rigen dicho vocabulario controlado.

Esta característica afecta a la representación de los descriptores, especialmente las visualizaciones obtenidas a través de Pajek. Estructuralmente la representación más clara es la de los descriptores KW+. En el caso de los MeSH, es posible que la excesiva interrelación de los grupos formados, a través de numerosos “descriptores puente” y el elevado número de clúster detectados, se deba a la rigidez de los descriptores y a la ausencia de términos nuevos que fomentarían la definición de los grupos y disminuiría la interrelación de los mismos. En cambio, en el caso de los KWA, la definición de los grupos es aún más compleja. Existe una excesiva interrelación entre los descriptores que impide detectar claramente las agrupaciones y la estructura temática del dominio. Quizá la baja representatividad de documentos indizados con KWA influya a la hora de visualizar el dominio.

En cuanto a los programas de visualización utilizados, el estudio ha mostrado que tanto las visualizaciones en Pajek como en VOSviewer, coinciden tanto en las líneas de investigación como en la localización de los descriptores más influyentes de la red. Pero además de coincidir, se complementan muy bien, ya que Pajek permite detectar de manera más clara la estructura relacional, mientras que VosViewer posibilita la detección de los grupos y los focos de investigación de una manera más plástica.

De cara a trabajos futuros, una vez detectadas las líneas de investigación más consolidadas en el periodo sería interesante detectar, mediante un análisis temporal evolutivo, las líneas de carácter temporal que reflejen los cambios en los focos de la investigación, así como, una imagen más dinámica del dominio.

Referencias

- BATAGELJ, V.; MRVAR, A. (2010). *Pajek 2.0: package for large network analysis*, disponible en: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> (accedido 19/01/2012).
- BOYACK, K.W., Börner, K.; KLAVANS, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, 79 (1), 45-60.
- BOYACK, K.W.; KLAVANS, R. (2010). Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? *Journal of the American Society for Information Science and Technology*, 61 (12), 2389-2404.
- BISHOP, A.E.; BUTTERY, L.D.K.; POLAK, J.M. (2002). Embryonic stem cells. *Journal of Pathology*, 197 (4), 424-429.
- CALLON, M.; COURTAIL, J.P.; TURNER, W.A.; BAUIN, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information Sur Les Sciences Sociales*, 22, 191-235.

- CAMBROSIO, A.; LIMOGES, C.; COURTIAL, J.P.; LAVILLE, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27 (2), 119-143.
- CANTOS-MATEOS, G.; VARGAS-QUESADA, B.; CHINCHILLA-RODRÍGUEZ, Z.; ZULUETA, M.A. (2012). Stem cell research: bibliometric analysis of main research areas through KeyWords Plus. *Aslib Proceedings*, 64 (6), 561-590.
- CAO, Q.L; BENTON, R.L.; WHITTEMORE, S.R. (2002). Stem cell repair of central nervous system injury, *Journal of Neuroscience Research*, 68 (5), 501-510.
- CHEN, C.; IBEKWE-SANJUAN, F.; HOU, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis, *Journal of the American Society for Information Science and Technology*. 61 (7), 1386-1409.
- CHEN, D.Z.; HUANG, M.H.; HSIEH, H.C.; LIN, C.P., (2011). Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Informetrics*, 5 (3), 400-412.
- COBO, M.J.; LÓPEZ-HERRERA, A.G.; HERRERA-VIEDMA, E.; HERRERA, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Infometrics*, 5 (1), 146-166.
- DI GIORGIO, F.P.; CARRASCO, M.A.; SIAO, M.C.; MANIATIS T.; EGGAN, K. (2007). Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. *Nature Neuroscience*, 10 (5), 608-614.
- DING, Chowdhury, G.; FOO, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, 37 (6), 67-78.
- GARFIELD, E. (1990). Keywords plus-ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette. *Current Contents*, 1, (32), 5-9.
- GARFIELD, E.; SHER, I.H. (1993). Keywords Plus™ – Algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44 (5), 298-299.
- GLÄNZEL, W.; THIJS, B. (2011). Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88 (1), 297-309.
- HAMMOND, J. (1949). Recovery and culture of tubal mouse ova. *Nature*, 163, (4131), 28-29.
- LEE, W.H. (2008). How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics*, 76 (3), 503-525.

- KAMADA, T.; KAWAI, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31 (1), 7-15.
- KESSLER, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14 (1), 10-25.
- LEYDESDORFF, L.; RAFOLS, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60 (2), 348-362.
- LEYDESDORFF, L.; WELBERS, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5 (3), 469-475.
- MARTÍNEZ SERRANO, A.; BJORKLUND, A. (1996). Protection of the neostriatum against excitotoxic damage by neurotrophin-producing, genetically modified neural stem cells, *Journal of Neuroscience*, 16 (15), 4604-4616.
- MOYA-ANEGÓN, F.; CHINCHILLA RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; GÓMEZ-CRISÓSTOMO, R.; HASSAN-MONTERO; VARGAS-QUESADA, B. (2009). *Indicadores Bibliométricos de la Actividad Científica Española: 2007*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).
- MOYA-ANEGÓN, F.; CHINCHILLA RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; GÓMEZ-CRISÓSTOMO, R.; GONZÁLEZ-MOLINA, A.; MUÑOZ-FERNÁNDEZ, F.; VARGAS-QUESADA, B. (2007a). *Indicadores bibliométricos de la actividad científica española: 1990-2004*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).
- MOYA-ANEGÓN, F.; VARGAS-QUESADA, B.; CHINCHILLA-RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; MUÑOZ-FERNÁNDEZ, F.J.; HERRERO-SOLANA, V. (2007b). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology*, 58 (14), 2167-79.
- NOYONS, E. C. M.; MOED, H. F.; LUWEL, M. (1999), Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50 (2), 115-131.
- QUIN, J. (2000). Semantic similarities between a keyword database and a controlled Vocabulary Database: an investigation in the antibiotic resistance Literature. *Journal of the American for Information Science (JASIS)*, 51 (3), 166-180.
- RAFOLS, I.; PORTER, A.; LEYDESDORFF, L. (2010). Science overlay maps: a new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61 (9), 1871-1887.
- RIP, A.; COURTIAL, P. (1984) Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6 (6), 381-400.

- SMALL, H.; UPHAM, S. P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79 (2), 365–375.
- SU, H.N.; LEE, P.C. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, 85 (1), 65-79.
- VAN ECK, N.J.; WALTMAN, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84 (2), 523-38.
- VAN RAAN, A.F.J.; TIJSEN, R.J.W. (1993). The neural net of neural network research. An exercise in bibliometric mapping. *Scientometrics*, 26 (1), 169-192.
- VARGAS-QUESADA, B.; MOYA-ANEGÓN, F.; CHINCHILLA-RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; GUERRERO-BOTE, V. (2008). Development of the Spanish scientific landscape: ISI web of science 1990-2005. *Profesional de la Información*, 17 (1), 22-37.
- VARGAS-QUESADA, B.; CHINCHILLA-RODRÍGUEZ, Z.; GONZÁLEZ-MOLINA, A.; MOYA-ANEGÓN, F. (2010). Showing the Essential Science Structure of a Scientific Domain and its Evolution. *Information Visualization*, 9, 288-300.
- WALTMAN, L.; VAN ECK, N.J.; NOYONS, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4 (4), 629-635.
- WASSERMAN, S.; FAUST, K. (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.
- WHITE, H.D.; GRIFFITH, B.C. (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163-171.
- WHITE, H.; MCCAIN, K. (1998). Visualizing a Discipline: an Author Co-Citation Analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science*, 49 (4), 327-375.
- WHITE, H.D. (2003). Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54 (5), 423-34.
- ZHAO, D.; STROTMANN, A. (2011). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, 87, 115-131.
- ZITT, M.; BASSECOULARD, E. (1996). Reassessment of co-citation methods for science indicators: Effect of methods improving recall rates. *Scientometrics*, 37 (2), 223–244.

ZITT, M.; LELU A.; BASSECOULARD, E. (2011). Hybrid citation-word representations in Science mapping: portolan charts or research fields? *Journal of the American society for information Science and Technology*, 62 (1), 19-39.

ZULUETA, M.A.; CANTOS-MATEOS, G.; SÁNCHEZ, C.; VARGAS-QUESADA, B. (2011). Research involving women and health in the Medline database, 1965-2005, Co-term analysis and visualization of main lines of research. *Scientometrics*, 88 (3), 679-706.

