



DEPARTAMENTO DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN

UNIVERSIDAD DE GRANADA

**“Análisis Estratégico de la Investigación
sobre Áreas Protegidas en España:
Ingeniería y Cartografía del Conocimiento”.**

TESIS DOCTORAL

José Pino Díaz

Granada. Abril de 2011.

Dirigida por el Dr. D. Evaristo Jiménez Contreras (1), la Dra.
Dña. Rosario Ruíz Baños (1) y el Dr. D. Rafael Bailón Moreno (2).
(1) Departamento de Biblioteconomía y Documentación y
(2) Departamento de Ingeniería Química.

A mis padres, que me inculcaron los valores del trabajo personal, de la responsabilidad y del estudio.

A mis hermanos, por su cariño, apoyo y confianza.

A mis hijos, Marina y Adrián.

A todas aquellas personas que gentilmente me prestaron ayuda y me brindaron su amistad.

Sea esta tesis un humilde legado.

Os quiero.

"Los seres humanos no nacen para siempre el día que sus madres los alumbran, sino que la vida les obliga otra vez y muchas veces a parirse a sí mismos"

El amor en los tiempos del cólera
(Gabriel García Márquez)

CONTENIDO

Índice de Figuras	11
Índice de Tablas	28
Índice de Ecuaciones	31
1. Resumen	33
2. Prefacio	36
3. Objeto de la tesis	38
4. Preliminares.....	40
4.1. Evaluación de la actividad científica y tecnológica	40
4.1.1. Índices de actividad	41
4.1.2. Indicadores relacionales	41
4.2. Gestión del conocimiento e inteligencia competitiva	42
4.3. Vigilancia estratégica, científica y tecnológica.....	44
4.4. Ingeniería y cartografía del conocimiento	44
4.5. Sistemas de conocimiento	47
4.6. Análisis de palabras asociadas (CO-WORD ANALYSIS).....	49
4.7. Teoría actor-red y Teoría de la traducción.....	51
4.8. Teoría de grafos	52
4.9. Trazado de grafos.....	54
4.9.1. Algoritmo de Kamada-Kawai	55
4.10. Espacio euclídeo	61
4.11. Espacio vectorial	62
4.11.1. Isobaricentro	65
4.12. Visualización de la información.....	66
4.13. Cartografía de la información	73

4.14.	Cartografía del conocimiento.....	77
4.15.	Áreas y espacios naturales protegidos.....	77
4.16.	Investigación en áreas y espacios naturales protegidos	78
4.16.1	Tesis doctorales sobre áreas protegidas	79
5.	Material y métodos	82
5.1.	Dominio documental.....	83
5.2.	Ámbito temporal	85
5.3.	Ámbito geográfico	86
5.4.	Copalred, Sistema de conocimiento.....	87
5.4.1.	Análisis PCAR.....	87
5.5.	Elección de la ocurrencia y de la coocurrencia mínimas	88
5.6.	Elección de los tamaños mínimo y máximo de las subredes	89
5.7.	Grupos o subredes estratégicas de investigación	91
5.7.1.	Algoritmos de agrupación	91
5.7.2.	Método gráfico de obtención de las subredes	93
5.8.	Visualización de las redes de investigación.....	94
5.8.1.	Centroide de la red tecnocientífica KK	96
5.8.2.	Centralidad nodal	97
5.9.	Centralidad y Densidad de las Subredes Estratégicas	103
5.9.1.	Cálculo de la centralidad de las Subredes	104
5.9.2.	Cálculo de la densidad de las Subredes	105
5.10.	Diagramas Estratégicos de las REDES tecnocientíficas.....	105
5.10.1.	Índice I-VECT de Importancia Estratégica.....	107
5.11	Parámetros VECT de las redes tecnocientíficas	107
5.12	Diagramas o mapas MDS	108
5.12.1.	Diagramas O mapas MDS de las redes tecnocientíficas.....	109

5.13.	Cartografía de las redes de investigación.....	114
6.	Resultados.....	116
6.1.	Base documental de la <i>Web of Science</i>	116
6.2.	Base de Conocimiento WoS_PCAR_53210 (1981-2005).....	119
6.2.1.	Autores.....	123
6.3.	Red Wos_PCAR_52310 (1981-2005).....	125
6.3.1.	Análisis estructural.....	139
6.3.2.	Análisis estratégico de la Red Wos_PCAR_53210.....	164
6.3.3.	Análisis dinámico de la Red WoS_PCAR_53210 (Copalred).....	227
6.3.4.	Parámetros VECT de la red WoS_PCAR_53210.....	239
6.3.5.	Cartografía del Conocimiento.....	244
6.3.6.	Mapa de Importancia Estratégica I-VECT de la Red WoS_PCAR_53210 (1981-2005).....	254
6.4.	Bases documentales deL IEDCYT (iCYT e ISOC).....	258
6.5.	Base de Conocimiento IEDCYT_PCAR_53210 (1981-2005).....	262
6.5.1.	Autores.....	269
6.5.2.	Revistas.....	270
6.6.	RED IEDCYT_PCAR_53210 (1981-2005).....	271
6.6.1.	Análisis estructural.....	298
6.6.2.	Análisis estratégico de la Red IEDCYT_PCAR_53210.....	328
6.6.3.	Análisis dinámico de la Red IEDCYT_PCAR_53210 (copalred)....	412
6.6.4.	Parámetros vect de la red IEDCYT_pcar_53210.....	425
6.6.5.	Cartografía del conocimiento.....	429
6.6.6.	Mapa de Importancia Estratégica I-VECT de la Red IEDCYT_PCAR_53210 (1981-2005).....	439
7.	Discusión de los resultados.....	443

8. Conclusiones.....	455
9. Bibliografía	480
10. Anexo: Mapas	493

ÍNDICE DE FIGURAS

Figura 1: Sistema de toma de decisiones en una organización inteligente. (Tomado de J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2001, http://freedownloadbooks.net/han-kamber-2001-ppt.html)	43
Figura 2: La Ingeniería y la Cartografía del Conocimiento en el Sistema de Toma de Decisiones de carácter estratégico de la organización. Realizada a partir de López Forniés, Fernandez Sora, Manchado Perez, Agustín Hernandez, & Zabala Diez, 2007.....	45
Figura 3: Fases del proceso Knowledge Discovery in Databases (KDD). (Tomado de José Hernández Orallo, http://www.dsic.upv.es/~jorallo/master/seminari.part.I.pdf)	45
Figura 4: Arquitectura de un sistema típico de minería de datos. Tomada de Gustavo Iván García, http://www.monografias.com/trabajos27/datamining/datamining.shtml#arquitectura	46
Figura 5: Clasificación de los algoritmos de agrupamiento en Minería de Datos Espaciales (MDE). (Tomado de Gustavo Iván García, http://www.monografias.com/trabajos27/datamining/datamining.shtml).....	47
Figura 6:: Esquema del Sistema Stanalist (Polanco, 2006). El módulo "Infometría" de Stanalist es un sistema de clasificación automática basado en dos programas alternativos, Neurodoc y Sdoc. Tomado de http://www.cincel.cl/documentos/Recursos/STANALYST2006.ppt	48
Figura 7: Cartografía de la red de actores de la controversia sobre la protección de las aves migratorias. Figuran las webs de los actores implicados y sus enlaces. Tomada de la web del <i>Centre de Sociologie de l'Innovation</i> , http://www.csi.ensmp.fr/ La Escuela Superior de Minas de París considera la importancia de la investigación y de la enseñanza de las controversias sociotecnológicas a sus alumnos.	52
Figura 8: Tomada de http://www.analytictech.com/networks/Teoriadegrafos.pdf	53
Figura 9: Tomada de http://es.wikipedia.org/wiki/Teoría_de_grafos	54
Figura 10: Ejemplo de funcionamiento de los algoritmos de los modelos dirigidos por fuerzas. Tomada de http://www-ma2.upc.edu/rsilveira/pubs/MScThesis.pdf	55
Figura 11: Aplicación del algoritmo KK al trazado de grafos.....	60

Figura 12: Diferentes interacciones del algoritmo KK un mismo grafo.....	61
Figura 13: Sistema de coordenadas cartesianas. http://es.wikipedia.org/wiki/Coordenadas_cartesianas	62
Figura 14: Utriusque cosmi maioris scilicet et minoris [...] historia, tomus II (1619), tractatus I, sectio I, liber X, De triplici animae in corpore visione. Fecha: 1619 (1619). Autor Robert Fludd. Tomada de http://upload.wikimedia.org/wikipedia/de/0/0c/RobertFuddBewusstsein17Jh.png	67
Figura 15: "El Árbol de la Ciencia". R. Llull, 1296, recurre a una analogía común en él: la comparación orgánica, en la que cada ciencia se representa como un árbol con raíces, tronco, ramas, hojas y frutos. Las raíces representan los principios básicos de cada ciencia; el tronco, la estructura; las ramas, los géneros; las hojas, las especies; y los frutos, los individuos, sus actos y sus finalidades. Los árboles llulianos expresan una clara sistematización del conocimiento que organiza y simplifica el estudio de las diversas disciplinas tratadas.Tomada de http://es.wikipedia.org/wiki/Ramon_Llull . Dendrograma. Tipo de representación gráfica o diagrama de datos en forma de árbol (dendro=árbol) que organiza los datos en subcategorías que se van dividiendo en otros hasta llegar al nivel de detalle deseado (asemejándose a las ramas de un árbol que se van dividiendo en otras sucesivamente). http://es.wikipedia.org/wiki/Dendrograma . Tomada de "Agrupamiento de los temas de investigación según las revistas de publicación" (J. Pino, 2005). ...	68
Figura 16: "Topic map" obtenido por medio de un programa denominado Star Tree Studio (basado en Java) en el que cada rama del conocimiento se subdivide en sus respectivas categorías. Proyecto Mapa del Conocimiento Humano, http://www.slideshare.net/AlphaCentauri77/mapa-del-conocimiento-humano	68
Figura 17: Los mapas conceptuales se estructuran en forma jerárquica en la que los conceptos más generales están en la raíz del árbol y a medida que vamos descendiendo por el mismo nos vamos encontrando con conceptos más específicos. http://www.infovis.net/printMag.php?num=141&lang=1 ...	69
Figura 18: "Ars Magna", R. Llull, 1290-95. Tomada de http://es.wikipedia.org/wiki/Ramon_Llull . Mapa mental de gestión de proyectos, 2008. http://www.wikimindmap.org . Un mapa mental es un diagrama usado para representar las palabras, ideas, tareas, u otros conceptos ligados y dispuestos radialmente alrededor de una palabra clave o de una idea central. Se utiliza para la generación, visualización, estructura y clasificación taxonómica de las ideas, y como ayuda interna para el estudio,	

organización, solución de problemas, toma de decisiones y escritura.
http://es.wikipedia.org/wiki/Mapa_mental69

Figura 19: Sociogramas. Reciben el nombre de sociogramas los grafos obtenidos en el análisis de redes sociales. Tomada de "Análisis de redes aplicado a la investigación forestal española" (Pino-Díaz J. , Análisis de redes sociales aplicado a la producción científica forestal española, ISI-1994/2003., 2005).....70

Figura 20: Cienciogramas. Reciben el nombre de cienciogramas los grafos obtenidos en Cienciometría. Fuente Atlas de la Ciencia del grupo de investigación Scimago. Tomada de <http://www.atlasofscience.net/pdf/atlas-of-science-quick-guide-v3.pdf> (Scimago, 2006)70

Figura 21: NewsMaps es uno de los mejores ejemplos de cartografía de la información disponible en la Web hoy en día. Los mapas NewsMaps son atractivos e interactivos, ofrecen un resumen del 'panorama' de grandes volúmenes de información textual, representado como colinas y valles, similar a los mapas topográficos del mundo real. Ofrecen mapas diarios de noticias internacionales y de EE.UU. NewsMaps ha sido desarrollado por Cartia, Inc. Tomada de http://www.mundi.net/maps/maps_015/ (NewsMaps: Topographic Mapping of Information)71

Figura 22: VxInsight es un sistema de conocimiento desarrollado por Sandia National Laboratories de Estados Unidos que emplea la metáfora del paisaje. En la imagen paisaje de la física nuclear. Tomada de <http://www.cs.sandia.gov/projects/VxInsight.html>71

Figura 23: John Old. Emplea su *Information Cartography* (Old J. , 2001), un Sistema de Información Geográfica para obtener Cartografía de la Información. Tomada de http://www.dcs.napier.ac.uk/~cs171/LJOld/InformationCartography/information_cartography.htm72

Figura 24: La red ilustra la interconexión de los países como resultado de la colaboración en los mismos proyectos IST y su nivel de actividad. Cuanto más grande es el círculo, el más alto es el nivel de actividad de un determinado país en los proyectos de IST. La conexión de dos líneas fuertes de los países señalan un enlace de investigación intensa. Código de color: los Estados miembros de la Unión Europea (rojo), los nuevos Estados miembros (naranja), otros países europeos (amarillo), el resto del mundo (verde). La misma información se transmitió a la ilustración en el lado derecho. Las montañas (rojo) representan un alto nivel de actividad en los proyectos IST. La relación de vecindad Vecindad indica una intensa investigación. Se ha utilizado un Sistema de Información Geográfica (SIG) para obtener la

impresión de paisaje. Fuente de datos: La Unión Europea, 5 ^o Programa Marco de proyectos IST. Tomada de (European Union. Fifth Framework Programme - IST Projects)	72
Figura 25: Diferentes tipos de mapas conceptuales. Tomada de http://scimaps.org/maps/browse/	73
Figura 26: Diferentes tipos de mapas de dominio. Tomada de http://scimaps.org/maps/browse/	74
Figura 27: Diferentes mapas cartográficos usados en trabajos científicos. Tomada de http://scimaps.org/maps/browse/	75
Figura 28: Esquema de un SIG. Imagen tomada de http://grafosistemas.com/deinteres.html	76
Figura 29: Datos de entrada y resultados que se obtienen en un SIG.	76
Figura 30: Interfaz del SIG ArcView 3.2 con los mapas en dos, 2D, y tres dimensiones, 3D, del nuevo conocimiento creado a partir del empleo del sistema de conocimiento CoPalRed.	77
Figura 31: Periodo de estudio de la tesis doctoral y situación respecto a eventos importantes acaecidos en las ANP de España.	86
Figura 32: Diagrama de flujo del Sistema de Conocimiento CoPalRed	87
Figura 33: Diagrama de frecuencias del nº de palabras clave por registro en la Base Documental IEDCYT. (Los datos pertenecen al periodo 1981-1989).....	89
Figura 34: Diagrama de frecuencias acumuladas del nº de palabras clave por registro en la Base Documental IEDCYT. (Los datos pertenecen al periodo 1981-1989).	89
Figura 35: Diagramas de frecuencias del nº de palabras clave por artículos en las bases ISOC e ICYT. Elaboración propia a partir de datos de Gil Leiva y Alonso-Arroyo (2005).....	90
Figura 36: Nº mínimo y máximo de descriptores obtenidos por Gil Leiva y Rodríguez Miño (1997) en la muestra documental de su trabajo.	90
Figura 37: Ejemplo que se encuentra en <i>Cognosfera</i> y que explica como actúa el algoritmo de enlaces simples para la formación de agrupaciones con un máximo de 5 palabras y la reconstrucción posterior de la red.	92
Figura 38: Proceso de agrupación sobre centros simples para el ejemplo de la Figura anterior. Se observa que el resultado es diferente ya que resulta un solo grupo con una estructura en estrella que será identificado mediante la	

palabra A. Ambos métodos pueden dar resultados algo diferentes a la hora de definir las subredes, pero al reconstruir la red global, el resultado es siempre el mismo.	93
Figura 39: Grafos de las Subredes del Componente Principal de la Red WoS obtenidas con Pajek suprimiendo los enlaces inferiores a 1819 y separando componentes.	94
Figura 40: Archivo de entrada a Pajek y grafo obtenido.	95
Figura 41: Coordenadas de los nodos o vértices de la Red.	96
Figura 42: Posición del centroide de la Red KK (solo se han mantenido visibles los nodos de la red)	97
Figura 43: Resultado de aplicar en tres diferentes interacciones el algoritmo KK a una misma red. Se comprueba como los dos grupos identificados mantienen la distancia con el centroide en cada interacción. El grupo naranja mejor relacionado con el resto de la red se mantiene cercano al centroide y el grupo rojo, con pocos enlaces al resto de la red, se mantiene alejado.	103
Figura 44: Cuadrantes del Diagrama Estratégico y caracterización de cada subred o tema según su pertenencia a un cuadrante.	106
Figura 45: MDS con matriz de índices de equivalencia de la Red_IEDCYT_PCAR_53210 (matriz de similitudes); la proximidad entre los nodos es medida de su índice de equivalencia (los nodos próximos son nodos con un alto valor de e).	113
Figura 46: MDS con matriz de índices de equivalencia de la Red_WoS_PCAR_53210 (matriz de similitudes); la proximidad entre los nodos es medida de su índice de equivalencia (los nodos próximos son nodos con un alto valor de e).	113
Figura 47: MDS con matriz de distancias geodésicas de la Red_IEDCYT_PCAR_53210. La proximidad entre los nodos es medida de su distancia geodésica (los nodos próximos son nodos con una distancia geodésica baja).	114
Figura 48: MDS con matriz de distancias geodésicas de la Red_WoS_PCAR_53210. La proximidad entre los nodos es medida de su distancia geodésica (los nodos próximos son nodos con una distancia geodésica baja).	114
Figura 49: Diagrama de flujo del procedimiento de construcción de mapas cartográficos de dominios documentales.	115

Figura 50: Diagrama de barras; "Producción anual de artículos científicos originarios de España sobre Áreas protegidas en la BD <i>Web of Science</i> ". ...	116
Figura 51: Diagrama de barras (frecuencias acumuladas); "Producción anual de artículos científicos originarios de España sobre Áreas protegidas en la BD <i>Web of Science</i> ".	117
Figura 52: Grafo de la Red WoS_PCAR_53210 (1981-2005). Se observan todos los nodos de valor de ocurrencia igual o mayor a cinco, enlazados o no.	140
Figura 53: Grafo de la Red WoS_53210 con sus dos componentes.	141
Figura 54: Subredes o grupos y centroide del componente principal.....	142
Figura 55: Grafos de Subredes de la Red ISI obtenidas con Copalred.	143
Figura 56: Grafos de las Subredes del Componente Principal de la Red WoS obtenidas con Pajek suprimiendo los enlaces inferiores a 1819 y separando componentes.	144
Figura 57: Grafos de la Subred <i>Delibes, M.</i>	145
Figura 58: Grafos de las Subredes <i>Doñana Nat. Park, Moreno, S., Age, Villafuerte, R. y Balbontín, J.</i>	146
Figura 59: Grafo de la Subred <i>Zunzunegui, M.</i>	147
Figura 60: Grafos de las Subredes <i>Aznalcóllar, y Contamination.</i>	148
Figura 61: Grafos de la Subredes <i>González, M.J. y Hernández, L.M.</i>	149
Figura 62: Grafos de la Subred <i>Soriguer, R.C.</i>	150
Figura 63: Grafos de las Subredes <i>Soils, Pb, Gómez-Ariza, J.L.</i>	151
Figura 64: Grafos de la Subred <i>Fire y Lloret, F.</i>	152
Figura 65: Grafos de la Subred <i>Lakes y Rojo, C.</i>	153
Figura 66: Grafos de la Subred <i>Communities, Sala, E. y Species richness.</i> ...	154
Figura 67: Grafos de la Subred <i>Ecology</i>	155
Figura 68: Grafos de la Subred <i>Communities.</i>	156
Figura 69: Grafos de la Subred <i>Arsenic, Hiraldo, F. y Baos, R.</i>	157
Figura 70: Grafos de las Subredes <i>Behavior y Braza, F.</i>	158

Figura 71: Grafos de las Subredes <i>Cryptogamie-Micologie(R)</i> y <i>Moreno,G.</i> .	159
Figura 72: Grafo del Componente Secundario de la Red <i>Capa,M.</i>	160
Figura 73: Enlaces exteriores entre las Subredes del componente principal de la Red ISI_PCAR_53210.	161
Figura 74: Área Estratégica de Investigación "Parque Nacional de Doñana" y Subredes que la forman.	162
Figura 75: Enlaces exteriores de las subredes del Área Estratégica de Investigación "Parque Nacional de Doñana".	163
Figura 76: Área estratégica de investigación "Álvarez Cobelas, M."	163
Figura 77: Área Estratégica de Investigación "Sala, E."	164
Figura 78: Situación de dos grupos, uno del cuadrante 1 y otro del cuadrante 3, en el Grafo y en el Diagrama Estratégico de la Red WoS_PCAR_53210. .	165
Figura 79: Situación de los grupos en el Grafo del Área de Investigación "Parque Nacional de Doñana" y en el Diagrama Estratégico de la Red ISI_PCAR_53210.	166
Figura 80: Diagrama Estratégico de la Red ISI_PCAR_53210 obtenido con Copalred.....	167
Figura 81: Diagrama Estratégico de la Red ISI_PCAR_53210 obtenido con el Método del Centroide.	170
Figura 82: Comparativa de los Diagramas Estratégicos de la Red WoS_PCAR_53210.	172
Figura 83: Áreas de Investigación de la Red WoS_PCAR_53210 y elementos singulares de investigación.	174
Figura 84: Análisis Estratégico del Área de Investigación Parque Nacional de Doñana.	176
Figura 85: Áreas Estratégicas de Investigación e Investigadores de las Subredes Estratégicas de Investigación de la Red ISI_PCAR_53210.	177
Figura 86: Áreas Estratégicas de Investigación e Investigadores de las Subredes Estratégicas de Investigación de la Red WoS_PCAR_53210.	179
Figura 87: Grafo de la Subred <i>Delibes, M.</i>	181
Figura 88: Grafo de la Subred <i>Aznalcóllar mine toxic spill.</i>	183

Figura 89: Grafo de la Subred González,, M. J.	185
Figura 90: Grafo de la Subred Soils.....	186
Figura 91: Grafo de la Subred Conservation.....	188
Figura 92: Grafo de la Subred Doñana National Park	190
Figura 93: Grafo de la Subred Communities.....	192
Figura 94: Grafo de la Subred Diversity.....	193
Figura 95: Grafo de la Subred Lakes.	198
Figura 96: Grafo de la Subred Ecology.	199
Figura 97: Grafo de la Subred Soriguer, R.C.	203
Figura 98: Grafo de la Subred Fire.....	204
Figura 99: Grafo de la Subred Behavior.....	206
Figura 100: Grafo de la Subred Zunzunegui.....	207
Figura 101: Grafo de la Subred Diets.	208
Figura 102: Grafo de la Subred Recruitment.	210
Figura 103: Grafo de la Subred Hiraldo, F.....	211
Figura 104: Grafo de la Subred Cryptogamie-Mycologie(R).....	212
Figura 105: Grafo de la Subred Moreno, S.	213
Figura 106: Grafo de la Subred Age.	215
Figura 107: Grafo de la Subred Science of the total environment (R).	218
Figura 108: Diagrama estraégico de la RED WoS_PCAR_53210 (Método del centroide).....	223
Figura 109: Serie periódica de las Subredes de la Red WoS_(ISI)_PCAR_53210 (Copalred).	227
Figura 110: Serie periódica de las Subredes (nodos) de la Red WoS_(ISI)_PCAR_53210 (Copalred).	228
Figura 111: Serie del nodo "Delibes, M." (Copalred).....	229
Figura 112: Serie del nodo "Populations" (Coalred)	230

Figura 113: Serie del nodo "Doñana National Park" (Coplared).....	231
Figura 114: Serie del nodo "SW Spain" (Coplared).....	232
Figura 115: Serie del nodo "Aznalcóllar Mine Toxic Spill" (Copalred).....	233
Figura 116: Serie del nodo "Soils" (Copalred).....	234
Figura 117: Serie del nodo "Patterns" (Copalred).....	235
Figura 118: Serie del nodo "Succession" (Copalred).....	236
Figura 119: Serie periódica de las Subredes de Investigadores de la Red WoS_A_53210 (Copalred).....	237
Figura 120: Serie periódica de las Subredes de Investigadores de la Red WoS_A_53210 (Copalred); se incluyen los investigadores que aparecen como nodos en las diferentes Subredes.....	238
Figura 121: Mapa cartográfico de la Red WoS_PCAR_53210 (Método del centroide). Se ha dibujado el Centroide de la Red y circunferencias de diferente distancia al centroide.....	246
Figura 122: Mapa cartográfico de Subredes Estratégicas de la Red WoS_PCAR_53210 (Método del centroide).....	247
Figura 123: Mapa cartográfico de Áreas Estratégicas de Investigación de la Red WoS_PCAR_53210. (Método del centroide).....	248
Figura 124: Correspondencia entre el Grafo de la Red de enlaces exteriores de las Subredes y el Mapa cartográfico de la Red. (Método del centroide) ..	249
Figura 125: Correspondencia entre el Mapa cartográfico y la imagen en 3D de la Red WoS_PCAR_53210 (Método del centroide).....	249
Figura 126: Imagen 3D de la Red WoS_PCAR_5320 donde se indican las diferentes Subredes (Método del centroide).....	250
Figura 127: Imagen 3D de la Red WoS_PCAR_5320 donde se indican los Investigadores de las diferentes Subredes (1) (Método del centroide).....	251
Figura 128: Imagen 3D de la Red WoS_PCAR_5320 donde se indican los Investigadores de las diferentes Subredes (2) (Método del centroide).....	252
Figura 129: Imagen 3D de la Red WoS_PCAR_5320 donde se indican los Investigadores del Área de Investigación "Parque Nacional de Doñana" (Método del centroide).....	253

Figura 130: Imagen 3D de la Red WoS_PCAR_5320 donde se indican las Palabras Clave del Área de Investigación "Parque Nacional de Doñana" (Método del centroide).	254
Figura 131: Tintas hipsométricas que indican elevación.	255
Figura 132: Simbología I-VECT.....	255
Figura 133: Mapa de Importancia Estratégica de la Red WoS_PCAR_53210 (1981-2005) con la simbología I-VECT (Método del centroide).	256
Figura 134: Mapa de Importancia Estratégica de la Red WoS_PCAR_53210 (1981-2005) con los nodos de la red y sus etiquetas. (Método del centroide)	257
Figura 135: Mapa de Importancia Estratégica de la Red WoS_PCAR_53210 (1981-2005) con los nodos de la red y sus etiquetas y las curvas de isovalor C-VECT (Método del centroide).....	258
Figura 136: Diagrama de barras; "Producción anual de artículos científicos originarios de España sobre Áreas protegidas en la BD IEDCYT (ICYT e ISOC).	259
Figura 137: Diagrama de barras (frecuencias acumuladas); "Producción anual de artículos científicos originarios de España sobre Áreas protegidas en la BD IEDCYT (ICYT e ISOC).....	260
Figura 138: Grafo de la Red IEDCYT_PCAR_53210 (1981-2005). Se observan todos los nodos de valor de ocurrencia igual o mayor a cinco, enlazados o no.	299
Figura 139: Grafo de la Red_IEDCYT_PCAR_53210 con sus seis componentes.	300
Figura 140: Proceso de obtención de las subredes en la Red_IEDCYT_PCAR_53210.....	301
Figura 141: Subredes o grupos y centroide del componente principal de la Red_IEDCYT_PCAR_53210.....	302
Figura 142: Grafos de Subredes de la Red_IEDCYT_PCAR_53210 obtenidas con el Sistema de conocimiento Copalred.	303
Figura 143: Grafos de las Subredes del Componente Principal de la Red IEDCYT_PCAR_53210 obtenidas con Pajek suprimiendo los enlaces inferiores a 1800 y separando componentes.	304
Figura 144: Grafos de las Subredes <i>Insecta-Coleoptera</i>	305

Figura 145: Grafos de las Subredes <i>Insecta-Lepidoptera</i>	306
Figura 146: Grafos de las Subredes <i>Insecta-Zabalegui, I.</i>	307
Figura 147: Grafos de la Subredes <i>Cárdenas Talaverón, A.M. y Coleoptera</i>	308
Figura 148: Grafos de las Subredes <i>Hymenoptera</i>	309
Figura 149: Grafos de la Subredes <i>Fungi y Boletín de la Sociedad Micológica de Madrid</i>	310
Figura 150: Grafos de las Subredes <i>Humedales y Rojo, C.</i>	311
Figura 151: Grafos de las Subredes <i>Humedales y Ecosistemas acuáticos</i> ...	312
Figura 152: Grafos de las Subredes <i>Pérez Latorre, A.V. y Cabezudo, B.</i>	313
Figura 153: Grafos de las Subredes <i>Cuadernos de Turismo y Araque Jiménez, E.</i>	314
Figura 154: Grafos de las Subredes <i>Cuadernos de Turismo y Actividades recreativas</i>	315
Figura 155: Grafos de las Subredes <i>Biogeografía y NOVOA, F.</i>	316
Figura 156: Grafos de las Subredes <i>Clemente Salas, L.</i>	317
Figura 157: Grafos de las Subredes <i>Flora y Valle Tendero, F.</i>	318
Figura 158: Grafos de las Subredes <i>Aguas subterráneas y Acuíferos</i>	319
Figura 159: Grafos de las Subredes <i>Revista de la Academia Canaria de Ciencias y Mollusca</i>	320
Figura 160: Grafos del Componente de la Red <i>Oxyura, Revista sobre las Zonas Húmedas</i>	321
Figura 161: Grafos del Componente de la Red <i>Clasificación de suelos</i>	322
Figura 162: Grafos del Componente de la Red <i>Aldezabal, A.</i>	323
Figura 163: Enlaces exteriores entre las Subredes del componente principal de la Red IEDCYT_PCAR_53210.	324
Figura 164: Enlaces exteriores entre las Subredes que forman el Área Estratégica de Investigación <i>Insectos</i>	326
Figura 165: Área Estratégica de Investigación "Turismos rural" y subredes que la forman.	327

Figura 166: Área Estratégica de Investigación "Ecosistemas acuáticos" y subredes que la forman.....	328
Figura 167: Diagrama Estratégico de la Red IEDCYT_PCAR_53210 obtenido con Copalred.....	329
Figura 168: Diagrama Estratégico de la Red IEDCYT_PCAR_53210 obtenido con el nuevo método.....	333
Figura 169: Comparativa de los dos Diagramas Estratégicos de la Red_IEDCYT_PCAR_53210.....	335
Figura 170: Áreas Estratégicas de Investigación del Componente Principal de la Red IEDCYT_PCAR_53210.	338
Figura 171: Área de Investigación "Insectos" de la Red IEDCYT_PCAR_53210.	339
Figura 172: Área Estratégica de Investigación "Turismo rural" de la Red IEDCYT_PCAR_53210.....	341
Figura 173: Área Estratégica de Investigación "Ecosistemas acuáticos" de la Red IEDCYT_PCAR_53210.	342
Figura 174: Área Estratégica de Investigación "Hongos" de la Red IEDCYT_PCAR_53210.....	343
Figura 175: Área Estratégica de Investigación "Gestión ambiental" de la Red IEDCYT_PCAR_53210.....	344
Figura 176: Área Estratégica de Investigación "Urbanismo" de la Red IEDCYT_PCAR_53210.....	345
Figura 177: Área Estratégica de Investigación "Valle, F." de la Red IEDCYT_PCAR_53210.....	346
Figura 178: Área Estratégica de Investigación "Moluscos" de la Red IEDCYT_PCAR_53210.....	347
Figura 179: Área Estratégica de Investigación "Cabezudo B." de la Red IEDCYT_PCAR_53210.....	348
Figura 180: Área Estratégica de Investigación "Musgos" de la Red IEDCYT_PCAR_53210.....	349
Figura 181: Área Estratégica de Investigación "Geomorfología" de la Red IEDCYT_PCAR_53210.....	350

Figura 182: Área Estratégica de Investigación "Acuíferos" de la Red IEDCYT_PCAR_53210.....	351
Figura 183: Área Estratégica de Investigación "Patrimonio Geológico" de la Red IEDCYT_PCAR_53210.	351
Figura 184: Área Estratégica de Investigación "Teledeteccion" de la Red IEDCYT_PCAR_53210.....	352
Figura 185: Área Estratégica de Investigación "contaminación de aguas" de la Red IEDCYT_PCAR_53210.	352
Figura 186: Área Estratégica de Investigación "Pastizales" de la Red IEDCYT_PCAR_53210.....	353
Figura 187: Área Estratégica de Investigación "Flora silvestre" de la Red IEDCYT_PCAR_53210.....	353
Figura 188: Grafo de la Subred Fungi (Copalred).....	356
Figura 189: Grafo de la Subred Insecta (Copalred).	357
Figura 190: Grafo de la Subred Cuadernos de Turismo (Copalred)	359
Figura 191: Grafo de la Subred Humedales (Copalred)	363
Figura 192: Grafo de la Subred Flora (Copalred).....	364
Figura 193: Grafo de la Subred Parques Naturales (Copalred).	366
Figura 194: Grafo de la Subred Impacto Ambiental (Copalred)	368
Figura 195: Grafo de la Subred Biogeografía (Copalred)	370
Figura 196: Grafo de la Subred Clemente Salas, L. (Copalred)	371
Figura 197: Grafo de la Subred Nueva Cita (Copalred).	372
Figura 198: Grafo de la Subred Aguas Subterráneas (Copalred).	374
Figura 199: Grafo de la Subred Oxyura-Revista sobre las zonas húmedas (Copalred)	376
Figura 200: Grafo de la Subred Sistemas de información geográfica (Copalred).	377
Figura 201: Grafo de la Subred Recursos naturales (Copalred).....	379
Figura 202: Grafo de la Subred Medio ambiente Canarias (Copalred)	382

Figura 203: Grafo de la Subred Patrimonio cultural (Copalred).	385
Figura 204: Grafo de la Subred Clasificación de suelos (Copalred).	386
Figura 205: Grafo de la Subred Hymenoptera (Copalred).....	387
Figura 206: Grafo de la Subred Patrimonio geologico (Copalred).	389
Figura 207: Grafo de la Subred Pérez Latorre, Andrés V. (Copalred).	392
Figura 208: Grafo de la Subred Cárdenas Talaverón, A.M. (Copalred).	393
Figura 209: Grafo de la Subred Revista de Pastos (Copalred).	395
Figura 210: Grafo de la Subred Revista de la Academia Canaria de Ciencias (Copalred).	397
Figura 211: Grafo de la Subred Aldezábal, Arantza (Copalred).	398
Figura 212: Grafo de la Subred Bañares Baudet, Ángel (Copalred).....	402
Figura 213: Grafo de la Subred Marismas (Copalred).	403
Figura 214: Grafo de la Subred Lichenes (Copalred).	404
Figura 215: Diagrama Estratégico de la Red IEDCYT_PCAR_53210.	407
Figura 216: Serie periódica de las Subredes de la Red_IEDCYT_PCAR_53210.	413
Figura 217: Serie del nodo "Conservación de la Naturaleza"	415
Figura 218: Serie del nodo "Parques Nacionales"	416
Figura 219: Serie del nodo "Espacios Naturales Protegidos"	417
Figura 220: Serie del nodo "Parques Naturales"	418
Figura 221: Serie del nodo "Reservas naturales"	419
Figura 222: Serie del nodo "Desarrollo sostenible"	420
Figura 223: Serie del nodo "Legislación"	421
Figura 224: Serie del nodo "Fauna".....	422
Figura 225: Serie del nodo "Fungi"	423
Figura 226: Serie periódica de subredes estratégicas de investigadores de la Red IEDCYT_A_33210	424

Figura 227: Serie periódica de las Subredes de Investigadores; se incluyen los investigadores que aparecen como nodos en las diferentes Subredes.	425
Figura 228: Palabras Clave de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210 y sus valores CVECT, DVECT y zVECT (1).....	426
Figura 229: Palabras Clave de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210 y sus valores CVECT, DVECT y zVECT (2).....	427
Figura 230: Investigadores de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210 y sus valores CVECT, DVECT y zVECT.....	428
Figura 231: Revistas de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210y su valores C-VECT, D-VECT y z-VECT.	429
Figura 232: Mapa cartográfico de la Red IEDCYT_PCAR_53210. Se ha dibujado el Centroides de la Red y circunferencias de diferente distancia al centroide.	431
Figura 233: Mapa cartográfico de Subredes Estratégicas de la Red IEDCYT_PCAR_53210.....	432
Figura 234: Mapa cartográfico de Áreas Estratégicas de la Red IEDCYT_PCAR_53210.....	433
Figura 235: Imagen zoom comparativa de de la zona central del mapa cartográfico y del grafo de la Red IEDCYT_PCAR_53210 con los enlaces exteriores entre subredes.	434
Figura 236: Imagen 3D de las Subredes Estratégicas de la Red IEDCYT_PCAR_53210 (1).....	435
Figura 237: Imagen 3D de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210 (2)	436
Figura 238: Imagen 3D de los Investigadores de las as Subredes Estratégicas de la Red IEDCYT_PCAR_53210.....	437
Figura 239: Imagen 3D de las Revistas de las as Subredes Estratégicas de la Red IEDCYT_PCAR_53210	438
Figura 240: Imagen 3D de los Centros de Investigación de las Subredes Estratégicas de la Red IEDCYT_PCAR_53210	439
Figura 241: Tintas hipsométricas que indican elevación.	440
Figura 242: Simbología I-VECT.....	440

Figura 243: Mapa de Importancia Estratégica de la Red IEDCYT_PCAR_53210 (1981-2005) con los símbolos I-VECT.	441
Figura 244: Mapa de Importancia Estratégica de la Red IEDCYT_PCAR_53210 (1981-2005) con los nodos de la red y sus etiquetas.....	441
Figura 245: Mapa de Importancia Estratégica de la Red IEDCYT_PCAR_53210 (1981-2005) con los nodos de la red y sus etiquetas y las curvas de isovalor C-VECT.	442
Figura 246: Diagrama estratégico de temas de investigación en el Campo de los Tensioactivos. Tomado de R. Bailón (2003).....	444
Figura 247: Top 100 de los investigadores en InformationScience (1988-1995). Tomada de (White & McCain, 1998).....	445
Figura 248: Agrupación de los temas estratégicos de investigación de la Ciencia Forestal española según autores comunes a los mismos. (Pino-Díaz J. , 2005).....	445
Figura 249: Cienciograma del Dominio Científico USA 2002. Tomado de Vargas-Quesada, De Moya Anegón, Chinchilla-Rodríguez y González-Molina (2006).	446
Figura 250: Grafo de los temas de investigación de los siete principales países sobre Conjuntos Difusos, subperiodo 2004-2008. Tomado de López-Herrera, Cobo, Herrera-Viedma, Herrera, Bailón-Moreno y Jiménez-Contreras (2009)	447
Figura 251: Mapa de curvas de cocitación de autores de Information Science. Tomado de L. John Old (2001)	448
Figura 252: Mapa cartográfico de teselas de "centralidad por autovalor" de la Red "Disciplinas de Ciencia y Tecnología" del dominio documental "Tesis doctorales sobre áreas protegidas leídas en España durante el periodo 1979-2004". Tomado de Pino-Díaz, José; Jiménez-Contreras, Evaristo; Ruíz-Baños, Rosario; Bailón-Moreno, Rafael (2009).	449
Figura 253: Mapa de Importancia Estratégica I-VECT de la Red WoS_PCAR_53210, 1981-2005.	452
Figura 254: Mapa de Importancia Estratégica I-VECT de la Red IEDCYT_PCAR_53210, 1981-2005.	453
Figura 255: Mapa de Importancia estratégica y Diagrama estratégico de la Red_WoS_PCAR_53210 (1981-2005).....	457

Figura 256: Mapa de Importancia estratégica y Diagrama estratégico de la Red_ISI_PCAR_53210 (1981-2005)	458
Figura 257: La Ingeniería y la Cartografía del Conocimiento en el proceso de Toma de Decisiones. Elaborado a partir de J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2001, http://freedownloadbooks.net/han-kamber-2001-ppt.html).....	460
Figura 258: Diagrama estratégico de la Red de investigación internacional ISI (Web of Science)	462
Figura 259: Áreas estratégicas de investigación del Componente principal de la Red WoS_ (ISI)_PCAR_53210, 1981-2005.	464
Figura 260: Producción anual de artículos conteniendo el término Aznalcóllar.	465
Figura 261: Producción de artículos conteniendo el término Delibes, M.	466
Figura 262: Diagrama estratégico de la Red de investigación nacional IEDCYT (ICYT+ISOC)	471
Figura 263: Áreas estratégicas de investigación del Componente principal de la red IEDCYT_PCAR_53210, 1981-2005.	473

ÍNDICE DE TABLAS

Tabla 1: Categorías UICN de las Áreas Naturales Protegidas.	78
Tabla 2: Élite de Revistas del Dominio documental "Investigación española en Áreas Protegidaso durante el periodo 1981-2005 indexada por la <i>Web of Science</i> "	117
Tabla 3: Élite de autores del Dominio documental "Investigación española en Áreas Protegidas durante el periodo 1981-2005 indexada por la <i>Web of Science</i> "	118
Tabla 4: Base de conocimiento PCAR de ocurrencia igual o mayor a cinco..	123
Tabla 5: Base de conocimiento de Autores de ocurrencia igual o mayor a cinco.....	124
Tabla 6: Base de conocimiento de las Revistas de ocurrencia igual o mayor a cinco.....	125
Tabla 7: Red WoS_PCAR_53210_(1981-2005)	139
Tabla 8: Parámetros generales de las subredes de la RED WoS_PCAR_53210 (Copalred).	168
Tabla 9: Parámetros generales de las Subredes (Método del centroide).	169
Tabla 10: Correspondencia entre las Subredes del Cuadrante nº 1 de los dos Diagramas Estratégicos de la Red WoS_PCAR_53210.	172
Tabla 11: Correspondencia entre las Subredes del Cuadrante nº 3 de los dos Diagramas Estratégicos de la Red WoS_PCAR_53210.	173
Tabla 12: Ranking de Importancia Estratégica de las Subredes de Investigación de la Red WoS-PCAR-53210 (Método del centroide).	224
Tabla 13: Ranking de Importancia Estratégica de las Palabras Clave de la Red WoS-PCAR-53210 (Método del centroide).	225
Tabla 14: Ranking de Importancia Estratégica de los Investigadores de la Red WoS-PCAR-53210 (Método del centroide).	226
Tabla 15: Ranking de Importancia Estratégica de las Revistas de la Red WoS-PCAR-53210 (método del centroide).	227
Tabla 16: Palabras Clave de las Subredes Estratégicas de la Red WoS_PCAR_53210 y sus valores C-VECT, D-VECT y z-VECT (Método del centroide) (1).	240

Tabla 17: Palabras Clave de las Subredes Estratégicas de la Red WoS_PCAR_53210 y sus valores C-VECT, D-VECT y z-VECT (Método del centroide) (2).....	241
Tabla 18: Investigadores de las Subredes Estratégicas de la Red WoS_PCAR_53210 y sus valores C-VECT, D-VECT y z-VECT (Método del centroide (!).	242
Tabla 19: Investigadores de las Subredes Estratégicas de la Red WoS_PCAR_53210 y sus valores C-VECT, D-VECT y z-VECT (Método del centroide (2).	243
Tabla 20: Revistas de las Subredes Estratégicas de la Red WoS_PCAR_53210 y sus valores C-VECT, D-VECT y z-VECT (Método del centroide).	244
Tabla 21: Élite de revistas del Dominio documental "Investigación española en Áreas Protegidas" indexadas en las BD ICYT e ISOC.....	260
Tabla 22: Élite de autores del Dominio documental "Investigación española en Áreas Protegidas" indexadas en las BD IEDCYT (ICYT e ISOC) (1981-2005).	262
Tabla 23: Ranking PCAR de ocurrencias igual o mayor a cinco. Base de conocimiento IEDCYT (ICYT e ISOC) (1981-2005).	269
Tabla 24: Ranking de Autores de ocurrencias igual o mayor a cinco de la Base de conocimiento IEDCYT (ICYT e ISOC) (1981-2005).	269
Tabla 25: Ranking de Revistas de ocurrencias igual o mayor a cinco de la Base de conocimiento IEDCYT (ICYT e ISOC).	271
Tabla 26: Red IEDCYT_PCAR_53210 (1981-2005).	298
Tabla 27: Parámetros generales de las Subredes de la Red_IEDCYT_PCAR_53210. (Copalred).....	330
Tabla 28: Parámetros generales de las Subredes de la Red_IEDCYT_PCAR_53210. (Método del centroide o método gráfico)	332
Tabla 29: Correspondencia entre las Subredes del Cuadrante nº 1 de los dos Diagramas Estratégicos de la Red_IEDCYT_PCAR_53210..	336
Tabla 30: Correspondencia entre las Subredes del Cuadrante nº 3 de los dos Diagramas Estratégicos de la Red_IEDCYT_PCAR_53210.	337
Tabla 31: Ranking de Importancia Estratégica de las Subredes de Investigación de la Red IEDCYT_PCAR_53210.	409

Tabla 32: Ranking de Importancia Estratégica de las Palabras Clave de la Red IEDCYT_PCAR_53210.....	410
Tabla 33: Ranking de Importancia Estratégica de los Investigadores de la Red IEDCYT_PCAR_53210.....	411
Tabla 34: Ranking de Importancia Estratégica de las Revistas de la Red IEDCYT_PCAR_53210.....	412
Tabla 35: Cuadro resumen del análisis estratégico de las redes.	459
Tabla 36: Periodos de mayor actividad de las Subredes Estratégicas de Investigación de la Red ISI_PCAR_53210 (1981-2005)	467
Tabla 37: Investigadores estratégicos de la Red WoS_PCAR_53210 1981-2005 (1).....	468
Tabla 38: Investigadores estratégicos de la Red WoS_PCAR_53210 1981-2005 (2).....	469
Tabla 39: Periodos de mayor actividad de las Subredes Estratégicas de Investigación de la Red IEDCYT_PCAR_53210 (1981-2005).	476
Tabla 40: Revistas de Importancia estratégica (I-VECT) en la investigación española de visibilidad nacional sobre áreas protegidas (1981-2005)	477
Tabla 41: Investigadores de Importancia estratégica (I-VECT) en la investigación española de visibilidad nacional sobre áreas protegidas (1981-2005).....	478

ÍNDICE DE ECUACIONES

Ecuación 1: Índice de Equivalencia entre dos palabras que coocurren.....	50
Ecuación 2: Fuerzas entre dos nodos conectados.	55
Ecuación 3: Energía potencial del enlace entre los nodos u y v.	56
Ecuación 4: Energía Total de la Red.	56
Ecuación 5: Minimización de la energía de la Red (I).	57
Ecuación 6: Minimización de la energía de la Red (II).	57
Ecuación 7: Cálculo del mínimo de energía (I).	57
Ecuación 8: Cálculo de la mínima energía local (II).	57
Ecuación 9: Método iterativo de Newton-Raphson.	58
<i>Ecuación 10: Gradiente de la energía del vértice v_m</i>	59
Ecuación 11: Coordenadas de un punto en el Plano Euclídeo.	61
Ecuación 12: Distancia euclídea.	61
Ecuación 13: Expresión de los vectores.	63
Ecuación 14: Vector nulo.	63
Ecuación 15: Vectores componentes cartesianos.	63
Ecuación 16: Suma de dos vectores.	64
Ecuación 17: Longitud o modulo de un vector.	64
Ecuación 18: Longitud de un vector en el plano euclídeo.....	64
Ecuación 19: Cálculo del baricentro.....	65
Ecuación 20: Cálculo de las coordenadas del Baricentro.....	66
Ecuación 21: Cálculo de las coordenadas del Isobaricentro.	66
Ecuación 22: Coordenadas del Centroide.	96
Ecuación 23: Centroide.....	98
Ecuación 24: Distancia euclídea.	99

Ecuación 25: Vector distancia nodal media de un nodo.....	99
Ecuación 26: Vector distancia nodal media del centroide.	99
Ecuación 27: Vector nodo-centroide.....	100
Ecuación 28: Módulo del vector distancia nodal media.....	101
Ecuación 29: Centralidad nodal de un nodo.....	101
Ecuación 30: Similitud de un nodo con el centroide.....	102
Ecuación 31: Centralidad nodal y distancia de un nodo al centroide.	103
Ecuación 32: Centralidad o Índice de cohesion externa de las subredes Copalred.....	104
Ecuación 33: Cálculo de la centralidad de las subredes del método gráfico.	104
Ecuación 34: Densidad o Índice de cohesion interna de las subredes Copalred.	105
Ecuación 35: Cálculo de la densidad de las subredes del método gráfico.....	105
Ecuación 36: Índice de importancia estratégica I-VECT.	107
Ecuación 37: Parámetro CTECT de los nodos.....	108
Ecuación 38: Parámetro DTECT de los nodos.	108
Ecuación 39: Parámetro ZTECT de los nodos.....	108
Ecuación 40: Índice de Importancia Estratégica I-VECT.....	222

1. RESUMEN

Las áreas protegidas o espacios naturales protegidos (parques nacionales, parques naturales, reservas naturales, etc.) son áreas naturales terrestres o marinas que destacan por su biodiversidad, su geodiversidad o sus paisajes prístinos bien conservados y poco humanizados. Desde finales del siglo XIX (Yellowstone, el parque más antiguo del mundo, fue creado en 1872; el Parque Nacional de Covadonga, creado en 1918, fue el primer espacio natural protegido español) las mejores muestras de la Naturaleza de la Tierra han sido objeto de protección y conservación por los Gobiernos de las Naciones. Así, la declaración legal de parques nacionales, de parques naturales, de reservas naturales o de otras distintas figuras de protección ha constituido desde entonces y hasta nuestros días uno de los ejes principales de las Políticas de Protección y de Conservación de la Naturaleza.

Las especies de flora o de fauna, las comunidades de seres vivos y los ecosistemas que se encuentran en las áreas protegidas han atraído históricamente a los investigadores de los campos científicos de Ciencias de la Vida y de Ciencias de la Tierra. Ya en etapas más recientes la incorporación de las Áreas Protegidas a los circuitos nacionales e internacionales de turismo de naturaleza, en lo que se ha venido en llamar ecoturismo o turismo verde, ha supuesto la realización de estudios e investigaciones encuadrables en las Ciencias Sociales. Por último, la declaración de áreas protegidas, su planificación y, cómo no, la gestión y administración de las mismas, ha centrado múltiples investigaciones tanto en Ciencias Jurídicas y de la Administración Pública como en Ciencias Tecnológicas y Ciencias Ambientales.

El acervo investigador sobre las áreas protegidas es, pues, diverso (monografías, tratados, manuales, artículos, tesis, etc.) y de amplia distribución académica, geográfica y temporal. Todo este conjunto documental, registrado e indexado en bases de datos bibliográficas, nacionales e internacionales, puede ser estudiado desde la perspectiva de la Infometría, la Ciencimetría, la Gestión del Conocimiento y/o la Inteligencia Competitiva. La presente tesis doctoral aborda el análisis estratégico del dominio documental "Investigación Española sobre Áreas Protegidas" desde el enfoque de la Vigilancia Estratégica Científica y Tecnológica (VECT), con la finalidad de servir de apoyo a la Toma de Decisiones en Política Científica sobre Áreas Protegidas.

La Ingeniería del Conocimiento y la Cartografía de la Información son disciplinas científicas recientes y en constante desarrollo, en las que convergen las Matemáticas, la Lingüística, la Informática y la Visualización de la información. Enfocadas a la creación y visualización de nuevo conocimiento a partir de la información contenida en las bases de datos documentales,

tienen un enorme y novedoso campo de aplicación en Gestión del Conocimiento, en Inteligencia Competitiva, en Evaluación de la Ciencia y la Técnica y en Política Científica y Tecnológica.

En esta tesis doctoral se realiza el "Análisis Estratégico de la Investigación Española sobre Áreas Protegidas" realizada en un periodo reciente de veinticinco años (1981-2005). Para ello se han estudiado dos grandes conjuntos documentales: uno internacional o de visibilidad internacional, los artículos científicos indexados en la *Web of Science*, y otro nacional o de visibilidad nacional, los artículos indexados en las bases de datos españolas del Instituto de Estudios Documentales sobre Ciencia y Tecnología, IEDCYT, (ICYT, ISOC e IME).

El empleo conjunto de técnicas KDD (*Knowledge Discovery Databases*), de análisis estratégico, de análisis de redes, de trazado de grafos y de cartografía, ha permitido crear y visualizar nuevo conocimiento, de carácter estratégico, sobre la Investigación Española en Áreas Protegidas. Los mapas obtenidos, denominados Mapas de Conocimiento Estratégico de las Redes Tecnocientíficas (Mapas CERT), son en sí mismos la imagen visual, uno de los productos de esta investigación; no obstante detrás de ellos existe, por un lado, un novedoso desarrollo de Ingeniería del Conocimiento (el "Método del Centroides de la Red Tecnocientífica y de la Centralidad Nodal") y, por otro lado, un "Método Gráfico de Visualización de las Subredes Estratégicas y de las Áreas Estratégicas de Investigación" mediante el empleo del Sistema de Conocimiento Copalred®, del software de análisis de redes Pajek® y del Sistema de Información Geográfica ArcView®.

El análisis VECT de las dos redes tecnocientíficas, la Red Internacional *Web of Science* y la Red Nacional IEDCYT, ha permitido conocer cuáles son sus subredes estratégicas y sus áreas estratégicas de investigación. Este nuevo conocimiento, de interés para la Toma de Decisiones en Política Científica y Tecnológica, ha sido descrito, diseccionado (en palabras clave, investigadores y revistas), evaluado y cartografiado. Para ello se han desarrollado los conceptos de "Centroides del Grafo de Mínima Energía de la Red" y de "Centralidad Nodal"; se ha implementado un nuevo método gráfico de análisis estructural y estratégico de las redes y de las subredes de investigación; se han obtenido nuevos parámetros de los términos y de las subredes: centralidad normalizada (CNECT), densidad normalizada (DNECT), e importancia absoluta (zNECT); y se han construido los diagramas estratégicos de las redes. Además se han elaborado los rankings de importancia estratégica (IVECT) de los términos (palabras clave, autores y revistas) y de las subredes de investigación y, para finalizar, se ha desarrollado una novedosa técnica de Cartografía del Conocimiento mediante "Mapas de Conocimiento Estratégico de las Redes Tecnocientíficas (Mapas CERT)":

mapas cartográficos 2D, imágenes 3D y "Mapas de Importancia Estratégica (Mapas IVECT)".

2. PREFACIO

Durante el largo periodo que va desde la aprobación del Plan de Trabajo de esta tesis, allá por el invierno de 2006, hasta el día de hoy, se han ido sucediendo pequeños hitos temporales y vitales que quisiera citar.

Iniciado el 2007 acabó mi larga etapa profesional en el Ministerio de Medio Ambiente como Director Adjunto de Parque Nacional, iniciada en 1995 en el Parque Nacional del Archipiélago de Cabrera (Islas Baleares) y continuada después en el Parque Nacional de Sierra Nevada (Granada-Almería) y comenzó otra en la Consejería de Medio Ambiente de la Junta de Andalucía.

A finales de 2007 se inició mi colaboración como Tutor de Energías Renovables y de Agricultura Ecológica en el Proyecto Aula Mentor del Ministerio de Educación.

En marzo de 2008 comencé a colaborar como Profesor Asociado en la Universidad de Málaga (nunca le estaré suficientemente agradecido a Emi, fue ella quien me animó a participar en el concurso). Desde entonces he tenido la fortuna de conocer y trabajar con gente estupenda, en especial Rocío y Carmen.

A finales de julio de 2008 nace Áreadoc, el blog académico del Área de Documentación de la Universidad de Málaga (al día de hoy ha recibido más de diez y ocho mil visitas provenientes de setentaiocho países distintos) .

Evaristo me invitó a formar parte del Grupo de Investigación Ec3 de la Universidad de Granada, acepté emocionado y me sentí muy honrado.

Presentamos una comunicación al IV (Zaragoza) y al V (Ávila) Congresos Forestales Españoles (una sobre producción científica forestal española y la otra sobre tesis doctorales de Áreas Protegidas).

Ingresé por promoción interna, mediante concurso-oposición, al Cuerpo Superior de Administradores Generales de la Junta de Andalucía.

A partir del Centro de Información y Documentación Científica (CINDOC) se crea el Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT).

El *Institute for Scientific information (ISI)* se integra en Thomson Reuters.

Presentamos una comunicación sobre la visualización de redes tecnocientíficas en el Seminario VSST2009 en Nancy (Francia) (reunión internacional sobre Vigilancia Estratégica, Científica y Tecnológica).

El Proyecto "Campus de Excelencia Internacional en Patrimonio Cultural y Natural" de las universidades andaluzas y coordinado por la Universidad de Jaén me seleccionó como miembro de su Red de Expertos en Patrimonio Cultural y Natural.

Presentamos la comunicación "Atlas de conocimiento estratégico de las redes tecnocientíficas de la investigación española sobre espacios naturales protegidos", al I Congreso Internacional "El Patrimonio Cultural y Natural como Motor de Desarrollo: Investigación e Innovación".

Enviamos tres artículos a las revistas de impacto JASIST, REDC y Scripta Nova. El artículo de REDC, "Evaluación de redes tecnocientíficas: la red española sobre Áreas Protegidas, según la Web of Science" ha sido aceptado para su publicación; el artículo de JASIST, "Strategic Knowledge Maps of the Techno-scientific Network (CERT Maps): A New Technology for Mapping Knowledge", ha sido revisado y está pendiente de reenvío y el artículo de Scripta Nova, "Las Ciencias Sociales en las redes tecno-científicas del IEDCYT sobre las Áreas Protegidas (1981-2005)", está pendiente de comunicación de la revista.

Durante todo este tiempo he trabajado en la tesis en cualquier momento y lugar: viviendas, hoteles, despachos, cafeterías, bibliotecas, terrazas de verano o vehículos. Para llevar adelante este trabajo he recibido facilidades, apoyo o ayuda de muchas personas, entre ellas: de Rocío, Carmen, Ma Antonia y Manolo Arjonilla del Departamento de Filología Griega, Estudios Árabes, Lingüística General y Documentación de la Universidad de Málaga; de los miembros del Grupo de Investigación de Evaluación de la Ciencia y de la Comunicación Científica (Ec3) de la Universidad de Granada; de los miembros del Departamento de Biblioteconomía de la Universidad de Granada y de amigos y compañeros del Parque Nacional de Sierra Nevada y de la empresa Tragsa. A todos mi más sincero agradecimiento.

Una mención especial se merecen los Tutores de la Tesis, Evaristo, Charo y Rafa; sin su dedicación, paciencia, consejos, recomendaciones o comentarios, habría sido imposible llegar a buen puerto.

Por último quisiera pedir perdón a mis hijos y a mi familia por los ratos de ausencia y de no dedicación (una y otra vez, y tantas veces más, con el portátil al hombro), deseo que el resultado haya merecido la pena.

3. OBJETO DE LA TESIS

La finalidad de esta Tesis Doctoral es analizar estratégicamente, describir y cartografiar la investigación sobre espacios naturales protegidos realizada durante un cuarto de siglo (1981/2005) en nuestro país, empleando técnicas propias de Ingeniería y Cartografía del Conocimiento, de Gestión del Conocimiento y de Inteligencia Competitiva y, especialmente, de Vigilancia Estratégica, Científica y Tecnológica.

El fin último es poner a disposición de los científicos, técnicos y gestores, un documento de carácter estratégico sobre la investigación científica y tecnológica española sobre áreas protegidas recogida en bases de datos multidisciplinares, nacionales e internacionales. Un documento estratégico dirigido a facilitar la Toma de Decisiones en las Administraciones, Organizaciones o Instituciones, por las personas responsables en Política Científica y Tecnológica o en Política de Espacios Naturales Protegidos, de la Nación o de las Comunidades Autónomas.

Al inicio de la tesis se plantearon los siguientes interrogantes:

- ¿Cuáles son las redes tecnocientíficas, nacional e internacional, de "Investigación sobre Áreas Protegidas en España" (Redes IAPE)?
- ¿Cuáles son las subredes y las áreas estratégicas de investigación de dichas redes?
- ¿Cuál es la importancia estratégica de cada una de las subredes IAPE?
- ¿Cuál es la posición estratégica de cada subred en la red tecnocientífica a la que pertenece?
- ¿Qué posición estratégica ocupan las palabras clave, los investigadores y las revistas?
- ¿Cuáles han sido las dinámicas de cambio de las subredes estratégicas y de los términos que las forman a lo largo del periodo estudiado?
- ¿Cuáles son los grafos que representan a las redes y las subredes estratégicas?
- ¿Cuál es la cartografía espacial más adecuada para visualizar el nuevo conocimiento estratégico obtenido del análisis KDD?
- ¿Cómo realizar mapas de conocimiento estratégico de las redes tecnocientíficas, síntesis y producto del análisis estratégico del dominio documental que representan?

Las respuestas a estas cuestiones principales y a otras secundarias que han ido surgiendo durante el desarrollo de la tesis, son las que se detallan y exponen en el presente trabajo.

4. PRELIMINARES

Para alcanzar los objetivos planteados en el punto anterior, se han empleado métodos y técnicas propios de diferentes disciplinas y subdisciplinas científicas: Evaluación de la Actividad Científica y Técnica, Gestión del Conocimiento e Inteligencia Competitiva, Vigilancia Estratégica, Científica y Tecnológica, Ingeniería del Conocimiento, Sistemas de Conocimiento, Análisis de Palabras Asociadas, Trazado de Grafos, Espacio Euclídeo, Espacio Vectorial, Visualización de la Información, Cartografía de la Información y Cartografía del Conocimiento .

Los lectores de esta tesis se van a encontrar un documento eminentemente visual ya que se han primado las imágenes, grafos y mapas de síntesis en la visualización del nuevo conocimiento creado.

Seguidamente, y sin ánimo de ser exhaustivo, se realiza una breve introducción de cada una de las materias y técnicas empleadas en la realización de este trabajo.

4.1. EVALUACIÓN DE LA ACTIVIDAD CIENTÍFICA Y TECNOLÓGICA

La evaluación de la actividad científica y tecnológica se realiza fundamentalmente sobre los documentos resultado de dicha actividad: artículos, informes, patentes, tesis doctorales, etc. Tiene su gran aplicación en la planificación, desarrollo y gestión de las políticas científicas regionales y nacionales.

Para la evaluación de la actividad científica y tecnológica se emplean métodos bibliométricos que analizan tanto los aspectos cuantitativos como los relacionales de la comunicación científica y técnica. Los objetivos de estos estudios son variados, desde conocer la producción de los investigadores, de los centros de investigación, de las revistas o de los países o regiones, hasta conocer los temas y grupos de investigación estratégicos de un Área de Conocimiento, de un Campo de la Ciencia y de la Técnica o de una región o país.

Actualmente se han impuesto las técnicas derivadas del estudio de las citas de los artículos científicos para evaluar la actividad de las revistas científicas y de los investigadores. Se construyen así, índices que clasifican a las revistas según su grado de penetración y aceptación en el mundo científico y que indirectamente otorgan un status diferenciador a los investigadores que publican en ellas; circunstancia comúnmente aceptada por la clase científica.

Ésto pone a disposición de las agencias evaluadoras de la actividad de los investigadores elementos que facilitan e impregnan su labor de la objetividad necesaria.

La evaluación de la comunicación científica y tecnológica se realiza siguiendo métodos que admiten diferentes clasificaciones; en este trabajo se ha empleado la clasificación que Jurado-Alameda, Bailón-Moreno y Ruíz-Baños (2002) emplean en su trabajo sobre el análisis de palabras asociadas aplicado a la evaluación de la actividad científica y técnica. Existen dos grandes conjuntos de índices: los Índices de Actividad y los Índices Relacionales; a su vez estos últimos se dividen en Índices Relacionales de Primera Generación e Índices Relacionales de Segunda Generación.

4.1.1. ÍNDICES DE ACTIVIDAD

Los **Índices de actividad** están basados en las leyes principales de la Bibliometría (Lotka, Bradford, etc.). En este grupo se encuentran los métodos utilizados para conocer la producción de los investigadores, de las revistas científicas, de los Centros de investigación, de las Áreas de Conocimiento y de las regiones o países.

A este grupo también pertenecen los métodos que analizan las referencias bibliográficas de los documentos. El índice más conocido de este grupo es el **Factor de Impacto** de las revistas científicas. Este índice clasifica y ordena las revistas según el resultado de ponderar el nº de citas recibido por el nº de artículos publicados. Se obtienen así listados de revistas científicas, ordenadas según su mayor o menor Factor de impacto, que son empleadas por las Agencias Nacionales o Regionales de Evaluación, en su labor de valoración de la actividad de los investigadores.

Las citas recibidas por los artículos científicos permiten, así mismo, realizar rankings de los diferentes investigadores de un área de conocimiento determinada.

4.1.2. INDICADORES RELACIONALES

En estos métodos se analiza la presencia conjunta de palabras, investigadores o revistas en los elementos de los artículos científicos: título, resumen, palabras clave, cuerpo del artículo o en la bibliografía. Estas relaciones de coocurrencia de palabras en el mismo artículo permiten construir y estudiar

redes de investigación, formadas por palabras clave, investigadores, revistas, disciplinas, centros de investigación o países.

Del estudio de las relaciones evidentes y explícitas se ocupan los **índices relacionales** llamados **de primera generación**. Un ejemplo de éstos es la **cocitación**, que se ocupa del análisis de los emparejamientos producidos entre investigadores o revistas al ser citados conjuntamente en la bibliografía de los artículos. En este caso dos descriptores están relacionados si aparecen citados en un documento. Las relaciones de cocitación son utilizadas para realizar las **Redes de Cocitación**.

Hay relaciones que solo son perceptibles tras el análisis de los conjuntos documentales mediante software apropiado basado en métodos ideados para tal fin, son los denominados **índices relacionales de segunda generación**. Es el caso del método de **análisis de las palabras asociadas**. En este método se analizan las relaciones de coocurrencia producidas entre los descriptores en el contenido del documento: título, autoría, palabras clave, resumen o cuerpo. Las relaciones de coocurrencia entre los descriptores son procesadas para realizar las **Redes de Coocurrencia** y a partir de éstas obtener los grupos o subredes más importantes.

4.2. GESTIÓN DEL CONOCIMIENTO E INTELIGENCIA COMPETITIVA

El conocimiento se puede clasificar según Nonaka y Takeuchi (1995) en **conocimiento explícito**, también llamado conocimiento objetivo (Popper, 1979), y **conocimiento tácito**. El conocimiento existente en las bases de datos bibliográficas es conocimiento explícito u objetivo (Reyes Meleán, 2005).

Hoy en día las organizaciones e instituciones de todo tipo disponen de una enorme cantidad de información en sus bases de datos. El adecuado uso que una organización da a esa información en aras a extraer nuevo conocimiento de interés, supone un nuevo factor de competitividad y de diferenciación entre las organizaciones modernas.

La gestión del conocimiento en una organización, entendida en sentido amplio, comprende todas aquellas actividades dedicadas a la captura, administración, almacenamiento, transferencia y difusión del conocimiento interno de la organización o del conocimiento externo de su interés. Comprende además la **generación de nuevo conocimiento** a partir del existente dentro de la propia organización o **conocimiento interno** (lo que se ha dado en llamar, en sentido estricto, **Gestión del Conocimiento**), o

fuera de ella dentro del ámbito de actuación o entorno de interés; en este caso se habla de **Inteligencia Competitiva**.

Las actividades y tareas propias de la Inteligencia Competitiva son aquellas que van dirigidas a tener un buen conocimiento de los factores que intervienen en la competitividad de las empresas o de las organizaciones. Prieto Viñuela (2004) citando a Porter (1990) distingue, según lo anterior, entre:

- **Vigilancia competitiva**, que se ocupa de los competidores,
- **Vigilancia comercial**, que se ocupa de los clientes y proveedores,
- **Vigilancia tecnológica**, que se ocupa de las nuevas tecnologías, y
- **Vigilancia del entorno social**, que se ocupa de la detección de los hechos que pueden tener incidencia en el futuro.

Una **organización inteligente** es aquella que integra la **Gestión del Conocimiento y la Inteligencia Competitiva** en la **Toma de Decisiones** (López Forniés, Fernández Sora, Manchado Pérez, Agustín Hernández, & Zabala Díez, 2007). El Sistema de Toma de Decisiones en una organización inteligente se ajusta a la Figura nº 1.

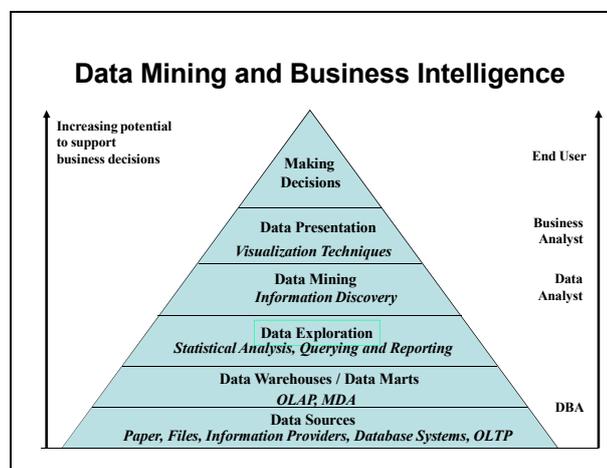


Figura 1: Sistema de toma de decisiones en una organización inteligente. (Tomado de J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2001, <http://freedownloadbooks.net/han-kamber-2001-ppt.html>)

4.3. VIGILANCIA ESTRATÉGICA, CIENTÍFICA Y TECNOLÓGICA.

El **Análisis Estratégico** consiste en esencia en la identificación y tipificación de los elementos principales y de las relaciones existentes entre ellos. Analizar estratégicamente es buscar lo esencial, lo fundamental, lo importante de todo sistema, prescindiendo de lo accesorio o poco relevante.

La **Vigilancia Estratégica, Científica y Tecnológica** es el conjunto de tareas que buscan proporcionar buena información a la persona idónea en el momento adecuado (Prieto Viñuela, 2004) y en el lugar adecuado; buena información, se entiende, sobre el entorno científico y tecnológico de la organización.

La **Planificación Estratégica** es el principal instrumento para la gestión de una organización a corto plazo. Es una herramienta para la **mejora continua** y para la **innovación** en la organización.

La Vigilancia Estratégica, Científica y Tecnológica se encuadra dentro de las actividades propias de la **Inteligencia Competitiva** de una organización. Va dirigida a analizar el conocimiento científico y tecnológico producido por su entorno de interés, con la finalidad de crear nuevo conocimiento de carácter estratégico y de utilidad en la **Toma de Decisiones**.

4.4. INGENIERÍA Y CARTOGRAFÍA DEL CONOCIMIENTO

La **Ingeniería del conocimiento** (*Knowledge engineering*) es la "Ingeniería que tiene por objetivo la creación de indicadores, métodos e instrumentos - (**Sistemas de Conocimiento**) - de naturaleza matemática, informática y lingüística, para el análisis y la representación cartográfica de la información científica y técnica" (Polanco, 1997).

La Ingeniería del Conocimiento es la rama de la **Inteligencia Artificial** dedicada a al diseño y desarrollo de los Sistemas Expertos, llamados **Sistemas de Conocimiento**, que permiten la generación de nuevo conocimiento a partir del conocimiento objetivo o explícito almacenado en las bases de datos.

La **Ingeniería y la Cartografía del Conocimiento** son de gran utilidad para las organizaciones inteligentes en investigación, desarrollo e innovación ya que facilita y mejora el funcionamiento de su **Sistema de Toma de Decisiones**. Los subsistemas de Gestión del Conocimiento, de Inteligencia Competitiva, de Vigilancia Tecnológica y de Diseño Industrial constituyen el Sistema de Toma de Decisiones de carácter estratégico de las organizaciones

para la Innovación (López Forniés, Fernandez Sora, Manchado Perez, Agustín Hernandez, & Zabala Diez, 2007) (ver figura nº 2).

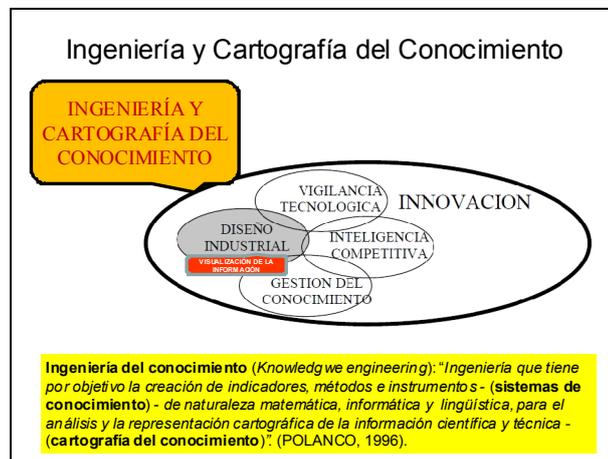


Figura 2: La Ingeniería y la Cartografía del Conocimiento en el Sistema de Toma de Decisiones de carácter estratégico de la organización. Realizada a partir de López Forniés, Fernandez Sora, Manchado Perez, Agustín Hernandez, & Zabala Diez, 2007.

Se conoce como **Knowledge Discovery in Databases (KDD)** la metodología de creación de nuevo conocimiento a partir de bases de datos bibliográficas (Pierret & Boutin, 2004). La Ingeniería del Conocimiento desarrolla sistemas basados, en todas o parte de sus fases, en la metodología KDD (Hernández Orallo, Juan Lizandra, Minaya Collado, & Monserrat Aranda) (ver Figura nº 3)

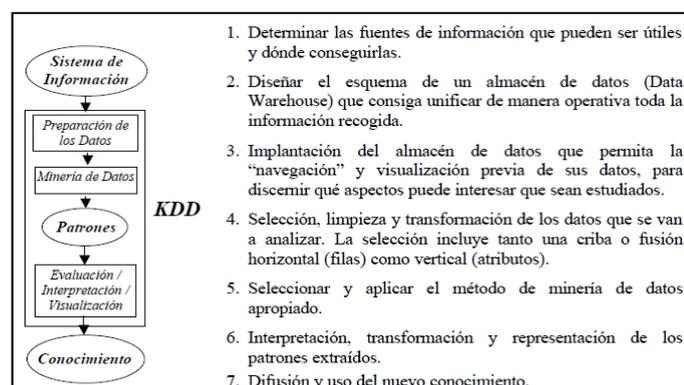


Figura 3: Fases del proceso Knowledge Discovery in Databases (KDD). (Tomado de José Hernández Orallo, <http://www.dsic.upv.es/~jorallo/master/seminari.part.I.pdf>)

Uno de las fases del proceso *KDD* es la **Minería de Datos (DM)**. Se conoce como minería de datos la extracción dirigida de la información existente en las bases de datos con el fin de descubrir patrones, relaciones o asociaciones para generar nuevo conocimiento (López Rodríguez, 2004)

Un sistema típico de minería de datos obedece al siguiente esquema, (ver Figura nº 4).

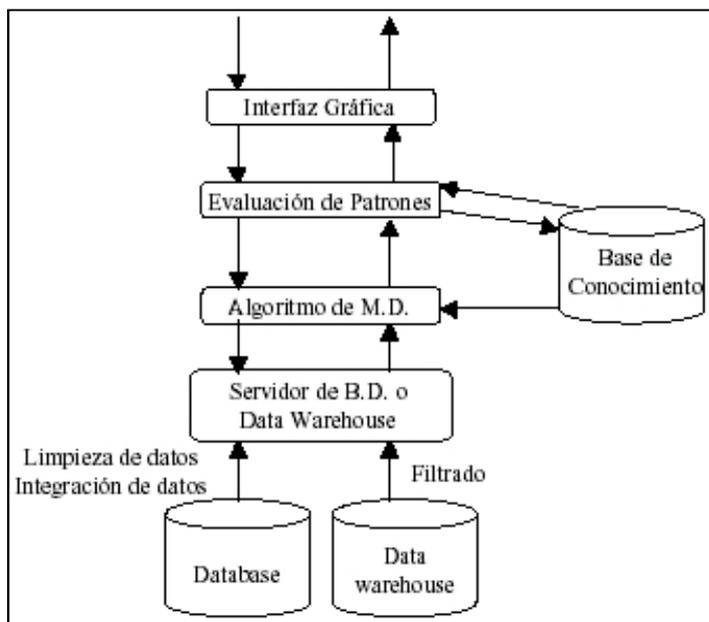


Figura 4: Arquitectura de un sistema típico de minería de datos. Tomada de Gustavo Iván García, <http://www.monografias.com/trabajos27/datamining/datamining.shtml#arquitect>.

Algunos tipos de Minería de Datos son:

- a) **Web mining**, que a su vez puede ser clasificada en:
 - a.1) **Web content mining** (minería del contenido web)
 - a.2) **Web structure mining** (minería de la estructura web)
 - a.3) **Web usage mining** (minería del uso web)
- b) **Text mining** (minería de datos textuales)
- c) **Spatial data mining** (minería de datos espaciales)

La Minería de texto a diferencia de la Recuperación de Información, con la cual no ha de confundirse, se ocupa de descubrir información no expresada en

ningún documento individual de la colección documental estudiada (Molina Félix, 2002).

Una de las técnicas empleadas en *Data Mining* es el **análisis de agrupamientos o cluster**. Se emplea para identificar dentro de la estructura general estudiada infraestructuras o subclases de elementos que guardan entre sí características comunes que los identifican y los diferencian de las demás. Se han definido diferentes algoritmos para identificar los agrupamientos; García (2005) realiza una clasificación de los algoritmos de agrupamiento en Minería de Datos Espaciales que se puede observar en el cuadro siguiente (ver figura nº 5)

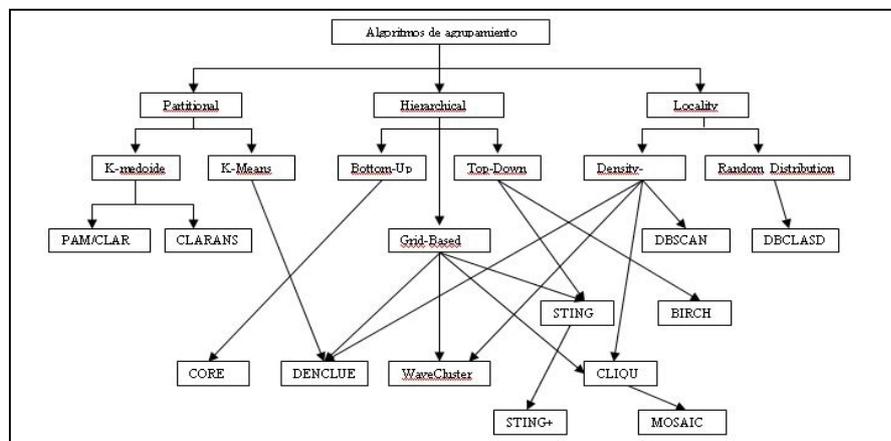


Figura 5: Clasificación de los algoritmos de agrupamiento en Minería de Datos Espaciales (MDE). (Tomado de Gustavo Iván García, <http://www.monografias.com/trabajos27/datamining/datamining.shtml>)

4.5. SISTEMAS DE CONOCIMIENTO

Los **Sistemas de Conocimiento** son Sistemas Expertos ideados para crear nuevo conocimiento a partir del conocimiento explícito presente en los documentos de las bases de datos. Siguen el proceso KDD en sus diferentes fases.

Tienen su origen en los **softwares cuantitativos** desarrollados para facilitar el análisis de grandes conjuntos documentales (artículos científicos, patentes, tesis doctorales, etc.)(ver Figura nº 6); éste es el propósito común de todos ellos (Larreina, Hernando, & Grisaleña, 2006). Su evolución ha sido constante durante los últimos años.

Permiten la creación de nuevo conocimiento en diferentes formatos de salida: listados basados en índices, grafos, diagramas, mapas, etc.

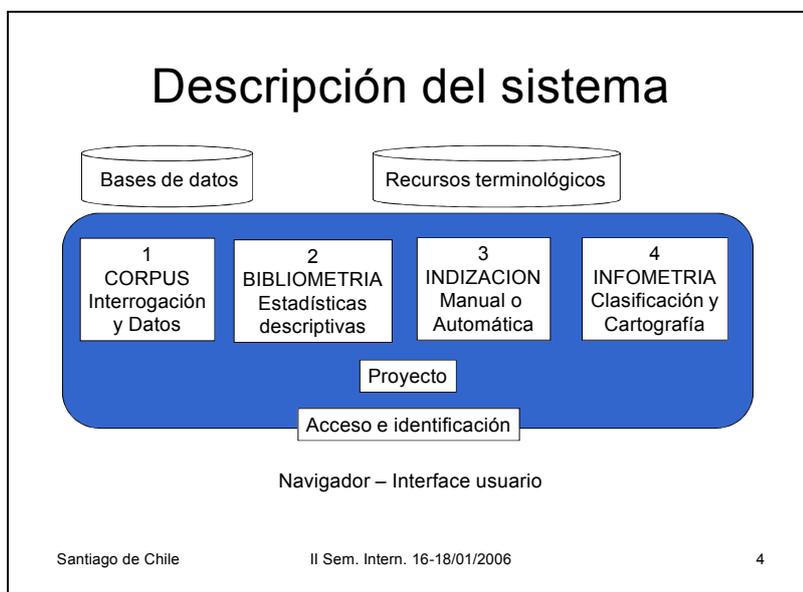


Figura 6:: Esquema del Sistema Stanalist (Polanco, 2006). El módulo "Infometría" de Stanalist es un sistema de clasificación automática basado en dos programas alternativos, Neurodoc y Sdoc. Tomado de <http://www.cinzel.cl/documentos/Recursos/STANALYST2006.ppt>

Algunos Sistemas de Conocimiento son:

- PatentLab-II (PatentLab-II. Extracting intelligence from patent data., 2001), <http://www.wisdomain.com>
- VantagePoint (2010), <http://www.thevantagepoint.com>
- Matheo Patent (Search, Analysis and Survey of Patents for Intellectual property, Innovation & Competitive Intelligence., 2010), <http://www.matheo-software.com>
- Matheo Analyzer (Matheo Analyzer : Database Analysis, Information Mapping, 2009), <http://www.matheo-software.com>
- Aureka (Aureka, 2010), <http://science.thomsonreuters.com/es/productos/aureka>
- PAT-LIST-WPI (PAT-LIST-WPI 3.0), <http://www.raytec.co.jp/EngPages/RaytecProd/EngPatListWpi.html>
- Delphion (Delphion, 2010), <http://science.thomsonreuters.com/es/productos/delphion>

- Patent Matrix (Patent Matrix - Claims Mapping - Competitive Landscape Analysis), <http://www.patentmatrix.com>
- Clearforest Analytics (ClearForest, 2010), <http://clearforest.com/>
- VxInsight (VxInsight), <http://www.cs.sandia.gov/projects/VxInsight.html>
- Tetralogie (Tetralogie), <http://atlas.irit.fr/TETRALOGIE/tetrajeu.htm>
- Bibexcel (Bibexcel), <http://www8.umu.se/inforsk/Bibexcel/>
- Analyst's notebook (i2's Intelligence-Led Operations Platform, 2010), <http://www.i2group.com/us>
- Data SOMining (DataSOMining), <http://www.dynamics.unam.edu/DinamicaNoLineal3/datasomining.htm>
- PROInTec (Giráldez Reyes, Díaz Pérez, & Armas Peñas, PROInTec: un software para el tratamiento inteligente de datos, 2008), <http://scielo.sld.cu/pdf/aci/v17n5/aci06508.pdf>
- Stanalist (Stanalyst, 2008), <http://stanalyst.inist.fr/>
<http://sites.google.com/site/xavierpolanco/stanalist2>
- CoPalRed (Sistema de Conocimiento CopalRed), <http://ec3.ugr.es/copalred/>

4.6. ANÁLISIS DE PALABRAS ASOCIADAS (CO-WORD ANALYSIS)

El investigador transmite su **conocimiento tácito** al resto de la clase científica en los documentos que escribe. Estos documentos se constituyen en **conocimiento explícito** disponible para otros investigadores.

Los documentos de los investigadores almacenados en las bases de datos, pueden ser analizados en su conjunto para descubrir y crear nuevo conocimiento mediante softwares especializados que se basan en métodos matemáticos, informáticos y lingüísticos. Uno de estos métodos es el **Análisis de Palabras Asociadas (Co-word analysis)**.

El Análisis de Palabras Asociadas:

- Se utiliza para estudiar la **estructura y la dinámica del conocimiento** en un dominio documental concreto.
- Es un método de **Knowledge Discovery in Databases (KDD)** (He, 1999), por el cual se crea nuevo conocimiento a partir de bases de datos bibliográficas. En este caso la fase de **Data Mining**

(DM) se denomina **Text Mining (TM)** (Fan, Wallace, Rich, & Zhang, 2006).

- Puede ser empleado de modo **retrospectivo**, **"a tiempo real"** o de forma **prospectiva**.
- Se basa en dos teorías sociológicas: la **Teoría Actor-Red** y la **Teoría de la Traducción**.
- Es considerado como **Índice relacional de 2ª generación** en **Evaluación de la Ciencia**
- Se utiliza para estudiar las **Redes Tecnocientíficas**.

El análisis de palabras asociadas se basa en la **coocurrencia de las palabras** en las unidades de análisis o de registro. Las coocurrencias se llevan sobre una matriz simétrica que posteriormente es normalizada por el índice de equivalencia para obtener valores entre 0 y 1.

La **matriz de coocurrencias** o **matriz de adyacencia** es una matriz simétrica que muestra en los puntos de corte entre filas y columnas, **c_{ij}** , las coocurrencias de los valores que encabezan cada fila **i** y cada columna **j** .

El **Índice de Equivalencia** o de asociación entre las palabras **i** y **j** se obtiene:

$$e_{ij} = \frac{c_{ij}^2}{c_i c_j}$$

Ecuación 1: Índice de Equivalencia entre dos palabras que coocurren.

donde:

e_{ij} es el índice de equivalencia o de asociación entre las palabras i y j

c_{ij} es el valor de co-ocurrencia de las palabras i y j

c_i es el valor de ocurrencia de la palabra i

c_j es el valor de ocurrencia de la palabra j .

Los **valores de e_{ij}** oscilan entre 0 y 1. Cuando dos palabras no aparecen nunca en los mismos documentos su coocurrencia es nula y por tanto su índice de equivalencia vale cero. Este índice es independiente del tamaño de la muestra.

La **matriz de enlaces o matriz normalizada** es la matriz que se obtiene de normalizar la matriz de adyacencia mediante el índice de equivalencia. Al normalizar la matriz de adyacencia se consigue visualizar mejor los enlaces entre las palabras clave, los investigadores y/o las revistas, ya que éstos no quedan enmascarados por los diferentes valores de ocurrencia de cada uno de ellos; así los enlaces fuertes, de valores igual a 1 o próximos a la unidad, son el exponente de uniones fuertes (pares de palabras clave que aparecen en los mismos artículos científicos; pares de investigadores que colaboran muy estrechamente; palabra clave e investigador que se repiten en los mismos artículos científicos o palabra clave, revista e investigador que coinciden a menudo, etc.)

4.7. TEORÍA ACTOR-RED Y TEORÍA DE LA TRADUCCIÓN

Ambas son **Teorías Sociotecnológicas** de la Ciencia y la Técnica, desarrolladas en la década de los ochenta principalmente por Bruno **Latour**, filósofo, sociólogo y antropólogo; Michel **Callon**, ingeniero y sociólogo; y Jhon **Law**, sociólogo.

La **Teoría Actor-Red** considera la Ciencia y la Técnica como una red de conocimiento, en la que un conjunto de **actores o "actantes"** (palabras clave, investigadores, centros de investigación, revistas, países, etc.) están relacionados entre sí mediante una **asociación cognitiva común**, constituyendo una estructura en forma de **red** (Figura nº 7).

Los actores son elementos dinámicos que cambian a lo largo del tiempo y que en ese cambio redefinen sus relaciones y por tanto la **red sociocognitiva** que conforman.

La **Teoría de la Traducción** también llamada **Sociología de la Traducción** (entendida traducción como **conversión, transformación, variación o cambio**), estudia los cambios que se producen en las relaciones entre los actores de la red. Estos cambios en las relaciones entre los actores producen la aparición, el fortalecimiento, el equilibrio, el debilitamiento y la desaparición de los actores. La aparición de nuevos actores se produce por emergencia o por convergencia; el fortalecimiento se produce por convergencia o por evolución incremental; el equilibrio, por evolución estable; el debilitamiento, por evolución decremental o por divergencia y la desaparición, por bifurcación

arista (enlace). Este valor puede indicar la presencia (1) o ausencia (0) de relación o puede coincidir con el nº de veces que se la relación.

- **Matriz normalizada.** Es una matriz cuadrada simétrica en la que las filas y las columnas son los vértices o nodos (los actores) y la intersección entre cada fila y cada columna de la matriz, el valor de la arista (enlace). Este valor del enlace indica la fortaleza de la relación (que toma valores entre 0, ausencia de relación, y 1, relación muy fuerte).
- **Matriz de distancias geodésicas.** Es una matriz cuadrada simétrica en la que las filas y las columnas son los vértices o nodos (los actores) y la intersección, entre cada fila y cada columna de la matriz, la **distancia geodésica** entre los dos vértices. Siendo la distancia geodésica el número de relaciones (aristas) en el camino más corto de un actor (vértice) a otro (ver Figura nº 8).

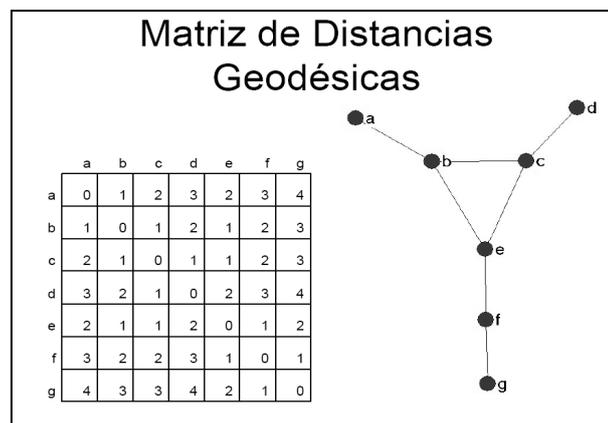


Figura 8: Tomada de <http://www.analytictech.com/networks/Teoriadegrafos.pdf>

Los grafos se puede clasificar como:

- **Conexos**, si cada par de vértices está conectado por un camino; es decir, si para cualquier par de vértices (a, b), existe al menos un camino posible desde a hacia b.
- **No conexos**, si no se da lo anterior.

Cada uno de los elementos de un grafo no conexo se denomina **componente** (ver Figura nº 9).

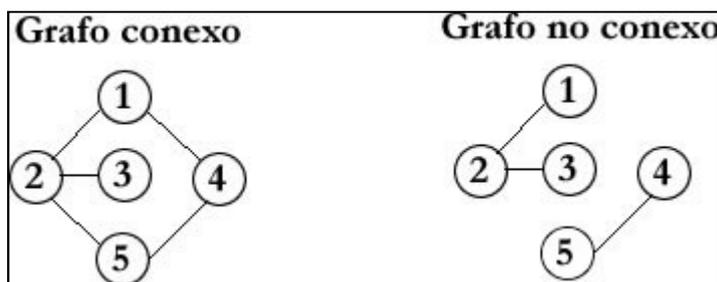


Figura 9: Tomada de http://es.wikipedia.org/wiki/Teoría_de_grafos

4.9. TRAZADO DE GRAFOS

El trazado de grafos consiste en encontrar para cada vértice una posición en el plano, de modo que el resultado final se adecúe a una estética predeterminada. Existen diversos procedimientos para trazar los grafos. Los criterios de trazado pueden ser variados:

- Minimizar cruces,
- Minimizar el área,
- Maximizar el ángulo entre ejes,
- Mostrar simetría,
- Mantener una longitud uniforme de las aristas,
- Que todos los nodos sean visibles al mismo tiempo,
- Utilizar el área de trazado lo máximo posible,
- Que los vértices no estén muy cerca entre sí,
- Que los vértices conectados estén cerca,
- Etc.

Dentro de las familias de algoritmos desarrollados para trazar los grafos existe el grupo de los **algoritmos dirigidos por fuerzas** (ver Figura nº 10).

Los algoritmos dirigidos por fuerzas modelizan el grafo como un sistema físico y buscan el equilibrio de ese sistema. Los algoritmos más importantes de este grupo son (Aiello & Silveira, 2004):

- **SE, Spring Embedder**, (Eades, 1984).
- **FR**, (Fruchterman y Reingold, 1991).
- **GEM**, (Frick et al., 1995).
- **SM**, (Sugiyama y Misue, 1995).
- **KK**, (Kamada y Kawai, 1989).
- **DH**, (Davidson y Harel, 1996).

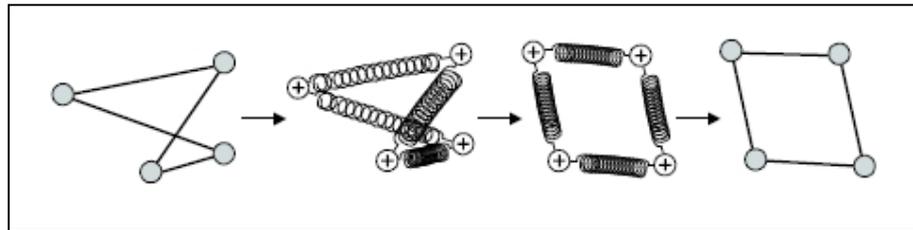


Figura 10: Ejemplo de funcionamiento de los algoritmos de los modelos dirigidos por fuerzas. Tomada de <http://www-ma2.upc.edu/rsilveira/pubs/MScThesis.pdf>

4.9.1. ALGORITMO DE KAMADA-KAWAI

Este algoritmo fue desarrollado por Kamada y Kawai (en adelante KK). Es un algoritmo apropiado para visualizar grafos ya que facilita la observación global de la red.

"KK proponen un modelo en el que hay un único tipo de fuerzas, que pueden ser vistas como resortes que conectan entre sí a todos los vértices. Los resortes obedecen a la ley de Hooke (fuerzas lineales) y cada resorte tiene una longitud ideal y una rigidez distinta. Uno de los aportes más importantes y originales de KK es que la distancia ideal de cada resorte es proporcional a la distancia geodésica entre los correspondientes vértices en la red. La distancia geodésica se refiere al número mínimo de enlaces entre dos vértices. Por lo tanto, el algoritmo KK es esencialmente un escalamiento multidimensional asociado a la matriz de distancias geodésicas.

KK trabaja directamente con la energía del sistema, y busca un estado de energía mínima es decir, una posición para los vértices del grafo en la cual la energía sea localmente mínima.

Hay un único tipo de fuerzas definido, que son resortes entre todo par de vértices u y v . Los mismos tienen una rigidez y longitud ideal determinada y producen el efecto de atraer a los vértices cuando el resorte está estirado, y repelerlos cuando está comprimido. La fuerza que se usa para modelar el resorte entre u y v es la siguiente:

$$f(u, v) = S_{uv} (d_{uv} - \delta_{uv})$$

Ecuación 2: Fuerzas entre dos nodos conectados.

Siendo:

d_{uv} distancia euclídea entre los nodos u y v ,

δ_{uv} distancia geodésica entre los nodos u y v ,

$S_{uv} = \frac{S}{\delta_{uv}^2}$ es un **parámetro de rigidez del resorte**, que es más fuerte para vértices cuya distancia geodésica es menor y decrece a medida que su distancia geodésica aumenta, y

S constante.

El algoritmo no trabaja con las fuerzas directamente, sino con la energía potencial del sistema (que es la integral de estas fuerzas), cuya expresión es:

$$E(u, v) = \int f(u, v) d(d_{uv} - \delta_{uv}) = \frac{1}{2} S_{uv} (d_{uv} - \delta_{uv})^2$$

Ecuación 3: Energía potencial del enlace entre los nodos u y v .

Así, la energía total del sistema es:

$$E = \sum_{u, v \in \bar{V} \atop u \neq v} \frac{1}{2} S_{uv} (d_{uv} - \delta_{uv})^2$$

Ecuación 4: Energía Total de la Red.

De esta manera, **el algoritmo busca la minimización de esa energía, a partir de las posiciones iniciales de los vértices** (Solís Arias, 2007)“:

$$\text{Min}(E) = \sum_{u,v \in \overline{V}, u \neq v} \frac{1}{2} S_{uv} (d_{uv} - \delta_{uv})^2$$

Ecuación 5: Minimización de la energía de la Red (I).

$$\text{Min}(E) = \sum_{u,v \in \overline{V}, u \neq v} \frac{1}{2} S_{uv} \left(\sqrt{(x_u - x_v)^2 + (y_u - y_v)^2} - \delta_{uv} \right)^2$$

Ecuación 6: Minimización de la energía de la Red (II).

Las condiciones necesarias para que las posiciones de los vértices estén en un mínimo local de la función E son:

$$\frac{\partial E}{\partial x_m}, \frac{\partial E}{\partial y_m} = 0, \quad \forall m \in V$$

Ecuación 7: Cálculo del mínimo de energía (I).

El objetivo del algoritmo KK es encontrar una posición para los vértices donde las derivadas parciales sean nulas, o lo mismo, buscar una raíz de las derivadas parciales:

$$\frac{\partial E}{\partial x_m} = \sum_{v \in \overline{V}, v \neq m} S_w \left\{ (x_m - x_v) - \frac{\delta_w (x_m - x_v)}{\sqrt{(x_m - x_v)^2 + (y_m - y_v)^2}} \right\}$$

$$\frac{\partial E}{\partial y_m} = \sum_{v \in \overline{V}, v \neq m} S_w \left\{ (y_m - y_v) - \frac{\delta_w (y_m - y_v)}{\sqrt{(x_m - x_v)^2 + (y_m - y_v)^2}} \right\}$$

Ecuación 8: Cálculo de la mínima energía local (II).

"Para esto utilizan una variante de la técnica Newton-Raphson. Primero notan que si bien hay $2n$ ecuaciones (Ecuación nº 8), no pueden usar directamente Newton-Raphson para $2n$ variables porque las variables no son independientes entre sí. Entonces deciden adoptar la siguiente estrategia: toman un vértice v_m por cada vez y suponen los otros fijos y buscan una raíz de las derivadas (Ecuación nº 8) considerando todas las otras posiciones constantes. De esta manera, el problema pasa a ser de 2 variables ($x_m; y_m$), sobre el cual usan Newton-Raphson para 2 variables (Burden & Faires, 1998)".

$$x_{n+1} = x_n - J(x)^{-1}F(x)$$

Ecuación 9: Método iterativo de Newton-Raphson.

Donde,

x_{n+1} y $x_n \in \mathbb{R}^2$, $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es la función cuya raíz se quiere encontrar,

$J(x) \in \mathbb{R}^{2 \times 2}$ es la matriz Jacobiano de F

En este caso:

$$F(x) = \begin{pmatrix} \frac{\partial E}{\partial x_m} \\ \frac{\partial E}{\partial y_m} \end{pmatrix}$$

y,

$$J(x) = \begin{pmatrix} \frac{\partial^2 E}{\partial x_m^2} & \frac{\partial^2 E}{\partial y_m \partial x_m} \\ \frac{\partial^2 E}{\partial x_m \partial y_m} & \frac{\partial^2 E}{\partial y_m^2} \end{pmatrix}$$

"En cada iteración, el algoritmo KK va a tomar un vértice v_m y va a buscar mediante Newton-Raphson una posición que será un mínimo de la función de energía E , donde las posiciones de todos los otros vértices han sido congeladas. Luego se repetirá este mismo proceso cambiando el vértice v_m . El pseudo-código del algoritmo es el siguiente:

KK(Grafo G)

1. Calcular distancias teóricas
2. Asignar una posición inicial a cada vértice de G
3. Mientras ($\max_i d_i > \epsilon$)

3.1. Tomar v_m tal que $d_m = \max_i d_i$

3.2. Mientras ($d_m > \epsilon$)

3.2.1. Calcular una nueva posición de v_m (método iterativo de Newton-Raphson) y recalcular d_m

$$d_m = \sqrt{\left(\frac{\delta E}{\delta x_m}\right)^2 + \left(\frac{\delta E}{\delta y_m}\right)^2}$$

Ecuación 10: Gradiente de la energía del vértice v_m

Donde d_m es la norma del gradiente de la energía del vértice v_m , o visto de otra forma, la magnitud de las fuerzas que actúan sobre v_m .

Se ha de notar que el vértice que se elige para ser movido en cada iteración del ciclo principal es el que está en mayor "desequilibrio", o el que está siendo afectado por la mayor fuerza.

Se ha de notar también que el paso 3.2. es exactamente la aplicación del método iterativo de Newton-Raphson, usando como criterio de corte un umbral en las derivadas (que es la función cuya raíz se quiere encontrar, así que es de esperar que d_m tienda a cero).

El algoritmo KK se centra en un único criterio estético, que las distancias en el trazado final sean iguales a las distancias teóricas. Es interesante ver que este criterio implica de cierta manera los criterios de trazado de longitud de aristas uniforme (esto es inmediato) y de vértices distribuidos uniformemente. Este último se puede entender pensando en que se evitan concentraciones de vértices, ya que si los mismos no están conectados y su distancia teórica es grande, se verán forzados a alejarse, y en caso de que estén conectados entonces está bien que estén cerca unos de otros, debido al criterio de las aristas de longitud uniforme. Esto explica que los resultados

obtenidos con KK sean muy buenos y muy similares a los obtenidos con otros algoritmos que se ocupan de los otros criterios estéticos como FR y GEM" (Aiello & Silveira, 2004).

El **algoritmo KK** tiende a igualar en el plano euclídeo (el plano de trazado de la red) las distancias entre los vértices con sus distancias geodésicas; como resultado, **"los vértices (conectados) similares se acercan y los diferentes se alejan"**. Además, si las aristas toman los valores de los índices de equivalencia de los vértices que conectan, observamos como los vértices de mayor similitud aparecen próximos y unidos por líneas de mayor grosor (ver Figura nº 11).

Los vértices similares aparecen cercanos constituyendo agrupaciones. Las agrupaciones también se posicionan en la red según su mayor o menor similitud; **"los grupos (conectados) similares se acercan y los diferentes se alejan"**.

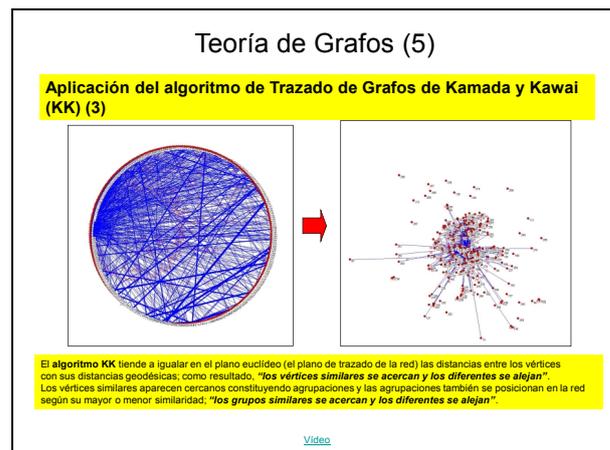


Figura 11: Aplicación del algoritmo KK al trazado de grafos.

En todas las diferentes interacciones en las que se aplique el algoritmo KK a una misma red se observa como **"los vértices (conectados) similares se acercan y los diferentes se alejan"** y **"los grupos (conectados) similares se acercan y los diferentes se alejan"** (ver Figura nº 12).

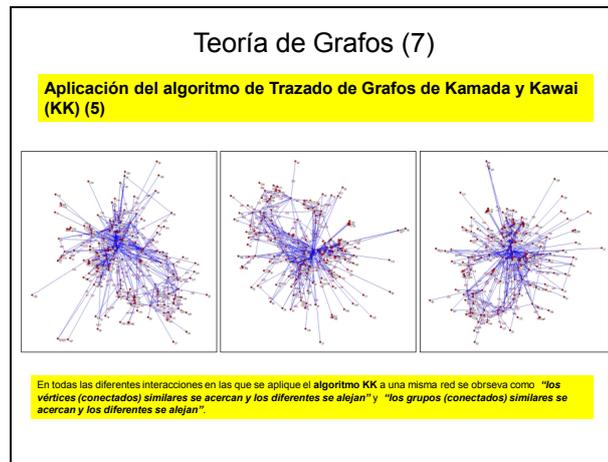


Figura 12: Diferentes interacciones del algoritmo KK un mismo grafo.

4.10. ESPACIO EUCLÍDEO

El Espacio Euclídeo es el espacio matemático n-dimensional. El Espacio Euclideo de dos dimensiones se denomina Plano Euclídeo.

El Espacio Euclídeo se puede considerar un Espacio Vectorial n-dimensional real.

En el Plano Euclídeo cada punto viene definido por dos coordenadas:

$$A (x_A, y_A)$$

$$B (x_B, y_B)$$

Ecuación 11: Coordenadas de un punto en el Plano Euclídeo.

de manera que la **distancia euclídea** entre esos puntos se define como:

$$d_{\overline{AB}} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

Ecuación 12: Distancia euclídea.

La distancia euclídea equivale a la longitud del segmento de recta que los une.

Sobre un plano euclídeo se puede trazar un **sistema de referencia**, sistema de coordenadas cartesianas, conformado por dos rectas perpendiculares, **ejes cartesianos**, que se cortan en un punto denominado **origen de coordenadas**. En el **sistema de coordenadas cartesianas** cada punto del plano puede ser situado por dos números (x, y) llamados **abscisa y ordenada**, que son las distancias ortogonales de ese punto a los ejes cartesianos (ver Figura nº 13).

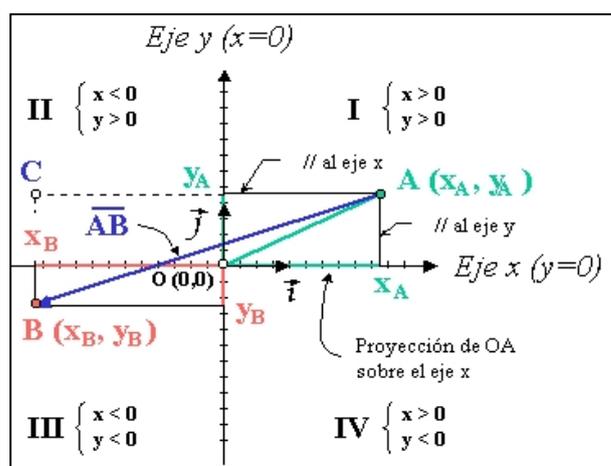


Figura 13: Sistema de coordenadas cartesianas.
http://es.wikipedia.org/wiki/Coordenadas_cartesianas

4.11. ESPACIO VECTORIAL

El plano euclídeo E^2 en el que se ha definido un sistema de coordenadas cartesianas se puede considerar como un espacio vectorial R^2 formado por **vectores**, que se representan gráficamente como segmentos de recta orientados o dirigidos.

Cada vector es una magnitud física que queda caracterizado por su **módulo o longitud** y por su **dirección, o ángulo que forma con los ejes de coordenadas**.

Cualquier vector de R^2 puede descomponerse en dos vectores paralelos a los ejes de coordenadas x e y, a los que se denomina **componentes del vector**. Los componentes de un vector se escriben entre paréntesis y separados por comas.

Cualquier vector a puede expresarse como combinación de sus componentes cartesianos.

$$\vec{a} = (\vec{a}_x, \vec{a}_y)$$

Ecuación 13: Expresión de los vectores.

El vector nulo es el vector:

$$\vec{0} = (\vec{0}_x, \vec{0}_y)$$

Ecuación 14: Vector nulo.

La suma de dos vectores es otro vector:

$$\vec{a} = (\vec{a}_x, \vec{a}_y)$$

$$\vec{b} = (\vec{b}_x, \vec{b}_y)$$

$$\vec{a} + \vec{b} = \vec{c}$$

Cuyos vectores componentes cartesianos son:

$$c = (\vec{c}_x, \vec{c}_y)$$

Ecuación 15: Vectores componentes cartesianos.

De tal manera que:

$$\vec{c}_x = (\vec{a}_x + \vec{b}_x)$$

$$\vec{c}_y = (\vec{a}_y + \vec{b}_y)$$

Ecuación 16: Suma de dos vectores.

La **longitud o módulo de un vector a** se determina a partir del Teorema de Pitágoras:

$$|\vec{a}| = \sqrt{|\vec{a}_x|^2 + |\vec{a}_y|^2}$$

Ecuación 17: Longitud o modulo de un vector.

En el plano euclídeo E^2 la longitud del vector a se expresa como la longitud del segmento de recta que une los puntos situados al inicio y al final del segmento (distancia euclídea entre dichos puntos). Si a estos dos puntos los denominamos I y F, y cada uno de ellos tiene las coordenadas:

$$\vec{a} = \overline{IF}$$

$$I (x_I, y_I)$$

$$F (x_F, y_F)$$

Resulta que:

$$|\vec{a}| = \sqrt{|\vec{a}_x|^2 + |\vec{a}_y|^2} = d_{IF} = \sqrt{(x_I - x_F)^2 + (y_I - y_F)^2}$$

Ecuación 18: Longitud de un vector en el plano euclídeo.

Y por tanto:

$$|\overline{a_x}|^2 = (x_I - x_F)^2$$

$$|\overline{a_y}|^2 = (y_I - y_F)^2$$

Los módulos o longitudes de los vectores solo toman valores positivos por tanto:

$$|a_x| = (x_F - x_I)$$

$$|a_y| = (y_F - y_I)$$

Por lo que se deduce que en un espacio euclídeo E^2 , en el que se ha definido un sistema de coordenadas cartesianas, asimilado al espacio vectorial R^2 , los módulos o longitudes de los vectores componentes coinciden con la distancia euclídea entre los puntos final e inicio del vector.

4.11.1. ISOBARICENTRO

Sea el espacio euclídeo E^2 en el que se ha definido un sistema de coordenadas cartesianas, asimilado al espacio vectorial R^2 .

Sean A_1, \dots, A_n , n puntos, y m_1, \dots, m_n , n masas de dichos puntos. Entonces el **baricentro** de los (A_i, m_i) es el punto B definido como sigue:

$$\overline{OB} = \frac{m_i \overline{OA_i}}{m_i}$$

Ecuación 19: Cálculo del baricentro.

Esta definición depende del punto O , que puede ser cualquiera. Si se toma el origen del plano o del espacio, se obtiene las coordenadas del baricentro, como promedio ponderado por los m_i , de las coordenadas de los puntos A_i :

$$x_B = \frac{x_{A_i}}{m_i}$$

$$y_B = \frac{y_{A_i}}{m_i}$$

Ecuación 20: Cálculo de las coordenadas del Baricentro.

Un **isobaricentro** es un baricentro con todas las masas iguales entre sí; es usual en tal caso tomarlas iguales a 1. Si no se precisan las masas, el baricentro es por defecto el isobaricentro.

$$x_{isoB} = \frac{x_{A_i}}{n} = \bar{x}_{A_i}$$

$$y_{isoB} = \frac{y_{A_i}}{n} = \bar{y}_{A_i}$$

Ecuación 21: Cálculo de las coordenadas del Isobaricentro.

Las coordenadas cartesianas del isobaricentro son la media de las coordenadas cartesianas de todos los puntos.

4.12. VISUALIZACIÓN DE LA INFORMACIÓN

Visualización es la formación en la mente de una imagen de un concepto. En este contexto, se entiende visualización como la representación gráfica de variables asociadas al concepto que se quiere visualizar (Dürsteler, 2000).

Visualización de la Información es el proceso de interiorización del conocimiento mediante la percepción de información (ver Figura nº 14).

La Visualización de la Información se beneficia básicamente de que:

- los seres humanos reciben información de forma eminentemente visual ya que es el sentido con mayor ancho de banda, es decir, que proporciona mayor cantidad de información.
- La capacidad simbólica del cerebro humano (Dürsteler, 2000).



Figura 14: Utriusque cosmi maioris scilicet et minoris [...] historia, tomus II (1619), tractatus I, sectio I, liber X, De triplici animae in corpore visione. Fecha: 1619 (1619). Autor Robert Fludd. Tomada de <http://upload.wikimedia.org/wikipedia/de/0/0c/RobertFuddBewusstsein17Jh.png>

Una **metáfora visual** es la representación de un sistema mediante atributos visuales propios de un sistema diferente que ya es familiar a los usuarios y que se comporta de una manera análoga (Dürsteler, 2000).

A continuación se muestran algunos ejemplos de las metáforas más usuales:

1) Metáfora del escritorio (imágenes de los archivadores y de las carpetas de documentos en el escritorio del sistema operativo Windows)

2) Metáfora del árbol (ver Figuras nº 15, 16 y 17)

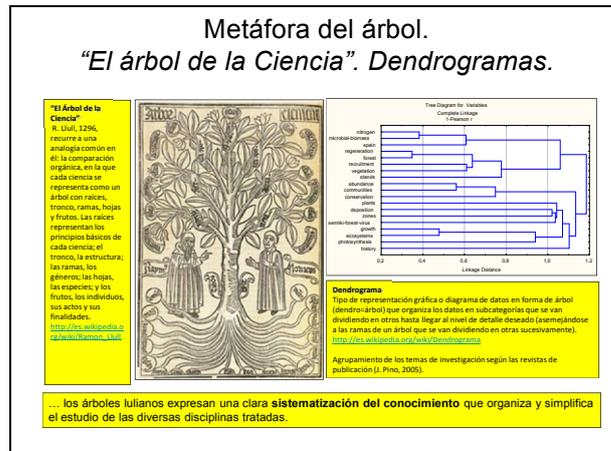


Figura 15: "El Árbol de la Ciencia". R. Llull, 1296, recurre a una analogía común en él: la comparación orgánica, en la que cada ciencia se representa como un árbol con raíces, tronco, ramas, hojas y frutos. Las raíces representan los principios básicos de cada ciencia; el tronco, la estructura; las ramas, los géneros; las hojas, las especies; y los frutos, los individuos, sus actos y sus finalidades. Los árboles lullianos expresan una clara sistematización del conocimiento que organiza y simplifica el estudio de las diversas disciplinas tratadas. Tomada de http://es.wikipedia.org/wiki/Ramon_Llull. Dendrograma. Tipo de representación gráfica o diagrama de datos en forma de árbol (dendro=árbol) que organiza los datos en subcategorías que se van dividiendo en otros hasta llegar al nivel de detalle deseado (asemejándose a las ramas de un árbol que se van dividiendo en otras sucesivamente). <http://es.wikipedia.org/wiki/Dendrograma>. Tomada de "Agrupamiento de los temas de investigación según las revistas de publicación" (J. Pino, 2005).

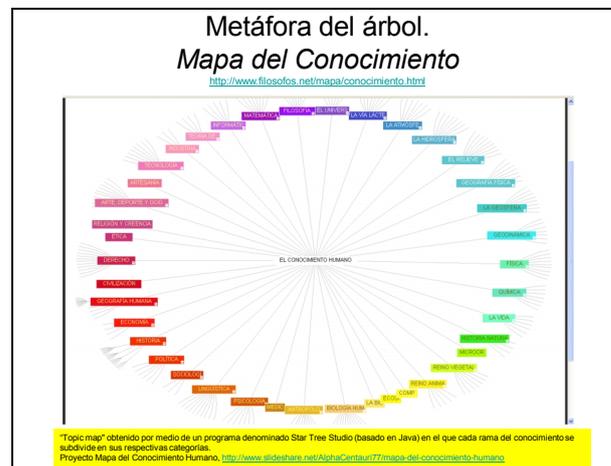


Figura 16: "Topic map" obtenido por medio de un programa denominado Star Tree Studio (basado en Java) en el que cada rama del conocimiento se subdivide en sus respectivas categorías. Proyecto Mapa del Conocimiento Humano, <http://www.slideshare.net/AlphaCentauri77/mapa-del-conocimiento-humano>

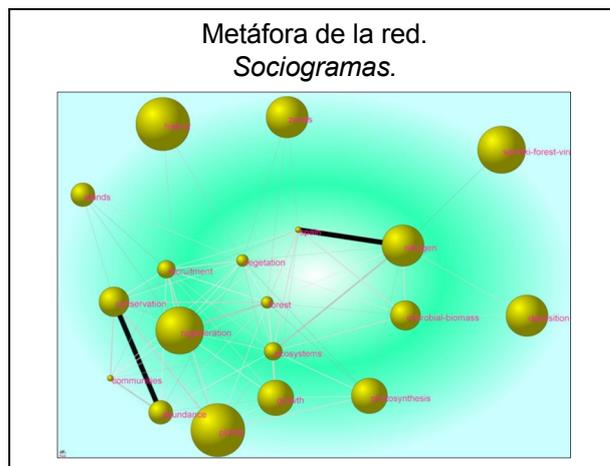


Figura 19: Sociogramas. Reciben el nombre de sociogramas los grafos obtenidos en el análisis de redes sociales. Tomada de “Análisis de redes aplicado a la investigación forestal española” (Pino-Díaz J. , Análisis de redes sociales aplicado a la producción científica forestal española, ISI-1994/2003., 2005)

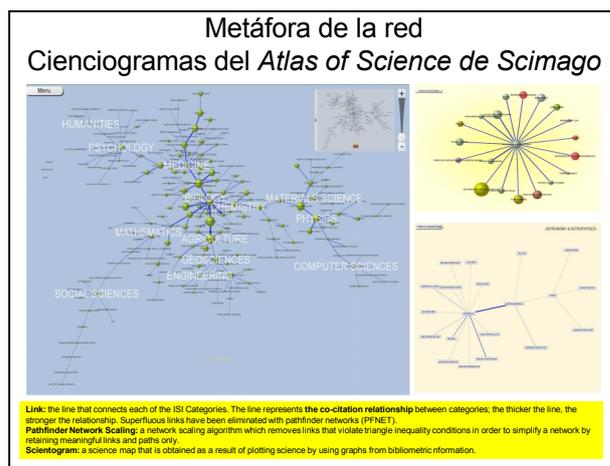


Figura 20: Cienciogramas. Reciben el nombre de cienciogramas los grafos obtenidos en Cienciometría. Fuente Atlas de la Ciencia del grupo de investigación Scimago. Tomada de <http://www.atlasofscience.net/pdf/atlas-of-science-quick-guide-v3.pdf> (Scimago, 2006)

4) Metáfora del paisaje (ver Figuras nº 21, 22 y 23). La metáfora del paisaje es muy útil para representar gran cantidad de datos de forma muy intuitiva ya que las nociones de distancia y altura son fácilmente comprensibles para la mayoría de la gente. Los mapas que utilizan esta metáfora, denominados mapas topográficos o cartográficos codifican varias variables: similitud, densidad, centralidad, etc.

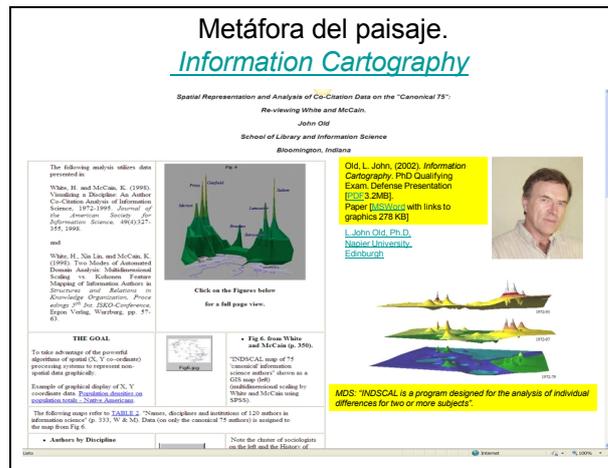


Figura 23: John Old. Emplea su *Information Cartography* (Old J. , 2001), un Sistema de Información Geográfica para obtener Cartografía de la Información. Tomada de http://www.dcs.napier.ac.uk/~cs171/LJOld/InformationCartography/information_cartography.htm

5) Metáforas de la red y del paisaje (ver Figura nº 24):

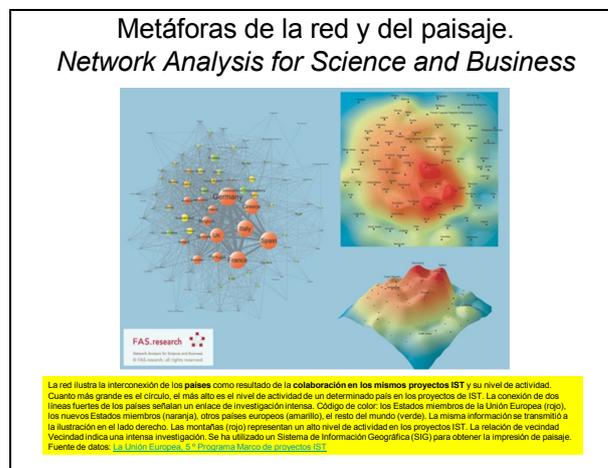


Figura 24: La red ilustra la interconexión de los países como resultado de la colaboración en los mismos proyectos IST y su nivel de actividad. Cuanto más grande es el círculo, el más alto es el nivel de actividad de un determinado país en los proyectos de IST. La conexión de dos líneas fuertes de los países señalan un enlace de investigación intensa. Código de color: los Estados miembros de la Unión Europea (rojo), los nuevos Estados miembros (naranja), otros países europeos (amarillo), el resto del mundo (verde). La misma información se transmitió a la ilustración en el lado derecho. Las montañas (rojo) representan un alto nivel de actividad en los proyectos IST. La relación de vecindad indica una intensa investigación. Se ha utilizado un Sistema de Información Geográfica (SIG) para obtener la impresión de paisaje. Fuente de datos: La Unión Europea, 5º Programa Marco de proyectos IST

datos: La Unión Europea, 5^o Programa Marco de proyectos IST. Tomada de (European Union. Fifth Framework Programme - IST Projects)

4.13. CARTOGRAFÍA DE LA INFORMACIÓN

La **visualización de información textual** mediante gráficos o mapas, de la traducción del término inglés *maps*, es una importante línea de investigación en Ciencimetría. Los métodos seguidos para la realización de estos mapas han sido muy diversos. Una primera aproximación permite clasificarlos en **mapas conceptuales** (*concep map*) y **mapas de dominio** (*domain map*) (ver figuras nº 25, 26 y 27).



Figura 25: Diferentes tipos de mapas conceptuales. Tomada de <http://scimaps.org/maps/browse/>



Figura 26: Diferentes tipos de mapas de dominio. Tomada de <http://scimaps.org/maps/browse/>

La **Cartografía de la Información** en su acepción más pura, es decir, entendida como la realización de mapas cartográficos de información textual es un concepto acuñado por Old (2002). Old emplea un **Sistema de Información Geográfica (SIG)** para visualizar cartográficamente los datos de cocitación de White y McCain. White y McCain realizaron un análisis de **escalamiento multidimensional (MDS)** a la matriz de datos de cocitación de un conjunto documental.

La cartografía espacial tiene una gran utilidad en los estudios del territorio y por ende en los trabajos científicos, de ahí su gran desarrollo (ver Figura nº 27).



Figura 27: Diferentes mapas cartográficos usados en trabajos científicos. Tomada de <http://scimaps.org/maps/browse/>

Fruto de este uso cotidiano, la técnica ha desarrollado softwares que facilitan la realización de mapas cartográficos y el tratamiento de una enorme cantidad de información espacial.

Un **Sistema de Información Geográfica** (SIG o GIS, en su acrónimo inglés) es una integración organizada de hardware, software, datos geográficos y personal, diseñado para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión (Grafo Sistemas, C.A. ¿Qué es SIG/GIS?, 2008) (ver Figura nº 28).



Figura 28: Esquema de un SIG. Imagen tomada de <http://grafosistemas.com/deinteres.html>

También puede definirse como un *modelo de una parte de la realidad referido a un sistema de coordenadas terrestre y construido para satisfacer unas necesidades concretas de información* (ver Figura nº 29).

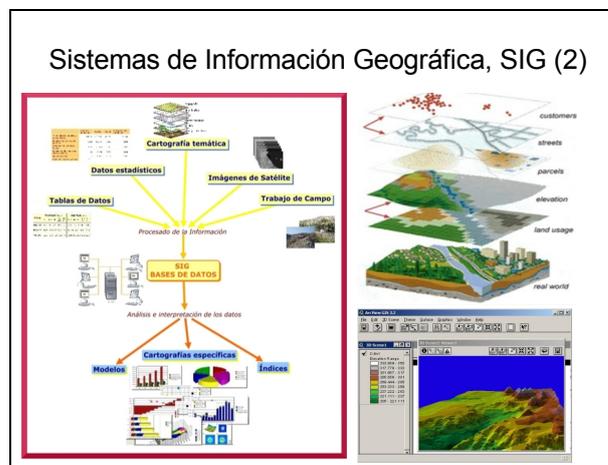


Figura 29: Datos de entrada y resultados que se obtienen en un SIG.

Old en su trabajo *“Information cartography”* emplea el SIG ArcView para realizar un mapa cartográfico a partir de la información espacial de un conjunto documental.

4.14. CARTOGRAFÍA DEL CONOCIMIENTO

Por **Cartografía del Conocimiento** se ha de entender la realización de mapas cartográficos, mediante Sistemas de Información Geográfica, que emplean la metáfora visual del paisaje para la visualización del nuevo conocimiento creado por Sistemas *KDD*. (ver Figura nº 30).

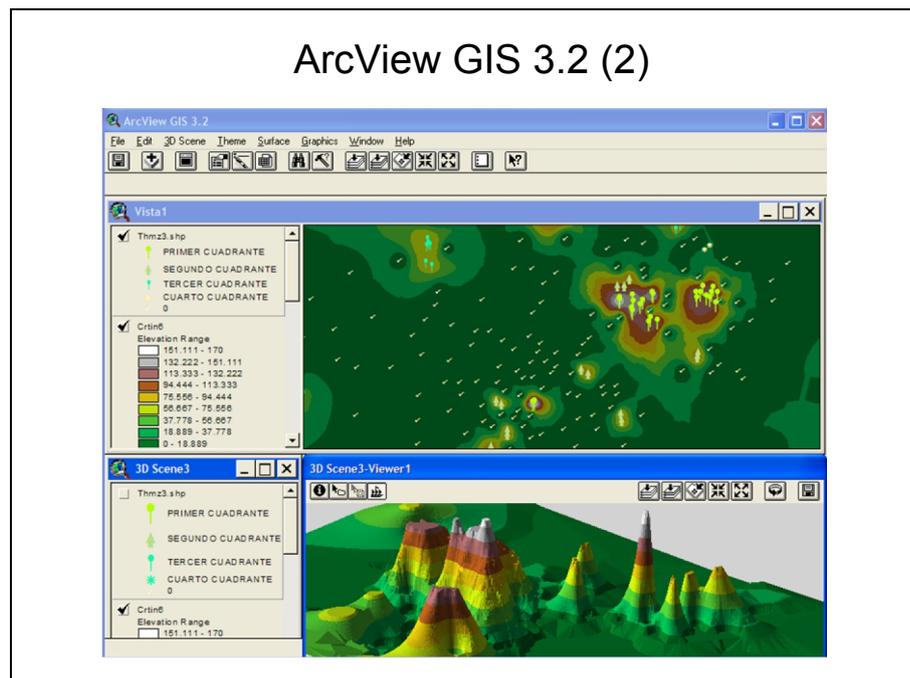


Figura 30: Interfaz del SIG ArcView 3.2 con los mapas en dos, 2D, y tres dimensiones, 3D, del nuevo conocimiento creado a partir del empleo del sistema de conocimiento CoPalRed.

4.15. ÁREAS Y ESPACIOS NATURALES PROTEGIDOS

Internacionalmente se conceptúa **Área Protegida** como el "área definida geográficamente que ha sido designada o regulada y administrada a fin de alcanzar objetivos específicos de conservación" (Convention on Biological Diversity), o bien como el "área de tierra y/o mar especialmente dedicada a la protección y mantenimiento de la diversidad biológica y de los recursos naturales y culturales asociados y gestionada mediante procedimientos legales y otros medios efectivos" (Directrices para la aplicación de las Categorías de Gestión de Áreas Protegidas). En España el concepto área protegida es sinónimo de **Espacio Natural Protegido (ENP)**.

La creación de Áreas Protegidas se han convertido en una herramienta eficaz para la conservación de la naturaleza y prueba de ello es que la superficie continental protegida en Europa superó el 12,70% del territorio. En España la superficie terrestre protegida se acerca a los seis millones de hectáreas, el 11,80% del territorio nacional (de Lucio Fernández, Múgica de la Guerra, Gómez-Limón García, Martínez Alandi, Puertas Blázquez, & Atauri Mezquida, 2007). El número de figuras legales de protección es numeroso y variado, en el caso de nuestro país existen 48 figuras distintas; las principales son Parque Nacional, Parque Natural y Reserva Natural, que suponen el 61,7% de la superficie total protegida y suman 434 de un total de 1.587 Espacios Naturales Protegidos. La mayor parte de la superficie protegida nacional, el 59,1%, se crea bajo las figuras de Parque Nacional y Parque Natural.

Existen diferentes tipos de Áreas Naturales Protegidas (ver Tabla nº 1).

Categoría UICN	Denominación	Descripción
Categoría I	Reserva natural integral	Espacio protegido gestionado principalmente con fines científicos o de protección de la vida silvestre
Categoría II	Parque Nacional	Espacio protegido gestionado principalmente para la protección de ecosistemas y recreo.
Categoría III	Monumento Natural	Espacio protegido gestionado principalmente para la conservación de rasgos naturales específicos.
Categoría IV	Área de gestión de hábitats y especies	Espacio protegido para la conservación mediante intervención gestora
Categoría V	Paisaje protegido terrestre/marino	Espacio protegido gestionado principalmente para la protección del paisaje terrestre/marino y el recreo
Categoría VI	Área protegida de recursos gestionados	Espacio protegido gestionado principalmente para el aprovechamiento sostenible de los recursos naturales

Tabla 1: Categorías UICN de las Áreas Naturales Protegidas.

4.16. INVESTIGACIÓN EN ÁREAS Y ESPACIOS NATURALES PROTEGIDOS

Del 6 al 7 de abril del 2000, organizado por EUROPARC España y la Diputación de Barcelona, se celebró el Seminario "La Investigación y el Seguimiento en los Espacios Naturales Protegidos del siglo XXI". En la monografía que recoge las ponencias presentadas se afirma que "...Una característica común a todos

los espacios protegidos es el difícil acceso a la numerosa y normalmente dispersa información existente..., puede estimarse que más del 50% de los estudios e investigaciones realizados no están publicados.” (Múgica de la Guerra, Gómez-Limón García, & De Lucio Fernández, 2002).

Con carácter general la tipología de los trabajos de investigación sobre ENP es variada, si bien predominan los trabajos de carácter descriptivo: censos de especies, cartografiado de especies y estudios de poblaciones. A nivel general abundan estudios de biología de poblaciones o comunidades y escasean estudios sobre aspectos socioeconómicos.

Un trabajo sobre “La literatura científica mundial sobre Parques Nacionales recogida por la Web of Science (Sci expanded, Ssci y Arts and Humanities Citation Index)” (Pino-Díaz J. , 2004) pone de manifiesto cómo los Parques Nacionales son objeto de estudio por trabajos encuadrables en diversas Disciplinas Científicas. Predominan los estudios de Ciencias de la Vida (Biología, Zoología, Ecología, etc), abundan también los estudios de Ciencias Forestales y de Ciencias de la Tierra (Geología, Hidrología, Limnología, etc.). Además aparecen algunos estudios de temática muy diversa: Contaminación ambiental, Cambio climático, Recreación del visitante, Turismo, Economía ambiental, Desarrollo sostenible, Educación ambiental, etc. Se comprueba, no obstante, que éstos últimos son poco numerosos, especialmente los documentos sobre Planificación, Gestión, Desarrollo sostenible, Uso público (turismo y recreación) y Actividades humanas. Lo anterior es significativo y caracterizador de que la investigación científica actual sobre los Parques Nacionales tiene por objeto el contenido y en raras ocasiones el continente. Entendiendo por contenido, las especies, poblaciones y comunidades animales y vegetales y los procesos naturales que en ellos existen o se dan, y por continente el marco de gestión (planificación, régimen de uso y protección, actividades de gestión etc.).

4.16.1 TESIS DOCTORALES SOBRE ÁREAS PROTEGIDAS

Pino (2004) en un trabajo bibliométrico en el que se analizan las tesis doctorales sobre Sierra Nevada recogidas en la base de datos TESEO comprueba como los grandes campos disciplinarios productores son Ciencias de la Tierra y del Espacio, con un 44% (26 tesis), y Ciencias de la Vida, con un 43% (24 tesis), y como las disciplinas más productivas son Botánica y Geología respectivamente, con 11 tesis cada una.

En otro trabajo posterior del mismo autor (Pino-Díaz J. , 2006), en el que se realiza el **análisis estratégico de las tesis doctorales sobre áreas protegidas leídas en España en el periodo 1979-2004**, se concluye que:

- La red de investigación "Miembros de Tribunal, Disciplinas y Subdisciplinas científicas de las Tesis Doctorales sobre Áreas Protegidas leídas en España (1979-2004)", está bien estructurada, es decir, presenta subredes motor, generalistas, adyacentes y emergentes.
- La Subred motor que destaca por su empuje investigador es "**Etología**" (M. Delibes de Castro, V. Castelló Losada, J.L. Tellería Jorge, L. Arias de Reina Martínez, F. Hiraldo Cano, Florentino de Lope y F. Braza Lloret, Etología, Etología Animal, Comportamiento Animal, Comportamiento de los Primates, Antropología Física, Zoología y Ecología Animal), detrás de la cual se sitúan eminentes científicos de la Estación Biológica de Doñana del CSIC, centro reconocido internacionalmente por su excelencia investigadora, o formados y procedentes de ella. Le acompaña en esta posición de liderazgo, la Subred "**Ciencias Geográficas**" (J. Bosque Sendra, E. Chuvieco Salinero, J. Martínez Vega, M.A. Díaz Muñoz, J.C. García Codrón, M. de Miro Orell, J.M. Rubio Recio, Ciencias Geográficas, Cartografía Geográfica, Geografía Física, Biogeografía, Geografía de los recursos Naturales y Usos del suelo), constituida por científicos especializados en cartografía geográfica y geografía de los recursos naturales.
- Las Subredes troncales de investigación son "**Botánica**", la más relacionada con el resto de subredes, "**Ciencias Forestales**", "**Hidrología**" y "**Química**". Dentro de este grupo también se sitúan "**Geografía**" y "**Geología**".
- Destaca como Subred emergente "**Ciencias Económicas**" debido al interés suscitado los últimos años por conocer los beneficios económicos que las áreas protegidas aportan a la sociedad tan de actualidad por las corrientes de opinión defensoras del desarrollo sostenible (turismo sostenible, actividades tradicionales compatibles con la conservación, conservación de la biodiversidad, etc.)
- Subredes muy bien cohesionadas internamente y que complementan el cuadro estratégico general son "**Ciencias Jurídicas y del Derecho**", "**Miracle Solé M. Rosa**" (parasitólogos), "**Ornitología**", "**Entomología**" y "**Edafología**". La presencia de "**Ciencias Jurídicas y del Derecho**" en el grupo de Subredes especializadas y accesorias a la red principal, está totalmente justificado, pues aún no tratándose de una disciplina experimental ni de estudio del territorio, la protección y gestión de las áreas protegidas se realiza mediante la promulgación de abundante legislación y normativa. El resto (parasitólogos, ornitólogos, entomólogos y edafólogos) se sitúan en este grupo debido a los parámetros constructivos de la red. Todas son ramas de Ciencias de la Vida y de la Tierra y del Espacio: Parasitología Animal, Ornitología y Entomología son Ciencias Zoológicas (Zoología aparece incluida en la subred motor "Etología"), pero quedan fuera de

esta subred por el algoritmo de agrupación sobre centros simples empleado en la formación de subredes, ya que al fijar como parámetro de construcción de las subredes un número máximo de catorce nodos, quedan fuera de la misma y pasan a constituir sus propias subredes.

- TESEO es fiel exponente de la transversalidad del objeto de estudio, Áreas Protegidas. TESEO a pesar de sus carencias (retrasos en la actualización de la base de datos, lagunas en la cobertura, falta de información en algunos campos, etc.) (Torres-Salinas, Delgado López Cózar, Jiménez Contreras, & Ruíz Pérez, 2005) es válida para este tipo de estudios.

De todo lo expuesto se deduce que el estudio de la investigación científica sobre áreas protegidas, tanto de tesis doctorales como de artículos científicos, pone de manifiesto que predominan las disciplinas "naturalísticas" (biología, zoología y geología) y que la variedad, riqueza, relevancia y baja producción de otras disciplinas es notoria y manifiesta.

5. MATERIAL Y MÉTODOS

El dominio documental y los ámbitos temporal y geográfico que abarca esta tesis son:

- Dominio Documental:

Artículos científicos sobre Áreas Naturales Protegidas (Parques Nacionales, Parques Naturales, Reservas Naturales, Áreas Protegidas o Espacios Naturales Protegidos) indexados en las BD del IEDCYT (ISOC, ICYT e IME) y de Thomson-Reuters (Web of Science:SCI-EXPANDED, SSCI e A&HCI)
- Ámbito temporal:

Artículos publicados durante un periodo de 25 años; desde el 01/01/1981 al 31/12/2005.
- Ámbito geográfico:

España como origen de los artículos.

Las diferentes fases de desarrollo de la presente tesis son las siguientes:

1. Gestión bibliográfica de los registros, depuración de los duplicados y de los registros no pertinentes y control de Autoridades.
2. Diseño de las Bases de Conocimiento (BC) de palabras clave (PC), de autores (A) y de revistas (R). Determinación de la ocurrencia mínima y de la coocurrencia mínima de los términos.
3. Diseño del tamaño de las subredes. Determinación del nº mínimo y máximo de términos.
4. Proceso *KDD*. Text Mining. Análisis de Palabras Asociadas en los campos PCAR y A.
5. Tratamiento estadístico de las matrices normalizadas de coocurrencias. Escalamiento Multidimensional (*MDS*). Obtención de los Diagramas *MDS* de las redes.
6. Análisis gráfico de las redes. Aplicación del Algoritmo de Kamada-Kawai y asignación de los índices de asociación a los valores de las líneas.

7. Obtención de las subredes mediante la aplicación del Algoritmo de Agrupación sobre Centros Simples (CoPalRed) y mediante el nuevo método gráfico de eliminación de los enlaces de valor inferior a un valor determinado (el resultado obtenido es similar al conseguido con el Algoritmo de Agrupación por Enlace Simple).
8. Cálculo del centroide de la red. Cálculo de las distancias euclídeas de los términos al centroide. Cálculo de la centralidad nodal de los términos.
9. Análisis estratégico de las redes. Construcción de los Diagramas Estratégicos de las Redes Tecnocientíficas.
10. Cartografía GIS. Construcción de los mapas 2D y 3D.
11. Visualización de mapas 3D en realidad virtual.

El software empleado ha sido:

- Gestión Bibliográfica, **Procite®** .
- Sistema de Conocimiento, **CoPalRed®**.
- Cálculo y Análisis Estadístico, **Microsoft Excel®** y **SPSS®**.
- Análisis y Visualización de Redes, **Pajek®**.
- Cartografía del Conocimiento, **ArcView® GIS 3.2**
- Realidad Virtual, **3D Scene** (herramienta de **3D Analyst**, extensión de **ArcView GIS 3.2**)

5.1. DOMINIO DOCUMENTAL

Con anterioridad se ha puesto de manifiesto que el objeto de estudio “investigación en áreas protegidas” es un concepto transversal u horizontal, en el que han participado o participan disciplinas y subdisciplinas académicas de muchos Campos de la Ciencia.

Los **criterios de selección** de las Bases de Datos Documentales utilizadas en la presente tesis han sido:

- Cobertura geográfica: artículos científicos con origen en España,

- Cobertura temporal: artículos científicos publicados entre 1981 y 2005, ambos inclusive, y
- Multidisciplinaridad: artículos científicos representativos de todos los Campos de la Ciencia.

Las bases de datos seleccionadas han sido:

- Las BD del Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT), **ICYT**, **ISOC** e **IME**,
- La BD de Thomson-Reuters, **Web of Science (SCI expanded, SSCI y Arts and Humanities)**

Todas ellas, cumplen ampliamente los requisitos o criterios de selección (multidisciplinaridad, cobertura temporal y cobertura geográfica), no así en el caso de las bases de datos de la Web of Science para el requisito de amplia representatividad y diversidad disciplinar, ya que es conocido el sesgo de éstas a favor de las Ciencias de la Vida y de las Ciencias de la Tierra y no así para las áreas aplicadas y tecnológicas ni para las Ciencias Sociales ni las Humanidades. No obstante la mayoría de los estudios sobre áreas protegidas proceden de las ciencias experimentales, que están muy bien representadas; además la Web of Science ha sido usualmente empleada en los estudios de Ciencimetría.

En las BD elegidas se han realizado las siguientes estrategias de búsqueda:

- Estrategia de búsqueda en las BD del *IEDCYT*: Artículos publicados entre 1981 y 2005 con origen en España en cuyos campos de título, palabras clave o resumen aparecen los términos: «*parque nacional o parques nacionales o parque natural o parques naturales o reserva natural o reservas naturales o área protegida o áreas protegidas o espacio natural protegido o espacios naturales protegidos*»
- Estrategia de búsqueda en las BD de la Web of Science: Artículos publicados entre 1981 y 2005 con origen en España en cuyos campos de *título, abstract, author-keyword* y *keyword Plus* aparecen los términos: «*natural reserve o natural reserves o natural preserve o natural preserves o national park o national parks o natural park o natural parks o protected area o protected areas o protected natural area o protected natural areas*»

Los **resultados** obtenido han sido:

- Documentos en las BD del IEDCYT:
 - ISOC : 405 docs.
 - ICYT : 604 docs.

- IME : 0 docs.
- Documentos en las BD de la *Web of Science*:
 - SCI-EXPANDED, SSCI e A&HCI: 724 docs.

Una vez realizado el filtrado de los datos obtenidos (eliminación de duplicados y de registros no pertinentes), los resultados se han reducido a:

- BD del IEDCYT: 902 documentos.
- BD de la *Web of Science*: 560 documentos.

Nos encontramos con un **dominio documental** que presenta las siguientes características:

- El dominio de análisis es **horizontal, multidisciplinar, representativo y diverso**.
- El dominio es **amplio**, cubre el periodo de 25 años de la literatura científica española de finales del siglo XX e inicio del siglo XXI.
- El conjunto documental recoge artículos publicados en el ámbito **nacional e internacional**.
- Es un conjunto documental **discreto**, en cuanto al número total de artículos.
- Es un dominio **vivo y creciente** aunque **no expansivo**.

5.2. ÁMBITO TEMPORAL

El presente estudio tiene como antecedente el trabajo de investigación tutelada "**Integración de sistemas de información textuales y espaciales. Análisis Estratégico de la Investigación sobre Áreas Protegidas en España**" (Pino-Díaz J. , 2006). En él se realiza el análisis estratégico de las Tesis Doctorales leídas en España en el periodo 1979-2004.

En el otoño de 2006, tras la obtención del diploma de Estudios Avanzados (DEA) se presentó el Plan de Trabajo de la presente Tesis, que fue aprobado por la Autoridad Académica a comienzos del 2007. Se determinó entonces realizar el análisis de la investigación española en áreas protegidas durante un periodo suficientemente representativo (ver Figura nº 31). Se eligió un periodo de veinticinco años, a contar desde el 31 de diciembre de 2005 hacia

atrás, asumiendo que los artículos fechados en ese año se encontraban en su totalidad indexados en las bases de datos fuente del estudio.



Figura 31: Periodo de estudio de la tesis doctoral y situación respecto a eventos importantes acaecidos en las ANP de España.

Se ha estudiado, por tanto, un periodo de veinticinco años, desde el 1 de enero de 1981 hasta el 31 de diciembre de 2005. A lo largo de este periodo se han considerado en los análisis dinámicos de las redes cinco subperiodos, de cinco años cada uno:

- 1º) Del 01/01/1981 al 31/12/1985
- 2º) Del 01/01/1986 al 31/12/1990
- 3º) Del 01/01/1991 al 31/12/1995
- 4º) Del 01/01/1996 al 31/12/2000
- 5º) Del 01/01/2001 al 31/12/2005

5.3. ÁMBITO GEOGRÁFICO

El dominio documental comprende artículos científicos cuyo origen es España: BD del IEDCYT, artículos de autores en instituciones de España, y BD de la *Web of science*, artículos en los que alguno de los autores tiene dirección institucional en España.

5.4. COPALRED, SISTEMA DE CONOCIMIENTO

Con el Sistema de Conocimiento **CoPalred 1.0** © (Bailón Moreno, 2003) se ha analizado la Base Documental extraída de las Bases de Datos fuente para crear nuevo conocimiento, no explícito, generado a partir de la información contenida en ellas. CoPalRed emplea el método de **Análisis de Palabras Asociadas** y realiza una depuración previa de la información y tres tipos de análisis (ver Figura nº 32):

- 1. Análisis estructural.** Pone de manifiesto la estructura en forma de red del campo científico en estudio, definiendo los actores y sus relaciones.
- 2. Análisis estratégico.** Sitúa a cada actor en una posición relativa dentro de la red, definiéndolo según la intensidad de sus relaciones externas (centralidad) y según su cohesión interna (densidad)
- 3. Análisis dinámico.** Analiza las transformaciones (traducciones- translaciones) de los actores con el tiempo. Identifica acercamientos, bifurcaciones, apariciones y desapariciones de los actores.

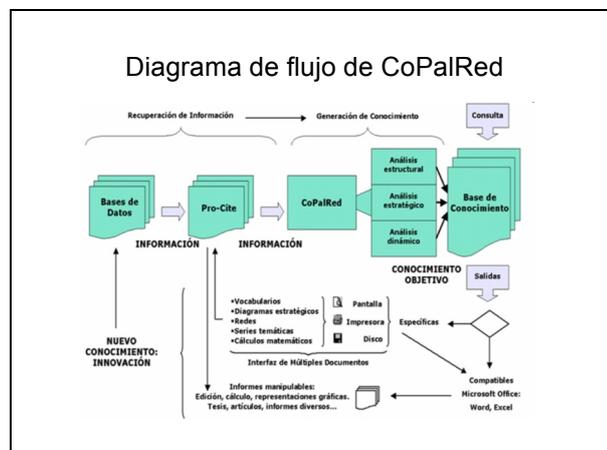


Figura 32: Diagrama de flujo del Sistema de Conocimiento CoPalRed

5.4.1. ANÁLISIS PCAR

Copalred permite la creación de un nuevo campo de registro, el campo **PCAR**. La asociación en un solo campo de estudio de las palabras clave (**PC**), de los autores (**A**) y de la revista de publicación (**R**) permite analizar las

coocurrencias de palabras de campos diferentes de la base de datos; se logra así evidenciar las relaciones **palabras_clave-autores**, **palabras_clave-revista** y **autores-revista**.

El establecimiento del nuevo campo PCAR crea un mayor número de relaciones de coocurrencia entre las palabras (PC, A y R) y esto hace que la red sea más tupida, lo cual mejora el análisis de la misma.

Es muy importante realizar un exhaustivo control de autoridades previo a la creación del campo PCAR. El control de autoridades debe identificar y diferenciar aquellas palabras que se repiten como palabra clave y como título de revista (ej. Ecología, palabra clave, y Ecología(R), revista) y examinar con especial cuidado los identificadores de los investigadores (ej. González, J. y González, M.J.)

Ejemplo del nuevo campo "PCAR" de la BD de la *Web of Science*:

"GORILLA-GORILLA/SOCIAL-STRUCTURE/LIFE-HISTORY-PATTERNS/IMPENETRABLE-NATIONAL-PARK/REPRODUCTIVE-SUCCESS/POPULATION-STRUCTURE/FEMALE-TRANSFER/ECOLOGY/PAN/EVOLUTION/PRIMATES/ROBBINS, M. M./BERMEJO, M./CIPOLLETTA, C./MAGLIOCCA, F./PARNELL, R. J./STOKES, E./AMERICAN-JOURNAL-OF-PRIMATOLOGY"

Ejemplo del nuevo campo "PCAR" de la BD de IEDCYT:

"INSECTA/LEPIDOPTERA/PAPILIONOIDEA/HESPERIOIDEA/FAUNA/RIQUEZA-BIOLÓGICA/FACTOR-ANTRÓPICO/ACLAREO/CONSERVACIÓN-DE-ESPECIES/JIMÉNEZ-VALVERDE, A./MARTÍN CANO, J./MUNGUIRA, M.L./ANIMAL-BIODIVERSITY-AND-CONSERVATION"

5.5. ELECCIÓN DE LA OCURRENCIA Y DE LA COOCURRENCIA MÍNIMAS

Teniendo en consideración el "bajo" número de registros obtenido de las BD fuente (902 documentos de IEDCYT y 560 documentos de la Web of Science) se ha optado por asumir un criterio de amplitud en lugar de un criterio de exhaustividad al elegir el umbral mínimo de ocurrencia. Igual razonamiento se ha considerado en la elección del número mínimo de coocurrencias entre pares de palabras asociadas.

- Nº mínimo de ocurrencias: $c_i = 5$

- Nº mínimo de coocurrencias: $c_{ij} = 3$

5.6. ELECCIÓN DE LOS TAMAÑOS MÍNIMO Y MÁXIMO DE LAS SUBREDES

Para elegir los tamaños mínimo y máximo de las subredes se ha optado por analizar el número de palabras clave de los registros y contrastar los resultados obtenidos con dos trabajos previos, uno sobre palabras clave y descriptores y el otro sobre indexación de documentos en bases de datos.

Para un periodo determinado elegido al azar, se han contado el número de palabras clave por artículo y se ha obtenido la siguiente gráfica (ver Figuras nº 33 y 34):

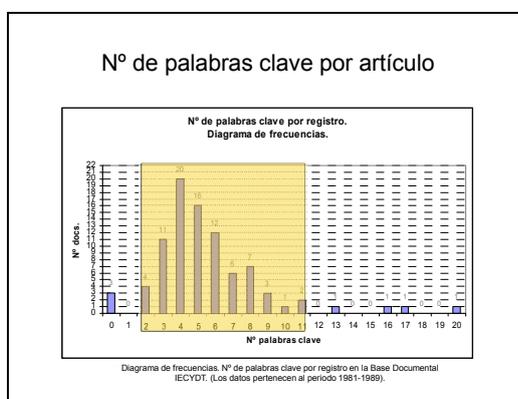


Figura 33: Diagrama de frecuencias del nº de palabras clave por registro en la Base Documental IEDCYT. (Los datos pertenecen al periodo 1981-1989).

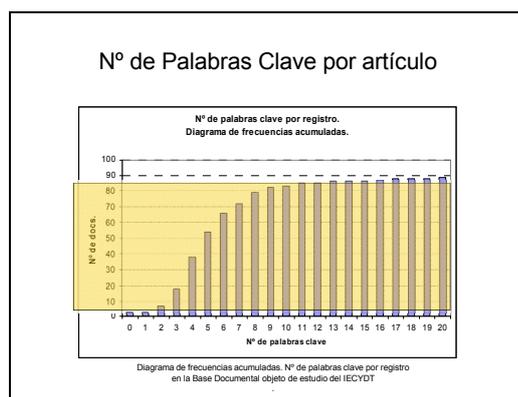


Figura 34: Diagrama de frecuencias acumuladas del nº de palabras clave por registro en la Base Documental IEDCYT. (Los datos pertenecen al periodo 1981-1989).

Del trabajo "La relación entre la palabras clave aportadas por autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT" (Alonso Arroyo & Gil Leiva, 2005), se han obtenido los datos de palabras clave por artículo y se ha elaborado la siguiente figura (ver Figura nº 35):

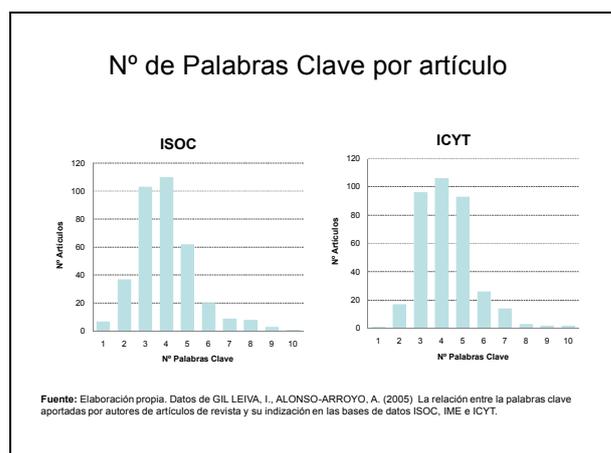


Figura 35: Diagramas de frecuencias del nº de palabras clave por artículos en las bases ISOC e ICYT. Elaboración propia a partir de datos de Gil Leiva y Alonso-Arroyo (2005).

Gil Leiva y Rodríguez Miñoz (1997) en su trabajo sobre el "Análisis de los descriptores de diferentes áreas de conocimiento indizadas en bases de datos de CSIC. Aplicación a la indización automática", obtiene los siguientes resultados (ver Figura nº 36):

Área	Nº mínimo de descriptores	Nº máximo de descriptores
Bib-Doc.	2	10
Medicina	1	6
Química	2	9
Biología	2	9
Psicología	2	12
Física	2	8

Figura 36: Nº mínimo y máximo de descriptores obtenidos por Gil Leiva y Rodríguez Miñoz (1997) en la muestra documental de su trabajo.

Tomando como referencia los datos anteriores se ha optado por fijar:

- Tamaño mínimo de subred: 2 nodos
- Tamaño máximo de subred: 10 nodos

5.7. GRUPOS O SUBREDES ESTRATÉGICAS DE INVESTIGACIÓN

Una vez fijados los tamaños mínimo y máximo de las subredes se ha procedido a realizar el agrupamiento de las palabras. Los grafos de las subredes difieren según el método o el algoritmo de agrupación empleado. En el presente trabajo se han realizado los agrupamientos mediante dos métodos: Algoritmo de Agrupación sobre Centro Simple, utilizado por el Sistema de Conocimiento Copalred, y método gráfico, eliminando los enlaces de valor inferior a un valor dado, con el software de análisis de redes Pajek.

5.7.1. ALGORITMOS DE AGRUPACIÓN

El Sistema de Conocimiento Copalred trabaja con el Algoritmo de Agrupación sobre Centro Simple para realizar las agrupaciones (subredes) de las palabras asociadas. Existe otro algoritmo que permite formar grupos empleando otra técnica diferente a la utilizada por el algoritmo anterior, es el Algoritmo de Agrupación por Enlace Simple.

a) **Algoritmo de clasificación por enlace simple.** *Los elementos de la matriz de asociaciones son ordenados en una lista decreciente según su índice de equivalencia. Esta lista está formada tan solo por aquellas palabras que tengan una ocurrencia mínima y pares de asociaciones también con una co-ocurrencia mínima preestablecidas. El programa recorre la lista desde el principio y va construyendo dobles, tripletes, etc. de palabras asociadas de forma que suministra un grafo conexo que no exceda de un valor máximo de palabras preestablecido (por ejemplo 10 ó 15) Cada vez que se obtiene un grafo, elimina las palabras de éste de la lista y comienza el proceso de construcción de nuevos grafos hasta agotar el total de palabras disponibles. La Figura nº 37 es un ejemplo extraído de (Courtial & Michelet, 1990) que se encuentra en *Cognosfera* (Ruíz-Baños & Bailón-Moreno, Cognosfera, 2005) y que explica los pasos que se siguen usando este algoritmo para la formación de agrupaciones con un máximo de 5 palabras y la reconstrucción posterior de la red.*

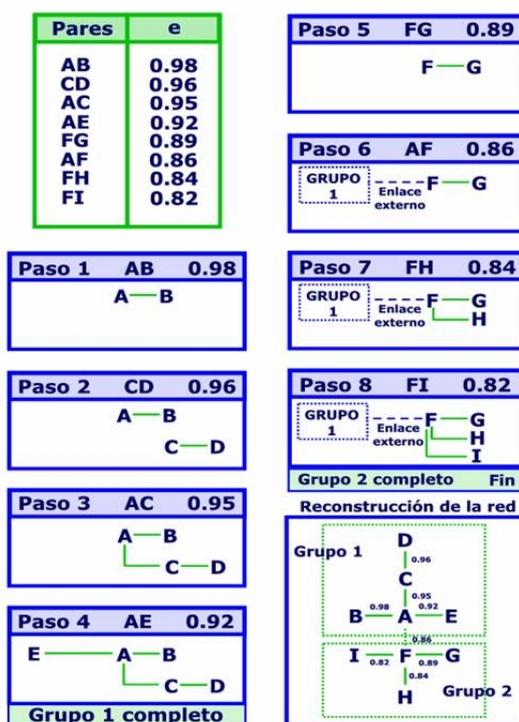


Figura 37: Ejemplo que se encuentra en *Cognosfera* y que explica como actúa el algoritmo de enlaces simples para la formación de agrupaciones con un máximo de 5 palabras y la reconstrucción posterior de la red.

b) **Algoritmo de Agrupación sobre Centros Simples.** Este algoritmo también ordena los pares de asociaciones por orden decreciente de índice de equivalencia y sólo pueden formar parte de esta lista las palabras con una ocurrencia mínima y los pares con una coocurrencia mínima establecidas previamente. El ordenador inicializa un contador para cada descriptor y comienza a recorrer la lista desde el principio incrementando el contador de las palabras que van apareciendo. Cuando el contador de una palabra alcanza un valor igual al número de palabras máximo estipulado para los temas menos uno, el algoritmo toma esta palabra como centro de una agrupación. El conjunto resultante estará formado por las uniones de esta palabra central y todas aquellas que se han asociado con ella. El resultado es una estructura en forma de estrella. Las palabras que han aparecido se eliminan de la lista y se comienza de nuevo el proceso para generar más agrupaciones. Si después de recorrer toda la lista ningún contador llega al valor máximo preestablecido, éste se disminuye en tantas unidades como sea necesario para formar una nueva agrupación. El proceso finaliza cuando el valor máximo del contador disminuya hasta un valor mínimo preestablecido o se terminen todas las

palabras de la lista ordenada de pares (Ruíz-Baños & Bailón-Moreno, Cognosfera, 2005) (ver Figura nº 38).

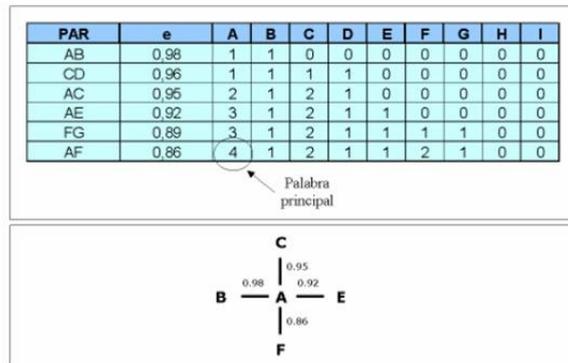


Figura 38: Proceso de agrupación sobre centros simples para el ejemplo de la Figura anterior. Se observa que el resultado es diferente ya que resulta un solo grupo con una estructura en estrella que será identificado mediante la palabra A. Ambos métodos pueden dar resultados algo diferentes a la hora de definir las subredes, pero al reconstruir la red global, el resultado es siempre el mismo.

5.7.2. MÉTODO GRÁFICO DE OBTENCIÓN DE LAS SUBREDES

Este es el segundo método de obtención de las subredes empleado en la presente tesis. Una vez trazado el grafo de la Red con el software de análisis de redes (algoritmo KK y similitudes como valores de las líneas), en primer lugar se eliminan los enlaces inferiores a un valor tal que se obtengan subredes con un número de nodos igual o inferior al límite máximo fijado en el estudio (en este caso, 10 nodos) y en segundo lugar se descompone la Red en subredes, para así obtener los grafos de las mismas por separado (ver Figura nº 39).

- 4) Se realiza el listado de pares de términos y sus valores e_{ij} de índice de equivalencia; los pares del listado deben superar el n° mínimo de coocurrencias fijado, $cmin_{ij} = 3$
- 5) Se prepara el archivo de texto de la Red (.txt) de entrada a Pajek (software de análisis de redes).
- 6) Se dibuja la Red en Pajek (se toman las opciones: "**los valores de las líneas son similitudes**", "**líneas de diferente ancho**") (ver Figura n° 40):
 - Se emplea el **algoritmo de Kamada-Kawai (KK)**
 - Se separan componentes
 - Se estudia cada componente de la red por separado
 - Para cada componente se realizan sucesivas eliminaciones de enlaces por debajo de un valor determinado, hasta obtener los grupos con un n° de nodos igual o inferior al máximo de nodos fijado (10 nodos).

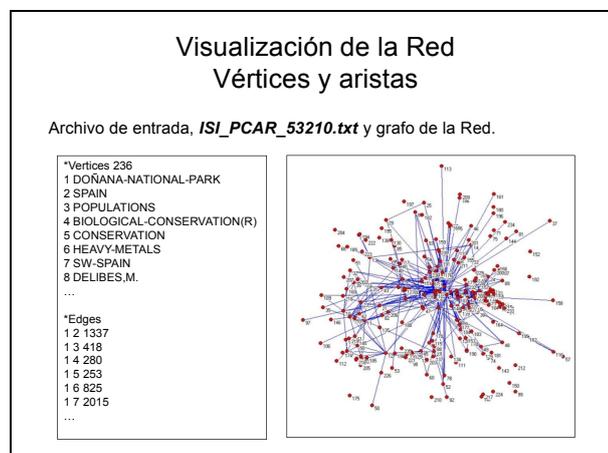


Figura 40: Archivo de entrada a Pajek y grafo obtenido.

Pajek permite obtener como archivo de salida las coordenadas (x, y) de los nodos (ver Figura n° 41)

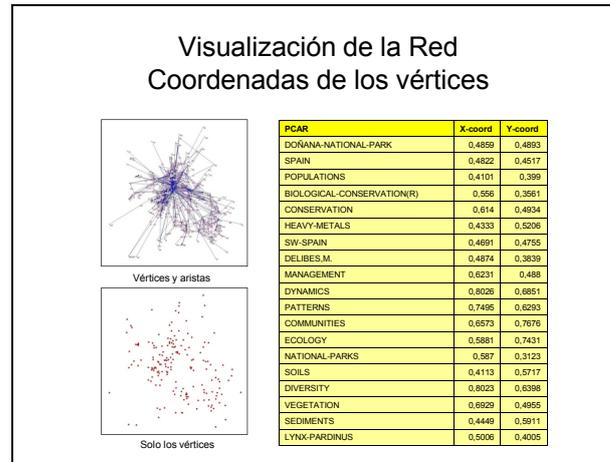


Figura 41: Coordenadas de los nodos o vértices de la Red.

5.8.1. CENTROIDE DE LA RED TECNOCIENTÍFICA KK

Se define el **Centroide de la Red** como el vértice imaginario que se corresponde con el isobaricentro de la Red KK estable de mínima energía.

Las **coordenadas del Centroide** de la Red-KK se obtienen aplicando la fórmula del isobaricentro de la red.

$$x_c = \frac{x_{A_i}}{n} = \overline{x_{A_i}}$$

$$y_c = \frac{y_{A_i}}{n} = \overline{y_{A_i}}$$

Ecuación 22: Coordenadas del Centroide.

Gráficamente se puede representar la posición del centroide (ver Figura nº 22).

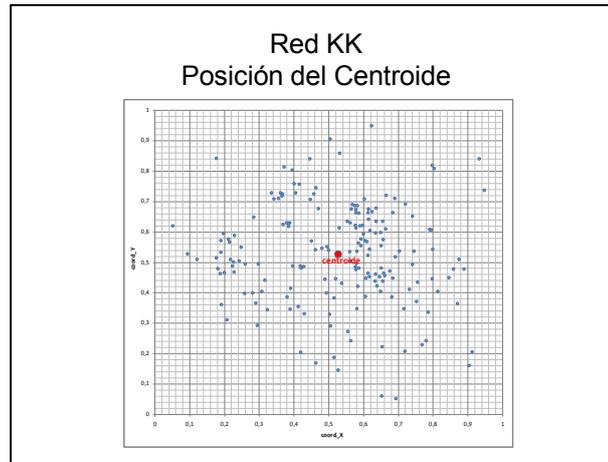


Figura 42: Posición del centroide de la Red KK (solo se han mantenido visibles los nodos de la red)

5.8.2. CENTRALIDAD NODAL

La **centralidad nodal** es un parámetro de gran utilidad en el estudio, descubrimiento y creación de nuevo conocimiento estratégico a partir de conjuntos documentales. Es un concepto innovador, muy cercano al concepto de **centralidad de las subredes** en el Análisis Estratégico de de las redes.

Junto con el parámetro de **densidad de las subredes**, permite realizar los análisis y mapas estratégicos de dichos conjuntos documentales; informes y mapas de gran interés en Vigilancia Estratégica Científica y Tecnológica, Inteligencia Competitiva o Evaluación de la Ciencia.

Su aplicación en Recuperación de Información permitiría clasificar y ordenar los resultados de una búsqueda de acuerdo a su pertinencia o afinidad temática. En este caso la pertinencia o afinidad temática sería directamente proporcional a la centralidad nodal.

5.7.2.1. JUSTIFICACIÓN MATEMÁTICA DEL CONCEPTO DE CENTRALIDAD NODAL

Sea el Grafo **G**:

$$G = \{N, (A, A_k)\}$$

formado por el conjunto de vértices o nodos:

$$\mathbf{N} = \{A_1, A_2, \dots, A_n\}$$

y el conjunto de aristas o pares relacionados $(A_j A_k)$.

Sea el valor de la arista entre dos nodos el valor del enlace de equivalencia de los dos nodos (medida de la similitud entre los nodos);

Sea el Plano Euclídeo E con un sistema de coordenadas cuyo origen es O $(0,0)$ y una abscisa máxima igual a 1 y una ordenada máxima igual a 1.

Sea la función KK (algoritmo de Kamada Kawai) sobre el grafo G , en el plano euclídeo E , que determina la posición de mínima energía de los nodos y de las aristas y permite la asignación a cada nodo de las coordenadas que corresponden a su situación en el plano euclídeo:

$$\text{Vértice o nodo } A_i, \text{ coordenadas } (x_{A_i}, y_{A_i})$$

$$0 < x_{A_i} < 1$$

$$0 < y_{A_i} < 1$$

Se define:

- a) El **Centroide del grafo G en el plano euclídeo E** , es el nodo que tiene por coordenadas, la media de las coordenadas del conjunto de nodos; tal que:

$$\text{Centroide, coordenadas } (x_c, y_c) = \left(\frac{\sum_{i=1}^N x_{A_i}}{N}, \frac{\sum_{i=1}^N y_{A_i}}{N} \right)$$

Ecuación 23: Centroide.

b) **Distancia euclídea entre dos nodos A y B del Grafo G:**

$$d_{\mathbf{E}_{AB}} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

Ecuación 24: Distancia euclídea.

Sea el Espacio Vectorial V formado por el conjunto de vectores posibles entre los vértices o nodos del grafo G:

$$V = \{\overrightarrow{A_1A_2}, \dots, \overrightarrow{A_{n-1}A_n}\}$$

Se define:

c) **Vector distancia nodal media de un nodo**, como el vector medio del conjunto de vectores que tienen por origen el nodo A_1 :

$$\overrightarrow{dV_{A_1A_i}} = \frac{\sum_{i=1}^n \overrightarrow{A_1A_i}}{n}$$

Ecuación 25: Vector distancia nodal media de un nodo.

d) **Vector distancia nodal media del centroide**, como el vector medio del conjunto de vectores que tienen por origen el centroide:

$$\overrightarrow{dV_{CA_i}} = \frac{\sum_{i=1}^n \overrightarrow{CA_i}}{n}$$

Ecuación 26: Vector distancia nodal media del centroide.

Se verifica que:

"El vector distancia nodal medio del nodo A_1 es el vector que lo une al centroide"

$$\overrightarrow{dV_{A_1 A_i}} = \frac{\sum_{i=1}^n \overrightarrow{A_1 A_i}}{n} = \frac{\sum_{i=1}^n (\overrightarrow{A_1 C} - \overrightarrow{CA_i})}{n} = \frac{n\overrightarrow{A_1 C} + \sum_{i=1}^n \overrightarrow{CA_i}}{n} = \frac{n\overrightarrow{A_1 C} + \mathbf{0}}{n} = \frac{n\overrightarrow{A_1 C}}{n} = \overrightarrow{A_1 C}$$

Ecuación 27: Vector nodo-centroide.

Demostración:

1)

$$\text{Vector } \overrightarrow{OC} \text{ coordenadas } (x_C, y_C) = \left(\frac{\sum_{i=1}^N x_{A_i}}{N}, \frac{\sum_{i=1}^N y_{A_i}}{N} \right)$$

$$\text{Vector suma } \sum_{i=1}^n \overrightarrow{OA_i} \text{ coordenadas } \left(\sum_{i=1}^n x_{A_i}, \sum_{i=1}^n y_{A_i} \right)$$

2)

$$\sum_{i=1}^n \overrightarrow{CA_i} = \sum_{i=1}^n (\overrightarrow{OA_i} - \overrightarrow{OC}) = \sum_{i=1}^n \overrightarrow{OA_i} - n\overrightarrow{OC} = 0$$

3)

$$\sum_{i=1}^n \overrightarrow{A_1 A_i} = \sum_{i=1}^n (\overrightarrow{A_1 C} - \overrightarrow{CA_i}) = n\overrightarrow{A_1 C} + \sum_{i=1}^n \overrightarrow{CA_i} = n\overrightarrow{A_1 C} + 0 = n\overrightarrow{A_1 C}$$