Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation

Félix de Moya-Anegón *, Benjamín Vargas-Quesada,
Zaida Chinchilla-Rodriguez, Elena Corera-Álvarez,
Victor Herrero-Solana, Francisco J. Munoz-Fernández

SCImago group, University of Granada, Library and Information Science Faculty, 18071 Granada, Spain

Abstract

We propose the use of *ISI-JCR* categories as units of cocitation and measurement for the construction of *heliocentric maps*. The use of a spatial metaphor allows us to illustrate, analyze and compare domains in terms of the categories and their interconnections or links. We can also move around within the structure of these domains for further analysis, and access the documents associated to the categories and to the links that cocite or relate them.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Information visualization; Domain analysis; Information retrieval; Cocitation

1. Introduction

In Moya Anegón et al. (2004), we reviewed the relevant literature of the past four decades in information visualization and proposed the use of class and subject category cocitation as a technique for the analysis and visualization of great domains. Departing from the same notion, the present paper puts forth the construction of *heliocentric maps* that make manifest the relationships among categories and the flux of information within and among them. Moreover, these maps yield the possibility of showing the documents *hidden* behind each category and the links that unite them.

---

* Corresponding author.
  E-mail address: felix@goliat.ugr.es (F. de Moya-Anegón).

The objective of this paper is to present a methodology for the visual representation and analysis of major scientific domains, defined by Hjorland and Albrechtsen (1995), in terms of the interactions between authors and their role in the scientific world that are reflected by citation. These representations, moreover, can be used as interfaces for information retrieval.

We shall begin by briefly explaining how the information surrounding ISI-JCR (The Thomson Corporation, 2004a) categories can be used as a unit of measurement and cocitation. We then describe the methodology developed for the generation of *heliocentric maps*, which facilitate the analysis of the structure of a given domain and enhance retrieval of documents associated with the central node and the links of each map. Finally, the results and conclusions of our work are presented: we show examples of maps based on the geographic domains of Spain, France and England with respect to the ISI-JCR categories, and comment on their distinctive elements as well as their features in common.

2. Category cocitation

Cocitation is a widely used and generally accepted technique for obtaining relational information about documents belonging to a domain. This relational information can be used to build maps that will represent, with a high degree of fidelity, the structure of the domain that the documents comprise. Because we strive to represent and analyze the structure of large domains, whether they be thematic, geographic or institutional, we propose the use of ISI-JCR cocitation categories as a tool for this purpose.

According to the above scheme, the references to document X manifest the relation of cocitation existing between the referenced documents and between their respective authors. By analogy, we can go one step further to relate the journals where these documents have been published, as well as the subject categories assigned by the ISI-JCR to each one of these journals. In this way, it is evidently easy to translate document cocitation into category cocitation. In fact, higher levels of grouping would also be possible.

3. Source of data

For strictly academic and investigative purposes, we downloaded from the Web of Science (The Thomson Corporation, 2004b)—more specifically from the Science Citation Index-Expanded (SCI-E), Social Science Citation Index (SSCI), and the Arts & Humanities Citation Index (A & HCI)—all those records for the year 2002 whose Address field included "*Spain*" or "*France*" or "*England*". These were introduced into an ad hoc relational database to make the necessary queries for category cocitation. At the time of our consultation, the database contained a total of 159,794 documents (articles, biographical items, book reviews, corrections, editorial materials, letters, meeting abstracts, news items and reviews) from 6584 journals, with the corresponding JCR categories assigned by the ISI itself.

4. Methodology

The adoption of the ISI-JCR classification as the unit of measurement and cocitation implies, first, facing those problems that are intrinsic to the classification itself. As illustrated in Fig. 1, the main idiosyncrasy of the ISI-JCR classification is that it may assign different categories to one single journal in view of the subject matter of the documents they publish, thus producing an error of accumulation in computing cocitation. This is what we will refer to as *latent cocitation*. That is, a single mention of a document published in a journal that happens to be assigned to more than one JCR category will inevitably set off the cocitation of that number of categories. For instance: the journal *Information Processing & Management* (*IPM*) belongs to the categories
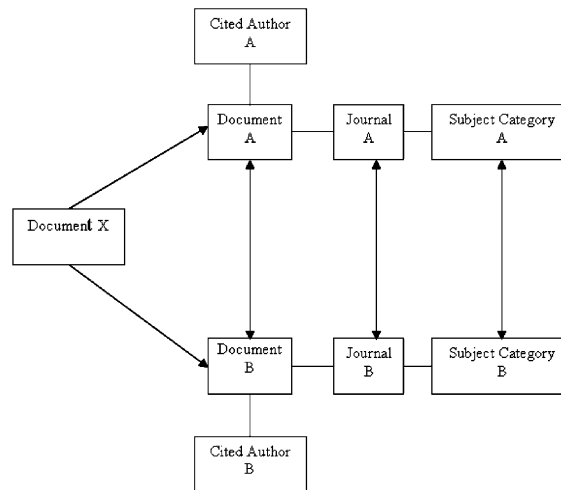
Fig. 1. Cocitation scheme.

Information Science & Library Science, and also to *Computer Science-Information Systems*, and therefore any one reference to a document in this journal will trigger the cocitation of both categories, resulting in an over-estimation of cocitation and implying a margin of error in the representation of the domain structure.

Eliminating this cocitation *latency* is a relatively easy matter. All we need to do is to group the categories cited by each one of the source documents, and calculate cocitation on the basis of that grouping. The result is *non-latent category cocitation*, where the latency inherent to the classification does not condition the representation of the structure of the domain we wish to analyze. This non-latent form of cocitation is the one we will use to generate heliocentric maps; and when we heretofore speak of cocitation, it should be understood in these terms.

Another drawback to bear in mind is the ISI category of *Multidiciplinary Sciences*, which includes journals such as *Science Nature, Endeavor* and *Interciencia*, among others. When for example an article about a specific discipline such as Genetics is published in one of these journals, it is not reflected in the map of its domain, but rather is labeled as "multidisciplinary". For the present study, in order to avoid the ensuing loss of information, we replace the category *Multidisciplinary Sciences* with the category that is most cited by the references of each one of the documents included therein. In those cases where the most cited category was *Multidisciplinary Sciences* itself, we retrieved each source document and assigned it a JCR category manually, on the basis of its title and abstract.

Because the multiple assignment of categories to one single journal introduces some distortion in the calculation of cocitation, we are currently working on a process that will allow us to categorize works individually in a way that is compatible with the ISI-JCR categorization. This would solve the problem of latent cocitation as well as facilitate the categorization of non-multidisciplinary work (which may erroneously be attributed to journals that the ISI-JCR classifies as multidisciplinary). For the time being, however, the intermediate formula described above was the best solution available.

Another obstacle to overcome is that of the normalization of the citation indexes throughout the field of disciplines included in the SCI, SSCI and A & HCI. This matter has already been dealt with by Small and Garfield (1985). In their wake, we standardize category cocitation by dividing the cocitation figure by the square root of the product of the frequency in citing the categories.

$$NCM(ij) = \frac{Cc(ij)}{\sqrt{c(i) \cdot c(j)}}$$

Normalized Cocitation Measurement (Salton & Bergmark, 1979) where Cc is the cocitation and C is the citation. Having cleared up these concepts, we can now focus on the main objective of building a set of *heliocentric maps* for as many categories as those proving productive during the period of study. The specific category to be analyzed will be situated in the center (*helios*) and the others will "revolve" around it, as *planets*, with the greater or lesser distance (*orbit*) in-between determined by the intensity of their relationship (cocitation). Each heliocentric map thus becomes an individualized characteristic representation of research into that specific domain.

4.1. Rendering the information

Due to limitations regarding space (Tufte, 1994; Tufte, 2001), it is practically impossible to clearly represent each one of the *heliocentric categories* with all its potential planets. Yet clarity in the representation is vital, as emphasized by Small (2000), and so the gain in simplicity may be worth the sacrifice of some connections in visualization. Bearing this in mind, we propose the application of a pruning mechanism, involving a variable threshold established by the particulars of the *heliocentric category* and its relationships or *orbits* with respect to the other categories or *planets*. The proposed cutoff value is the mean plus the standard deviation of all the cocitation values greater than zero—already standardized—of the heliocentric category with respect to its planets. The planets with a threshold lower than the established one will be eliminated from the map and do not appear in the final representation. For instance, the heliocentric map shown in Fig. 2 depicts "*Information Science & Library Science*" in Spain, with all the planets that surpass the aforementioned cutoff point. This pruned map does not imply that the planets included are
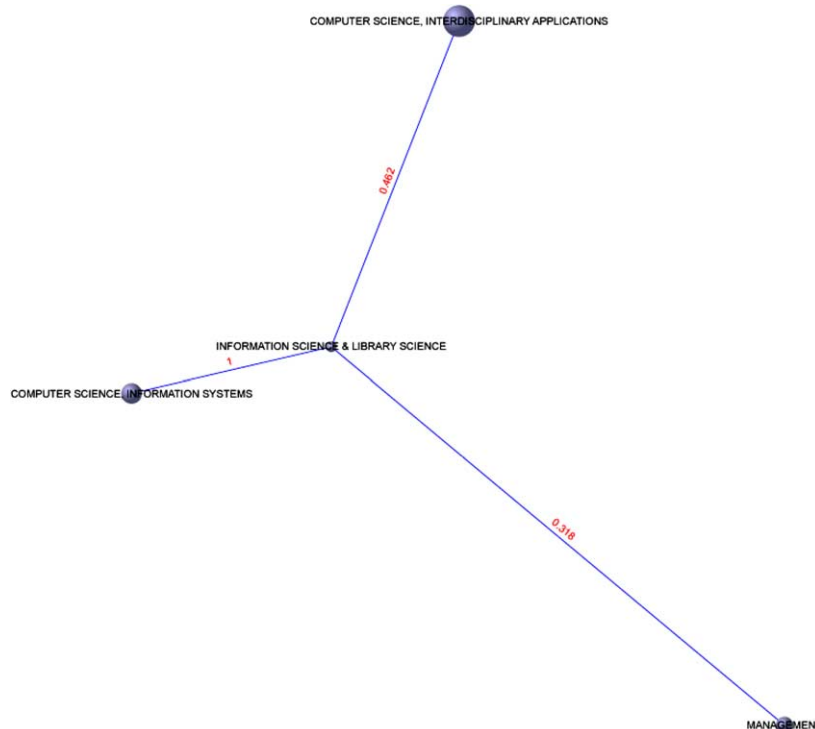

Fig. 2. Heliocentric map of Information Science & Library Science in Spain.

the only ones bearing some relation with the heliocentric category; but rather that they are the most significant ones in that domain, by virtue of their surpassing a very high cocitation threshold.

It is highly probable that, as a consequence of the multidisciplinarity of science, more *planets* exist below the established threshold and some of them may even be significant and important for the analysis of that domain. For example, and to continue with the case of Information Science & Library Science, we arrive at Fig. 3 by applying a cutoff value equal to the mean, whereas Fig. 4 has no cutoff point. This means that all the planets of the heliocentric category appear in Fig. 4. Of course, in both examples, the maps obtained appear somewhat complicated due to the lack of space, a direct consequence of the increased information.

The *heliocentric maps* of Figs. 3 and 4 were selected partly because they are related with the material we work with, but above all because of their limited number of *planets*. Normally, a heliocentric map would have many more planets. We cannot offer a statistical or algorithmic explanation as to why the cutoff point proposed is used instead of some other value, nor defend it as the ideal one. Yet we can say that after carrying out a detailed analysis of the representations and contemplating multiple cutoff points, the mean plus the standard deviation proved to be the threshold guaranteeing the best balance between map clarity, the relevance of information supplied by the domain, and very limited informational noise produced by the appearance of planets with a low degree of cocitation.

In generating these graphs, we used the algorithm of Kamada and Kawai (1989). This algorithm automatically generates non-directed graphs on a plane, guided by esthetic criteria: it minimizes the number of crossed links, reflects the symmetries of the graph, distributes the nodes in a uniform manner over the available space and makes all the links homogeneous with regards to length. Thus, we fix a node in the central position that corresponds with that of each one of the heliocentric categories (the productive ones in the domain in question). Unlike Kamada Kawai on this point, we preferred to interpret the cocitation values of the planets with respect to the central category as similarities, which means that the links become of
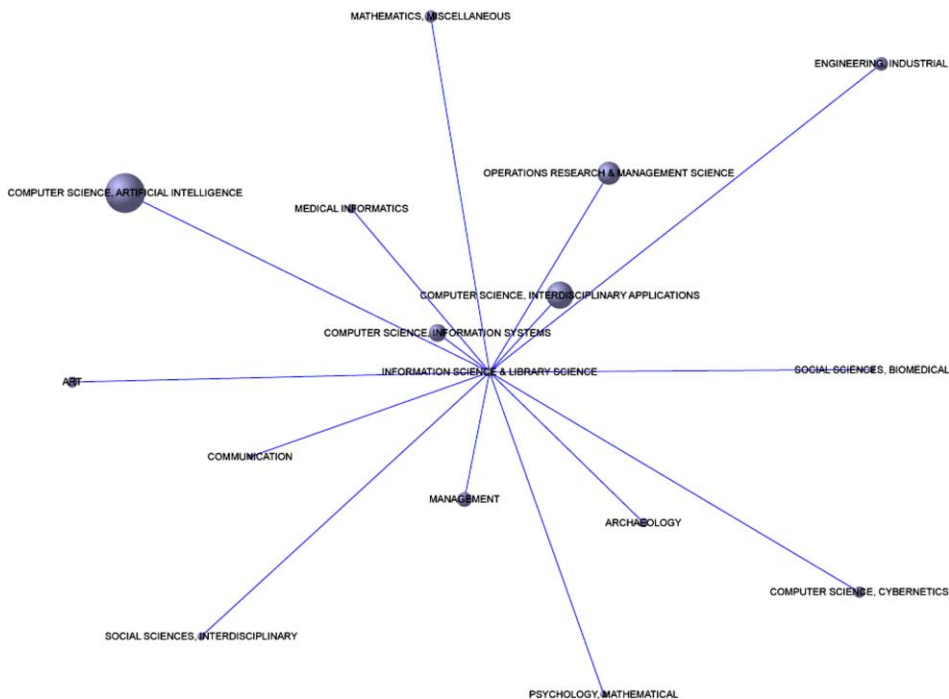

Fig. 3. Heliocentric map of Information Science & Library Science in Spain with a threshold value equal to the mean.

Fig. 4. Heliocentric map of Information Science & Library Science in Spain with no cutoff point.

variable length. To emphasize the distance among planets, a maximum value for cocitation is established as 1, and the rest of the values are made proportional with reference to this maximum. In this way we can see at first glance which planets are the closest ones to the heliocentric category, and are therefore most closely related with it.

The resulting map is exported to *Scalable Vector Graphic* (*SVG*) format (W3C, 2004), which allows us to zoom in or move vertically or horizontally over the maps. In turn, the code is subjected to a series of modifications. In the first place, the nodes of each map are tagged with the names corresponding to each one of the *ISI-JCR* categories. Then, for each map, the size of these categories is made proportional to the number of documents produced in them. In this way categories with only minor scientific production are made perfectly visible. Third, the hyperlinks needed in the links and in the central category are inserted to allow the retrieval of information associated with them. At the same time, we insert the hyperlinks needed in the rest of the categories to facilitate navigating through the structure of the domain.

In the case of our study, representations in two dimensions are used, but there is no reason why not to use three dimensions.

4.2. Information retrieval

As mentioned in the methodology section, each heliocentric map includes—in the helios and in the links with its planets—hyperlinks that make it possible for us to click into a relational database built *ad hoc*. This database would gather up the bibliographic information of the documents dealt with.

There are two means of retrieving and accessing this information. The first is tied to the *heliocentric category* itself. A simple query, for each map, shows the works pertaining to that category ordered according

to document type. The second would be an ordering of the documents in view of the orbits existing between the heliocentric category and its planets by relevance of cocitation.

The ordering by relevance of cocitation is achieved by combining the three queries into one main query, to determine the relevance of the documents and therefore their order as well. The first-degree query of relevance shows, for each orbit, the works cocited by the heliocentric category and their corresponding planet, but pertaining necessarily to that heliocentric category. According to our criteria, these works are the most relevant ones, as they reflect contributions made from the very *heliocentric category* itself, in tight relation with each of the *planets* or disciplines that orbit around it. This serves to highlight the coinciding points between disciplines or research fronts from the point of view of the heliocentric category. The second-degree query, then, shows the works that, while cocited by the heliocentric category and its corresponding planet, appear published in the planet category, thus indicating documents that exhibit confluence between categories or research fronts originating in the planet. Third and finally, the works cocited by categories that do not belong to either the heliocentric category nor to its planets mainly represent sporadic or tangential relationships between disciplines from a totally separate standpoint. In order to set off the different degrees of relevance of the works retrieved here, the output of each of the three types of queries is shown with a different colored background.

5. Results

To facilitate the understanding of results for the reader, in the first place we give a general analysis of the Spanish domain, using as an example several *heliocentric maps* of that domain. Then we compare the domains of Spain, France and England, also on a general level, by looking at some of the more characteristic or unusual heliocentric maps produced.

5.1. Analysis of a domain

Fig. 2 shows a snapshot of the disciplines that make up the main lines of research in Information Science & Library Science in Spain for the year 2002: *Information Systems and Computer Science*, *Management*, *Computer Science-Information Systems* and *Computer Science-Interdisciplinary Application*. Obviously, in this snapshot map the heliocentric category or object of study is Information Science & Library Science, and the planets with a greater level of cocitation and nearby orbits are, in the following order, *Computer Science-Information Systems*, *Computer Science-Interdisciplinary Applications* and *Management*. (The numbers appearing next to each link indicate the degree of standardized cocitation of each planet with respect to the heliocentric category.) We can also easily see from the size of the heliocentric category and its different planets that the category with the greatest proportion of output is *Computer Science-Interdisciplinary Applications*. It is followed by *Computer Science-Information Systems* and *Management*, whereas the lowest level of production corresponds to *Information Science & Library Science*.

Direct access to the information contained in each heliocentric category and in its orbits allows us to observe, for example in Fig. 5, a snapshot of the top twenty references obtained upon activating the query of the heliocentric category of Fig. 2.

In keeping with the example of Information Science & Library Science, Fig. 6 provides another *snapshot* of the top-cocited 48 documents published in 2002, obtained by activating the orbit between that heliocentric category and *Computer Sciences-Information Systems*, ordered according to the relevance of cocitation.

Finally, and to round out the Library & Information Science perspective of the examples, the graph of Fig. 7 shows the map accessed by activating the hyperlink of the category *Computer Science-Information Systems* of the map in Fig. 2, thus calling up a new heliocentric map that likewise enables the user to navigate and retrieve information.

SPAIN - INFORMATION SCIENCE & LIBRARY SCIENCE - DOCUMENT LIST

Title	Journal	Year	N°	Vol	Pag	Type of doc.
Document organization using Kohonen's algorithm	INFORM PROCESS MANAG	2002	38	1	79-89	Article
An algorithm for term conflation based on tree structures	J AM SOC INF SCI TEC	2002	53	3	199-208	Article
A context vector model for information retrieval	J AM SOC INF SCI TEC	2002	53	3	236-249	Article
Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country	SCIENTOMETRICS	2002	53	2	195-206	Article
Multivariate evaluation of Spanish educational research journals	SCIENTOMETRICS	2002	55	1	87-102	Article
Research productivity of scientists in consolidated vs. non-consolidated teams: The case of Spanish university geologists	SCIENTOMETRICS	2002	55	1	137-156	Article
Methods for the analysis of the uses of scientific information: The case of the University of Extremadura (1996-7)	LIBRI	2002	52	2	99-109	Article
Creating e-books in a distributed and collaborative way	ELECTRON LIBR	2002	20	4	288-295	Article
Participative knowledge production of learning objects for e-books	ELECTRON LIBR	2002	20	4	296-305	Article
Obtaining feedback for indexing from highlighted text	ELECTRON LIBR	2002	20	4	306-313	Article
A system to generate electronic books on programming exercises	ELECTRON LIBR	2002	20	4	314-321	Article
EIS evolution in large Spanish businesses	INFORM MANAGE-AMSTER	2002	40	1	41-50	Article
A test of genetic algorithms in relevance feedback	INFORM PROCESS MANAG	2002	38	6	793-805	Article
Automatic extraction of relationships between terms by means of Kohonen's algorithm	LIBR INFORM SCI RES	2002	24	3	235-250	Article
The effect of team consolidation on research collaboration and performance of scientists. Case study of Spanish university researchers in Geology	SCIENTOMETRICS	2002	55	3	377-394	Article
Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies	J MED LIBR ASSOC	2002	90	4	411-430	Article
Vertical integration of sciences: an approach to a different view of knowledge organization	J INFORM SCI	2002	28	5	397-405	Article
Geographic information systems for science and technology indicators	RES EVALUAT	2002	11	3	141-148	Article
Assessing quality of domestic, scientific journals in geographically oriented disciplines: scientists' judgements versus citations	RES EVALUAT	2002	11	3	149-154	Article
Strategies and models for teaching how to use information: Guide for teachers, librarians and archivists	J ACAD LIBR	2002	28	1-2	77-78	Book Review

Fig. 5. Spanish documents under the category Library Science and Information Science, year 2002.

In our opinion, the map of *Information Sciences* & *Library Sciences* proves logical and realistic, reflecting the actual state of research in Spain. This is confirmed by the existence of the published works obtained from the orbit query. Furthermore, this opinion was backed up by a number of Spanish experts in Computational Sciences whom we consulted for their assessment of the heliocentric maps created.

5.2. Comparison of domains

We see that both Spain and France produced scientific works in 216 categories, while England shows output in two more categories, for a total of 218. The differences, therefore, are minimal and are shown in Table 1.

Generally speaking, the three geographic domains we look at here are quite similar, with equally well developed socioeconomic and technological foundations. France and England share G8 status (the group of the eight most industrialized countries of the world). France and Spain, on the other hand, share a Mediterranean culture that distinguishes them from the Anglo Saxon background in many respects. In other words, a domain structure is bound to reflect certain differences insofar as levels of development, economics, and scientific policy. And so it is.

Only two heliocentric categories or disciplines—*Astronomy & Astrophysics*, and *Physics-Particles & Fields* (Figs. 8 and 9)—fully coincide with regard to the situations of their planets in all three geographical

SPAIN - INFORMATION SCIENCE & LIBRARY SCIENCE - COMPUTER SCIENCE, INFORMATION SYSTEMS

Title	Journal	Year	Vol.	Nº	Pages	Doc. type
Participative knowledge production of learning objects for e-books	ELECTRON LIBR	2002	20	4	296-305	Article
EIS evolution in large Spanish businesses	INFORM MANAGE-AMSTER	2002	40	1	41-50	Article
Document organization using Kohonen's algorithm	INFORM PROCESS MANAG	2002	38	1	79-89	Article
A test of genetic algorithms in relevance feedback	INFORM PROCESS MANAG	2002	38	6	793-805	Article
A context vector model for information retrieval	J AM SOC INF SCI TEC	2002	53	3	236-249	Article
Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies	J MED LIBR ASSOC	2002	90	4	411-430	Article
Automatic extraction of relationships between terms by means of Kohonen's algorithm	LIBR INFORM SCI RES	2002	24	3	235-250	Article
Methods for the analysis of the uses of scientific information: The case of the University of Extremadura (1996-7)	LIBRI	2002	52	2	99-109	Article
Assessing quality of domestic, scientific journals in geographically oriented disciplines: scientists' judgements versus citations	RES EVALUAT	2002	11	3	149-154	Article
Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country	SCIENTOMETRICS	2002	53	2	195-206	Article
Multivariate evaluation of Spanish educational research journals	SCIENTOMETRICS	2002	55	1	87-102	Article
The effect of team consolidation on research collaboration and performance of scientists. Case study of Spanish university researchers in Geology	SCIENTOMETRICS	2002	55	3	377-394	Article
Automated code generation of dynamic specializations: an approach based on design patterns and formal techniques	DATA KNOWL ENG	2002	40	3	315-353	Article
Dual grid: A new approach for robust spatial algebra implementation	GEOINFORMATICA	2002	6	1	57-76	Article
The power of a pebble: Exploring and mapping directed graphs	INFORM COMPUT	2002	176	1	1-21	Article
Context-sensitive rewriting strategies	INFORM COMPUT	2002	178	1	294-343	Article
Practical algorithms for deciding path ordering constraint satisfaction	INFORM COMPUT	2002	178	2	422-440	Article
A fully syntactic AC-RPO	INFORM COMPUT	2002	178	2	515-533	Article
Parallel evolutionary algorithms can achieve super-linear performance	INFORM PROCESS LETT	2002	82	1	7-13	Article
Lower bounds on the information rate of secret sharing schemes with homogeneous access structure	INFORM PROCESS LETT	2002	83	6	345-351	Article
Bases for the development of LAST: a formal method for business software requirements specification	INFORM SOFTWARE TECH	2002	44	2	65-75	Article
Derived types and taxonomic constraints in conceptual modeling	INFORM SYST	2002	27	6	391-409	Article
Internet usage and competitive advantage: the impact of the Internet on an old economy industry in Spain	INTERNET RES	2002	12	5	391-401	Article
Fuzzy ARTMAP and back-propagation neural networks based quantitative structure-property relationships (QSPRs) for octanol-water partition coefficient of organic compounds	J CHEM INF COMP SCI	2002	42	2	162-183	Article
Using molecular quantum similarity measures under stochastic transformation to describe physical properties of molecular systems	J CHEM INF COMP SCI	2002	42	2	317-325	Article
An integrated SOM-fuzzy ARTMAP neural system for the evaluation of toxicity	J CHEM INF COMP SCI	2002	42	2	343-359	Article
Modeling large macromolecular structures using promolecular densities	J CHEM INF COMP SCI	2002	42	4	847-852	Article
Polarizabilities of solvents from the chemical composition	J CHEM INF COMP SCI	2002	42	5	1154-1163	Article
Molecular Quantum Similarity-based QSARs for binding affinities of several steroid sets	J CHEM INF COMP SCI	2002	42	5	1185-1193	Article
In silico studies toward the discovery of new anti-HIV nucleoside compounds with the use of TOPS-MODE and 2D/3D connectivity indices. 1. Pyrimidyl derivatives	J CHEM INF COMP SCI	2002	42	5	1194-1203	Article
Parallel algorithms for graph cycle extraction using the cyclical conjunction operator	J CHEM INF COMP SCI	2002	42	6	1398-1406	Article
Cyclical conjunction: An efficient operator for the extraction of cycles from a graph	J CHEM INF COMP SCI	2002	42	6	1415-1424	Article
Integrating genomics into health information systems	METHOD INFORM MED	2002	41	1	25-30	Article
Space-Time receivers for GSM radio interfaces in subway tunnel environments	WIREL COMMUN MOB COM	2002	2	7	719-733	Article
Bases for the development of LAST: a formal method for business software requirements specification (vol 44, pg 65, 2002)	INFORM SOFTWARE TECH	2002	44	6	393-393	Correction
On binary 1-perfect additive codes: Some structural properties	IEEE T INFORM THEORY	2002	48	9	2587-2592	Letter
Repairing syntax errors in LR parsers	ACM T PROGR LANG SYS	2002	24	6	698-710	Article
Bit-parallel arithmetic implementations over finite fields GF(2(m)) with reconfigurable hardware	ACTA APPL MATH	2002	73	3	337-356	Article
Determination of topological indexes with highest influence on the molecular-structure of saturated acyclic hydrocarbons	AFINIDAD	2002	59	501	504-515	Article
Splitting a Delaunay triangulation in linear time	ALGORITHMICA	2002	34	1	39-46	Article
Graded similarity-based semantics for nonmonotonic inferences	ANN MATH ARTIF INTEL	2002	34	1-3	89-105	Article
Isomorphism classes of genus-2 hyperelliptic curves over finite fields	APPL ALGEBR ENG COMM	2002	13	1	57-65	Article
Graded codes	APPL ALGEBR ENG COMM	2002	13	2	107-120	Article
Goodness-of-fit tests based on Rao's divergence under sparseness assumptions	APPL MATH COMPUT	2002	130	2-3	265-283	Article
Luminescence properties of thin films prepared by laser ablation of Nd-doped potassium gadolinium tungstate	APPL PHYS A-MATER	2002	74	1	109-113	Article
Composition and optical properties of thin films prepared by laser ablation of Nd : KGW	APPL SURF SCI	2002	186	1-4	469-473	Article
Using similarity criteria to make issue trade-offs in automated negotiations	ARTIF INTELL	2002	142	2	205-237	Article
Empirical evaluation of a hybrid intelligent monitoring system using different measures of	ARTIF INTELL MED	2002	24	1	71-96	Article

Fig. 6. Documents associated with the link between Library Science and Information Science and Computer Science & Information Systems, year 2002.
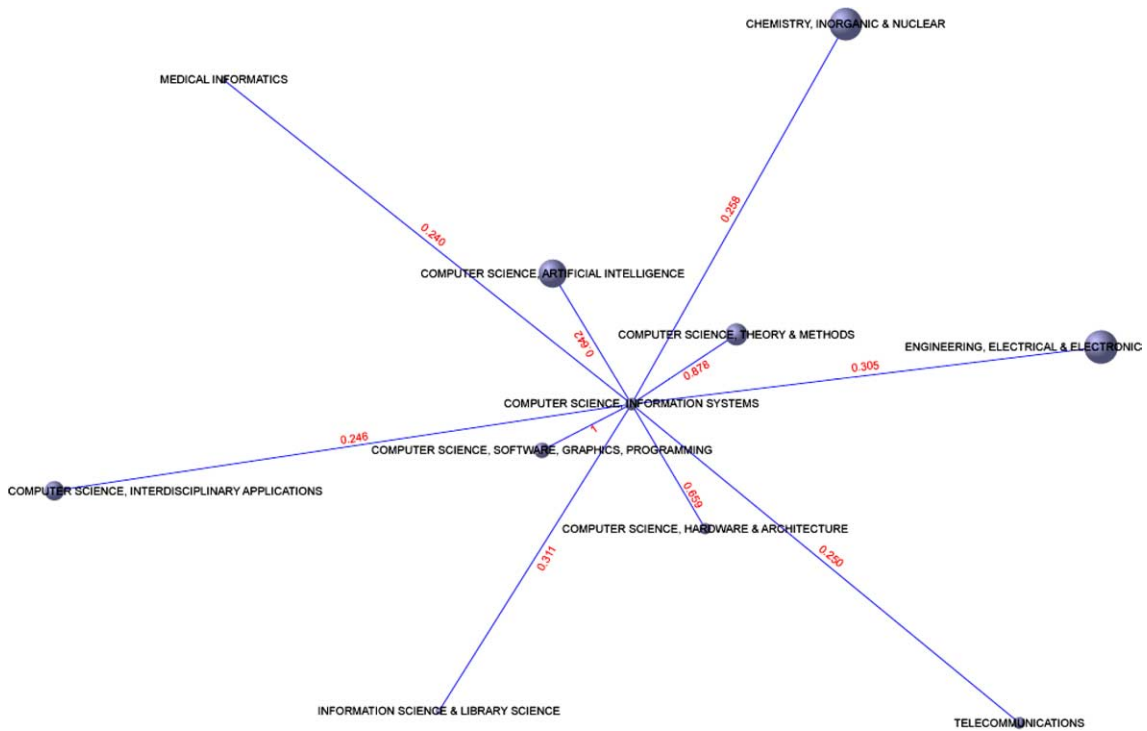
Fig. 7. Heliocentric map of Computer Science-Information Systems in Spain.

Table 1
Categories for which output does not exist in all three countries

Spain	France	England
	Literature, slavic	Literature, slavic
Nursing		Nursing
		Criminology & penology

domains, though the orbits are slightly different. This coincidence indicates that these two disciplines are well consolidated, with very stable research fronts that highly resemble each other. If we wish, however, to detect differences among the three different domains, all we have to do is reduce the threshold or cutoff value of cocitation during the map construction stage. This will give us representations with a greater amount of information, though they will also be more difficult to visualize.

Such highly coincidental cases are surprising, as one would naturally assume a certain degree of similarity among the planets appearing in each map and in their orbits, but not to such a noteworthy extent from country to country. The heliocentric maps of *Psychology*, *Sport Science*, and *Tropical Medicine* (respectively Figs. 10–12) exhibit an intermediate degree of similarity. We can also find cases at the other extreme, such as *Law* (Fig. 13), where practically all the planets differ substantially from one country to the next. This suggests a certain lack of representativity of the information contained in the *ISI* databases in the case of the social sciences and humanities, including *Law*.

We should underline that the comparison of domains will reveal the coincidence of the planets and identify common research fronts for the three countries studied. In turn, the length of the orbits with respect to
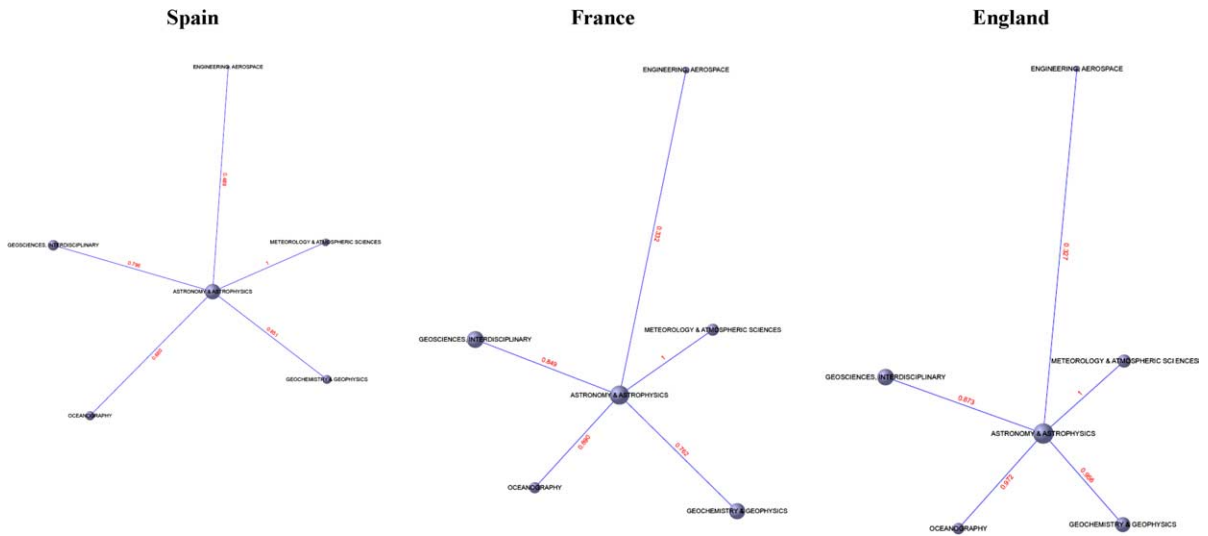
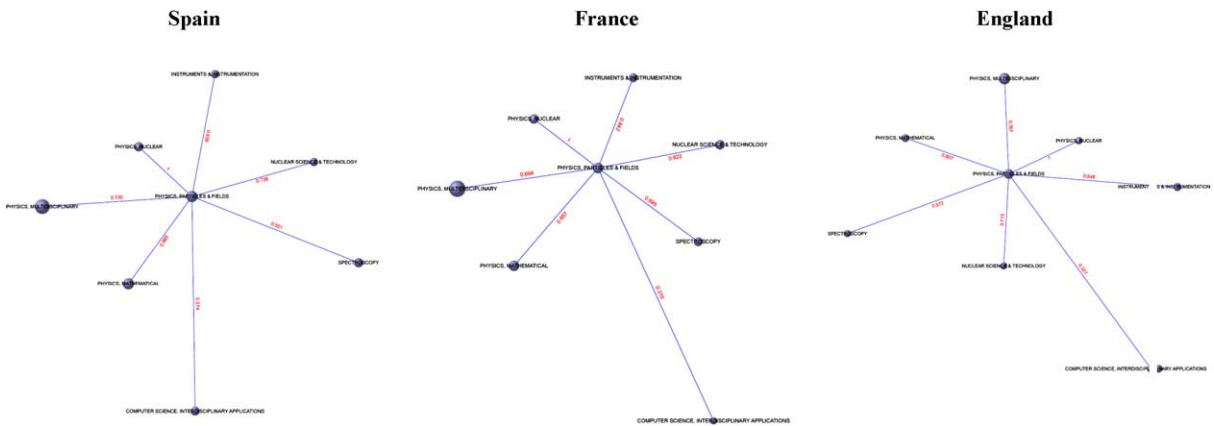Fig. 8. Heliocentric maps of Astronomy & Astrophysics.

Fig. 9. Heliocentric maps of Physics-Particles & Fields.

the heliocentric category indicates their degree of relevance, that is, the amount of interchange of documents between the central category and its planets. We should also consider the degree of each heliocentric category from the perspective of analysis of social networks (Wasserman & Faust, 1998), which reflects the degree of multidisciplinarity of each heliocentric category or discipline in each domain.

We shall now compare the three domains for the category Sport Science from Fig. 11. The decreasing proximity of each planet with respect to the heliocentric category is shown in Table 2. Five planets constitute the nucleus or common research front in Sport Sciences in all three countries: *Rehabilitation*, *Physiology*, *Orthopedics*, *Applied Psychology* and *Respiratory System*. The planets with research fronts that vary from one country to the next are indicated by dark shading. Others are shared by two out of the three and are lightly shaded. As far as nodes are concerned, Spain has 7, France 8, and England has a total of 10, suggesting a greater level of multidisciplinarity in the case of English Sport Science. Under similar
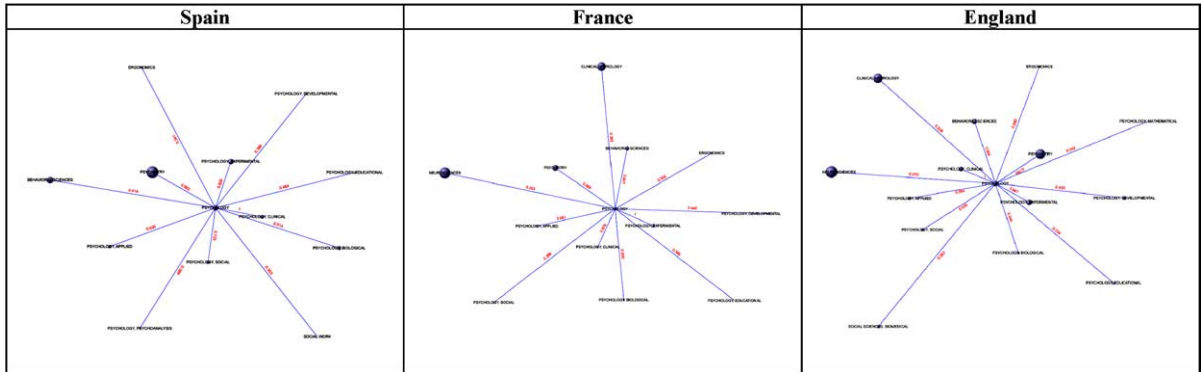
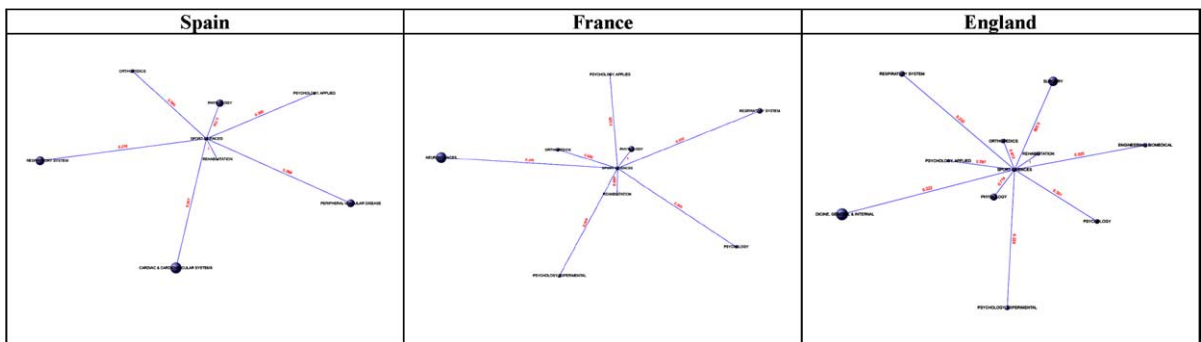Fig. 10. Heliocentric maps of Psychology.

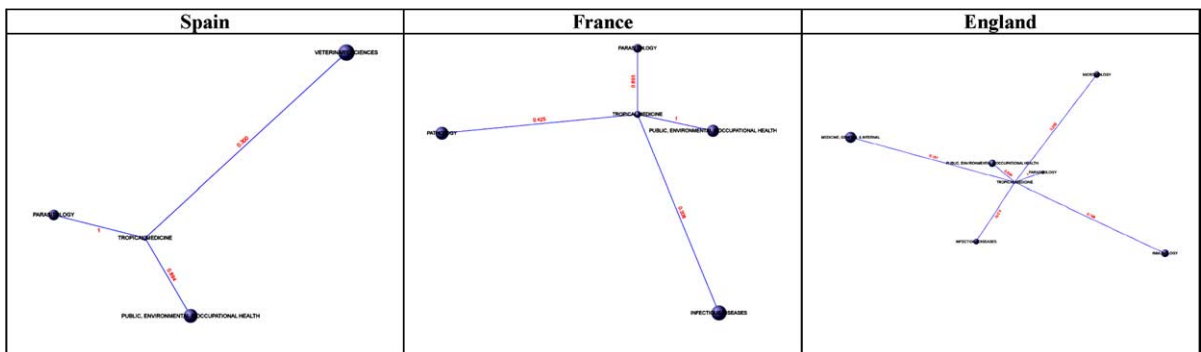

Fig. 11. Heliocentric maps of Sport Sciences.

Fig. 12. Heliocentric maps of Tropical Medicine.

circumstances and during the same time period, England has a slightly wider range of research fronts than France or Spain.

In view of the results published here, and with the understanding that the degree of heliocentricity of each domain is directly proportional to its level of multidisciplinary when a comparison is made of the
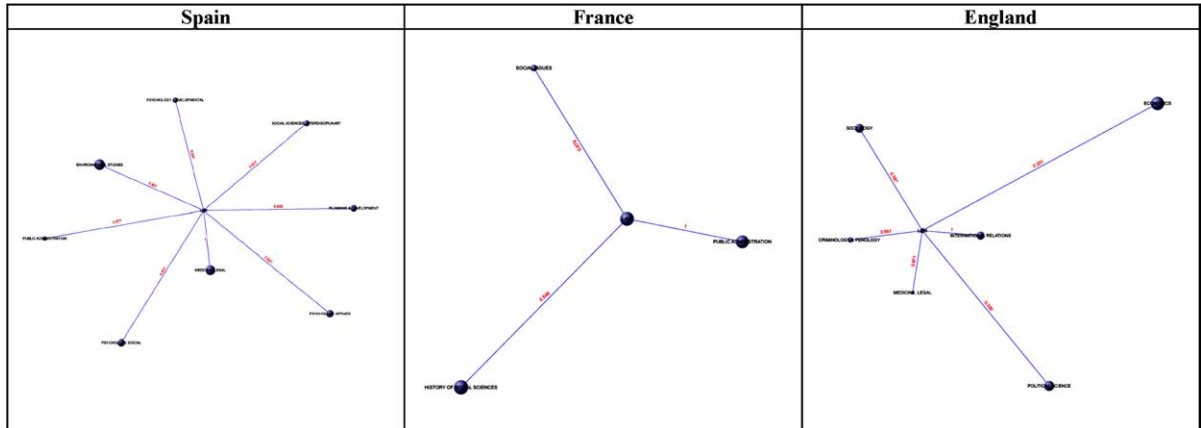
Fig. 13. Heliocentric maps of Law.

Table 2
Sport Science planets ordered by closeness

Spain	France	England
Rehabilitation	Physiology	Rehabilitation
Physiology	Rehabilitation	Orthopedics
Orthopedics	Orthopedics	Physiology
Psychology Applied	Psychology Applied	Psychology Applied
Peripheral Vascular Disease	Psychology Experimental	Psychology
Respiratory System	Psychology	Surgery
Cardiac & Cardiovascular System	Respiratory System	Engineering Biomedical
	Neuro sciences	Respiratory System
		Psychology Experimental
		Medicine General & Internal

same category in different countries, we may assert that the category *Sports* Science is more multidisciplinary in England than it is in France or Spain. The same can be said of *Tropical Medicine*, and of *Psychology*. Further consideration of our maps (not all included here) reveals that England is the country with more multidisciplinary categories as well. What is most relevant in our opinion, however, is not who "wins" in multidisciplinarity, but simply the fact that this method facilitates our making such observations within or among different domains.

6. Conclusions

We are well aware of the fact that our reliance on the *ISI-JCR* classification as an element of cocitation entails some bias and limitations. We are willing to acknowledge the possible claims that this classification is arbitrary, ever-changing and beyond our control. Granted, if we had used some other taxonomy, the representations obtained would not be exactly the same, yet we suspect that they would be essentially similar. These reflections serve to highlight the value of the methodology used, as they show that great domains can be represented even under somewhat unfavorable conditions.

It is reasonable to propose this methodology as perfectly valid for the representation and analysis of large domains of knowledge or information from a social point of view, and that the renderings be used as interfaces for information retrieval. The cutoff values used in the construction of the maps may be adjusted depending on the user's objective, whether it be domain analysis per se or the use of the heliocentic map as an interface.

Furthermore, as a general conclusion, the research efforts reflected in our maps are not distributed uniformly over disciplines or over countries. The time period we analyze here is too short to show the evolution of research in a country. However, year to year snapshots might be used to observe and compare the evolution of science or progress in other domains. Groupings of three or four years, in our opinion, would provide this sort of realistic insight.

Interested readers may consult http://www.atlasofscience.net, at its third level of navigation or browsing, for all the heliocentric maps of Spain, featuring the characteristics for information retrieval mentioned in this article for the period 1990–2002.

Acknowledgement

References

Hjorland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science (JASIS), 46*(6), 400–425.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31*(1), 7–15.

Moya Anegón, F. d., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics, 61*(1), 129–145.

Salton, G., & Bergmark, D. (1979). A citation study of computer science literature. *Professional Communication, IEEE Transaction, PC-22*, 146–158.

Small, H. (2000). Charting pathways through science: exploring Garfield's vision of a unified index to science. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: a Festschrift in honor of Eugene Garfield* (pp. 449–473). Medford, N.J.: Information Today.

Small, H., & Garfield, E. (1985). The geography of science: disciplinary and national mappings. *Journal of Information Science, 11*(4), 147–159.

The Thomson Corporation (2004a). *ISI Journal Citation Reports*. Available from: <http://isiknowledge.com> (visited: 09/07/2004).

The Thomson Corporation (2004b). *ISI Web of Science*. Available from: <http://isiknowledge.com> (visited: 09/07/2004).

Tufte, E. R. (1994). *Envisioning information*. Cheshire: Graphics Press.

Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire: Graphics Press.

W3C (2004). *Scalable Vector Graphics (SVG) 1.1 Specification*. Available from: <http://www.w3.org/TR/SVG11/> (visited: 10/09/2004).

Wasserman, S., & Faust, K. (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.