Co-Network Analysis

Evaristo Jiménez-Contreras[*], Emilio Delgado-López-Cózar[*], Rafael Ruiz-Pérez[*], Mercedes de la Moneda-Corrochano[*], Rosario Ruiz-Baños[*] and Rafael Bailón-Moreno[**]

evaristo@ugr.es
[*]Departamento de Biblioteconomía y Documentación. Colegio Máximo de Cartuja.
Universidad de Granada. 18071 - Granada. Spain.

[**]Departamento de Ingeniería Química. Facultad de Ciencias.
Universidad de Granada. 18071- Granada. Spain.

Abstract
Here, we present a new method for studying the network of actors, called co-network analysis. It consists of considering actors as vectors whose components are networks previously constructed by co-word analysis. This enables actors to be linked according to their similarity in terms of their assignment to the networks, inheriting moreover the centrality and density values that the networks possess. In this way, it is possible to construct graphic representations in which the thematically similar actors appear in strategic positions which are also similar. That is, it is an improved combination of the classical strategic diagram of co-word analysis and MDS analysis. It is also possible to establish a generality index of the actors. This technique is applicable to research policies, scientific and technological monitoring, or sociological analysis.

Introduction
It is well-known that a scientific document can be represented by a group of key words or descriptors. If each descriptor represents a dimension, the complete vocabulary used to represent the set of documents gives us the dimensionality of the vectorial space in which the document is a vector (Lee, Kim, & Lee, 1993; Jones, 1993; Bartell, Cottrell, & Belew, 1995; Ding, Chowdhury, & Foo, 2001; Tafani, 1976; Belkin & Croft, 1987).

It is possible to relate these documents by the scalar product of the vectors, this technique being applied for the recovery of information from document databases and for the generation of document networks. Documents are recovered, according to Salton, relating a vector-interrogation with the document vectors, recovering those that exceed the minimum threshold of similarity among the vectors (Salton, 1979). Also, the crossed probability that they have common descriptors can be calculated, enabling thematic networks to be constructed according to Co-Word Analysis (Callon, Courtial, & Laville, 1991).

These third-generation relational techniques of representation are based on the use of word vectors as a means of comparison. The first generation are the indicators of non-relational activity, mainly productivity and citations, and the second generation are the relational indicators based on co-citations (Callon, Courtial, & Penan, 1995). For example, with co-word analysis, thematic networks are constructed by division into subnetworks of descriptors. With each subnetwork, that is really a network by itself, we can associate a list of actors as researchers, journals or institutions. This actors produce or publish documents thematically similar to these subnetworks, but between these actors no thematic relationship is established.

Aim
A proposal to relate actors and generate networks would be to consider the actors as vectors with coordinates that are descriptors and to compare them; but we have a more advanced proposal, and as will be explained later, more fruitful. This will be to consider vectors of networks or vectors of vectors for which the coordinates are not scalar, but are also vectors. This latter technique, which we call co-network analysis, we shall say is fourth generation since it is at an organizational level a step higher than the third generation, and it is the one that will be presented for the first time here. The networks that can be constructed can be of any type of actor: researcher, journal, laboratory, etc..

Proceedings of ISSI 2005, pages 417–425.
Edited by P. Ingwersen and B. Larsen.

And its applications are very diverse, such as research policies, scientific and technological monitoring, or sociological analysis.

Material and Methods

To present this new methodology called co-network analysis, we shall consider the case in which the actors are journals. Each journal is represented by vectors with coordinates which are networks of words, these networks being previously determined as themes or research lines by co-word analysis.

For this, all the documents are downloaded from the database ISOC-Psychology of the Centro de Información y Documentación Científica (CINDOC) of the Consejo Superior de Investigaciones Científicas (CSIC) for the period 1993-2002. The body of documents is formed by 7902 articles published by a total of 250 journals, of which a fraction will be considered as actors belonging to the area of psychology and the rest as occasional actors.

Previous co-word analysis

The knowledge system CoPalRed© was used. The parameters used for the analysis were: occurrence and co-occurrence, minimum of 10, which implies a free vocabulary of 779 descriptors. The size of the cluster was established at a minimum threshold of 4 and a maximum of 10 descriptors. The subnetworks generated were the main editorial lines followed by the bulk of the Spanish journals and these serve us as components of the vectors that represent these same journals. Clearly, key words are not used but rather networks.

Algorithm of co-network analysis

The steps to follow are:

1. Generation of a database from the documents downloaded from the ISOC-Psicología database.

2. Co-word analysis by the appropriate software, such as CoPalRed©, which enables us to relate the subnetworks with the actors that generate them: countries producing the documents, research centres and labs, researchers, and journals. In our case, we have chosen the journals as the object of study. In general, we shall consider an actor *i*, for example the journal *i*, and a theme *j*, represented by a network and which is the result of co-word analysis. The weight of the actor *i* will be defined with respect to the theme *j*, p_{ij}, as the quotient between the number of articles published by the actor that contain at least a certain quantity of common links with the network of the theme a_{ij} and the total number of articles published by the actor a_i:

$$p_{ij} = \frac{a_{ij}}{a_i} \qquad \text{eq. 1}$$

The weight of the actor is the way to condense into one number the fraction of the network *j* that is part of the actor *i*.

3. Each actor is represented by a vector, $\overrightarrow{R_i}$, with as many dimensions or coordinates as subnetworks that constitute the overall network of research in the scientific field considered. The value of the coordinates is the weight, p_{ij}, of the actor of each network considered.

$$\overrightarrow{R_i} = (p_{i1}, p_{i2}, p_{i3}, ...) \qquad \text{eq. 2}$$

4. If we take into account all the actors, we construct a matrix of actors *x* networks (Table 1)

Table 1.- Matrix actors *x* networks

	Network 1	Network 2	...	Network J
Actor 1	P_{11}	P_{12}	...	P_{1J}
Actor 2	P_{21}	P_{22}	...	P_{2J}
...	...	...	...	...
Actor I	P_{I1}	P_{I2}	...	p_{IJ}

5. The actors are compared on the basis of the scalar product of their vectors. For the actor *s* and the actor *t*, the scalar product of the two would be:

$$\overrightarrow{R_s} \bullet \overrightarrow{R_t} = \left|\overrightarrow{R_s}\right|\left|\overrightarrow{R_t}\right| \cos\alpha_{st} \qquad \text{eq. 3}$$

If two actors—represented by their vectors whose *J* coordinates are related with their thematic content—are similar, their vectors should also be similar. In geometric terms, this is interpreted by stating that the vectors are parallel or almost parallel. This implies that the angle α between them is zero or close to zero, and therefore its cosine is equal to or very close to unity. On the contrary, very different actors are represented by very different vectors with practically right angles —if the coordinates are strictly positive— and the cosine practically null or null. Therefore, it is a standardized index appropriate to compare actors. From the previous equation, the cosine of the angle formed by the actor s with the actor *t*, considering *J* dimensions, is determined with the following expression (using the values in Table 1):

$$\cos\alpha_{st} = \frac{\sum_{j=1}^{J} p_{sj} p_{tj}}{\sqrt{\sum_{j=1}^{J} p_{sj}^2}\sqrt{\sum_{j=1}^{J} p_{tj}^2}} \qquad \text{eq. 4}$$

Since:

$$\overrightarrow{R_s} \bullet \overrightarrow{R_t} = \sum_{j=1}^{J} p_{sj} p_{tj} \qquad \text{eq. 5}$$

$$\left|\overrightarrow{R_s}\right| = \sqrt{\sum_{j=1}^{J} p_{sj}^2} \qquad \text{eq. 6}$$

and

$$\left|\overrightarrow{R_t}\right| = \sqrt{\sum_{j=1}^{J} p_{tj}^2} \qquad \text{eq. 7}$$

6. From the cross of all the actors, we construct a symmetrical square matrix of actors *x* actors whose cells contain all the cosine combinations. As the cosine is a measure of similarity or nearness, the matrix is also called an adjacency matrix. Table 2 shows this matrix. Note that the main diagonal consistently contains 1, since each actor is exactly equal to itself according to reflexive property. Furthermore, only half has been filled, since the other half is the mirror image.

Table 2.- Matrix of actors *x* actors (adjacency matrix)

	Actor 1	Actor 2	...	Actor S
Actor 1	1	$\text{Cos}\alpha_{12}$	...	$\text{Cos}\alpha_{1S}$
Actor 2		1	...	$\text{Cos}\alpha_{2S}$
...			1	
Actor S				1

7. For each actor *i*, we have calculated the mean density, as the weighted measure of the density of the terms or networks to which it is assigned. The weighting criterion is the weight of the actor with the network considered, p_{ij}:

$$\overline{d}_i = \frac{\sum_{j=1}^{J} p_{ij} d_j}{\sum_{j=1}^{J} p_{ij}} \qquad \text{eq. 8}$$

This index gives us an idea, for example, of the development of the editorial lines of a journal or the implication in the scientific field of a laboratory.

8. Analogously, for each actor, we have calculated its mean centrality:

$$\overline{c}_i = \frac{\sum_{j=1}^{J} p_{ij} c_j}{\sum_{j=1}^{J} p_{ij}} \qquad \text{eq. 9}$$

The mean centrality of the actor measures the breadth of the spectrum of themes that it publishes. It is also related to its generalist character, although for this latter consideration, a new index has been formulated, called the generality index, g_i.

9. The generality index of an actor i, g_i, is defined as the quotient between the number of networks to which the actor is assigned—that is, the number of non-null components of its vector, t_i, beginning between the total number of themes, J, comprising the scientific area considered.

$$g_i = \frac{t_i}{J} \qquad \text{eq. 10}$$

10. From the adjacency matrix, the mean centrality and mean density values, as well as productivity, are processed with a specific software created for this study. We have succeeded in preparing a graphic representation of the network of actors based on the similarity of the thematic networks that share these actors. This network, together with the previously mentioned indicators of mean density, mean centrality, and generality index are the fundamental output of the analysis of co-network analysis.

Results and Analysis

The results shown next are for the case indicated previously of Spanish psychological journals. From a total of 250 in the database of the ISOC, we selected those with relative productivity in the field of psychology greater than 30% (total of 90). Of these, 67 are considered by the ISOC as strictly belonging to this field because all the published articles are gathered as concerning psychology.

Thematic network of Psychology

CoPalRed©, beginning from the ISOC database and by means of a Co-Word analysis, it becomes clear that the scientific field under study consists of 20 sub-networks. Table 3 lists the names given to these sub-networks according to the central word and Figure 1 gives the drawing of some examples of them.

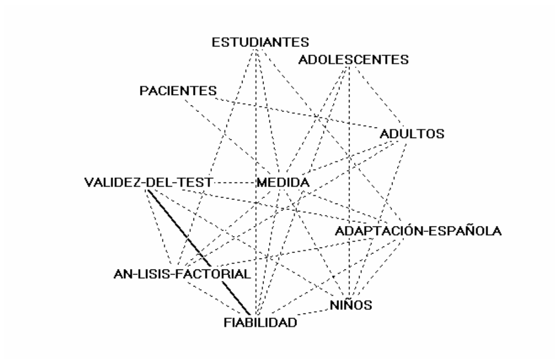
As a means of comparison, the networks of descriptors are more explanatory than a series of primary unfiltered descriptors. In addition, they provide additional information that the descriptors alone do not offer, such as centrality and density.

Table 3.- Sub-networks of the scientific field of Psychology. Example of weight of two journals

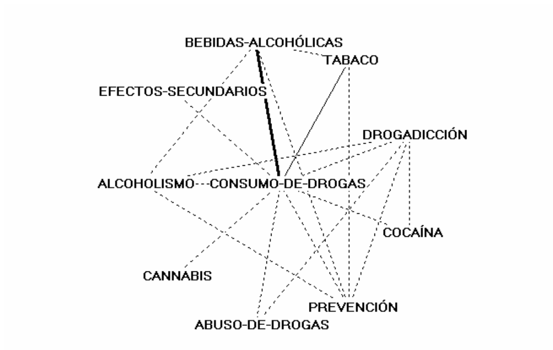
Network	**Centrality**	**Density**	**Journal's Weight**	
			Psicothema	***Gramma***
Network 1: Medida (Measurement)	28,6	9,8	0,307	0,045
Network 2: Consumo de Drogas (Drug consumption)	11,3	9,1	0,014	0,000
Network 3: Psicoanálisis (Psychoanalysis)	5,6	6,0	0,003	0,000
Network 4: Aprendizaje (Learning)	9,7	3,3	0,038	0,015
Network 5: Tratamiento (Treatment)	14,4	5,4	0,046	0,015
Network 6: Memoria (Memory)	10,6	2,1	0,034	0,000
Network 7: Programas de Intervencion (Programs of intervention)	10,5	3,5	0,013	0,000
Network 8: Mujeres (Women)	13,3	4,4	0,018	0,015
Network 9: Teorías (Theories)	8,6	2,2	0,056	0,000
Network 10: Trastornos Mentales (Mental disorders)	3,4	1,6	0,001	0,000
Network 11: Actitudes (Attitudes)	3,8	7,3	0,020	0,000
Network 12: Historia de La Psicología (History of psychology)	2,9	4,2	0,004	0,000
Network 13: Análisis Estadístico (Statistical analysis)	1,6	9,1	0,041	0,000
Network 14: Sintomatología (Symptom)	5,4	1,4	0,001	0,000
Network 15: Factores de Riesgo (Factors of Risk)	4,5	2,1	0,003	0,000
Network 16: Español (Spanish)	0,1	19,9	0,004	0,000
Network 17: Empleo (Employment)	3,7	7,5	0,003	0,000
Network 18: Vejez (Old age)	1,0	6,9	0,004	0,000
Network 19: Deporte (Sport)	1,4	9,3	0,010	0,000
Network 20: Grafología (Graphology)	0,9	10,9	0,000	0,591

Journals as network vectors

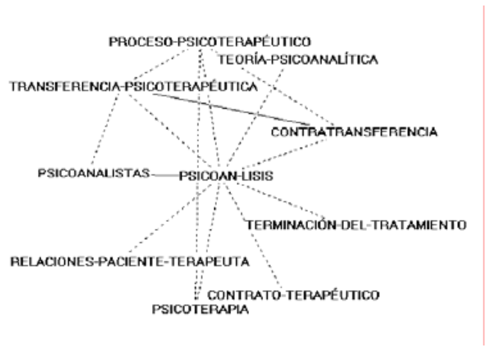
With CoPalRed©, we have identified the networks that are associated with each journal. For the present study, we have followed the least restrictive criterion to assign the documents to the themes. That is, a document of a journal belongs to a theme or network if this document contains a single link to the network that represents this theme. The fraction of the articles of the journal that satisfy this condition will therefore be the weight of that journal with respect to that network. The process is repeated for each of the networks, thus constructing the vector of the journal. Table 3 shows the weights of two journals that have been taken as examples, and therefore the columns of each are the vectors that represent them. We note that *Psicothema* is assigned to 19 of the 20 networks (considering a threshold of a weight of 0.001) whereas *Gramma*, a journal dedicated to Graphology, hardly in 5 networks, with a special influence of course in the network "Graphology" ("Grafología") It is evident that *Psicothema*, having representative documents in practically all the networks, can be considered as a general journal (its index of generality, according to Eq. 10, is 19/20 = 0.95, which in percentage is 95%). On the other hand, *Gramma* is a very specialized journal, it generality index being low (5/20 = 0.25, i.e. 25%)



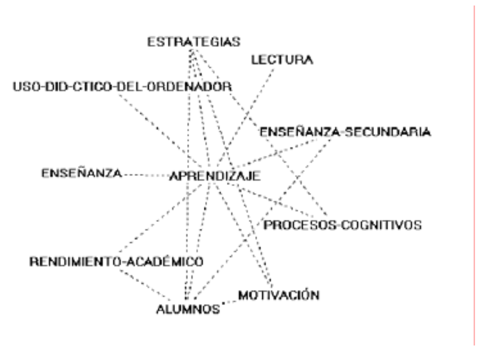
Network 1: “Medida” (Measurement))



Network 2: “Consumo de Drogas” (Drug consumption)



Network 3: “Psicoanálisis” (Psychoanalysis)



Network 4: “Aprendizaje” (Learning)

Figure 1.-Examples of thematic networks in the field of Psychology.

The weight of *Psicothema* in network 1 “Measured” (“Medida”) is 0.307 (30.7%), a value resulting from the quotient between 218 articles belonging to the network and 711 total articles published. On the other hand, *Gramma* has 3 of the 66 total articles published, and thus its weight, according to Eq. 1 is only 0.045 (4.5%)

Table 4.- Mean centrality and mean density of the journals Psicothema and Gramma

Journal	**Mean centrality,** $\overline{c}_i$	**Mean density,** $\overline{d}_i$
Psicothema	18.4	3.5
Gramma	7.4	10.4

The mean density and centrality of these journals are calculated according to Eq. 8 and 9. The results are shown in Table 4.

In a way similar to that used with the journals *Psicothema* and *Gramma*, the vectors of all the journals can be determined, together with their generality indices as well as the mean centrality and density values.

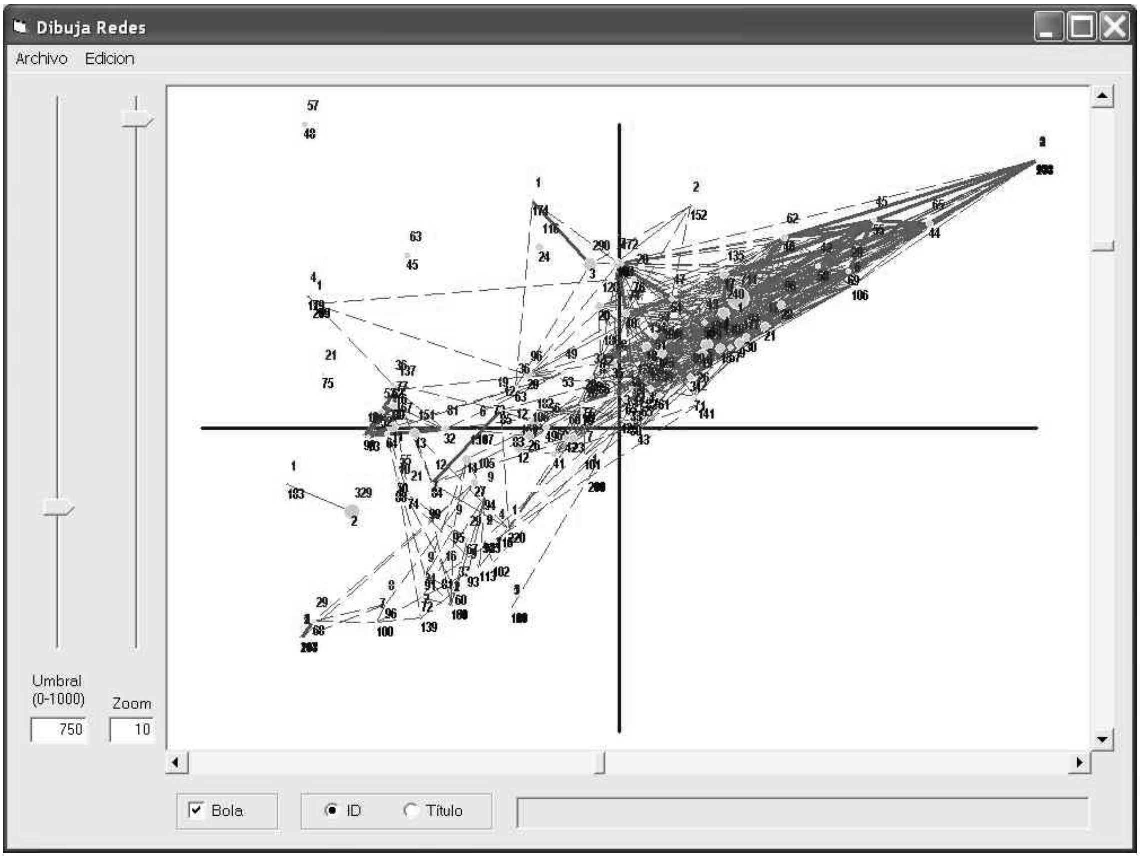


Figure 2.-Network of psychology journals. The links for which the cosine is equal or greater than 0.750 are drawn.

In addition, from the scalar product of all these vectors, and using Eq. 4, we can construct the adjacency matrix. This matrix represents the network of actors, which, superimposed over the positions of the journals in the strategic diagram, calculated previously, leads to an image as in Figure 2. The strategic diagram is a cartesian diagram where axis "x" is the centrality and the axis "y" the density.

Groups of journals according to their thematic similarity

Modifying the minimum threshold of the cosine to draw the links, it is possible to find groupings of journals especially linked together. For example, when we represent only the links with the cosine equal to or greater than 0.950, the image produced is that of Figure 3. This representation shows some strongly linked groups, of which we can highlight the group of general journals and the Psychiatry journals. In the upper left-hand part, with a high mean density and low mean centrality, the journal *Gramma* appears isolated for its high specialization. On the other hand, the journal *Psicothema* is within the group of general journals in the upper right-hand quadrant.

Figure 4 presents a detail of the 17 most productive journals. We can appreciate with total clarity the relationships that make up the network of the most elite journals of psychology in Spain.

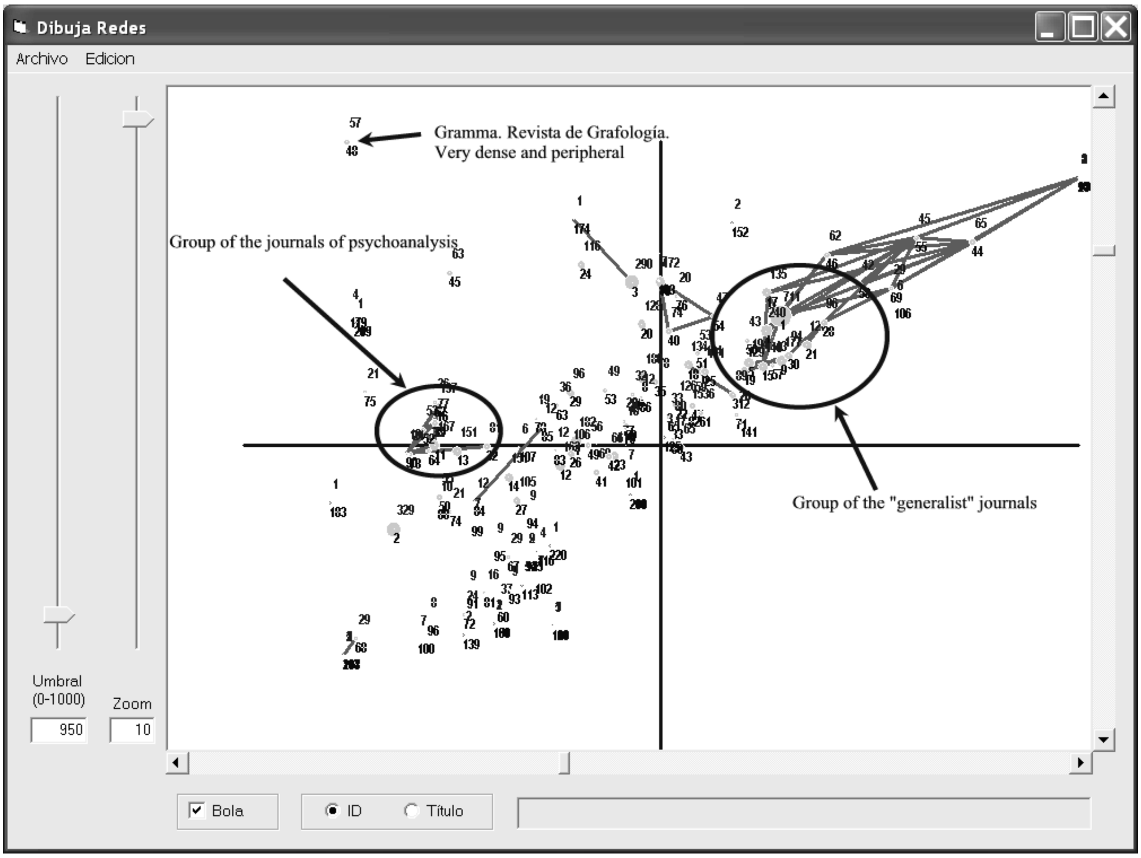


Figure 3.- Network of psychology journals. The links with a cosine of equal to or more than 0.950 are drawn.

Conclusions

The co-network analysis presented here is an appropriate scientometric tool for studying the networks of actors that produce documents (countries, research centres and laboratories, businesses, and researchers) or that publish them (e.g. scientific journals).

This analysis, based on networks, presents numerous advantages, among which we can highlight:

--Noise is eliminated in the comparison, since the groupings of the descriptors have previously passed through a filter that has identified them as the most pertinent ones.

--The subnetworks also provide a basic classification matrix of the actors.

--It is possible to establish, in this way, a quantitative criterion of generality or specialization.

--Co-word analysis assigns to each subnetwork two indices: density and centrality. This enables an evaluation of the degree of cohesion of the actors themselves from the mean density and the centrality of the themes to which they are assigned.

--Finally, the mean centrality and density values serve as axes of a bidimensional representation of the network of actors, similar to an MDS analysis, but with the advantage that the axes have an intrinsic meaning that enables interpretations not only in terms of pure similarity or nearness, but also in terms of cohesion, as indicated in the foregoing point. That is, we have achieved an improved combination of the strategic classical diagram of co-word analysis and MDS analysis.

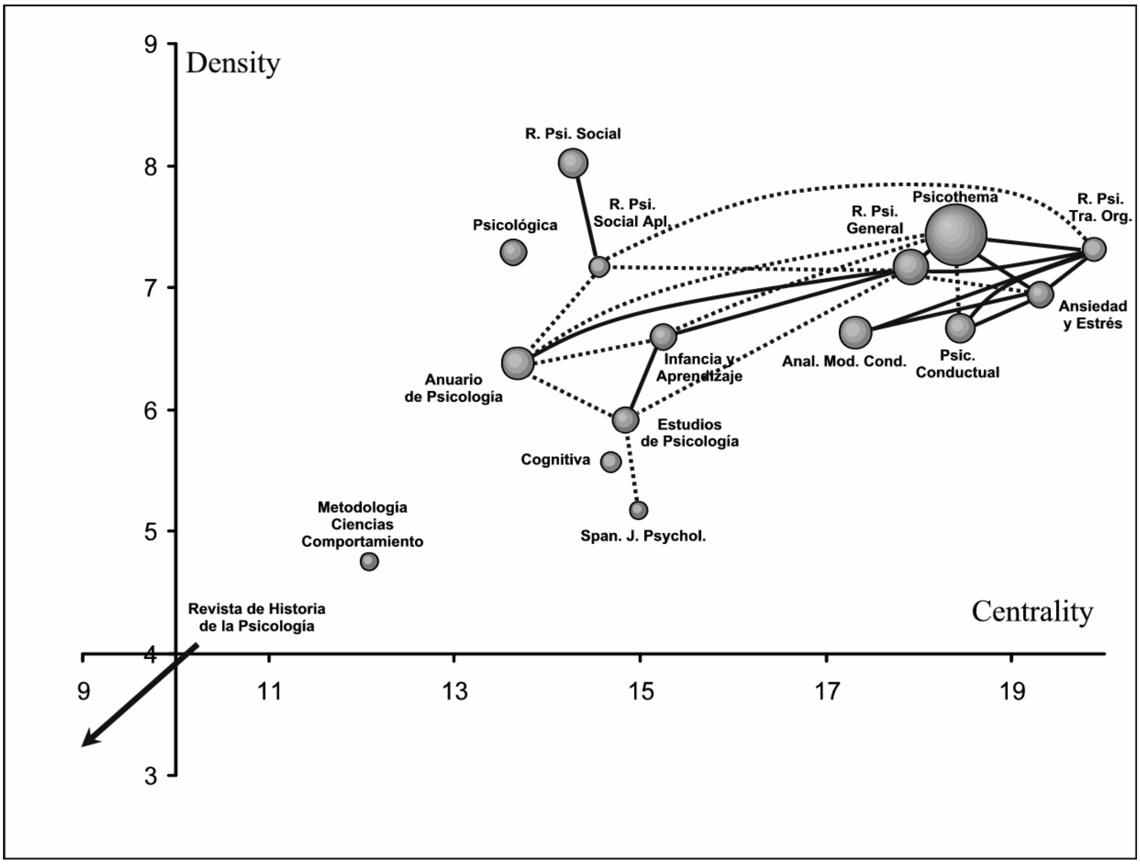


Figure 4.- Detail of the network of the most productive journals of the upper right-hand quadrant.

In short, co-network analysis represents a qualitative leap in the techniques of representing networks of actors and can undoubtedly be applied in research policies, scientific and technological monitoring or sociological analysis.

References

Bartell, B. T., Cottrell, G. W., & Belew, R. K. (1995). Representing documents using an explicit model of their similarities. *Journal of the American Society for Information Science*, 46(4), 254-271.

Belkin, N. J., & Croft, B. W. (1987). Retrieval Techniques. *Annual Review of Information Science and Technology (ARIST),* 22, 107-145.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics*, 22(1), 155-205.

Callon, M., Courtial, J. P., & Penan, H. (1995). *Cienciometría. La medición de la actividad científica: De la bibliometría a la vigilancia tecnológica*. Gijón: Trea.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management,* 37, 817-842.

Jones, S. (1993). A thesaurus data model for an intelligent retrieval system. *Journal of Information Science*, 19, 167-178.

Lee, J. H., Kim, M. H., & Lee, Y. J. (1993). Information retrieval based on conceptual distance in is-a hierarchies. *Journal of Documentation*, 49(2), 188-207.

Salton, G. (1979). Mathematics and information retrieval. *Journal of Documentation*, 35(1), 1-29.

Tafani, V. (1976). A fuzzy model of document retrieval systems. *Information Processing & Management,* 12, 177-187.