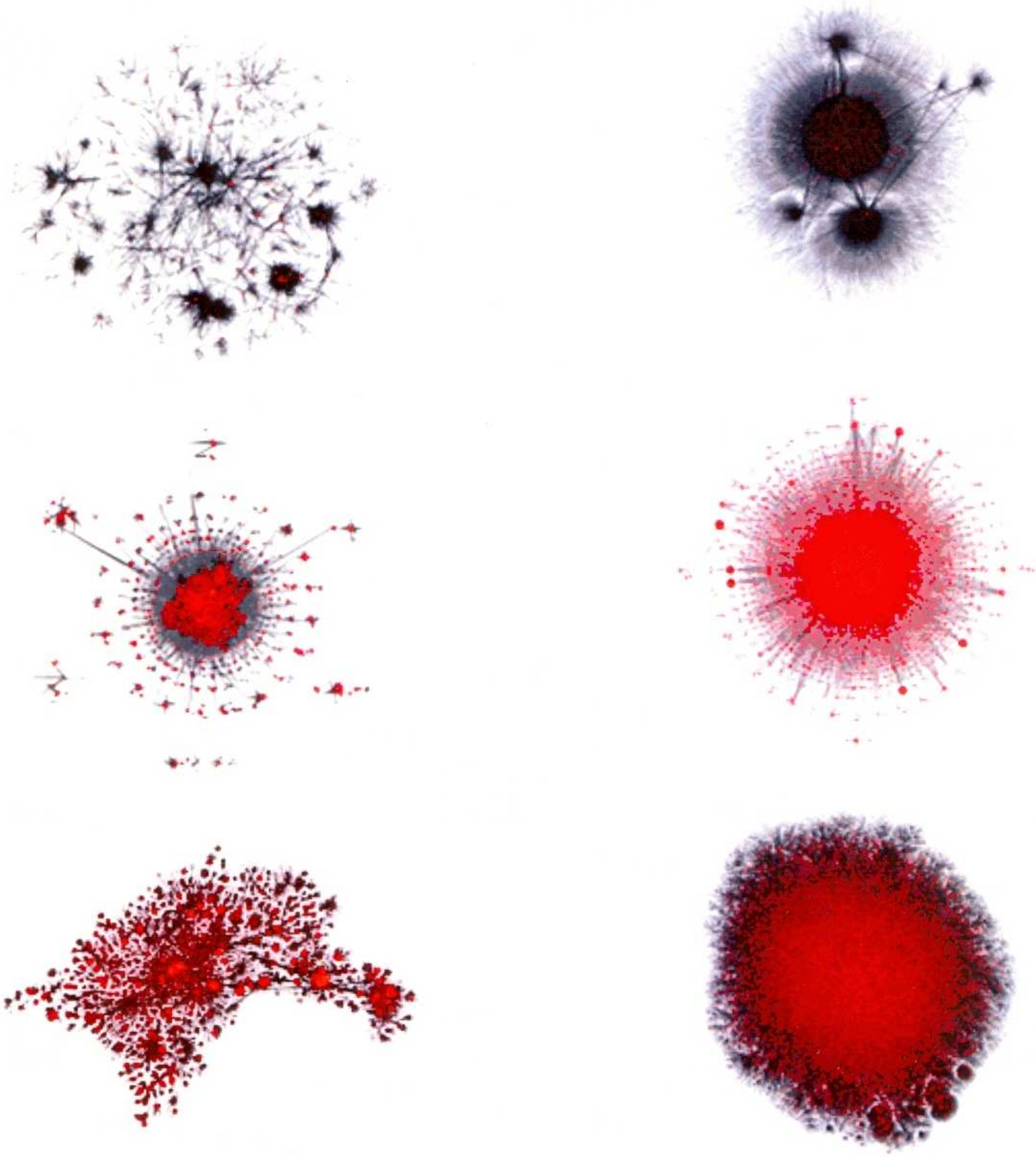
Information Visualization

Edited by

Ebad Banissi, Liz Stuart, Theodor G. Wyeld, Mikael Jern, Gennady Andrienko, Nasrullah Memon, Reda Alhajj, Remo Aslak Burkhard, Georges Grinstein, Dennis Groth, Anna Ursyn, Jimmy Johansson, Camilla Forsell, Urska Cvek, Marjan Trutschl, Francis T. Marchese, Carsten Maple, Andrew J. Cowell, and Andrew Vande Moere



Hubs and Authorities in a Spanish Co-authorship Network

Mª Antonia Ovalle-Perandones, Antonio Perianes-Rodriguez, Carlos Olmeda-Gomez,
Department of Library and Information Science. Carlos III University. Getafe, Spain
movalle@bib.uc3m.es, aperiane@bib.uc3m.es, olmeda@bib.uc3m.es

Abstract

How can the prestige of research centres forming part of scientific co-authorship networks focused on Physiology and Pharmacology be measured?

This paper attempts to answer that question on the basis of a bibliometric analysis of Spanish scientific production in these areas of research between 1995 and 2005 as listed in Thomson Reuters' Science Citation Index Expanded. An affinity index is used to measure the asymmetric co-authorship relationships between any two institutions on the collaboration network to obtain the hub and authority values for the leading institutions. The spatial distribution of network nodes is mapped with the Kamada Kwai algorithm. The findings identify the centres of greatest prestige from the standpoint of co-authorship of scientific papers.

1. Introduction

Many co-authorship networks have been studied to explore the structural properties of scientific collaboration [1]-[2] by social scientists drawn to the subject by the awareness that such networks contain all the ingredients of small worlds in their make-up [3], while also reflecting the dynamic aspects that govern the development of such complex systems [4]. Their findings have shown that in these scale-free networks development is governed by the principle of preferential attachment [5] and the vertex degree and connection strength distributions by a power law [6].

This article examines the values of hubs and authorities [7] found from Web of Science data for the network of pharmacological papers co-authored by Spanish companies and Spanish public research bodies such as universities and hospitals. It constitutes a continuation of prior studies on proposals to contribute to the measurement of co-authorship network actors' status or influence [8]-[9].

1.1 Related work

Today's interest in the analysis of the factors that contribute to node status in the context of bibliometric citation or scientific co-authorship networks is the result of the success of the Google page ranking algorithm [10]. Generically based on an iterative process of calculating both the number of links received by a website and the status of the sites hosting those links, it has become a standard for evaluating website status.

Based on this approach, new bibliometric indicators have been suggested to evaluate academic publications or the impact of their authors [11]-[12] against the backdrop of a review of the ways to measure the influence of academic publications and the agents and organizations constituting the scientific system. Katy Börner used weighted links and graph visualization techniques to analyze research teams through co-authorship networks with a view to identifying a new scientific field such as information visualization [13]. Leydesdorf detected the emergence of a world-wide core of countries that have collaborated most intensely since the nineteen nineties [14].

2. Method

2.1 Data collection

The data used in this study were drawn from the *Science Citation Index Expanded (SCI-Expanded)* database contained in Thomson Reuters' Web of Knowledge, downloaded in January 2008. In the first phase, all types of papers in which Spain appeared in the address field and which were published from 1995 to 2005 were retrieved from the base. In the second phase, a sub-set of papers was defined to include only those with standardized company addresses. A total of 1 557 papers (articles, biographical items, book reviews, corrections, editorial materials, letters, meeting abstracts, news items and reviews) published by Spanish research bodies were retrieved, 760 of which had been written jointly. Each paper was assigned to an institutional sector based on individual authors' institutional affiliation. The following classification was used: private enterprise, health system, university system, government, Spanish National

978-0-7695-3733-7/09 $25.00 © 2009 IEEE
DOI 10.1109/IV.2009.87

IEEE computer society

Research Council (CSIC), CSIC mixed centres, public research bodies (EPI) and others not classifiable in any of the aboce categories.

One of the problems that arises in bibliometric analyses of scientific disciplines is the criteria for classifying papers by scientific area. In large-scale analyses, the only practical way to allocate papers by area consists in using the subject categories into which the ISI's *Journal Citation Reports* (JCR) divides the journals where they are published. These ISI categories have subsequently been subdivided to establish more refined schemes such as the ANEP classification chosen for the present analysis. ANEP, the Spanish National Agency for Evaluation and Prospective Studies, is a Ministry of Science and Innovation body under the aegis of the Secretariat of State for Universities [15]. Co-author distribution by sectors is given in Table 1, while Table 2 shows the bibliometric parameters.

Table 1. Author distribution by sector

No. institutions	Sector	%
194	Private enterp.	41,28
175	Health	37,23
45	University	9,57
20	Government	4,26
14	CSIC	2,98
11	Others	2,34
7	Mixed centres	1,49
4	EPI	0,85

Table 2. Bibliometric parameters defining the Spanish pharmacology network (1995-2005)

Measure	Value
Total No. nodes	470
Total No. papers (all types) (a)	1 557
Total No. papers with inter-institutional collaboration (b)	760
Total No. papers with international collaboration	135
Co-authorship index	6,56
International co-authorship index	7,64
Total citations (a)	2 494
Total citations (b)	2 112
Total (b) cited	228
Total (b) not cited	532
Total citations per paper (a)	1,60
Total citations per paper (b)	2,78
Largest No. of authors of a single paper	27

2.2 Tools

Variations on organizations' names may distort the results of bibliometric analysis: different spellings, typographical errors, misuse of upper case, abuse of initials or abbreviations or mistakes in transliteration. To obviate these difficulties, *ad hoc* software was used that avoids homonymy by combining author and institution and synonymy by combining author and paper and corrects the lack of precision in institutional denominations [16].

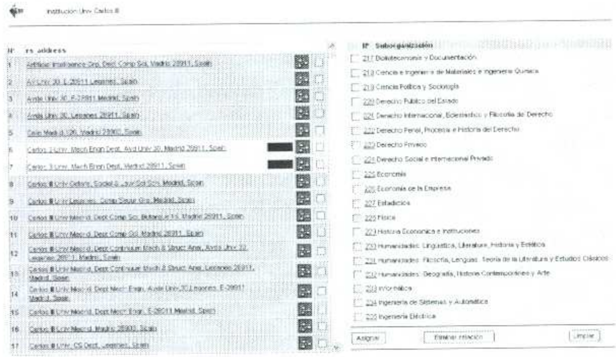


Figure 1. Software for refining author affiliation

2.3 Matrix generation

Calculating co-authorship from the database described at the paper level initially yielded symmetric or 1-mode matrices. Institutions were counted using full accounting, which attributes a value of "1" to each institutional author of an article whenever an institution appears in the set of papers. Co-authorship was therefore defined in terms of authors' institutional affiliation. Absolute co-authorship values were used in the calculations.

When applying visualization techniques to bibliometric co-authorship networks, one aspect to be borne in mind is the graphic representation of the direction of the relationship or link established by collaborating universities, and the effectiveness of that collaboration. The existence of collaboration between two countries, institutions or persons implies reciprocity, but provides no insight into the degree of dependence of one or the other. The degree of dependence may vary among organizations, for collaboration may not be symmetric. Confirmation or reciprocity is an important property of links in network analysis. Confirmation is not defined simply by the existence of the link, but by the degree to which the value of reciprocity is the same in the various nodes in the network.

Such dissimilarity in the degree of collaboration between universities is represented by computing the asymmetric collaboration rate and mapping the inter-university collaboration network, in which asymmetry is denoted by the differences in the direction of the arrows between nodes. This indicator, borrowed from the affinity index used to measure asymmetric relations between two countries [17], was adapted here to estimate asymmetric collaboration between two organizations.

It was calculated from formulas used to measure the direction of cooperation between any two nodes, as follows:

$$TCA(Insti_1 \rightarrow Insti_2) = \frac{COA(Insti_1 \leftrightarrow Insti_2)}{COA(Insti_1 \leftrightarrow total_{class})} \times 100$$

$$TCA(Insti_2 \rightarrow Insti_1) = \frac{COA(Insti_2 \leftrightarrow Insti_1)}{COA(Insti_2 \leftrightarrow total_{class})} \times 100$$

where TCA is the asymmetric collaboration rate between institutions 1 and 2, COA ($Insti_1 \leftrightarrow Insti_2$) is the total number of papers co-authored by institutions 1 and 2 and COA ($Insti_1 \leftrightarrow Total_{class}$) is the total set of papers co-authored by institution 1. This yields a directed network from which prestige values can be computed for the nodes.

The present study calculated the values of hubs and authorities. Nodal hub score is proportional to the combined authority score, and authority score is proportional to the combined hub score of in-neighbors. That is, nodal hub score becomes higher initially if the node has more out-neighbors, but it is affected by the authority scores of its out-neighbors. So, if a node has many out-neighbors which have low authority scores, hub score of that node will be low. Nodal authority score is similar to hub score, but it is affected by in-neighbors. In addition, authority score of a node is also affected by hub scores of its in-neighbors. In the present study, two weights, Xv, $Yv \in [0,1]$, were computed for each vertex v to determine its value as an authority and a hub. Vertex V is regarded to be a better authority than vertex U if $Xv>Xu$. Weights were computed according to network solving the eigenvector problems of matrices AA^T (hubs) and $A^T A$ (authorities) [18]. The Kamada Kwai algorithm [19] was used to map two asymmetric collaboration networks (Figure 2) for institutions with more than four co-authored papers.

3. Results

The findings are given in Tables 3 and 4, where institutions are ranked in descending order of their authority/hub weights.

Table 3. Top pharmacological authorities

Institution	Acronym	Author. weight
Clin & Provincial Hospital	HCPB	0,366
La Fe University Hospital	HULF	0,352
Valle Hebron Gen Univ Hosp	HGUVH	0,316
Univ of Barcelona	UB	0,223
Ramon y Cajal Hospital	HRYC	0,216
Virgen Macarena Univ. Hosp	HUVM	0,205
Miguel Servet Univ. Hosp.	HUMS	0,190
Autonomaous U. Barcelona	UAB	0,175
Lozano Blesa Univ. Hosp.	HLB	0,167
Sta Creu & Sant Pau Univ. Hosp.	HUSCSP	0,166
Salamanca Hosp. Complex	HSAL	0,164
Univ of Valencia	UV	0,153
San Carlos Univ. Hosp.	HUSCM	0,147
Virgen de las Nieves Hosp. Complex	HVLN	0,121
Municipal Inst Med Res.	IMIMB	0,121
Asturias Hosp. Cent.	HCAO	0,118
GLAXO WELLCOME	GLAXM	0,116
Dr Negrin Hosp. Complex	HDNEG	0,115
Univ of Oviedo	UNIOVI	0,111

Table 4. Top pharmacological hubs

Institution	Acronym	Hub weight
Grp Arkopharma	ARKO	0,268
Gynaecol Clin	GYNCLIN	0,268
Leon Hosp Complex	HLEO	0,226
Canary Univ Hosp.	HUCAN	0,226
Malalties Cardiovasc Clin. Inst.	ICLINMC	0,226
Santa Cristina Univ. Hosp.	HUSCR	0,166
Basurto Hosp.	HBAS	0,162
Dexeus Barcelona Univ. Inst.	IDEX	0,150
Island Maternity Hosp. Complex	HMI	0,148
Sant Joan Alacant Univ. Hosp.	HSJA	0,137
Getafe Univ. Hosp.	HUGET	0,137
Badajoz Univ. Hosp. Complex	HUB	0,133
Asturias Hosp. Cent.	HCAO	0,130
Oncol Inst.	IOSS	0,127
Santiago de Compostela Univ. Hosp. Complex	HUSC	0,120
Guadalajara Univ. Hosp.	HUGUA	0,114
San Pedro Alcantara Hosp.	HSPAC	0,113
Galdakao Hosp.	HGAL	0,113
Miguel Servet Univ. Hosp.	HUMS	0,107

In Figure 2, nodes are only connected by lines when at least four papers were written jointly for publication by researchers affiliated with the two respective institutions. The initial network of 470 nodes was thus reduced to 116. In both figures, the size of the node is proportional to the volume of co-authored production, the intensity of the colour of the link indicates the dimension of the asymmetric collaboration rate and the direction of the arrow denotes the direction of collaboration among network nodes. In Figure 2, the colour of the node indicates the sector to which the institution is assigned.

4. Discussion

The network studied contained 470 institutions that co-authored 760 papers on pharmacology published between 1995 and 2005. The resulting network illustrates how institutions inter-relate in terms of the degree of scientific co-authorship. The diagram generated is asymmetric: some institutions co-authored studies with different types of organizations, whereas some research centres collaborated nearly exclusively with other research centres.

This initial exploratory paper identifies the centres that roused the greatest interest as partners. The University and Provincial Hospital at Barcelona, La Fe University Hospital at Granada, and Valle de Hebrón General University Hospital at Barcelona proved to be good authorities, followed at a significant distance by the University of Barcelona (0,22) and Ramon & Cajal Hospital at Madrid (0,21). Health system institutions prevail in the list of top institutional authorities, along with four universities (U. of Barcelona, Autonomous U. of Barcelona, U. of Valencia and U. of Oviedo). The network in Figure 2 can be used as a basis for discussion of a number of interesting characteristics of the structure of the Spanish co-authorship network and contributes to the understanding of the mechanisms used to create co-authorship links by different types of organizations engaging in scientific production in the area.

5. Conclusions

The present preliminary analysis of co-authorship data establishes the prestige of the university hospitals at Barcelona, Seville and Zaragoza in the Spanish physiology and pharmacology network. The main network authorities, located there, have co-authorship ties with universities in Barcelona and Oviedo, while their working relations with private enterprise are much less intense. The algorithm presumes that a good hub is an organization that connects to many others and a good authority an organization to which many others connect. These provisional results constitute a stimulus to continue the study of real co-authorship networks, and to apply the findings of social network analysis.

References

[1] M.E.J..Newman, "Scientific collaboration networks. I. network construction and fundamental results", *Physical Review E* 64, 016131, 2001.

[2] M.E.J. Newman, "Coauthorship networks and patterns of scientific collaboration",*Proceedings of the National Academy of Sciences of the United States of America* 101,April 6, 2004, pp. 5200-5205.

[3] D. Watts, *Small Worlds: The dynamics of networks between order and randomness,* Pricenton, Princeton University Press, 1999.

[4] S. N. Dorogovtsev, and J.F.F. Mendes, "Evolution of networks", *Advances in Physics* 51 1079, 2002. [arXiv:cond-mat/0106144v2].

[5] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations", *Physica A,* 311, 2002, pp. 590-614.

[6] Ch. Li, and G. Chen, "Network connection strengths: another power-law?", November, 2003 [ArXiv:cond-mat/0311333].

[7] J. M Kleinberg, "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, 46 (5), 1999, pp. 604-632.

[8] A. Perianes-Rodriguez, C. Olmeda-Gomez, and F. de Moya-Anegon, "Detecting research groups in coauthorship networks",*Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth Collnet Meeting*, Berlin, 2008.

[9] C. Olmeda-Gomez, A. Perianes-Rodriguez, MªA. Ovalle-Perandones, and F. de Moya-Anegon, "Comparative analysis of university-government-enterprise co-authorship networks in three scientific domains in the region of Madrid", *Information Research*, (13) 3, paper 352, 2008. [Available at http://InformationR.net/ir/13-3/paper352.html].

[10] S. Brin, and L. Page, "The anatomy of a large-scale hypertextual web search engine", *Computer Networks and ISDN systems,* 30 (1-7), 1998, pp. 107-117.

[11] J. Bollen, M.A. Rodríguez, and H. van de Sompel, "Journal status", *Scientometrics,* 69 (3), 2006, pp. 669-687.

[12] X. Liu, J. Bollen, M.L. Nelson, and H. van de Sompel, "Co-authorship network in the digital library research community", *Information Processing & Management,* 41, 2005, pp. 1462-1480.

[13] K. Börner, L. Dall'Asta, W. Ke, and A. Vespignani, "Studying the emerging global brain: analyzing and visualizing the impact of co-authorship teams", *Complexity,* 10 (4), 2005, pp. 57-67.

[14] L. Leydesdorff, and C.Wagner, "International collaboration in science and the formation of a core group". *Journal of Informetrics* 2 (4), 2008, pp. 317-325.

[15] ANEP, "Áreas temáticas". Madrid, Ministerio de Educación y Ciencia. 2006, [Available at http://www.micinn.es/ciencia/jsp/plantilla.jsp?area=anep&id=24&contenido=/anep/htm/areas.html]

[16] C. Gálvez, and F. de Moya-Anegón, "Standardizing formats of corporate source data", *Scientometrics*, 70 (1), 2007, pp. 3-26.

[17] M. Zitt, E. Bassecoulard, and Y. Okubo, "Shadows of past in international cooperation: collaboration profiles of the top five producers of science", *Scientometrics*, 47 (3), 2000, pp. 627-657.

[18] A. Mrvar, *Network analysis using Pajek.* Liubliana University, 2000.

[19] T. Kamada, and S. Kawai, "An algorithm for drawing general undirected graphs", *Information Processing Letter* 31 (1), September, 1994, pp. 31-43.

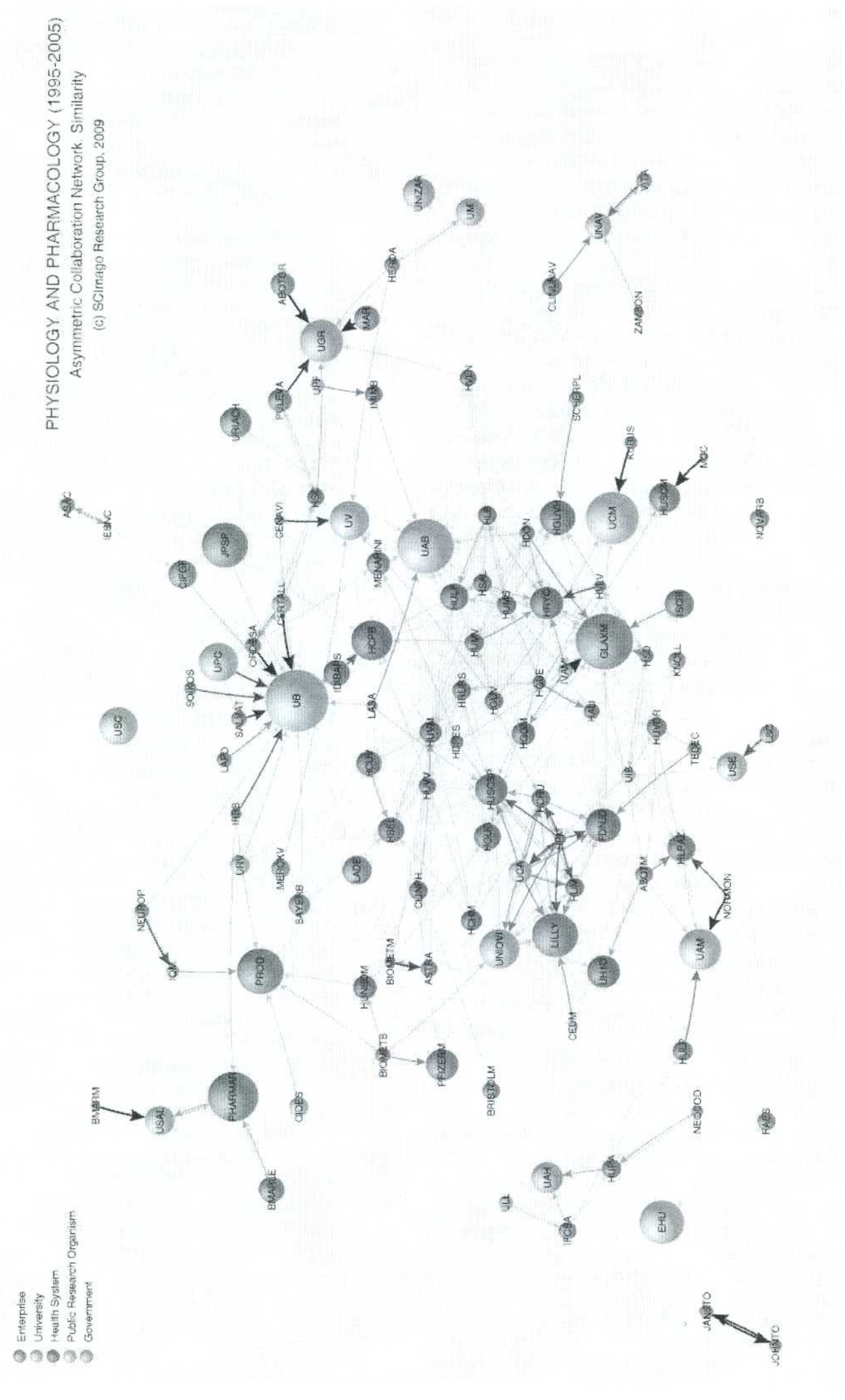


Figure 2. Physiology and Pharmacology asymmetric collaboration network