**Visualization of networks of international collaboration**

**F. Moya-Anegón [a], Z. Chinchilla-Rodríguez, B. Vargas-Quesada, A. González-Molina**

[a] *SCImago Research Group. Departamento de Biblioteconomía y Documentación. Universidad de Granada, Colegio Máximo de Cartuja, s/n, 18071 Granada, Spain*

This study shows visual displays of scientific information from which different aspects of international collaboration may be analyzed. The first aim is to identify the degree of "internationality" of research by highlighting fluxes of knowledge in the form of publications. This allows us, in turn, to identify the main geographical axes, to show the relationships of the analyzed domains with other countries, discovering the relative strength of these relations, and to see how they might affect the *visibility* of the work on the virtual horizon. Two final applications are foreseen: the detection of significant differences that can help characterize the publishing patterns of a given geographical domain or system of knowledge interchange, and the generation of visualizations that act as interfaces for domain analysis in general.

Key words: Bibliometrics, Scientific Output, Scientific Collaboration, Visualization, Domain Analysis

## 1 INTRODUCTION

Up until the eighties, approximately, scientific output was largely the harvest of individual initiative, moved forward by intellectual energy, though with some social component as well. Collaboration in the scientific realm is a reflection of the interaction of individual networks, which may also channel the work of institutional or global networks.[1, 2] At present, a number of scientifically well developed countries are intent upon promoting collaboration at all levels —national, international, interregional or institutional— through productive measures in scientific policy. It is now widely understood that sustainable development is the best foundation for the socioeconomic independence of less developed countries. Scientific cooperation favors the mobility of researchers and knowledge, and contributes to the common cultural and scientific wealth of participating parties.[3] In this sense, the networks that set up international collaboration have a dual and complementary role in the interchange of ideas: they vertebrate the scientific structure of knowledge, and they simplify and facilitate the implementation of policies for promoting international collaboration (mobility, means, and extension). This would explain the growing interest in analyzing cross-national research efforts, as a consolidated scientific *glasnost* or aperture, and their visibilty in terms of the impact on the scientific community. To do so, we make more easily visible the networks behind the generation/vertebration of scientific and technological knowledge.

The present contribution describes a methodology for the partial analysis of the patterns of international collaboration, in different geographical domains with common fields of research, by providing visual representations of scientific output. After identifying the most international line of investigation, as revealed in fluxes of knowledge and publication, we establish which are the strongest geographical axes. This allows us to likewise show the international relations of a given domain, the intensity of these relations, and their repercussions in output according to the type of underlying collaboration. Thus we can detect important differences that serve to characterize and represent the behavior of each individual domain or of the system in which they take part.

## 2 MATERIAL AND METHODS

The information used came from the *Science Citation Index, Social Sciences Citation Index* and *Arts & Humanities* databases of the Web of Science (Thomson Scientific). From the field *Address*, we gathered all the documents pertaining to each of the countries represented with output in 2004. The Impact Factor according to the JCR was taken into account. The registers were exported to a relational database, after which we normalized the data at all levels: institutional, thematic, geographic, and with regards to the JCR Impact Factor.[4, 5] To count records, we used the full recount, in which a document signed by more than one institution or country is considered to belong to each. The types of collaboration are:

- no collaboration (document with authors from a single institution)
- national collaboration (authors from more than one national institution)
- international collaboration (at least one author from another country and institutional affiliation)

The assignment of ISI categories to vast thematic domains follows the classification now in vigor in the ANEP (*Agencia Nacional de Evaluación y Prospectiva*)[6]. The indicators presented, in either tabular or graphic form, are: output in that area, number of collaborating countries, output from international

collaboration (documents undersigned by more than one foreign country), and Impact Factor within its particular field and adjusted according to type of collaboration.

At this point we were able to use the data to build a heliocentric network for the area Agriculture, taking in different Latin American countries as the central node. The volume is proportional to the percentage of documents published in collaboration with that country. The color signals the membership of each country to a geographic region. Orbiting around the node, at a greater or lesser distance, are those countries with which collaboration took place, and their relationship is represented with a line whose length is inversely proportional to visibility. In other words, the countries closest to the center achieve more visibility by virtue of their scientific impact, and the farther ones wield less impact. And so it is readibly visible with which countries one publishes more (greater volume) and with which it manages to make more impact on the scientific community (closer to the center).

Moreover, in order to compare the visibility of the intellectual association with a given country, three concentric circles can be seen with the relative values of impact according to the type of collaboration. These circles represent the national reference to visibility of the published documents: no collaboration (green), national collaboration (blue), and international collaboration (red). Thus we can situate the countries in terms of their membership to a peripheral circle (less visible), and determine which ones are above the average impact according to type of scientific interchange. The visual information is complemented with a table showing the indicators of study and structured into two parts —information about the field and the country, and data regarding the analysis of correlation among variables.

## 3 RESULTS AND DISCUSSION

Table 1 gives the results of the analysis of collaboration both at the thematic level (left side of Table) and in geographic terms (right). The country with the greatest volume of output in the area is Cuba, with 16%, followed by Mexico and Argentina with similar shares. Spain lags far behind with just under 3%. At the same time however, Mexico presents the greatest rates of international collaboration, followed by Brasil and Cuba, and then Spain (with 27.85%).

The country with the best results in tems of visibility in the field is Spain, with Mexico in second place. Meanwhile, we see Spain has twice the output of Brazil, which in turn doubles the output of Mexico. Cuba, then, takes last place as far as the visibility of its research studies in Agriculture.

With respect to the patterns of collaboration of each country in all the thematic areas, the proportion of documents produced without any collaboration in Brazil, Spain and Mexico is over 42%, while in Cuba its does not reach 25%. Brazil, showing the greatest rate of output without collaboration, is also the country evidencing more collaborative efforts with its own national institutions (54%), followed by Spain. In international collaboration, however, Cuba stands out with over half of its scientific output produced in conjunction with another country (54.76%). The country with the greatest visibility in all the fields of scientific publication for the year 2004 is Spain, followed in order by Mexico, Cuba and Brazil.

Table 1. Basic Indicators

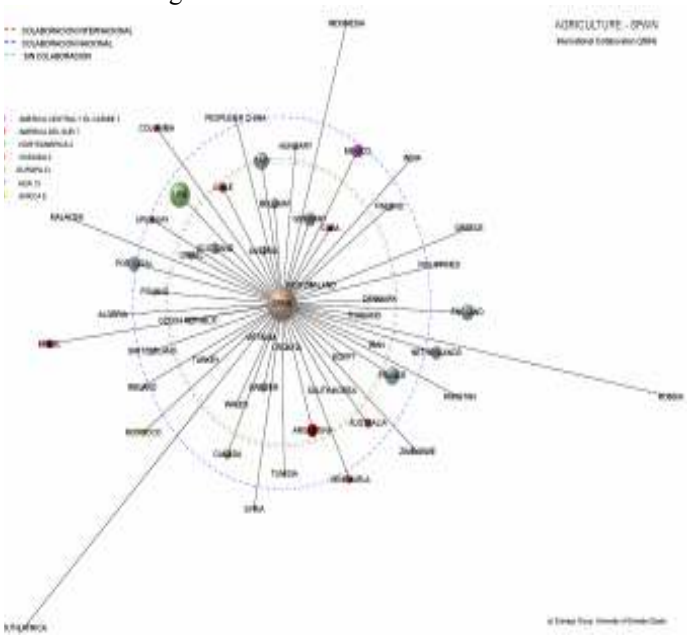
| | ANEP AREA - AGRICULTURE | | | | | | | | | COUNTRY | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **# countries** | **% countries** | **ndoc** | **% ndoc** | **col-int** | **% int** | **fin-al** | **fin-nac** | **fin-int** | countries | ndoc | %al | fin-al | % dom | fin-dom | %int | fin-int | fin |
| Spain | 52 | 39,39 | 1034 | 2,92 | 288 | 27,85 | **1,20** | 1,18 | 1,19 | 132 | 35412 | 42,71 | 1,02 | 30,31 | 1,09 | 33,53 | **1,13** | 1,08 |
| Brazil | 66 | **53,66** | 1577 | 8,52 | 499 | 31,64 | 0,69 | 0,75 | **0,86** | 123 | 18507 | **45,99** | 0,80 | **54,01** | 0,84 | 29,97 | **0,90** | 0,83 |
| Mexico | 48 | 48,98 | 728 | 9,24 | 282 | **38,74** | 0,90 | 0,94 | **1,05** | 98 | 7876 | 42,88 | 0,94 | 25,49 | 0,96 | 37,97 | **1,04** | 0,98 |
| Cuba | 16 | 26,67 | 115 | **16,11** | 36 | 31,30 | 0,66 | 0,79 | **0,92** | 60 | 714 | 24,71 | 0,77 | 21,10 | 0,89 | **54,76** | **1,01** | 0,92 |

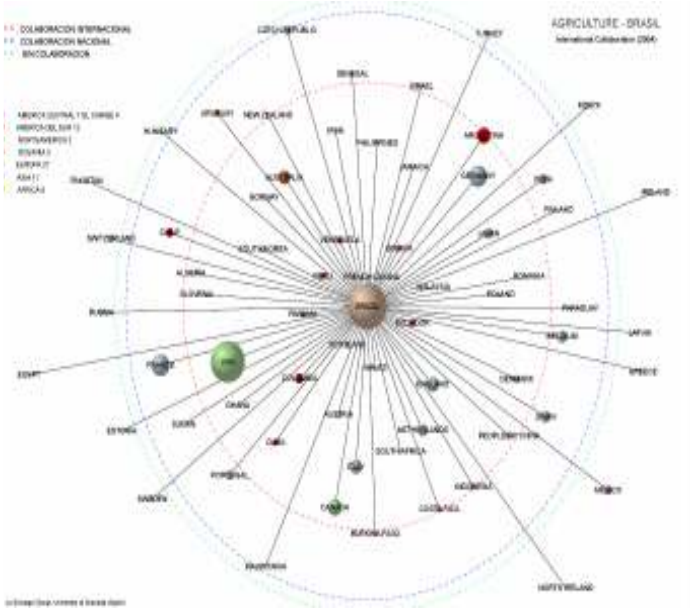
(ANEP area - **#countries:** number of countries with which it collaborates; %countries: percentage of collaborating countries with respect to national total; **ndoc:** number of documents in field; **%ndoc:** percentage of documents in that field with respect to national total; **col-int:** number of documents involving international collaboration**; %int:** percentage of documents in collaboration in that thematic area; **fin:** normalized impact factor for the area; //COUNTRY- **country:** total number of countries with which one collaborates, in all fields; **Ndoc:** output in 2004; **%alone:** proportion of documents signed by only one institution; **fin-alone:** normalized impact factor for documents without collaboration; **% dom:** proportion of documents signed by more than one national institution; **fin-nat:** normalized impact factor for the documents of national collaboration; **%int:** proportion of documents signed by more than one country; **fin-int:** normalized impact factor for documents of international collaboration; **fin:** normalized impact factor for documents produced by each country in all thematic areas).

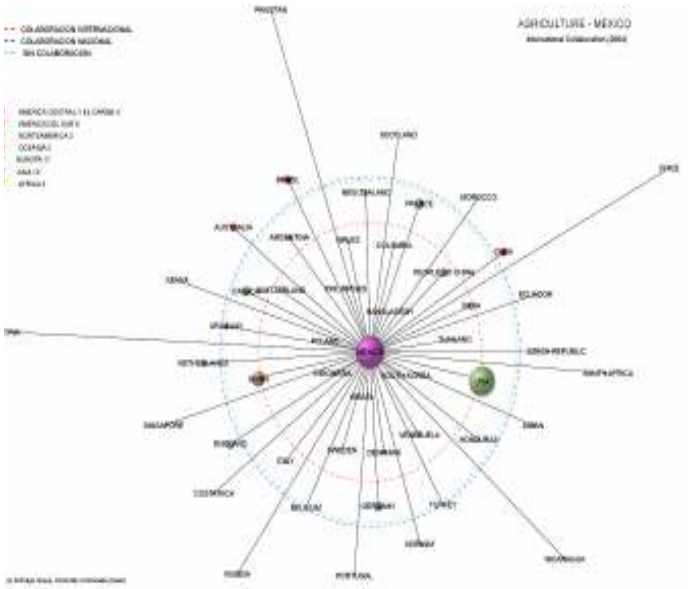
### Percentage of the thematic area with respect to the national total and rates of international collaboration (%ndoc vs % ncol)

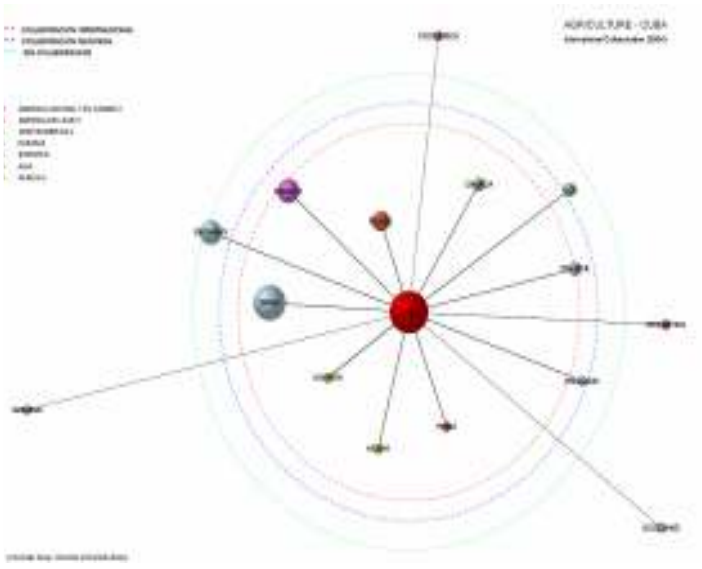
Cuba, with two times the scientific output of Brazil, presents a similar rate of overall collaboration (institutional or international). Spain lies at the other extreme: with a lesser total volume of output, it

notwithstanding shows rates just slightly lower than the other countries for the proportion of collaborative efforts. Therefore, there is a lack of correlation between the volume of production and the rate of international collaboration (Pearson coefficient r = 0.29). Meanwhile, if we compare the ratios of international collaboration in the field of agriculture with those of the country in general, Brazil and Mexico are seen to collaborate with foreign institutions well above the national average (colums % int). Yet in Cuba (and Spain, but to a lesser degree) there are areas that are "more international" than Agriculture.




**Number of countries within the realm of collaboration (countries vs ncol)**

One fairly widespread notion surrounding collaboration is that an inversely proportional relationship exists between the volume of output and that of collaboration. This is due to the need of smaller countries, with limited resources, to establish external contacts, and the contrasting capacity of the larger countries to exploit their own resources on an internal or autonomous level. For the smallest domains, then, it is particularly useful to count on the participation of larger countries with more sophisticated technology and better consolidated systems of scientific production. This hypothesis is fulfilled in the countries included in our study (r = 0.73), Cuba being the country with the least volume of output and with the

highest proportion of collaboration. Moreover, there is hardly any overlapping of Cuban documents produced in either national or international collaboration (that is, the two types of collaboration tend to be mutually exclusive); while Brazil has the highest rate of national collaboration and also the greatest overlap.

This pattern can vary, depending on the thematic areas involved, and in the case of Agriculture we might wonder why the volume of international output is related with the number of participating countries. Brazil and Spain establish the widest networks of collaboration (with 66 and 52 countries, respectively), but the intensity of relations is greater in Mexico, which is associated with fewer countries (48) yet shares in more collaborative efforts. Cuba is the country with the most limited network of associations, despite having the highest index of production in Agriculture of the four, and contrary to Brazil. These relationships are reflected clearly in the maps (lines), along with the geographic areas (color of nodes) for each country.

Asia constitutes an important scientific partner for Mexico, Spain and Brazil (respectively associating with 13, 13 and 11 Asian contries). Cuba has no bonds with Asia. Similarly, Oceania has no productive links with any the four countries studied here. Brazil is the country collaborating most with African countries (9) and with Europe (27), followed by Spain (respectively 6 and 21). Brazil also collaborates with nearly all of South America (12 countries) and Central America (4), logical in view of the geographic proximity [8, 9] as well as linguistic and cultural roots.[10, 11] Spain collaborates with 7 South American countries, but has contacts with just one Central American nation. In Mexico, likewise, the stronger relations are those held with geographically close neighbors.

**Relationship between the rates of collaboration and visibility (ncol vs fir)**

Another pattern of behavior described in the literature is that productive correlation can be seen between the impact factor of the journal of publication and the participation of more than one author (individual or institutional).[12] The greater the number of associates, the greater the impact of their documents; and moreover, the visibility obtained from foreign collaboration is supposedly greater than with merely institutional or domestic inter-institutional work [13-16]. Generally speaking, this phenomenon in confirmed by our results (Table 1: column fin-al, fin-do y fin-int-country), and the position of the orbits in the maps (according to the proximity to the central node, either red, blue or green). However, the countries with more associate authors are not the ones with the greatest visibility ($r = 0.12$).

On the other hand, in the field studied here we see an unusual inverted pattern for Spain. Note that the documents signed without collaboration achieve values of impact higher than those produced in national (institutional) collaboration, which in turn are greater in impact than the international documents. This leads us to point to Spain as a valuable potential scientific associate. This finding is supported by the fact that the countries collaborating most at the international level are Mexico and Brazil; and yet Spain, despite being the least collaborative, manages to be more visible in overall output.

With respect to the representations among the countries studied, if we comapre their positions in each map, we see with which countries the greatest visibility is attained (proximity to the central node) regardless of the volume of output (size of sphere) mentioned above. Thus we confirm that Spain is a better associate for Brazil than vice versa —the documents co-authored with Brazil are situated in the orbit of its national impact, whereas for Spain this visibility is below the minimal impact level reached in the area (periphery). With Cuba (international orbit) and with Mexico (national orbit) we see more benefits in terms of visibility.

Finally, if we focus on the international orbit, the top associates can quickly be identified. Cuba produces documents with greater impact in conjunction with Brazil, Peru, Canada, Uganda, Kenya, Spain and France. With Sweden, Scotland and Costa Rica it obtains the lowest values. Spain collaborates with 26 countries, with which it obtains a level of impact much greater than the national average. These include 14 European associates, 6 Asian ones, and 3 South American countries. For Brazil, the Central and South American associates are strong points both in production and in visibility, with the exceptions of Chile, Uruguay, Paraguay and Mexico. In Mexico, these beneficial relations with neighbors are not apparent; only Venezuela and Columbia act as visible allies. In general, the US and Spain seem to be determinant in the mean value of international collaboration.

## 4 CONCLUSIONS

The analysis of scientific collaboration affords us relational information that complements bibliometric data. Social Network Analysis, adopted in the framework of Domain Analysis[17] allows us to combine bibliometric input with the schematic display of the domain, making apparent the flows of knowledge in a way that appeals to human intuition and can be easily reproduced at any aggregate level (individual, thematic, institutional, regional, national, etc.). We believe it will prove useful for researchers and for policy makers. It stands as a strategic tool for decision-making processes regarding scientific alliances, as it sheds light on the degree of aperture of the domain at hand, its intellectual ties with neighboring terrain, and the final repercussions in terms of productivity and visibility.

At the same time, this study has led us to wonder which specialized areas are "more cooperative", what institutions are behind the collaboration and the function in the network; what sort of national or centralized policies serve to foment scientific interchange of this sort. Hopefully, we will soon have input from other sources that will help us recreate the broader context and perceive this information in more solid form.

## REFERENCES


[1] Kyvik, S. y Larsen, I. M. (1994) International Contact and Research Performance. Scientometrics 29(1):161-172.

[2] Kretschmer, H. (1993) Measurement of Social-Stratification - A Contribution to the Dispute on the Ortega Hypothesis. Scientometrics. 26(1):97-113.

[3] Ministerial Declaration on International Science and Technology Co-operation for Sustainable Development (23-24 October 2003).*Reunión de Ministros de Ciencia y Tecnología. OCDE*: DSTP/STP(2003)34

[4] Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Herrero-Solana, V., Muñoz-Fernández, F.J., Vargas-Quesada, B. (2005) Indicadores bibliométricos de la actividad científica española – 2004. Madrid: Fundación Española para la Ciencia y la Tecnología.

[5] Chinchilla Rodríguez, Z y Moya-Anegón, F. La investigación científica española (1995-2002): una aproximación métrica. Granada: Universidad, 2006

[6] Agencia Nacional de Evaluación y Prospectiva. Accesible en: http://www.mec.es/ciencia/jsp/plantilla.jsp?area=anep&id=24 Consultado 10-3-2005

[7] Melin, G. y Persson, O. (1996) Studying Research Collaboration Using Co-Authorships. *Scientometrics* 36, (3): 363-377.

[8] Bridgstock, M. (1991) The Quality of Single and Multiple Authored Papers - An Unresolved Problem. *Scientometrics* 21 (1):37-48.

[9] Katz, J.S. (1994) Geographical Proximity and Scientific Collaboration. *Scientometrics* 31, (1):31-43.

[10] Lewison, G. y Igic, R. (1999) Yugoslav Politics, Ethnic Cleansing and Co-Authorship in Science. *Scientometrics* 44,(2):183-192.

[11] Markusova, V. A. (2000) Collaboration between Russian and Western Scientists. *Second Berlin Workshop on Scientometrics and Informetrics: Collaboration in Science and in Technology*; Free University Berlin. COLLNET: Collaboration Network. (2000); 1-2 Sep

[12] Lewison, G. y Cunningham, P. (1991) Bibliometric Studies for the Evaluation of Trans-National Research. *Scientometrics* 21, (2):223-244

[13] Katz, J.S. y Martin, B.R. (1997) What Is Research Collaboration. *Research Policy* 26 (1):1-18.

[14] Narin, F., Stevens, K. y Whitlow, E.S. (1991) Scientific Cooperation in Europe and the Citation of Multinationally Authored Papers. *Scientometrics* 21, (3):313-323.

[15] Katz, J.S. y Hicks, D. (1997b) How Much Is a Collaboration Worth? A Calibrated Bibliometric Model. *Scientometrics* 40, (3): 541-554.

[16] Gómez, I.; Fernández, M.T.; Méndez, A. (1995) Collaboration patterns of Spanish publications in different research areas and disciplines. En M.E.D. Koening y A. Bookstein (eds*.) Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*. River Forest. Estados Unidos, p. 187-196.

[17] Hjorland, B. y Albrechtsen, H. (1995) Toward a New Horizon in Information-Science: Domain Analysis. Journal of the American Society for Information Science. 46(6):400-425.

[18] Moya-Anegón, F., Herrero-Solana, V., Vargas-Quesada, B.,Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F. J., Guerrero-Bote, V. , Olmeda-Gómez, C. (2004) Atlas de la Ciencia Española: propuesta de un sistema de información científica. Revista Española de Documentación Científica, 27 (1): 11-29

[19] Atlas of Science http://www.atlasofscience.net
[20] Grupo SCImago. http://www.scimago.es