Methodological Annex – Chapter 4

Analysis of scientific production based on publications in specialized journals

1. Conceptualization of quantitative indicators for the analysis of science

Quantitative indicators do not represent a "truth" regarding the condition of science and technology, but approximations to reality or an incomplete expression of it. The approach used in constructing such indicators should be comparative. Although absolute numbers are not indicative in themselves, they may be significant when suitably compared and interpreted. Nor is it possible to adopt an absolute scale, since scientific output relates to the expectations of the society in which it is developed. It is also important to avoid excessive reliance on numbers whose validity has not been sufficiently established, especially for new situations (Kondo, 1998; Trzesniak, 1998; Jarneving, 2005).

The relations between scientific production and the factors that influence it are probabilistic (stochastic) rather than deterministic, as are its results and impacts. For example, growth in scientific output by a specific country cannot reliably be attributed to any single cause. In general, such growth is simultaneously driven by various factors, including the quantity and quality of scientists, investment, public policies etc. These factors have indirect impacts and it is difficult to quantify the results of each one (Trzesniak, 1998).

Despite the importance of such indicators, the complexity of their construction and use has so far prevented the emergence of procedures with sufficient international recognition and deployment (Spinak 1998; Van Leeuwen; Moed, 2005; BIREME, 2008). On the other hand, science input indicators, albeit also complex, are well-established thanks to the following OECD guidelines: (a) the *Frascati Manual* (FECYT, 2003), which focuses mainly on human and financial resources in research and development (R&D), while also including general recommendations on S&T personnel, patents, bibliometrics, balance of payments, high-tech industries and products, and the information society; (b) the *Oslo Manual* (OECD, 2005), which deals with technological innovation indicators; and (c) the *Canberra Manual* (OECD, 1995), on human resources in S&T. More recently, the Iberoamerican-Interamerican Network of Science & Technology Indicators (Ricyt) produced a manual on suitable indicators for the real conditions prevailing in Latin American and Caribbean countries (Ricyt, 2001; Leta & Cruz, 2003). Despite the lack of widely recognized procedures for the construction of science indicators, the *Frascati Manual* recommends another OECD publication by Okubo (1997) on bibliometric indicators.

2. Characteristics of SCIE, SSCI and A&HCI databases

Science Citation Index Expanded (SCIE), *Social Science Citation Index* (SSCI) and *Arts & Humanities Citation Index* (A&HCI) are major multidisciplinary world databases originally developed by ISI and now owned by Thomson Reuters. They are the largest courses of bibliographic data on world scientific production and the most widely used to construct bibliometric science indicators (Okubo, 1997). In Brazil they are freely available to researchers with access to the CAPES journal portal (Portal de Periódicos) as *Web of Science*, which contains data for more than 39 million publications and their respective citations (Thomson Reuters, 2008a). SCIE indexes more than 6,650 journals, adding some 19,000 new records per week. It covers current data as well as partial retrospective data since 1990, mainly in exact, biological and earth sciences. SSCI indexes 1,950 complete journals and parts of 3,500 other publications, adding some 2,900 records per week and covering partial retrospective data since 1956. A&HCI indexes some 1,160 journals and parts of 6,800 other publications, adding 2,300 new records per week and covering partial retrospective data since 1975 (Thomson Reuters, 2008a).

In aggregate, SCIE, SSCI and A&HCI are multidisciplinary in their coverage, but it is difficult to discern a balance in terms of the representativity of the various knowledge areas owing to differences in habits and traditions with regard to publication and citation, as well as language barriers. Physics, chemistry and biomedicine are considered well covered, whereas there are problems in geosciences, biology, engineering, mathematics and other areas (Nederhof, 2006). Because English is the predominant language of indexed publications, moreover, there are grounds for the interpretation that publications from non-Anglophone countries are under-represented, especially in the humanities (Leta, Glanzel & Thijs, 2006). Although journal articles predominate, a wide variety of documents are indexed, including reviews, abstracts, editorials and, in smaller numbers, books, musical compositions and art works, among others. However, the indicators analyzed in this chapter were constructed using only articles, letters, notes and *reviews* (NSB, 2002, 2004, 2006; European Commission, 2003), potentially entailing under-representation of knowledge areas in the humanities in which authors prefer to publish books (Prat, 1998; Spinak, 1998; Macias-Chapula, 1998; Targino & Garcia, 2000). Knowledge areas were not compared directly because of the unbalanced coverage of publications by knowledge area, the predominance of English, and the prevalence of articles among indexed publications, among other reasons. However, it is valid to compare publications from different countries in specific knowledge areas and to analyse the growth in publications by knowledge area over time (Faria, Gregolin & Hoffmann, 2007).

Products developed using SCIE and SSCI for the analysis of scientific production include: (a) *National Citation Reports* (NCR), an electronic database of publications for specific countries, containing bibliographic information for each publication, as well as the total and annual citations to each publication (Thomson Reuters, 2009a); (b) *Journal Citation Reports* (JCR), which focuses on assessments and comparisons of journals indexed by SCIE and SSCI based on the citations received by the articles they contain, including journal rankings by absolute numbers of citations and impact factors, an indicator as valued by the scientific community as it is controversial (Thomson Reuters, 2009b; Dong, Loh & Mondry, 2005); and (c) *Essential Science Indicators* (ESI), which contains statistics by country, institution, researcher, journal and knowledge area on SCIE- and SSCI-indexed journals and citations for the past ten years (Thomson Reuters, 2009c).

In conjunction, the SCIE and SSCI databases and the analytical products derived from them offer positive features for the construction of indicators, including multidisciplinary coverage, detailed bibliographic records, identification of all co-authors, availability of citation data, and a large number of indexed publications by Brazilian authors. The number of publications indexed by SCIE and SSCI for 2006 with at least one researcher affiliated to a Brazilian institution among their authors totals 22,410 (FAPESP, 2005; Faria, Gregolin & Hoffmann, 2007).

There are also limitations: the databases were created to disseminate science and not as an aid to the construction of indicators, although they are widely used for this purpose (Adam, 2002); they focus on mainstream world science and tend to leave subjects of regional interest aside (Spinak, 1998); the distribution of journals by knowledge area does not correspond to the global distribution (Leta & Cruz, 2003); and they are biased in favour of English-language journals (Luwell, 1999; Velho, 2008).

Despite these limitations, until recently no other databases were available with similar attributes for the production of bibliometric indicators and with the same combination of quantity and coverage for Brazilian publications. Now, however, there is Scopus, launched in 2004 by Elsevier with the aim of serving as an alternative for the construction of indicators.

3. Characteristics of Scopus, SciELO and selected specialized databases

It is important to use multiple databases when constructing bibliometric indicators, since no single database provides complete coverage of the world's scientific literature and each one has deficiencies and limitations. Thus, whenever possible it is valid and recommendable to use various different sources, even though this increases the complexity of the analysis and may produce partially discordant results (Okubo, 1997). Scopus, SciELO, Compendex, Inspec, PubMed, *Biological Abstracts* and *Sociological Abstracts* are some of the databases that can be used to produce bibliometric indicators. They can be classified into two groups: (i) Scopus and SciELO are multidisciplinary, include information on the affiliations of all co-authors and quantify citations to indexed publications; (ii) Compendex, Inspec, PubMed, *Biological Abstracts* and *Sociological Abstracts* specialize in some knowledge, display affiliation information only for first-named authors, and do not quantify citations. Scopus and SciELO can be used for broad concurrent studies of several knowledge areas, analysis of collaboration between institutions and countries, and citation studies. The specialized databases can be used for in-depth analysis of the specific knowledge areas they cover. More information on the databases mentioned is set out below.

Scopus. Produced since 2004 by Elsevier, Scopus is a multidisciplinary, multiaffiliation, multilanguage bibliographic database with significant coverage of world scientific literature. It has so far indexed some 33 million publications, of which 16 million since 1996, when more consistent coverage of journals began (Elsevier, 2008a). The indexed publications come from more than 15,000 journals – some 3,400 in biological sciences, 5,300 in health sciences, 5,500 in exact sciences, and 2,800 in human sciences. More than 1,000 of the indexed scientific journals are open-access, including the titles in SciELO. Scopus also covers monographs, conference proceedings, patents, trade magazines, and scientific web pages selected according to quality standards. Like SCIE and SSCI, it contains citations made and received by indexed publications, enabling its use for the construction of citation indicators (De Moya-Anegón et al., 2007). It also offers features that facilitate both searches and production of bibliometric indicators, such as Citation Tracker, to investigate citations received by publications, authors and institution; *Author Identifier*, to identify and standardize author names; and *Affiliation Identifier*, to identify and standardize institution names, and to calculate the *h*-index, an indicator based on the number of publications and the frequency with which they are cited and recently valued to represent the influence of journals, researchers, institutions and countries (Elsevier, 2008b; Van Raan, 2006; Vinkler, 2007). The SCImago Journal & Country Rank (http://www.scimagojr.com) is a portal offering unpaid access to journal and country visibility indicators developed by SCImago, a research group at the University of Granada, Spain, from the information contained in Scopus, including several types of indicators relating to publications, citations and networks etc. for the analysis of journals and countries over time and by knowledge area. Scopus can be accessed via the CAPES journal portal (Portal de Periódicos).

SciELO. SciELO is an electronic library launched in Brazil in 1996. Its Brazilian collection, **SciELO Brazil**, used for this chapter, covers a selection of scientific journals and offers access to full texts. It is the result of a collaborative project involving FAPESP, the Latin American & Caribbean Center for Health Science Information (BIREME) and more recently the National Council for Scientific & Technological Development (CNPq). It also offers a methodology for the preparation, storage, dissemination and evaluation of scientific production in electronic format (SciELO, 2008). SciELO has expanded internationally and there are now national collections in 11 other Latin American and Caribbean countries, as well as Spain and Portugal. It also has thematic collections in public health and social sciences administered by BIREME. Among the main results of the creation of SciELO Brazil are increased dissemination of Brazilian scientific production and enhancement of its visibility and credibility at home and abroad (Packer, 1998; FAPESP, 2002, 2005; SciELO, 2008). It contains some 90,000 publications from 209 Brazilian journals in agricultural sciences, biological sciences, chemistry, engineering, geosciences, health sciences and human sciences. It has expanded rapidly and now contains a significant number of publications for the analysis of Brazilian scientific production, albeit not yet as many as SCIE and SSCI or Scopus. For 2006 SCIE and SSCI contain some 19,000 Brazilian publications, Scopus some 23,000 and SciELO Brazil some 12,000. SciELO's sites offer a number of quantitative reports on indexed publications and citations received, as well as access to publications, all of which is very useful for the construction of indicators.

However, there are difficulties that need to be understood. Because it is an eminently national database – about 85% of the publications in SciELO Brazil had at least one Brazilian author in 2006 – its use for positioning Brazilian scientific production compared with other countries is limited. SciELO's regional portal, launched in 2008, integrates indicators for all national and thematic collections, and in future will enable analysis of regional collaboration. The fact that the retrieval of data from searches of SciELO occurs via bibliographic references rather than bibliographic records, in which data are presented in defined fields, prevents the use of custom bibliometrics software, but as coverage of SciELO's journals by Scopus proceeds it will shortly be possible to retrieve bibliographic records via this database. A consequence of this, for example, will be the possibility of producing indicators for knowledge areas more quickly. The growing volume of data, diversity of languages, coverage focusing on national journals and quality of indexation and standardization make SciELO an important source for the production of indicators on Brazilian scientific production, with a tendency for this importance to increase as the collection grows. SciELO can be accessed via its regional portal or via the CAPES journal portal (Portal de Periódicos).

Compendex. Produced by Engineering Information Inc., currently part of the Elsevier group, Compendex contains some 10 million bibliographic records from 1969 to present. It covers articles, books, journals, trade magazines and conference proceedings in engineering and technology, serving as a resource for current awareness, new product information, technological forecasting and competitive intelligence. It of-

1. Dynamic controlled vocabulary of descriptors that are semantically and generically related, covering a specific knowledge area.

fers an online thesaurus[1] to permit searching or browsing of more relevant sets of records and thus improve search results. Compendex can be accessed via the CAPES journal portal (Portal de Periódicos) (Engineering Information, 2008a, 2008b).

Inspec. Produced by the Institution of Engineering and Technology (IET), Inspec contains some 9 million bibliographic records for articles, books and other documents from 1969 to now in physics, electrical engineering and electronics, computers and control, information technology for business, and mechanical and production engineering. Inspec can be accessed via the CAPES journal portal (Portal de Periódicos) (Ovid, 2008).

PubMed. Maintained by the U.S. National Library of Medicine (NLM) at the National Institutes of Health (NIH), PubMed is a bibliographic database with over 18 million indexed publications in journals covering health and biological sciences, including 16 million publications in the MEDLINE database (NLM, 2008a, 2008b). PubMed can be accessed via the CAPES journal portal (Portal de Periódicos), via OVID and via BIREME's Virtual Health Library.

Biological Abstracts. Produced by Thomson Reuters, *Biological Abstracts* indexed more than 11.3 million world publications in life sciences (Thomson Reuters, 2008c) and can be accessed via the CAPES journal portal (Portal de Periódicos).

Sociological Abstracts. Produced by CSA, *Sociological Abstracts* indexes world literature in sociology and related areas, covering more than 1,800 journals as well as books, abstracts, dissertations and conference papers from 1952 to now, with more than 840,000 records. *Sociological Abstracts* can be accessed via the CAPES journal portal (Portal de Periódicos) (CSA, 2008).

4. Complexity of using science citation indicators

Despite growing interest in science citation indicators, several factors make their construction and use complex, especially the following (Okubo, 1997; Adam, 2002; Macias-Chapula, 1998; Velho, 2008): (a) differences between journals in terms of audiences, staff sizes, language barriers, procedures, cultures and visibility by knowledge area and market; (b) citations for reasons not related to the significance of a publication, such as tactics, tributes to eminent scholars, references to methodologies, criticism of errors, controversy, self-citation etc.; (c) reviews and short communications, which do not represent original contributions and may bias the indicators; (d) technical errors in the sources or in information processing.

With regard to knowledge areas, for example, articles in biomedicine are more frequently cited than articles in mathematics or medicine (ADAM, 2002). A study of citations to journals covered by ISI found that 55% of the papers published between 1981 and 1985 in journals indexed by the database received no citations at all in the five years after they were published (Hamilton, 1990). Another study by the same author demonstrated non-citation in the five years after publication for more than 72% of papers in engineering, 75% in social sciences, and 92%-99.9% of most papers in most arts and humanities disciplines (Hamilton, 1991).

Recent research also shows that for various reasons not associated with the quality of their work scientists in peripheral countries publish in local journals and thus have less visibility and are less cited. The main reasons are language barriers, the expense of international publication, a nationalist wish to strengthen domestic journals, a target audience who does not read foreign journals, lack of experience as a student abroad, and a national or regional research focus (Velho, 2008).

It is worth noting that there are two strongly opposed schools of thought about citations. Some experts argue that citation analysis contributes to objective decision making. Others, however, claim exactly the opposite, pointing to errors and deviations associated with the interpretation of citation indicators (Adam, 2002) and the difficulties of using SCIE and SSCI, the only databases that could be used until 2004, when Scopus was launched, and still the most widely used in citation research.

5. Methodology of boxes on research networks (section 5)

5.1 Box 4.2: Nanotechnology research networks

The data presented in this box on Brazilian scientific publications in nanotechnology and nanoscience come from *Web of Science* (WoS), which is accessible via the CAPES journal portal (Portal de Periódicos). They were extracted from a larger dataset relating to the entirety of Brazil's scientific production indexed by WoS, retrieved using the search expression "AD=brazil" for the period 2002-06 and prepared specifically for the production of indicators, first by selecting 519 publications for the period classified under the category "Nanoscience & Nanotechnology". The data were treated and quantified using VantagePoint bibliometric software to generate a co-occurrence matrix of institutions to which the authors of the publications in question were affiliated. The matrix was processed using Ucinet and NetDraw to map the collaboration network presented in Chart 4.1.

Data on nanotechnology patents generated in Brazil were collected from Derwent Innovations Index, also accessible via the CAPES journal portal. Of the 101 patents analyzed, 73 were issued to Brazilian organizations. The other 28 are registered only in the name of private individuals. The search expression used was "(IP=(B82*) OR TS=(nano* NOT (nanometer* OR nanosecond* OR nano2 OR nano3))) AND PN=(BR*)." The search was restricted to the period 1998-2006. Patents were then selected with only Brazil as the priority country. The data were treated in a similar manner to those for scientific publications, resulting in a co-occurrence matrix of patent holders and the patent collaboration network presented in Chart 4.2.

5.2 Box 4.4: Sugarcane genomics and breeding scientific cooperation network

The data presented in this box on scientific publications in sugarcane breeding come from *Web of Science* (WoS), which is accessible via the CAPES journal portal (Portal de Periódicos). The following parameters were used to search for publications: (a) the expression TS=(Sugarcane OR "Sugar Cane" OR Saccharum) AND TS=(Gene OR EST OR cDNA OR Genome OR Genomic OR Transcriptome OR array OR chip OR Genetic OR Marker OR polymorphism OR expression OR SUCEST); (b) year published: 1998-2006; (c) databases selected: SCIE and SSCI; (d) document types: articles, letters, notes and reviews. The search returned 690 publications, of which 139 had at least one author affiliated with a Brazilian organization. The data were treated and quantified using VantagePoint bibliometric software to generate a co-occurrence matrix of institutions to which the authors of the publications in question were affiliated. The matrix was processed using Ucinet and NetDraw to map the collaboration network presented in Chart 4.4.

5.3 Box 4.5: Theoretical and conceptual aspects

The description of these networks is part of a broader context of S&T research and was strongly based on the work of Dal Poz (2006), where the discussions mentioned are set out in detail.

The analytical tools used for the study included work on sociotechnical networks, which seeks to integrate economic and sociological views with technological production. The conceptual approach included a definition of the theoretical subfields that involve the study of networks and the dimensions considered in the analysis presented in this box, especially the borders of the networks in question.

The literature associated with these studies conceptualizes networks as the result of dynamic evolutionary processes, which can best be understood from a perspective that takes into account the following elements:

- A social element, involving individuals, public and private organizations, and interactions among these individuals and organizations in specific contexts;
- An economic element, involving the interactions and transactions that transform knowledge and resources in order to achieve certain benefits;
- A technological element, involving technology and technology transfer to obtain competitive advantages;
- A sociotechnical element, involving the ways in which society and technology influence each other.

This is the context for analyzing the evolution of sociotechnical networks and the social actors involved in the innovation process. In particular, these elements can be used to identify and define the actors and their roles in technological arrangements independently from the geographical dimension.

Interest in the use of this approach has increased for at least two reasons. One is the emergence of new organizational forms of structuring industries, which stress elements such as the horizontal and lateral links between firms. The other is the emergence of new technologies such as IT that make possible less rigid arrangements of organizational structures (Nohria & Eccles, 1992).

Network "theory" seeks to explain the relations between social and economic behavior from the standpoint of technological development. The approach is therefore useful for the construction of indicators capable of casting new light on intrinsic relations that are hard to measure with traditional S&T indicators.

The literature on networks displays a persistent distinction between the economic and sociological viewpoints, although they are increasingly brought together analytically and conceptually.

A number of key bibliographic references presented at the end of this chapter can be considered the academic basis for the context of this box.

The sociology of networks (Callon, 1992; Larédo, 1998; Larédo, Mustar & Callon, 1993; Bell & Callon, 1994) is the main approach used in the study, although economics and research on national innovation systems are also important.

5.3.1 Algebraic indicators

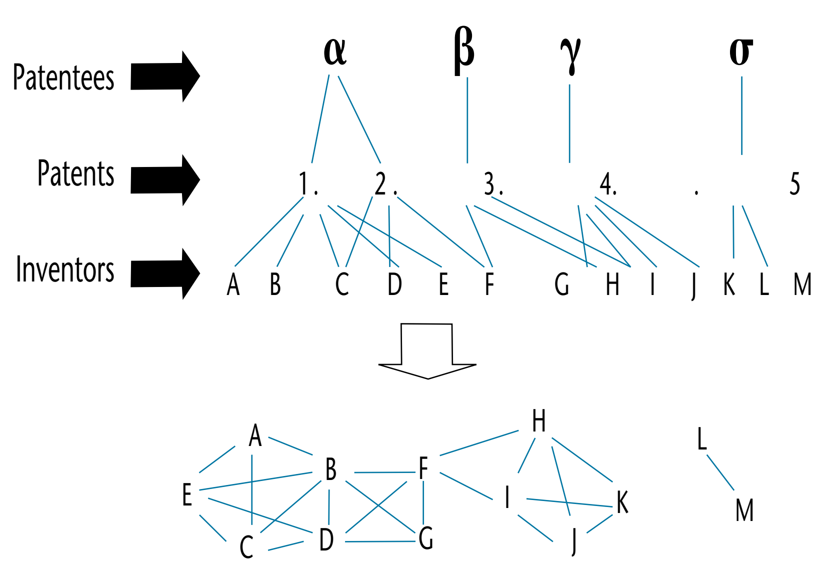
The relations among actors in a network can be expressed visually with the aid of specialized social network analysis software such as Pajek or UCINET, or in algebraic form. The most frequently used indicators – density, centrality and betweenness or connectivity – are defined below.

Algebraic representation starts with the positions of the actors in a square matrix, with citers in the columns and cited authors in the rows. Such matrices link pairs of elements as shown in Chart M4.1 and Table M4.1 to plot the citations (marked 1) received by a hypothetical patent X by later patents and non-citations (marked 0).

In a matrix representation, k represents the k^{th} generic network being analyzed, and $z\ ijk$ represents the value of the relationship between the i^{th} actor and the j^{th} actor ($z\ ijk \geq 1$ if the i^{th} actor is connected to the j^{th} actor; *otherwise*, $z\ ijk = 0$).

The frequency with which a patent is cited by later patents is an important indicator of its market value. Thus the results of the analysis of a patent's citations by later patents are used to investigate the history of some of the outputs obtained by the most relevant ac-

Graph M4.1
Graph of binary relationships among patentees, patents and inventors and geodesic representation of network nodes



Note: Based on Figure 1, "Bipartite graph of patents and inventors," in Balconi, Breschi & Lissoni (2004).

Table M4.1
Matrix of later citations of a patent by different inventors

Patent X Patentee Y	Matrix of later citations of a patent by different inventors				
	Patentee A	Patentee B	Patentee C	Patentee D	Patentee E
Later citation (cited) 1	0	1	1	0	1
Later citation (cited) 2	0	0	1	1	1

Source: Dal Poz (2006).

tors in the network, i.e. the actors who stand out for the frequency with which they are cited.

The assumption is that inventors A to M are interconnected by the knowledge embodied in patents, although the degrees of relationship among them vary.

5.3.2 Network density

Network density is an indicator of the degree of dynamism of a network rather than of any of its individual members. It is calculated by dividing the number of ties in a generic network k by the number of possible ties minus the relationship of each node with itself, N^2-N. Density values range between 0 for no connections, and 1 for all the possible ties among members of a network. Because only links between different actors are considered, self-referential links (between an actor and itself) are discounted. In the case in question, there are no co-authorships between researchers and themselves.

The density of this network is 0.1549, meaning that the existing links represent 15.49% of the possible relationships. Evidently not all institutions are expected to collaborate with all others, which would give a density equal to 1.0, the highest possible value. Algebraic values are not useful when analyzed in isolation, as explained below.

The utility of density as an indicator increases in proportion to the amount of comparative evidence obtained and depends on a comparison with other indicators, including the most traditional. Moreover, it makes more sense when used as one of a group of analytical comparatives. An example may illustrate this idea. Certain knowledge areas, such as the life sciences, require collaboration because of the growing complexity of research and the need for certain competencies to be introduced in order to achieve “publishable” results. In such a case, if after a series of actions to develop specific areas of S&T the classic indicators of scientific production in terms of publication do not respond positively, taking into account the temporality issues that must be observed in the analysis, then collaboration density may point to the path that policy making should take by demonstrating the need to marry research funding policy with collaboration.

Thus the algebraic value of this network’s density can strengthen the analysis, for example, of indicators for Brazilian scientific output in terms of publications in the same period, especially if there is a need for collaboration to bolster the dynamics of the area concerned.

This indicator measures the degree of involvement of the actors (in this case, co-author institutions) with other actors in the same network. It is given by the algebraic value of all contacts involving actor a divided by all real contacts found in the network.

An analysis of network configuration and algebraic indicators for the intrinsic relationships among the actors present in the network can show that networks have limits, that they involve certain actors and not others, and that the legal framework for S&T policy making and the factors influencing the formulation and implementation of S&T policy make a difference in terms of helping to determine whether certain actors are or are not included in the dynamics of world technology appropriation.

5.3.3 Connectivity

Connectivity or betweenness is an indicator of the frequency of relationships used to measure the proximity between actors, as well as the importance of network actors in terms of their number and location.

When networks are analyzed, the distance between actors is given not by geography but by the intensity of their relationships. Thus the more connected they are, the less will be the social distance between the actors in the network. When institutions that publish scientific articles in collaboration are mapped, the closest are those with the strongest interaction (i.e. those that publish most frequently as co-authors).

Location within the network also has significance. Actors positioned at or close to the “core” of the network are those with the largest number of ties to other actors and can therefore be considered most important or dynamic. Actors with relatively few connections are on the network periphery.

Within the group of actors classified as most important or dynamic, however, there is a crucial distinction: some actors have ties with others that are also dynamic, and these are central to the network, whereas some actors have ties with others that are not dynamic, and these are positioned between the core and the periphery.

5.3.4 Geodesic distance

This indicator is defined by the number of edges that separate two distinct actors in a network. For example, in the scientific collaboration co-authorship network, USP has a direct tie to Unicamp. The geodesic distance between these two institutions is 1. Unicamp has a direct tie to Iapar. However, Iapar does not have direct ties to USP, although they do relate via Unicamp. The geodesic distance between USP and Iapar is therefore 2.

An important observation should be made about the use of this indicator. It is necessary to take into account whether the orientation of the links between the vertices is specified or directed, i.e. whether the links are arcs or arrows rather than mere lines. Depending on the nature of these relations, the indicator may need to be adjusted.

5.3.5 Centrality of actors

Some actors stand out for the number of ties they have with other actors in the same network. These can be said to display a high degree of centrality. Other central actors may not have as many ties, but their positions in certain sectors of the network means they play the role of connectors between different components or subsectors, so that they can be said to have betweenness centrality.

Centrality shows how central certain actors in a network are in terms of their high degree of involvement with other actors in the same network. It is calculated by dividing all the contacts involving actor a by all the real contacts found in the network. Like density, centrality ignores self-reference, since it is valid only for cases in which a ≠ i. Furthermore, by the same argument as for density, it is necessary to make adjustments to this indicator.

If the ties are directed, centrality is configured as follows:

$$C_a = \frac{\sum_{i=1}^{N} (Z_{ia} + Z_{ai})}{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ijk}\right)}$$

$i \neq a$ (where i means all other actors except a).

Thus the centrality of actor a is measured by dividing the number of ties from this actor to other actors and from other actors to this actor (direction matters in this case) by the total number of ties in network k as a whole.

If direction does not matter, centrality is as follows:

$$C_a = \frac{\sum_{i=1}^{N} (Z_{ia})}{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ijk}\right)}$$

$i \neq a$ (where i means all other actors except a).

Thus the centrality of actor a is measured in this case by dividing the number of ties between this actor and other actors (whatever the direction) by the total number of ties in network k as a whole.

The following table presents some centrality indicators for the network.

Given the temporal search window deployed, it can be inferred that certain S&T funding programs, such as the Genome Projects launched in 1997 (FAPESP's *Xylella Genome and Brazilian Genomes*), resulted in collaborative scientific production after almost a decade. The geographic overlap of research funding and development policies is clear: most of the institutions in the above list are in São Paulo State and their aggregate centrality matches the state's predominance in the network's activities. Centrality also serves to measure the importance and roles of specific institutions in these situations.

5.4. Box 4.6: Biophotonics research networks

The study of biophotonics capabilities and networks in Brazil was based on a sample survey that showed the existence of a productive and active community with many multidisciplinary collaborations in R&D, including ramifications in the business sector. This demonstrates the area's strong potential in Brazil.

Capabilities, competencies and collaborative networks were detected by means of searches in *Web of Science* using the keywords shown at the end of this section (Chart M4.1) and the word "brazil" as part of the institution's address. The keywords were extracted from *Handbook of Biomedical Optics* (http://www.crcpress.com/product/isbn/9781420090369); and *Introduction to Biophotonics* by Paras Prasad (http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471287709.html) so as to guarantee coverage of all the main areas of biophotonics. The searches returned 1,730 publications. The period was 1995-2006, which includes all years covered by *Web of Science*. No records were found of publications in the area prior to 1983.

Publications with overseas reprint addresses were removed, leaving 1,388, of which 1,080 had only Brazilian addresses. To investigate the location of capabilities and competencies, addresses were broken down to the department level, adding data for similar entities in São Paulo's state universities (instituto, faculdade) and for their equivalents at federal institutions, where departments are called institutos. For example, a physics department at a federal university corresponds to a physics institute at a state university in São Paulo. However, an anatomy department, for example, does not appear as an isolated unit but as part of a school (faculdade) of medicine. Non-university institutions, such as Instituto Ludwig or IPEN, for example, appear as units. Chart M4.2 shows the level of aggregation for the representative units in this study. It lists the institutions in São Paulo with more than 1% of the 1,388

Table M4.2
Centrality of key institutions ñ Brazil, 2000-2007

Institution		Centrality index
1	Universidade de São Paulo – USP	0.1096
2	Fundação Oswaldo Cruz – Fiocruz	0.0559
3	Universidade Estadual Paulista – Unesp	0.0669
4	Universidade Federal do Rio de Janeiro – UFRJ	0.0417
5	Universidade Estadual de Campinas – Unicamp	0.0461
6	Universidade Federal do Rio Grande do Sul – UFRGS	0.0406
7	Universidade Federal de São Paulo – Unifesp	0.0230
8	Universidade Federal de Minas Gerais – UFMG	0.0164
9	Empresa Brasileira de Pesquisa Agropecuária – Embrapa	0.0241
10	Universidade Federal do Paraná – UFPR	0.0252
11	Hospital A.C. Camargo	0.0230
12	Instituto Ludwig de Pesquisa sobre o Câncer	0.0307
13	Universidade de Brasília – UnB	0.0121
14	Instituto Butantan	0.0121

Source: Dal Poz (2006).

publications, along with the percentage of publications in which their names figure among the co-authors' institutions of origin. The criterion used to identify competencies in the area, especially in São Paulo State, considering the 1,080 publications that had Brazilian authors only, was to exclude institutions outside São Paulo with less than 1% and include all institutions in the state. This left 165 institutions, which were analyzed using Pajek social network analysis software. The large number of publications and institutions involved demonstrates the size of the biophotonics sector in Brazil today.

Chart M4.1
List of keywords used to identify competencies in biophotonics ñ Brazil

Keywords used to identify competencies in biophotonics
biophotonic (1)
laser and therap (1)
laser and surger (1)
endoscop (1) and surgery
laser and cardiovasc (1)
image guided surgery
minimally invasive surgery
laser and angioplast (1)
lasik
laser and imaging
laser and (medicin (1) or biology (1) or lifescienc (1))
optics and (biomedic (1) or medicin (1) or biology (1) or lifescienc (1))
optic (1) and tweezers
optical and diagnos (1)
optic (1) and glucos (1) and sensor
optic (1) and biosensor (1)
optic (1) and blood
optical biopsy
optical and tomograph (1)
optical and coheren (1) and tomograph (1)
two-photon (1) and microscop (1)
fluoresc (1) and biomedic (1)
fluorescen (1) and tag (1)
fluoresc (1) and DNA
biochip (1) or microarray (1)
flow and cytomet (1)
genomics or proteomics
quantum and dot (1) and colloidal
raman and (biomedic (1) or medicin (1) or biology (1))
photodynamic (1) and therap (1)
optic (1) and near-field and (snom or nsom)

(1) Indicates that the search captured different forms of the root word.

Table M4.3
Web of Science-indexed publications on biophotonics by institution in São Paulo State, 1983-2006

Position	Institution	Publications on biophotonics	
		Abs. nos.	%
1	Unifesp - Medicina - São Paulo - SP	171	12.32
2	USP - Medicina - São Paulo - SP	134	9.65
3	USP - Medicina - Ribeirão Preto - SP	87	6.27
4	USP - Biologia - São Paulo - SP	83	5.98
5	Unicamp - Medicina - Campinas - SP	82	5.91
6	USP - Química - São Paulo - SP	78	5.62
7	Univap - IPD - São José dos Campos - SP	42	3.03
8	USP - Física - São Carlos - SP	33	2.38
9	Unicamp - Biologia - Campinas - SP	33	2.38
10	USP - Química - São Carlos - SP	29	2.09
11	USP - Química - Ribeirão Preto - SP	28	2.02
12	USP - Farmácia - Ribeirão Preto - SP	27	1.95
13	Unicamp - Física - Campinas - SP	25	1.80
14	Hosp. AC Camargo - Medicina - São Paulo - SP	25	1.80
15	Inst. Ludwig - Medicina - São Paulo - SP	22	1.59
16	USP - Odontologia - São Paulo - SP	18	1.30
17	USP - Matemática - São Paulo - SP	18	1.30
18	Butantan - São Paulo - SP	18	1.30
19	USP - Farmácia - São Paulo - SP	17	1.22
20	USP - Biologia - Ribeirão Preto - SP	16	1.15
21	Unesp - Biologia - Botucatu - SP	16	1.15
22	IPEN - São Paulo - SP	15	1.08
23	USP - Física - São Paulo - SP	14	1.01
24	Hosp. Albert Einstein - Medicina - São Paulo - SP	14	1.01

Source: Fragnito et al. (2007).

Chart M4.2
Detailed legend to Chart 4.7

1 Unifesp – SP
2 USP – Med – SP
3 Unicamp – Med – Camp
4 USP – Med – Rib Preto
5 UFRJ – ICB – RJ
6 USP – Quim – SP
7 Fiocruz – RJ
8 UFRJ – Biofis – RJ
9 USP – ICB – SP
10 Univap – IPD – SJCamp
11 UFMG – Med – BH
42 UFSC – Biol – Florianópolis
43 UFMG – Fis – BH
44 USP – Fis – SP
45 USP – Bioinformática – SP
46 Unesp – Vet – Jaboticabal
47 UMC – Biol – M Cruzes
48 UFSCar – Biol – S Carlos
49 Butantan – SP
50 USP – Fis – R Preto
51 Unicamp – Quim – Camp
52 Unesp – Fis – SJR Preto
83 UFPE – Fis – Recife
84 UFPE – Biofis – Recife
85 UFOP – Farm – Ouro Preto
86 UFMG – Vet – BH
87 UFGO – Quim – Goiânia
88 UFGO – Med – Goiânia
89 UFCE – Med – Fortaleza
90 Senac – SP
91 Santa Casa – Med – SP
92 PUC – RS – Odonto – POA
93 H Serv Estado – SP
124 Onze Junho – SP
125 I S Abujamra – SP
126 I ProGastro – Campinas
127 I Pasteur – SP
128 I Onco Pediat – SP
129 I L Souza Lima – SP
130 I Emilio Ribas – SP
131 IB Cont Canc – SP
132 IB Cont Canc – SP
133 I Biológico – SP
134 IAC – Campinas

(CONTINUED ON NEXT PAGE)

Chart M4.2 (continued)
Detailed legend to Chart 4.7

12 Fiocruz – MG	53 Uerj – Med – RJ	94 USP – Cena – Piracicaba	135 H Socorro – SP
13 Unicamp – Bio – Camp	54 UEL – Biol – Londrina	95 Unesp – Odont – SJ Camp	136 H S Marcelina – SP
14 USP – Quim – S Carlos	55 LNLS – Camp	96 Unesp – Farm – Araraq	137 H S Helena – SP
15 USP – Fis – S Carlos	56 USP – Vet – SP	97 UFSCAR – Quim – S Carlos	138 H P Biygton – SP
16 Unicamp – Fis – Camp	57 USP – Odonto – R Preto	98 USP – E Elét – S Carlos	139 H João XXIII – SP
17 UFRGS – Med – POA	58 USP – Odonto – Bauru	99 USF – Med – B Paulista	140 H Iamspe – SP
18 UFMG – Biol – BH	59 USP – ESALQ – Piracicaba	100 UFSCar – Fis – S Carlos	141 H Equinos SGFoz – SP
19 Inst Nac Câncer – RJ	60 Unicamp – E Elét – Camp	101 Cons Retina – Araraq	142 H E Vasconcelos – SP
20 USP – Quim – R Preto	61 Unesp – Biol – SJR Preto	102 USP – Matem – S Carlos	143 H Benef Port – SP
21 Inst Ludwig – SP	62 UNB – Fis – DF	103 USP – Enferm – R Preto	144 H A Carvalho – SP
22 H AC Camargo – SP	63 UFPE – Odonto – Recife	104 USP – Comput – SP	145 H A Carvalho – Jaú
23 USP – Farm – R Preto	64 UENF – Biol – Campos – RJ	105 USP – Comput – S Carlos	146 H Oswaldo Cruz – SP
24 UFBA – Odont – Salvador	65 Fiocruz – BA	106 Unimep – Piracicaba	147 Gatromed – SP
25 USP – Biol – SP	66 USP – SP	107 Unimar – Odonto – Mar	148 Fundecitrus – Araraq
26 UNB – Biol – DF	67 Unesp – Odonto – Araraq	108 Unicamp – E Biomed	149 F ONCOCTR – SP
27 UFMG – ICB – BH	68 UNB – Med – DF	109 Unicamp – Comput	150 F ProSangue – SP
28 USP – Odont – SP	69 UFRJ – Farm – RJ	110 Unesp – Biol – R Claro	151 Fifty Med Res – SP
29 UFRGS – Biol – POA	70 Santa Casa – POA – RS	111 Laser Vis Cumica – SP	152 F MCS Vidigal – SP
30 IPEN – SP	71 Lab Fleury – SP	112 I Invest Imunol – SP	153 F ABC – Med – SP
31 USP – Matem – SP	72 IAC – Cordeirópolis – SP	113 I Brugnera – SP	154 Exército Bras – SP
32 USP – Biol – R Preto	73 H Sírio-Libanês – SP	114 H Oswaldo Cruz – SP	155 Embrapa – R Preto
33 UFRJ – Med – RJ	74 Excimer Las HS Cruz – SP	115 H Olhos Araraquara	156 Embrapa – Cordeirop
34 Uerj – Biol – RJ	75 Unicamp – Odonto – Pirac	116 H Heliópolis – SP	157 Centro Universitário Salesiano – Lorena
35 USP – Farm – SP	76 Unesp – Quim – Araraq	117 Fatec – SP	158 Copersucar – SP
36 I Adolfo Lutz – SP	77 Unesp – Med – Botucatu	118 Eye Clin Day H – SP	159 Clin Miyake – SP
37 H Albert Einstein – SP	78 UFV – Biol – Vicosa	119 Embrapa – S Carlos	160 Cetesb – SP
38 Embrapa – DF	79 UFU – Quim – Uberlândia	120 C Odonto S Leop – Camp	161 AB Laser Cir – SP
39 Unesp – Biol – Botucatu	80 UFRJ – RJ	121 Alellyx – Campinas	162 Grp Lapar&Cir – SP
40 UFBA – Med – Salvador	81 UFRJ – Fis – RJ	122 Skopia Endo – SP	163 C Ocular Laser ABC – SP
41 I Olhos Goiânia – Goiânia	82 UFRJ – Biol – RJ	123 S Paulo ENT Clin – SP	