

Anexo Metodológico – Capítulo 4

Análise da produção científica a partir de publicações em periódicos especializados

1. Conceituação sobre indicadores quantitativos para análise da ciência

Os indicadores quantitativos não representam uma “verdade” sobre o estado da ciência e da tecnologia, mas aproximações da realidade ou uma expressão incompleta dela. A abordagem dos indicadores deve ser comparativa. Embora os valores absolutos não sejam indicativos em si, podem ter significado quando são feitas comparações e interpretações. Também não é possível ter-se uma escala absoluta, porque existe uma relação da ciência produzida com as expectativas da sociedade na qual ela se desenvolve. Deve-se, inclusive, evitar o excesso de confiança em números de validade insuficientemente estabelecida, principalmente para novas situações (KONDO, 1998; TRZESNIAK, 1998; JARNEVING, 2005).

As relações de causa e efeito da produção científica com os fatores que a influenciam são probabilísticas (estocásticas), e não determinísticas, o mesmo ocorrendo nas relações com os efeitos resultantes. Por exemplo, um aumento da produção científica de um país dificilmente pode ser atribuído a uma causa bem determinada, em geral sendo influenciado simultaneamente por diversos fatores contributivos (quantidade e qualificação de cientistas, investimentos, políticas públicas e outras). O efeito dos fatores é indireto e a quantificação dos resultados referentes a cada um deles, difícil (TRZESNIAK, 1998).

Apesar de sua importância, a complexidade da construção e emprego desses indicadores faz com que, até o momento, não se tenham procedimentos com o suficiente reconhecimento e uso internacional (SPINAK 1998; VAN LEEUWEN; MOED, 2005; BIREME, 2008). Enquanto isso, a formulação e o uso dos indicadores de insumos, embora também complexos, estão mais bem estabelecidos nos seguintes manuais de orientação editados pela OCDE: a) Manual de Frascati (FECYT, 2003), que engloba principalmente gastos e pessoal dedicado à pesquisa e desenvolvimento (P&D), e anexos com recomendações gerais sobre pessoal em C&T, patentes, bibliometria, balanço de

pagamentos, indústrias e produtos de alta tecnologia e sociedade da informação; b) Manual de Oslo (OECD, 2005), sobre indicadores de inovação tecnológica; e c) Manual de Canberra (OECD, 1995), sobre recursos humanos ligados a C&T. Mais recentemente, a Rede Ibero-Americana de indicadores e ciência e tecnologia (Ricyt) organizou um manual sobre indicadores para a realidade dos países da América Latina e Caribe (RICYT, 2001; LETA; CRUZ, 2003). Apesar da falta de procedimentos de amplo reconhecimento voltados para indicadores de produção científica, o Manual de Frascati indica como importante referência a publicação da OCDE, elaborada por Okubo (1997), sobre indicadores bibliométricos.

2. Características das bases de dados SCIE, SSCI e A&HCI da Thomson Reuters

As bases de dados *Science Citation Index Expanded* (SCIE), *Social Science Citation Index* (SSCI) e *Arts & Humanities Citation Index* (A&HCI), originalmente elaboradas pelo *Institute for Scientific Information* (ISI), que foi absorvido pela Thomson Reuters, são as maiores fontes de dados bibliográficos científicos mundiais e as mais empregadas para a construção de indicadores bibliométricos de produção científica (OKUBO, 1997). No Brasil, estão disponíveis para consulta livre a pesquisadores com acesso ao Portal de Periódicos da Capes, reunidas no produto *Web of Science*, que contém dados de mais de 39 milhões de publicações e respectivas referências citadas (THOMSON REUTERS, 2008a). A base *Science Citation Index Expanded* indexa mais de 6650 periódicos, com crescimento de aproximadamente 19 mil novos registros por semana e abrange dados atuais e retrospectivos (parciais) desde 1900, principalmente nas áreas de ciências exatas, biológicas e da terra. A *Social Science Citation Index* possui mais de 1950 periódicos na íntegra e partes de outras 3500 publicações, com crescimento de aproximadamente 2900 novos registros por semana e dados retrospectivos (parciais) desde 1956. A *Arts*

& *Humanities Citation Index* indexa aproximadamente 1.160 periódicos e partes de outros 6800 periódicos, com crescimento de 2300 novos registros por semana e dados retrospectivos (parciais) desde 1975 (THOMSON REUTERS, 2008a).

Em conjunto, SCIE, SSCI e A&HCI têm cobertura multidisciplinar, mas há dificuldade de equilíbrio na representatividade das várias áreas do conhecimento, devido a diferentes hábitos e tradições de publicação e citação, além de barreiras linguísticas. Física, Química e Biomedicina são consideradas com boa representação, enquanto há problemas em Geociências, Biologia, Engenharia, Matemática e outras (NEDERHOF, 2006). Quanto ao idioma das publicações indexadas, há predominância do inglês, o que dá margem para a interpretação de que há sub-representação das publicações de países cujo idioma nacional não seja o inglês, principalmente nas áreas de humanidades (LETA; GLANZEL; THIJIS, 2006). Embora predominem os artigos de periódicos, há grande variedade de tipos de documentos indexados, incluindo revisões, resumos, editoriais e, em menor quantidade, livros, composições musicais, obras de arte e outros. No entanto, para fins de elaboração de indicadores, tem-se optado pela seleção de artigos (*articles*), cartas (*letters*), notas (*notes*) e revisões (*reviews*) (NSB, 2002, 2004, 2006; EUROPEAN COMMISSION, 2003), o que pode levar a uma sub-representação das áreas de humanidades em que há preferência pelas publicações em livros (PRAT, 1998; SPINAK, 1998; MACIAS-CHAPULA, 1998; TARGINO; GARCIA, 2000). O desequilíbrio na cobertura das publicações por áreas do conhecimento, a predominância do idioma inglês e dos artigos nas publicações indexadas são alguns dos principais motivos para não serem feitas comparações diretas entre áreas do conhecimento. No entanto, são válidas as comparações de publicações de diferentes países por área do conhecimento e a evolução das publicações por área do conhecimento ao longo do tempo (FARIA; GREGOLIN; HOFFMANN, 2007).

A partir da SCIE e da SSCI são desenvolvidos produtos voltados para a análise da publicação científica, com destaque para: a) *National Citation Reports* (NCR), uma base de dados eletrônica das publicações de um país específico, que contém a informação bibliográfica de cada publicação, além de contagens sobre a citação total e ano a ano, recebida por publicação (THOMSON REUTERS, 2009a); b) *Journal Citation Reports* (JCR), voltado para a avaliação e comparação de periódicos presentes nas bases de dados SCIE e SSCI a partir das citações que seus artigos recebem, incluindo a ordenação dos periódicos por número absoluto de citações e por fator de impacto, indicador tão valorizado pela comunidade científica

quanto polêmico (THOMSON REUTERS, 2009b; DONG; LOH; MONDRY, 2005) e c) *Essential Science Indicators* (ESI), que contém estatísticas por países, instituições, pesquisadores, periódicos e áreas do conhecimento sobre os dados dos últimos dez anos de publicações e citações indexadas na SCIE e na SSCI (THOMSON REUTERS, 2009c).

Associadas, as bases SCIE e SSCI e os produtos analíticos delas derivados apresentam características positivas para a produção de indicadores: cobertura multidisciplinar; registros bibliográficos detalhados; identificação de todos os coautores das publicações; disponibilidade de dados sobre as citações e boa quantidade de publicações de autores brasileiros indexadas: SCIE e SSCI em conjunto contam com 22410 publicações de 2006 tendo ao menos um pesquisador de instituições brasileiras entre seus autores (FAPESP, 2005; FARIA; GREGOLIN; HOFFMANN, 2007). Há também limitações: as bases foram criadas para a divulgação científica, e não para a produção de indicadores, embora sejam amplamente usadas para isso (ADAM, 2002); favorecem a representação do *mainstream* – “corrente principal” – da ciência mundial, preterindo temas de interesse regional (SPINAK, 1998); a distribuição dos periódicos por áreas do conhecimento não corresponde à mundial (LETA; CRUZ, 2003) e existe um viés para a cobertura de periódicos em inglês (LUWELL, 1999; VELHO, 2008).

Apesar das limitações, até recentemente não se dispunha de outra base com atributos semelhantes para a produção de indicadores bibliométricos e com a mesma combinação de quantidade e abrangência de publicações brasileiras. Lançada em 2004 pela Elsevier, a base de dados Scopus tem procurado posicionar-se como uma alternativa para a produção de indicadores.

3. Características das bases Scopus, SciELO e de bases especializadas selecionadas

Na construção de indicadores bibliométricos é importante o emprego de múltiplas bases de dados, uma vez que nenhuma delas tem cobertura completa da literatura mundial e cada qual apresenta suas deficiências e limitações. Desse modo, sempre que possível, é válido e tem sido recomendado empregar diferentes fontes, apesar do aumento da complexidade e possibilidade de obtenção, inclusive, de resultados em parte discordantes (OKUBO, 1997). Scopus, SciELO, Compendex, Inspec, PubMed, *Biological Abstracts* e *Sociological Abstracts* são algumas das bases que podem ser utilizadas para a elaboração de indicadores biblio-

métricos. Elas podem ser classificadas em dois grupos: (i) Scopus e SciELO são multidisciplinares, apresentam informação sobre a afiliação de todos os coautores e quantificam as citações recebidas pelas publicações indexadas; (ii) Compendex, Inspec, PubMed, *Biological Abstracts* e *Sociological Abstracts* são especializadas em algumas áreas do conhecimento, apresentam informação apenas sobre a afiliação do primeiro coautor e não quantificam as citações. As bases Scopus e SciELO permitem estudos amplos e simultâneos de diversas áreas do conhecimento, estudos de colaboração entre instituições e países, e estudos de citação. As bases especializadas permitem análises mais aprofundadas em suas áreas de cobertura. A seguir, mais informações sobre as bases mencionadas:

Scopus – produzida a partir de 2004 pela Elsevier, é uma base de dados bibliográfica multidisciplinar, multiafiliação e multi-idioma, com cobertura expressiva da literatura científica mundial. Ao todo, são cerca de 33 milhões de publicações indexadas, sendo 16 milhões a partir de 1996, ano que representa o início de uma cobertura mais consistente dos periódicos (ELSEVIER, 2008a). As publicações são provenientes de mais de 15 mil periódicos, sendo cerca de 3400 de Ciências biológicas, 5300 de Ciências da saúde, 5500 de Ciências exatas e 2800 de Ciências humanas. Dentre as revistas científicas indexadas, há mais de 1 000 títulos de acesso aberto, dentre os quais os títulos da SciELO (*Scientific Electronic Library Online*). Há também cobertura de monografias, anais de congressos e eventos científicos, patentes e outras fontes de informação científica disponíveis na internet, selecionadas segundo padrões de qualidade. Assim como SCIE e SSCI, possui dados sobre as citações feitas e recebidas pelas publicações indexadas, o que permite seu uso para elaboração de indicadores de citação (DE MOYA-ANEGÓN et al., 2007). Ela conta também com recursos que facilitam tanto a busca como a elaboração de indicadores bibliométricos, como o *Citation Tracker*, para o acompanhamento das citações recebidas por publicações, autores ou instituições; o *Author Identifier*, para identificação e padronização de nomes de autores; o *Affiliation Identifier*, para identificação e padronização de nomes de instituições e o cálculo do índice H, indicador baseado no número de publicações e na frequência com que são citadas, bastante valorizado recentemente para representar a influência de periódicos, pesquisadores, instituições e países (ELSEVIER, 2008b; VAN RAAN, 2006; VINKLER, 2007). A partir dos dados da Scopus, o grupo de pesquisa SCImago, da Universidade de Granada, Espanha, elabora e disponibiliza gratuitamente no *site SCImago Journal & Country Rank* (<http://www.scimagojr.com/>) diversos indicadores de publicação, citação e ligação, para a análise de periódicos e países ao

longo do tempo e por áreas do conhecimento. Scopus está disponível para acesso via Portal de Periódicos da Capes.

Scientific Electronic Library Online (SciELO) – é uma biblioteca eletrônica criada no Brasil em 1996 que, em sua coleção brasileira –, a **SciELO Brasil**, empregada neste trabalho – abrange uma coleção selecionada de periódicos científicos brasileiros, com disponibilização de textos completos. É resultado de um projeto colaborativo envolvendo a FAPESP, a Bireme (Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde) – e, mais recentemente, o CNPq, que tem por objetivo o desenvolvimento de metodologia para preparação, armazenamento, disseminação e avaliação da produção científica em formato eletrônico (SCIELO, 2008). A SciELO se expandiu internacionalmente e hoje existem coleções nacionais em outros 11 países da América Latina e Caribe, Espanha e Portugal, além de coleções temáticas nas áreas de Saúde pública e Ciências sociais, coordenadas pela Bireme. Entre os principais resultados da criação da SciELO Brasil estão a maior disseminação da produção científica brasileira e a promoção da sua visibilidade e credibilidade nacional e internacional (PACKER, 1998; FAPESP, 2002, 2005; SCIELO, 2008). Contém aproximadamente 90 mil publicações de 209 periódicos brasileiros das áreas de Ciências agrícolas, Ciências biológicas, Química, Engenharia, Geociências, Ciências da saúde e Ciências humanas. A base tem crescido rapidamente e conta com número de publicações relevante para a análise da produção científica brasileira, embora ainda inferior ao da SCIE e SSCI ou Scopus. Em relação a publicações brasileiras de 2006, há aproximadamente 19 mil na SCIE e SSCI, cerca de 23 mil na Scopus e 12 mil na SciELO Brasil. Há nos *sites* da SciELO uma série de relatórios quantitativos sobre as publicações indexadas, as citações recebidas e o acesso às publicações, muito úteis para a elaboração de indicadores. No entanto há dificuldades a serem compreendidas. Por ser uma base de dados eminentemente nacional – em 2006 cerca de 85% das publicações da SciELO Brasil tinham ao menos um autor brasileiro – o seu uso para o posicionamento da publicação brasileira frente a de outros países é limitado. O portal regional da SciELO, lançado em 2008, integra os indicadores de todas as coleções nacionais e temáticas e permitirá futuramente a análise da colaboração regional. Ademais, o fato de a recuperação de dados resultantes de buscas na SciELO ocorrer por meio de referências bibliográficas, e não por registros bibliográficos, em que os dados são apresentados em campos definidos, constitui uma dificuldade para o uso de *softwares* especializados em bibliometria. No entanto, com a progressão da cobertura dos periódicos da SciELO pela base Scopus, em breve deverá

ser possível a recuperação dos registros bibliográficos por meio dessa base. Uma consequência, por exemplo, será a possibilidade de maior agilidade para elaboração de indicadores de áreas do conhecimento. O volume crescente de dados, a diversidade de idiomas, a cobertura focada nos periódicos nacionais e a qualidade na indexação e padronização dos dados fazem da SciELO uma base importante para a produção de indicadores sobre a publicação científica brasileira, com tendência de aumento dessa importância com o crescimento da coleção. Disponível no portal regional SciELO ou via Portal de Periódicos da Capes.

Compendex – produzida pela Engineering Information Inc., atualmente parte do grupo Elsevier, contém cerca de 10 milhões de registros bibliográficos desde 1969: artigos, livros e outros documentos, em várias áreas especializadas de engenharia e tecnologia, incluindo por exemplo informações sobre novos produtos, previsão tecnológica e inteligência competitiva. Possui um tesouro¹ *online* que facilita as buscas e melhora os resultados da pesquisa. Disponível no Portal de Periódicos da Capes (ENGINEERING INFORMATION, 2008a, 2008b).

Inspec – produzida pelo Institution of Engineering and Technology (IET), contém aproximadamente 9 milhões de registros bibliográficos de artigos, livros e outros documentos, desde 1969, em áreas especializadas de física, engenharia elétrica e eletrônica, computação e controle e tecnologia de informação. Disponível no Portal de Periódicos da Capes (OVID, 2008).

PubMed – base de dados bibliográfica produzida pela *National Library of Medicine* (NLM) do U.S. National Institutes of Health (NIH) e que reúne mais de 18 milhões de publicações de periódicos das áreas de saúde e ciências biológicas, incluindo 16 milhões de publicações da base MEDLINE (NLM, 2008a, 2008b). Disponível no Portal de Periódicos da Capes, pela OVID e também pela Bireme, na Biblioteca Virtual em Saúde.

Biological Abstracts – produzida pela Thomson Reuters, a base indexa mais de 11,3 milhões de publicações internacionais na área de ciências da vida (THOMSON REUTERS, 2008c). A *Biological Abstracts* está disponível para acesso via Portal de Periódicos da Capes.

Sociological Abstracts – produzida pela empresa CSA, indexa literatura internacional em sociologia e áreas afins de mais de 1800 periódicos e também de livros, resumos, dissertações e artigos de congressos, desde 1952, num total de 840 mil registros. A *Sociological Abstracts* está disponível para acesso via Portal de Periódicos da Capes (CSA, 2008).

4. Fatores de complexidade do emprego de indicadores de citação científica

Apesar do interesse crescente pelos indicadores de citação científica, existem muitos fatores que tornam complexa sua construção e seu emprego, podendo-se destacar os seguintes (OKUBO, 1997; ADAM, 2002; MACIAS-CHAPULA, 1998; VELHO, 2008): a) as diferenças de audiência, tamanhos de equipes, barreiras linguísticas, procedimentos e culturas por áreas do conhecimento, mercado e visibilidade das revistas; b) citações por motivos que não a relevância do trabalho, incluindo táticas ou tributos a eminências, referências a metodologias, crítica a erros ou controvérsias, autocitação etc. c) trabalhos de revisão ou comunicação rápida, que não representam contribuições originais e podem viesar os indicadores; ou, ainda, d) a ocorrência de erros técnicos nas fontes ou no processamento de informações.

No que diz respeito às áreas do conhecimento, por exemplo, artigos em biomedicina são frequentemente mais citados do que artigos em matemática ou medicina (ADAM, 2002). Em estudo sobre citações dos artigos publicados em periódicos científicos cobertos pelas bases do ISI no período de 1981 a 1985 foi verificado que em torno de 55% dos artigos não foram citados uma única vez nos cinco anos após a publicação (HAMILTON, 1990). Estudo complementar do mesmo autor indicou para a área de engenharia uma proporção superior a 72% de artigos não citados nos cinco anos seguintes; em ciências sociais, 75%; enquanto em artes e humanidades a maioria das disciplinas não teve citação para 92 a 99,9% dos artigos publicados (HAMILTON, 1991).

Estudos recentes também têm mostrado que, por diversas razões não associadas à qualidade do trabalho, cientistas de países periféricos publicam seus trabalhos em periódicos locais e assim têm menos visibilidade e são menos citados, destacando-se as seguintes: barreiras da língua, despesas da publicação internacional, sentimento nacionalista de reforço a periódicos nacionais, audiência-alvo que não lê periódicos estrangeiros, falta de experiência estudantil no exterior e o caráter nacional ou regional focalizado na pesquisa (VELHO, 2008).

Vale mencionar que existem duas fortes correntes antagônicas a respeito das citações: enquanto alguns especialistas apontam as análises de citações como um

1. Vocabulário controlado e dinâmico de descritores que se relacionam semântica e genericamente, cobrindo um ramo específico de conhecimento.

procedimento que contribui para a objetividade nas decisões, outros consideram justamente o contrário, devido aos erros e desvios associados à construção e interpretação desse tipo de indicador (ADAM, 2002), bem como as eventuais dificuldades de uso das bases SCIE e SSCI, as únicas que poderiam ser empregadas até 2004 com a criação da base Scopus e até hoje as mais utilizadas nesses estudos.

5. Metodologia dos boxes sobre redes de pesquisa (seção 5)

5.1 Box 4.2: Redes de pesquisa em nanotecnologia

Os dados de publicações científicas brasileiras na área de nanotecnologia e nanociência são provenientes da base *Web of Science* (WoS), disponível no Portal de Periódicos da Capes. Eles foram extraídos de um conjunto maior de dados, relativo a toda a produção científica brasileira indexada na WoS, recuperado por meio da busca “AD=brazil”, com restrição ao período 2002 a 2006 e preparado especificamente para a produção de indicadores. Foram selecionadas 519 publicações do período classificadas na categoria “Nanoscience & Nanotechnology”. Os dados foram tratados e quantificados utilizando-se o *software* bibliométrico VantagePoint para gerar uma matriz de co-ocorrência de instituições autoras das publicações. A matriz foi processada nos *softwares* de análise e representação de redes Ucinet e NetDraw, resultando na rede de colaboradores em publicações apresentada na Figura 4.1.

Os dados das patentes sobre nanotecnologia geradas no Brasil foram coletados a partir da base Derwent Innovations Index, também disponível no Portal de Periódicos da Capes. Das 101 patentes analisadas, 73 têm organizações brasileiras como titulares. As 28 demais estão registradas em nome apenas de pessoas físicas. A expressão de busca utilizada foi “(IP=(B82*) OR TS=(nano* NOT (nanometer* OR nanosecond* OR nano2 OR nano3))) AND PN=(BR*)”, com restrição ao período 1998 a 2006 e posterior seleção das patentes tendo apenas o Brasil como país de prioridade. Os dados foram tratados de forma similar às publicações científicas, resultando numa matriz de co-ocorrência de titulares de patentes e na rede de colaboradores em patentes apresentada na Figura 4.2.

5.2 Box 4.4: Rede de cooperação científica de genômica e melhoramento genético da cana-de-açúcar

Os dados de publicações científicas sobre o tema melhoramento genético são provenientes da base *Web of Science* (WoS), disponível no Portal de Periódicos da Capes. Para a busca de publicações foram empregados os seguintes parâmetros: a) expressão 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) ano de publicação (Year Published): 1998-2006; c) bases selecionadas: SCIE e SSCI e d) tipos de documentos: artigos, *letters*, notas e *reviews*. Foram recuperadas 690 publicações, das quais 139 apresentaram contribuição de ao menos um autor vinculado a uma organização brasileira. Os dados foram tratados e quantificados utilizando-se o *software* bibliométrico VantagePoint para gerar uma matriz de co-ocorrência de instituições autoras das publicações. A matriz foi processada nos *softwares* de análise e representação de redes Ucinet e NetDraw, resultando na rede de colaboradores em publicações apresentada no Gráfico-Box 4.2.

5.3 Box 4.5: Redes: aspectos teóricos e conceituais

As redes apresentadas se inserem no amplo contexto conceitual dos estudos de C&T e foram fortemente baseadas no trabalho de Dal Poz (2006), no qual as discussões apontadas encontram-se detalhadas.

O trabalho assume como parte do instrumental de análise os estudos do campo das *redes sociotécnicas*, que aplicam esforços na integração de visões econômicas e sociológicas acerca da produção tecnológica. O recorte da abordagem conceitual utilizada passa pela definição dos subcampos teóricos que estudam redes, das dimensões consideradas na análise que este box realiza, o que inclui as fronteiras das redes que são aqui estudadas.

A literatura associada a estes estudos assume as redes como o resultado de processos evolutivos dinâmicos, que podem ser mais bem entendidos a partir de uma perspectiva que leva em conta seus elementos:

- Social, envolvendo indivíduos, organizações públicas e privadas, e as interações de todos eles em contextos específicos;
- Econômico, dizendo respeito às interações e transações que transformam conhecimento e recursos para que certos ganhos sejam alcançados;

- Tecnológico, referindo-se às tecnologias e sua transferência, para garantia de vantagens competitivas;
- Sociotécnico, que focaliza as maneiras pelas quais a sociedade e a tecnologia se influenciam mutuamente.

Este contexto permite pensar a evolução das redes socioeconômicas e de atores sociais envolvidos na geração de inovação. Em particular, permitem identificar e definir tais atores e sua dinâmica nos arranjos tecnológicos, independentemente da dimensão geográfica que tenham.

O interesse pela utilização desta abordagem vem crescendo, por pelo menos duas razões: a emergência de novas formas organizacionais de estruturação das indústrias, que destaca elementos como as ligações horizontais e laterais entre as firmas, e o surgimento das novas tecnologias, como informática, que tornaram possível a existência de arranjos menos rígidos das estruturas organizacionais (NOHRIA; ECCLES, 1992).

O desenvolvimento das “teorias” ou estudos de redes visa explicar, do ponto de vista do desenvolvimento tecnológico, as relações entre comportamentos sociais e econômicos. Por isso, a abordagem se mostra útil para a elaboração de indicadores capazes de iluminar, sob um novo ponto de vista, relações intrínsecas pouco tangíveis por indicadores tradicionais de C&T.

A literatura sobre redes demonstra que a distinção entre a visão econômica e a sociológica permanece, ainda que a aproximação analítica e conceitual entre elas venha crescendo.

Algumas referências bibliográficas seminais são apresentadas ao final deste capítulo, podendo ser consideradas como a base acadêmica do conteúdo do box.

A vertente sociológica das redes (CALLON, 1992; LARÉDO, 1998; LARÉDO; MUSTAR, CALLON, 1993;

BELL; CALLON, 1994) será tomada como a principal forma de abordagem para o trabalho, sem contudo deixar de considerar as abordagens econômicas e provenientes dos estudos que se referem aos *sistemas nacionais de inovação*.

5.3.1 Indicadores algébricos

As relações entre os atores de uma rede podem ser expressas de forma visual, com o auxílio de *softwares* especializados em construção de redes como o *software* PAJEK ou o UCINET, ou de forma algébrica. Os indicadores mais utilizados – densidade, centralidade e conectividade da rede são definidos a seguir.

A representação algébrica parte do posicionamento dos atores numa matriz quadrada (por convenção, os citantes estão colocados nas colunas e os citados nas linhas). Tais matrizes relacionam, dois a dois, os elementos da Figura M4.1 apresentada logo a seguir, como ilustrado na Tabela M4.1, representação matricial das citações (marcadas com 1) posteriores recebidas por uma patente X, hipotética, e não citações (marcadas com 0).

Na representação matricial, k representa a rede genérica $k^{\text{ésimo}}$, sob análise, e z_{ijk} representa o valor da relação do ator $i^{\text{ésimo}}$ com o ator $j^{\text{ésimo}}$ ($z_{ijk} \geq 1$ se o ator $i^{\text{ésimo}}$ está ligado ao ator $j^{\text{ésimo}}$; de outra forma, $z_{ijk} = 0$).

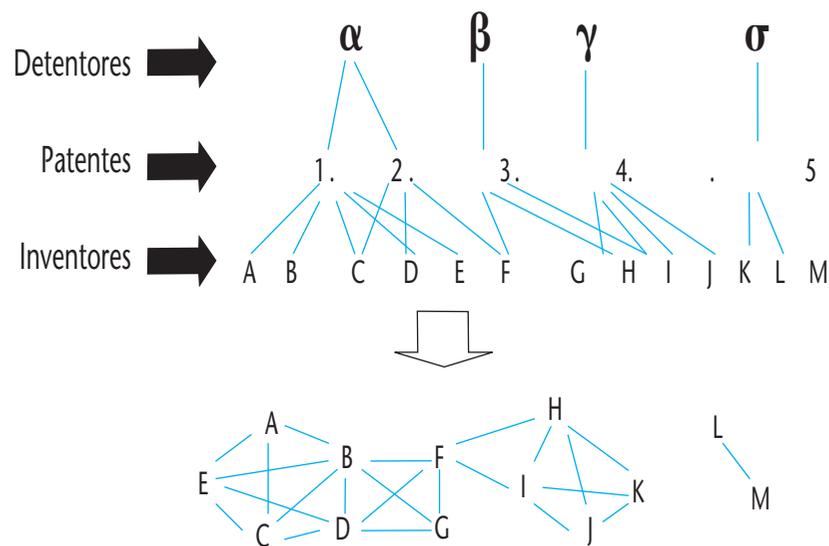
A frequência com que certa patente é citada por patentes posteriores é um importante indicador do valor de mercado desta patente, de modo que os resultados da análise de citação de uma patente por outras, posteriores, serão utilizados para averiguar a história de alguns produtos obtidos pelos atores mais relevantes da rede, ou seja, os atores que se diferenciam em termos da frequência com que são citados.

O pressuposto é que os inventores A até M estão vinculados entre si por meio do conhecimento incorporado nas patentes, mas que o grau de relacionamento entre eles varia.

Tabela M4.1
Matriz de citações posteriores de uma patente por diferentes inventores

Patente X Detentor Y	Matriz de citações posteriores de uma patente por diferentes inventores				
	Detentor A	Detentor B	Detentor C	Detentor D	Detentor E
Citação posterior (citado) 1	0	1	1	0	1
Citação posterior (citado) 2	0	0	1	1	1

Figura M4.1
Gráfico de relacionamentos binários entre detentores, patentes e inventores e a representação geodésica da rede de vínculos



Nota: A figura baseia-se na Figura 1 “Bipartite graph of patents and inventors”, elaborada por Balconi, Breschi e Lissoni (2004).

5.3.2 Indicador de densidade da rede

Indicador que expõe o grau de dinamismo de uma certa rede e não de um ou outro ator individual. O cálculo é feito dividindo-se o número de relações que existem na rede genérica ‘k’ pelo número de relações possíveis – ‘N² - N’ – já subtraídas as relações de cada vértice com ele mesmo. Pode assumir valores de zero a um: de nenhuma conexão a todas aquelas possíveis entre atores de uma rede. Uma vez que só são considerados os vínculos entre atores diferentes, são descontados os vínculos de autorreferência, ou seja, os vínculos de um ator para consigo mesmo. No caso em questão, não há coautoria de um pesquisador consigo mesmo.

O indicador de densidade desta rede é 0,1549, ou seja, as relações existentes representam 15,49% das relações possíveis. Obviamente não se esperaria que todas as instituições colaborassem com todas as outras, num grau máximo, tornando a densidade da rede ao valor máximo, igual a 1,0. Os valores algébricos não são úteis quando analisados de modo isolado, como é explicado a seguir.

A utilidade deste indicador cresce quanto mais evidência comparativa for obtida e depende de um contexto comparativo de indicadores, incluindo os mais tradicionais. O indicador passa também a fazer mais sentido quando utilizado dentro de um grupo de comparativos analíticos. Um exemplo pode ilustrar

esta ideia. Certas áreas, como no caso das ciências da vida, a colaboração científica é necessária, dada a complexidade crescente da pesquisa e a necessidade de complementação de certas competências para alcançar resultados “publicáveis”. Se, num caso destes, após uma série de ações de fomento a certas áreas de C&T, os indicadores clássicos de produção científica de publicação não estiverem respondendo positivamente, considerando-se as questões de temporalidade que devem ser observadas para a análise, o índice de densidade das colaborações poderá iluminar os caminhos do *policy making*, demonstrando a necessidade de aliar a política de fomento à colaboração.

O valor algébrico da densidade desta rede poderá, então, fortalecer a análise, por exemplo, dos indicadores de produção de publicações brasileiras no mesmo período, caso a dinâmica desta área seja pautada pela necessidade de colaboração.

Este indicador demonstra o grau de envolvimento dos atores (neste caso, das instituições dos coautores) com os outros atores da mesma. É dado pelo valor algébrico de todos os contatos que envolvem o ator “a”, dividido por todos os reais contatos encontrados na rede.

A análise da configuração da rede e dos indicadores algébricos acerca das relações intrínsecas dos atores presentes na rede poderá demonstrar que as redes apresentam limites, envolvem certos atores e não outros, e que os cenários de ordem jurídica, legal e re-

lativos à formulação e implementação de políticas de C&T fazem a diferença para que certos atores estejam ou não incluídos na dinâmica de apropriação tecnológica mundial.

5.3.3 Conectividade

Este indicador examina a frequência das relações para estabelecer a proximidade entre os atores, e a quantidade das relações para determinar a localização dos atores dentro da rede, em termos de importância.

Quando as redes são analisadas, a distância entre os atores não é dada pela dimensão geográfica, mas pela intensidade com que eles se relacionam. Assim, quanto mais conectados, menor a distância social desses atores dentro da rede. No caso do mapeamento das instituições que publicaram artigos científicos em conjunto, os institutos mais próximos são os que têm uma interação mais forte, ou seja, publicaram juntos com grande frequência.

A localização dentro da rede também tem significado: atores posicionados no “núcleo” da rede ou próximos do “núcleo” são os que têm uma quantidade grande de conexões com outros atores e, portanto, podem ser classificados como os mais importantes ou mais dinâmicos. Já os atores pouco relacionados situam-se na periferia da rede.

Mas dentro do grupo de atores classificados como mais importantes (ou dinâmicos) há uma distinção fundamental: há atores que estão relacionados com outros que também são dinâmicos (estes atores serão centrais na rede) e há atores que estão relacionados com outros que são pouco dinâmicos (estes atores se posicionarão entre o núcleo e a periferia).

5.3.4 Distância geodésica

Este indicador é definido pelo número de arestas que separa dois atores distintos de uma rede. Por exemplo, na rede de colaboração científica de coautorias, a USP tem relação direta com a Unicamp. A distância geodésica entre estas duas instituições é 1,0; a Unicamp, por sua vez, tem relação com o Iapar. Mas o Iapar não tem relação direta com a USP: somente por intermédio da Unicamp; assim, a distância geodésica entre a USP e o Iapar é, portanto, 2,0.

Uma observação importante deve ser feita quanto à utilização deste indicador. Deve-se observar se a direção das relações entre os diferentes vértices importa, ou seja, se as conexões estão em forma de arco ou de seta. Dependendo da natureza destas relações, haverá a necessidade de ajustar o indicador.

5.3.5 Centralidade dos atores

Alguns atores se destacam pelo número de vínculos que apresentam, exibindo alto grau de centralidade, enquanto outros podem não apresentar tantos vínculos, mas estar posicionados em determinados setores das redes que lhes conferem um papel de conectores entre diferentes componentes ou subsetores.

O indicador de centralidade mostra como certos atores centrais da rede apresentam alto grau de envolvimento com os outros atores da rede. É dado por todos os contatos que envolvem o ator “a” dividido por todos os reais contatos encontrados na rede. Da mesma forma que o indicador de densidade, este indicador não leva em conta as autorreferências, já que só vale para os casos nos quais $a \neq i$. Além disso, pelo mesmo argumento usado na densidade, é preciso fazer ajustes neste indicador.

Se a direção da relação importar o indicador terá a seguinte conformação:

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

$$i \neq a \text{ (i são todos os demais atores, exclusive a)}$$

Significa dizer que a centralidade do ator ‘a’ é medida pelo número de relações que ele tem com os demais atores e os demais atores têm com ele (neste caso, a direção importa), dividido pelo número total de relações da rede ‘k’ toda.

Se a direção da relação não importar o indicador terá a seguinte conformação:

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

$$i \neq a \text{ (i são todos os demais atores, exclusive a)}$$

Significa dizer que a centralidade do ator ‘a’ é medida pelo número de relações que ele tem com os demais atores ($Z_{ia} = Z_{ai}$ e, neste caso, a direção não importa) dividido pelo número total de relações da rede ‘k’ toda.

A tabela a seguir apresenta alguns índices de centralidade da rede.

Tabela M4.2
Índices de centralidade das instituições – Brasil – 2000-2007

Instituições	Índice de centralidade
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

Fonte: Dal Poz (2006).

Dado o espaço temporal de busca, é possível inferir que determinados programas de fomento à C&T, como os Projetos-Genoma iniciados no ano de 1997 (Genoma *Xylella*, da FAPESP, e Genomas Brasileiros) resultaram, após quase uma década, em produção científica colaborativa. O espectro geográfico da sobreposição das políticas de fomento fica claro: instituições do Estado de São Paulo são majoritárias na lista acima, e seus índices de centralidade, se somados, conformam a predominância desse estado nas atividades-tema. O indicador também permite mensurar a importância e o papel relativo de determinadas instituições nestes cenários.

5.4. Box 4.6: Redes de pesquisa em biofotônica

O levantamento das competências e o estudo sobre a comunidade na área de biofotônica no Brasil basearam-se numa amostragem que revelou uma comunidade produtiva e atuante, com muitas colaborações multidisciplinares de P&D, incluindo ramificações na área empresarial. Isso demonstra o grande potencial dessa área no Brasil.

Levantou-se a existência das competências e das redes de colaborações por meio de uma pesquisa na base *Web of Science*, com as palavras-chave mostradas ao final deste anexo (Quadro M4.1) e a palavra “brazil” no endereço da instituição. As palavras-chave foram extraídas do *Handbook of Biomedical Optics*, da CRC

Press, e do livro *Introduction to Biophotonics*, de Paras Prasad, da Wiley Interscience, para garantir a presença de todas as áreas representativas da biofotônica. O resultado foram 1 730 trabalhos publicados. O período pesquisado abarcou todos os anos cobertos pela *Web of science*, isto é, de 1955 a 2006. Entretanto, não houve nenhum registro de qualquer trabalho publicado nessa área antes de 1983.

Após extrair os trabalhos com “reprint address” no exterior, sobraram 1 388 trabalhos, dos quais 1 080 continham apenas endereços brasileiros. No intuito de investigar a localização da competência, decidiu-se processar os endereços quebrados no âmbito de departamento. Agregaram-se os dados ao que se considera um instituto ou faculdade nas universidades estaduais de São Paulo, tomando o cuidado de encontrar o equivalente nas instituições federais, nas quais os departamentos correspondem aos institutos. Nesses casos, por exemplo, um departamento de física em uma universidade federal representa uma unidade equivalente a um instituto de física nas universidades estaduais paulistas. Entretanto, um departamento de anatomia, por exemplo, não aparece como uma unidade isolada, mas sim agregado como faculdade de medicina. Instituições não universitárias, como Instituto Ludwig ou Ipen, por exemplo, aparecem como uma unidade. O Quadro M4.2 mostra o nível de agregação das unidades representativas desse estudo. Nele observa-se uma

lista das instituições de São Paulo com mais de 1% das 1 388 publicações e o percentual dos trabalhos em que seu nome aparece entre as instituições de origem dos coautores. Com o objetivo de identificar as competências na área, principalmente no Estado de São Paulo, e considerando-se apenas os 1 080 trabalhos com autores exclusivamente brasileiros, utilizou-se o seguinte cri-

tério de corte: instituições fora do Estado de São Paulo com menos de 1% foram desconsideradas, mantendo todas as instituições do Estado. Sobraram 165 instituições, com as quais realizamos a análise de redes utilizando o *software* Pajek. O grande número de publicações e de instituições envolvidas demonstra o alcance que a biofotônica já tem hoje no Brasil.

Quadro M4.1

Lista de palavras-chave para a identificação de competências em biofotônica – Brasil

Palavras-chave para a identificação de competências em biofotônica

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) Indica que a busca capturou diferentes formas de escrita da palavra.

Tabela M4.3
Publicações na área de biofotônica indexadas na base Web of Science, segundo instituições – Estado de São Paulo – 1983-2006

Posição	Instituições	Publicações na área de biofotônica	
		N ^{os} Abs.	%
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

Fonte: Fragnito et al. (2007).

Quadro M4.2
Legenda detalhada da Figura 4.7

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

(CONTINUA)

Quadro M4.2

Legenda detalhada da Figura 4.7

11 UFMG – Med – BH	52 Unesp – Fis – SJR Preto	93 H Serv Estado – SP	134 IAC – Campinas
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 UNLS – 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 Biypton – 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 – Cordeiro
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	