UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

**ANÁLISE ESPACIAL DA PRODUÇÃO E DAS REDES DE COLABORAÇÃO
CIENTÍFICA NO BRASIL: 1990-2010**

Otávio José Guerci Sidone

Orientador: Prof. Dr. Eduardo Amaral Haddad

SÃO PAULO

2013

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Reinaldo Guerreiro
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Joaquim José Martins Guilhoto
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

OTÁVIO JOSÉ GUERCI SIDONE

**ANÁLISE ESPACIAL DA PRODUÇÃO E DAS REDES DE COLABORAÇÃO
CIENTÍFICA NO BRASIL: 1990-2010**

Dissertação apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Eduardo Amaral Haddad

Versão Corrigida

(versão original disponível na Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2013

FICHA CATALOGRÁFICA

Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Sidone, Otávio José Guerci

Análise espacial da produção e das redes de colaboração
científica no Brasil : 1990-2010 / Otávio José Guerci Sidone.
– São Paulo, 2013.

166 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2013.
Orientador: Eduardo Amaral Haddad.

1. Economia regional 2. Cientometria 3. Desenvolvimento
econômico regional 4. Política tecnológica I. Universidade de
São Paulo. Faculdade de Economia, Administração e Contabilidade
II. Título.

CDD – 338.09

**Aos meus pais, pelo empenho e
esforço dedicados à minha
formação educacional.**

AGRADECIMENTOS

Agradeço especialmente aos meus pais pelo suporte e conforto oferecidos, além do apoio incondicional à continuidade dos meus estudos. Agradeço também aos meus irmãos pelo carinho de sempre.

Agradeço ao Professor Eduardo Amaral Haddad pela amizade e cumplicidade, pelo incentivo constante e oferecimento de novas oportunidades desde os tempos de minha graduação, e pela atenção e dedicação na orientação desse trabalho. O convívio durante esse período fez aumentar ainda mais o sentimento de estima em relação ao seu profissionalismo e excelência nas atividades de docência e pesquisa acadêmica, bem como em relação à conduta respeitosa nos relacionamentos profissionais.

Agradeço aos Professores Hélio Nogueira da Cruz e Carlos Roberto Azzoni pela atenção dispensada e pelas preciosas contribuições tanto em conversas informais como no exame de qualificação, por meio de valorosas ideias e comentários pertinentes para o melhor encaminhamento deste trabalho.

Agradeço ao Professor Jesús Mena-Chálco pela disposição e auxílio na elaboração do banco de dados utilizado no estudo, além do tratamento cordial e solícito ao longo das várias etapas da pesquisa, o que constituiu num verdadeiro exemplo da importância da colaboração na ciência moderna.

Agradeço ao Sílvio de Paula, coordenador técnico do Departamento de Informática da Reitoria da USP, pela atenção e disposição dispensada na delicada e importante etapa preliminar de obtenção dos dados utilizados.

Agradeço aos colegas do Núcleo de Economia Regional e Urbana da USP (NEREUS): Alexandre Nunes, Ana Barufi, Denise Imori, Eliane Teixeira, Juliana Scriptore, Marina Camargo, Rafael Perez, Renato Vieira, Ricardo Lopes e Weslem Faria, aos colegas de pós-graduação do IPE-USP e ao amigo Caio Haib, pelo companheirismo nos imprescindíveis momentos de descontração.

Agradeço ao apoio institucional do Programa de Pós-Graduação em Economia da FEA/USP, pelo provimento de estrutura durante o curso, e ao NEREUS, pelo ambiente agradável proporcionado para a condução das atividades de pesquisa, caracterizado pela possibilidade de interação colaborativa entre os pesquisadores e

pelo o compartilhamento de conhecimentos técnicos. Por fim, agradeço ao suporte financeiro da CAPES concedido durante o período de estudos.

RESUMO

O crescimento acelerado da produção científica brasileira nos anos recentes foi acompanhado pela expansão das colaborações científicas domésticas. Neste estudo, olhamos mais atentamente esse assunto na tentativa pioneira de identificar padrões espaciais da produção e colaboração científica no Brasil, e avaliar o papel da proximidade geográfica na determinação das interações entre os pesquisadores brasileiros. Por meio de uma base única composta por mais de um milhão de pesquisadores registrados na Plataforma Lattes e de sete milhões de publicações científicas, coletamos e consolidamos informações sobre as colaborações científicas inter-regionais em termos de redes de coautorias entre 1.347 municípios brasileiros ao longo do período compreendido entre 1990 e 2010, o que permitiu uma abrangência de dados e perspectiva de análise inéditas na literatura. Os efeitos da distância geográfica nas redes de colaboração são mensurados para as diferentes áreas do conhecimento por meio da estimação de modelos de interações espaciais. Os principais resultados sugerem fortes evidências de um processo de desconcentração espacial da produção científica nos últimos anos associado à expansão das redes de colaboração e ao aumento da participação de autores das regiões cientificamente menos tradicionais, tais como Sul e Nordeste. Ademais, também encontramos evidência de que a distância ainda desempenha papel crucial na determinação da intensidade dos fluxos de conhecimento nas redes de colaboração científica no Brasil, embora a magnitude do efeito varie entre as redes das diferentes áreas do conhecimento. Por exemplo, verificamos que o distanciamento de 200 quilômetros entre dois pesquisadores implica na redução média de 22% ou 45% na probabilidade de haver colaboração entre eles, caso eles sejam de Linguística, Letras e Artes ou Ciências Exatas e da Terra, respectivamente.

ABSTRACT

Recent years have witnessed an accelerated growth of Brazilian scientific production that was accompanied by an expansion of domestic research collaboration. In this paper we look more closely at this issue in a pioneering attempt to identify spatial patterns of research production and collaboration in Brazil, and to measure the role of geographical proximity in determining interaction between Brazilian researchers. Using a unique database comprised of over one million researchers registered in the Brazilian Lattes Platform and seven million scientific publications, we collect and consolidate information on interregional research collaboration in terms of co-authorship networks among 1,347 Brazilian cities over the period between 1990 and 2010, which enabled a range of data and analysis perspective unprecedented in literature. The effects of geographical distance on research collaboration are measured for different knowledge areas under the estimation of spatial interaction models. The main results suggest strong evidence of spatial de-concentration of scientific production in the last years with expansion of research collaboration networks and an increased participation of authors in scientifically less traditional regions, such as South and Northeast. Moreover, we also find evidence that distance still plays a crucial role in determining the intensity of knowledge flows in scientific collaboration networks in Brazil, although the magnitude of such effects varies among networks of different knowledge areas. For instance, we found that the distancing of 200 kilometers between two researchers implies an average reduction of 22% or 45% on probability of collaboration among them, if they are of Linguistics, Letters and Arts or Exact and Earth Sciences, respectively.

SUMÁRIO

1.	INTRODUÇÃO	9
2.	ECONOMIA DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO	15
2.1.	Óticas de análise da atividade científica	16
2.1.1.	Publicação	17
2.1.2.	Citação	20
2.1.3.	Especialização	21
2.1.4.	Localização	22
2.1.5.	Colaboração	23
2.1.5.1.	Internacional	24
2.1.5.2.	Doméstica	26
2.2.	Cientometria espacial	28
2.2.1.	Geografia econômica evolucionária	30
2.2.2.	Dimensões de proximidade	32
2.2.3.	Proximidade e cientometria	37
2.3.	Colaboração científica e coautorias	38
2.3.1.	Redes de coautorias	39
3.	METODOLOGIA	43
3.1.	Análise de redes sociais e espaciais	43
3.1.1.	Representação das redes de colaboração científica	44
3.1.2.	Métricas	45
3.2.	Modelos gravitacionais de interações espaciais	46
3.2.1.	Determinantes da colaboração científica	47
3.2.2.	Especificação e estimação	50
4.	BANCO DE DADOS	57
4.1.	Matrizes intermunicipais de colaboração científica	57
4.1.1.	Currículo Lattes	57
4.1.2.	Procedimento de elaboração das matrizes	60
4.1.2.1.	<i>ScriptLattes</i>	60
4.1.2.2.	Identificação das coautorias	61
4.1.2.3.	Identificação dos municípios	62
4.1.2.4.	Identificação das áreas de conhecimento	65
4.1.2.5.	Método de contabilização das coautorias	66
4.2.	Principais características dos dados	67
5.	RESULTADOS	71
5.1.	Análise espacial da produção e colaboração científica no Brasil	72
5.1.1.	Quantificação da perda de informação	72
5.1.2.	Características estruturais da produção científica municipal no Brasil	73
5.1.3.	Características estruturais das redes espaciais de colaboração científica	85
5.2.	Avaliação do papel da distância geográfica nas redes de colaboração científica	96
6.	CONSIDERAÇÕES FINAIS	107
	REFERÊNCIAS	111
	APÊNDICE A. MÓDULOS DE PRODUÇÃO CIENTÍFICA NO CL	119
	APÊNDICE B. GRANDES ÁREAS DO CONHECIMENTO SEGUNDO O CNPQ	121
	APÊNDICE C. MAPAS DA PRODUÇÃO CIENTÍFICA MUNICIPAL NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO	125

APÊNDICE D. EVOLUÇÃO DOS PRINCIPAIS MUNICÍPIOS ASSOCIADOS À PRODUÇÃO CIENTÍFICA NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO	135
APÊNDICE E. MAPAS DA COLABORAÇÃO CIENTÍFICA INTERMUNICIPAL NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO	145
APÊNDICE F. PRINCIPAIS LIGAÇÕES DE COLABORAÇÃO CIENTÍFICA NO BRASIL NO TRIÊNIO 2007-2009: POR ÁREA DO CONHECIMENTO	153

LISTA DE TABELAS

Tabela 1 – Número de pesquisadores associados a cada grande área do conhecimento.....	66
Tabela 2 – Evolução do percentual de publicações e colaborações sem localização municipal.....	72
Tabela 3 – Evolução dos principais municípios associados à produção científica no Brasil: todas as áreas do conhecimento.....	76
Tabela 4 – Número de municípios por faixa de produção científica.....	80
Tabela 5 – Estatísticas descritivas das matrizes trienais de colaboração científica intermunicipal: todas as áreas do conhecimento.....	85
Tabela 6 – Evolução das principais ligações intramunicipais de colaboração científica no Brasil em triênios selecionados.....	88
Tabela 7 – Evolução das principais ligações intermunicipais de colaboração científica no Brasil em triênios selecionados.....	89
Tabela 8 – Evolução da distância geográfica média nas colaborações científicas entre pesquisadores dos 105 municípios que mais colaboraram no triênio 2007-2009.....	96
Tabela 9 – Estimativas dos modelos de Poisson e binomial negativo para os triênios selecionados: todas as áreas do conhecimento.....	100
Tabela 10 – Estimativas dos modelos ZIP e ZINB nos triênios selecionados: todas as áreas do conhecimento.....	101
Tabela 11 – Estimativas dos modelos de Poisson e binomial negativo nos triênios selecionados, controlada pelo total de colaborações no triênio 1992-1994: todas as áreas do conhecimento.....	102

LISTA DE FIGURAS

Figura 1 – Matriz de adjacência em redes de colaboração científica inter-regional.....	44
Figura 2 – Evolução do total anual de publicações científicas únicas analisadas.....	62
Figura 3 – Evolução do total da publicação científica municipal no Brasil.....	73
Figura 4 – Evolução da participação relativa das áreas do conhecimento no total da publicação científica municipal no Brasil.....	74
Figura 5 – Taxa de crescimento da publicação científica municipal no Brasil entre triênios selecionados: por área do conhecimento.....	74
Figura 6 – Mapas da produção científica municipal no Brasil em triênios selecionados.....	78
Figura 7 – Mapas da produção científica municipal no Brasil no triênio 2007-2009 e localização dos <i>campi</i> de universidades públicas (federais e estaduais).....	79
Figura 8 – Curvas de localização municipal da produção científica de todas as áreas do conhecimento em triênios selecionados.....	82
Figura 9 – Curvas de localização municipal da produção científica por área do conhecimento: triênios 1992-1994 e 2007-2009.....	82
Figura 10 – Mapas da rede de colaboração científica no Brasil no triênio 2007-2009 e dos graus municipais na rede.....	87
Figura 11 – Evolução da quantidade de municípios (a), do número de arestas (b) e do grau médio municipal (c) nas redes anuais de colaboração científica no Brasil.....	91
Figura 12 – Evolução da densidade (a), do diâmetro (b) e da transitividade (c) nas redes anuais de colaboração científica no Brasil.....	92
Figura 13 – Evolução do diâmetro (a) e da transitividade (b) nas redes trienais de colaboração científica no Brasil.....	93
Figura 14 – Efeito da distância geográfica sobre a probabilidade de colaboração científica no triênio 2007-2009: por área do conhecimento.....	103

LISTA DE ABREVIATURAS E SIGLAS

C&T&I: Ciência, Tecnologia e Inovação
 CA: *Chemical Abstracts*
 CAB: *Commonwealth Agricultural Bureaux*
 CAPES: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
 CNPq: Conselho Nacional do Desenvolvimento Científico e Tecnológico
 CL: Currículo Lattes
 Embrapa: Empresa Brasileira de Pesquisa Agropecuária
 FAPESP: Fundação de Amparo à Pesquisa do Estado de São Paulo
 Fiocruz: Fundação Oswaldo Cruz
 FP: *European Framework Programme*
 IBGE: Instituto Brasileiro de Geografia e Estatística
 INPI: Instituto Nacional da Propriedade Industrial
 INPA: Instituto Nacional de Pesquisas da Amazônia
 INPE: Instituto Nacional de Pesquisas Espaciais
 ITA: Instituto Tecnológico de Aeronáutica
 MCTI: Ministério da Ciência e Tecnologia e Inovação
 MQO: Mínimos Quadrados Ordinários
 NSB: *National Science Board*
 OCDE: Organização para a Cooperação e Desenvolvimento Econômico
 P&D: Pesquisa e Desenvolvimento
 PIB: Produto Interno Bruto
 PL: Plataforma Lattes
 RS: *Royal Society*
 SciELO: *Scientific Electronic Library Online*
 UERJ: Universidade do Estado do Rio de Janeiro
 UFMG: Universidade Federal de Minas Gerais
 UFPE: Universidade Federal de Pernambuco
 UFPR: Universidade Federal do Paraná
 UFRGS: Universidade Federal do Rio Grande do Sul
 UFRJ: Universidade Federal do Rio de Janeiro
 UFSC: Universidade Federal de Santa Catarina
 UFSCar: Universidade Federal de São Carlos
 UFV: Universidade Federal de Viçosa
 UnB: Universidade de Brasília
 UNESCO: *United Nations Educational, Scientific and Cultural Organization*
 Unesp: Universidade Estadual Paulista
 Unicamp: Universidade Estadual de Campinas
 Unifesp: Universidade Federal de São Paulo
 USP: Universidade de São Paulo
 WoS: *Web of Science*
 ZIP: *Zero-Inflated Poisson Model*
 ZINB: *Zero-Inflated Negative Binomial Model*

1. INTRODUÇÃO

É notável o aumento recente do interesse de especialistas por temas relacionados à inovação tecnológica, visto que essa constitui um processo imprescindível para o crescimento e desenvolvimento econômico, principalmente devido a seu papel determinante como motor da competitividade industrial nas economias capitalistas. Como a inovação consiste no mecanismo responsável pela transformação do conhecimento em aplicações economicamente eficientes, a geração e difusão de conhecimento tornam-se elementos indispensáveis na economia moderna, isto é, a criação de valor depende cada vez mais da melhor utilização do conhecimento, qualquer que seja o nível de desenvolvimento dos países. Ademais, a intensa concentração espacial dos processos de criação e difusão de conhecimento evidencia a essencialidade da incorporação do espaço geográfico em meio à caracterização e compreensão das relações entre pesquisa científica, inovação tecnológica e desenvolvimento regional.

A atuação das universidades figura como um mecanismo impulsionador fundamental para o desenvolvimento econômico regional. Além de promover o aumento da renda e exercer funções tradicionais de ensino e difusão do conhecimento, elas podem propiciar o fomento da atividade inovativa. No âmbito da economia regional, existem fatores que podem condicionar a magnitude dos impactos das universidades, tais como a concentração geográfica de firmas inovativas e institutos de pesquisa, o tamanho da região, a presença de economias de aglomeração, a estrutura industrial, entre outros, uma vez que esses estão diretamente associados ao número de firmas, à concentração de firmas de alta tecnologia e ao desenvolvimento da infraestrutura de conhecimento, os quais podem influenciar o processo de transferência de conhecimento de maneira determinante. Compreender a importância relativa desses fatores é indispensável para o delineamento de políticas efetivas de desenvolvimento regional (GOLDSTEIN 2009; VARGA; PARAG, 2009).

Entretanto, dentre os efeitos propiciados pelas universidades, o potencial de criação de novos conhecimentos faz delas componentes centrais de um sistema de inovação, já que

são responsáveis pela produção de conhecimento com potencial de serem utilizados de maneira economicamente úteis. Isso torna a verificação da ocorrência e dimensão espacial dos *spillovers* de conhecimento (entendido como o processo pelo qual o conhecimento acadêmico é compreendido e utilizado pelas firmas do setor privado) tema importante nos estudos sobre transferência de conhecimento (VARGA; PARAG, 2009, PONDS *et al*, 2009). Porém, diante da dificuldade inerente de medição da capacidade das firmas regionais absorverem o conhecimento gerado nas universidades, os esforços de pesquisa têm-se concentrado na atuação da geografia nesse processo de interação, tendo encontrado evidências de localização espacial dos *spillovers* de conhecimento, *i.e.*, a intensificação dos mecanismos de aprendizagem e transmissão de conhecimento entre universidades e firmas é favorecida pela proximidade geográfica (JAFFE *et al*, 1993; AUDRESTCH; FELDMAN, 1996; ANSELIN *et al*, 2000; ÁCS *et al*, 2002).¹ Todavia, a literatura recente parece caminhar rumo à explicação e interpretação dos *spillovers* por meio da análise das redes colaborativas entre firmas e universidades, e entre as próprias universidades, de maneira a determinar a influência das redes na explicação dos processos de criação e difusão do conhecimento (PHELPS *et al*, 2012). Além disso, é esperado que as interações colaborativas entre os pesquisadores sejam localizadas no espaço, já que a elaboração e o progresso das atividades de pesquisa usualmente exigem encontros e discussões frequentes entre as partes envolvidas (PAN *et al*, 2012). Assim, tornam-se fundamentais as análises inseridas no ramo da cientometria espacial (responsável pela caracterização das atividades científicas no espaço geográfico), principalmente por meio de estudos acerca da configuração espacial dos fluxos de conhecimento e dos efeitos da proximidade geográfica sobre as redes de colaboração científica, uma vez que tal perspectiva poderia promover a potencialização da eficiência de políticas que estimulem o espraiamento do conhecimento para atividades inovativas.²

¹ Usualmente a identificação e mensuração dos *spillovers* de conhecimento é realizada por meio da abordagem da função de produção de conhecimento, também chamada de Jaffe-Feldman-Varga. A partir da construção de uma função de produção, essa abordagem procura explicar a intensidade da atividade inovativa regional (medida pelo número de patentes) como resultante da utilização de insumos (investimentos públicos e privados em P&D) e características dos sistemas de inovação regionais (ÁCS, 2009). Exemplos de aplicação desse ferramental analítico podem ser encontrados na revisão de literatura de Drucker e Goldstein (2007, p.32).

² Embora as percepções e ideias da geografia da inovação sejam interessantes, sua extensão é limitada para a compreensão dos fenômenos relacionados à geografia da colaboração científica, na medida em que a pesquisa científica e a inovação industrial estão imersos em ambientes de regras e incentivos que

Quanto à produção de conhecimento, a trajetória recente da evolução da atividade científica mundial é caracterizada pela transformação da geografia da pesquisa. A partir da tradicional organização no eixo Europa, Estados Unidos e Japão, a ciência voltou-se para a configuração de uma extensa rede global, onde a ascensão da produção de países emergentes, tais como China, Brasil e Índia, outrora de dimensões reduzidas, passou a desempenhar papel de destaque na produção de conhecimento no mundo (RS, 2011). Além disso, o crescimento acelerado da produção desses países parece ser sustentado pelo aumento das colaborações domésticas, *i.e.*, mediante o processo de intensificação dos esforços colaborativos entre pesquisadores localizados dentro dos territórios nacionais.

Logo, a colaboração na produção de conhecimento científico tornou-se elemento central das políticas de Ciência, Tecnologia e Inovação (C&T&I), visto que os dispêndios das universidades podem resultar em diferentes intensidades de fluxos de conhecimento, dependendo da estrutura e integração das redes científicas, e a articulação das redes colaborativas pode influenciar de maneira determinante a qualidade e produtividade científica (VARGA; PARAG, 2009; PONDS *et al*, 2009). No caso brasileiro, além das necessidades de se estimular o gasto em Pesquisa e Desenvolvimento (P&D) por parte do setor privado e de se promover a internacionalização das universidades, um dos maiores desafios enfrentados pelas políticas de C&T&I consiste na necessidade de se estimular a difusão da excelência científica dos grandes centros urbanos do Sudeste para centros menos privilegiados doutras regiões (CRUZ; CHAIMOVICH, 2010), de maneira a atenuar a intensa desconcentração regional da atividade científica ao longo do território nacional (BARROS, 2000; ALBUQUERQUE *et al*, 2002, 2005; FAPESP, 2011; CHIARINI *et al*, 2013), bem como de associar o crescimento da atividade científica ao incremento de sua qualidade (HELENE; RIBEIRO, 2011). Acreditamos que tais desafios podem ser encarados de maneira mais eficiente através da percepção de que estão intrinsecamente ligados à constituição e funcionamento das redes de colaboração científica no país, uma vez que a colaboração é o principal mecanismo de

moldam comportamentos bastante diferentes. Enquanto os atores da produção de conhecimento na academia procuram maximizar sua difusão, na indústria o objetivo é precisamente sua minimização (GOLDSTEIN, 2009; PONDS *et al*, 2009).

difusão do conhecimento e está estreitamente associada a maior qualidade da produção científica e ao maior valor econômico dos novos conhecimentos criados (SINGH; FLEMING, 2010).

A necessidade da melhor compreensão do papel desempenhado pela geografia nas redes de colaboração científica possui diversas razões práticas. Primeiramente, pode fundamentar decisões internas à comunidade acadêmica no tocante à escolha de parceiros colaboradores para a otimização do impacto e visibilidade da pesquisa. Em segundo lugar, pode auxiliar a condução de políticas públicas quanto à alocação de recursos para a realização de projetos colaborativos entre regiões, de maneira a aumentar a qualidade da produção científica a partir de determinada quantidade dispendida no financiamento desses (PAN *et al*, 2012). Por fim, permite a identificação das regiões mais proeminentes no processo de produção compartilhada de conhecimento, o que pode facilitar a capacidade de acesso por parte das firmas às fontes de conhecimento, fator fundamental para a realização de atividades inovativas.

Em meio à discussão apresentada, verifica a imprescindibilidade da compreensão da importância da geografia na produção científica e na disposição dos fluxos de conhecimento no Brasil, em termos do papel da proximidade geográfica como determinante das interações colaborativas entre os pesquisadores, temas ainda não explorados para o caso brasileiro, principalmente devido à enorme dificuldade no tocante à obtenção de dados apropriados. Logo, esses consistem nos principais objetivos deste trabalho, fundamentado na tentativa pioneira de caracterização da colaboração científica inter-regional no país, no nível da articulação entre pesquisadores dos municípios brasileiros.³

Sumariamente, este estudo visa encarar três objetivos primordiais. Primeiramente, analisaremos a produção científica do Brasil sob a perspectiva espacial, com o intuito de examinar a evolução tanto do total da produção científica entre 1990 e 2010 como de

³ A literatura sobre colaborações científicas inter-regionais ainda é escassa (Katz (1994) (Reino Unido), Liang e Zhu (2002) (China), Danell e Persson (2003) (Suécia), Okubo e Zitt (2004) (França e regiões de fronteira), Ponds *et al* (2007) (Holanda), Jones *et al* 2008 (EUA), Hoekman *et al* (2009, 2010) (Europa), Boshoff (2010) (África), Acosta *et al* (2011) (Europa)), porém, é crescente a importância dada a essa perspectiva de análise em meio as características da evolução da ciência moderna.

sua distribuição geográfica ao longo do território nacional, de maneira a evidenciar as regiões de maior produção de conhecimento no país e averiguar a possível existência de um processo de desconcentração espacial no período. Em segundo lugar, exploraremos os padrões geográficos dos fluxos de conhecimento no país, de maneira a estudar a evolução da colaboração científica entre 1990 e 2010, identificar a localização dos principais fluxos e as regiões com maior perfil colaborativo, além de verificar a existência de diferenciações nos padrões geográficos dos fluxos em cada área do conhecimento particular. Em terceiro lugar, investigaremos o papel da distância geográfica nas redes de colaboração científica, com o intuito de examinar o papel da distância geográfica como fator impeditivo à colaboração por meio da tentativa de mensuração de seu efeito, o que permitirá a verificação da hipótese de declínio de sua importância e de diferenciação de seu entre as áreas do conhecimento.

Mais especificamente, a estratégia adotada consiste na quantificação da produção científica dos municípios brasileiros e na elaboração de redes de coautorias entre esses ao longo do período 1990-2010 para diferentes áreas do conhecimento científico, a partir da análise das informações de mais de sete milhões de publicações científicas obtidas mais de um milhão de Currículos Lattes (CNPQ, 2012). Assim, permitir-se-á tanto a caracterização espacial da evolução da produção científica brasileira, bem como a identificação de padrões geográficos nas redes de colaboração científica, de maneira a promover o entendimento do papel da geografia na intermediação dos fluxos de conhecimento pelo território nacional, primordialmente para domínios científicos caracterizados pelo maior potencial de aplicabilidade às atividades inovativas no país.

2. ECONOMIA DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

Na economia moderna do conhecimento é crescente a atenção voltada para a melhoria da compreensão de assuntos relacionados às políticas de C&T&I, de modo a auxiliar na formulação de políticas que levem em conta as características peculiares de cada país. O Brasil insere-se de maneira proeminente nesse contexto, visto que são diversas as iniciativas dos órgãos estatísticos nacionais no aperfeiçoamento da mensuração dos esforços em atividades de Pesquisa e Desenvolvimento (P&D) e de indicadores de desempenho do sistema nacional de inovação.⁴

Os dispêndios mundiais em P&D crescem mais rapidamente do que o crescimento econômico mundial, o que é evidência da concentração de esforços em direção à constituição de verdadeiras economias do conhecimento, com intensiva capacitação tecnológica (NSB, 2012). Contudo, a análise da composição dos gastos mundiais em P&D revela a enorme heterogeneidade entre países desenvolvidos, emergentes e em desenvolvimento. Os EUA lideram os gastos totais, porém, países emergentes como Brasil, China, Índia, México e África do Sul estão alocando cada vez mais recursos em P&D, liderados pela triplicação da intensidade dos dispêndios em P&D da China (HOLLANDERS; SOETE, 2010; RS, 2011; NSB, 2012).⁵

Além da heterogeneidade entre os países do mundo, os esforços em C&T&I são bastante concentrados espacialmente em um pequeno número de localidades dentro dos territórios nacionais. Nos Estados Unidos, cerca de 65% dos dispêndios em P&D são realizados em apenas 10 dos 50 estados americanos.⁶ No Brasil, o Estado de São Paulo é responsável por 40% do total de gastos efetuados no país (HOLLANDERS; SOETE,

⁴ P&D é uma categoria que se refere ao trabalho criativo realizado de forma sistemática com o objetivo de aumentar o estoque de conhecimento, inclusive sobre o homem, a cultura e a sociedade, e usá-lo para desenvolver novas aplicações (FAPESP, 2011a).

⁵ A intensidade em P&D é definida como a fração do PIB que é dispendida nessa atividade. Assim, o valor para a China revela uma trajetória de crescimento sustentado quando levamos em conta o crescimento médio do PIB de 12% nesse período. Entre 2002 e 2008, a intensidade de P&D do país cresceu 50%. Já no Brasil e Índia, esse valor tornou-se estável, apesar do crescimento econômico. Em 2008, a intensidade de P&D no Brasil atingiu 1,09%, valor maior do que o observado nos países da América Latina, mas menor do que a média dos países da OCDE (2,28%) e União Europeia (1,77%) (HOLLANDERS; SOETE, 2010; NSB, 2012).

⁶ O Estado da Califórnia, individualmente, é responsável por mais de 20% do total (RS, 2011).

2010; FAPESP, 2010).⁷ O Estado de São Paulo é responsável por cerca de um terço do PIB do país e possui longa tradição no financiamento a educação superior e pesquisa, sendo aquele que recebe maior montante de recursos da esfera federal.⁸ Entretanto, a heterogeneidade espacial dos sistemas de inovação regionais e o potencial da C&T&I para o desenvolvimento regional são reconhecidos pelo governo brasileiro, o que tornam o fortalecimento e integração da capacidade científica objetivos primordiais de políticas, de forma a estimular as regiões menos favorecidas (CRUZ; CHAIMOVICH, 2010).

A heterogeneidade dos dispêndios nos insumos necessários para a realização das atividades de C&T&I certamente provoca dessemelhanças tanto na quantidade como na localização do produto dessas. No âmbito da ciência, tais disparidades são evidentes.

2.1. Óticas de análise e evolução recente da atividade científica

Quanto à quantidade e localização da atividade científica, é importante destacar os principais aspectos relacionados à dinâmica das mudanças no padrão científico mundial, de modo a apontar as tendências recentes da evolução da ciência a partir dos anos 1990 e a caracterização do Brasil nesse cenário.

Primeiramente, verifica-se o crescimento acelerado da produção científica mundial, aliado a uma tendência acentuada à intensificação das colaborações entre os pesquisadores. Além disso, observa-se intenso processo de globalização da atividade científica, em que a localização tradicionalmente concentrada dessa atividade dá espaço a uma organização mais abrangente, de modo que a produção científica espraia-se para novas regiões de intensa atividade produtiva.

⁷ Muitos estados brasileiros buscam assemelhar-se ao Estado de São Paulo, o qual possui maior tradição científica, baseada numa segurança financeira e autonomia em relação ao governo federal. Essa separação da esfera federal é fundamental, uma vez que reduz o impacto de ambientes macroeconômicos instáveis sobre o montante de recursos direcionados ao financiamento das atividades de pesquisa (CRUZ; CHAIMOVICH, 2010).

⁸ Isso ocorre devido à localização das maiores universidades e de importantes agências de fomento científico como, por exemplo, a FAPESP (CRUZ; CHAIMOVICH, 2010).

Em segundo lugar, observa-se o crescimento espetacular da produção científica dos países emergentes, tais como China, Índia e Brasil, o que fornece a esses países enorme destaque no cenário científico mundial, e faz emergir a possibilidade da constituição de uma nova ordem mundial na atividade científica nos próximos anos.

2.1.1. Publicação

A atividade científica é praticada por milhões de pesquisadores ao redor do mundo e resulta na publicação anual de cerca de 25.000 periódicos científicos. O número total de artigos científicos publicados cresceu de 460.200 em 1988 para 788.300 em 2009.⁹ Tal tendência de crescimento manifestou-se, primordialmente, na última década, uma vez que a partir de 2002, houve praticamente a duplicação do dispêndio global anual em P&D, juntamente à elevação do número de pesquisadores, o que resultou no crescimento de 35% da publicação científica mundial (RS, 2011; NSB, 2012).

A produção científica mundial é caracterizada por um recente processo de desconcentração geográfica, contrariamente à hipótese dominante no debate científico de que as chamadas cidades globais constituem verdadeiros polos de produção de conhecimento no mundo (GROSSETTI *et al*, 2012).

A paulatina perda de hegemonia científica dos países tradicionalmente líderes mundiais é evidente. Entre 1996 e 2008, Estados Unidos, Japão e Rússia perderam 20%, 22% e 24% de suas participações no total da produção mundial, respectivamente. Embora os Estados Unidos ainda liderem a produção científica mundial em termos absolutos, sua participação relativa na produção mundial é aquela que apresenta maior queda nos últimos anos, chegando ao patamar de 28%, o que significa que o país tem crescido num ritmo mais lento do que a média mundial. Em processo de desaceleração semelhante, os países da União Europeia, hoje responsáveis por cerca de 37% da produção mundial, também presenciam a redução anual de sua contribuição, apesar de que alguns países tradicionais na produção científica, como Alemanha e Inglaterra, conseguem manter

⁹ O número de publicações científicas é o indicador mais utilizado de produção científica, na medida em que as publicações refletem as novas contribuições ao estoque de conhecimento.

suas parcelas relativamente constantes. Ademais, o Japão, país que experimentou crescimento acelerado na década de 1980, apresenta expressivo declínio de sua produção científica, tanto em termos de volume como de participação no total mundial (RS; 2011; HOLLANDERS; SOETE, 2010; ADAMS; KING, 2009).

Concomitantemente, observa-se um surpreendente crescimento da atividade científica desempenhada nos países em desenvolvimento. O grupo dos BRIC (Brasil, Rússia, Índia e China) aumentaram enormemente seus dispêndios em P&D e assistem a intensificação dos esforços em C&T&I resultar num expressivo crescimento do volume da produção científica, o que proporcionou o aumento de sua participação de 17% para 24% no total da produção mundial.¹⁰ Logo, esse conjunto de países pode ser especialmente chamado de economias cientificamente emergentes.¹¹

Dentre os emergentes, a China indubitavelmente se destaca por seu crescimento espetacular. Em 1995, o país era o décimo quarto em termos da produção mundial. Já em 2007, sua produção ultrapassou a do Japão, transformando o país na segunda maior potência, somente atrás dos Estados Unidos. Em 2009, a produção chinesa alcançou participação superior a 10% no total mundial, seguida por Japão e Alemanha, com cerca de 8% cada (NSB, 2012; HOLLANDERS; SOETE, 2010).

Embora distante do padrão de crescimento chinês, o Brasil adquire crescente relevância no cenário científico internacional, com crescimento da produção bastante superior à média mundial. Entre 1981 e 2008, o número de publicações de pesquisadores brasileiros indexados nas tradicionais bases de dados internacionais *Web of Science* (WoS) e *Scopus* cresceu de 2.000 para mais de 30.000, o que acarretou no crescimento da parcela da contribuição brasileira de 0,8% em 1992 para 2,7% em 2008.¹² No mesmo período, o Brasil passou da 27^a para a 14^a colocação mundial entre os países que mais

¹⁰ A Rússia é única exceção, já que sua produção científica presenciou uma década de acentuado declínio. Entre 2002 e 2008, o país reduziu sua participação no total mundial de 3,5% para 2,7%, o que provocou a queda da sétima para a décima terceira posição no *ranking* da produção científica mundial (HOLLANDERS; SOETE, 2010; NSB, 2012).

¹¹ Segundo extrapolações lineares das tendências observadas de produção científica as produções científicas desses países, individualmente, devem superar as da França e do Japão até 2020 (RS, 2011).

¹² Entretanto, o aumento da cobertura de periódicos brasileiros por parte de bases de dados internacionais nos últimos anos do período apresentado provoca dificuldade na avaliação da dimensão exata do crescimento da produção científica brasileira (CRUZ; CHAIMOVICH, 2010; LETA, 2011).

produzem conhecimento no mundo. É interessante observar que tal processo de crescimento deu-se principalmente na última década, com crescimento médio anual ao redor de 8%. No período entre 1996 e 2008, o Brasil foi o terceiro país no mundo que apresentou maior média anual de crescimento da produção científica. A intensificação dessa trajetória de crescimento deu-se entre 2002 e 2008, período em que o crescimento da produção brasileira foi de cerca de 110% (GLÄNZEL *et al*, 2006; ADAMS; KING, 2009; CRUZ; CHAIMOVICH, 2010; GROSSETTI *et al*, 2012; RS, 2011).¹³

Regionalmente, o crescimento da produção dos países da América Latina também foi proeminente, na medida em que a parcela do total mundial produzido passou de 1,7% em 1990, para 3,8% em 2002, e 4,9% em 2008. Semelhantemente ao padrão brasileiro, a aceleração do crescimento deu-se nos anos recentes, principalmente entre 1997 e 2007, período em que a produção mais que dobrou.¹⁴ A coincidência dos períodos de crescimento acelerado é esperada, já que o crescimento da produção latino-americana sustenta-se, principalmente, no desempenho dos países mais dinâmicos, caso do Brasil, Argentina e México. Mais especificamente, o desempenho brasileiro tornou-o responsável por quase metade da produção anual latino-americana (GLÄNZEL *et al*, 2006; ADAMS; KING, 2009; HOLLANDERS; SOETE, 2010; ALBORNOZ *et al* 2010).¹⁵ A importância produtiva e o alto grau de colaboração internacional com os outros países mais dinâmicos da região fazem do Brasil elemento determinante na configuração da emergente rede científica regional na América Latina.

¹³ Leta (2011) apresenta uma revisão dos trabalhos de cientometria no Brasil, principalmente daqueles que analisam detalhadamente a evolução da produção científica brasileira.

¹⁴ Conforme descrito anteriormente, os valores apresentados referem-se à análise realizadas a partir de informações das tradicionais bases de dados internacionais *WoS* e *Scopus*; porém, tendência semelhante pode ser observada ainda de maneira mais acentuada por meio de outras bases de dados. Como exemplos, a parcela latino-americana na produção mundial praticamente dobrou entre 1997 e 2007, de acordo com a base francesa multidisciplinar *Pascal*, cresceu de 5,4% para 7,8% segundo a base CAB, e de 1,5% para 2,0% na base CA. Outras bases nas áreas de biologia, medicina, engenharia e física apontam tendências semelhantes (ALBORNOZ *et al*, 2010). Conforme será descrito posteriormente na seção 4.1.3, a avaliação efetiva da produção científica latino-americana é imperfeita, uma vez que parte significativa da produção dá-se em espanhol e português, o que faz com que não esteja representada nas principais bases internacionais, caracterizadas pela inclusão de periódicos de língua inglesa. Logo, é possível que a trajetória de crescimento apontada esteja subestimada.

¹⁵ O Brasil ampliou sua participação nas publicações latino-americanas de 41% para 47% entre 1997 e 2007 (ALBORNOZ *et al*, 2010).

2.1.2. Citação

Embora seja fundamental a avaliação de tendências referentes à quantidade de publicações científicas, também é necessária a tentativa de avaliação da qualidade da ciência produzida. Na área da cientometria, ramo do conhecimento que estuda temas relacionados ao desenvolvimento das atividades científicas, essa ambiciosa tarefa fica a cargo das análises de citações. Seu objetivo principal é a identificação e mensuração da influência dos artigos sobre a comunidade acadêmica, uma vez que o ato de um pesquisador citar trabalhos anteriores pode ser interpretado como a admissão explícita da influência do conhecimento outrora descoberto sobre suas atividades de pesquisa, o que torna tal indicador uma medida indireta da qualidade e relevância dos artigos científicos.

Análises cientométricas costumam revelar aumento sistemático na quantidade de citações ao redor do mundo, tendência determinada pela acentuada difusão do conhecimento promovida pelo desenvolvimento dos meios de comunicação, pelo crescente número de eventos científicos, e pelo aumento do perfil colaborativo e multidisciplinar próprios da atividade científica moderna (VANZ, 2009). Como exemplo, o crescimento acelerado da produção científica da China foi acompanhado pelo incremento no número médio de citações, principalmente nas áreas de engenharia, ciência da computação e química. Entretanto, apresentou-se um padrão peculiar de citações: houve queda na proporção das citações internacionais (NSB, 2012). Tal comportamento indica que grande parte da utilização do conhecimento gerado no país esteja ocorrendo internamente à comunidade científica chinesa.¹⁶

¹⁶ A avaliação do impacto dos artigos dá-se por meio da contagem das citações durante um determinado período de tempo seguinte ao ano de publicação. No caso do Brasil, o impacto médio dos artigos de pesquisadores nacionais, avaliado durante o período de 2 anos após o ano de publicação, cresceu de 1,45 citações por artigo, em 2000, para 2,05 em 2007 (UNESCO, 2010).

2.1.3. Especialização

A importância relativa entre as áreas disciplinares na produção científica também apresenta dissimilaridades entre os países. Embora a comparação entre a especialização de países seja difícil e imperfeita, evidenciam-se alguns padrões ao redor do mundo.¹⁷

Dentre os países tradicionais, observam-se padrões de especialização mais evidentes. A França possui especialização em matemática, e possui destaque nas ciências da terra e do espaço, semelhantemente à Alemanha. No Japão desenvolvem-se, primordialmente, áreas ligadas à tecnologia, como física, química e engenharia. Já os Estados Unidos e Inglaterra são os maiores especialistas mundiais em biomedicina e em ciências da terra e do espaço. Contudo, contrariamente aos padrões verificados nos países tradicionais (citados acima), os países emergentes, tais como China, Índia, Brasil e Rússia, apresentam maior diversidade. A Rússia possui especialização em engenharia, física, matemática e ciências da terra e do espaço; a Coreia do Sul busca desenvolver áreas como a medicina e a biologia; e a China verifica seu crescimento acelerado solidificado nas ciências tradicionais, como física, matemática e engenharia, embora seja evidente uma reorientação recente para as ciências da vida e da saúde, principalmente a química; enquanto que na Índia, destacam-se a engenharia, biologia e química. (ADAMS; KING, 2009; HOLLANDERS; SOETE, 2010; NSB, 2012; RS, 2011). Diferentemente do padrão de especialização científica dos países emergentes, o Brasil possui reconhecimento internacional primordialmente em disciplinas relacionadas às ciências da vida e recursos naturais. Particularmente, destacam-se artigos relacionados às áreas de medicina, física, biologia, química e agricultura.¹⁸

De modo abrangente, os padrões de especialização científica parecem estar associados a diversos fatores, tais como a necessidade de solução de problemas locais (medicina tropical), a presença de oportunidades geográficas (ciências da terra e do espaço e

¹⁷ Para uma descrição da complexidade dos problemas relacionados veja FAPESP (2011a, p. 4-11).

¹⁸ O crescimento da pesquisa brasileira nessas áreas levou à denominação do país de “*the natural knowledge economy*”, i.e., a economia do conhecimento natural. Na área de medicina, o reconhecimento internacional dá-se, principalmente, nas áreas de medicina tropical e parasitologia, em que a produção brasileira é superior a 10% da produção mundial. Outras áreas, como agricultura e medicina veterinária, são responsáveis por cerca de 5% da produção mundial (ADAMS; KING, 2009).

biologia), existência de afinidades históricas e culturais (matemática e física), e concentração da atividade industrial (química) (UNESCO, 2010).

2.1.4. Localização

Conforme vimos, a atividade científica é distribuída de maneira bastante dessemelhante no mundo, visto que a geografia internacional da pesquisa é marcada pela intensa concentração em poucos países, embora se verifiquem evidências de transformações rumo à desconcentração espacial, na medida em que se configura uma extensa rede global de produção de conhecimento. Ademais, concentrações espaciais também são observadas internamente aos territórios nacionais, onde é comum a evidência de padrões heterogêneos de localização geográfica da atividade científica.

Entre os principais países produtores de conhecimento, a localização das atividades científicas é caracterizada por intensa heterogeneidade espacial, *i.e.*, a produção de conhecimento dá-se de maneira assaz concentrada em seus territórios nacionais. Na Rússia, a capital Moscou é origem de metade da produção científica do país. Cidades como Praga, na República Tcheca, Budapeste, na Hungria, e Buenos Aires, na Argentina, localizam cerca de 40% de suas produções nacionais, enquanto Londres, na Inglaterra, Pequim, na China, e Paris, na França, são responsáveis por cerca de 20% dos totais nacionais. Tal concentração espacial sistemática leva à utilização crescente das cidades como unidades espaciais de análise em estudos cientométricos (RS, 2011).

No Brasil, também se verifica enorme heterogeneidade espacial das atividades de pesquisa científica, onde o padrão regional da distribuição das publicações científicas e dos pesquisadores é altamente concentrado na região Sudeste, com destaque às capitais dos estados. Como exemplo, a cidade de São Paulo concentra cerca de 20% da produção científica brasileira. Durante a década passada, a cidade cresceu 21 posições na lista das cidades que mais produzem conhecimento no mundo, passando a figurar dentre os 20 municípios que mais produzem conhecimento no mundo (RS, 2011), e destacou-se entre as cidades que mais apresentaram crescimento na produção científica e no padrão de

citações (MATTHIESSEN *et al*, 2010). Esses fatos são reflexos do crescimento acelerado da produção científica brasileira e da importância da cidade como a capital do estado com maior tradição científica no país.¹⁹

No caso brasileiro, a concentração espacial está diretamente relacionada à localização dos *campi* das universidades públicas, primordialmente as estaduais e federais, uma vez que essas são responsáveis pela maioria da produção científica brasileira, padrão típico de países em desenvolvimento. Além disso, a colaboração doméstica é dominada pela interação entre universidades públicas, e entre essas e institutos públicos de pesquisa, na medida em que esse conjunto de instituições emprega a maioria dos pesquisadores brasileiros. Em 2009, somente sete universidades, localizadas nas regiões Sudeste e Sul do país, foram responsáveis por cerca de 60% da produção publicada em periódicos internacionais.²⁰ Dentre elas, quatro possuem *campi* universitários localizadas no Estado de São Paulo. Em 2009, a USP produziu cerca de um quarto da produção brasileira seguida pela UNESP e Unicamp com cerca de 8% (GLÄNZEL *et al*, 2006; LETA *et al*, 2006; UNESCO, 2010; FAPESP, 2011a, 2011b).

2.1.5. Colaboração

A ciência moderna possui como característica principal o aumento do perfil colaborativo, verificado em todas as suas áreas. Atualmente, cerca de três quartos dos artigos científicos produzidos no mundo estão associados a autores de diferentes instituições, *i.e.*, são provenientes de colaborações entre pesquisadores (RS, 2011). Desse total, cerca de 44% é oriundo de esforços colaborativos entre pesquisadores de diferentes países, e 56% de colaborações entre pesquisadores em território nacional (GLÄNZEL; SCHUBERT, 2004; RS, 2011). Sua importância é notável, uma vez que as análises cientométricas usualmente apontam para a colaboração como um fator

¹⁹ A heterogeneidade do sistema de inovação paulista é evidente. Como indicador de sua especificidade, a intensidade do dispêndio total em P&D foi de 1,52% do PIB paulista em 2008. Esse valor é maior do que aqueles calculados tanto para o Brasil como para outros países em crescimento como China, Rússia e Índia. Em valores absolutos, o dispêndio total em P&D do estado supera o de países como Argentina, Israel e Bélgica (FAPESP, 2011).

²⁰ A parcela conjunta gerada por USP, Unicamp, UNESP, UFMG, UFRGS, UFRJ e Unifesp era de 60% em 2000, subiu para 66% em 2003, para 71% em 2007, e voltou para 60% em 2009 (UNESCO, 2010).

potencial de acréscimo de qualidade aos trabalhos científicos (SINGH; FLEMING, 2010).

2.1.5.1. Internacional

Em 1988, apenas cerca de 8% da produção científica mundial era fruto de parcerias internacionais. Já em 2009, tal parcela subiu para cerca de 23%, em grande parte sustentada pelo crescimento da colaboração nos países de maior produção (RS, 2011). Como exemplo, as colaborações entre os países na Europa aumentaram e espraiaram-se por todo o continente nesse período, principalmente devido a existência de políticas explícitas de estímulo à colaboração na região.²¹

Cerca de um quarto da produção científica dos Estados Unidos é proveniente de colaborações internacionais. Isso faz do país o principal parceiro nas colaborações internacionais ao redor do mundo, em grande parte devido a sua enorme capacidade produtiva, de forma que pesquisadores norte-americanos estejam envolvidos em 17% de toda a produção mundial oriunda de esforços colaborativos internacionais (RS, 2011).

O Brasil encaixa-se no quadro de tendência ao aumento da colaboração científica. Contudo, o produto das colaborações entre pesquisadores brasileiros e estrangeiros tem se mantido constante nos últimos anos, com períodos de ligeira queda (VANZ, 2009; FAPESP, 2011a).²²

Os Estados Unidos figuram como principal parceiro científico do Brasil. Pesquisadores norte-americanos participaram em cerca de 11% da produção brasileira entre 2003 e

²¹ Os *Framework Programmes* (FPs) são os principais instrumentos de integração dos sistemas nacionais de inovação na Europa, e consistem em projetos colaborativos internacionais em P&D que visam integrar a base científica e tecnológica para fomentar a competitividade internacional e promover pesquisas de apoio às políticas da União Europeia (SCHERNGELL; BARBER, 2009).

²² O aumento do perfil colaborativo da ciência brasileira possui algumas características bastante peculiares, tais como os efeitos causados pela atuação de agências de fomento (CAPES, CNPq) na avaliação dos programas de pós-graduação, em que são definidas metas exigentes de produção acadêmica. Embora incentive o crescimento da produção científica, a necessidade de publicar pode também provocar o estímulo ao compartilhamento de autorias (VANZ, 2009; LETA, 2011).

2007, seguidos pelos ingleses, franceses e alemães, com cerca de 3% cada.²³ Dentre os países latino-americanos, Argentina, México e Chile colaboraram em cerca de 1,7%, 0,8% e 0,7% das publicações brasileiras, respectivamente (GLÄNZEL; SCHUBERT, 2004; GLÄNZEL *et al*, 2006; VANZ, 2009; ADAMS; KING, 2009).²⁴

Dentre os próprios países emergentes, verifica-se crescimento no número de interações em pesquisa colaborativa. No âmbito da rede formada entre os países da América Latina, as colaborações entre Brasil e Argentina, e Argentina e México, estão entre as maiores do mundo em termos de indicadores da intensidade colaborativa entre países. Já entre os países asiáticos a colaboração é bem maior do que a esperada, principalmente entre Índia, Japão, China, Coreia do Sul, Cingapura e Taiwan, rede que tem como produto principal publicações na área de engenharia.²⁵ As evidências de ampliação dessas redes regionais, fundamentada no potencial produtivo dos países emergentes, pode ser indício da constituição de uma nova dinâmica na atividade científica mundial nos próximos anos. Nesse contexto, a sustentação da produção científica mundial no longo prazo parece depender de forma crucial da expansão e integração das redes na América Latina e Ásia (NSB, 2012; ADAMS; KING, 2009; RS, 2011).

A importância da colaboração internacional é amplamente reconhecida na formulação de políticas de C&T&I, visto que pode viabilizar o aumento da qualidade da ciência doméstica por meio da absorção de novos conhecimentos, além de consistir em maneira eficaz de acesso ao conhecimento desenvolvido a partir de investimentos em pesquisa de outros países. Isso faz com que a colaboração na atividade científica seja interpretada diferentemente de um jogo de soma zero²⁶, na medida em que o crescimento acelerado dos países emergentes não esteja atrelado a possíveis reduções da atividade científica nos países tradicionais. De modo contrário, a ambição de vultosos investimentos num

²³ As colaborações internacionais seguem os padrões de especialização dos países, uma vez que a maior parte da colaboração entre pesquisadores brasileiros e americanos dá-se na área de ciências biológicas, enquanto com os franceses há predominância das ciências exatas (VANZ, 2009).

²⁴ Todavia, o Brasil possuiu a menor participação de colaboração internacional na comparação com esses países (GLÄNZEL *et al*, 2006). Dentre os países europeus, destaque pode ser feito a Portugal, que mais que dobrou suas colaborações com o Brasil nos últimos anos, crescimento que pode estar relacionado a fatores de aproximação linguística e cultural entre os países (RS, 2011).

²⁵ Enquanto a colaboração entre China e Índia permaneceu praticamente constante, houve recentemente aumento expressivo das parcerias entre a Índia, Japão e Coreia do Sul (NSB, 2012).

²⁶ Nome dado à situação em que a vitória, ou melhoria de bem-estar por parte de um jogador está inequivocamente condicionada à derrota ou perda de bem-estar do outro jogador.

futuro próximo, por parte dos países cientificamente emergentes, favorece que o apoio à colaboração internacional seja estratégia política prioritária nos países mais tradicionais, uma vez que a criação de novas oportunidades pode estimular o crescimento de suas atividades científicas internas (RS, 2011; ADAMS; KING, 2009).

Apesar dos benefícios inerentes a qualquer espécie de colaboração científica entre pesquisadores, boa parte das iniciativas rumo à internacionalização da ciência é fundamentada em resultados empíricos que apontam a existência de uma relação positiva entre a colaboração e o impacto contabilizado pelas citações (GLÄNZEL, 2001; LETA; CHAIMOVICH, 2002). Todavia, é interessante a percepção de que a relação entre internacionalização e aumento da qualidade, mensurada pela frequência de citação, é mais tênue do que aparenta. Isso se dá na medida em que a coautoria múltipla pode exercer um impacto de rede, já que é provável que o trabalho consolidado seja visto por um maior número de pessoas, de modo a resultar na ampliação do número de citações (RS, 2011; VANZ, 2009). Logo, a utilização desse indicador deve ser realizada com cautela, já que seu aumento pode não estar associado diretamente à maior qualidade, e sim à ampliação dos mecanismos de difusão internos à comunidade científica.

2.1.5.2. Doméstica

A colaboração doméstica consiste na interação de esforços colaborativos entre pesquisadores localizados dentro dos limites nacionais. Possui importância reduzida nas caracterizações típicas da evolução da produção científica mundial, visto que o crescimento da produção é costumeiramente atrelado ao crescimento da importância da colaboração internacional (VANZ, 2009; NSB, 2012). Contudo, algumas questões merecem atenção especial. Será que a tendência rumo à internacionalização da ciência está provocando a redução da importância relativa das colaborações domésticas pelo mundo? Ainda mais importante, será que o crescimento acelerado dos países emergentes deu-se impulsionado pela internacionalização da ciência nesses países?

A princípio, observa-se que a evolução da colaboração internacional apresenta padrões dissimilares entre os países. Por um lado, a colaboração internacional dos Estados Unidos cresceu de maneira acelerada. Entre 1990 e 2010, a parcela da produção norte-americana oriunda de colaborações internacionais saltou de 12% para 30%, crescimento superior ao da participação média da colaboração internacional pelo mundo, que passou de 10% para 22%. Fenômeno semelhante ocorre nos países europeus tradicionais na produção científica, o que pode ser indício de que o crescimento da participação média da colaboração internacional no mundo é, em grande parte, impulsionado pelo acelerado processo de internacionalização da ciência nesses países. Por outro lado, apesar da participação da colaboração doméstica nos Estados Unidos ter se mantido praticamente constante, tal participação cresceu ao redor do mundo, numa tendência de crescimento semelhante ao crescimento da colaboração internacional. Fenômeno semelhante ocorre nos países europeus cientificamente tradicionais, onde a participação da colaboração doméstica apresenta tendência de estagnação ou declínio (NSB, 2012). Tais considerações estimulam a análise mais cuidadosa sobre a evolução da relação entre a colaboração internacional e doméstica nos países cientificamente emergentes.

É interessante notar que os países cientificamente emergentes costumam apresentar níveis de colaboração internacional muito menores do que o observado nos países tradicionais. A participação da colaboração doméstica sobre a produção científica dos países emergentes manteve-se praticamente constante, ao redor de 70% desde 1990, *i.e.*, cerca de três quartos da produção científica de países como China, Índia, Brasil são oriundos dos esforços colaborativos entre os pesquisadores locais, sem nenhuma colaboração por parte de estrangeiros (LETA; CHAIMOVICH, 2002; RS, 2011).²⁷ Logo, o crescimento acelerado da produção científica nesses países parece não estar tão associado à intensificação da colaboração internacional, e sim diretamente ligado ao crescimento também acelerado da colaboração doméstica.²⁸

²⁷ Como exemplo, o grau de internacionalização da produção da China manteve-se praticamente constante ao redor de 25%, bastante distante daquele observado nos países tradicionais na produção científica. Outros países de crescimento científico acelerado, como Coreia do Sul e Turquia também apresentaram comportamento semelhante (NSB, 2012; RS, 2011).

²⁸ Como exemplo, a colaboração interestadual no Brasil cresceu 79,4% entre 2002 e 2006, bem acima do crescimento total da produção científica (FAPESP, 2011a).

Nesse contexto, é importante compreender que as diferenças nos padrões de internacionalização são oriundas da disponibilidade de recursos, escala da comunidade científica em cada país, e de fatores associados à proximidade entre os colaboradores.²⁹ No caso particular dos países emergentes, alguns fatores auxiliam na explicação do baixo padrão de internacionalização da colaboração científica. Por um lado, a existência de barreiras linguísticas e de grandes distâncias geográficas e culturais certamente dificulta a interação colaborativa com parceiros estrangeiros. Por outro lado, a dimensão territorial continental e o elevado número de pesquisadores nesses países podem favorecer a busca por colaboradores dentro dos limites territoriais nacionais.

A verificação da intensificação dos esforços colaborativos domésticos, aliada a fatores geográficos e culturais que atuam como mecanismos incentivadores desse processo, ressalta a importância da colaboração doméstica como um dos principais motores do crescimento científico nos países emergentes. Além disso, o fomento da articulação entre pesquisadores nacionais na produção compartilhada de conhecimento técnico-científico pode desempenhar papel primordial no tocante aos desafios das políticas de C&T&I, uma vez que consiste num instrumento essencial para a desconcentração regional da produção científica e o incremento de qualidade do produto gerado. Logo, faz-se oportuna a tentativa de promover a melhor compreensão dos determinantes das redes de colaboração nesses países, como é o caso do Brasil.

2.2. Cientometria espacial

A tentativa de estudar os padrões geográficos da colaboração científica normalmente faz uso do ferramental da área de geografia econômica, primordialmente o ramo chamado de geografia econômica evolucionária. Esta seção visa caracterizar os principais

²⁹ A maior interação com pesquisadores estrangeiros é facilitada pela maior disponibilidade de recursos que auxiliem a comunicação entre os pesquisadores, assim, a baixa disponibilidade de recursos pode dificultar ou até mesmo inviabilizar a colaboração internacional. Já a menor escala da comunidade científica local pode consistir num forte mecanismo de estímulo à colaboração internacional. Uma vez que dificulta a colaboração doméstica. Quanto à proximidade espacial, países pequenos apresentam usualmente níveis mais elevados de internacionalização da ciência. Como exemplo, Bélgica, Holanda e Dinamarca tiveram mais da metade da produção entre 2004 e 2008 oriunda de colaborações internacionais (RS, 2011).

aspectos dessa abordagem, de modo a destacar os motivos que a fazem alternativa aos métodos tradicionais de geografia econômica, além de apresentar o ferramental tipicamente empregado em análises espaciais de redes de colaboração científica.

A geografia econômica pode ser definida como o campo de estudo que lida com a identificação dos determinantes da distribuição irregular das atividades econômicas no espaço. De maneira sucinta, o desafio de responder as questões pertinentes é enfrentado por duas vertentes teóricas principais: a geografia econômica institucional e a nova geografia econômica. Embora ambas sejam bastante influentes e correntemente aplicadas na área, interpretam diferentemente as questões usuais e proporcionam métodos de análise de certa maneira conflitantes. Por um lado, a geografia econômica institucional é dominada por pesquisadores com sólida base de conhecimento geográfico e com certa familiaridade à economia institucional, de modo que as irregularidades referentes à distribuição da riqueza entre os territórios no espaço são interpretadas como oriundas de diferenças institucionais entre eles. Por outro lado, a nova geografia econômica teve seu desenvolvimento associado à economia neoclássica e interpreta as desigualdades espaciais das atividades econômicas como advindas de processos de aglomeração, os quais são fontes de retornos crescentes.³⁰ Embora as abordagens e interpretações não sejam necessariamente divergentes, a geografia econômica é caracterizada pelo distanciamento e dificuldade de conciliação ou aproximação entre as duas correntes, principalmente devido à existência de duas incomensurabilidades determinantes: a distinção entre os procedimentos metodológicos adotados e a diferenciação teórica entre os pressupostos fundamentais a respeito da interpretação do comportamento econômico. No ramo metodológico, a geografia econômica institucional costuma realizar pesquisas indutivas utilizando estudos de caso e dispensa inicialmente a utilização de modelagem formal, uma vez que enfatiza as especificidades intrínsecas aos locais analisados. Por outro lado, a nova geografia econômica descarta as pesquisas de estudos de caso e faz uso de modelos formais, de forma dedutiva, fundamentados em análises de equilíbrio provenientes do

³⁰ A nova geografia econômica é caracterizada por notável sucesso teórico, mas também pela dificuldade de aplicações práticas que verifiquem a validade das proposições teóricas enunciadas. Contudo, os casos de sucesso costumam confirmar muitos de seus resultados teóricos. Um panorama sobre a história e desafios da nova geografia econômica pode ser encontrado em Fujita e Krugman (2004).

comportamento de agentes econômicos representativos. Já quanto à interpretação do comportamento econômico observado, a utilização de modelos formais pela nova geografia econômica busca explicar os padrões geográficos das atividades econômicas a partir de ações tomadas pelos agentes econômicos individuais, advindas dos processos racionais de maximização de utilidade por parte dos indivíduos e de maximização do lucro por parte das firmas. De maneira contrária, a geografia econômica institucional baseia-se na premissa de que o comportamento econômico deve ser compreendido como fruto de escolhas delimitadas por regras, o que implica na suposição de racionalidade limitada dos agentes, visto que seus comportamentos econômicos são totalmente dependentes da maneira que os arranjos institucionais moldam e guiam suas escolhas. Dessa maneira, o fato das instituições estarem imersas em locais reais e geograficamente determinados torna as localidades como unidades relevantes de análise (BOSCHMA; FRENKEN, 2007).³¹

Diante do contexto de dificuldade na conciliação entre as vertentes teóricas descritas, emergiu uma nova corrente, a geografia econômica evolucionária, a qual pode ser considerada uma terceira abordagem no âmbito de pesquisa em geografia econômica.

2.2.1. Geografia econômica evolucionária

A geografia econômica evolucionária pode ser entendida como produto da síntese entre a economia evolucionária e a geografia econômica. Assim, enquanto as outras vertentes destacam o processo de mobilidade dos fatores ou enaltecem a unicidade das instituições em territórios específicos, a geografia econômica evolucionária propõe a interpretação da economia como um processo evolutivo que se desdobra no espaço e no tempo. Isso faz com que a compreensão das fontes de desigualdades no desenvolvimento econômico seja realizada no contexto do processo schumpeteriano de destruição criativa em diferentes níveis de agregação espacial (cidades, regiões, países,

³¹ Pode-se dizer que as instituições não desempenham papel determinante nos modelos neoclássicos, sendo encontradas somente de maneira vaga e implícita (BOSCHMA; FRENKEN, 2007).

continentes).³² Assim, a fundamentação teórica sob a noção schumpeteriana da inovação como agente principal das transformações endógenas ao sistema econômico produz interesse por temas ligados à inovação tecnológica, principalmente relacionados à dimensão espacial dos *spillovers* de conhecimento da academia para o setor industrial, e da formação de redes interativas entre os agentes de um sistema nacional de inovação. De maneira geral, as abordagens dessa nova corrente teórica apresentam avanços importantes em relação à interpretação do sistema econômico. Por um lado, diferem da nova geografia econômica, na medida em que superam os limites impostos por essa, tanto advindos dos pressupostos sobre o comportamento dos agentes econômicos, como da redução da importância da geografia à existência de custos de transporte. Por outro lado, a interpretação das diferenças territoriais como provenientes não somente de diferenças institucionais, mas também dependentes de fatores evolucionários (como as distintas trajetórias históricas das firmas e indústrias de um território específico) promove o distanciamento em relação à abordagem institucional tradicional. As aplicações abarcam diferentes níveis de análise dos atores: unidades microeconômicas (firmas), macroeconômicas (redes e sistemas espaciais) e estruturas intermediárias (indústrias), bem como a análise de unidades espaciais. Nesse sentido, as localidades determinadas nos sistemas espaciais, como cidades, regiões ou países, podem ser caracterizadas em meio a suas posições nas redes espaciais, de modo a permitir a análise de mudanças estruturais ao longo do tempo (BOSCHMA; FRENKEN, 2007).

Pode-se afirmar que a geografia econômica evolucionária possui avanços metodológicos significativos, resultantes da tentativa de conciliação entre as metodologias tradicionais, por meio da utilização de ferramentais ligados a ambas as áreas, desde estudos de caso, métodos econométricos, exercícios teóricos de modelagem e avaliação de políticas. Além dessas, destaca-se o emprego do ferramental da análise de redes sociais, o qual permite analisar a estrutura e evolução das redes de interações, bem como o efeito da posição das unidades de análise nas redes sobre seu próprio desempenho. Dentre as possíveis aplicações, pode-se dizer que a análise de redes sociais consiste no mecanismo fundamental para a compreensão dos fenômenos relacionados

³² O processo schumpeteriano de destruição criativa pode ser entendido como o processo de mutação industrial que revoluciona incessantemente o interior da estrutura econômica, destruindo continuamente os antigos elementos e criando novos (SCHUMPETER, 1961).

aos *spillovers* de conhecimento, uma vez que elas atuam como principais veículos para a transmissão e difusão do conhecimento. De maneira abrangente, a partir da percepção de que as transformações econômicas não ocorrem de maneiras uniformemente distribuída no espaço, a geografia econômica evolucionária utiliza os conceitos e ideias da economia evolucionária para o entendimento da mudança da configuração econômica ao longo do tempo, e ressalta a importância da localização espacial das atividades econômicas, uma vez que temas geográficos podem ser determinantes na compreensão da natureza e trajetória de evolução dos sistemas econômicos. Portanto, ao entender o processo contínuo de geração e difusão do conhecimento como fator que impulsiona a evolução do sistema econômico, torna-se fundamental a exploração das redes de geração e difusão do conhecimento no espaço geográfico. Dessa maneira, contrariamente aos estudos de *spillovers* de conhecimento que tratam o espaço geográfico como uma caixa-preta, a análise volta-se à determinação do papel da proximidade geográfica como facilitadora da formação de ligações em redes de conhecimento (BOSCHMA; FRENKEN, 2007).

2.2.2. Dimensões de proximidade

Recentemente, o interesse pelo potencial das análises de redes em estudos empíricos vem crescendo de maneira acelerada, e seu emprego em pesquisas sobre inovação contribuiu para o surgimento de uma ampla variedade de conceitos e metodologias. No entanto, a incorporação de conceitos na geografia da inovação ainda está em processo, *i.e.*, a geografia econômica evolucionária ainda não utiliza esse ferramental de maneira consolidada (BOSCHMA; FRENKEN, 2010) e a melhor utilização desse instrumental para a análise da evolução e desempenho das redes de inovação e de conhecimento está intimamente atrelada ao entendimento do conceito de proximidade, a qual é compreendida de forma mais abrangente do que sua simplificação intuitiva à proximidade geográfica. Nesse contexto, é possível lidar com uma das questões mais proeminentes no campo da geografia econômica: a determinação do impacto da proximidade geográfica no processo de geração e difusão do conhecimento. Contudo, a

importância da proximidade geográfica não pode ser verificada de maneira isolada, mas sim examinada comparativamente a outras dimensões de proximidade.

O desenvolvimento de conceitos relacionados à proximidade é proveniente da chamada escola francesa de proximidade.³³ A principal contribuição dessa abordagem consiste na tentativa de dissociação entre a proximidade física (geográfica) e outras formas de proximidade. A partir dessa interpretação, e alicerçado numa extensa revisão sobre a literatura do papel da proximidade sobre a colaboração em atividades inovativas, Boschma (2005) elaborou as definições e distinções entre as dimensões geográfica, cognitiva, organizacional, social e institucional de proximidade. Apresentaremos definições sucintas dessas dimensões, de modo a destacar suas importâncias na análise espacial da colaboração científica.

A proximidade geográfica (ou física) é medida pela distância espacial entre as unidades de análise. Isso pode ser realizado por meio de medidas contínuas ou categóricas. Medições contínuas são realizadas em termos absolutos, através da medição da distância em quilômetros (HOEKMAN *et al*, 2009), ou em termos relativos, através da medição do custo ou tempo de viagem. Já a medição de maneira categórica dá-se através da construção de uma variável binária que diferencie as relações domésticas e estrangeiras (PONDS *et al*, 2007). A fraca proximidade geográfica pode ser encarada como indicador de ausência de externalidades espaciais, uma vez que é esperado que o crescimento da proximidade geográfica implique na maior interação entre os atores de uma rede. Assim, no caso da colaboração científica, espera-se que a proximidade geográfica seja um condicionante das relações entre os pesquisadores, já que a redução da distância facilita o contato entre eles, e deve, portanto, estimular a colaboração (FRENKEN *et al*, 2009).

Já a proximidade institucional lida com o fato de que os atores de uma rede podem partilhar do mesmo arcabouço institucional, seja esse de ordem formal (leis e regras) ou informal (valores, normas e hábitos culturais). Assim, pode ser entendida como o grau de similaridade entre as estruturas de incentivos que moldam a conduta e os principais

³³ Essa denominação refere-se ao grupo de economistas franceses interessados na análise da evolução espacial das atividades econômicas (CARRINCAZEUX *et al*, 2008).

objetivos dos pesquisadores, já que tais estruturas podem direcionar seus comportamentos de maneira sistemática. Nos estudos de cientometria espacial, é usual a diferenciação entre as seguintes esferas institucionais: universidades, indústria e governo (FRENKEN *et al*, 2005; PONDS *et al*, 2007), em virtude dos distintos papéis desempenhados por esses no processo de geração e difusão do conhecimento. Enquanto as universidades são as principais responsáveis pela transformação primária de pesquisas em publicações, os pesquisadores no setor privado são os principais responsáveis na transformação dos resultados de pesquisas em conhecimento passível de ser comercializado e os pesquisadores no governo oferecem caminhos para a transformação de conhecimento em políticas públicas de maneira eficiente, o que faz das relações entre essas esferas institucionais tema fundamental na literatura da economia da C&T&I (FRENKEN *et al*, 2009).

A proximidade cognitiva pode ser entendida como o grau de similaridade pelo qual dois pesquisadores compartilham da mesma base de conhecimento. Sua importância consiste na ideia de que a efetividade da transferência de conhecimento numa colaboração requer certo nível de capacidade absorptiva para a identificação, processamento, compreensão, interpretação e exploração do novo conhecimento. Assim, faz-se necessário que as bases cognitivas dos pesquisadores sejam próximas o suficiente para o sucesso na comunicação e entendimento por ambas as partes, de modo que a proximidade cognitiva entre pesquisadores funcione como pré-requisito essencial para o processo interativo de aprendizado e colaboração. Logo, consiste num mecanismo fundamental engajador de interações de conhecimento, tanto em redes de atividades inovativas, como também nas colaborações científicas (BOSCHMA, 2005). Inicialmente, é esperado que o aumento da proximidade cognitiva entre os pesquisadores contribua para a maior efetividade e qualidade da pesquisa colaborativa. Entretanto, a colaboração entre pesquisadores com maior distância cognitiva pode levar a contribuições relativamente radicais ou até mesmo à consolidação de novas áreas científicas (FRENKEN *et al*, 2009).

Já a dimensão organizacional da proximidade refere-se ao conjunto de interdependências entre as unidades da rede que possuam algum tipo de ligação institucional ou econômica em relação a determinado sistema de controle (BOSCHMA,

2005). Ocorre, por exemplo, quando as unidades conectadas estão submetidas a um mesmo sistema de controle financeiro central. No âmbito da colaboração científica, essa dimensão pode ser definida pelo grau de similaridade pelo qual dois pesquisadores estão sujeitos à mesma estrutura de controle hierárquico. Isso é de fundamental importância para a coordenação das atividades de pesquisa, visto que a colaboração torna-se mais difícil em situações em que ambos os pesquisadores estejam submetidos a diferentes sistemas de regras e padrões existentes em suas organizações. Na avaliação empírica, é possível a construção de uma variável que identifique a dicotomia entre pessoas que trabalhem para a mesma, ou para diferentes organizações, estrutura comumente empregada em estudos sobre o impacto da colaboração científica sobre citações (FRENKEN *et al*, 2009).

Além das dimensões citadas, a proximidade social pode ser definida pela intensidade em que dois pesquisadores possuem relações de amizade entre si. Sua importância advém do fato de que a intensificação das relações de amizade pode facilitar a interação através da criação de uma relação de confiança entre os pesquisadores, o que é fundamental na continuidade de projetos de pesquisa complexos. Essas relações sociais podem se originar da vida privada ou profissional dos pesquisadores envolvidos, e sua mensuração é analisada tipicamente nos trabalhos empíricos que identificam padrões de citação entre pesquisadores (FRENKEN *et al*, 2009). Contudo, o excesso de proximidade social pode não ser benéfico, pois pode levar a perda da racionalidade econômica, visto que as relações de mercado podem ficar intrinsecamente ligadas às sociais (BOSCHMA, 2005).

As dimensões da proximidade descritas são essenciais por fornecerem soluções para o problema fundamental de coordenação, relacionado à incerteza inerente ao processo de geração de conhecimento e inovação. Entretanto, é crucial o entendimento das dimensões de proximidade como passíveis de serem substituídas entre si, o que possibilita que a proximidade excessiva seja capaz de exercer impactos negativos sobre a interação entre os agentes, *i.e.*, o conceito de proximidade inclui a ideia de que a maior proximidade em uma dimensão pode implicar em distanciamento em outra. Como exemplo, a proximidade geográfica é necessária em muitas formas de interações

científicas, mas é esperado que seja menos importante caso dois pesquisadores estejam próximos sob a dimensão cognitiva. Dessa caracterização decorre o chamado paradoxo da proximidade, o qual retrata a existência de um nível ótimo de proximidade entre os agentes, uma vez que o aumento da proximidade entre dois agentes em determinado dimensão pode provocar maior distanciamento em outra dimensão.

Outro aspecto interessante é a possibilidade de interação entre as diferentes dimensões de proximidade ao longo do tempo. Como exemplo, a proximidade geográfica pode fortalecer as outras dimensões de proximidade com o passar do tempo, já que pesquisadores podem desenvolver proximidade cognitiva durante um período de contato próximo, e quando eles se movem para outras organizações, a proximidade física e organizacional desaparece, mas a proximidade cognitiva e social construída no passado permite que eles continuem a colaborar efetivamente. Logo, a avaliação dos papéis de proximidade necessita da adoção de uma abordagem dinâmica, na medida em que as interações de substitutabilidade e complementariedade entre as dimensões de proximidade modificam o papel da proximidade geográfica ao longo do tempo.

A proximidade geográfica por si só não é condição necessária nem suficiente para o processo de aprendizagem (BOSCHMA, 2005), mas a maior proximidade geográfica consiste num importante mecanismo facilitador do processo de interação, principalmente quando é simultânea a outras dimensões da proximidade. Em teoria, a ocorrência de proximidade geográfica concomitantemente a um limiar mínimo de proximidade cognitiva é suficiente para o sucesso da interação e aprendizagem.

A partir da discussão acima emerge a ideia de que as outras formas de proximidade podem atuar como substitutos da proximidade geográfica, o que torna complexa a análise e verificação empírica de seus efeitos. O impacto efetivo da proximidade geográfica somente pode ser avaliado com êxito quando controlado pelas outras dimensões de proximidade, de forma a promover o isolamento entre essas, já que elas podem atuar como substitutas eficazes da proximidade geográfica (BOSCHMA; FRENKEN, 2010).

2.2.3. Proximidade e cientometria

As análises espaciais da ciência, área denominada de cientometria espacial, visam verificar empiricamente de que forma a atividade científica, primordialmente a geração e difusão do conhecimento, são localizados espacialmente. Logo, a determinação de padrões espaciais da atividade científica consiste no tema central da área. Dentro desse contexto, a análise dos efeitos da proximidade geográfica sobre a configuração dos padrões espaciais constitui-se problema típico. Contudo, a pesquisa na área ainda é bastante fragmentada, caracterizada pela adoção de diversas metodologias, níveis distintos quanto às unidades de análise (pesquisadores, institutos de pesquisa, regiões, países), noções de distância e abordagens exploratórias. Além disso, as análises usuais não levam em consideração as diferentes formas de proximidade que determinam os padrões de colaboração científica, o que prejudica de forma determinante os resultados alcançados. Como exemplo, estudos que apontavam o maior impacto da colaboração internacional sobre citações, quando comparado à colaboração doméstica, tiveram seus resultados abruptamente reduzidos quando houve controle de outras dimensões de proximidade, como a social e organizacional (FRENKEN *et al*, 2009). Logo, é indispensável que sejam incorporadas as múltiplas dimensões de proximidade em análises nessa área de pesquisa, *i.e.*, deve-se incluir o maior número possível de dimensões de proximidade de maneira, uma vez que, como veremos, as dimensões podem estar correlacionadas.³⁴ Entretanto, a incorporação das diferentes dimensões de proximidade está atrelada à dificuldade de obtenção dos dados necessários para as unidades de análise selecionadas.³⁵ Ponds *et al* (2007) verificaram o papel da proximidade geográfica mediado pela proximidade institucional e concluíram que a colaboração científica em pesquisa entre organizações institucionalmente diferentes é mais geograficamente localizada do que entre organizações semelhantes. Já para a análise particular de determinadas disciplinas, o papel da inclusão da dimensão da

³⁴ Por exemplo, as proximidades linguística, étnica, ideológica e etária também podem desempenhar papel na determinação das interações científicas, na medida em que o domínio da mesma língua, a similaridade entre os costumes, culturas e ideologias, e até mesmo a proximidade entre as idades dos pesquisadores podem facilitar a colaboração científica entre eles.

³⁵ Qualquer tentativa de mensuração de um tipo de proximidade para quaisquer n unidades de análise requer a observação de $\left[\frac{(n^2-n)}{2}\right]$ pares de relações entre as n unidades.

proximidade cognitiva foi considerado de menor importância (HOEKMAN *et al*, 2009; PONDS *et al*, 2007).

Dentre as possibilidades metodológicas, a análise estatística de regressão permite a utilização das dimensões de proximidade como variáveis independentes e as interações científicas como variável dependente, mensurada pela frequência de colaborações, frequência de citações ou fluxos de mobilidade. Isso é feito essencialmente por meio da aplicação dos denominados modelos gravitacionais (PONDS *et al*, 2007; FRENKEN *et al*, 2009; HOEKMAN *et al*, 2009). Conforme veremos, a maior vantagem dessa abordagem é que as interações científicas são avaliadas num ferramental estatístico multivariado, de modo a permitir que as múltiplas dimensões de proximidade sejam consideradas simultaneamente. Assim, a avaliação correta do efeito de uma forma particular de proximidade (por exemplo, a proximidade geográfica) é alcançada de maneira mais precisa, devido ao isolamento e controle do efeito das outras dimensões. Ademais, incorporam-se as interações entre as dimensões de proximidade, de modo a lidar com a possível correlação significativa entre elas, o que pode ser interpretado como uma mudança dinâmica oriunda do reforço mútuo entre as dimensões de proximidade.

2.3. Colaboração científica e coautorias

Conforme apresentado, a compreensão e identificação de padrões de funcionamento das redes científicas são importantes para o entendimento do processo de geração e difusão do conhecimento na sociedade. A estrutura cognitiva da ciência manifesta-se por meio da rede de relacionamentos entre a comunidade científica e a ciência moderna caracteriza-se pela crescente pesquisa envolvendo esforços colaborativos, de forma que sua unidade organizacional típica passou de indivíduos isolados para a constituição de grupos de pesquisa. Apesar de mais comum nas ciências naturais, observa-se crescimento de publicações em colaboração em todas as áreas do conhecimento científico (GLÄNZEL; SCHUBERT, 2004).

São diversos os motivos que os pesquisadores possuem para colaborar. Na busca pela excelência da pesquisa, a transferência do conhecimento e de habilidades, por meio do compartilhamento do trabalho com colegas de destaque, pode promover o aumento da qualidade e visibilidade da pesquisa, as reduções do tempo dispensado e da possibilidade de erros, a obtenção ou ampliação de financiamentos para as atividades de pesquisa, entre outros (RS, 2011; VANZ, 2009).³⁶

As redes de colaboração científica podem ser analisadas sob óticas diversas, desde relações entre pesquisadores da academia e do setor industrial, entre pesquisadores de áreas científicas diferentes, ou até mesmo sob o ponto de vista de como as interações entre os pesquisadores distribuem-se pelo espaço geográfico (MALI *et al*, 2012).³⁷

2.3.1. Redes de coautorias

Dentre os mecanismos responsáveis pela articulação das relações sociais entre a comunidade científica, as redes de coautorias são particularmente importantes, uma vez que são indicadores dos fluxos de conhecimento entre os pesquisadores. No entanto, representam apenas uma faceta do processo de colaboração, pois são numerosos os casos em que as colaborações não resultam em publicações em coautoria (KATZ; MARTIN, 1997).³⁸

Embora seja indicador imperfeito, os estudos sobre redes de colaborações científicas usualmente utilizam dados sobre coautorias em publicações como medida de colaboração (WANG *et al*, 2005). Isso ocorre devido às vantagens inerentes a esse tipo de dado. Primeiramente, é elevado e crescente o estoque de artigos científicos, o que torna possível a extração e utilização de amplo conjunto de informações, bem como a coleta de dados observados em longos períodos de tempo. Em segundo lugar, é

³⁶ Bozeman e Lee (2005) apresentam diversos motivos que levam os pesquisadores a colaborar, apontados pela literatura especializada.

³⁷ Beaver (2001) faz uma análise sobre a evolução dos estudos sobre colaboração científica.

³⁸ A coautoria não é o único meio de colaboração científica. Outras formas possíveis são: compartilhamento e atividades de edição de publicações, organização conjunta de conferências científicas, supervisão compartilhada e desenvolvimento conjunto de projetos (MALI *et al*, 2012).

permitido alto grau de detalhamento das informações disponíveis (título, nome dos pesquisadores, afiliações institucionais, resumo, citações), o que permite diversas perspectivas de análise, tais como o agrupamento por área, período ou localização geográfica dos autores. No entanto, além das outras formas possíveis de colaboração científica, é importante a percepção de que os esforços de pesquisas voltados à produção de artigos não necessariamente resultam em publicações. Isso pode ocorrer devido à rejeição dos trabalhos por parte dos revisores técnicos, às restrições de tempo, ou ao custo de submissão à publicação. Outro fator importante é que as publicações podem simplesmente não fornecer contribuição efetiva ao processo de acúmulo de conhecimento, *i.e.*, a raridade de citações de um trabalho pode ser indício do pequeno valor adicionado ao estoque de conhecimento (HOEKMAN *et al*, 2009).

A evolução dos estudos das redes de coautoria deu-se paralelamente ao desenvolvimento de novas abordagens metodológicas quantitativas, desde a comparação entre estatísticas descritivas de séries temporais como também abordagens determinísticas e estocásticas.³⁹ Mais especificamente, o crescimento dos estudos de redes de coautorias foi facilitado pela ampla disponibilidade de bases de dados bibliográficos e pelo desenvolvimento e sofisticação de técnicas da análise de redes sociais, fomentadas pelos avanços matemáticos e computacionais em teoria dos grafos.⁴⁰ Contudo, embora forneçam resultados assaz interessantes sobre os padrões dos fluxos de conhecimento, os trabalhos sobre redes de coautorias recebem menos atenção do que aqueles que analisam redes de citações. As diferenças entre os objetos de estudo são evidentes, já que as redes de citações não se configuram como redes sociais entre indivíduos, pois não capturam a estrutura das interações sociais entre os pesquisadores, as quais são descritas em análises de redes de coautoria.⁴¹ Logo, essas últimas possuem informações mais importantes sobre os padrões de colaboração, além de permitirem tanto a localização dos autores em estruturas mais abrangentes da comunidade

³⁹ Veja Mali *et al* (2012) para um panorama sobre a evolução metodológica da área.

⁴⁰ Conforme veremos na seção 3.1, a análise de redes sociais refere-se à aplicação da teoria dos grafos no estudo de relacionamentos sociais, de modo a lidar com as características mensuráveis das estruturas das redes. A pesquisa na área teve forte expansão associada à facilitação no uso extensivo de dados empíricos e na criação de imagens gráficas (MALI *et al*, 2012).

⁴¹ Metodologicamente, os vértices (ou nós) de redes de coautorias usualmente representam pesquisadores e a ligação entre dois vértices (arestas) dá-se na existência de coautoria, enquanto que nas redes de citações, os vértices representam publicações e a ligação entre eles é direcionada entre determinada publicação e as outras que a citam.

científica, como a realização de análises particulares entre as diferentes áreas do conhecimento, nas quais as especificidades intrínsecas de cada área resultam em padrões bastante dissimilares daqueles da rede como um todo (MALI *et al*, 2012).⁴²

Embora esteja intimamente ligada ao crescimento da ciência, principalmente nos países cientificamente emergentes, a análise de colaborações científicas entre pesquisadores internamente aos territórios nacionais figura como tema ainda pouco explorado na literatura (VANZ, 2009; MALI *et al*, 2010; SCHERNGELL; HU, 2011). Isso se deve, em grande parte, à importância reduzida que é dada às colaborações domésticas quando comparadas às colaborações internacionais, já que é ainda arraigada a ideia de que o produto resultante de colaborações internacionais seja usualmente publicado em prestigiosos periódicos internacionais e receba maior número de citações quando comparado àquele publicado em periódicos de circulação nacional, embora o produto resultante da colaboração doméstica consiga muitas vezes obter sucesso semelhante (GLÄNZEL; SCHUBERT, 2004).⁴³ Ademais, a qualidade e viabilidade de estudos de colaborações domésticas está diretamente atrelada à necessidade de um conjunto complexo de informações, o que torna indispensável o uso complementar de sistemas locais de informações. Assim, a inexistência ou dificuldade de acesso às informações necessárias implicam na utilização exclusiva das grandes bases de dados bibliográficos internacionais, o que contribui para a menor magnitude e visibilidade dos resultados obtidos, especialmente para pequenos sistemas científicos nacionais, devido, principalmente, à orientação internacional e viés de língua inglesa dessas bases.⁴⁴

Quanto à análise espacial das colaborações científicas, o tamanho do país pode desempenhar papel determinante na comparação entre as colaborações domésticas e internacionais. Enquanto as distâncias geográficas entre unidades colaborativas podem ser enormes em países territorialmente extensos, a ultrapassagem das fronteiras nacionais é facilitada em países menores. Portanto, da mesma maneira que a

⁴² A desconsideração da diferenciação sistemática dos padrões de publicação entre as áreas do conhecimento pode provocar grandes equívocos nas análises que visam identificar padrões espaciais de publicações ou coautorias.

⁴³ Contudo, a maior visibilidade no cenário internacional não necessariamente reflete o impacto e relevância do conteúdo de artigos científicos na solução de problemas específicos ou locais.

⁴⁴ Para mais detalhes a respeito dos aspectos relacionados ao uso de bases de dados internacionais, veja seção 4.2.

colaboração internacional é mecanismo essencial em países pequenos, ou com pequenas comunidades científicas (tanto pela facilidade geográfica na ultrapassagem das fronteiras nacionais como pela possível dificuldade em localizar especialistas dentro de seus territórios), o contrário ocorre em países de grandes dimensões territoriais, uma vez que a colaboração doméstica é facilitada em comparação à internacional devido à vantagens de escala referentes à localização de diversos institutos de pesquisa nacionais, o que leva a maior oportunidades de colaboração com pesquisadores locais, e, assim, acaba por estimular a colaboração doméstica (TIJSSEN, 2008; CHINCHILLA-RODRÍGUEZ *et al*, 2010). Diante disso, torna-se fundamental a análise espacial das colaborações científicas ao longo do território brasileiro. Para tal finalidade, faz-se necessária a descrição dos conceitos e técnicas a serem utilizadas posteriormente na análise das redes de coautorias em publicações científicas por parte dos pesquisadores brasileiros.

3. METODOLOGIA

3.1. Análise de redes sociais e espaciais

É crescente o número de estudos sobre redes de conhecimento que fazem uso do ferramental da análise de redes sociais, uma vez que as interações sociais que permeiam a constituição dessas redes são fundamentais para a explicação dos processos de criação, difusão e transferência de conhecimento (PHELPS *et al*, 2012).⁴⁵ Este capítulo possui o objetivo de apresentar sucintamente algumas definições básicas relacionadas à terminologia tipicamente empregada na análise de redes sociais, a fim de facilitar tanto a compreensão acerca da maneira específica de organização e consolidação do banco de dados utilizado como também a caracterização descritiva da estrutura e dinâmica das redes de colaborações científicas que será realizada posteriormente.

O arcabouço analítico do ferramental da chamada análise de redes sociais pode ser encarado como um dos ramos do conhecimento que mais se encaminha rumo à interdisciplinariedade, uma vez que permite uma enorme gama de distintas aplicações.⁴⁶ A análise de redes sociais refere-se à aplicação da teoria dos grafos no estudo de relacionamentos sociais, de modo a lidar com as características mensuráveis das estruturas das redes.⁴⁷ Teoricamente, a análise de redes sociais possui como principal fundamento a ideia de representar as relações entre indivíduos por meio de redes ou grafos, em que uma rede consiste num conjunto de nós (elemento básico de uma rede) conectados por um conjunto de arestas, as quais são ligações entre dois nós. Os nós podem ser especificados como indivíduos, grupos, firmas, organizações, patentes, unidades geográficas, entre outros.⁴⁸ Já as arestas podem possuir direcionamento (cada par de nós possui um nó de origem e outro de destino) e ponderação (cada aresta possui um valor numérico associado) (NEWMAN, 2010).

⁴⁵ Bergman (2009) apresenta um panorama sobre o uso do ferramental de redes sociais em estudos espaciais sobre redes de conhecimento e inovação.

⁴⁶ Para exemplos sobre as diferentes aplicações veja Newman (2010).

⁴⁷ A denominação de grafos é mais comum na literatura especializada em matemática.

⁴⁸ Usualmente a denominação análise de redes sociais refere-se ao caso particular do funcionamento de redes entre indivíduos. Para outros exemplos de nós veja Newman (2010, p.110).

3.1.1. Representação das redes de colaboração científica

Uma das formas típicas de representação de um grafo (rede) baseia-se na denominada matriz de adjacência, as quais consistem em matrizes quadradas de dimensão $(n \times n)$ onde os i nós ($i = 1, \dots, n$) são representados nas linhas e colunas, e cada entrada contém os valores referentes à intensidade das ligações entre os nós (arestas).

Dentre os diversos tipos de redes, as redes de colaborações científicas consistem em redes não-direcionadas (a colaboração entre os pesquisadores A e B não impõe qualquer determinação de direcionamento entre eles) com arestas ponderadas (cada ligação possui um valor numérico associado que determina a magnitude das interações).⁴⁹ Desse modo, no caso particular das redes de colaboração científica, cada célula (i, j) da matriz de adjacência possui o valor da intensidade de colaborações entre os pesquisadores i e j .

No âmbito da análise cientométrica espacial, é usual a associação entre os pesquisadores e unidades geográficas, de maneira a permitir a elaboração de uma nova rede de conexões entre as próprias unidades geográficas, as quais passam a funcionar como novos nós. Nesse caso, é criada uma nova matriz de adjacência composta pelos valores inter-regionais totais de colaborações entre pesquisadores das unidades geográficas i e j , a qual é uma matriz simétrica (devido ao não-direcionamento das ligações) e contém em sua diagonal principal o total de colaborações intrarregionais, *i.e.*, entre pesquisadores localizados internamente à mesma unidade geográfica (Figura 1).

	Região i	Região j	Região k
Região i	a	b	c
Região j	b	d	e
Região k	c	e	f

Figura 1 – Matriz de adjacência em redes de colaboração científica inter-regional

⁴⁹ O não-direcionamento das redes de colaborações implica na simetria da matriz de adjacência, uma vez que o total de colaborações entre os pesquisadores i e j é o mesmo que entre os pesquisadores j e i .

3.1.2. Métricas

A maneira mais simples de caracterização das redes em trabalhos empíricos dá-se por meio da análise descritiva dos valores das métricas, as quais consistem na mensuração de propriedades topológicas dos grafos. Contudo, a incorporação do espaço geográfico na configuração das redes, denominadas de redes espaciais, provoca efeitos determinantes em suas estruturas topológicas, visto que as restrições do espaço podem afetar suas propriedades (BARTHÉLEMY, 2011). Assim, fazem-se necessárias algumas considerações a respeito das métricas utilizadas.

De maneira sucinta, as métricas globais permitem avaliar as características estruturais da configuração de uma rede como um todo, enquanto que as métricas individuais procuram avaliar e comparar o papel desempenhado por seus nós.

Dentre as principais métricas globais utilizadas, o tamanho da rede consiste no número total de nós em uma rede e fornece uma medida de sua abrangência em termos da quantidade de diferentes nós. Já o número total de arestas permite a avaliação do tamanho da rede em termos da quantidade de ligações entre os nós envolvidos. Já o grau médio, métrica comumente utilizada nas análises de redes espaciais, retrata a média aritmética entre os graus dos nós individuais, onde o grau de um nó é definido pelo número de arestas adjacentes, *i.e.*, de nós conectados a ele. Assim, figura como um indicador da articulação média entre os participantes da rede. A densidade de uma rede consiste na razão entre o número de ligações existentes na rede e o total de ligações possíveis entre os nós.⁵⁰ O diâmetro possui o propósito de verificar a maior distância entre dois nós internos à rede, com a particularidade de considerar também o peso das arestas. Já a transitividade (ou coeficiente de clusterização) mensura a probabilidade de nós próximos estarem conectados, de maneira a representar uma medida de formação de *clusters* na rede (JACKSON, 2008; NEWMAN, 2010).

⁵⁰ A densidade apresenta informação semelhante ao grau médio, uma vez que pode ser expressa como a razão entre esse e o número total de nós da rede menos um.

3.2. Modelos gravitacionais de interações espaciais

Com o intuito de identificar os determinantes geográficos dos padrões dos fluxos de conhecimento no Brasil, verificados a partir das redes de colaboração científica entre os pesquisadores brasileiros, faz-se necessária a modelagem da estrutura espacial da colaboração científica. Para tal feito, optamos por utilizar a abordagem dos modelos gravitacionais de interações espaciais, descritos a seguir.

A interação espacial é o processo pelo qual atores em diferentes pontos do espaço físico realizam contatos, enquanto que a denominação gravitacional refere-se à modelagem análoga à Lei da Gravitação Universal de Isaac Newton de 1687, a qual estabeleceu que a força entre dois corpos é proporcional às massas e função decrescente da distância entre eles, de acordo com uma lei quadrado-inversa (MAGGIONI *et al*, 2007).

A utilização da família dos modelos gravitacionais e de suas extensões abrange diversas áreas do conhecimento, sendo fundamentais em análises estatísticas que buscam identificar as fontes de disparidades regionais em diversos fenômenos oriundos de interações humanas, uma vez que permitem a verificação da hipótese de que a distância geográfica seja a principal responsável por determinado fenômeno ou se existem outros fatores que explicam os padrões de interação observados. Dessa maneira, os modelos gravitacionais figuram como ferramenta poderosa de análise empírica para explicar as interações sociais de acordo com a existência de forças de atração e de afastamento, assim, esses modelos são frequentemente utilizados para a modelagem de dados econômicos que representam diferentes tipos de fluxos de origem-destino através do espaço geográfico, tais como a modelagem de fluxos comerciais (comércio internacional), monetários (investimento estrangeiro direto e turismo), migratórios e até de casamentos (HOEKMAN *et al*, 2010; MAGGIONI *et al*, 2007).

Na literatura econômica, os modelos gravitacionais são empregados, principalmente, no âmbito dos estudos sobre comércio internacional. Nesse caso particular, a ideia fundamental é a de que o fluxo de comércio bilateral entre dois países deve ser proporcional às suas massas econômicas (Produto Interno Bruto [PIB] ou população) e

inversamente relacionado à distância geográfica entre eles.⁵¹ Além do notável sucesso da maior parte da literatura que os utiliza no estudo da relação entre a distância geográfica e o fluxo bilateral de comércio de bens, esses modelos também passaram a ser aplicados em estudos empíricos que investigam outros tipos de relações socioeconômicas, tais como os fluxos de conhecimento, por meio da citação de patentes, e a internacionalização das colaborações em P&D, por meio dos FPs (SCHERNGELL; BARBER, 2011).

3.2.1. Determinantes da colaboração científica

No contexto específico de modelagem dos fluxos de colaboração científica, os modelos gravitacionais foram utilizados para explicar a intensidade das colaborações científicas entre regiões na Holanda (PONDS *et al*, 2007), China (LIANG; ZHU, 2002; WANG *et al*, 2005; SCHERNGELL; HU, 2011) e na Europa (HOEKMAN *et al*, 2009).

Inicialmente, o modelo teórico pode ser representado por:

$$Y_{ij} = X_{ij} + \varepsilon_{ij} \quad ; \quad i, j = 1, \dots, n \quad (3.1)$$

Onde Y_{ij} é a variável dependente estocástica correspondente a matriz de fluxos de colaboração observados entre duas regiões i e j (onde $y_{ij} \geq 0$ representa cada elemento não-negativo da matriz), X_{ij} é uma função que captura a relação estocástica entre outras variáveis aleatórias amostrais a partir de uma distribuição de probabilidade específica, e ε_{ij} é o termo de erro que varia entre todos os pares inter-regionais (i, j) . É assumido que o termo de erro é independente e identicamente distribuído (i.i.d.), portanto, é válida a seguinte propriedade: $E[\varepsilon_{ij} | y_{ij}] = 0$ (SCHERNGELL; HU, 2011).

⁵¹ Nesse contexto, a análise empírica desde os anos 1960 apresentou sinais estatisticamente positivos (e próximos à unidade) dos parâmetros referentes ao PIB de países exportadores e importadores, enquanto o sinal do parâmetro da distância geográfica mostrou-se negativo e significativo, o que corrobora as hipóteses do modelo (MAGGIONI *et al*, 2007).

O modelo descrito consiste numa estrutura muito abrangente. Para a análise empírica é necessária especificação mais precisa para as variáveis explicativas X_{ij} . O fato de lidarmos com observações entre unidades espaciais discretas torna a perspectiva do modelo de interação espacial do tipo gravitacional alternativa assaz apropriada, uma vez que esse modelo é caracterizado pela distinção formal entre três tipos de funções capazes de explicar a variação das interações inter-regionais: duas funções que caracterizam as regiões de origem e destino, e outra função responsável por mensurar o grau de separação espacial entre as regiões. Assim, como nosso interesse reside na mensuração da importância relativa das características das regiões de origem e destino, e das medidas de distância sobre a determinação dos fluxos de colaborações, optamos pela especificação de X_{ij} como um modelo gravitacional geral (SCHERNGELL; HU, 2011; HOEKMAN *et al*, 2010; 2012; SCHERNGELL; BARBER, 2009; 2011; PONDS *et al*, 2007):

$$X_{ij} = O_i \cdot D_j \cdot S(u_{ij}) \quad ; \quad i, j = 1, \dots, n \quad (3.2)$$

Onde O_i é uma função que caracteriza a região i de interação, D_j é uma função que caracteriza a região j de interação e $S(u_{ij})$ é uma função que caracteriza a separação espacial entre as regiões i e j de interação. Desse modo, a variável dependente X_{ij} passa a ser constituída por um conjunto de elementos específicos a cada região e por medidas de separação espacial entre as duas regiões de origem e destino.

As funções O_i e D_j , as quais caracterizam as regiões de origem e destino da interação, representam as massas de origem e destino no modelo gravitacional e podem ser especificadas por meio de “funções de poder”, de acordo com a teoria clássica da interação espacial (SEN; SMITH, 1995), o que faz com que passem a assumir a forma das funções abaixo:

$$O_i = O(o_i, \alpha_1) = o_i^{\alpha_1} \quad ; \quad i = 1, \dots, n \quad (3.3)$$

$$D_j = D(d_j, \alpha_2) = d_j^{\alpha_2} \quad ; \quad j = 1, \dots, n \quad (3.4)$$

Onde o_i e d_j são variáveis que mensuram características específicas das regiões i e j , e os termos α_1 e α_2 são os parâmetros a serem estimados nas duas especificações. Em nosso caso, não existe determinação da direção dos fluxos entre as regiões, na medida em que as interações são fruto de colaborações, as quais não possuem qualquer espécie de direcionamento. Dessa maneira, nenhuma distinção possível entre as variáveis de origem e destino é passível de ser aplicada nesse contexto, *i.e.*, as variáveis são simétricas. Logo, ocorrerá o caso especial de igualdade entre os parâmetros das duas regiões i e j ($\alpha_1 = \alpha_2$). Ademais, é esperado que as estimativas desses parâmetros sejam significantes e próximas à unidade. Assim, o produto entre O_i e D_j pode ser simplesmente interpretado como o número total de colaborações distintas esperadas entre as duas regiões i e j (SCHERNGELL; BARBER, 2011).

Embora as características intrínsecas às regiões de interação assumam importância decisiva, o foco principal da análise consiste no melhor entendimento do papel das dissimilaridades entre as regiões na determinação dos fluxos de colaboração científica. Assim, a distinção e a caracterização das dimensões de proximidade fazem com que o interesse volte-se, primordialmente, à mensuração da importância relativa entre as medidas de distância sobre os fluxos de colaboração. Dessa maneira, merece atenção especial o modo que o termo de separação espacial $S(u_{ij})$ é especificado.

A especificação tradicional do termo de separação espacial dá-se por meio da forma funcional multivariada exponencial, como reproduzido a seguir:

$$S(u_{ij}) = S(u_{ij}, \beta) = \exp\left[\sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}\right] ; \quad i, j = 1, \dots, n \quad (3.5)$$

Nessa representação, as variáveis $u_{ij}^{(k)}$ representam k medidas de separação espacial entre as regiões i e j , e o termo β_k refere-se ao conjunto de k parâmetros desconhecidos associados a cada uma das k medidas de separação espacial entre i e j .⁵² Por conseguinte, a estimação desses parâmetros permitirá a medição da força relativa entre

⁵² A descrição das medidas de distância utilizadas, definidas a partir das diferentes dimensões de proximidade, será realizada na seção 5.2.

as medidas de separação, de modo a determinar a importância relativa entre as dimensões de proximidade.

É interessante notar que tanto os parâmetros α_1 e α_2 quanto os parâmetros β_k podem ser interpretados como elasticidades de y_{ij} com respeito à variável de origem o_i , à variável de destino d_j e às variáveis de separação $u_{ij}^{(k)}$.

A incorporação da função de separação espacial é fundamental, uma vez que permite que a modelagem da escolha dos parceiros de colaboração não seja baseada somente no número de possíveis colaborações por si só (oriundo da interação das características intrínsecas às duas regiões), mas também dependa do grau de separação espacial entre as regiões, refletido pelas medidas de distância entre elas.

A partir da introdução das especificações das funções das características da origem (3.3), de destino (3.4) e da função de separação entre as regiões (3.5) na equação do modelo inicial (3.2), obtemos o modelo empírico a ser estimado:

$$Y_{ij} = \alpha_0 + o_i^{\alpha_1} \cdot d_j^{\alpha_2} \cdot \exp\left[\sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}\right] + \varepsilon_{ij} \quad ; \quad i, j = 1, \dots, n \quad (3.6)$$

Na expressão acima, além dos parâmetros α_1 , α_2 e β_k apresentados anteriormente, é introduzido um novo parâmetro (α_0), termo constante que visa captar a linearidade da regressão.

3.2.2. Especificação e estimação

A adoção do procedimento correto de estimação do modelo descrito na equação (3.6) dá-se posteriormente à decisão de qual deve ser a melhor especificação estatística do modelo. Entretanto, tal processo de decisão não é trivial, sendo a escolha baseada na realização de uma sequência de testes de especificação, descritos a seguir.

Inicialmente, a análise preliminar quanto à natureza dos dados já permite considerações importantes, uma vez que os valores numéricos resultantes da contabilização das coautorias possuem característica bastante peculiar: são valores inteiros e não-negativos, *i.e.*, pertencem ao conjunto dos números naturais.⁵³

Tal característica faz com que não seja apropriada a aplicação de uma especificação log-normal sobre a equação do modelo (3.6) e a consequente estimação dos parâmetros pelo tradicional método de mínimos quadrados ordinários (MQO). Primeiramente, muitas das hipóteses estatísticas necessárias para o sucesso na estimação não se sustentariam nesse contexto, o que poderia levar a estimativas ineficientes, viesadas, ou até mesmo inconsistentes (LONG; FREESE, 2001). Além da inadequação do método de MQO, o procedimento também traria outros problemas, uma vez que a extração dos logaritmos da equação do modelo faria com que as estimativas dos logaritmos da variável dependente fossem subestimadas para os maiores valores, e a verificação de frequências inter-regionais de colaborações nulas teria que ser artificialmente contornada, já que a especificação log-normal não é capaz de lidar com zeros.⁵⁴ Entretanto, o maior problema do procedimento citado é o fato de que o processo gerador dos dados analisados está muito distante da suposição de que a contagem das coautorias seja gerada a partir de uma distribuição log-normal em torno de seu valor médio e com uma variância constante (HOEKMAN *et al*, 2010).

Todavia, as deficiências da especificação log-normal e os fortes pressupostos necessários à aplicação de MQO podem ser contornados por meio da especificação das coautorias como contagens discretas. Portanto, a primeira decisão em busca da melhor especificação dá-se por meio da interpretação mais apropriada do modelo como de dados de contagem.

A interpretação do modelo como de contagem assume que o processo de geração dos dados produza somente valores inteiros não-negativos. A partir disso, é usual admitir que a quantidade de colaborações segue uma distribuição de Poisson. Tradicionalmente,

⁵³ O conjunto dos números naturais é entendido aqui como o conjunto dos números inteiros não-negativos, o que admite o valor zero como pertencente a esse conjunto.

⁵⁴ A soma de uma unidade aos valores nulos é o método mais usual de contorno a esse problema.

essa suposição é considerada uma boa alternativa para a descrição de valores inteiros não-negativos. De modo geral, a utilização de modelos pertencentes à família de modelos de Poisson resolve as deficiências técnicas de MQO, reconhece explicitamente a natureza inteira e não-negativa da variável dependente, além de permitir que as estimativas de máxima-verossimilhança dos parâmetros possam ser interpretadas como elasticidades (WINKELMANN, 2008; SCHERNGELL; HU, 2011).

A densidade de Poisson é dada pela seguinte expressão:

$$\Pr(Y_{ij} = y_{ij} | X_{ij}) = \frac{e^{-\mu_{ij}} \cdot \mu_{ij}^{y_{ij}}}{y_{ij}!}; \quad i, j = 1, \dots, n; \quad y_{ij} = 0, 1, 2, \dots \quad (3.7)$$

onde μ_{ij} representa o conjunto de variáveis dependentes do modelo empírico (3.6):

$$\mu_{ij} = X_{ij} = \exp[\alpha_0 + \alpha_1 \cdot \log(o_i) + \alpha_2 \cdot \log(d_j) + \sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}] \quad (3.8)$$

A função de distribuição de Poisson possui a propriedade estatística de equidispersão, definida pela igualdade entre a média e variância condicionais (WINKELMANN, 2008, p.8; HILBE, 2011, p.77). Assim, segue-se diretamente que:

$$\mu_{ij} = V[Y_{ij} | X_{ij}] = E[Y_{ij} | X_{ij}] \quad (3.9)$$

Para nossos propósitos, a igualdade acima consiste na característica crucial da distribuição de Poisson. Caso essa hipótese não seja rejeitada pelos dados observados, é possível assumir que eles sejam gerados a partir de um processo de Poisson, e o modelo especificado pode ser estimado consistentemente pelo método de máxima verossimilhança padrão (CAMERON; TRIVEDI, 1998; WINKELMANN, 2008).⁵⁵

Contudo, é possível que a contagem de colaborações científicas entre pares regionais desvie-se de um padrão de Poisson para a geração dos dados, já que é comum que a

⁵⁵ A não-linearidade do modelo faz com que os valores das estimativas sejam encontrados através do método de iterações de Newton-Raphson (HILBE, 2011).

distribuição desses valores não satisfaça à propriedade de equidispersão.⁵⁶ Usualmente, as análises empíricas apontam para a presença de superdispersão nos dados observados, situação em que a variância condicional é superior à média condicional (WINKELMANN, 2008, p.8; HILBE, 2011, p.140).⁵⁷ Isso normalmente decorre do problema de heterogeneidade não-observada, uma vez que as variáveis independentes especificadas não são capazes de captar toda a heterogeneidade nos dados por meio da função da média condicional, *i.e.*, na predição das ocorrências de colaborações inter-regionais (CAMERON; TRIVEDI, 1998, p.96). Tal situação usualmente advém do fato de que a quantidade observada de termos nulos na variável dependente (ausência de colaborações) é superior àquela quantidade assumida pela distribuição de Poisson. Logo, a presença de heterogeneidade não-observada, quando não capturada pelas variáveis do modelo, implica em superdispersão, o leva a estimativas viesadas dos parâmetros e invalidade dos testes de hipóteses usuais, uma vez que os erros-padrão são subestimados (HILBE, 2011; WINKELMANN, 2008; CAMERON; TRIVEDI, 1998).⁵⁸

A possibilidade de detecção do problema de heterogeneidade não-observada nos dados analisados sugere a procura por um modelo que se adapte melhor à estrutura dos dados observados, o que nos leva à escolha do modelo binomial negativo, que consiste na principal alternativa ao modelo de Poisson, por também pertencer à mesma família de modelos, alternativa típica nos trabalhos empíricos sobre colaborações científicas (HOEKMAN *et al*, 2010; SCHERNGELL; BARBER, 2011; SCHERNGELL; HU, 2011; SCHERNGELL; LATA, 2011). A escolha por essa classe de modelo dá-se por sua capacidade de lidar com a heterogeneidade não-observada, uma vez que é especificado um parâmetro adicional (parâmetro de heterogeneidade) a fim de possibilitar que a variância condicional exceda o valor da média esperada, o que permite a acomodação da superdispersão observada nos dados. A expressão da densidade da distribuição binomial negativa é dada pela seguinte expressão:

⁵⁶ Diferentemente do que ocorre com outras distribuições paramétricas, a violação da hipótese de equidispersão é suficiente para a violação da hipótese de um processo gerador de dados de Poisson (WINKELMANN, 2008, p.8).

⁵⁷ A superdispersão também pode ser entendida como a verificação de excessiva correlação nos dados analisados (HILBE, 2011 p.64).

⁵⁸ A invalidade do pressuposto da equidispersão possui consequências qualitativas semelhantes à invalidade do pressuposto da homoscedasticidade no modelo clássico de regressão linear, com o agravante de que a magnitude dos efeitos sobre os erros-padrão e estatísticas *t* reportados podem ser ainda maiores (CAMERON; TRIVEDI, 1998, p.77).

$$\Pr(Y_{ij} = y_{ij} | X_{ij}) = \frac{\Gamma(y_{ij} + \alpha^{-1})}{y_{ij}! \Gamma(\alpha^{-1})} \cdot \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}}\right)^{\alpha^{-1}} \cdot \left(\frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}}\right)^{y_{ij}} \quad (3.10)$$

Onde $\Gamma(\cdot)$ é a função gama e α é o parâmetro de heterogeneidade. Assim, a variância condicional do modelo passa a ser expressa por:⁵⁹

$$V[Y_{ij} | X_{ij}] = \mu_{ij}(1 + \alpha \cdot \mu_{ij}) \quad ; \quad \mu_{ij} = E[Y_{ij} | X_{ij}] \quad (3.11)$$

Observa-se que é permitida uma forma mais geral de heterogeneidade, alternativamente à igualdade entre média e variâncias condicionais impostas pela distribuição de Poisson (LONG; FREESE, 2001; WINKELMANN, 2008). Dessa maneira, o principal elemento de decisão de especificação entre os dois modelos consiste na verificação da presença de heterogeneidade não-observada. Contudo, o modelo de Poisson consiste numa especificação particular do modelo binomial negativo, referente à situação em que o parâmetro de heterogeneidade seja igual a zero (LONG; FREESE, 2001), o que permite a elaboração de um teste de hipótese tradicional em que a hipótese nula de equidispersão é testada contra a alternativa, caso em que ocorre superdispersão (CAMERON; TRIVEDI, 1998, pp.77-8).⁶⁰ Logo, a verificação da significância estatística da estimativa do parâmetro de heterogeneidade α , estimado a partir do modelo binomial negativo, permite a decisão entre os dois modelos, de modo que sua significância aponta para a validação do modelo binomial negativo.

Outro problema de especificação refere-se à quantidade excessiva de zeros nos dados observados, o que pode figurar como fonte adicional de heterogeneidade não-observada, na medida em que a ocorrência de valores nulos pode ser bastante superior àquela passível de acomodação pelos modelos de Poisson e binomial negativo. Contudo, tal problema pode ser contornado por meio da utilização das versões infladas de zeros do modelo de Poisson (ZIP) e binomial negativo (ZINB)⁶¹, os quais admitem uma estrutura

⁵⁹ A expressão apresentada refere-se à especificação *Negbin II* (WINKELMANN, 2008, p.134), a qual é a mais utilizada na pesquisa empírica (HILBE, 2011, p.346).

⁶⁰ Dentre os testes possíveis, destaca-se o Teste de Wald, o qual utiliza a estatística t reportada para a estimativa α no modelo binomial negativo (CAMERON; TRIVEDI, 1998).

⁶¹ As siglas ZIP (*Zero-Inflated Poisson Model*) e ZINB (*Zero-Inflated Negative Binomial Model*) referem-se às versões infladas de zeros dos modelos de Poisson e binomial negativo.

de média condicional que diferencia os valores nulos e positivos (HILBE, 2011, p.370).⁶² Dessa maneira, é usual a estimação e comparação entre as especificações apresentadas a fim de decidir quanto à melhor especificação estatística do modelo gravitacional de interações espaciais para a aplicação na análise espacial das redes de colaboração científica.

⁶² Basicamente, a modelagem dos valores nulos dá-se por meio de um processo binário (*probit* ou *logit*), enquanto os valores positivos são modelados por um processo de contagem (Poisson ou binomial negativo) (HILBE, 2011, p.370).

4. BANCO DE DADOS

A discussão teórica do arcabouço da análise formal é seguida pela descrição e discussão detalhada a respeito dos dados utilizados. Os objetivos primordiais desse capítulo são explicar a fonte das informações, a definição da unidade espacial de análise, o procedimento de coleta, contabilização e periodicidade dos dados utilizados, bem como discutir suas principais características e limitações.

4.1. Matrizes intermunicipais de colaboração científica

Conforme apresentado, tanto as técnicas de análise das redes espaciais como a estimação do modelo empírico (equação 3.6) necessitam de dados que constituam a matriz Y_{ij} , a qual é composta pelos valores correspondentes aos fluxos observados de colaborações científicas entre duas regiões i e j ($y_{ij} \geq 0$), sendo que esses são usualmente mensurados pela contagem de coautorias em publicações científicas por parte dos pesquisadores de diferentes regiões.

4.1.1. Currículo Lattes

Em nosso caso, os dados de coautorias foram extraídos a partir das informações sobre publicações científicas existentes no sistema de Currículos Lattes (CL), parte integrante da Plataforma *Lattes* (PL) do CNPq, a qual consiste num verdadeiro sistema nacional de informações de C&T&I, implantado e mantido pelo governo brasileiro para o gerenciamento de informações relacionadas aos pesquisadores, instituições e atividades de pesquisa no país.⁶³

⁶³ Basicamente, a consolidação da PL representa a integração das bases de dados dos CL, Grupos de Pesquisa, e Instituições, num único sistema de informações (CNPQ, 2012).

O sistema de CLs figura como o núcleo da PL e consiste num abrangente sistema de informações curriculares, constituído por um enorme acervo de informações individuais de pesquisadores, professores, estudantes e profissionais de todas as áreas do conhecimento no Brasil. Logo, possui importância crucial para o balizamento de atividades nos processos de planejamento, gestão e operacionalização das agências de fomento federais e estaduais, das fundações estaduais de apoio à C&T&I e das instituições de ensino superior e institutos de pesquisa, uma vez que fornece informações confiáveis para a análise de mérito e competência dos pesquisadores, avaliação de programas de pós-graduação e análise de pleitos de financiamentos. Ademais, o sistema possui caráter indispensável como mecanismo auxiliar na formulação e avaliação de políticas públicas, principalmente por parte do Ministério da Ciência e Tecnologia e Inovação (MCTI), na medida em que permite a extração de diagnósticos precisos e sistêmicos sobre o desenvolvimento das atividades de C&T&I no país (CNPQ, 2012).

A extensa dimensão da cobertura informacional deve-se ao fato do sistema ter se tornado padrão nacional no registro das atividades acadêmicas e profissionais da comunidade científica brasileira. Tal abrangência foi alcançada por meio de contínuos esforços por parte do CNPq, aliado à facilidade de que o registro e publicação do histórico acadêmico tornou-se hábito entre os pesquisadores brasileiros. Internamente à comunidade científica, a declaração e atualização das informações curriculares é motivada pela necessidade quanto à divulgação das atividades desempenhadas, seja para fins de pleito de recursos junto às fontes de fomento científico (sob a forma de bolsas ou auxílios para projetos de pesquisa), como para o reconhecimento perante aos pares da própria comunidade científica. Ademais, a disponibilização pública das informações curriculares e de grupos de pesquisa via *web* e a utilização das informações pelas universidades em meio às decisões sobre titulações e promoções dos professores estimulam a inserção correta e a veracidade dos dados publicados.⁶⁴ Logo, o estabelecimento de um mecanismo de incentivos para os pesquisadores preencherem e

⁶⁴ Embora o processo de atualização dos currículos ainda não seja realizado de forma contínua pelos usuários, a disponibilização pública das informações possibilita o monitoramento da qualidade das informações divulgadas, uma vez que a decisão quanto às alocações de recursos para pesquisas depende crucialmente da comparação direta entre os currículos dos pesquisadores pleiteadores.

atualizarem corretamente suas informações proporcionou grande visibilidade, credibilidade e reconhecimento internacional à base de dados do sistema de CLs, destacada como um dos melhores bancos de dados existentes no mundo sobre pesquisadores, modelo de sucesso a ser seguido internacionalmente (LANE, 2010).

A versão atual do CL consiste num formulário eletrônico a ser preenchido individualmente pelo usuário. Sua estrutura permite o cadastro das atuações e produções desenvolvidas numa ampla gama de atividades relacionadas à C&T&I. Contudo, a autonomia dada ao usuário no registro manual de suas próprias informações pode acarretar em erros de preenchimento de diversas espécies, os quais usualmente geram inconsistências na análise e validação de informações. Diante disso, a forma de cadastro e de preenchimento dos campos dos currículos tem sido modificada, num constante processo de aperfeiçoamento por parte do CNPq, com o intuito de minimizar a necessidade da digitação manual de nomes por parte do usuário, de modo a reduzir, ou até mesmo impossibilitar, a ocorrência de erros.⁶⁵

A disposição dos campos de preenchimento do CL dá-se por meio de módulos.⁶⁶ No tocante às publicações científicas, o usuário deve preencher o título completo da publicação de sua autoria (identificação dos autores, título completo, periódico, local e ano). Dessa maneira, torna-se possível identificar a ocorrência de coautoria por meio da comparação direta entre os títulos registrados pelos usuários. A definição da estratégia adotada e a descrição do procedimento computacional de identificação e contabilização das coautorias são realizadas a seguir.

⁶⁵ O CNPq mantém-se ativo na busca por novos mecanismos de controle que promovam o direcionamento do preenchimento das informações. Como exemplo, a versão atual do CL (disponibilizada em 2012) oferece ao usuário uma lista de nomes de pesquisadores possivelmente relacionados durante a elaboração da citação bibliográfica.

⁶⁶ O processo de aperfeiçoamento faz com que a apresentação das interfaces do CL sofra alterações ao longo do tempo. A versão atual do CL contém doze módulos principais, os quais são subdivididos em diversos campos passíveis de preenchimento.

4.1.2. Procedimento de elaboração das matrizes

4.1.2.1. *ScriptLattes*

Os CLs são disponibilizados publicamente no portal da PL na *web* (CNPQ, 2012). Embora o acesso às informações individuais seja imediato, o acesso sistemático à completude da base de dados do sistema de CLs não é possível, o que torna o esforço na coleta de informações o maior empecilho para a análise de grande volume de dados. Enquanto a compilação manual de grande volume de currículos é repetitiva e sujeita a falhas, a automatização do processo torna-o extremamente rápido e menos susceptível às últimas.⁶⁷ O desafio computacional resultou no desenvolvimento do *ScriptLattes*, uma ferramenta computacional que, a partir de uma lista de códigos identificadores de CLs e da definição do tipo de informação desejada, é capaz de extrair as informações dos CLs e gerar relatórios com os resultados, além de insumos computacionais para aplicações posteriores, principalmente para análises de redes sociais (cálculo de métricas e visualização das redes de coautoria).⁶⁸

Para nossos propósitos, foi utilizada uma modificação no *ScriptLattes* capaz de gerar as redes de colaboração científica, por meio de um algoritmo que visa identificar as ligações de coautorias de maneira mais eficiente. Dessa maneira, tornou-se factível o processamento de grandes quantidades de CLs, de modo que fossem analisadas informações de 1.131.912 CLs, o que significa que é atingida grande representatividade do banco de dados do sistema.⁶⁹

⁶⁷ A automatização da análise de grandes grupos de indivíduos requer esforço computacional especializado, uma vez que é necessária a extração de vasta quantidade das informações de interesse a partir dos currículos individuais, processo conhecido como mineração de dados.

⁶⁸ O *ScriptLattes* é um *software* livre em constante aperfeiçoamento que tem sido adaptado e utilizado por diversas agências de fomento, institutos de pesquisa e universidades para auxiliar os processos de avaliações institucionais (disponibilizado em <http://scriptlattes.sourceforge.net/>).

⁶⁹ Os CLs foram rastreados por meio de seus códigos de identificação no dia 05 de maio de 2011. O procedimento completo de mineração dos dados é descrito em Digiampietri *et al* (2011).

4.1.2.2. Identificação das coautorias

A modificação realizada sobre o *ScriptLattes* permite estabelecer ligações de coautoria entre pesquisadores caso exista uma produção comum entre eles divulgadas em seus CLs. O procedimento consiste na busca por semelhanças entre os títulos das produções acadêmicas cadastradas no conjunto dos CLs dos pesquisadores selecionados. No âmbito da análise das redes sociais, cada pesquisador é representado por um nó e a detecção de uma relação de coautoria é representada por uma ligação entre os nós. Após a identificação da coautoria, a contabilização das ligações pode ser feita de maneira simples, ponderada, normalizada ou por frequências, como veremos adiante.

Mais especificamente, a informação sobre as coautorias foi extraída a partir das informações contidas no módulo “Produção Bibliográfica” do CL.⁷⁰ Dentre os campos passíveis de preenchimento desse módulo, a análise da identificação e contabilização das coautorias foi realizada sobre as informações contidas em quatro campos específicos: “artigos completos publicados em periódicos”; “trabalhos completos publicados em anais de congressos”; “livros publicados, organizados ou edições”; e “capítulos de livros publicados”.

A Figura 2 apresenta o total anual de publicações científicas que foram analisadas nos processos de identificação e contabilização das coautorias, em que o agrupamento trienal ao longo do período 1990-2010 foi possível a partir da informação do ano de publicação verificado nos títulos completos das produções científicas. Observa-se que há uma tendência de crescimento acelerado das publicações ao longo do período.⁷¹ Para o período completo (1990-2010), o total de publicações únicas analisadas foi de 7.351.957.

⁷⁰ Na versão atual do CL, a produção acadêmica é subdividida entre *Produção Bibliográfica*, *Técnica e Artística/ Cultural*. A classificação completa dos campos contidos no módulo *Produção Bibliográfica* do CL encontra-se no Apêndice A.

⁷¹ A notável ruptura da tendência de crescimento no ano de 2010 está fortemente associada ao fato de que os dados foram coletados em maio de 2011. Assim, é provável que muitos pesquisadores ainda não tivessem atualizado seus CLs de modo a incluir as publicações do ano anterior (2010).

Figura 2 – Evolução do total anual de publicações científicas únicas analisadas

4.1.2.3. Identificação dos municípios

Por constituir-se o tema central em cientometria espacial, o procedimento de busca por informações de localização geográfica é descrito a seguir de maneira mais detalhada, de modo a explicitar os principais problemas metodológicos evidenciados pela literatura.

Usualmente, as informações dos endereços dos autores são obtidas nas próprias publicações científicas ou por meio da consulta do endereço dos autores em bases de dados complementares. Contudo, a utilização dessa espécie de dados para a construção de bases agregadas ao nível da unidade espacial selecionada enfrenta problemas metodológicos substanciais. Inicialmente, o pesquisador já se defronta com uma questão delicada e crucial: a determinação da unidade espacial relevante para a análise. Embora a maior parte dos estudos opte por agregar endereços a partir das áreas administrativas oficialmente definidas (países, estados, províncias, microrregiões, municípios), as redes científicas constituem-se sistemas complexos de entrelaçamento e rompimento das fronteiras formais, o que torna difícil o convencimento de que as fronteiras percebidas pelos pesquisadores na decisão sobre a possibilidade de colaboração coincidam com os

limites administrativos impostos por qualquer espécie de divisão territorial oficial. Todavia, a possibilidade de organização dos dados por meio de outro tipo de classificação é cerceada pela demanda de esforços extras significativos e pela sujeição a críticas diversas, decorrentes do grau de arbitrariedade inerente a qualquer classificação proposta. Logo, a agregação das unidades amostrais entre alguma das unidades espaciais tradicionais acaba por ser a escolha usual (FRENKEN *et al*, 2009). Dentre as unidades possíveis, a utilização de municípios como nós das redes de colaboração científica é alternativa recentemente empregada (RS, 2011; PAN *et al*, 2012), assim, optamos pela escolha dos municípios brasileiros como unidade de análise, uma vez que a utilização de estados ou regiões no caso brasileiro empobreceria bastante a análise espacial e da importância do papel da distância geográfica, principalmente pela necessidade de ampla agregação das observações e pela dificuldade de cálculo maneira satisfatório da distância entre estados ou regiões.⁷² Entretanto, ao invés de localizar os coautores a partir de seus endereços nas publicações (ou por meio da associação entre nomes e endereços com bases complementares) fez-se uso direto da informação sobre os endereços profissionais dos pesquisadores declarados nos CLs.⁷³

Inicialmente, foi projetado e desenvolvido um programa para a extração das informações sobre a identificação da localização municipal nos CLs. No entanto, a autonomia dada ao usuário no preenchimento do campo “endereço” gerou dificuldades, uma vez que se revelou significativa a quantidade de erros nos dados selecionados, provenientes de erros de digitação, utilização de abreviaturas regionais e alocação errônea de municípios aos seus respectivos estados.⁷⁴ Diante disso, tornou-se necessária a padronização dos nomes dos municípios, a qual se baseou na sistematização da forma de digitação do nome dos municípios e na identificação correta dos estados associados a partir da elaboração de uma lista de ocorrência que sumarizou todas as entradas de

⁷² Frenken *et al* (2009) afirma que os estudos sobre padrões de colaboração científica entre municípios eram bastante raros na época. Porém, trabalhos recentes já abarcam tal perspectiva, como é o caso de Pan *et al* (2012).

⁷³ Os problemas decorrentes dessa escolha serão discutidos posteriormente na seção 4.2.

⁷⁴ Como exemplos, mesmo após a transformação preliminar dos nomes em letras minúsculas e sem nenhuma espécie de acentuação, foram encontradas 28 versões diferentes associadas ao nome do município do Rio de Janeiro/RJ, as versões “Floripa” para Florianópolis/SC e “BH” e “Belô” para Belo Horizonte/MG, além de associações errôneas entre a capital nacional Brasília/DF e seis estados distintos.

nomes distintos.⁷⁵ A associação de uma única entrada com o nome correto do município e estado correspondente permitiu a consolidação de uma lista de ocorrências correta, a qual estabeleceu ligação unilateral precisa entre o código de cada CL e o município de localização. Desse modo, foi determinada a unidade espacial de análise, uma vez que a unidade de análise passou de indivíduos para o nível municipal, com a atribuição de centenas de códigos de CLs a municípios únicos. Essa associação permitiu que o *ScriptLattes*, usualmente aplicado a indivíduos, fosse implementado para a identificação e quantificação de coautorias entre os municípios brasileiros.

Infelizmente, grande parte do total de currículos analisados (cerca de 43,22%) não apresentou nenhuma informação a respeito do município de localização no campo “endereço”, o que inviabilizou a identificação da localização geográfica desses pesquisadores. Contudo, acreditamos que a relevância do problema deve ser avaliada por meio da quantificação da perda de informação em termos do número de coautorias sem localização municipal. Para tal fim, optou-se pela atribuição da entrada “indeterminado” aos CLs sem nenhuma identificação municipal e essa nova entrada foi acrescentada à lista de ocorrências consolidada, de modo a mensurar a perda de informação resultante da ausência da identificação dos municípios nos CLs efetivamente em termos do total de produção e colaboração científica.⁷⁶ A lista de ocorrências dos municípios brasileiros declarados nos CLs analisados totalizou 4.616 entradas, o que significa que o total de CLs foi atribuído a 4.615 municípios diferentes.

⁷⁵ A correção dos problemas foi realizada por meio da utilização da lista oficial dos nomes dos municípios brasileiros do IBGE. Após a transformação preliminar dos nomes, a lista de ocorrências identificava 5.321 entradas de diferentes nomes de municípios. Após o processo de padronização, a nova lista contava com 4.615 entradas. Erros menos frequentes, como a descrição do nome de bairros de grandes cidades e de *campi* universitários, também foram identificados e corretamente contornados.

⁷⁶ Como veremos, essa estratégia permitiu que o processo de identificação e contagem de coautorias identificasse a quantidade de coautorias entre municípios indeterminados e os demais municípios como também a quantidade de coautorias entre municípios indeterminados.

4.1.2.4. Identificação das áreas de conhecimento

De modo a permitir a diferenciação dos padrões de colaborações científicas entre as diferentes áreas do conhecimento, as coautorias entre pesquisadores foram identificadas e contabilizadas a partir da discriminação da área do conhecimento desses.

No preenchimento do CL, cabe ao pesquisador informar a grande área do conhecimento que abarca o conteúdo de suas atividades científicas.⁷⁷ Entretanto, é permitido que o usuário escolha mais de uma grande área do conhecimento, o que é condizente com a tendência do conhecimento científico rumo à interdisciplinaridade. Na versão do CL analisada, é permitida a escolha entre as seguintes grandes áreas: *Ciências Agrárias* (AGR), *Ciências Biológicas* (BIO), *Engenharias* (ENG), *Ciências Exatas e da Terra* (EXT), *Ciências Humanas* (HUM), *Linguística, Letras e Artes* (LLA), *Ciências da Saúde* (SAU) e *Ciências Sociais Aplicadas* (SOC).⁷⁸

A extração das informações acerca da área de atuação de cada pesquisador permitiu a associação entre os pesquisadores e a área de conhecimento por eles declarada, da maneira apresentada na Tabela 1, em que se observa que 76,7% dos pesquisadores identificados foram associados a uma única área do conhecimento. Logo, a associação entre pesquisador, município e área do conhecimento permitiu a elaboração de matrizes de coautorias intermunicipais diferenciadas por área do conhecimento.⁷⁹

⁷⁷ Na classificação adotada pelo CNPq, o conhecimento é distribuído entre oito grandes áreas bem definidas e uma grande área denominada “Outros”, a qual abarca as áreas não pertencentes a nenhuma das anteriores. A identificação das grandes áreas e de suas respectivas áreas está descrita no Apêndice B.

⁷⁸ Também foi considerada a grande área “Outros” (OUT), composta por áreas do conhecimento não classificadas dentre as oito grandes áreas citadas. A fim de simplificar a argumentação, optamos por ocultar a denominação “grandes” na referência às grandes áreas do conhecimento.

⁷⁹ Os pesquisadores foram associados a todas as áreas do conhecimento declaradas em seus CLs. Assim, optou-se pela contagem superestimada daqueles pesquisadores que possuem mais de uma área declarada, uma vez que suas publicações foram analisadas na identificação e contabilização de coautorias nas diferentes áreas. Logo, a somatória das matrizes de todas as áreas particulares (calculadas separadamente) possui valores superiores àqueles observados na matriz total em que não há discriminação das áreas do conhecimento.

Tabela 1 – Número de pesquisadores associados a cada grande área do conhecimento

	<i>Total de Pesquisadores</i>	<i>Pesquisadores associados a uma única área</i>
<i>AGR</i>	92.927	59.484 (64,0%)
<i>BIO</i>	128.104	65.593 (51,2%)
<i>ENG</i>	120.993	70.075 (57,9%)
<i>EXT</i>	176.114	102.372 (58,1%)
<i>HUM</i>	270.149	141.376 (54,2%)
<i>LLA</i>	99.129	53.151 (53,6%)
<i>SAU</i>	272.783	206.772 (75,8%)
<i>SOC</i>	264.230	164.037 (62,1%)
Total	1.131.912	868.250 (76,7%)

4.1.2.5. Método de contabilização das coautorias

Entre os métodos tradicionais de contagem de coautorias, a contagem completa e fracionada são os métodos mais utilizados. O método de contagem completa (*full-counting*), usualmente escolhido em trabalhos empíricos de análise de coautorias, baseia-se na ideia de que cada unidade de análise (autores ou regiões) deve receber uma unidade de crédito pela participação em determinada publicação científica (SCHERNGELL; BARBER, 2011).⁸⁰ Como alternativa, o método de contagem fracionada realiza contagem ponderada, de forma que os valores das ligações entre duas regiões são divididos pela quantidade de ligações entre as regiões dos pesquisadores envolvidos, *i.e.*, o crédito da coautoria é dividido proporcionalmente entre os coautores. Mais comum em análises de citações, esse método subestima o impacto de artigos oriundos da colaboração entre pesquisadores de diversas regiões, enquanto que o método de contagem completa superestima tal contagem. Logo, para nossos propósitos, optamos pela utilização do método de contagem completa das coautorias nas publicações científicas.

⁸⁰ Como exemplo, a partir da observação de uma publicação científica fruto da colaboração entre os pesquisadores I, II e III, localizados nas regiões A, A e B, respectivamente, são contabilizados os valores 1 para o par de regiões (A, A) e 2 para o par (A, B). Por simetria, o par (B, A) também recebe valor 2.

4.2. Principais características dos dados

Conforme descrito anteriormente, cada entrada da matriz de colaborações intermunicipais possui o valor da intensidade de coautorias entre os pesquisadores dos municípios i e j .⁸¹ Assim, a partir da identificação e contabilização das coautorias, cada célula (i, j) da matriz possui a informação da quantidade de coautorias entre os pesquisadores dos municípios i e j .

A declaração da informação sobre o ano de publicação (contida no título completo das publicações) permitiu a elaboração de matrizes anuais de coautorias. Logo, foram construídas matrizes de coautorias para 10 grupos (para cada uma das nove áreas do conhecimento e outra para todas as áreas) e para todos os anos do período 1990-2010.⁸² A identificação de 4.615 municípios determinou, inicialmente, que a dimensão das matrizes de colaboração, n , fosse igual a 4.615.⁸³ Portanto, cada matriz é composta por 10.651.420 valores.⁸⁴ Embora tenham sido identificados 4.615 municípios com CLs associados, após a contabilização das coautorias observamos que os CLs 3.268 municípios não apresentaram nenhuma observação de publicações em coautoria. Logo, as matrizes intermunicipais analisadas possuíam 1.347 municípios.

Cabe destacar as principais características dos dados utilizados que possibilitaram a superação satisfatória de problemas usualmente apontados pela literatura especializada.

Primeiramente, destaca-se a abrangência da amostra utilizada, tanto em relação ao número de pesquisadores e publicações analisadas quanto ao amplo período de coleta de

⁸¹ Como destacamos anteriormente, as matrizes descritas estão diretamente associadas às matrizes de adjacência da teoria dos grafos, as quais se baseiam na representação dos nós nas linhas e colunas, e cada entrada demonstra a intensidade das ligações entre eles. Para n regiões, a matriz possui dimensão $(n \times n)$, onde as i regiões ($i = 1, \dots, n$) são representadas nas linhas e as j regiões ($j = 1, \dots, n$) nas colunas.

⁸² Para os anos mais recentes, principalmente no caso de 2010, é possível que a contabilização realizada esteja subestimada, visto que a contabilização correta depende da frequência da atualização dos CLs.

⁸³ A princípio, para a quantificação da perda de informação resultante da não-localização de pesquisadores, a matriz construída possuía dimensão 4.616, por também incluir a entrada “indeterminado”, que foi excluída para as análises posteriores.

⁸⁴ Para cada matriz, a simetria faz com que o total de entradas distintas seja dado por: $n + (n - 1) + (n - 2) + \dots + 2 + 1 = [n \cdot (n + 1)]/2$, assim, a matriz com $n = 4615$, possui 10.651.420 entradas distintas.

dados, uma vez que o volume de dados utilizados é bastante superior àquele usualmente analisado.⁸⁵

Em segundo lugar, a cientometria espacial defronta-se com o problema de que as informações coletadas sobre os endereços normalmente referem-se aos institutos de pesquisa e não aos próprios autores, o que pode fazer com que publicações com múltiplos endereços possam estar associadas a autores com múltiplas afiliações ou que conduzam pesquisas em determinado instituto e eventualmente se moveram para outros institutos (FRENKEN *et al*, 2009). Por exemplo, pode acontecer de pesquisadores em visita temporária escolherem registrar seu próprio instituto e pesquisadores registrarem as organizações de subvenção (financiamento) ao invés do instituto onde a pesquisa é realmente realizada. Em relação aos institutos de pesquisas e empresas, as localidades de suas sedes podem ser registrados em vez das unidades filiais onde as pesquisas foram efetivamente realizadas.⁸⁶ Contudo, esses problemas são, em certa medida, contornados em nossa análise, uma vez que o procedimento de localização geográfica foi realizado diretamente por meio das informações dos endereços declaradas pelos autores.

A fonte de dados utilizada merece destaque especial, na medida em que a grande maioria dos estudos cientométricos faz uso de bases de dados internacionais, o que produz duas grandes limitações.⁸⁷ Primeiramente, essas bases apresentam certo viés, uma vez que cobrem principalmente os periódicos de língua inglesa, assim, a qualidade dos dados parte do pressuposto de que os periódicos não indexados sejam locais, os quais publicam suas pesquisas na língua nacional oficial. Em segundo lugar, é bastante baixa a cobertura da produção das áreas de ciências sociais e humanidades

⁸⁵ Scherngell e Hu (2011) interpretam como uma amostra aleatória a observação de 758.682 fluxos de colaborações na China para o ano de 2007, contabilizados a partir de metade das publicações de pesquisadores chineses naquele ano. Hoekman *et al* (2010) analisaram dados de colaboração entre regiões da Europa para o período entre 2000 e 2007, em que a amostra com maior fluxo de colaborações observadas (524.155) deu-se em 2007. Vanz (2009) admite a pequena fração da produção brasileira analisada por meio de bases de dados internacionais.

⁸⁶ Embora a afiliação múltipla de um indivíduo possa ser interpretada como o estabelecimento de colaborações entre elas, tal situação difere-se substancialmente do envolvimento colaborativo entre pesquisadores de diferentes organizações.

⁸⁷ Dentre as principais, a base de dados *Web of Science* (WoS), produzida pela *Thomson Reuters*, indexa aproximadamente 10.000 periódicos “revisados pelos pares” e é considerada uma das mais abrangentes e confiáveis fontes de informação sobre atividades de pesquisa básica entre países e áreas do conhecimento (ADAMS; KING, 2009).

(HOEKMAN *et al*, 2010). Isso ocorre, principalmente, devido às características intrínsecas a essas áreas, nas quais a produção científica dá-se predominantemente sob a forma de capítulos de livros e a tradução para o inglês é muitas vezes inviabilizada pela dificuldade de tradução exata de termos e expressões contidos no material, o que faz com que essas áreas possuam viés para a publicação na língua local.

Logo, acreditamos que a utilização de bases de dados internacionais não permite a avaliação completa da atividade científica brasileira, uma vez que nos países em desenvolvimento o novo conhecimento é predominantemente publicado e divulgado por periódicos locais, muitos dos quais não são incorporados às bases internacionais por não possuírem circulação internacional (UNESCO, 2010). Além disso, apesar da língua inglesa figurar como a língua franca da pesquisa científica, a universalização linguística da ciência ainda encontra obstáculos, o que faz com que a língua predominante nos periódicos brasileiros seja a língua portuguesa, principalmente no caso de revistas das áreas de ciências humanas e sociais aplicadas. Assim, a avaliação efetiva dos padrões de atividade científica na América Latina não é tarefa trivial, uma vez que parte significativa da produção regional dá-se em espanhol e português.⁸⁸

Diante dessas peculiaridades, destaca-se a coleta de informações a partir de dados do sistema de CLs, o que permite que seja extensa a cobertura da publicação científica brasileira, uma vez que são contabilizadas as coautorias nos artigos publicados em periódicos de circulação nacional. Logo, contorna-se a exclusão usual desses últimos na contabilização a partir de bases de dados internacionais. Ademais, a inclusão da produção bibliográfica sob a forma de livros e capítulos permite a melhor avaliação da produção e colaboração científica nas áreas de ciências sociais e humanidades.

Entre as limitações do procedimento utilizado, a principal refere-se à contabilização somente de coautorias domésticas entre pesquisadores brasileiros, *i.e.*, não é realizada a

⁸⁸ Existem cerca de 13.000 periódicos em espanhol e 5.000 em português em circulação na América Latina e situação semelhante ocorre na China, fonte do maior número de publicações em línguas diferentes da inglesa, o que pode obscurecer os resultados dos levantamentos usuais sobre a colaboração científica nesses países (RS, 2011). Com o intuito de aumentar a visibilidade da produção científica brasileira, a FAPESP e o *Latin American and Caribbean Center on Health Sciences Information* criaram o portal de acesso livre SciELO em 1999 (UNESCO, 2010).

identificação e contabilização das colaborações internacionais. Como o procedimento computacional efetiva a coautoria quando compara e identifica títulos semelhantes em CLs diferentes, não é possível identificar aquelas colaborações entre pesquisadores brasileiros e estrangeiros, devido à inexistência de CLs desses últimos. Portanto, uma publicação produzida com o auxílio de pesquisadores estrangeiros só é contabilizada em nossa análise caso ocorra simultaneamente alguma colaboração entre pesquisadores brasileiros, o que efetiva a existência de colaboração doméstica.

No entanto, embora a natureza dos dados imponha certas limitações para a análise, o próprio procedimento de extração de informações do sistema de CLs para a elaboração das matrizes intermunicipais de coautoria baseia-se na validade de algumas hipóteses. Primeiramente, a utilização de informações de endereços para a localização geográfica é baseada na suposição de que o registro dos endereços dos pesquisadores corresponda à localização onde a pesquisa foi verdadeiramente conduzida. Esse consiste num problema metodológico tradicional em ciências espaciais, porém, não há razões para se acreditar que essa suposição não seja razoável (FRENKEN *et al*, 2009). Em segundo lugar, é pressuposto que a informação da localização geográfica dos pesquisadores observada seja a verdadeira localização do pesquisador durante todos os anos analisados, *i.e.*, a contabilização da coautoria entre os autores de determinada publicação referente ao ano de 1990 é atribuída aos municípios com base na informação da localização geográfica dos pesquisadores extraída em 2011. Assim, não se leva em consideração a possível migração de pesquisadores ao longo do território nacional. Embora a migração seja problema importante e incontornável em nossa análise, consideramos os dados utilizados apropriados e bastante úteis para os objetivos do estudo, uma vez que muitos dos problemas apontados são, em certa medida, contornados ou reduzidos pelo amplo tamanho da amostra utilizada.

5. RESULTADOS

Conforme apresentado, os fluxos de conhecimento (sejam entre indivíduos, organizações ou entidades geográficas) tornaram-se importantes objetos de estudo em geografia econômica, em que as descrições dos padrões geográficos da produção de conhecimento, bem como das articulações entre os pesquisadores de diversas localidades, consistem no principal objetivo, uma vez que a capacidade de acesso por parte das firmas às fontes de conhecimento é fator fundamental para a realização de atividades inovativas (MAGGIONI; UBERTI, 2011a; TER WAL, 2011). Dessa maneira, o melhor entendimento sobre a maneira que os pesquisadores colaboram nos processos de criação e difusão do conhecimento pode potencializar o efeito de políticas de fomento à transmissão do conhecimento da academia para o setor privado, na medida em que esse processo tende a ser localizado geograficamente, primordialmente para domínios científicos e tecnológicos caracterizados pela existência de conhecimentos tácitos e complexos, os quais não podem ser completamente codificados sob a forma de artigos científicos (FELDMAN; AVNIMELECH, 2011).⁸⁹

Com o intuito de compreender o processo de produção e difusão do conhecimento científico no caso brasileiro, nosso foco estará primeiramente centrado na análise da evolução espacial da produção científica no país, de maneira a verificar quais regiões são as maiores produtoras e como a evolução da produção científica deu-se no espaço. Em seguida, serão descritas as principais características estruturais e topológicas das redes de colaboração científica entre os municípios com o intuito de analisar descritivamente a evolução da articulação entre os pesquisadores na configuração das redes. Por fim, a dimensão do papel desempenhado pela distância geográfica nas interações colaborativas entre pesquisadores brasileiros nas redes será avaliada por meio da estimação de modelos gravitacionais.

⁸⁹ Esforços recentes concentram-se no desenvolvimento de novas técnicas de visualização de mapas geográficos de redes científicas entre cidades e institutos de pesquisa, de modo a facilitar a identificação e localização dos fluxos de conhecimento ao longo do espaço (LEYDESDORFF; PERSSON, 2010).

5.1. Análise espacial da produção e colaboração científica no Brasil

5.1.1. Quantificação da perda de informação

Inicialmente, é necessário quantificar mais precisamente a perda de informação anunciada anteriormente. Como vimos, não foi possível a identificação da localização geográfica de grande parte do total de CLs analisados. Todavia, acreditamos que tal perda de informação não deve ser diretamente quantificada a partir do número de CLs sem localização geográfica, uma vez que para os nossos propósitos é necessário saber qual é a importância relativa desses indivíduos em atividades científicas.⁹⁰ Dessa maneira, a relevância do problema deve ser dada pela quantificação da perda de informação em termos do número de publicações e de coautorias sem localização municipal.

A Tabela 2 mostra a quantidade relativa de publicações e coautorias contabilizadas, mas que deixaram de ser incluídas na análise devido à impossibilidade de associação da localização geográfica. Observa-se que os valores oscilam entre 7% e 13,5%, sendo que a maior perda de informação dá-se a partir do quarto triênio selecionado, o qual consiste no período de aceleração do crescimento da publicação científica brasileira. Logo, embora não seja desprezível, acreditamos que a perda de informação não acarreta grandes prejuízos no tocante à análise estrutural da evolução da produção e colaboração científica entre os pesquisadores brasileiros.

Tabela 2 – Evolução do percentual de publicações e colaborações sem localização municipal

	<i>1992-1994</i>	<i>1995-1997</i>	<i>1998-2000</i>	<i>2001-2003</i>	<i>2004-2006</i>	<i>2007-2009</i>
<i>Produção</i>	7,1%	7,0%	7,0%	7,9%	9,2%	10,6%
<i>Colaboração</i>	8,1%	8,0%	8,8%	10,2%	11,9%	13,5%

Fonte: Elaboração própria.

⁹⁰ No Brasil, praticamente todas as pessoas envolvidas com pesquisa, desde pesquisadores ativos até estudantes de graduação e pós-graduação possuem CL. Enquanto os primeiros são os maiores responsáveis pela publicação científica, existem muitos casos de CL sem qualquer informação sobre produção bibliográfica, situação típica de estudantes de graduação que ainda estão iniciando seus estudos.

5.1.2. Características estruturais da produção científica municipal no Brasil

A partir do procedimento de coleta de dados descrito anteriormente, foi possível a obtenção de informações anuais sobre a produção científica dos municípios brasileiros ao longo do período 1990-2010, o que possibilita algumas considerações importantes a respeito dos padrões de localização geográfica da atividade científica no país. Logo, o objetivo desta seção é apresentar uma análise descritiva da maneira pela qual o produto da atividade científica brasileira dispõe-se ao longo do território nacional. Mais especificamente, apresentaremos a evolução da produção científica dos municípios brasileiros, de modo a identificar os locais de maior produção de conhecimento, e, principalmente, verificar a existência de evidências de desconcentração espacial da atividade científica no país.

Inicialmente, faz-se necessária a descrição precisa do significado dos dados a serem analisados, uma vez que a determinação de padrões espaciais da atividade científica possui como característica crucial a maneira que se dá a alocação do produto da atividade científica entre as unidades espaciais. Isso acontece porque tal processo de alocação é permeado por dificuldades, já que na maioria das vezes a associação entre uma publicação e uma localidade específica não é passível de ser determinada de maneira direta e única. Logo, qualquer tentativa de mensuração do produto da atividade científica em cada localidade depende das decisões específicas tomadas pelo pesquisador.

O principal problema na determinação da produção municipal ocorre devido à existência de coautorias entre pesquisadores de municípios distintos, na medida em que não é possível associar a publicação a um único município. Para lidar com esse problema, foi adotada a estratégia de mensurar a participação dos pesquisadores em publicações, ao invés de mensurar o total de publicações efetivas. Como exemplo, para uma publicação em coautoria entre dois pesquisadores, um de São Paulo/SP e outro de Recife/PE, foi contabilizada uma unidade de participação para cada município envolvido, de maneira que o total apurado (duas participações) superestima o total da publicação efetiva (uma única publicação). Portanto, os valores analisados não

correspondem ao total de publicações efetivas, mas sim ao total de participações absolutas em publicações por parte de pesquisadores associados a cada município.⁹¹ Embora a contabilização fracionada⁹² pareça ser a escolha mais intuitiva à primeira vista (por manter o total de publicações efetivas), ambos os procedimentos de contabilização fracionada e completa são encontrados na literatura, e há consenso de que a ordem da produção das regiões independe do método de contabilização escolhido e de que os esforços regionais nas publicações colaborativas são subestimados quando se utiliza o método de contagem fracionada (OSBORNE; HOLLAND, 2009; GROSSETTI *et al*, 2012; MICHELS *et al*, 2013). Dessa maneira, como o foco do trabalho é a análise espacial da produção científica brasileira (e não sua quantificação), optamos por considerar a participação absoluta em publicações como indicador da atividade científica municipal, com o devido cuidado na interpretação desses valores, na medida em que seu crescimento pode estar associado tanto ao aumento da produção efetiva como ao aumento do perfil colaborativo.

A Figura 3 apresenta a evolução do total da produção científica dos municípios brasileiros (mensurada por meio da participação em publicações científicas dos pesquisadores de cada município). A Figura 4 apresenta a contribuição relativa de cada área do conhecimento para o total. Já a Figura 5 mostra a taxa de crescimento da publicação entre os triênios selecionados para o total e para cada área do conhecimento.

Inicialmente, verifica-se uma tendência de crescimento do total da produção municipal ao longo dos triênios apresentados (Figura 3). A contribuição relativa de cada área do conhecimento na produção total não apresentou grandes transformações ao longo do período (Figura 4), uma vez que, embora os valores para cada área sejam diferentes, as taxas de crescimento aproximam-se da tendência geral de desaceleração do crescimento do total da publicação (Figura 5).

⁹¹ Esse método aproxima-se ao método de contabilização utilizado na contagem das coautorias, conforme discutido na seção 4.1.2.5.

⁹² Nesse método, a contabilização dá-se pela contribuição relativa de cada autor em determinada publicação, *i.e.*, no caso de dois coautores, é atribuído o valor $\frac{1}{2}$ a cada autor.

Contudo, conforme afirmado, a interpretação desses resultados deve ser bastante cautelosa, por não ser possível a determinação imediata das causas do crescimento observado, já que tal comportamento pode se originar tanto da aceleração da produção científica em termos de publicações efetivas, bem como do aumento do perfil colaborativo entre os pesquisadores no período. Apesar de estudos cientométricos apontarem o crescimento acelerado da produção científica no país⁹³, é possível que tenha ocorrido um aumento do perfil colaborativo da ciência brasileira no período.

Figura 3 – Evolução do total da publicação científica municipal no Brasil⁹⁴

⁹³ Os resultados e referências são discutidos na seção 2.1.1..

⁹⁴ Conforme argumentado, os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, devido à existência de publicações em coautoria.

Figura 4 – Evolução da participação relativa das áreas do conhecimento no total da publicação científica municipal no Brasil

Figura 5 – Taxa de crescimento da publicação científica municipal no Brasil entre triênios selecionados: por área do conhecimento

Embora os totais apresentados sejam importantes, a análise individual da evolução da produção municipal permite uma melhor avaliação da maneira que a produção é organizada espacialmente ao longo do território nacional. Os dez principais municípios produtores de conhecimento no Brasil em cada triênio, para o total das áreas e por área do conhecimento são apresentados na Tabela 3 e nas Tabelas D1 a D9 (Apêndice D), respectivamente.

Os valores apresentados evidenciam a enorme heterogeneidade espacial das atividades científicas no país, altamente concentradas na região Sudeste, semelhantemente ao encontrado em FAPESP (2011a, 2011b). Para o total das áreas, observa-se na Tabela 3 que somente dois municípios não pertencentes à região Sudeste (Recife/PE e Brasília/DF) figuram entre os principais produtores de conhecimento no país.⁹⁵ De um modo geral, ocorre pouca variação entre os primeiros municípios da lista, em que predomina a presença de municípios bastante populosos e sedes de universidades públicas (estaduais e federais), tais como São Paulo/SP (USP, Unifesp), Rio de Janeiro/RJ (UFRJ, UERJ), Porto Alegre/RS (UFRGS), Belo Horizonte/MG (UFMG), Campinas/SP (Unicamp), Curitiba/PR (UFPR), Recife/PE (UFPE), Florianópolis/SC (UFSC), Brasília/DF (UnB), Ribeirão Preto/SP (USP), São Carlos/SP (USP, UFSCar), entre outros. Assim, a pouca variação entre os principais produtores é esperada, uma vez que os centros de pesquisa instalados nesses municípios são historicamente consolidados e apresentam intenso nível de atividade e produção de destaque no cenário científico nacional e internacional. Padrão semelhante também é verificado para as áreas do conhecimento isoladamente (Tabelas D1 a D9).⁹⁶

⁹⁵ Para o triênio 2007-2009, somente 12 municípios não pertencentes à região Sudeste estão entre os 50 maiores produtores de conhecimento no país.

⁹⁶ Além das universidades públicas, diversos institutos pesquisa tais como: Embrapa (agricultura), Fiocruz (saúde), INPA (biodiversidade), Instituto Butantan (biologia e biomedicina), Instituto Adolfo Lutz (saúde pública), ITA (engenharia aeronáutica), INPE (áreas espacial e do ambiente terrestre), entre outros, também possuem papel de destaque no cenário científico nacional e internacional.

Tabela 3 – Evolução dos principais municípios associados à produção científica no Brasil: todas as áreas do conhecimento

Total		Total		Total	
1992-1994		1995-1997		1998-2000	
São Paulo/SP	71.013	São Paulo/SP	112.539	São Paulo/SP	161.991
Rio de Janeiro/RJ	37.100	Rio de Janeiro/RJ	64.109	Rio de Janeiro/RJ	102.309
Campinas/SP	20.045	Porto Alegre/RS	32.944	Porto Alegre/RS	52.446
Porto Alegre/RS	18.228	Campinas/SP	31.283	Campinas/SP	47.949
Belo Horizonte/MG	14.420	Belo Horizonte/MG	25.432	Belo Horizonte/MG	44.633
Ribeirão Preto/SP	10.388	São Carlos/SP	18.501	São Carlos/SP	28.430
São Carlos/SP	9.034	Florianópolis/SC	15.077	Brasília/DF	25.925
Brasília/DF	8.540	Brasília/DF	15.054	Florianópolis/SC	25.713
Recife/PE	7.583	Ribeirão Preto/SP	14.925	Recife/PE	24.500
Florianópolis/SC	7.334	Recife/PE	14.035	Curitiba/PR	24.253
2001-2003		2004-2006		2007-2009	
São Paulo/SP	226.688	São Paulo/SP	313.910	São Paulo/SP	352.541
Rio de Janeiro/RJ	146.139	Rio de Janeiro/RJ	193.348	Rio de Janeiro/RJ	215.550
Porto Alegre/RS	82.101	Porto Alegre/RS	115.614	Porto Alegre/RS	132.622
Belo Horizonte/MG	64.294	Campinas/SP	90.575	Belo Horizonte/MG	113.487
Campinas/SP	64.194	Belo Horizonte/MG	89.293	Campinas/SP	95.089
Brasília/DF	43.711	Curitiba/PR	61.462	Curitiba/PR	75.125
Curitiba/PR	41.102	Brasília/DF	61.003	Recife/PE	72.119
São Carlos/SP	40.628	Recife/PE	56.519	Florianópolis/SC	70.322
Recife/PE	38.781	São Carlos/SP	56.372	Brasília/DF	65.963
Florianópolis/SC	37.763	Florianópolis/SC	54.094	Ribeirão Preto/SP	65.252

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

A heterogeneidade espacial no caso brasileiro faz com que a desconcentração espacial das atividades de pesquisa científica seja um dos objetivos primordiais das políticas de C&T&I no país, uma vez que pode potencializar o desenvolvimento regional das áreas menos favorecidas.⁹⁷ Como exemplo recente, o programa de expansão das universidades federais foi responsável pela criação de 14 novas universidades e mais de 100 novos *campi* universitários espalhados pelo país, principalmente nas regiões economicamente menos favorecidas (Ministério da Educação, 2013).⁹⁸ Nesse contexto, a quantificação do crescimento da produção científica municipal abre espaço para questões acerca de como esse processo de crescimento ocorreu geograficamente. Em outras palavras, é fundamental a verificação da possível existência de um concomitante processo de desconcentração espacial da produção científica no país.

A Figura 6 apresenta os mapas da produção científica municipal nos triênios de 1992-1994 e 2007-2009. Observa-se que a produção de conhecimento está melhor distribuída espacialmente no triênio 2007-2009, visto que é elevado o número de municípios, outrora com produção praticamente insignificante (11-100), que passaram a apresentar produção em níveis intermediários (101-10.000) nesse período. Já a Figura 7 apresenta o mapa com a introdução da estrutura dos *campi* das universidades públicas (federais e estaduais) no país em 2009. Observa-se que é nítida a associação entre a localização dos *campi* universitários e a produção municipal de conhecimento.⁹⁹ Os mapas C1 a C9 (Apêndice C) apresentam as distribuições espaciais da produção em cada área do conhecimento. Embora em intensidades distintas, observa-se que a melhor distribuição da produção é fenômeno sistemático para todas as áreas do conhecimento.

⁹⁷ Como já destacado, um exemplo importante é a realização de pesquisas direcionadas ao enfrentamento de problemas locais.

⁹⁸ Informações acerca do programa de expansão das universidades federais podem ser encontradas em <http://reuni.mec.gov.br/>.

⁹⁹ Tal resultado está de acordo com o apresentado por Leta *et al* (2006), em que se verificou que entre 1991 e 2003 mais de 80% da produção científica brasileira indexada na base WoS era proveniente de universidades públicas.

Figura 6 – Mapas da produção científica municipal no Brasil em triênios selecionados

Figura 7 – Mapas da produção científica municipal no Brasil no triênio 2007-2009 e localização dos *campi* de universidades públicas (federais e estaduais)

A Tabela 4 apresenta a estrutura da distribuição dos municípios por diferentes faixas de produção, em que se verifica o crescimento sistemático do número de municípios em cada faixa. Todavia, a partir da aparente melhoria da distribuição da produção não é possível verificar a existência de um processo sistemático de desconcentração espacial entre os períodos considerados, uma vez que também houve crescimento da produção dos municípios maiores produtores, o qual poderia ter sido suficiente para acarretar a concentração espacial da produção. Assim, faz-se necessária a comparação temporal entre as porcentagens acumuladas da produção científica municipal e a quantidade de municípios envolvidos, com o intuito de verificar se o processo de crescimento verificado deu-se de maneira menos concentrada no espaço.

Tabela 4 – Número de municípios por faixa de produção científica

<i>Total de publicações</i>	<i>1992-1994</i>	<i>1995-1997</i>	<i>1998-2000</i>	<i>2001-2003</i>	<i>2004-2006</i>	<i>2007-2009</i>
0	4.795	4.561	4.286	3.870	3.527	3.374
1 a 10	413	507	620	821	1.000	1.097
11 a 100	225	310	380	475	545	552
101 a 1.000	93	123	193	269	330	358
1.001 a 10.000	35	54	65	99	127	137
10.001 a 100.000	6	11	20	31	35	45
100.001 a 1.000.000	0	1	2	2	3	4

A Figura 8 apresenta as curvas de localização da produção científica para os 200 municípios que mais produziram conhecimento em cada triênio selecionado.¹⁰⁰ Observa-se que no triênio 1992-1994, 90% da produção do país estava concentrada em 48 municípios, ao passo que essa proporção estava distribuída em 102 municípios no triênio 2007-2009. Logo, é possível concluir que há evidências de que o crescimento da

¹⁰⁰ As curvas de localização são elaboradas a partir da colocação, no eixo vertical, das porcentagens acumuladas da variável-base (emprego, PIB, etc.) numa determinada atividade (indústria, setor, etc.) por região, e, no eixo horizontal, das porcentagens acumuladas do total das atividades em todas as regiões. Assim, se os dois conjuntos de porcentagens forem idênticos, a curva coincidirá com a diagonal de 45 graus a partir da origem; caso ocorram divergências, a curva de localização irá se deslocar para cima e para a esquerda, em que a magnitude do deslocamento corresponde à concentração espacial da atividade (HADDAD, 1989, pp.237-8).

produção científica no Brasil deu-se de maneira espacialmente desconcentrada no período analisado.¹⁰¹ Portanto, verifica-se que o padrão de desconcentração no caso brasileiro parece seguir aquele observado em outros países, tais como Rússia, França, Espanha e China, onde o desenvolvimento de seus sistemas de produção científica nacionais parece seguir uma tendência de desconcentração espacial, principalmente baseada no crescimento da produção nas cidades secundárias, *i.e.*, naquelas de produção intermediária (GROSSETTI *et al*, 2012).

Embora seja evidente a existência de um processo de desconcentração da produção científica total, é fundamental a compreensão da maneira que a produção científica de cada área do conhecimento é concentrada espacialmente, além da verificação de possíveis diferenças entre os processos de desconcentração da produção científica de cada área particular. A Figura 9 exibe as curvas de localização da produção científica em cada área do conhecimento nos triênios de 1992-1994 e 2007-2009, em que é possível notar a diferença na concentração espacial da produção em cada área, bem como a evolução do processo de desconcentração espacial. Inicialmente, vemos que a produção científica em Ciências Agrárias apresenta a menor concentração espacial, seguida por Ciências Biológicas e Ciências Humanas, enquanto a área de Ciências da Saúde apresenta a maior concentração espacial. Na comparação entre os diferentes períodos, verifica-se que o processo de desconcentração da produção científica é generalizado, ou seja, comum a todas as áreas do conhecimento.¹⁰²

¹⁰¹ É importante destacarmos que o processo observado de desconcentração espacial da produção científica está subestimado devido à contabilização da produção municipal por meio da participação em publicações científicas, já que tal método favorece os municípios de maior produção, pois esses estão mais envolvidos nas redes de colaboração científica (GROSSETTI *et al*, 2012).

¹⁰² Mais especificamente, o processo de desconcentração espacial da produção científica deu-se de maneira acelerada até o triênio 2004-2006, uma vez que no triênio seguinte (2007-2009) houve forte desaceleração desse processo, e até mesmo sua reversão nos casos da produção das áreas de Ciências Agrárias e Ciências Biológicas.

Figura 8 – Curvas de localização municipal da produção científica de todas as áreas do conhecimento em triênios selecionados

Figura 9 – Curvas de localização municipal da produção científica por área do conhecimento: triênios 1992-1994 e 2007-2009

5.1.3. Características estruturais das redes espaciais de colaboração científica

Em meio à descrição e discussão sobre o padrão espacial da produção científica no Brasil, vimos que o crescimento da produção está estreitamente associado a um processo de desconcentração espacial. Contudo, é difícil identificar os determinantes desse fenômeno, uma vez que tal processo possui como fontes tanto o crescimento efetivo do total de publicações científicas (as quais podem estar melhor distribuídas geograficamente) como o crescimento da colaboração científica entre pesquisadores de diferentes regiões. Diante disso, o objetivo desta seção é explorar a evolução das redes espaciais de colaboração científica no país, de modo a destacar os principais fluxos de conhecimento ao longo do espaço geográfico e analisar o papel da expansão das redes como determinante do processo de desconcentração espacial da atividade científica no país.

Conforme apresentado, o estudo de fluxos de conhecimento usualmente faz uso do ferramental da análise de redes sociais.¹⁰³ Nesse sentido, a união entre o arcabouço teórico de redes sociais e a consideração do espaço geográfico é essencial, e ocorre a partir da agregação das informações iniciais às unidades espaciais.¹⁰⁴ Conforme descrito anteriormente, os pesquisadores brasileiros analisados foram associados a seus municípios, de modo a permitir a construção de redes intermunicipais de colaboração científica.

Embora a análise de redes consista num vasto campo de estudo, optamos por descrições no âmbito de abordagem da estática comparativa, a qual considera que as estruturas das

¹⁰³ Maggioni e Uberti (2011a) apresentam uma revisão da literatura econômica recente que utiliza redes sociais na análise de fluxos de conhecimento, enquanto que Ter Wal e Boschma (2009) destacam a perspectiva de enorme potencial desse ferramental analítico nessa área particular.

¹⁰⁴ Mais especificamente, a ligação entre redes e geografia pode ser realizada por meio de duas maneiras. Primeiramente, é possível acoplar o ferramental de redes na geografia e interpretar as relações entre os agentes como influenciadas, ou condicionadas, pelas características de seus vizinhos. Essa estratégia faz uso das técnicas de econometria espacial para a verificação da presença e dimensão da autocorrelação entre as variáveis no modelo. Outra estratégia é a possibilidade da inserção direta da geografia nas redes, por meio da construção de redes em que os nós são unidades geográficas. Nesse contexto, a importância da distância geográfica torna-se determinante, e sua estimação é realizada por meio de modelos gravitacionais (MAGGIONI; UBERTI, 2011a).

redes desenvolvem-se no tempo discreto, tornando possíveis análises estatísticas e comparações entre as propriedades estruturais da mesma rede ao longo do tempo.¹⁰⁵

Inicialmente, a Tabela 5 apresenta estatísticas descritivas das matrizes trienais de colaboração científica intermunicipal trienais para o total das áreas do conhecimento. A análise dos valores entre os triênios evidencia algumas tendências gerais da evolução da colaboração doméstica no Brasil.

Primeiramente, observa-se uma tendência acentuada de crescimento da colaboração científica no país, tanto em termos do total de colaborações intermunicipais e intramunicipais, como de seus valores médios. O total de colaborações científicas saltou de 547.249 no triênio 1992-1994 para 9.445.399 no triênio 2007-2009, visto que as colaborações intramunicipais passaram de 317.810 para 1.037.274, enquanto que as intermunicipais passaram de 229.439 para 8.408.125. Ademais, destaca-se o crescimento acelerado das colaborações nos triênios de 2001-2003 e 2004-2007, o que sustenta a hipótese inicial de que o crescimento acelerado da produção científica nos países emergentes está estreitamente associado ao crescimento da colaboração doméstica.

Em segundo lugar, a tendência de crescimento da colaboração concomitantemente associado ao crescimento do erro-padrão médio (medida de variabilidade absoluta) é indício da existência e crescimento da heterogeneidade entre os valores das matrizes de colaboração, uma vez que alguns pares municipais apresentam elevado número de ligações, diferentemente dos valores baixos ou nulos da maioria dos pares intermunicipais.

¹⁰⁵ Embora seja bastante intuitiva e facilmente aplicável, essa abordagem possui importantes limitações, uma vez que as comparações são realizadas a partir da discretização temporal exógena de um processo contínuo (decidido pelo pesquisador ou imposto pela disponibilidade dos dados), além de não serem examinados os determinantes das mudanças estruturais das redes (MAGGIONI; UBERTI, 2011a). Portanto, o objetivo dessa abordagem deve ser puramente descritivo, em que não se permitem testes de qualquer hipótese específica, mas consiste numa valiosa etapa preliminar para a compreensão da evolução das redes ao longo do tempo.

Tabela 5 – Estatísticas descritivas das matrizes trienais de colaboração científica intermunicipal: todas as áreas do conhecimento

	1992-1994	1995-1997	1998-2000	2001-2003	2004-2006	2007-2009
Ligações Totais						
<i>Elementos</i>	21.307.456	21.307.456	21.307.456	21.307.456	21.307.456	21.307.456
<i>Soma</i>	776.688	1.727.464	4.461.722	9.474.504	16.566.804	23.578.972
<i>Média</i>	0,01	0,03	0,10	0,19	0,29	0,39
<i>Erro-padrão médio</i>	0,51	1,02	2,73	4,56	7,11	9,82
<i>Mínimo</i>	0	0	0	0	0	0
<i>Máximo</i>	121.746	192.684	424.281	609.379	964.061	1.343.394
Ligações Positivas						
<i>Elementos</i>	6.180	10.708	18.005	27.698	33.785	34.120
<i>Soma</i>	776.688	1.727.464	4.461.722	9.474.504	16.566.804	23.578.972
<i>Média</i>	52,6	68,4	118,6	145,2	184,5	253,0
<i>Erro-padrão médio</i>	698,8	844,6	1.800,2	1.881,6	2.554,8	3.440,3
<i>Mínimo</i>	1	1	1	1	1	1
<i>Máximo</i>	121.746	192.684	424.281	609.379	964.061	1.343.394
Ligações Intramunicipais						
<i>Elementos</i>	4.616	4.616	4.616	4.616	4.616	4.616
<i>Soma</i>	317.810	397.459	501.475	629.060	822.800	1.037.274
<i>Média</i>	27,9	55,6	131,4	226,2	367,0	512,6
<i>Erro-padrão médio</i>	793,1	1.245,8	3.292,3	4.269,2	6.483,6	8.755,9
<i>Mínimo</i>	0	0	0	0	0	0
<i>Máximo</i>	121.746	148.942	169.210	192.684	243.509	293.754
Ligações Intermunicipais						
<i>Elementos</i>	21.302.840	21.302.840	21.302.840	21.302.840	21.302.840	21.302.840
<i>Soma</i>	229.439	549.202	1.537.106	3.439.905	6.078.783	8.408.125
<i>Média</i>	0,00	0,01	0,04	0,07	0,11	0,14
<i>Erro-padrão médio</i>	0,11	0,25	0,78	1,29	1,90	2,57
<i>Mínimo</i>	0	0	0	0	0	0
<i>Máximo</i>	10.792	23.113	59.698	100.515	175.033	239.351

A elaboração das redes intermunicipais de colaboração científica permite a visualização da configuração espacial dos fluxos de conhecimento no Brasil. Como exemplo, a Figura 10 exibe um mapa com a caracterização de tais fluxos para o total das áreas do conhecimento no triênio 2007-2009, e outro mapa com a dimensão relativa dos graus dos municípios na rede de colaboração científica. Para as áreas do conhecimento isoladamente, os mapas das redes de colaboração são apresentados no Apêndice E.¹⁰⁶

As Tabelas 6 e 7 mostram a evolução dos principais fluxos de conhecimento intramunicipais e intermunicipais para os triênios selecionados. Já o Apêndice F apresenta os principais fluxos de conhecimento no triênio 2007-2009 em cada área do conhecimento. Como exemplo, para o total das áreas, foram contabilizadas 1.343.394 colaborações entre pesquisadores de São Paulo/SP (Tabela 6) e 76.716 colaborações entre pesquisadores de Campinas/SP e São Paulo/SP (Tabela 7) no triênio 2007-2009.

A análise da Figura 10 simultaneamente ao detalhamento da magnitude dos fluxos apresentado nas Tabelas 6 e 7 permite extrair resultados interessantes. De um modo geral, observa-se a localização sistemática dos fluxos na região Sudeste do país. Assim, é encontrado padrão similar àquele verificado na análise da produção científica municipal, uma vez que as principais ligações intermunicipais e intramunicipais dão-se, primordialmente, no âmbito das capitais dos estados do Sudeste e de municípios no interior do Estado de São Paulo que sediam universidades tradicionais. Fora desse contexto, verifica-se a intensificação das colaborações intramunicipais em Recife/PE, Fortaleza/CE, Goiânia/GO, Brasília/DF, Viçosa/MG e Santa Maria/RS, municípios que também abrigam universidades públicas tradicionais. Já o Apêndice F revela a diferenciação entre os padrões de localização dos principais fluxos de colaborações entre as áreas do conhecimento, já que em áreas como Ciências Biológicas e Ciências da Saúde os fluxos concentram-se, principalmente, no eixo São Paulo/SP, Campinas/SP e Ribeirão Preto/SP, enquanto que em Ciências Agrárias os fluxos são mais esparsos ao longo do território, em que municípios de diversos estados figuram dentre os principais.

¹⁰⁶ Os mapas apresentados foram elaborados no ambiente computacional do *software Gephi* (disponível em <https://gephi.org/>).

Rede de Colaboração Científica (2007-2009)

Graus Municipais na Rede (2007-2009)

Figura 10 – Mapas da rede de colaboração científica no Brasil no triênio 2007-2009 e dos graus municipais na rede

Tabela 7 – Evolução das principais ligações intermunicipais de colaboração científica no Brasil em triênios selecionados

1992-1994	Total	1995-1997	Total
Campinas/SP – São Paulo/SP	5.682	Campinas/SP – São Paulo/SP	9.890
Rio de Janeiro/RJ – São Paulo/SP	3.883	Rio de Janeiro/RJ – São Paulo/SP	9.500
Niterói/RJ – Rio de Janeiro/RJ	3.793	Niterói/RJ – Rio de Janeiro/RJ	7.199
Ribeirão Preto/SP – São Paulo/SP	2.607	Porto Alegre/RS – São Paulo/SP	4.682
Florianópolis/SC – São Paulo/SP	2.107	Ribeirão Preto/SP – São Paulo/SP	4.097
Araraquara/SP – São Paulo/SP	1.971	Belo Horizonte/MG – São Paulo/SP	4.085
Belo Horizonte/MG – São Paulo/SP	1.765	São Carlos/SP – São Paulo/SP	3.984
Porto Alegre/RS – São Paulo/SP	1.597	Curitiba/PR – São Paulo/SP	3.738
Botucatu/SP – São Paulo/SP	1.457	S. J. dos Campos/SP – São Paulo/SP	3.643
1998-2000		2001-2003	
Ribeirão Preto/SP – São Paulo/SP	40.727	Ribeirão Preto/SP – São Paulo/SP	48.657
Campinas/SP – São Paulo/SP	30.672	Campinas/SP – São Paulo/SP	41.538
Botucatu/SP – São Paulo/SP	22.587	Goiânia/GO – Brasília/DF	37.518
Rio de Janeiro/RJ – São Paulo/SP	15.839	Rio de Janeiro/RJ – São Paulo/SP	36.168
Piracicaba/SP – São Paulo/SP	14.249	Niterói/RJ – Rio de Janeiro/RJ	26.363
Niterói/RJ – Rio de Janeiro/RJ	12.563	São Carlos/SP – São Paulo/SP	22.649
Porto Alegre/RS – São Paulo/SP	10.139	Botucatu/SP – São Paulo/SP	20.108
São Carlos/SP – São Paulo/SP	9.532	Santa Maria/RS – Porto Alegre/RS	17.987
Belo Horizonte/MG – São Paulo/SP	9.173	Porto Alegre/RS – São Paulo/SP	17.057
2004-2006		2007-2009	
Campinas/SP – São Paulo/SP	72.698	Campinas/SP – São Paulo/SP	76.716
Ribeirão Preto/SP – São Paulo/SP	72.375	Ribeirão Preto/SP – São Paulo/SP	74.078
Rio de Janeiro/RJ – São Paulo/SP	56.346	Niterói/RJ – Rio de Janeiro/RJ	75.224
Niterói/RJ – Rio de Janeiro/RJ	41.536	Rio de Janeiro/RJ – São Paulo/SP	72.500
Goiânia/GO – Brasília/DF	35.948	Seropédica/RJ – Rio de Janeiro/RJ	65.348
Porto Alegre/RS – São Paulo/SP	33.655	Porto Alegre/RS – São Paulo/SP	47.343
Botucatu/SP – São Paulo/SP	31.152	Santa Maria/RS – Porto Alegre/RS	39.252
Santa Maria/RS – Porto Alegre/RS	30.151	Santo André/SP – São Paulo/SP	35.694
São Carlos/SP – São Paulo/SP	26.444	Curitiba/PR – São Paulo/SP	32.692

No tocante à análise da configuração espacial dos fluxos de conhecimento em cada área do conhecimento (Apêndice E), também é possível observar tanto a expansão das redes como a intensificação de suas ligações em todas as áreas, porém em proporções diferentes. A caracterização dos graus municipais ressalta a importância relativa dos municípios na articulação das relações na rede, *i.e.*, uma aproximação para o papel individual desempenhado por eles no processo de difusão do conhecimento na rede científica.¹⁰⁷

A análise da evolução dos valores de algumas métricas globais das redes de colaboração científica, apresentada nas Figuras 11 e 12, também permite obter resultados interessantes a respeito da articulação entre os pesquisadores brasileiros na produção compartilhada de conhecimento técnico-científico.

A Figura 11 (a) mostra a quantidade de municípios conectados nas redes de colaboração científica anuais para cada área de conhecimento. Em 1990, a rede de colaboração era restrita a pesquisadores de apenas 345 dos municípios brasileiros. Já em 2009, a rede de colaboração intermunicipal passou a contar com 1.348 municípios. Logo, é notável a expansão da rede de colaborações, uma vez que novos municípios foram incorporados à rede com o passar do tempo. A Figura 11 (b) apresenta o crescimento do total de arestas (ligações) nas redes de colaboração. Já a Figura 11 (c) apresenta o grau médio dos municípios na rede de colaboração em cada ano, em que se observa o crescimento desse valor de 0,8 em 1990 para 7,5 em 2009.¹⁰⁸ De maneira sucinta, a evolução dessas métricas é indício da expansão espacial das relações nas redes de colaboração, e mais precisamente, da intensificação das relações rumo a maior integração. Também é notável a diferenciação entre a evolução das métricas para as redes de cada área do conhecimento em particular, uma vez que a expansão e integração das redes ocorrem em intensidades distintas.

¹⁰⁷ Como exemplo, pode-se verificar que o município de Viçosa/MG (sede da UFV) possui o maior grau na rede de colaborações em Ciências Agrárias.

¹⁰⁸ Como descrito anteriormente, o grau de um nó em uma rede mede a quantidade de arestas distintas que estão ligadas a esse nó. O grau médio é a média dos graus dos nós individuais, assim, o valor 7,5 significa que, em média, cada município da rede está ligado a 7,5 municípios diferentes.

Figura 11 – Evolução da quantidade de municípios (a), do número de arestas (b) e do grau médio municipal (c) nas redes anuais de colaboração científica no Brasil

Figura 12 – Evolução da densidade (a), do diâmetro (b) e da transitividade (c) nas redes anuais de colaboração científica no Brasil

Figura 13 – Evolução do diâmetro (a) e da transitividade (b) nas redes trienais de colaboração científica no Brasil

A Figura 12 (a) apresenta a evolução dos valores referentes à densidade das redes de colaboração, em que se observa tendência semelhante à verificada no caso do grau médio.¹⁰⁹ Já as Figuras 12 (b) e 12 (c) apresentam o diâmetro e a transitividade das redes anuais. Contudo, a discretização temporal das redes entre anos não permite evidências acerca do comportamento temporal dessas métricas. Como alternativa, tais métricas foram calculadas segundo a periodização trienal das redes de colaboração (Figura 13). A partir dessa nova perspectiva, observa-se que os valores do diâmetro apresentam tendência de declínio com o passar do tempo (Figura 13 (a)) enquanto que a transitividade tem seu valor aumentado (Figura 13 (b)). Portanto, há evidência de uma tendência de redução da distância entre os nós (municípios) pertencentes à rede e de aumento da probabilidade de nós próximos estarem conectados.

¹⁰⁹ Conforme apresentado, a densidade de uma rede consiste na razão entre as ligações existentes e as possíveis, assim, é simplesmente o grau médio dividido por $(n-1)$.

O procedimento de discretização temporal trienal das matrizes intermunicipais de colaborações científicas deu-se por meio do agrupamento das matrizes anuais ao longo do período 1990-2010 em seis triênios (1992-1994, 1995-1997, 1998-2000, 2001-2003, 2004-2006 e 2007-2009), de acordo com os triênios selecionados pela avaliação trienal da CAPES. Por um lado, a determinação da janela temporal ideal em análises cientométricas é tema de debate na literatura, em que o período de um ano é encarado como muito curto para a verificação de transformações estruturais. Por outro lado, nas últimas décadas, o modelo de avaliação da produção científica no Brasil esteve baseado na determinação de padrões internacionais a serem perseguidos pelos pesquisadores brasileiros, o que funciona como um verdadeiro mecanismo de incentivo à pesquisa científica (VANZ, 2009; LETA, 2011; LEITE *et al*, 2011). Portanto, acreditamos que a avaliação trienal no Brasil pode ser encarada como uma janela ideal para análises cientométricas, uma vez que há estímulo para que os pesquisadores tentem concretizar seus esforços de pesquisa realizados em determinado triênio, transformando-os em publicações até o fim deste período, a fim de que o trabalho não seja contemplado somente numa avaliação futura.

5.2. Avaliação do papel da distância geográfica nas redes de colaboração científica

Como vimos, é provável que a proximidade geográfica exerça papel determinante na formação das redes de colaboração científica. Contudo, a medição de seus efeitos não é trivial, já que as redes se modificam e se desenvolvem ao longo do tempo.¹¹⁰

De acordo com a descrição do modelo realizada na seção 3.2, os parâmetros α_1 e α_2 e β_k estão associados às variáveis de origem (o_i) e destino (d_j) e às k variáveis de separação ($u_{ij}^{(k)}$) do modelo empírico a ser estimado:

$$Y_{ij} = \alpha_0 + o_i^{\alpha_1} \cdot d_j^{\alpha_2} \cdot \exp\left[\sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}\right] + \varepsilon_{ij} \quad ; \quad i, j = 1, \dots, n \quad (3.6)$$

¹¹⁰ Ainda não existem estudos sobre a dinâmica das mudanças nas estruturas das redes espaciais ao longo do tempo, o qual consiste na principal direção da evolução da pesquisa na área (TER WAL; BOSCHMA, 2009).

Em nosso caso, as variáveis de origem (o_i) e destino (d_j) foram medidas pelo total de publicações científicas em cada município.¹¹¹ Assim, espera-se que o total de colaborações entre pesquisadores dos municípios i e j (y_{ij}) dependa positivamente do total de publicações em cada município, visto que quanto maior a produção científica de dois municípios, maior deve ser a probabilidade de haver colaboração entre pesquisadores localizados nesses municípios.

Quanto às variáveis de separação, duas medidas foram utilizadas. Primeiramente, foi construída uma matriz de distância geográfica, mensurada de maneira contínua, em que cada elemento $u_{ij}^{(1)}$ apresenta o cálculo da distância em quilômetros (km) entre os dois municípios i e j .¹¹² Cabe ressaltar que as distâncias intramunicipais $u_{ii}^{(1)}$ (termos da diagonal principal da matriz de distâncias) foram calculadas por meio da fórmula de Bröcker (1989), como é usual na literatura.¹¹³

Além da distância geográfica, foi introduzida uma segunda variável de separação. A partir da atribuição do valor $u_{ij}^{(2)} = 1$ aos pares de municípios i e j em que ambos possuem *campi* de universidades públicas (e zero caso contrário), foi construída uma matriz que representa a distância institucional entre os municípios brasileiros.¹¹⁴ Assim, é esperado que o fato de dois municípios possuírem instalações de *campi* de universidades públicas seja diretamente proporcional ao aumento da probabilidade de colaboração científica entre pesquisadores desses municípios.

Para as estimações dos modelos gravitacionais, foi utilizada uma amostra reduzida de 105 municípios, conjunto selecionado e correspondente aos municípios que mais colaboraram no triênio 2007-2009, responsáveis por 87,6% do total de publicações e por 77,4% do total de coautorias contabilizadas no período. A escolha pela redução da amostra deve-se ao fato

¹¹¹ Conforme destacado anteriormente, optou-se pela contabilização da participação absoluta de cada pesquisador no caso de publicações em coautoria.

¹¹² Devido à inexistência de dados consolidados sobre as distâncias entre os municípios brasileiros, optamos pela medição da menor distância entre eles. Isso se deu por meio da aplicação da fórmula da distância geodésica aos dados de latitude e longitude dos centroides dos municípios.

¹¹³ Segundo Bröcker (1989), a distância intrarregional é função da área do município i (A_i) como abaixo:

$$u_{ii}^{(1)} = \left(\frac{2}{3}\right) \cdot \left(\frac{A_i}{\pi}\right)^{0,5}$$

¹¹⁴ A construção da matriz da distância institucional para cada triênio deu-se por meio da identificação dos municípios brasileiros que possuíam *campi* de universidades públicas (estaduais ou federais) no primeiro ano de cada triênio analisado.

de que a matriz completa de colaborações (com 1.347 municípios) apresentou 97,8% de valores nulos, quantidade tão elevada que inviabilizaria a estimação dos modelos.¹¹⁵ Logo, a utilização dos fluxos de colaboração entre os 105 municípios totalizou 11.025 observações.

Inicialmente, é interessante observar que houve aumento da distância média das colaborações entre pesquisadores dos 105 municípios selecionados (Tabela 8). Tal fato sugere a hipótese de que o papel impeditivo exercido pela distância geográfica nas redes de colaboração científica tenha reduzido sua importância com o passar do tempo.

Tabela 8 – Evolução da distância geográfica média nas colaborações científicas entre pesquisadores dos 105 municípios que mais colaboraram no triênio 2007-2009

	<i>1992-1994</i>	<i>1995-1997</i>	<i>1998-2000</i>	<i>2001-2003</i>	<i>2004-2006</i>	<i>2007-2009</i>
<i>Distância (km)</i>	357,8	387,6	425,4	490,5	493,9	465,2

Os resultados das estimativas dos modelos de Poisson e binomial negativo para os 105 municípios brasileiros que mais colaboraram no triênio 2007-2009 são apresentados na Tabela 9.¹¹⁶ Alguns aspectos importantes merecem atenção especial.

Conforme esperado, as estimativas das medidas de massa (origem e destino) são estatisticamente significantes e próximas de um para todos os triênios selecionados, o que é indício da boa especificação desses modelos. Ademais, observa-se que as estimativas do modelo de Poisson não apresentam grande variabilidade temporal, o que poderia levar à interpretação de que a intensidade de colaborações tenha sido semelhante ao longo do período de tempo analisado, dada certa quantidade de publicações. Contudo, cabe ressaltar que tal interpretação não deve ser realizada de maneira direta, uma vez que o método utilizado para a contabilização das publicações faz com que crescimento das publicações esteja de alguma forma associado ao crescimento da colaboração científica. Assim, quanto à análise das estimativas dos coeficientes das medidas de massa, devemos nos deter somente ao indício de boa especificação do modelo, e não de seus valores intrínsecos.

¹¹⁵ A determinação exata de 105 municípios deu-se a partir da verificação da existência de um terço de valores nulos nesse caso.

¹¹⁶ As estimações foram realizadas no ambiente computacional do *software STATA 11.2*.

Também se observa na Tabela 9 a significância estatística e o sinal positivo das estimativas referentes à distância institucional para todos os triênios. Tal resultado era esperado e sustenta a hipótese inicial, visto que o fato de dois municípios possuírem instalações de *campi* de universidades públicas aumenta a probabilidade de haver colaboração científica entre pesquisadores desses municípios.

Conquanto os resultados apresentados sejam importantes, nosso interesse principal recai sobre o efeito da distância geográfica sobre a probabilidade de colaboração. Observa-se na Tabela 9 que as estimativas da distância geográfica são estatisticamente significantes e com sinal negativo para todos os triênios selecionados. Tal resultado corrobora a hipótese de que o aumento da distância entre dois pesquisadores reduz a probabilidade de colaboração entre eles, *coeteris paribus*.¹¹⁷ Todavia, a interpretação das estimativas nos modelos de contagem (não-lineares) não é tão imediata como no modelo clássico de regressão linear. O valor encontrado para o triênio 2007-2009 (-0.0017769) significa que um aumento da distância em 100 quilômetros (km) entre dois pesquisadores reduz, em média, 16,3% a probabilidade de haver colaboração entre eles. Como o efeito não é linear, um aumento da distância em 300 km (600 km) reduz a probabilidade de haver colaboração em 41,3% (65,6%), em média.¹¹⁸

Além da análise da significância estatística das estimativas dos coeficientes, algumas observações a respeito da comparação entre suas magnitudes também são pertinentes. Inicialmente, esperamos que o papel desempenhado pela distância geográfica tenha se reduzido com o passar do tempo, visto que o avanço tecnológico nos meios de transporte e telecomunicações propiciou uma revolução nos relacionamentos entre as pessoas, por permitir a comunicação instantânea e o envio e compartilhamento de arquivos eletrônicos, fatores que atuam como mecanismos facilitadores para a ocorrência de interações entre os pesquisadores. Entretanto, tal hipótese não foi corroborada pelos resultados das estimações do modelo apresentadas na Tabela 9, já que é possível observar que a tendência inicial de

¹¹⁷ Resultado semelhante foi encontrado em Scherngell e Hu (2011) e em Pan *et al* (2012).

¹¹⁸ O efeito de um acréscimo na variável explicativa sobre a variável dependente (de contagem), em termos percentuais (mantendo-se constante as demais variáveis), é calculado pela seguinte expressão (LONG; FREESE, 2001, p.232):

$$100 \cdot \frac{E(y|x, x_k + \delta) - E(y|x, x_k)}{E(y|x, x_k)} = 100 \cdot [\exp(\beta_k \cdot \delta) - 1]$$

redução das estimativas dos coeficientes associados à distância geográfica apresenta uma ruptura, principalmente no último triênio analisado (2007-2009), em que ocorre aumento do valor da estimativa em relação ao triênio anterior (2004-2006).¹¹⁹

Quanto à comparação entre os modelos, observa-se na Tabela 9 que a estimativa do parâmetro de heterogeneidade é estatisticamente significativa, o que é evidência da presença de superdispersão nos dados analisados. Assim, a escolha da versão binomial negativa, em contrapartida à de Poisson, parece ser a especificação mais adequada para o modelo de dados de contagem proposto.

A Tabela 10 apresenta os resultados das versões infladas de zeros dos modelos de Poisson e binomial negativo. De maneira geral, os resultados assemelham-se àqueles apresentados na Tabela 9, visto que as estimativas dos coeficientes são estatisticamente significantes. Os testes de Vuong apontam para a escolha das versões infladas de zeros, quando esses são comparados aos modelos iniciais. Já o teste da razão de verossimilhança, o qual permite a escolha entre os modelos ZIP e ZINB, aponta para a escolha do modelo binomial negativo inflado de zeros (ZINB).¹²⁰

Conforme analisado, as estimativas dos parâmetros associados às variáveis explicativas utilizadas são estatisticamente significantes. Entretanto, é possível que outras dimensões de proximidade sejam fundamentais na determinação da intensidade das interações de colaborações científicas entre os pesquisadores, ou seja, é possível que existam outras variáveis explicativas importantes, mas que estão omitidas no modelo. Tal situação pode fazer com que o modelo estimado esteja sujeito ao problema de endogeneidade, visto que é possível que o termo de erro seja composto de variáveis (omitidas) que estejam correlacionadas de alguma maneira às variáveis explicativas contidas no modelo.

Com o intuito de tentar contornar esse problema, foram estimados novamente os modelos de Poisson e binomial negativo para o total das áreas do conhecimento nos triênios selecionados, mas com a inclusão de uma nova variável explicativa para controlar o

¹¹⁹ Semelhantemente, os resultados do estudo de Hoekman *et al* (2010) não corroboram a hipótese de que o efeito da distância geográfica tenha se reduzido com o passar do tempo.

¹²⁰ O teste da razão de verossimilhança permite a comparação entre os modelos ZIP e ZINB por meio da seguinte estatística:

$$LR = -2(\mathcal{L}_{ZIP} - \mathcal{L}_{ZINB}) \sim \chi^2(1, LR)/2$$

problema de endogeneidade: o total de coautorias no triênio 1992-1994. A ideia empregada consiste no fato de que, ao incluir o total de coautorias no período inicial como variável explicativa na modelagem das coautorias nos períodos seguintes, as outras dimensões de proximidade que atuam na determinação das interações entre os pesquisadores devem ser correlacionadas com as coautorias no período inicial. Logo, a inclusão do total de coautorias defasadas (associadas a períodos anteriores), permite, de alguma maneira, a consideração simultânea de outras dimensões de proximidade como variáveis explicativas, de modo a retirá-las do termo de erro, e, assim, o problema de correlação entre o termo de erro e as variáveis explicativas (endogeneidade) pode ser reduzido, ou até mesmo extinguido. Os resultados desses modelos estendidos são apresentados na Tabela 11.¹²¹ De maneira sucinta, verifica-se na Tabela 11 que todas as conclusões a respeito do sinal e significância estatísticas das estimativas dos coeficientes realizadas anteriormente também são válidas para as estimações do modelo estendido. Inicialmente, observa-se que a nova variável incluída possui coeficiente estatisticamente significativo, o que pode ser indício da presença de endogeneidade nos modelos anteriores. Verifica-se também que os valores das estimativas dos coeficientes no modelo estendido são menores do que os encontrados anteriormente.

¹²¹ Foram estimados modelos a partir do triênio 1995-1997, uma vez que a nova variável explicativa incluída consiste no total de coautorias no triênio 1992-1994.

Tabela 9 – Estimativas dos modelos de Poisson e binomial negativo para os triênios selecionados: todas as áreas do conhecimento

	1992-1994	1995-1997	1998-2000	2001-2003	2004-2006	2007-2009
Poisson						
Origem – Destino ($\alpha_1 = \alpha_2$)	0,82127*** (0,07994)	0,77950*** (0,06784)	0,82201*** (0,06838)	0,78581*** (0,0579629)	0,79731*** (0,0544)	0,78859*** (0,05789)
Distância Geográfica (β_1)	-0,00195*** (0,00030)	-0,00192*** (0,00025)	-0,00174*** (0,00020)	-0,00153*** (0,00016)	-0,00154*** (0,00015)	-0,00177*** (0,00017)
Distância Institucional (β_2)	0,36443*** (0,1769)	0,41621*** (0,14363)	0,16094*** (0,12549)	0,31369*** (0,10611)	0,29401*** (0,10686)	0,42871*** (0,11636)
Constante (α_0)	-7,8571*** (1,3045)	-7,2514*** (1,1495)	-7,8745*** (1,2150)	-7,3404*** (1,0601)	-7,5977*** (1,0349)	-7,3564*** (1,1167)
Binomial Negativo						
Origem – Destino ($\alpha_1 = \alpha_2$)	0,85214*** (0,02546)	0,81110*** (0,02155)	0,74925*** (0,02520)	0,72744*** (0,02193)	0,73588*** (0,02193)	0,64376*** (0,2798)
Distância Geográfica (β_1)	-0,00080*** (0,00006)	-0,00086*** (0,00004)	-0,00077*** (0,00003)	-0,00077*** (0,00001)	-0,00083*** (0,0004)	-0,00088*** (0,0003)
Distância Institucional (β_2)	0,20469*** (0,09759)	0,1340*** (0,07653)	0,25959*** (0,06870)	0,18238*** (0,06341)	0,07382*** (0,05982)	0,20527*** (0,05890)
Constante (α_0)	-8,8486*** (0,25238)	-8,2565*** (0,24356)	-7,2966*** (0,3405)	-6,7090*** (0,27935)	-6,8110*** (0,31460)	-5,0248*** (0,45363)
Heterogeneidade (α)	6,0820* (0,19144)	5,0892* (0,11670)	4,5624* (0,09137)	3,8480* (0,06647)	3,6189* (0,06441)	3,7508* (0,06666)

Notas: *i*) $n^2=11.025$ observações; *ii*) os erros-padrão estão entre parênteses; *iii*) ***, ** e * referem-se às estimativas estatisticamente significantes aos níveis de significância de 0,001, 0,01 e 0,05, respectivamente.

Tabela 10 – Estimativas dos modelos ZIP e ZINB nos triênios selecionados: todas as áreas do conhecimento

	1992-1994	1995-1997	1998-2000	2001-2003	2004-2006	2007-2009
ZIP						
Origem – Destino ($\alpha_1 = \alpha_2$)	0,78074 ^{***} (0,00092)	0,75185 ^{***} (0,0006)	0,80186 ^{***} (0,00041)	0,76963 ^{***} (0,00030)	0,78123 ^{***} (0,00023)	0,77849 ^{***} (0,00020)
Distância Geográfica (β_1)	-0,00192 ^{***} (4,00.10 ⁻⁶)	-0,00188 ^{***} (2,59.10 ⁻⁶)	-0,00172 ^{***} (1,51.10 ⁻⁶)	-0,00152 ^{***} (9,39.10 ⁻⁷)	-0,00153 ^{***} (7,19.10 ⁻⁷)	-0,00176 ^{***} (6,51.10 ⁻⁷)
Distância Institucional (β_2)	0,35437 ^{***} (0,00334)	0,39968 ^{***} (0,00231)	0,15269 ^{***} (0,00139)	0,30125 ^{***} (0,00100)	0,28582 ^{***} (0,00359)	0,41281 ^{***} (0,00064)
Constante (α_0)	-7,05112 ^{***} (0,01261)	-6,00882 ^{***} (0,00882)	-7,44622 ^{***} (0,00612)	-6,98269 ^{***} (0,00449)	-7,30067 ^{***} (0,00359)	-7,12039 ^{***} (0,00315)
Vuong (ZIP x Poisson)	17,29 ^{***}	17,07 ^{***}	20,38 ^{***}	20,45 ^{***}	22,92 ^{***}	23,29 ^{***}
ZINB						
Origem – Destino ($\alpha_1 = \alpha_2$)	0,53245 ^{***} (0,13615)	0,56452 ^{***} (0,01171)	0,57718 ^{***} (0,01119)	0,59960 ^{***} (0,01128)	0,62453 ^{***} (0,01146)	0,56556 ^{***} (0,01171)
Distância Geográfica (β_1)	-0,00069 ^{***} (0,00002)	-0,00075 ^{***} (0,00002)	-0,00071 ^{***} (0,00002)	-0,00073 ^{***} (0,00001)	-0,00077 ^{***} (0,00002)	-0,00084 ^{***} (0,00001)
Distância Institucional (β_2)	0,06398 ^{***} (0,04710)	0,03475 ^{***} (0,04102)	0,13497 ^{***} (0,03725)	0,09068 ^{***} (0,03516)	0,02596 ^{***} (0,03403)	0,12315 ^{***} (0,03586)
Constante (α_0)	-3,43762 ^{***} (0,14343)	-3,89693 ^{***} (0,13169)	-4,02405 ^{***} (0,13119)	-4,22713 ^{***} (0,13833)	-4,59020 ^{***} (0,14309)	-3,37807 ^{***} (0,14616)
Heterogeneidade (α)	1,6661 [*] (0,03434)	1,7251 [*] (0,02983)	1,8355 [*] (0,02733)	1,8653 [*] (0,02503)	1,8665 [*] (0,02419)	2,0805 [*] (0,02615)
Vuong (ZINB x Bin. Neg.)	83,05 ^{***}	99,67 ^{***}	92,65 ^{***}	73,58 ^{***}	61,97 ^{***}	53,86 ^{***}
Razão de Verossimilhança (ZIP x ZINB)	ZINB ^{***}	ZINB ^{***}	ZINB ^{***}	ZINB ^{***}	ZINB ^{***}	ZINB ^{***}

Notas: i) $n^2=11.025$ observações; ii) os erros-padrão estão entre parênteses; iii) ***, ** e * referem-se às estimativas estatisticamente significantes aos níveis de significância de 0,001, 0,01 e 0,05, respectivamente; iv) ZIP (*Zero-Inflated Poisson Model*) e ZINB (*Zero-Inflated Negative Binomial Model*) referem-se às versões infladas de zeros dos modelos de Poisson e Binomial Negativo.

Tabela 11– Estimativas dos modelos de Poisson e binomial negativo nos triênios selecionados, controlada pelo total de colaborações no triênio 1992-1994: todas as áreas do conhecimento

	1995-1997	1998-2000	2001-2003	2004-2006	2007-2009
<i>Poisson</i>					
Origem – Destino ($\alpha_1 = \alpha_2$)	0,66042 ^{***} (0,02677)	0,70998 ^{***} (0,03091)	0,69243 ^{***} (0,02643)	0,70839 ^{***} (0,02593)	0,69218 ^{***} (0,02887)
Distância Geográfica (β_1)	-0,00162 ^{***} (0,00018)	-0,00149 ^{***} (0,00014)	-0,00136 ^{***} (0,00012)	-0,00138 ^{***} (0,00013)	-0,00159 ^{***} (0,00015)
Distância Institucional (β_2)	0,42902 ^{***} (0,14652)	0,17579 ^{**} (0,13859)	0,34449 ^{***} (0,10963)	0,31687 ^{***} (0,11000)	0,44880 ^{***} (0,11931)
Colaborações Defasadas	0,000015 ^{***} (1,71.10 ⁻⁶)	0,000014 ^{***} (1,60.10 ⁻⁶)	0,000014 ^{***} (1,70.10 ⁻⁶)	0,000013 ^{***} (1,77.10 ⁻⁶)	0,000014 ^{***} (1,95.10 ⁻⁶)
Constante (α_0)	-5,2851 ^{***} (0,2524)	-5,9106 ^{***} (0,5057)	-5,6409 ^{***} (0,4531)	-5,9099 ^{***} (0,4628)	-5,4888 ^{***} (0,5204)
<i>Binomial Negativo</i>					
Origem – Destino ($\alpha_1 = \alpha_2$)	0,74508 ^{***} (0,02490)	0,69183 ^{***} (0,0268)	0,67688 ^{***} (0,0242)	0,68449 ^{***} (0,0247)	0,57955 ^{***} (0,0306)
Distância Geográfica (β_1)	-0,00074 ^{***} (0,00005)	-0,00066 ^{***} (0,00004)	-0,00071 ^{***} (0,00004)	-0,00075 ^{***} (0,00004)	-0,00078 ^{***} (0,00004)
Distância Institucional (β_2)	0,13552 [*] (0,06754)	0,26920 ^{***} (0,06017)	0,18305 ^{***} (0,05882)	0,07824 [*] (0,05628)	0,20852 ^{***} (0,05435)
Colaborações Defasadas	0,00141 ^{**} (0,00064)	0,00117 ^{**} (0,00046)	0,00088 ^{**} (0,00038)	0,00085 ^{**} (0,00038)	0,00109 ^{**} (0,00046)
Constante (α_0)	-7,5930 ^{***} (0,27508)	-6,6500 ^{***} (0,35502)	-6,0495 ^{***} (0,31510)	-6,0793 ^{***} (0,35752)	-4,0916 ^{***} (0,48631)
Heterogeneidade (α)	4,9194 [*] (0,1189)	4,4436 [*] (0,0879)	3,7816 [*] (0,0664)	3,5593 [*] (0,0646)	3,6694 [*] (0,0655)

Notas: i) $n^2=11.025$ observações; ii) os erros-padrão estão entre parênteses; iii) ***, ** e * referem-se às estimativas estatisticamente significantes aos níveis de significância de 0,001, 0,01 e 0,05, respectivamente.

Figura 14 – Efeito da distância geográfica sobre a probabilidade de colaboração científica no triênio 2007-2009: por área do conhecimento

A Figura 14 evidencia o fato de que o efeito da distância geográfica sobre a probabilidade de colaboração não é linearmente proporcional à distância. De acordo com a figura, para o total das áreas do conhecimento, o distanciamento de dois pesquisadores em 100 km provoca a redução em 16%, em média, da probabilidade de haver colaboração entre eles. Já o distanciamento em 400 km reduz praticamente pela metade a probabilidade de colaboração.¹²²

Ademais, observa-se na Figura 14 que o efeito da distância varia consideravelmente entre as redes de colaborações científicas das diferentes áreas do conhecimento. Como exemplo, o distanciamento de 400 km entre dois pesquisadores reduz em cerca de 40% a probabilidade de haver colaboração caso eles sejam da área de Linguística, Letras e Artes, enquanto o impacto chega a 65% no caso de Ciências Agrárias e Ciências Exatas e da Terra.

¹²² É interessante notar que a proximidade total (distância geográfica nula) está associada à verificação da probabilidade de 100% de colaboração, *i.e.*, a distância nula não possui impacto sobre a probabilidade de colaboração. Entretanto, cabe ressaltar que interpretações a partir do modelo devem excluir tal situação, visto que sempre admitimos alguma distância geográfica entre dois pesquisadores, mesmo que ambos sejam do mesmo município.

A verificação de diferenças no papel impeditivo exercido pela distância geográfica nas redes de colaboração científica de cada área do conhecimento era esperada, já que a transmissão de informações em cada área possui especificidades intrínsecas, caracterizadas pela existência de conhecimentos tácitos e complexos, o que pode exigir um número maior de encontros entre os pesquisadores ao longo do período da pesquisa e ampliar as dificuldades advindas do maior distanciamento espacial entre eles. Além disso, a própria condução de pesquisas em cada área do conhecimento possui características inerentes, *i.e.*, é esperado que o desenvolvimento recente dos meios de comunicação e de transporte tenha contribuído enormemente para o compartilhamento de informações e facilidades de locomoção, de maneira a reduzir o papel impeditivo da distância geográfica sobre as colaborações entre pesquisadores da área de Linguística, Letras e Artes. Todavia, outras áreas do conhecimento são caracterizadas pela execução de atividades de pesquisa em instalações físicas, tais como laboratórios, pela utilização intensa de instrumentação, e pela necessidade de encontros frequentes, o que faz com que a distância geográfica ainda desempenhe papel impeditivo fundamental na determinação do número de interações colaborativas entre os pesquisadores, caso das áreas de Ciências Agrárias e Ciências Exatas e da Terra.

6. CONSIDERAÇÕES FINAIS

A criação, difusão e transmissão do conhecimento científico para atividades tecnológicas são processos complexos e economicamente fundamentais, visto que a pesquisa científica básica desempenha papel determinante como impulsionador de avanços tecnológicos e, por conseguinte, da evolução do sistema econômico e do desenvolvimento regional. Tais processos são espacialmente localizados, o que evidencia a importância de análises cientométricas que visem à identificação de padrões de configuração espacial dos fluxos de conhecimento e da atuação desempenhada pela geografia na determinação das interações entre os pesquisadores nas redes de geração e difusão do conhecimento. A colaboração na produção de conhecimento é o motor da evolução recente da ciência moderna e consiste num mecanismo capaz de melhorar a qualidade e o valor econômico dos resultados obtidos, o que torna a análise espacial das redes de colaboração científica uma área de pesquisa emergente nos estudos sobre redes de conhecimento, principalmente por permitir o enfrentamento de uma das questões mais proeminentes no campo da geografia econômica: a determinação dos efeitos da proximidade geográfica nos processos de geração e difusão do conhecimento e de inovação.

Sob a perspectiva de compreender a importância do papel da geografia na intermediação das interações entre os pesquisadores nos processos de produção e difusão do conhecimento no Brasil, foram construídas redes de colaboração científica entre 1.347 municípios brasileiros, a partir da extração, consolidação e mensuração do total de coautorias em publicações científicas cadastradas no conjunto de informações disponíveis nos Currículos Lattes de 1.131.912 pesquisadores. A metodologia empregada permitiu a análise espacial da produção e da colaboração científica em todas as áreas do conhecimento ao longo do período compreendido entre 1990 e 2010.

A partir dessa inédita perspectiva de análise da produção científica brasileira, algumas evidências revelaram-se importantes. Primeiramente, verificamos o crescimento acelerado tanto da produção como da colaboração científica em todas as áreas do conhecimento, porém, com indícios de desaceleração. Em segundo lugar, vimos que a geografia da produção e colaboração científica no país é marcada por intensa heterogeneidade espacial,

com concentração sistemática da produção e dos fluxos de conhecimento nos municípios das regiões Sudeste e Sul, com destaque às capitais dos estados e aos municípios que sediam universidades públicas (federais e estaduais) consolidadas no cenário científico nacional. Além disso, a peculiaridade quanto ao perfil colaborativo de cada domínio científico permitiu a caracterização de padrões particulares quanto à configuração espacial dos fluxos de conhecimento. Em terceiro lugar, destacou-se a evidência de que o recente crescimento da produção científica brasileira parece ter sido acompanhado por um processo de desconcentração espacial, em que os municípios de produção intermediária ganharam destaque.

Quanto ao papel desempenhado pela distância geográfica na articulação das redes de colaboração científica, os resultados apontam para o fato de que esta ainda é determinante na dinâmica de interação entre os pesquisadores brasileiros, e não há evidências de que seu efeito tenha diminuído com o passar do tempo, apesar do desenvolvimento recente dos meios de transporte e comunicação. Mais precisamente, o aumento de 100 km de distância entre dois pesquisadores implica na redução, em média, de 16% da probabilidade de haver colaboração entre eles. Ademais, verificou-se que a magnitude do efeito impeditivo da distância na articulação das redes de colaboração científica varia consideravelmente entre os pesquisadores das diferentes áreas do conhecimento, devido, em grande parte, às características intrínsecas das atividades de pesquisa de cada área particular. É importante ressaltarmos a interpretação de que a proximidade geográfica por si só, embora seja fundamental, não deve ser entendida como condição necessária e suficiente para a interação entre os pesquisadores nos processos de criação e difusão do conhecimento, mas a maior proximidade geográfica consiste num mecanismo facilitador eficaz para a ocorrência dessas interações, principalmente quando simultânea à verificação de outras dimensões da proximidade, tais como a institucional, cognitiva e social.

Por fim, acreditamos que as informações obtidas por meio da abordagem utilizada e a perspectiva de análise adotada poderiam contribuir significativamente para a melhor compreensão da importância das redes de colaboração científica no tocante à formulação de políticas de C&T&I no Brasil por meio de três aspectos fundamentais: primeiramente, a percepção de que os objetivos de desconcentração regional da atividade científica e desenvolvimento regional estão estreitamente vinculados ao funcionamento das redes de

produção e difusão de conhecimento ao longo do território nacional; em segundo lugar, a noção de que, além da importância indiscutível da colaboração internacional, o estímulo a maior interação entre os pesquisadores brasileiros no processo de criação do conhecimento científico por meio da colaboração doméstica poderia desempenhar papel determinante em relação ao propósito de incremento de qualidade à produção científica, primordialmente nas regiões cientificamente menos tradicionais; e finalmente, a ideia de que o aumento da capacidade de acesso de firmas inovativas às fontes de conhecimento pode ter seu efeito potencializado em termos de atividades inovativas caso essas fontes estejam integradas às redes de produção compartilhada de conhecimento técnico-científico.

REFERÊNCIAS

ACOSTA, Manuel *et al.* *Factors affecting inter-regional academic scientific collaboration within Europe: the role of economic distance.* **Scientometrics**, v. 87, n. 1, p. 63-74, 2011.

ÁCS, Zoltán J. *Jaffe-Feldman-Varga: the search for knowledge spillovers.* In: VARGA, Attila. (Ed.). **Universities, knowledge transfer and regional development: geography, entrepreneurship and policy.** Cheltenham: Edward Elgar, 2009, p. 36-56.

ÁCS, Zoltán J. *et al.* *Patents and innovation counts as measures of regional production of new knowledge.* **Research Policy**, v. 31, p. 1069-1085, 2002.

ADAMS, Jonhatan; KING, Christopher. **Global research report: Brazil.** Leeds: Evidence, 2009.

ALBORNOZ, Mario *et al.* *Latin America.* In: UNESCO (Ed.). **Science report 2010.** Paris: Unesco Publishing, 2010, p. 77-101.

ALBUQUERQUE Eduardo *et al.* *A distribuição espacial da produção científica e tecnológica brasileira: uma descrição de estatísticas de produção local de patentes e artigos científicos.* **Revista Brasileira de Inovação**, v. 1, n. 2, p. 225-25, 2002.

ALBUQUERQUE Eduardo *et al.* *Produção científica e tecnológica das regiões metropolitanas brasileiras.* **Revista de Economia Contemporânea**, v. 9, n. 3, p. 615-642, 2005.

ANSELIN, Luc *et al.* *Geographical spillovers and university research: a spatial econometric perspective.* **Growth and Change**, v. 31, p. 501-515, 2000.

AUDRETSCH; David B.; FELDMAN, Maryann P. *R&D spillovers and the geography of innovation and production.* **The American Economic Review**, v. 86, n. 3, p. 630-640, 1996.

BARROS, Fernando A. F. de. *Os Desequilíbrios Regionais da Produção Técnico-Científica.* **São Paulo em Perspectiva**, v. 14, n. 3, p. 12-19, 2000.

BARTHÉLEMY, Marc. *Spatial networks.* **Physics Reports**, v. 499, p. 1–101, 2011.

BEAVER, Donald Deb. *Reflections on scientific collaboration (and its study): past, present, and future.* **Scientometrics**, v. 52, n. 3, p. 365-377, 2001.

BERGMAN, Edward, M.. *Embedding network analysis in spatial studies of innovation.* **Annals of Regional Science**, v. 43, p. 559-565, 2009.

BOSCHMA, Ron. *Proximity and innovation: a critical assessment.* **Regional Studies**, v. 39, n. 1, p. 61-74, 2005.

BOSCHMA, Ron A.; FRENKEN, Koen. *Introduction: applications of evolutionary economic geography.* In: FRENKEN, Koen. (Ed.). **Applied evolutionary economics and economic geography.** Cheltenham: Edward Elgar, 2007, p. 1-24.

BOSCHMA, Ron A.; FRENKEN, Koen. The spatial evolution of innovation networks: a proximity perspective. In: BOSCHMA, Ron; MARTIN, Ron. (Eds.). *The handbook of evolutionary economic geography*. Cheltenham: Edward Elgar, 2010, p. 120-135.

BOSCHMA, Ron A.; MARTIN, Ron. The aims and scope of evolutionary economic geography. In: BOSCHMA, Ron; MARTIN, Ron. (Eds.). *The handbook of evolutionary economic geography*. Cheltenham: Edward Elgar, 2010, p. 120-135.

BOSHOF, Nelius. South–South research collaboration of countries in the Southern African Development Community (SADC). *Scientometrics*, v. 84, n. 2, p. 481-503, 2010.

BOZEMAN, Barry; LEE, Sooho. *The Impact of Research Collaboration on Scientific Productivity*. *Social Studies of Science*, v. 35, n. 5, p. 673-702, 2005.

BRÖCKER, Johannes. *Partial Equilibrium Theory of Interregional Trade and the Gravity Model*. *Papers in Regional Science*, v. 66, n. 1, 1989, p. 7-18.

CAMERON, A. Colin; TRIVEDI, Pravin K. *Regression analysis of count data*. Cambridge: Cambridge University Press, 1998.

CARRINCAZEUX, Christophe; LUNG, Yannick; VICENTE, Jérôme. *The scientific trajectory of the French school of proximity: interaction and institution-based approaches to regional innovation systems*. *European Planning Studies*, v. 16, n. 5, p. 617-628, 2008.

CASELLA, George; BERGER, Roger L. *Inferência estatística*. São Paulo: Cengage Learning, 2010.

CHIARINI Tulio *et al*, A geografia da produção de novos conhecimentos: a dinâmica do ‘quarteto científico’ no Brasil, 2000 a 2010. *Revista Economia & Tecnologia*, v. 9, n. 3, p. 137-172, 2013.

CHINCHILLA-RODRÍGUEZ, Zaida *et al*. *New approach to the visualization of international scientific collaboration*. *Information Visualization*, v. 9, n. 4, p. 277–287, 2010

CNPq – CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. *Plataforma Lattes*. Brasília. Disponível em: <http://www.lattes.cnpq.br/>. Acesso em: out. 2012.

CRUZ, Carlos Henrique de Brito; CHAIMOVICH, Hernan. *Brazil*. In: UNESCO (Ed.). *Science report 2010*. Paris: Unesco Publishing, 2010, p. 103-121.

DALTON, Rex. *High hopes for Brazilian science*. *Nature*, v. 465, n. 10, p. 674-674, 2010.

DANELL, Rickard; PERSSON, Olle. *Regional R&D activities and interactions in the Swedish Triple Helix*. *Scientometrics*, v. 58, n. 2, p. 203-218.

DIGIAMPIETRI, Luciano A. *et al.* Minerando e caracterizando dados de currículos lattes, 2011. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**. Curitiba, 2012, p. 1-12.

DRUCKER, Joshua; GOLDSTEIN, Harvey. *Assessing the regional economic development impacts of universities: a review of current approaches*. **International Regional Science Review**, v. 30, n. 1, p. 20-46, 2007.

FAPESP – FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. Análise da produção científica a partir de publicações em periódicos especializados. In: **Indicadores de ciência, tecnologia e inovação em São Paulo 2010**. São Paulo, 2011, v. 1, cap. 4.

_____. Dimensão regional dos esforços de ciência, tecnologia e inovação no Estado de São Paulo. In: **Indicadores de ciência, tecnologia e inovação em São Paulo 2010**. São Paulo, 2011, v. 1, cap. 8.

FELDMAN, Maryann P.; AVNIMELECH, Gil. *Knowledge spillovers and the geography of innovation – revisited: a 20 years' perspective on the field on geography of innovation*. In: AUDRETSCH, David B. *et al* (Eds.). **Handbook of Research on Innovation and Entrepreneurship**. Cheltenham: Edward Elgar, 2011, p. 150-160.

FRENKEN, Koen *et al.* *Spatial scientometrics: towards a cumulative research program*. **Journal of Informetrics**, v.3, p. 222-232, 2009.

FUJITA, Masahisa; KRUGMAN, Paul. *The new economic geography: past, present and the future*. **Papers in Regional Science**, v. 83, p. 139-164, 2004.

GLÄNZEL, Wolfgang. *National characteristics in international scientific co-authorship relations*. **Scientometrics**, v. 51, n. 1, p. 69-115, 2001.

GLÄNZEL, Wolfgang *et al.* Science in Brazil. Part 1: A macro-level comparative study. **Scientometrics**, v. 67, n. 1, p. 67-86, 2006.

GLÄNZEL, Wolfgang; SCHUBERT, András. *Analyzing scientific networks through co-authorship*. In: MOED, Henk, F. *et al* (Eds.). **Handbook of Quantitative Science and Technology Research**. New York: Springer, 2005, p. 257-276.

GOLDSTEIN, Harvey A.. *What we know and what we don't know about the regional economic impacts of universities*. In: VARGA, Attila. (Ed.). **Universities, knowledge transfer and regional development: geography, entrepreneurship and policy**. Cheltenham: Edward Elgar, 2009, p. 11-35.

GOYAL, Sanjeev. **Connections: an introduction to the economics of networks**. Princeton: Princeton University Press, 2007.

GROSSETTI, Michel *et al.* *The geographical deconcentration of scientific activities (1987-2010)*. In: **17th International Conference of Science and Technology Indicators**, p. 348-356, 2012.

HADDAD, Paulo R.. Medidas de localização e de especialização. In: HADDAD, Paulo R. (Org.). **Economia regional: teorias e métodos de análise**. Fortaleza: Banco do Nordeste, 1989, p. 225-247.

HELENE, André Frazão; RIBEIRO, Pedro Leite. *Brazilian scientific production, financial support, established investigators and doctoral graduates*. **Scientometrics**, v. 89, p. 677-686, 2011.

HENDERSON, Rebecca *et al.* *Universities as a source of commercial technology: a detailed analysis of university patenting, 1965-1988*. **The Review of Economic and Statistics**, v. 80, n. 1, p. 119-127, 1998.

HILBE, Joseph M.. **Negative Binomial Regression**. Cambridge: Cambridge University Press, 2011.

HOEKMAN, Jarno *et al.* *Acquisition of European research funds and its effect on international scientific collaboration*. **Journal of Economic Geography**, p. 1-30, 2012.

HOEKMAN, Jarno *et al.* *Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe*. **Research Policy**, v. 39, p. 662-673, 2010.

HOEKMAN, Jarno *et al.* *The geography of collaborative knowledge production in Europe*. **The Annals of Regional Science**, v. 43, p. 721-738, 2009.

HOLLANDERS, Hugo; SOETE, Luc. *The growing role of knowledge in the global economy*. In: UNESCO (Ed.). **Science report 2010**. Paris: Unesco Publishing, 2010, p. 1-27.

JACKSON, Matthew O. **Social and Economic Networks**. New Jersey: Princeton University Press, 2008.

JAFFE, A. B. *et al.* *Geographic localization of knowledge spillovers as evidenced by patent citations*. **Quarterly Journal of Economics**, v. 108, n. 3, p. 577-598.

JONES, Benjamin, F. *et al.* *Multi-university research teams: shifting impact, geography, and stratification in science*. **Science**, v. 322, p. 1259-1262, 2008.

KATZ, J. Sylvan. *Geographical Proximity and Scientific Collaboration*. **Scientometrics**, v. 31, n. 1, p. 31-463, 1994.

KATZ, J. Sylvan.; MARTIN, Ben R. *What is research collaboration?* **Research Policy**, v. 26, p. 1-18, 1997.

LANE, Julia. *Let's make science metrics more scientific*. **Nature**, v. 454, n. 25, p. 488-489, 2010.

LEITE, Paula *et al.* *A new indicator for international visibility: exploring Brazilian scientific community*. **Scientometrics**, v. 88, p. 311-319, 2011.

LETA, Jacqueline. Indicadores de desempenho, ciência brasileira e a cobertura das bases informacionais. **Revista USP**, v. 89, p. 72-77, 2011.

LETA, Jacqueline; CHAIMOVICH, Hernan. *Recognition and international collaboration: the Brazilian case*. **Scientometrics**, v. 53, n. 3, p. 325-335, 2002.

LETA, Jacqueline *et al.* *Science in Brazil. Part 2: Sectoral and institucional research profiles*. **Scientometrics**, v. 67, n. 1, p. 87-105, 2006.

LEYDESDORFF, Loet; PERSSON, Olle. *Mapping the geography of science: distribution patterns and networks of relations among cities and institutes*. **Journal of the American Society for Information Science and Technology**, v. 61, n. 8, p. 1622-1634, 2010.

LIANG, Liming; ZHU, Ling. *Major factors affecting China's inter-regional research collaboration: regional scientific productivity and geographical proximity*. **Scientometrics**, v. 55, n. 2, p. 287-316, 2002.

LONG, J. Scott; FREESE, Jeremy. *Regression models for categorical dependent variables using Stata*. College Station: Stata Press, 2001.

JONES *et al.* *Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science*. **Science**, v. 322, p. 1259-1262, 2008.

MAGGIONI, Mario A. *et al.* *Space versus networks in the geography of innovation: a European analysis*. **Papers in Regional Science**, v. 86, n. 3, p. 471-494, 2007.

MAGGIONI, Mario A.; UBERTI, Teodora E. *Networks and geography in the economics of knowledge flows*. **Quality & Quantity**, v. 45, p. 1031-1051, 2011.

MALI, Franc *et al.* *Co-authorship trends and collaboration patterns in the Slovenian sociological community*. **Corvinus Journal Of Sociology And Social Policy**, v. 1, n. 2, p. 29-50, 2010.

MALI, Franc *et al.* *Dynamic scientific co-suthorship networks*. In: SCHARNHORST, Andrea *et al* (Eds.). **Models of science dynamics**. Berlin: Springer-Verlag, 2012, p. 195-232.

MASSARD, Nadine. *Networks and geography in the economics of knowledge flows: a comment*. **Quality & Quantity**, v. 45, p. 1053-1057, 2011.

MATTHIESSEN, Christian W. *et al.* *World cities of scientific knowledge: systems, networks and potential dynamics. An analysis based on bibliometric indicators*. **Urban Studies**, v. 47, n. 9, p. 1879-1897, 2010.

MENA-CHALCO, Jesús. P.; CESAR JUNIOR, R. M. *ScriptLattes: an open-source knowledge extraction system from the Lattes platform*. **Journal of the Brazilian Computer Society**, v. 15, p. 31-39, 2009.

MENA-CHALCO, Jesús P. *et al.* *Brazilian bibliometric co-authorship networks*. **Journal of the American Society for Information Science and Technology**, 2013 (in press).

MICHELS, Carolinet *al.* ***Performance and structures of the German science system 2012***. Berlin: ExpertenkommissionForschung und Innovation, 2013.

NSB – NATIONAL SCIENCE BOARD. ***Science and engineering Indicators 2012***.Arlington, VA: NSB, 2012. Disponível em: <http://www.nsf.gov/statistics/seind12/>>.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATIONAND DEVELOPMENT. ***Measuring innovation: a new perspective***. OECD Publishing, 2010.

OKUBO Yoshiki; ZITT, Michel. *Searching for research integration across europe: a closer look at international and inter-regional collaboration in France*. **Science and Public Policy**, v. 31, p. 213–226, 2004.

OSBORNE, Jason W.; HOLLAND, Abigail. *What is authorship, and what should it be? a survey of prominent guidelines for determining authorship in scientific publications*. **Practical Assessment, Research & Evaluation**, v. 14, n. 15, 2009.

PAN, Raj Kumar; KASKI , Kimmo; FORTUNATO, Santo. *World citation and collaboration networks: uncovering the role of geography in science*. **Scientific Reports**, v. 2, n. 902, 2012.

PHELPS, Corey *et al.* *Knowledge, networks, and knowledge networks: a review and research agenda*. **Journal of Management**, v. 38, n. 4, p. 1115-1166, 2012.

PONDS, Roderik *et al.* *Internationalization and regional embedding of scientific research in the Netherlands*. In: VARGA, Attila. (Ed.). ***Universities, knowledge transfer and regional development: geography, entrepreneurship and policy***. Cheltenham: Edward Elgar, 2009, p. 109-137.

PONDS, Roderik *et al.* *The geographical and institucional proximity of research collaboration*. **Papers in Regional Science**, v. 86, n. 3, p. 423-444, 2007.

REGALADO, Antonio. *Brazilian science: riding a gusher*. **Science**, v. 330, p. 1306-1312, 2010.

ROY, John, R.. ***Spatial interaction modeling***. Berlin: Springer, 2004.

RS – ROYAL SOCIETY. ***Knowledge, networks and nations: global scientific collaboration in the 21st century***. London: The RS, 2011.

SCHERNGELL, Thomas; BARBER, Michael J. *Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU Framework Programme*. **Annals of Regional Science**, v. 46, p. 247-266, 2011.

SCHERNGELL, Thomas; BARBER, Michael J. *Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework Programme*. **Papers in Regional Science**, v. 88, n. 3, p. 531-547, 2009.

SCHERNGELL, Thomas; LATA, Rafael. *Towards an integrated European research area? Findings from eigenvector spatially filtered spatial interaction models using European Framework Programme data*. **Papers in Regional Science**, 2012.

SCHERNGELL, Thomas; HU, Yuanjia. *Collaborative knowledge production in China: regional evidence from a gravity model approach*. **Regional Studies**, v. 45, n. 6, p. 755-772, 2011.

SCHUMPETER, Joseph, A.. **Capitalismo, socialismo e democracia**. Rio de Janeiro: Editora Fundo de Cultura, 1961.

SCIELO Brasil. Disponível em: <http://www.scielo.br/>. Acesso em: 14 out. 2012.

SEN, Ashish K.; SMITH, Tony E. **Gravity Models of Spatial Interaction Behavior**. New York: Springer, 1995.

SINGH, Jasjit; FLEMING, Lee. *Lone inventors as sources of breakthroughs: myth or reality?* **Management Science**, v. 56, p. 41-56, 2010.

TER WAL, Anne L. J. *Networks and geography in the economics of knowledge flows: a commentary*. **Quality & Quantity**, v. 45, p. 1059-1063, 2011.

TER WAL, Anne L. J.; BOSCHMA, Ron A. *Applying social network analysis in economic geography: framing some key analytic issues*. **Annals of Regional Science**, v. 43, p. 739- 756, 2009.

TIJSEN, Robert J. *Are we moving towards an integrated European Research Area? Some macro-level bibliometric perspectives*. In: KRETSCHMER, H.; HAVEMANN, F. (Eds.). **Proceedings of WIS 2008**. Berlin, 2008, p. 1-8.

UNESCO - UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. **Science report 2010**. Paris: Unesco Publishing, 2010. Disponível em: <http://www.unesco.org/new/en/natural-sciences/science-technology/prospective-studies/unesco-science-report/>.

VANZ, Samile Andréa de Souza. **As redes de colaboração científica no Brasil (2004-2006)**. Porto Alegre, 2009. 204 p. Tese (Doutorado) — Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul.

VARGA, Attila; PARAG, Andrea. *Academic knowledge transfers and the structure of international research networks*. In: VARGA, Attila. (Ed.). **Universities, knowledge transfer and regional development: geography, entrepreneurship and policy**. Cheltenham: Edward Elgar, 2009, p. 138-159.

WALTMAN, Ludo *et al.* *Globalisation of science in kilometres*. **Journal of Informetrics**, v. 5, p. 574-582, 2011.

WANG, Yan *et al.* *Scientific collaboration in China as reflected in co-authorship*. **Scientometrics**, v. 62, n. 2, p. 183-198.

WINKELMANN, Rainer. *Econometric analysis of count data*. Berlin: Springer, 2008.

WOOLDRIDGE, Jeffrey M. *Econometric analysis of cross section and panel data*. Massachusetts: The MIT Press, 2002.

ZHOU, Ping *et al.* *Regional analysis on Chinese scientific output*. **Scientometrics**, v. 81, n. 3, p. 839-857, 2009.

APÊNDICE A. MÓDULOS DE PRODUÇÃO CIENTÍFICA NO CL

Produção Científica

Produção Bibliográfica

Artigos completos publicados em periódicos

Artigos aceitos para publicação

Livros publicados, organizados ou edições

Capítulos de livros publicados

Texto em jornais ou revistas

Trabalhos publicados em anais de congressos

Apresentação de trabalho e palestra

Partitura musical

Tradução

Prefácio, posfácio

Outras produções bibliográficas

Produção Técnica***Produção Artística/Cultural***

Fonte: CNPq (2012).

APÊNDICE B. GRANDES ÁREAS DO CONHECIMENTO SEGUNDO O CNPQ

Ciências Humanas

Filosofia	Sociologia
Antropologia	Arqueologia
História	Geografia
Psicologia	Educação
Ciência política	Teologia

Ciências Agrárias

Agronomia	Recursos florestais e engenharia florestal
Engenharia agrícola	Ciência e tecnologia de alimentos
Medicina veterinária	Recursos pesqueiros e engenharia de pesca
Zootecnia	

Engenharias

Engenharia de minas	Engenharia de materiais e metalúrgica
Engenharia elétrica	Engenharia mecânica
Engenharia química	Engenharia sanitária
Engenharia de produção	Engenharia nuclear
Engenharia de transportes	Engenharia naval e oceânica
Engenharia aeroespacial	Engenharia biomédica
Engenharia civil	

Linguística, Letras e Artes

Linguística	Artes
Letras	

Continuação

Ciências de Saúde

Odontologia

Enfermagem

Saúde coletiva

Educação física

Medicina

Farmácia

Nutrição

Fonoaudiologia

Fisioterapia e terapia ocupacional

Ciências Exatas e da Terra

Probabilidade e estatística

Astronomia

Química

Oceanografia

Matemática

Ciência da computação

Física

Geociências

Ciências Biológicas

Genética

Zoologia

Morfologia

Bioquímica

Farmacologia

Microbiologia

Biologia geral

Botânica

Ecologia

Fisiologia

Biofísica

Imunologia

Parasitologia

Ciências Sociais Aplicadas

Administração

Arquitetura e urbanismo

Demografia

Museologia

Economia

Planejamento urbano e regional

Ciência da informação

Comunicação

Continuação

Serviço social	Economia doméstica
Desenho industrial	Turismo
Direito	
Outros	
Administração rural	Carreira militar
Carreira religiosa	Ciências
Biomedicina	Ciências atuariais
Ciências sociais	Decoração
Desenho de moda	Desenho de projetos
Diplomacia	Engenharia de agrimensura
Engenharia cartográfica	Engenharia de armamentos
Engenharia mecatrônica	Engenharia têxtil
Estudos sociais	História natural
Química industrial	Relações internacionais
Relações públicas	Secretariado executivo
Administração hospitalar	

Fonte: CNPQ (2012).

APÊNDICE C. MAPAS DA PRODUÇÃO CIENTÍFICA MUNICIPAL NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO

Mapa C1. Produção científica municipal no Brasil: Engenharias

1992-1994

2007-2009

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C2. Produção científica municipal no Brasil: Ciências Humanas**1992-1994** **2007-2009**

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C3. Produção científica municipal no Brasil: Ciências Sociais Aplicadas
1992-1994

2007-2009

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C4. Produção científica municipal no Brasil: Linguística, Letras e Artes**1992-1994** **2007-2009**

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C5. Produção científica municipal no Brasil: Ciências Agrárias
1992-1994

2007-2009

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C6. Produção científica municipal no Brasil: Ciências Exatas e da Terra**1992-1994** **2007-2009**

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C7. Produção científica municipal no Brasil: Ciências da Saúde
1992-1994

2007-2009

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C8. Produção científica municipal no Brasil: Ciências Biológicas**1992-1994** **2007-2009**

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Mapa C9. Produção científica municipal no Brasil: Outras áreas**1992-1994** **2007-2009**

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

APÊNDICE D. EVOLUÇÃO DOS PRINCIPAIS MUNICÍPIOS ASSOCIADOS À PRODUÇÃO CIENTÍFICA NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO

Tabela D1 – Evolução dos principais municípios associados à produção científica no Brasil: Engenharias

<i>Total</i>		<i>Total</i>		<i>Total</i>	
1992-1994		1995-1997		1998-2000	
Rio de Janeiro/RJ	9.615	Rio de Janeiro/RJ	16.009	Rio de Janeiro/RJ	25.659
São Paulo/SP	7.172	São Paulo/SP	13.164	São Paulo/SP	20.537
Campinas/SP	4.985	São Carlos/SP	8.826	São Carlos/SP	13.531
São Carlos/SP	4.171	Campinas/SP	8.270	Campinas/SP	12.487
Florianópolis/SC	3.034	Florianópolis/SC	6.543	Florianópolis/SC	11.089
Belo Horizonte/MG	2.636	Porto Alegre/RS	5.497	Porto Alegre/RS	8.672
Porto Alegre/RS	2.578	Belo Horizonte/MG	5.262	Belo Horizonte/MG	8.447
S. J. dos Campos/SP	2.035	S. J. dos Campos/SP	4.053	S. J. dos Campos/SP	7.122
Curitiba/PR	1.436	Curitiba/PR	2.806	Curitiba/PR	4.943
Recife/PE	1.046	Brasília/DF	2.280	Recife/PE	3.821
2001-2003		2004-2006		2007-2009	
Rio de Janeiro/RJ	37.353	Rio de Janeiro/RJ	46.922	Rio de Janeiro/RJ	46.204
São Paulo/SP	29.304	São Paulo/SP	39.753	São Paulo/SP	38.525
São Carlos/SP	18.745	São Carlos/SP	25.064	São Carlos/SP	25.997
Campinas/SP	15.806	Campinas/SP	22.118	Campinas/SP	21.462
Florianópolis/SC	14.428	Florianópolis/SC	17.817	Florianópolis/SC	21.341
Porto Alegre/RS	12.331	Belo Horizonte/MG	17.662	Belo Horizonte/MG	19.368
Belo Horizonte/MG	12.093	Porto Alegre/RS	17.098	Porto Alegre/RS	18.153
S. J. dos Campos/SP	10.020	S. J. dos Campos/SP	14.019	Curitiba/PR	15.247
Curitiba/PR	8.924	Curitiba/PR	13.592	S. J. dos Campos/SP	14.429
Recife/PE	7.031	Campina Grande/PB	10.646	Recife/PE	12.607

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D2 – Evolução dos principais municípios associados à produção científica no Brasil: Ciências Humanas

<i>Total</i>		<i>Total</i>		<i>Total</i>	
1992-1994		1995-1997		1998-2000	
São Paulo/SP	7.816	São Paulo/SP	12.710	São Paulo/SP	18.306
Rio de Janeiro/RJ	6.020	Rio de Janeiro/RJ	10.608	Rio de Janeiro/RJ	16.841
Porto Alegre/RS	2.880	Porto Alegre/RS	4.824	Porto Alegre/RS	7.944
Campinas/SP	2.301	Belo Horizonte/MG	4.041	Belo Horizonte/MG	6.345
Belo Horizonte/MG	2.189	Campinas/SP	3.790	Campinas/SP	5.450
Brasília/DF	1.956	Brasília/DF	3.299	Brasília/DF	5.378
Salvador/BA	1.190	Florianópolis/SC	2.306	Florianópolis/SC	3.894
Florianópolis/SC	1.099	Salvador/BA	1.912	Recife/PE	3.381
Curitiba/PR	1.019	Recife/PE	1.888	Curitiba/PR	3.226
Niterói/RJ	959	Curitiba/PR	1.779	Salvador/BA	3.158
2001-2003		2004-2006		2007-2009	
São Paulo/SP	26.194	São Paulo/SP	36.328	São Paulo/SP	41.531
Rio de Janeiro/RJ	24.972	Rio de Janeiro/RJ	32.877	Rio de Janeiro/RJ	37.936
Porto Alegre/RS	12.241	Porto Alegre/RS	18.463	Porto Alegre/RS	21.552
Belo Horizonte/MG	9.575	Belo Horizonte/MG	13.260	Belo Horizonte/MG	17.927
Brasília/DF	8.962	Brasília/DF	12.609	Brasília/DF	13.896
Campinas/SP	7.346	Curitiba/PR	11.031	Curitiba/PR	12.978
Curitiba/PR	6.608	Campinas/SP	10.567	Florianópolis/SC	12.182
Florianópolis/SC	6.435	Florianópolis/SC	9.346	Campinas/SP	11.547
Salvador/BA	5.083	Recife/PE	7.614	Recife/PE	10.139
Recife/PE	5.074	Salvador/BA	7.181	Santa Maria/RS	9.784

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D3 – Evolução dos principais municípios associados à produção científica no Brasil: Ciências Sociais Aplicadas

Total		Total	Total		
1992-1994		1995-1997	1998-2000		
São Paulo/SP	6.226	São Paulo/SP	10.408	São Paulo/SP	16.011
Rio de Janeiro/RJ	4.008	Rio de Janeiro/RJ	7.005	Rio de Janeiro/RJ	11.545
Porto Alegre/RS	1.881	Porto Alegre/RS	3.020	Porto Alegre/RS	5.468
Brasília/DF	1.677	Brasília/DF	2.946	Brasília/DF	5.115
Campinas/SP	1.414	Belo Horizonte/MG	2.794	Belo Horizonte/MG	4.841
Belo Horizonte/MG	1.112	Campinas/SP	2.292	Florianópolis/SC	3.624
Salvador/BA	931	Florianópolis/SC	2.048	Campinas/SP	3.436
Curitiba/PR	842	Salvador/BA	1.667	Curitiba/PR	3.080
Florianópolis/SC	776	Curitiba/PR	1.651	Recife/PE	2.818
Recife/PE	685	Recife/PE	1.429	Salvador/BA	2.751
2001-2003		2004-2006	2007-2009		
São Paulo/SP	25.747	São Paulo/SP	40.354	São Paulo/SP	42.740
Rio de Janeiro/RJ	17.113	Rio de Janeiro/RJ	23.509	Rio de Janeiro/RJ	25.893
Porto Alegre/RS	9.082	Porto Alegre/RS	15.207	Belo Horizonte/MG	17.213
Brasília/DF	8.356	Belo Horizonte/MG	12.813	Porto Alegre/RS	16.157
Belo Horizonte/MG	7.960	Brasília/DF	11.778	Curitiba/PR	13.257
Curitiba/PR	6.321	Curitiba/PR	10.403	Florianópolis/SC	13.244
Florianópolis/SC	5.889	Florianópolis/SC	8.772	Brasília/DF	13.168
Campinas/SP	4.890	Campinas/SP	7.163	Recife/PE	9.254
Recife/PE	4.835	Recife/PE	6.800	Salvador/BA	7.359
Salvador/BA	4.574	Salvador/BA	6.512	Campinas/SP	7.160

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D4 – Evolução dos principais municípios associados à produção científica no Brasil: Linguística, Letras e Artes

Total		Total	Total		
1992-1994		1995-1997	1998-2000		
São Paulo/SP	2.766	São Paulo/SP	3.944	São Paulo/SP	5.254
Rio de Janeiro/RJ	1.606	Rio de Janeiro/RJ	3.084	Rio de Janeiro/RJ	4.378
Belo Horizonte/MG	782	Belo Horizonte/MG	1.406	Belo Horizonte/MG	2.222
Porto Alegre/RS	780	Porto Alegre/RS	1.248	Porto Alegre/RS	2.037
Campinas/SP	675	Campinas/SP	1.134	Campinas/SP	1.479
Florianópolis/SC	399	Florianópolis/SC	690	Florianópolis/SC	1.101
Brasília/DF	379	Salvador/BA	651	Salvador/BA	994
Salvador/BA	360	Brasília/DF	615	Brasília/DF	918
Niterói/RJ	292	Niterói/RJ	580	Niterói/RJ	847
Curitiba/PR	262	Curitiba/PR	451	Recife/PE	754
2001-2003		2004-2006	2007-2009		
São Paulo/SP	7.312	São Paulo/SP	9.781	São Paulo/SP	11.429
Rio de Janeiro/RJ	6.365	Rio de Janeiro/RJ	8.027	Rio de Janeiro/RJ	8.563
Porto Alegre/RS	2.999	Belo Horizonte/MG	4.005	Belo Horizonte/MG	4.818
Belo Horizonte/MG	2.998	Porto Alegre/RS	3.629	Porto Alegre/RS	4.128
Campinas/SP	1.966	Florianópolis/SC	2.711	Florianópolis/SC	3.826
Florianópolis/SC	1.837	Campinas/SP	2.622	Campinas/SP	2.979
Brasília/DF	1.613	Brasília/DF	2.082	Salvador/BA	2.424
Salvador/BA	1.510	Curitiba/PR	1.954	Curitiba/PR	2.414
Curitiba/PR	1.261	Recife/PE	1.906	Brasília/DF	2.378
Niterói/RJ	1.101	Salvador/BA	1.898	Recife/PE	2.040

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D5 – Evolução dos principais municípios associados à produção científica no Brasil: Ciências Agrárias

	Total		Total	Total
	1992-1994	1995-1997	1998-2000	
Viçosa/MG	4.248	Viçosa/MG	8.413	Viçosa/MG
Campinas/SP	3.323	Piracicaba/SP	5.162	Campinas/SP
Piracicaba/SP	2.762	Campinas/SP	5.103	Piracicaba/SP
São Paulo/SP	2.688	São Paulo/SP	4.931	Lavras/MG
Botucatu/SP	2.237	Lavras/MG	4.761	São Paulo/SP
Lavras/MG	2.205	Brasília/DF	3.524	Belo Horizonte/MG
Brasília/DF	2.182	Jaboticabal/SP	3.479	Brasília/DF
Jaboticabal/SP	2.049	Porto Alegre/RS	3.117	Botucatu/SP
Londrina/PR	1.510	Botucatu/SP	3.078	Porto Alegre/RS
Santa Maria/RS	1.494	Belo Horizonte/MG	3.022	Jaboticabal/SP
	2001-2003	2004-2006	2007-2009	
Viçosa/MG	26.120	Viçosa/MG	32.801	Viçosa/MG
Lavras/MG	14.375	São Paulo/SP	16.022	Lavras/MG
São Paulo/SP	13.105	Lavras/MG	15.742	São Paulo/SP
Piracicaba/SP	11.884	Piracicaba/SP	15.656	Santa Maria/RS
Campinas/SP	11.689	Campinas/SP	15.144	Piracicaba/SP
Brasília/DF	10.595	Santa Maria/RS	14.048	Jaboticabal/SP
Jaboticabal/SP	10.033	Jaboticabal/SP	13.929	Botucatu/SP
Porto Alegre/RS	9.535	Brasília/DF	13.370	Campinas/SP
Santa Maria/RS	9.363	Botucatu/SP	12.327	Recife/PE
Pelotas/RS	9.310	Porto Alegre/RS	11.955	Brasília/DF

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D6 – Evolução dos principais municípios associados à produção científica no Brasil: *Ciências Exatas e da Terra*

	Total	Total	Total
1992-1994		1995-1997	1998-2000
São Paulo/SP	11.486	Rio de Janeiro/RJ	Rio de Janeiro/RJ
Rio de Janeiro/RJ	10.707	São Paulo/SP	São Paulo/SP
Campinas/SP	6.436	São Carlos/SP	São Carlos/SP
São Carlos/SP	5.095	Campinas/SP	Campinas/SP
Porto Alegre/RS	4.037	Porto Alegre/RS	Porto Alegre/RS
S. J. dos Campos/SP	3.411	S. J. dos Campos/SP	S. J. dos Campos/SP
Belo Horizonte/MG	2.730	Belo Horizonte/MG	Belo Horizonte/MG
Recife/PE	2.559	Recife/PE	Recife/PE
Florianópolis/SC	1.836	Florianópolis/SC	Florianópolis/SC
Fortaleza/CE	1.659	Curitiba/PR	Curitiba/PR
2001-2003		2004-2006	2007-2009
Rio de Janeiro/RJ	40.398	Rio de Janeiro/RJ	Rio de Janeiro/RJ
São Paulo/SP	36.101	São Paulo/SP	São Paulo/SP
São Carlos/SP	19.228	São Carlos/SP	São Carlos/SP
Campinas/SP	18.308	Campinas/SP	Campinas/SP
Porto Alegre/RS	18.218	Porto Alegre/RS	Porto Alegre/RS
Belo Horizonte/MG	11.688	S. J. dos Campos/SP	Belo Horizonte/MG
S. J. dos Campos/SP	11.561	Belo Horizonte/MG	S. J. dos Campos/SP
Curitiba/PR	10.238	Recife/PE	Recife/PE
Recife/PE	10.134	Curitiba/PR	Curitiba/PR
Brasília/DF	9.176	Florianópolis/SC	Florianópolis/SC

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D7 – Evolução dos principais municípios associados à produção científica no Brasil: Ciências da Saúde

Total		Total	Total
1992-1994		1995-1997	1998-2000
São Paulo/SP	37.178	São Paulo/SP	São Paulo/SP
Rio de Janeiro/RJ	8.428	Rio de Janeiro/RJ	Rio de Janeiro/RJ
Ribeirão Preto/SP	7.190	Porto Alegre/RS	Porto Alegre/RS
Porto Alegre/RS	6.803	Ribeirão Preto/SP	Ribeirão Preto/SP
Belo Horizonte/MG	4.680	Campinas/SP	Belo Horizonte/MG
Campinas/SP	4.244	Belo Horizonte/MG	Campinas/SP
Salvador/BA	2.059	Salvador/BA	Curitiba/PR
Recife/PE	1.881	Recife/PE	Salvador/BA
Botucatu/SP	1.810	Curitiba/PR	Recife/PE
Curitiba/PR	1.658	Botucatu/SP	Fortaleza/CE
2001-2003		2004-2006	2007-2009
São Paulo/SP	108.863	São Paulo/SP	São Paulo/SP
Rio de Janeiro/RJ	37.663	Rio de Janeiro/RJ	Rio de Janeiro/RJ
Porto Alegre/RS	26.112	Porto Alegre/RS	Porto Alegre/RS
Ribeirão Preto/SP	20.250	Ribeirão Preto/SP	Ribeirão Preto/SP
Belo Horizonte/MG	17.744	Belo Horizonte/MG	Belo Horizonte/MG
Campinas/SP	15.165	Campinas/SP	Campinas/SP
Curitiba/PR	10.923	Curitiba/PR	Curitiba/PR
Salvador/BA	9.615	Fortaleza/CE	Fortaleza/CE
Recife/PE	8.160	Salvador/BA	Florianópolis/SC
Fortaleza/CE	8.039	Recife/PE	Recife/PE

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D8 – Evolução dos principais municípios associados à produção científica no Brasil: Ciências Biológicas

Total		Total	Total
1992-1994		1995-1997	1998-2000
São Paulo/SP	14.529	São Paulo/SP	22.404
Rio de Janeiro/RJ	10.100	Rio de Janeiro/RJ	15.675
Ribeirão Preto/SP	3.712	Porto Alegre/RS	6.096
Belo Horizonte/MG	3.529	Belo Horizonte/MG	5.777
Porto Alegre/RS	2.993	Campinas/SP	4.813
Campinas/SP	2.936	Ribeirão Preto/SP	4.630
Brasília/DF	2.193	Brasília/DF	3.584
Recife/PE	1.713	Recife/PE	3.311
Curitiba/PR	1.421	Viçosa/MG	3.080
Botucatu/SP	1.374	São Carlos/SP	2.490
2001-2003		2004-2006	2007-2009
São Paulo/SP	47.118	São Paulo/SP	67.678
Rio de Janeiro/RJ	35.543	Rio de Janeiro/RJ	47.465
Porto Alegre/RS	17.662	Porto Alegre/RS	25.581
Belo Horizonte/MG	14.143	Belo Horizonte/MG	19.945
Campinas/SP	11.624	Campinas/SP	17.483
Ribeirão Preto/SP	10.814	Ribeirão Preto/SP	17.070
Brasília/DF	10.713	Brasília/DF	14.922
Recife/PE	8.618	Recife/PE	12.281
Viçosa/MG	8.529	Fortaleza/CE	11.232
Fortaleza/CE	7.365	Curitiba/PR	10.821
</			

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

Tabela D9 – Evolução dos principais municípios associados à produção científica no Brasil: Outras áreas

Total		Total	Total
1992-1994		1995-1997	1998-2000
São Paulo/SP	571	São Paulo/SP	918
Rio de Janeiro/RJ	361	Rio de Janeiro/RJ	612
Porto Alegre/RS	303	Porto Alegre/RS	552
Brasília/DF	160	Brasília/DF	308
Campinas/SP	113	Belo Horizonte/MG	233
Belo Horizonte/MG	103	Campinas/SP	230
Recife/PE	77	Curitiba/PR	162
Curitiba/PR	76	Maringá/PR	135
Fortaleza/CE	75	Recife/PE	131
Florianópolis/SC	47	Fortaleza/CE	99
2001-2003		2004-2006	2007-2009
São Paulo/SP	2.192	São Paulo/SP	3.193
Rio de Janeiro/RJ	1.778	Rio de Janeiro/RJ	2.686
Porto Alegre/RS	1.141	Porto Alegre/RS	1.310
Brasília/DF	907	Brasília/DF	1.251
Curitiba/PR	622	Curitiba/PR	878
Campinas/SP	582	Fortaleza/CE	819
Recife/PE	472	Campinas/SP	816
Belo Horizonte/MG	402	Belo Horizonte/MG	771
Fortaleza/CE	350	Recife/PE	676
Maringá/PR	312	Piracicaba/SP	531

Obs: os valores correspondem ao somatório da participação em publicações científicas por parte dos pesquisadores localizados geograficamente. Como a participação é contabilizada de forma absoluta (unitária), os valores não correspondem ao total da publicação efetiva, uma vez que no caso de publicações em coautoria, foi atribuída uma unidade de participação para cada autor.

APÊNDICE E. MAPAS DA COLABORAÇÃO CIENTÍFICA INTERMUNICIPAL NO BRASIL EM TRIÊNIOS SELECIONADOS: POR ÁREA DO CONHECIMENTO

Engenharias

Rede de Colaboração Científica (1992-1994)

Graus Municipais na Rede (1992-1994)

Rede de Colaboração Científica (2007-2009)

Graus Municipais na Rede (2007-2009)

*Ciências Humanas**Rede de Colaboração Científica (1992-1994)* *Graus Municipais na Rede (1992-1994)* *Rede de Colaboração Científica (2007-2009)* *Graus Municipais na Rede (2007-2009)*

Ciências Sociais Aplicadas

Rede de Colaboração Científica (1992-1994)

Graus Municipais na Rede (1992-1994)

Rede de Colaboração Científica (2007-2009)

Graus Municipais na Rede (2007-2009)

*Linguística, Letras e Artes**Rede de Colaboração Científica (1992-1994)* *Graus Municipais na Rede (1992-1994)* *Rede de Colaboração Científica (2007-2009)* *Graus Municipais na Rede (2007-2009)*

Ciências Agrárias

Rede de Colaboração Científica (1992-1994)

Graus Municipais na Rede (1992-1994)

Rede de Colaboração Científica (2007-2009)

Graus Municipais na Rede (2007-2009)

*Ciências Exatas e da Terra****Rede de Colaboração Científica (1992-1994)*** ***Graus Municipais na Rede (1992-1994)*** ***Rede de Colaboração Científica (2007-2009)*** ***Graus Municipais na Rede (2007-2009)***

*Ciências da Saúde****Rede de Colaboração Científica (1992-1994)*** ***Graus Municipais na Rede (1992-1994)*** ***Rede de Colaboração Científica (2007-2009)*** ***Graus Municipais na Rede (2007-2009)***

Ciências Biológicas

Rede de Colaboração Científica (1992-1994)

Graus Municipais na Rede (1992-1994)

Rede de Colaboração Científica (2007-2009)

Graus Municipais na Rede (2007-2009)

APÊNDICE F. PRINCIPAIS LIGAÇÕES DE COLABORAÇÃO CIENTÍFICA NO BRASIL NO TRIÊNIO 2007-2009: POR ÁREA DO CONHECIMENTO

	<i>Total</i>		<i>Total</i>
<i>Engenharias</i>			
Rio de Janeiro/RJ	67.030	Araraquara/SP – São Carlos/SP	9.838
São Paulo/SP	51.931	Niterói/RJ – Rio de Janeiro/RJ	5.896
São Carlos/SP	34.956	S. J. dos Campos/SP – São Paulo/SP	5.537
Belo Horizonte/MG	29.292	São Carlos/SP – São Paulo/SP	4.935
Campinas/SP	28.655	Campinas/SP – São Paulo/SP	4.696
<i>Ciências Humanas</i>			
Rio de Janeiro/RJ	18.103	Niterói/RJ – Rio de Janeiro/RJ	2.596
São Paulo/SP	16.446	Rio de Janeiro/RJ – São Paulo/SP	2.343
Porto Alegre/RS	13.031	Campinas/SP – São Paulo/SP	2.181
Belo Horizonte/MG	6.514	Santa Maria/RS – Porto Alegre/RS	1.484
Brasília/DF	6.166	São Carlos/SP – São Paulo/SP	1.319
<i>Ciências Sociais Aplicadas</i>			
São Paulo/SP	21.538	Ribeirão Preto/SP – São Paulo/SP	2.230
Florianópolis/SC	14.058	Rio de Janeiro/RJ – São Paulo/SP	1.644
Belo Horizonte/MG	10.197	Santa Maria/RS – Porto Alegre/RS	1.557
Porto Alegre/RS	9.873	Campinas/SP – São Paulo/SP	1.545
Rio de Janeiro/RJ	7.704	João Pessoa/PB – Recife/PE	982
<i>Linguística, Letras e Artes</i>			
São Paulo/SP	1.635	Curitiba/PR – Florianópolis/SC	374
Rio de Janeiro/RJ	723	Niterói/RJ – Rio de Janeiro/RJ	240
Porto Alegre/RS	606	Campinas/SP – São Paulo/SP	159
Belo Horizonte/MG	592	Florianópolis/SC – São Paulo/SP	116
Florianópolis/SC	478	Rio de Janeiro/RJ – São Paulo/SP	97

*Continuação**Ciências Agrárias*

Viçosa/MG	105.844	Pelotas/RS – Capão do Leão/RS	23.598
São Paulo/SP	51.827	Pirassununga/SP – São Paulo/SP	12.736
Santa Maria/RS	51.086	Lavras/MG – Viçosa/MG	10.562
Lavras/MG	48.120	Seropédica/RJ – Rio de Janeiro/RJ	9.426
Botucatu/SP	45.928	Alegre/ES – Viçosa/MG	8.885

Ciências Exatas e da Terra

Rio de Janeiro/RJ	100.202	Niterói/RJ – Rio de Janeiro/RJ	23.422
São Paulo/SP	89.977	Araraquara/SP – São Carlos/SP	21.439
S. J. dos Campos/SP	73.570	S. J. dos Campos/SP – São Paulo/SP	18.622
São Carlos/SP	62.066	São Carlos/SP – São Paulo/SP	9.900
Porto Alegre/RS	43.570	Rio de Janeiro/RJ – São Paulo/SP	8.485

Ciências da Saúde

São Paulo/SP	781.883	Ribeirão Preto/SP – São Paulo/SP	37.556
Ribeirão Preto/SP	154.847	Rio de Janeiro/RJ – São Paulo/SP	28.787
Rio de Janeiro/RJ	141.640	Campinas/SP – São Paulo/SP	26.846
Porto Alegre/RS	130.612	Porto Alegre/RS – São Paulo/SP	25.823
Belo Horizonte/MG	99.421	Niterói/RJ – Rio de Janeiro/RJ	20.482

Ciências Biológicas

São Paulo/SP	358.113	Seropédica/RJ – Rio de Janeiro/RJ	49.284
Rio de Janeiro/RJ	305.472	Campinas/SP – São Paulo/SP	29.028
Porto Alegre/RS	152.739	Ribeirão Preto/SP – São Paulo/SP	27.408
Belo Horizonte/MG	145.475	Rio de Janeiro/RJ – São Paulo/SP	26.277
Ribeirão Preto/SP	99.466	Niterói/RJ – Rio de Janeiro/RJ	9.693

Fonte: elaboração própria.