

**NEREUS**

Núcleo de Economia Regional e Urbana  
da Universidade de São Paulo  
The University of São Paulo  
Regional and Urban Economics Lab

**PADRÕES DE COLABORAÇÃO CIENTÍFICA NO BRASIL:  
O ESPAÇO IMPORTA?**

Otávio J. G. Sidone  
Eduardo A. Haddad  
Jesús Mena-Chalco

**TD Nereus 09-2013**  
São Paulo  
2013

# Padrões de Colaboração Científica no Brasil: O Espaço Importa?

Otávio J. G. Sidone, Eduardo A. Haddad e Jesús Mena-Chalco

**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 de identificar padrões espaciais da 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, nós coletamos e consolidamos informações sobre as colaborações científicas inter-regionais em termos de redes de coautorias. 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 colaboração científica nos últimos anos com o aumento da participação de autores das regiões cientificamente menos tradicionais. 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.

## 1. Introdução

Pode-se afirmar que as universidades desempenham um papel crucial em um sistema de inovação, por sua capacidade de produzir novos conhecimentos passíveis de serem aproveitados em aplicações economicamente eficientes, e assim, propiciar o fomento da atividade inovativa e o desenvolvimento econômico regional. Dessa maneira, o estudo das relações entre pesquisa científica e inovação, por meio dos processos de geração e difusão do conhecimento, consiste em tema recorrente em economia da inovação e geografia econômica.

Os estudos sobre transferência de conhecimento que procuraram analisar a atuação da geografia na ocorrência e dimensão dos *spillovers* de conhecimento (processo pelo qual o conhecimento acadêmico é compreendido e utilizado pelas firmas do setor privado) encontraram evidências de localização espacial, ou seja, a intensificação da transmissão de conhecimento entre universidades e firmas é favorecida pela proximidade geográfica (JAFPE *et al*, 1993; AUDRESTCH; FELDMAN, 1996; ANSELIN *et al*, 2000; ÁCS *et al*, 2002; VARGA; PARAG, 2009, PONDS *et al*, 2009), o que sugere que a aproximação física pode facilitar a capacidade de acesso das firmas inovativas às fontes

de conhecimento e potencializar as atividades inovativas (MAGGIONI; UBERTI, 2011; TER WAL, 2011). Entretanto, a literatura recente caminha rumo à explicação e interpretação dos *spillovers* por meio da análise das redes entre firmas e universidades e entre as próprias universidades (FRENKEN *et al*, 2009). Isso é de suma importância, uma vez que a compreensão da 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).<sup>1</sup>

As análises espaciais da ciência, inseridas no ramo denominado de cientometria espacial, possuem como objetivo principal o estudo da disposição geográfica dos fluxos de conhecimento (sejam entre indivíduos, organizações ou entidades geográficas), ou seja, das articulações entre os pesquisadores nos processos de geração e difusão do conhecimento. Essa abordagem é fundamental, uma vez que a atividade científica é distribuída de maneira bastante desigual tanto entre países como dentro de seus territórios, onde é comum a evidência de padrões de localização geográfica caracterizados por intensa heterogeneidade espacial.

Embora ainda distante do padrão de crescimento da China, o Brasil adquire crescente relevância no cenário científico internacional, com crescimento da produção bastante superior à média mundial. 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; LETA, 2011; GROSSETTI *et al*, 2012; RS, 2011).

---

<sup>1</sup> São crescentes os esforços no desenvolvimento de técnicas de visualização de mapas geográficos de redes científicas entre cidades e institutos de pesquisa para facilitar a localização dos fluxos de conhecimento (LEYDESDORFF; PERSSON, 2010).

A ciência moderna possui como característica principal o aumento do perfil colaborativo em todas suas áreas, visto que cerca de 70% dos artigos produzidos atualmente no mundo estão associados a autores de diferentes instituições e, entre esses, cerca de 44% é oriundo de esforços colaborativos entre pesquisadores de diferentes países e 56% de colaborações entre pesquisadores em território nacional.<sup>2</sup> Particularmente no caso dos países cientificamente emergentes, tais como China, Brasil e Índia, o crescimento acelerado da produção está diretamente associado à intensificação dos esforços colaborativos entre pesquisadores localizados dentro do território nacional. (GLÄNZEL; SCHUBERT, 2004; RS, 2011). Nesse contexto, 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 dos 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 promover-se 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 de outras regiões (CRUZ; CHAIMOVICH, 2010), bem como de se associar o crescimento da atividade científica ao incremento de sua qualidade (HELENE; RIBEIRO, 2011). 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 difusão do conhecimento e está estreitamente associada a maior qualidade da produção científica.

Contudo, a tentativa de compreensão da importância da geografia na produção do conhecimento, principalmente por seu papel na intermediação das interações entre os

---

<sup>2</sup> São diversos os motivos que os pesquisadores possuem para colaborar. Como exemplos, o compartilhamento do trabalho pode promover a transferência de conhecimentos e habilidades, o aumento da qualidade e visibilidade da pesquisa, reduções do tempo dispensado e da ocorrência de erros, além da obtenção ou ampliação de financiamentos (RS, 2011; VANZ, 2009).

pesquisadores brasileiros em colaborações científicas, ainda é tema inexplorado pela literatura. O entendimento da articulação entre as redes científicas pode auxiliar o processo de formulação de políticas de C&T&I no Brasil, uma vez que o aumento da capacidade de acesso de firmas inovativas às fontes de conhecimento pode ter seu efeito potencializado em termos de atividade inovativas caso essas últimas estejam integradas às redes de produção compartilhada de conhecimento técnico-científico.

Sob essa perspectiva, analisamos como o crescimento da produção e da colaboração científica entre os pesquisadores de todas as áreas do conhecimento ao longo do período entre 1990 e 2010 deu-se no espaço geográfico brasileiro, mais especificamente, ao nível municipal. Adicionalmente, verificamos a importância da distância geográfica como fator impeditivo às interações entre os pesquisadores nas redes de colaborações científicas. É importante destacar que, no nosso entendimento, este consiste no primeiro estudo realizado a partir do tratamento de uma quantidade extremamente grande de dados associados a mais de um milhão de pesquisadores brasileiros cadastrados na Plataforma Lattes.

Além dessa introdução, a estrutura do artigo inicia-se pela descrição detalhada do procedimento de coleta dos dados sobre a colaboração científica no Brasil (seção 2). Em seguida, são apresentadas as principais características da evolução espacial da produção de conhecimento e das redes de colaborações científicas entre os municípios brasileiros (seção 3). A dimensão do papel desempenhado pela distância geográfica na configuração das redes espaciais é avaliada por meio de um modelo gravitacional (seções 4 e 5) e, por fim, algumas das principais conclusões do estudo são enunciadas (seção 6).

## **2. Colaboração Científica no Brasil**

Dentre os mecanismos responsáveis pela articulação das relações sociais na comunidade científica, as redes de coautorias são particularmente importantes, já que são indicadores dos fluxos de conhecimento entre os pesquisadores. No entanto, representam apenas

uma faceta da colaboração<sup>3</sup>, pois não necessariamente as colaborações resultam em publicações em coautoria (KATZ; MARTIN, 1997).<sup>4</sup>

Embora seja indicador imperfeito, os estudos sobre redes de colaborações científicas normalmente utilizam dados referentes à contabilização de coautorias em publicações como medida quantitativa da colaboração científica (WANG *et al*, 2005). Sob a perspectiva da cientometria espacial, faz-se necessária a agregação das coautorias entre indivíduos em unidades geográficas, o que permite a elaboração de matrizes inter-regionais de colaborações científicas, compostas pelos valores dos fluxos observados entre as regiões *i* e *j*, os quais são mensurados a partir da contagem de coautorias em publicações científicas entre pesquisadores localizados nessas regiões.

Nesta seção, o principal objetivo é explicar a fonte das informações sobre colaborações científicas (coautorias) no Brasil, a definição da unidade espacial de análise, o procedimento de coleta, contabilização, periodicidade dos dados e suas principais características e limitações.

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 (CNPQ, 2012).<sup>5</sup> 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, o que tornou o

---

<sup>3</sup> Outras formas possíveis de colaboração científica são: compartilhamento de atividades de edição, organização conjunta de conferências científicas, supervisão compartilhada e desenvolvimento conjunto de projetos de pesquisa (MALI *et al*, 2012).

<sup>4</sup> Os esforços voltados à produção de artigos podem não resultar em publicações devido à rejeição dos trabalhos por parte dos revisores técnicos, às restrições de tempo, ou ao custo de submissão (HOEKMAN *et al*, 2009).

<sup>5</sup> O sistema de CL consiste num abrangente sistema de informações curriculares de pesquisadores, professores, estudantes e profissionais de todas as áreas do conhecimento e possui importância crucial 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 de pesquisa, por fornecer informações confiáveis para a análise de mérito e competência dos pesquisadores, avaliação de programas de pós-graduação e análise de pleitos de financiamentos (CNPQ, 2012).

sistema padrão nacional no registro das atividades acadêmicas e profissionais da comunidade científica.<sup>6</sup> Logo, o estabelecimento de um mecanismo de incentivos para o preenchimento e atualização correta das informações proporcionou credibilidade e reconhecimento internacional ao sistema de CL, modelo de sucesso a ser seguido internacionalmente (LANE, 2010).

Os CL são disponibilizados publicamente no portal da PL na *web* (CNPQ, 2012), porém, embora o acesso às informações individuais seja imediato, não é possível o acesso sistemático à completude da base de dados, o que torna o esforço na coleta de informações o maior empecilho para a análise de grande volume de dados e a necessidade de automatização do processo. Como veremos, o procedimento descrito a seguir deu-se por meio do processamento das informações de 1.131.912 de CL.<sup>7</sup>

Para nossos propósitos, foi utilizada uma versão modificada do *ScriptLattes*<sup>8</sup>, capaz de estabelecer ligações de coautoria entre pesquisadores caso exista uma produção comum entre eles divulgadas em seus CL. Basicamente, o procedimento consiste na comparação direta entre os títulos completos das produções acadêmicas (identificação dos autores, título da publicação, periódico, local e ano) registradas nos CL dos pesquisadores em busca de similaridades.<sup>9</sup> De maneira mais precisa, a identificação e contabilização das coautorias deu-se a partir das informações contidas em quatro campos específicos do módulo *Produção Bibliográfica*<sup>10</sup>: artigos completos publicados em periódicos; trabalhos completos publicados em anais de congressos; livros publicados, organizados ou edições; e capítulos de livros publicados, o que totalizou a análise de 7.351.957 produções acadêmicas distintas publicadas entre 1990 e 2010.

---

<sup>6</sup> A declaração e atualização das informações curriculares pelos pesquisadores são motivadas pela necessidade quanto à divulgação de suas atividades, 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.

<sup>7</sup> Com o intuito de atingir grande representatividade do banco de dados, os CL foram rastreados por meio de seus códigos de identificação. O procedimento completo de mineração dos dados é descrito em Digiampietri *et al* (2011).

<sup>8</sup> O *ScriptLattes* é um *software* livre (<http://scriptlattes.sourceforge.net/>) capaz de extrair informações de CL selecionados e de gerar relatórios de resultados e insumos computacionais para análises de redes sociais. Em processo contínuo de desenvolvimento, o programa tem sido adaptado e utilizado por diversas agências de fomento, institutos de pesquisa e universidades para auxiliar os processos de avaliações institucionais (MENA-CHALCO; CÉSAR JÚNIOR, 2009).

<sup>9</sup> No âmbito da análise das redes sociais, cada pesquisador é representado por um nó e a detecção de uma relação de coautoria entre eles (verificada pela presença de títulos semelhantes nos CL) é representada por uma ligação entre os nós (aresta).

<sup>10</sup> Na versão atual do CL, a produção acadêmica é subdividida entre *Produção Bibliográfica*, *Técnica e Artística/ Cultural*.

Após a identificação das coautorias, as ligações entre os pesquisadores foram contabilizadas por meio do método de contagem completa (*full-counting*), segundo o qual cada unidade de análise (autores ou regiões) recebe uma unidade de crédito pela participação na publicação científica (SCHERNGELL; BARBER, 2011). Como exemplo, para 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.<sup>11</sup>

O procedimento de busca por informações de localização geográfica dos pesquisadores consiste no tema central em cientometria espacial e merece destaque.<sup>12</sup> Em nosso caso, optamos pela escolha dos municípios brasileiros como unidade de análise.<sup>13</sup> Entretanto, ao invés de localizar os coautores a partir de seus endereços nas publicações (ou com o auxílio de bases complementares) fez-se uso direto da informação sobre os endereços profissionais dos pesquisadores declarados nos CL. Inicialmente, foi projetado e desenvolvido um programa para a extração das informações sobre a localização municipal nos CL. No entanto, a autonomia dada ao usuário no preenchimento do campo “endereço profissional” gerou dificuldades, uma vez que se revelou significativa a quantidade de problemas provenientes de erros de digitação, utilização de abreviaturas regionais e alocação errônea de municípios aos seus respectivos estados. Assim, fez-se necessária a padronização dos nomes e a identificação correta dos estados correspondentes, de maneira a permitir o estabelecimento de uma ligação unilateral precisa entre os códigos dos CL e os 4.615 municípios encontrados.<sup>14</sup>

---

<sup>11</sup> 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 (crédito da coautoria é dividido proporcionalmente entre os coautores). Mais comum em análises de citações, tal método subestima o impacto de artigos de coautores de diversas regiões, enquanto que o método de contagem completa superestima tal contagem. Como é usual nos trabalhos empíricos sobre colaboração científica, optamos pela utilização do método de contagem completa.

<sup>12</sup> A determinação da unidade espacial é cerceada de problemas metodológicos, já que as redes científicas consistem em sistemas complexos de entrelaçamento e rompimento das fronteiras formais, em que é difícil que as fronteiras percebidas pelos pesquisadores coincidam com os limites administrativos oficiais (países, estados, municípios). Todavia, a arbitrariedade inerente a qualquer classificação torna o uso das unidades espaciais oficiais escolha usual (FRENKEN *et al*, 2009).

<sup>13</sup> É crescente o uso de municípios como unidades de análise em cientometria espacial (RS, 2011), porém, são bastante raros os estudos de colaboração científica entre municípios (FRENKEN *et al* 2009).

<sup>14</sup> Para grande parte dos CL analisados (49%), a ausência de informação inviabilizou a identificação da localização geográfica do pesquisador. Contudo, a relevância da perda de informação deve ser avaliada em termos do total de coautorias sem localização, assim, foi atribuída a entrada “indeterminado” aos CL sem identificação municipal.



De modo a permitir a diferenciação dos padrões de colaborações científicas entre as diferentes Grandes Áreas do conhecimento (classificação realizada pela CAPES), as coautorias foram identificadas e contabilizadas a partir da discriminação no CL da grande área do conhecimento que abarca o conteúdo das atividades científicas desenvolvidas pelos pesquisadores. 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)*, *Ciências Sociais Aplicadas (SOC)*.<sup>15</sup> A extração dessas informações permitiu a associação entre os pesquisadores e a grande área de conhecimento por eles declarada, conforme apresentado na Tabela 1, onde se observa que 76,7% dos pesquisadores identificados foram associados a uma única área 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 grande área</i>
AGR	92.927	59.484 (64,0%)
BIO	128.104	65.593 (51,2%)
ENG	120.993	70.075 (57,9%)
EXT	176.114	102.372 (58,1%)
HUM	270.149	141.376 (54,2%)
LLA	99.129	53.151 (53,6%)
SAU	272.783	206.772 (75,8%)
SOC	264.230	164.037 (62,1%)
<b>Total</b>	<b>1.131.912</b>	<b>868.250 (76,7%)</b>

Assim, a associação entre pesquisador, município e área do conhecimento permitiu a elaboração de matrizes de coautorias intermunicipais diferenciadas por área do conhecimento. De maneira resumida, cada célula  $(i, j)$  da matriz possui a informação sobre a quantidade de coautorias entre pesquisadores dos municípios  $i$  e  $j$ .<sup>16</sup>

<sup>15</sup> Para os casos de mais de uma área declarada, foi feita a associação entre o pesquisador e todas as áreas declaradas. Também foi considerada a grande área “Outros”, composta por áreas do conhecimento não classificadas dentre as oito citadas.

<sup>16</sup> Essas matrizes são denominadas de matrizes de adjacência em teoria dos grafos, em que os vértices são representados nas linhas e colunas, e cada entrada apresenta a intensidade das ligações entre dois vértices. Assim, para  $n$  regiões, a matriz possui dimensão  $(n \times n)$ , onde as  $i$  regiões são representadas nas linhas e as  $j$  regiões nas colunas  $(i, j = 1, \dots, n)$ .

Por fim, a declaração da informação sobre o ano de publicação (contida no título completo das publicações) permitiu a periodização anual das matrizes de coautorias. Logo, o conjunto consolidado de dados de coautorias consiste em 210 matrizes de coautorias intermunicipais (associadas ao total das áreas e cada área particular, e aos anos compreendidos entre 1990 e 2010) de dimensão 4.615, composta por 10.651.420 valores.<sup>17</sup> Embora os CL tenham sido associados a 4.615 municípios, 3.268 municípios não apresentaram publicações em coautoria, o que permitiu a redução da dimensão das matrizes para 1.347.<sup>18</sup>

Algumas características dos dados utilizados merecem destaque, uma vez que possibilitaram a superação satisfatória de vários problemas usualmente apontados pela literatura especializada.

Primeiramente, destaca-se a abrangência da amostra utilizada, tanto em relação ao número de pesquisadores quanto ao amplo período de coleta de dados, uma vez que o volume de dados analisados é bastante superior ao usualmente analisado.<sup>19</sup>

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 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).<sup>20</sup> Esse problemas são, em certa medida, contornados, já que a localização geográfica deu-se por meio das informações dos endereços declarados pelos autores.

---

<sup>17</sup> 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, para  $n = 4615$ , teremos 10.651.420 entradas distintas.

<sup>18</sup> Para os anos mais recentes, é possível que a contabilização esteja subestimada, pois é dependente da atualização dos CL.

<sup>19</sup> Scherngell e Hu (2011) contabilizaram 758.682 coautorias em metade das publicações de pesquisadores chineses em 2007. Já Hoekman *et al* (2010) contabilizaram 524.155 coautorias entre regiões da Europa em 2007, na análise a partir de 2000.

<sup>20</sup> Pesquisadores em visita temporária podem escolher o registro de seu instituto ou de organizações de subvenção (financiamento) ao invés do instituto onde a pesquisa é realmente realizada. Em relação aos institutos de pesquisas e empresas, as localidades das sedes podem ser registradas ao invés das filiais onde as pesquisas foram efetivamente realizadas.

Em terceiro lugar, destaca-se a qualidade da fonte de dados para a análise da ciência brasileira, na medida em que a grande maioria dos estudos cientométricos faz uso de bases de dados internacionais, o que produz duas grandes limitações.<sup>21</sup> Primeiramente, essas bases apresentam certo viés, por cobrirem principalmente os periódicos de língua inglesa, o que exige o pressuposto de que os periódicos não indexados sejam locais ou domésticos, os quais publicam suas pesquisas na língua nativa.<sup>22</sup> Em segundo lugar, é bastante baixa a cobertura da produção das áreas de ciências sociais e humanidades (HOEKMAN *et al*, 2010), devido, principalmente, 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, o que faz com que essas áreas possuam viés para a publicação na língua local. Logo, a utilização de bases de dados internacionais não permite a avaliação completa da produtividade científica brasileira, uma vez que nos países em desenvolvimento o novo conhecimento é publicado e divulgado por periódicos locais, muitos dos quais não são incorporados 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 nas áreas de ciências humanas e sociais aplicadas.

Diante dessas peculiaridades, destaca-se a coleta de informações a partir de dados do sistema de CL, o que permite a extensa cobertura da publicação científica brasileira, com a contabilização das coautorias tanto de artigos publicados em periódicos de circulação nacional como da produção bibliográfica sob a forma de livros e capítulos de livros, de modo a permitir uma melhor avaliação da produção científica das áreas de ciências sociais e humanidades.

Entre as limitações do procedimento utilizado, a principal consiste na restrição à contabilização apenas das coautorias domésticas, ou seja, não é possível a identificação

---

<sup>21</sup> Como exemplo, a base de dados *Web of Science* (WoS) indexa aproximadamente 10.000 periódicos e é considerada uma das mais abrangentes e confiáveis fontes de informação sobre as atividades científicas no mundo (ADAMS; KING, 2009).

<sup>22</sup> É consensual a ideia de que esse tipo de publicação possua perfil menos colaborativo e que a colaboração deva ser mais sensível a vieses espaciais, assim, é possível que o viés espacial seja subestimado em análises de dados de bases internacionais.

e contabilização das colaborações entre brasileiros e estrangeiros.<sup>23</sup> Assim, uma publicação produzida em parceria com pesquisadores estrangeiros só é contabilizada caso ocorra simultaneamente colaboração entre pesquisadores brasileiros (colaboração doméstica).

A natureza inerente aos dados também impõe certas limitações. Primeiramente, a utilização de informações de endereços para a localização geográfica baseia-se na suposição de que o registro dos endereços dos pesquisadores corresponda à localização onde a pesquisa foi verdadeiramente conduzida. Em segundo lugar, é pressuposto que a informação observada da localização geográfica dos pesquisadores seja a verdadeira localização do pesquisador durante todos os anos analisados.<sup>24</sup> Assim, a estrutura de coleta dos dados não considera a possível migração de pesquisadores ao longo do território nacional. Contudo, os problemas não invalidam a utilização dos dados, uma vez que muitos dos problemas apontados são, em certa medida, contornados ou reduzidos pelo amplo tamanho da amostra utilizada.

### **3. Padrões Locacionais de Produção e Colaboração Científica no Brasil**

Inicialmente, a perda de informação associada à impossibilidade de identificação da localização geográfica em cerca de 49% do total de CL analisados é quantificada mais precisamente pela importância relativa desses indivíduos em atividades científicas, ou seja, em termos do número de publicações e de coautorias contabilizadas sem localização municipal.<sup>25</sup> Observa-se na Tabela 2 que, embora não seja desprezível, a perda de informação não acarreta em grandes prejuízos no tocante à análise estrutural da evolução da produção e colaboração científica.

---

<sup>23</sup> Como o procedimento computacional efetiva a coautoria ao identificar títulos semelhantes em CL diferentes, as colaborações entre pesquisadores brasileiros e estrangeiros não é possível devido à inexistência de CL desses últimos.

<sup>24</sup> Como exemplo, a contabilização da coautoria entre os autores de determinada publicação referente ao ano de 1996 é atribuída aos municípios com base na informação da localização geográfica dos pesquisadores extraída em 2011.

<sup>25</sup> No Brasil, praticamente todas as pessoas envolvidas com pesquisa, desde pesquisadores ativos até estudantes de graduação possuem CL. Enquanto os primeiros são os maiores responsáveis pela publicação científica, existem muitos casos de CL sem informações sobre produção bibliográfica, situação típica de estudantes de graduação que ainda estão iniciando seus estudos.

**Tabela 2. Percentual de Publicações e Colaborações sem Localização Municipal**

	1992-1994	1995-1997	1998-2000	2001-2003	2004-2006	2007-2009
<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%

A determinação de padrões espaciais da atividade científica possui como característica crucial a maneira pela qual o produto é alocado entre as unidades espaciais, processo muitas vezes permeado por dificuldades, já que a associação entre uma publicação e uma localidade específica não é passível de ser determinada de maneira direta e única.<sup>26</sup> Em nosso caso, foi adotada a estratégia de mensurar a participação dos pesquisadores em publicações, ao invés de mensurar o total de publicações efetivas. Assim, a partir de uma publicação em coautoria entre dois pesquisadores de municípios distintos, 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 publicação). Desse modo, 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.<sup>27</sup> Embora a contabilização fracionada<sup>28</sup> pareça ser a escolha mais intuitiva à primeira vista (por manter o total de publicações efetivas), tal método também apresenta problemas, já que a suposição implícita de que cada autor contribui igualmente ao produto final é discutível (OSBORNE; HOLLAND, 2009).<sup>29</sup> Ambos os procedimentos são encontrados na literatura, em que a comparação entre os métodos mostra que a ordem da produção das regiões independe do método de contabilização escolhido, sendo consensual a ideia de que os esforços regionais nas publicações colaborativas são subestimados quando se utiliza o método de contagem fracionada (GROSSETTI et al, 2012; MICHELS et al, 2013). Dessa maneira, como o intuito é a análise espacial da produção científica brasileira (e não a quantificação precisa de sua evolução), optamos por considerar a participação absoluta em publicações como indicador da atividade científica municipal,

<sup>26</sup> 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 diretamente a publicação a um único município.

<sup>27</sup> Esse método aproxima-se ao método de contabilização utilizado na contagem das coautorias, discutido anteriormente.

<sup>28</sup> Nesse método, a contabilização dá-se pela contribuição relativa de cada autor em determinada publicação, ou seja, no caso de coautores de  $n$  municípios, é atribuído o valor  $(1/n)$  para cada ligação entre os municípios envolvidos.

<sup>29</sup> Por exemplo, a ordenação dos autores segundo suas contribuições relativas é fenômeno típico das áreas de maior perfil colaborativo, o que torna problemática a suposição de contribuição igualitária entre eles.

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.

Embora tenham sido construídas matrizes intermunicipais de coautorias para cada ano do período 1990-2010, optamos pelo agrupamento das matrizes em 6 triênios, de acordo com os triênios selecionados pela avaliação trienal da CAPES.<sup>30</sup> Por um lado, a determinação da janela temporal ideal em análises cientométricas é tema de debate na literatura, em que predomina o uso de períodos entre 2 e 5 anos. Por outro lado, o modelo de avaliação da produção científica brasileira nas últimas décadas esteve baseado na determinação de padrões internacionais a serem perseguidos pelos pesquisadores, o que funciona como um verdadeiro mecanismo de incentivo à pesquisa científica (VANZ, 2009; LETA, 2011).<sup>31</sup>

A Figura 1 apresenta a participação na produção total e a taxa de crescimento da produção de cada uma das áreas do conhecimento. Nota-se que a contribuição relativa de cada área para a produção total não apresentou grandes transformações ao longo do período, enquanto que se verifica um crescimento acelerado do total da produção ao longo dos triênios selecionados, mas com as taxas de crescimento de cada área se aproximando à tendência geral de desaceleração do crescimento da produção total.<sup>32</sup>

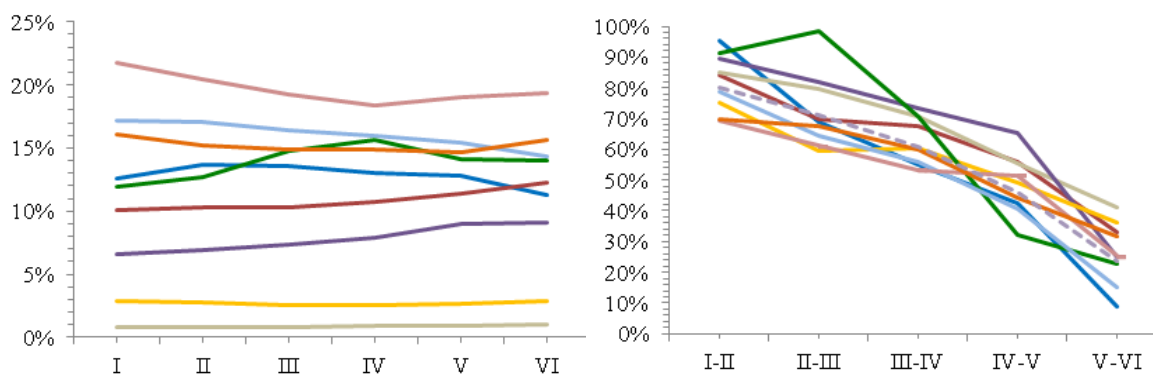
### **Figura 1. Participação e Taxas de Crescimento da Produção Científica das Áreas do Conhecimento**

---

<sup>30</sup> Os triênios selecionados são: I: 1992-1994; II: 1995-1997; III: 1998-2000; IV: 2001-2003; V: 2004-2006; VI: 2007-2009.

<sup>31</sup> Acreditamos que a janela trienal no caso brasileiro é bastante razoável, 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 os esforços não sejam contemplados somente numa avaliação futura.

<sup>32</sup> Conforme afirmado, a interpretação deve ser bastante cautelosa, por não ser possível a determinação imediata das causas do crescimento observado, já que tal comportamento pode ter sido gerado tanto da aceleração da produção científica em termos de publicações efetivas, bem como do aumento do perfil colaborativo entre os pesquisadores no período.



Triênios: I: 1992-1994; II: 1995-1997; III: 1998-2000; IV: 2001-2003; V: 2004-2006; VI: 2007-2009

— ENG — HUM — OUT — SOC — LLA — AGR — EXA — SAU — BIO - - - Total

**Tabela 3. Evolução dos Principais Municípios Associados à Produção Científica no Brasil: Todas as Áreas do Conhecimento**

	<i>Total</i>		<i>Total</i>		<i>Total</i>
	<b>1992-1994</b>		<b>1995-1997</b>		<b>1998-2000</b>
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
	<b>2001-2003</b>		<b>2004-2006</b>		<b>2007-2009</b>
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.

Quanto à análise individual, os dez principais municípios associados ao total da produção científica no Brasil em cada triênio são apresentados na Tabela 3. Os valores

apresentados evidenciam a enorme heterogeneidade espacial das atividades científicas no país, altamente concentradas na região Sudeste<sup>33</sup>, uma vez que somente dois municípios não pertencentes a tal região (Recife/PE e Brasília/DF) figuram entre os principais produtores de conhecimento.<sup>34</sup> 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<sup>35</sup> (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. A pequena 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 atividades e produção de destaque no cenário científico nacional e internacional.<sup>36</sup>

A heterogeneidade espacial no caso brasileiro evidencia a preocupação em relação à necessidade de desconcentrar as atividades de pesquisa científica ao longo do país, uma vez que isso pode potencializar o desenvolvimento regional de áreas menos favorecidas.<sup>37</sup> Nesse contexto, torna-se fundamental a compreensão da maneira que o crescimento da produção científica brasileira deu-se no espaço geográfico.

O Mapa 1 apresenta a produção científica municipal nos triênios de 1992-1994 e 2007-2009. Observa-se que a produção de conhecimento está melhor distribuída espacialmente no segundo período.<sup>38</sup> Para o triênio 2007-2009, é introduzida a estrutura

---

<sup>33</sup> Resultado semelhantemente foi encontrado em FAPESP (2011a, 2011b).

<sup>34</sup> No triênio 2007-2009, 38 municípios pertencentes ao Sudeste figuravam entre os 50 maiores produtores de conhecimento.

<sup>35</sup> 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 (RS, 2011) e destacou-se entre as cidades mundiais que mais apresentaram crescimento na produção científica e no padrão de citações (MATTHIESSEN *et al*, 2010).

<sup>36</sup> 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), entre outros, desenvolvem atividades de pesquisa de alto nível e reconhecidas no cenário científico nacional e internacional.

<sup>37</sup> Entre outros motivos, a desconcentração espacial das atividades de pesquisa permite o direcionamento de esforços para o enfrentamento de problemas locais.

<sup>38</sup> É elevado o número de municípios que passaram da faixa de baixa produção (11-100) em 1992-1994 para o nível de produção intermediária (101-10.000) em 2007-2009.



dos *campi* das universidades públicas (federais e estaduais) vigente em 2009, onde é nítida a associação entre suas localizações e a produção científica municipal.<sup>39</sup>

Todavia, a verificação da existência de um processo sistemático de desconcentração espacial entre os períodos considerados é confirmada pela construção de curvas de localização da produção científica para os 200 municípios de maior produção em cada triênio selecionado (Figura 2). 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, há evidências de que o crescimento da produção científica no Brasil deu-se de maneira espacialmente desconcentrada no período analisado, semelhantemente ao padrão observado em outros países, tais como Rússia, França, Espanha e China, onde o desenvolvimento de seus sistemas de produção científica parece seguir uma tendência de desconcentração espacial, baseada, fundamentalmente, no crescimento da produção nas cidades secundárias, caracterizadas por um nível intermediária produção científica (GROSSETTI *et al*, 2012).<sup>40</sup> A existência de um processo de desconcentração espacial é comum a todas as áreas do conhecimento, porém, intensidades distintas, conforme observado na comparação entre as curvas de localização na Figura 3.

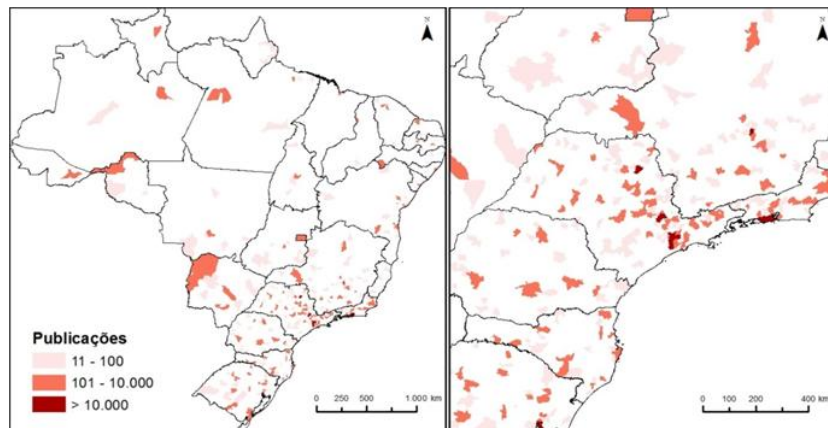
---

<sup>39</sup> Embora não sejam apresentados, a melhor distribuição espacial foi evidenciada pelos mapas de todas as áreas.

<sup>40</sup> É importante destacarmos que o processo observado de desconcentração espacial da atividade científica está subestimado devido à contabilização da produção municipal por meio da participação em publicações científicas, uma vez que tal método favorece os municípios de maior produção, pois esses estão envolvidos de maneira sistemática na rede de colaborações científicas (GROSSETTI *et al*, 2012). Verificamos também que o processo generalizado de desconcentração espacial deu-se para todas as áreas do conhecimento até o triênio 2004-2006, mas no triênio seguinte (2007-2009) houve forte desaceleração do processo de desconcentração, e até mesmo sua reversão nos casos das Ciências Agrárias e Ciências Biológicas.

## Mapa 1. Produção Científica Municipal no Brasil em Triênios Seleccionados

1992-1994



2007-2009

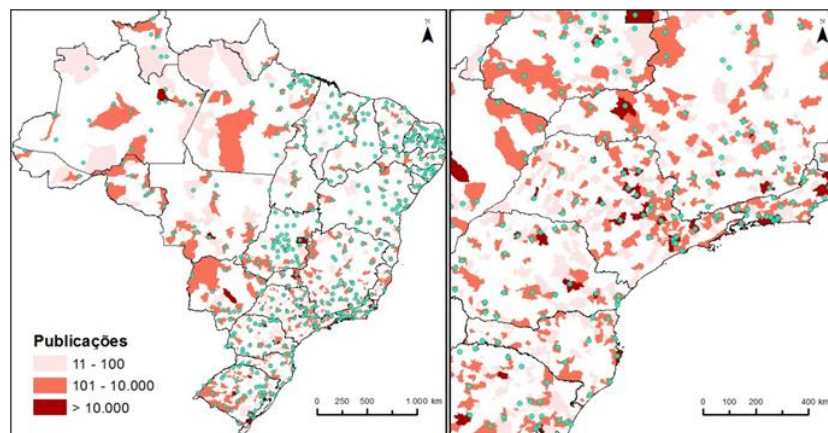
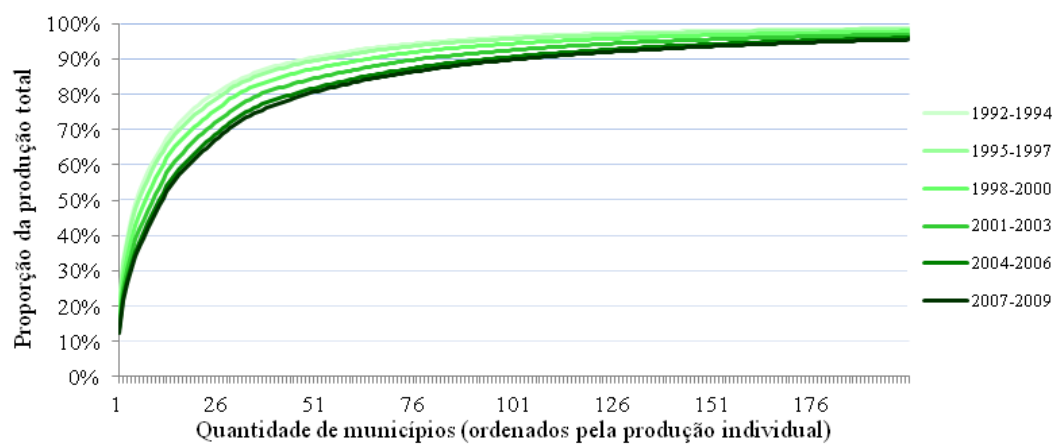
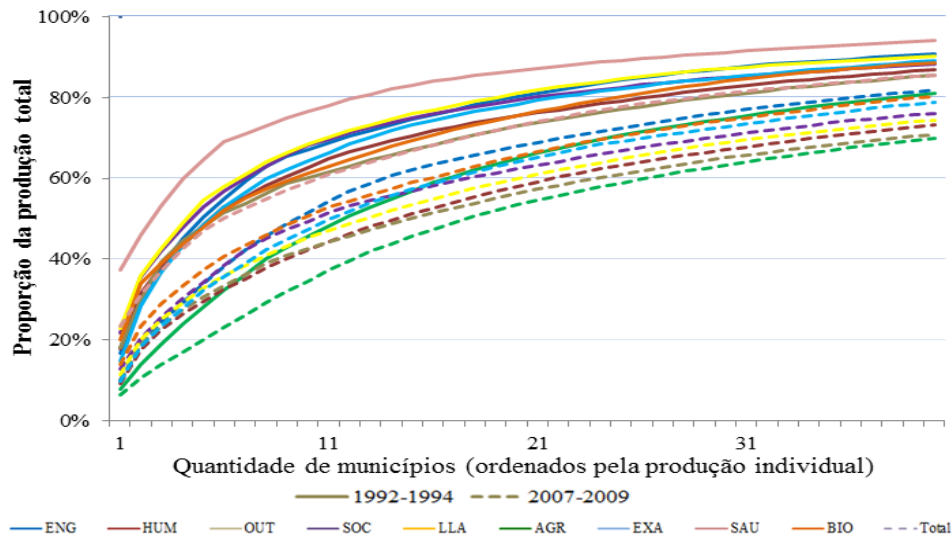


Figura 2. Evolução das Curvas de Localização Municipal da Produção Científica Total



**Figura 3. Evolução das Curvas de Localização Municipal da Produção Científica:  
por Área do Conhecimento**



A análise das matrizes de coautorias revela algumas tendências gerais acerca da evolução das redes de colaboração científica no País. Primeiramente, verifica-se uma tendência acentuada de crescimento das colaborações, tanto em termos da quantidade total de colaborações intermunicipais e intramunicipais, como de seus valores médios. O total de colaborações científicas, mensuradas pelas coautorias, saltou de 547.249 no triênio 1992-1994 para 9.445.399 no triênio 2007-2009 (as colaborações intramunicipais passaram de 317.810 para 1.037.274, enquanto que as intermunicipais passaram de 229.439 para 8.408.125). Nesse processo de crescimento, destaca-se o período de aceleração nos triênios de 2001-2003 e 2004-2007, o que evidencia a importância da colaboração doméstica como um dos motores do crescimento acelerado da produção científica brasileira nesse período.

**Tabela 4. Principais Ligações Intermunicipais de Colaboração Científica no Brasil**

1992-1994		1995-1997	
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

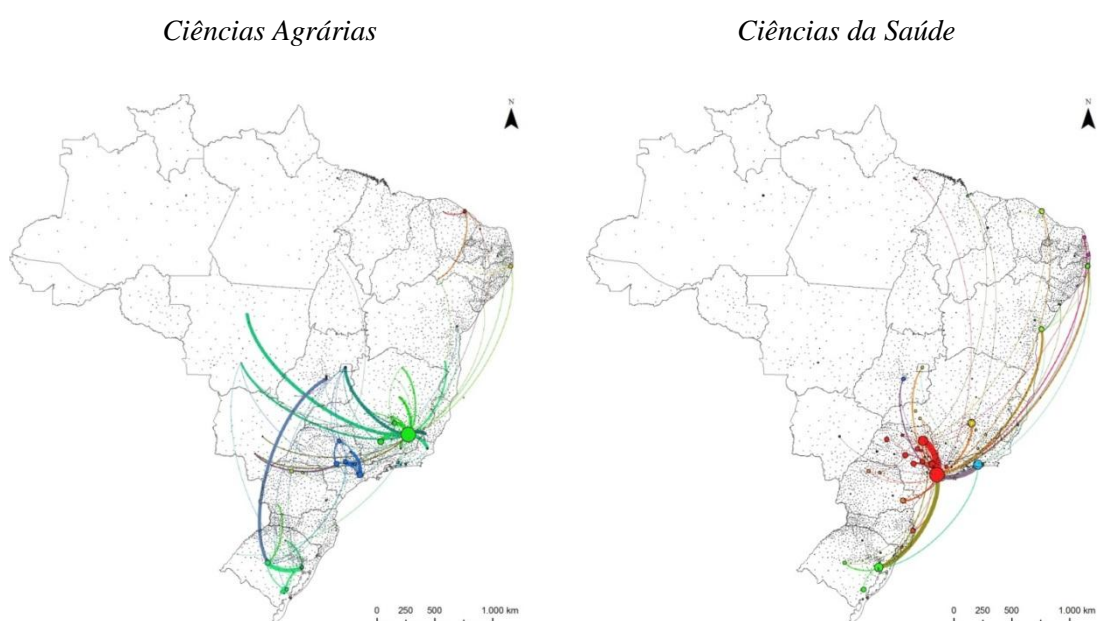
A Tabela 4 apresenta a evolução dos principais fluxos de conhecimento intermunicipais nos triênios selecionados para o total das áreas, em que a maior ligação deu-se na contabilização de 76.716 colaborações entre pesquisadores de Campinas/SP e São Paulo/SP no triênio 2007-2009. De um modo geral, observa-se a localização sistemática dos fluxos na região Sudeste do país (padrão similar àquele verificado na análise da produção científica municipal), uma vez que as principais ligações intermunicipais e intramunicipais<sup>41</sup> dão-se, primordialmente, no âmbito das capitais dos estados do Sudeste e de municípios no interior 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.

<sup>41</sup> Também foram contabilizadas as ligações intermunicipais de colaborações científicas (diagonal principal da matriz de coautorias)

De modo a facilitar a visualização dos fluxos de conhecimento ao longo do território, a Figura 4 apresenta mapas com os 100 principais fluxos de conhecimento das áreas de Ciências Agrárias e da Saúde para o período completo (1990-2010).<sup>42</sup> A comparação entre os mapas evidencia a diferença entre o padrão geográfico das colaborações em cada área.

Por fim, a análise da evolução dos valores de algumas métricas globais das redes de colaboração também permite resultados interessantes. A Figura 5 mostra a quantidade de municípios conectados (medida tamanho da rede) e o grau médio dos municípios<sup>43</sup> (medida de integração da rede) nas redes de colaboração de cada área do conhecimento. Observa-se que com o passar do tempo houve tanto a expansão das redes de colaborações (com a incorporação de novos municípios às redes) como a intensificação das relações nas redes (crescimento do grau médio dos municípios<sup>44</sup>).

**Figura 4. Principais Fluxos de Colaborações em Ciências Agrárias e Ciências da Saúde no Período 1990-2010**

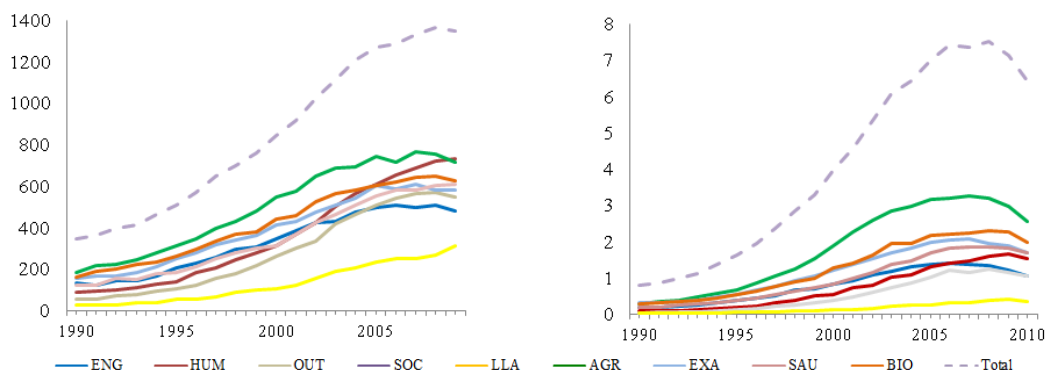


<sup>42</sup> Apesar de não apresentados, também foram construídos mapas semelhantes para as outras áreas do conhecimento.

<sup>43</sup> O grau de um nó mede a quantidade de arestas distintas ligadas a ele. O grau médio é a média dos graus dos nós individuais.

<sup>44</sup> O grau médio de 7,5 em 2009 significa que cada município da rede está ligado, em média, a outros 7,5 municípios distintos.

**Figura 5. Evolução da Quantidade de Municípios e do Grau Médio nas Redes de Colaboração Científica**



#### 4. Modelo de Interações Espaciais

A ampliação das redes espaciais de colaboração científica e a intensificação de suas relações abre espaço para a discussão sobre o papel da distância geográfica na articulação entre os pesquisadores ao longo do território nacional. O objetivo dessa seção é a quantificação do efeito da distância geográfica como fator impeditivo às colaborações científicas entre os pesquisadores brasileiros.

A avaliação correta do efeito de uma forma particular de proximidade, como, por exemplo, a proximidade geográfica, é alcançada de maneira mais precisa por meio de um ferramental multivariado, devido à possibilidade de isolamento e controle do efeito das outras dimensões de proximidade.<sup>45</sup> Nesse contexto, optamos pela modelagem da estrutura espacial dos fluxos de colaboração científica por meio da abordagem dos modelos gravitacionais de interações espaciais<sup>46</sup>, procedimento comumente utilizado nos estudos de ciëntometria espacial.<sup>47</sup> Basicamente, o modelo de interação espacial

<sup>45</sup> A definição usual das dimensões de proximidade é apresentada em Boschma (2005).

<sup>46</sup> Os modelos gravitacionais são fundamentais na identificação das fontes de disparidades regionais existentes em fenômenos oriundos de interações humanas, pois permitem a verificação da hipótese de que a distância geográfica seja a principal responsável ou se existem outros determinantes que explicam os padrões de interação observados.

<sup>47</sup> Os modelos gravitacionais explicaram a intensidade das colaborações científicas entre regiões na Holanda (PONDS *et al.*, 2007), China (WANG *et al.*, 2005; SCHERNGELL; HU, 2011) e na Europa (HOEKMAN *et al.*, 2009).

gravitacional é caracterizado pela distinção formal entre três tipos de funções capazes de explicar a variação das interações inter-regionais em um modelo de regressão:<sup>48</sup>

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

$$X_{ij} = O_i \cdot D_j \cdot S(u_{ij}) \quad ; \quad i, j = 1, \dots, n \quad (2)$$

As funções  $O_i$  e  $D_j$  caracterizam as regiões de interação  $i$  e  $j$ , e podem ser especificadas por meio de “funções de poder”, de acordo com a teoria clássica da interação espacial (SEN; SMITH, 1995). Já a especificação tradicional do termo de separação espacial  $S(u_{ij})$  dá-se por meio da forma funcional multivariada exponencial. Assim, as funções assumem as formas a seguir:

$$O_i = O(o_i, \alpha_1) = o_i^{\alpha_1} \quad ; \quad i = 1, \dots, n \quad (3)$$

$$D_j = D(d_j, \alpha_2) = d_j^{\alpha_2} \quad ; \quad j = 1, \dots, n \quad (4)$$

$$S(u_{ij}) = S(u_{ij}, \beta) = \exp\left[\sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}\right] \quad ; \quad i, j = 1, \dots, n \quad (5)$$

Onde  $o_i$  e  $d_j$  são variáveis que mensuram características específicas das regiões  $i$  e  $j$  e as variáveis  $u_{ij}^{(k)}$  representam  $k$  medidas de separação espacial entre as regiões  $i$  e  $j$ . Já os termos  $\alpha_1$  e  $\alpha_2$  são os parâmetros a serem estimados nas duas especificações<sup>49</sup> e o termo  $\beta_k$  refere-se ao conjunto de  $k$  parâmetros desconhecidos associados a cada uma das  $k$  medidas de separação espacial entre  $i$  e  $j$ . Por meio da substituição das especificações no modelo inicial, obtemos o modelo empírico a ser estimado:

$$Y_{ij} = \alpha + o_i^{\alpha_1} \cdot d_j^{\alpha_2} \cdot \exp\left[\sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}\right] + \varepsilon_{ij} \quad ; \quad i, j = 1, \dots, n \quad (6)$$

<sup>48</sup> 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.

<sup>49</sup> Como nas colaborações científicas não existe determinação da direção dos fluxos entre as regiões (interações são fruto de colaborações sem direcionamento), as variáveis de origem e destino são simétricas e, assim,  $\alpha_1 = \alpha_2$ , onde se espera que as estimativas sejam significantes e próximas à unidade. Logo, o produto entre  $O_i$  e  $D_j$  pode ser simplesmente interpretado como o número total de possíveis colaborações distintas entre as duas regiões  $i$  e  $j$  (SCHERNGELL; BARBER, 2011).

É importante notar que a natureza dos dados de coautorias, caracterizados por valores inteiros e não-negativos, torna inapropriada a aplicação de uma especificação log-normal sobre a equação do modelo (4.6), e a consequente estimação dos parâmetros pelo tradicional método de mínimos quadrados ordinários (MQO) (LONG; FREESE, 2001).<sup>50</sup> Todavia, as deficiências da especificação log-normal e os fortes pressupostos necessários à aplicação de MQO podem ser contornados por meio da interpretação do modelo como de dados de contagem, em que é assumido 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, dada pela seguinte expressão:<sup>51</sup>

$$\Pr(Y_{ij} = \gamma_{ij} \mid X_{ij}) = \frac{e^{-\mu_{ij}} \cdot \mu_{ij}^{\gamma_{ij}}}{\gamma_{ij}!}; \quad i, j = 1, \dots, n; \quad \gamma_{ij} = 0, 1, 2, \dots \quad (7)$$

onde  $\mu_{ij}$  representa o conjunto de variáveis dependentes do modelo empírico (6):

$$\mu_{ij} = X_{ij} = \exp[\alpha_0 + \alpha_1 \cdot \log(o_i) + \alpha_2 \cdot \log(d_j) + \sum_{k=1}^K \beta_k \cdot u_{ij}^{(k)}] \quad (8)$$

A função de distribuição de Poisson possui a propriedade estatística de equidispersão, definida pela igualdade entre a média e variância condicionais.<sup>52</sup> 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. 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 distribuição desses valores não satisfaça à propriedade de equidispersão<sup>53</sup>, situação que normalmente

---

<sup>50</sup> Além da inadequação do método de MQO, 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).

<sup>51</sup> De modo geral, a família de modelos de Poisson resolve as deficiências técnicas de MQO, reconhece explicitamente a natureza inteira e não-negativa da variável dependente, e permitem que as estimativas de máxima-verossimilhança dos parâmetros possam ser interpretadas como elasticidades (WINKELMANN, 2008; SCHERNGELL; HU, 2011).

<sup>52</sup>  $\mu_{ij} = V[Y_{ij} \mid X_{ij}] = E[Y_{ij} \mid X_{ij}]$

<sup>53</sup> Diferentemente de outras distribuições paramétricas, a violação da hipótese de equidispersão é suficiente para a violação da hipótese de um processo gerador de dados de Poisson, e é usual em análises empíricas a verificação da presença de superdispersão (variância condicional superior à média



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. A superdispersão leva a estimativas viesadas dos parâmetros e invalidade dos testes de hipóteses usuais, uma vez que os erros-padrão são subestimados (WINKELMANN, 2008; HILBE, 2011).<sup>54</sup>

Nesse contexto, a alternativa típica nos trabalhos empíricos sobre colaborações científicas é a utilização do modelo binomial negativo (HOEKMAN *et al*, 2010; SCHERNGELL; BARBER, 2011; SCHERNGELL; HU, 2011; SCHERNGELL; LATA, 2011), capaz de lidar com a heterogeneidade não-observada por meio da inclusão de um parâmetro adicional (parâmetro de heterogeneidade) que permite a acomodação da superdispersão observada nos dados. As expressões da densidade da distribuição binomial negativa e da variância condicional seguem abaixo:<sup>55</sup>

$$\Pr(Y_{ij} = \gamma_{ij} | X_{ij}) = \frac{\Gamma(\gamma_{ij} + \alpha^{-1})}{\gamma_{ij}! \Gamma(\alpha^{-1})} \cdot \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}}\right)^{\alpha^{-1}} \cdot \left(\frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}}\right)^{\gamma_{ij}} \quad (10)$$

$$V[Y_{ij} | X_{ij}] = \mu_{ij}(1 + \alpha \cdot \mu_{ij}) \quad ; \quad \mu_{ij} = E[Y_{ij} | X_{ij}] \quad (11)$$

Onde  $\Gamma(\cdot)$  é a função gama e  $\alpha$  é o parâmetro de heterogeneidade.<sup>56</sup>

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, a verificação da significância estatística da estimativa do parâmetro de heterogeneidade  $\alpha$  (estimado a partir do modelo binomial negativo) permite a decisão entre os dois modelos (LONG; FREESE, 2001).

---

condicional), entendida como a verificação de excessiva correlação nos dados analisados (WINKELMANN, 2008; HILBE, 2011)

<sup>54</sup> 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 (WINKELMANN, 2008).

<sup>55</sup> A expressão apresentada refere-se à especificação *Negbin II* (WINKELMANN, 2008, p. 134), a qual é a mais utilizada na pesquisa empírica (HILBE, 2011).

<sup>56</sup> Observe que é permitida uma forma mais geral de heterogeneidade (alternativa à igualdade entre média e variâncias condicionais impostas pela distribuição de Poisson) Dessa maneira, o principal elemento de decisão de especificação entre os dois modelos consiste na verificação da presença de heterogeneidade não-observada (WINKELMANN, 2008).

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 inflacionadas de zeros do modelo de Poisson (ZIP) e binomial negativo (ZINB)<sup>57</sup>, os quais admitem uma estrutura de média condicional que diferencia os valores nulos e positivos (HILBE, 2011).<sup>58</sup>

## 5. Resultados

Nessa seção, descreveremos as variáveis utilizadas e as estimativas dos parâmetros do modelo 6. As variáveis de origem ( $o_i$ ) destino ( $d_j$ ) foram medidas pelo total de publicações científicas em cada município, em que é esperado que o total de colaborações entre pesquisadores dos municípios  $i$  e  $j$  ( $\gamma_{ij}$ ) dependa positivamente do total de publicações em cada município, já que quanto maior a produção científica de um município, maior deve ser a probabilidade de haver colaboração.

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  $j$ .<sup>59</sup> É esperado que a distância desempenhe papel impeditivo nas interações entre os pesquisadores, porém, é possível que seu efeito tenha se reduzido com o passar do tempo devido as maiores facilidades de comunicação. Além dessa medida, foi introduzida uma segunda variável de separação, com intuito de mensurar a proximidade institucional entre os municípios. A partir da atribuição do valor  $u_{ij}^{(2)} = 1$

---

<sup>57</sup> 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.

<sup>58</sup> 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).

<sup>59</sup> 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. Já as distâncias intramunicipais  $u_{ii}^{(1)}$  (termos da diagonal principal da matriz de distâncias) foram calculadas como função da área do município  $i$  ( $A_i$ ), por meio da fórmula de Bröcker (1989):  $u_{ii}^{(1)} = \left(\frac{2}{3}\right) \cdot \left(\frac{A_i}{\pi}\right)^{0,5}$

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.<sup>60</sup> Logo, é esperado que o fato de dois municípios possuírem instalações de *campi* de universidades públicas aumente a probabilidade de haver colaboração científica entre os pesquisadores desses municípios.

Foi utilizada uma amostra reduzida de 105 municípios para as estimações dos modelos de Poisson, binomial negativo, ZIP e ZINB (correspondente ao conjunto dos municípios que mais colaboraram no triênio 2007-2009)<sup>61</sup>, uma vez que a matriz completa de colaborações (1.347 municípios) apresentou 97,8% de valores nulos, quantidade tão elevada que inviabilizaria as estimações. Logo, a observação de fluxos de colaboração entre os 105 municípios totalizou 11.025 observações.

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 5.<sup>62</sup> Conforme esperado, as estimativas das medidas de massa (origem e destino) são estatisticamente significantes e próximas de 1 para todos os triênios selecionados, o que é indício da boa especificação desses modelos. Também se observa 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 de 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, o interesse principal recai sobre a estimativa do efeito da distância geográfica sobre a probabilidade de colaboração. Observa-se na Tabela 5 que essas estimativas são estatisticamente significantes e com sinal negativo para todos os triênios selecionados, o que corrobora a hipótese de que o aumento da distância entre dois pesquisadores reduz a probabilidade de colaboração entre eles, *coeteris paribus*. Todavia, a interpretação das estimativas nos modelos de contagem (não-lineares) não é tão imediata como no modelo clássico de

---

<sup>60</sup> A matriz da distância institucional para cada triênio resulta da identificação dos municípios brasileiros que possuíam *campi* de universidades públicas (estaduais ou federais) no primeiro ano de cada triênio analisado.

<sup>61</sup> Esses foram responsáveis por 87,6% e 77,4% do total de publicações e coautorias contabilizadas no período.

<sup>62</sup> As estimações foram realizadas no ambiente computacional do *software* STATA 11.2.

regressão linear, assim, 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 (600) Km reduz a probabilidade de haver colaboração em 41,3% (65,6%), em média.<sup>63</sup> Contudo, a hipótese esperada de que o efeito da distância geográfica teria diminuído com o passar do tempo não foi corroborada pelos resultados das estimações, o que é evidência de que a distância ainda desempenha papel determinante na articulação das redes de colaboração científica.<sup>64</sup>

Por fim, observa-se na Figura 6 que o efeito da distância geográfica sobre a probabilidade de colaboração não é proporcional à distância de maneira linear<sup>65</sup> e varia consideravelmente entre as redes de colaboração científicas das diferentes áreas do conhecimento. Como exemplo, o distanciamento de 400 quilômetros 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 é chega a 65% caso sejam de Ciências Agrárias ou Ciências Exatas e da Terra.

---

<sup>63</sup> 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):  $100 \cdot \frac{E(y|x, x_k + \delta) - E(y|x, x_k)}{E(y|x, x_k)} = 100 \cdot [\exp(\beta_k \cdot \delta) - 1]$

<sup>64</sup> Resultado semelhante também é encontrado em Hoekman *et al* (2010).

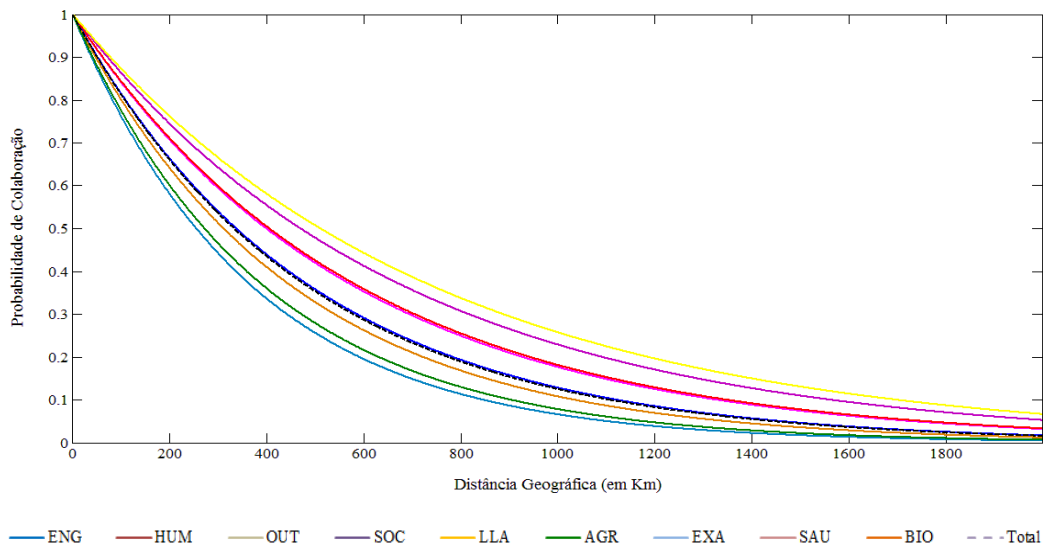
<sup>65</sup> De acordo com a figura, para o total das áreas do conhecimento, o distanciamento de dois pesquisadores em 100 Km provoca a redução em 16%, em média, da probabilidade de haver colaboração entre eles. Já o distanciamento em 400 quilômetros reduz praticamente pela metade (50%) a probabilidade de haver colaboração. Também é interessante notar que a proximidade total (distância geográfica nula) está associada à verificação da probabilidade de 100% de colaboração, ou seja, a distância nula não possui impacto sobre a probabilidade de colaboração. Entretanto, cabe ressaltar que o modelo exclui essa situação, uma vez que sempre admitimos alguma distância geográfica entre dois pesquisadores, mesmo que ambos sejam de um único município.

**Tabela 5. Estimativas dos Modelos de Poisson, Binomial Negativo, ZIP e ZINB para os Triênios Selecionados: Total das 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 ( $\beta_1$ )	-0,00195*** (0,0003065)	-0,00192*** (0,00025)	-0,00174*** (0,00020)	-0,00153*** (0,00016)	-0,00154*** (0,00015)	-0,00177*** (0,00017)
	Distância Institucional ( $\beta_2$ )	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 ( $\alpha_0$ )	-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 ( $\beta_1$ )	-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 ( $\beta_2$ )	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 ( $\alpha_0$ )	-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 ( $\alpha$ )	6,082* (0,19144)	5,0892* (0,1167)	4,5624* (0,09137)	3,8480* (0,06647)	3,6189* (0,06441)	3,7508* (0,06666)
ZIP	Origem – Destino ( $\alpha_1 = \alpha_2$ )	0,78074*** (0,00092)	0,75185*** (0,0006)	0,80186*** (0,00041)	0,76963*** (0,00030)	0,7812*** (0,00023)	0,77849*** (0,00020)
	Distância Geográfica ( $\beta_1$ )	-0,0019*** (4,00.10 <sup>-6</sup> )	-0,00188*** (2,59.10 <sup>-6</sup> )	-0,00172*** (1,51.10 <sup>-6</sup> )	-0,00152*** (9,39.10 <sup>-7</sup> )	-0,00153*** (7,19.10 <sup>-7</sup> )	-0,00176*** (6,51.10 <sup>-7</sup> )
	Distância Institucional ( $\beta_2$ )	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 ( $\alpha_0$ )	-7,0511*** (0,01261)	-6,0088*** (0,0088)	-7,4462*** (0,00612)	-6,9826*** (0,00449)	-7,300*** (0,00359)	-7,1203*** (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*** (0,1361)	0,56452*** (0,01171)	0,57718*** (0,01119)	0,59961*** (0,01128)	0,62453*** (0,01146)	0,56556*** (0,01171)
	Distância Geográfica ( $\beta_1$ )	-0,00068*** (0,00002)	-0,00075*** (0,00002)	-0,00071*** (0,00002)	-0,00073*** (0,00002)	-0,00077*** (0,00010)	-0,00083*** (0,00002)
	Distância Institucional ( $\beta_2$ )	0,06398*** (0,04710)	0,03475*** (0,04102)	0,13497*** (0,03725)	0,09068*** (0,03516)	0,02596*** (0,03403)	0,12313*** (0,03586)
	Constante ( $\alpha_0$ )	-3,4376*** (0,14343)	-3,8969*** (0,13169)	-4,0241*** (0,13119)	-4,2271*** (0,13833)	-4,5902*** (0,14309)	-3,3780*** (0,14616)
	Heterogeneidade ( $\alpha$ )	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	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.

**Figura 6. Efeito da Distância Geográfica sobre a Probabilidade de Colaboração Científica no Triênio 2007-2009, por Área do Conhecimento**



## 6. Conclusões

A tentativa de compreensão dos padrões espaciais da produção e colaboração científica no Brasil entre 1990 e 2010 evidenciou alguns resultados 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, 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 nas regiões Sudeste e Sul, com destaque às capitais dos estados. Contudo, foram encontradas evidências de que o processo de crescimento da produção científica no País foi caracterizado pela desconcentração espacial, em que os municípios de produção intermediária ganharam destaque.

Quanto ao papel desempenhado pela distância geográfica na articulação das redes de colaborações científicas, os resultados evidenciam o fato de que esta ainda é determinante na interação entre os pesquisadores brasileiros, uma vez que o aumento de 100 km de distância entre dois pesquisadores implica na redução, em média, de 16% da probabilidade de haver colaboração. Por fim, verificamos que o efeito da distância varia

entre redes das diferentes áreas do conhecimento e não foram encontradas evidências de que seu efeito tenha diminuído com o passar do tempo.

Sob a perspectiva de estudos futuros, pretendemos analisar as métricas individuais dos componentes das redes espaciais de colaboração científica, de maneira a identificar os principais municípios responsáveis pelas transformações estruturais das redes, e verificar a existência de possíveis direcionamentos sistemáticos dos fluxos de conhecimento para determinadas regiões. Tais informações poderiam ser úteis, na medida em que poderiam contribuir significativamente para a melhor compreensão da importância da colaboração científica para os propósitos de política científica e tecnológica no país, principalmente quanto aos objetivos de desconcentração regional da atividade científica e desenvolvimento regional.

## Referências

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.

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.

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

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.

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.

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.

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.

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

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

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.
- 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.
- 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.
- LETA, Jacqueline. *Indicadores de desempenho, ciência brasileira e a cobertura das bases informacionais.* **Revista USP**, v. 89, p.72-77, 2011.
- LETA, Jacqueline *et al.* *Science in Brazil. Part 2: Sectoral and institutional 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.
- LONG, J. Scott; FREESE, Jeremy. *Regression models for categorical dependent variables using Stata.* College Station: Stata Press, 2001.
- MALI, Franc *et al.* *Dynamic scientific co-authorship networks.* In: SCHARNHORST, Andrea *et al* (Eds.). **Models of science dynamics.** Berlin: Springer-Verlag, 2012, p. 195-232.
- 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.
- MAGGIONI, Mario A.; UBERTI, Teodora E. *Networks and geography in the economics of knowledge flows.* **Quality & Quantity**, v. 45, p. 1031-1051, 2011.
- 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*).

- 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.
- PONDS, Roderik *et al.* *The geographical and institucional proximity of research collaboration*. **Papers in Regional Science**, v. 86, n.3, p. 423-444, 2007.
- 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; HU, Yuanjia. *Collaborative knowledge production in China: regional evidence from a gravity model approach*. **Regional Studies**, v.45, n.6, p. 755-772, 2011.
- SEN, Ashish K.; SMITH, Tony E. *Gravity Models of Spatial Interaction Behavior*. New York: Springer, 1995.
- TER WAL, Anne L. J. *Networks and geography in the economics of knowledge flows: a commentary*. **Quality & Quantity**, v. 45, p. 1059-1063, 2011.
- 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.
- 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.