

# A distância geográfica ainda importa para a formação de redes científicas no Brasil?

Luís Fabiano Farias Borges<sup>1</sup>

## Resumo:

Este estudo analisa se a distância geográfica influenciou a formação de redes científicas no Brasil entre 2010 e 2019. O efeito geográfico sobre as colaborações científicas foi estimado por modelos gravitacionais a partir de dados extraídos da Plataforma Lattes, cujos resultados revelam que os parâmetros de proximidade geográfica e institucional foram estatisticamente significativos para as redes científicas. Após construir uma ampla rede científica com mais de 1.200 municípios brasileiros, observou-se que o fluxo de conhecimento ainda está concentrado nas regiões Sul e Sudeste do país. A concentração regional pode ser verificada pela formação de clusters, visto que 105 municípios concentraram mais de 80% dos artigos e das colaborações científicas durante o período em estudo. Os resultados sugerem a necessidade de discussões sobre políticas científicas, de modo a buscar maior coesão regional sob a perspectiva da cientometria espacial.

**Palavras-chave:** Redes Científicas; Modelos Gravitacionais; Plataforma Lattes; Cientometria Espacial.

## Abstract:

This study analyses if geographical distance influenced the formation of scientific networks in Brazil from 2010 to 2019. Geographical effect on scientific collaborations was estimated by gravity models through data extracted from the Lattes Platform, whose results reveal that geographical and institutional proximities are statistically significant in scientific networks. After constructing a broad scientific network with more than 1,200 municipalities, it is observed that the knowledge flow is still concentrated on Southeastern and Southern Brazil. The regional concentration can be verified by cluster formations, insofar as 105 municipalities concentrated more than 80% of papers and scientific collaborations during the period in analysis. The results suggested the importance of discussions on scientific policies in order to achieve more regional cohesion from the perspective of spatial scientometrics.

**Keywords:** Scientific Networks; Gravity Models; Lattes Platform; Spatial Scientometrics.

**JEL CLASSIFICATION:** C31, C35, O38

**Área 10** - Economia Regional e Urbana

---

<sup>1</sup> Chefe de Divisão de Monitoramento de Resultados da Diretoria de Relações Internacionais da CAPES e doutorando do Programa de Pós-Graduação em Modelagem Computacional de Sistemas da Universidade Federal de Tocantins.

## INTRODUÇÃO

Poucas invenções foram baseadas na ciência antes do século XIX, o que atesta a complexidade do trinômio CT&I (ciência, tecnologia e inovação). A crescente influência da ciência sobre a tecnologia levou à concepção equivocada de que tecnologia seria sinônimo de pesquisa aplicada em oposição à pesquisa básica (FREEMAN, 1977). Langrish (1974) sugere a presença de ciclos ao tratar da relação entre ciência e tecnologia. Sob sua perspectiva, a tecnologia teria contado pouco com a teoria científica na primeira metade do século XIX, ainda que técnicas científicas tenham sido empregadas. A segunda metade do século XIX teria sido caracterizada por uma série de *inputs* científicos que visavam à solução de problemas da sociedade. Por fim, a primeira metade do século XX teria sido caracterizada pelo triunfo dos métodos empíricos na indústria.

Com a internet das coisas, houve uma sensação difusa de que todas as localidades estão interconectadas. É fato notório que inovações tecnológicas tendem a reduzir custos de transação, mas o processo inovativo também guarda relação com outros fatores relevantes. A Escola Francesa de Proximidade contribui para essa discussão ao tratar dos efeitos da proximidade em atividades socioeconômicas, considerando a literatura sobre geografia econômica evolucionária e sistema regional de inovação, conforme discutem os autores Carrincazeaux, Lung e Vicente (2008).

Na Europa, os *Framework Programmes* viabilizam projetos colaborativos em pesquisa e desenvolvimento (P&D) capazes de integrar os sistemas regionais de inovação e de promover sinergia entre ciência e tecnologia (SCHERNGELL; BARBER, 2011; SIDONE, 2018). No Brasil, estudos sobre sistemas regionais de inovação revelam que os centros mais avançados estão localizados em São Paulo, Rio de Janeiro, Santa Catarina e Rio Grande do Sul (ROCHA; FERREIRA, 2004), mas o país ainda carece de instrumentos internos robustos de integração de P&D.

Em cientometria espacial, a pesquisa científica pode ser discutida com base em três esferas institucionais: empresa, governo e universidade (PONDS; VAN OORT; FRENKEN, 2007). Nesse contexto, destacam-se os *spillovers* de conhecimento (efeito transbordamento), cuja perspectiva supõe que a produção acadêmica e científica pode servir como insumo para inovação e ser absorvida pelo setor privado. No caso brasileiro, De Castro Araújo e García (2019) demonstraram que *spillovers* inter-regionais estão associados à inovação, na medida em que a proximidade entre microrregiões inovadoras afeta positivamente a inovação local, considerando a *proxy* de patentes.

O crescimento da comunidade científica do Brasil é refletido pela quantidade de publicações da comunidade acadêmica (GLÄNZEL; LETA; THIJS, 2006; LETA; CHAIMOVICH, 2002), o que demonstra que as universidades públicas são o centro de pesquisa no país, e não o setor privado. Em virtude da concentração espacial dos *spillovers*, estudos sobre proximidade geográfica em cientometria podem trazer elementos para discussões sobre a interação entre empresa, governo e universidade.

Observa-se aumento do perfil colaborativo na produção científica mundial (GLÄNZEL; SCHUBERT, 2005; SIDONE, 2018), ainda que áreas do conhecimento apresentem diferenças significativas quanto ao número de coautores e à frequência de publicações (ELSEVIER, 2018). Estudos sobre redes de colaborações estão inseridas nesse contexto, cuja *proxy* mais utilizada é a coautoria (JEONG; CHOI; KIM, 2013;

LEYDESDORFF; BORNMANN; WAGNER, 2019; SIDONE, 2018; WAGNER; JONKERS, 2017).

Não obstante, a aferição de colaboração científica pela presença de coautores deve ser analisada com cautela, pois há casos de colaborações que não são revertidas em coautoria de *papers*. Ademais, colaborações científicas nem sempre são classificadas de forma clara, como, por exemplo, duas instituições domésticas que colaboram com uma instituição estrangeira, pois implica tanto colaboração *internacional* quanto *intranacional* (KATZ; MARTIN, 1997). Dentro de um país, as redes científicas são similares porque um município pode interagir com outros (*inter*) ou pode haver interações dentro do próprio município (*intra*). Essa complexa rede de colaborações científicas pode ser analisada por efeitos geográficos (ACOSTA et al., 2011; KATZ, 1994; SCHERNGELL; HU, 2011; SIDONE; HADDAD; MENA-CHALCO, 2016) e por grafos computacionais, nos quais os nós são os pesquisadores e as coautorias são as arestas (CHEN; ZHANG; FU, 2019; MENA-CHALCO et al., 2014; YE; WU; WANG, 2008).

A literatura enfatiza que países com grandes dimensões territoriais tendem a estabelecer colaboração nacional, enquanto países menores são mais propensos à colaboração internacional (CHINCHILLA-RODRÍGUEZ et al., 2010; DAVIDSON FRAME; CARPENTER, 1979; ROYAL SOCIETY, 2011; SIDONE, 2018). Em contrapartida, existem outros fatores significativos, como, por exemplo, a quantidade de pesquisadores que os países dispõem (FRENKEN; HARDEMAN; HOEKMAN, 2009). Há crescente interesse por essa literatura em países que apresentam alta colaboração nacional, a exemplo do Brasil (MENA-CHALCO et al. 2014; HADDAD; MENA-CHALCO; SIDONE, 2017; SIDONE, HADDAD; MENA-CHALCO, 2016) e da China (WANG et al., 2005).

Ao discutir proximidade e inovação no âmbito da geografia econômica, Boschma (2005) define cinco dimensões de proximidade: geográfica, institucional, social, organizacional e cognitiva. Efeito geográfico sobre a produção de conhecimento científico é assunto pouco explorado na literatura, mas estudos de cientometria espacial contribuem sobremaneira para a discussão por permitirem separar os efeitos dessas dimensões de proximidade sobre a interação entre pesquisadores na rede colaborativa (SIDONE; HADDAD; MENA-CHALCO, 2016). Em análise estatística de regressões, modelos gravitacionais permitem utilizar dimensões de proximidade como variável explicativa (e.g distância geográfica) e as interações científicas como variável resposta (e.g colaborações científicas), (FRENKEN; HARDEMAN; HOEKMAN, 2009; HOEKMAN; FRENKEN; TIJSSEN, 2010; PONDS; VAN OORT; FRENKEN, 2007).

Sidone, Haddad e Mena-Chalco (2016) investigaram ampla rede científica no Brasil, com base em dados extraídos da Plataforma Lattes entre 1992 e 2009. Os pesquisadores realizaram extração de 1.131.912 currículos, totalizando 7.351.957 materiais acadêmicos de graduandos, mestres e doutores. Nesse sentido, a presente pesquisa tem como objetivo complementar o período analisado por Sidone, Haddad e Mena-Chalco (2016), considerando uma análise espacial de 2010 a 2019 (cinco biênios). Para tanto, serão analisados apenas os artigos científicos publicados por doutores para compor a rede, com base nas dimensões de proximidade geográfica e institucional. Desse modo, os resultados desta pesquisa poderão auxiliar futuros estudos sobre sistemas regionais de inovação.

Além desta introdução, esta pesquisa está estruturada nas seguintes seções: métodos (modelo empírico e base de dados), resultados, discussão e conclusão.

## MÉTODOS

### Modelo empírico

Grafos podem ser representados por matriz de adjacência, na qual os nós da rede são constituídos por linhas e colunas de dimensão  $n \times n$ . As redes de colaborações científicas são grafos não direcionados e apresentam arestas ponderadas, as quais determinam o valor da intensidade de colaborações entre pesquisadores (MENA-CHALCO et al., 2014; NEWMAN, 2001). Ao associar pesquisadores e unidade geográfica, a matriz de adjacência é constituída pelos totais de colaboração intermunicipal das unidades geográficas  $i$  e  $j$  e pelos totais de colaborações intramunicipais na diagonal principal. As distâncias intermunicipais  $u_{ij}^{(1)}$  foram calculadas em quilômetros (km), com o auxílio do índice espacial hierárquico hexagonal da biblioteca H3 do Python (UBER, 2020). As distâncias intramunicipais  $u_{ii}^{(1)}$ , por sua vez, seguiram o método proposto por Bröcker (1988):

$$\left(\frac{2}{3}\right) \left(\frac{A}{\pi}\right)^{0.5}$$

Os modelos de interação espacial descrevem funções que caracterizam a origem  $i$  e o destino  $j$  da interação, bem como a separação entre duas regiões  $i$  e  $j$ . O modelo teórico geral é representado pela seguinte equação:

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

$Y_{ij}$  é a variável dependente estocástica concernente à matriz dos fluxos de colaboração científica entre as regiões  $i$  e  $j$ .  $X_{ij}$  é uma função que capta a relação estocástica entre outras variáveis aleatórias a partir de uma distribuição de probabilidade. Por fim,  $\varepsilon_{ij}$  é o termo de erro entre todos os pares inter-regionais  $i$  e  $j$ .  $X_{ij}$  é especificado como função de covariáveis e corresponde ao modelo gravitacional geral, pois o objetivo é mensurar a importância relativa das regiões (origem e destino), bem como a distância na determinação dos fluxos de colaborações científicas:

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

As funções  $O_i$  e  $D_j$  caracterizam as interações das regiões origem  $i$  e destino  $j$ , podendo ser descritas por *power functions* de acordo com a interação espacial clássica (SEN; SMITH, 1995). A especificação do termo de separação espacial  $S(u_{ij})$  ocorre na forma funcional exponencial multivariada, permitindo que outras dimensões de proximidade sejam consideradas no modelo simultaneamente. Assim, as funções tomam as seguintes formas:

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

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

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

As variáveis  $o_i$  e  $d_j$  medem características específicas das regiões  $i$  e  $j$ , e a variável  $u_{ij}^{(k)}$  representa  $k$  medidas da separação espacial entre as regiões  $i$  e  $j$ . Os termos  $\alpha_1$  e  $\alpha_2$  são parâmetros a serem estimados, e o termo  $\beta_k$  refere-se ao conjunto de  $k$  parâmetros associados a cada uma das  $k$  medidas de separação entre as regiões  $i$  e  $j$ . Considera-se o caso especial  $\alpha_1 = \alpha_2$  porque as variáveis de origem e de destino são simétricas e a direção do fluxo das colaborações científicas entre regiões é um grafo não direcionado. Os parâmetros  $\alpha_1$ ,  $\alpha_2$  e  $\beta_k$  podem ser interpretados como elasticidades de  $y_{ij}$  com relação às variáveis de origem  $o_i$ , destino  $d_j$  e separação  $u_{ij}^{(k)}$  (SCHERNGELL; HU, 2011).

Ao substituir as equações do modelo inicial, obtém-se o seguinte modelo empírico a ser estimado:

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

As variáveis de origem ( $o_i$ ) e destino ( $d_j$ ) são mensuradas pelo total de publicações científicas de cada município. Espera-se que as colaborações entre as municipalidades  $i$  e  $j$  dependam positivamente do total de publicações de cada município. Como o número de publicações em uma região aumenta a probabilidade de colaboração com outras regiões, a estimativa para os termos de massa tende a aproximar-se da unidade ( $\alpha_1 = \alpha_2$ ).

Foram utilizados dois parâmetros para as variáveis de separação. O primeiro parâmetro ( $\beta_1$ ) consiste na construção de uma matriz de distância geográfica na qual cada elemento  $u_{ij}^{(1)}$  é calculado pela distância em quilômetros (km) entre dois municípios  $i$  e  $j$ . O segundo parâmetro ( $\beta_2$ ) mede a relação institucional entre os municípios, conforme método descrito por Sidone, Haddad, and Mena-Chalco (2016). Ao construir a matriz institucional para os municípios, impõe-se o valor  $u_{ij}^{(2)} = 1$  para os municípios  $i$  e  $j$  se ambos apresentarem universidade pública (sede federal ou estadual), caso contrário  $u_{ij}^{(2)} = 0$ . O critério do campus sede seguiu os dados do censo do INEP<sup>2</sup>.

Os modelos de contagem podem ser utilizados para problemas de especificação em modelos de interação espacial, em vez do método de Mínimos Quadrados Ordinários, uma vez que não é adequada a especificação log-normal na equação 6 (LONG; FREESE, 2001; SIDONE; HADDAD; MENA-CHALCO, 2016). Problemas de especificação log-normal podem ser evitados ao recorrer a modelos de dados de contagem, assumindo que o processo de geração de dados produz apenas números inteiros e não negativos. Assim, o número de colaborações científicas pode seguir a distribuição de Poisson de acordo com a seguinte expressão:

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

O conjunto de variáveis independentes do modelo empírico 6 é representado por:

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

A função de distribuição de Poisson apresenta a propriedade estatística da equidispersão, caracterizada pela igualdade entre variância e média condicionais. O modelo pode ser estimado pelo método de máxima verossimilhança padrão, caso os

<sup>2</sup> <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>.

dados observados sejam gerados a partir de um processo de Poisson. Se a contagem das colaborações científicas regionais não for compatível com a distribuição de Poisson padrão, não será satisfeita a propriedade de equidispersão e o processo de Poisson será violado. A superdispersão gera viés na estimação e invalida o teste de hipótese (CAMERON; TRIVEDI, 2013; HILBE, 2011; WINKELMANN, 2008).

A Binomial Negativa é alternativa comum em estudos empíricos sobre colaborações científicas (HOEKMAN; FRENKEN; TIJSSEN, 2010; SCHERNGELL; BARBER, 2011; SCHERNGELL; HU, 2011; SIDONE; HADDAD; MENA-CHALCO, 2016), pois lida com a heterogeneidade não observada mediante a inclusão do parâmetro  $\alpha$ . A expressão de densidade da distribuição binomial negativa (9) e a variância condicional (10)<sup>3</sup> são as seguintes:

$$\Pr (Y_{ij} = y_{ij} | X_{ij}) = \frac{\Gamma(y_{ij} + \alpha^{-1})}{y_{ij}! \Gamma(\alpha^{-1})} \cdot \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \cdot \left( \frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{y_{ij}} \quad (9)$$

$$V [Y_{ij}|X_{ij}] = \mu_{ij} + \alpha \mu_{ij}^2; \quad \mu_{ij} = E [Y_{ij}|X_{ij}] \quad (10)$$

Onde  $\Gamma(\cdot)$  é a função gamma, e  $\alpha$  é o parâmetro de heterogeneidade.

## Base de dados

O *curriculum vitae* de graduandos, mestres e doutores que atuam no Brasil pode ser acessado pela Plataforma Lattes (<http://lattes.cnpq.br>). Não obstante, é necessário mencionar limitações quanto aos dados da Plataforma Lattes. Primeiramente, não há acesso direto à base completa, o que demanda ferramentas computacionais para obter e processar volumosa quantidade de dados sobre pesquisadores. Em segundo lugar, não é possível quantificar colaboração internacional. Como a coautoria é processada pela identificação de currículos distintos nos pares de colaboração, não é possível contabilizar publicação feita por um pesquisador brasileiro em colaboração com estrangeiros que atuam no exterior. Em terceiro lugar, não é possível captar o efeito da migração de pesquisadores pelo território nacional.

O desafio computacional na extração de dados sobre pesquisadores brasileiros resultou na criação do *scriptlattes*<sup>4</sup> (MENA-CHALCO; CESAR JUNIOR, 2009), cujos procedimentos de extração de dados foram detalhados por Digiampietri et al. (2012) e Mena-Chalco et al. (2014). Esses procedimentos permitem identificar publicações acadêmicas, Lattes ID, endereço profissional dos pesquisadores, entre outras informações relevantes. A partir dessas extrações de dados, estudos propuseram métodos de análise da rede científica brasileira mediante grafos e modelos gravitacionais (HADDAD; MENA-CHALCO; SIDONE, 2017; MENA-CHALCO et al., 2014; SIDONE; HADDAD; MENA-CHALCO, 2016). Para gerar ampla base de dados dos currículos Lattes, é necessário fazer adaptações do *scriptlattes* (código aberto).

A agregação por unidades geográficas é comum em estudos de cientometria espacial (FRENKEN; HARDEMAN; HOEKMAN, 2009), o que permite utilizar cidades como nós das redes colaborativas (PAN; KASKI; FORTUNATO, 2012). Nesta pesquisa, o Lattes ID e o endereço profissional declarado no Lattes foram utilizados

<sup>3</sup> A equação 10 refere-se à especificação Negbin II (LONG; FREESE, 2001).

<sup>4</sup> Informação disponível em: <http://scriptlattes.sourceforge.net>.

como *proxies* para localizar os coautores, conforme método descrito por Mena-Chalco e César Júnior (2014); Sidone, Haddad e Mena-Chalco (2016); e Haddad, Mena-Chalco e Sidone (2017). Foi adotado o tradicional método de contagem completa (*full-counting*) de coautorias, assumindo que cada nó da rede recebe uma unidade pela participação em determinada publicação científica (SCHERNGELL; BARBER, 2011).

A base de dados inicial contempla 232.966 doutores ativos<sup>5</sup>, distribuídos em 3.187 municípios distintos. Foi selecionada amostra reduzida de 105 municípios para comparar os biênios na modelagem, pois a matriz completa dos municípios seria inviável para a estimação dos modelos de contagem, em razão da quantidade excessiva de valores nulos.

Foram analisadas oito grandes áreas do conhecimento de acordo com a classificação da Plataforma Lattes: Ciências Agrárias, Ciências Biológicas, Engenharias, Ciências Exatas e da Terra, Ciências Humanas, Ciências da Saúde, Ciências Sociais Aplicadas e Linguística, Letras e Artes. Para construir matrizes de coautorias intermunicipais por área de conhecimento, foi necessário associar pesquisador, município e grande área do conhecimento, de acordo com a declaração registrada pelo pesquisador na Plataforma Lattes.

A rede científica nacional seguiu o modelo gravitacional, cujos métodos são comuns em estudos de ciencometria espacial (ACOSTA et al., 2011; FRENKEN; HARDEMAN; HOEKMAN, 2009; HOEKMAN; FRENKEN; TIJSEN, 2010; SCHERNGELL; BARBER, 2011; SCHERNGELL; HU, 2011; SIDONE; HADDAD; MENA-CHALCO, 2016).

## ANÁLISE E RESULTADOS

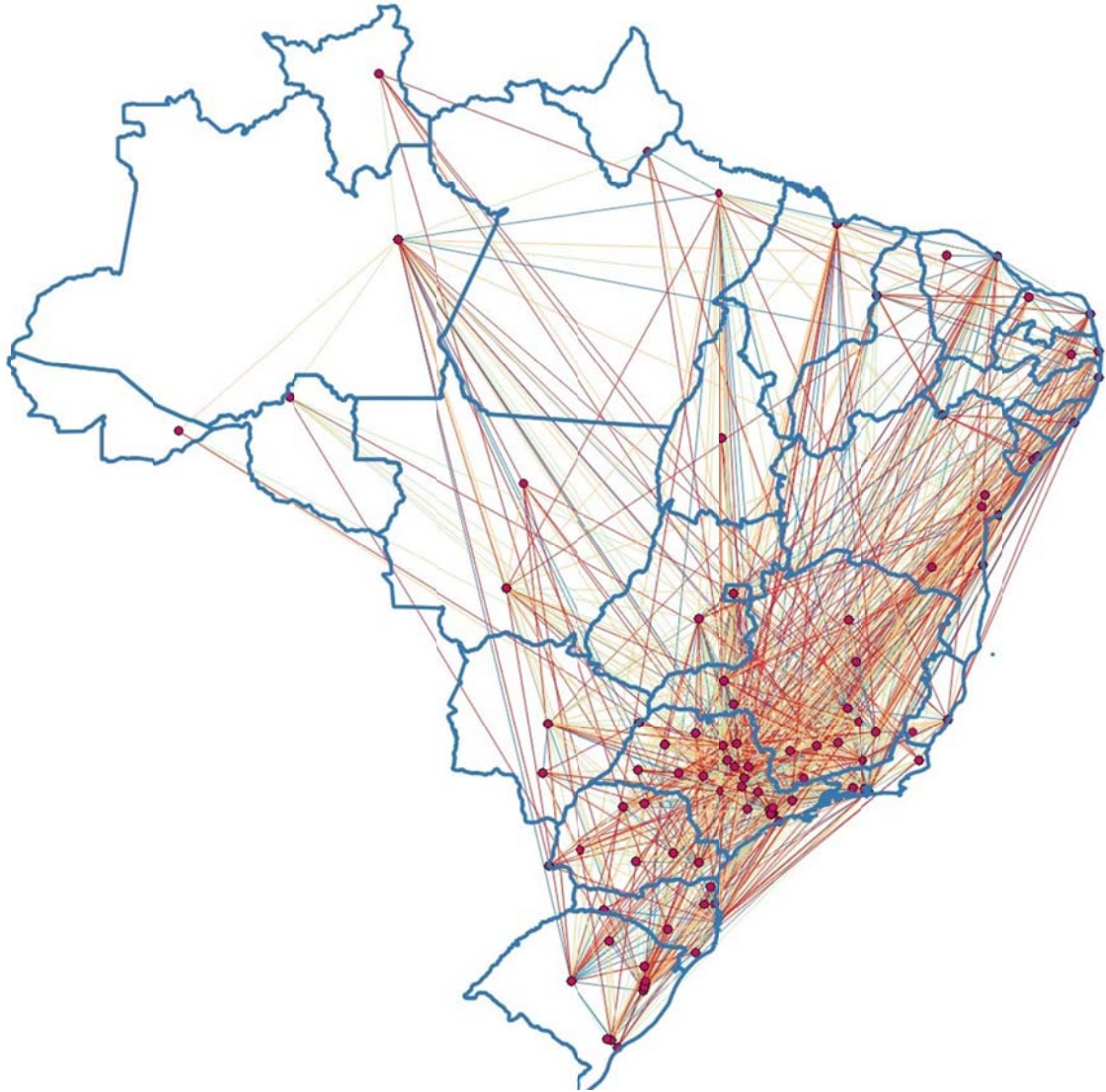
Foi utilizada uma base de dados com 232.966 doutores ativos (distribuídos em 3.187 municípios) para construir a rede de colaborações científicas. Entretanto, 51.155 pesquisadores não declararam o endereço profissional e apenas 173.584 pesquisadores declararam tanto o endereço profissional quanto a grande área do Lattes. Após testar a existência dos pares de colaboração, foram identificados 136.939 doutores ativos distribuídos em 1.271 municípios, o que corresponde ao dado estabilizado. Foram detectadas 1.636.031 publicações, mas apenas 694.159 possuíam número de autores maior do que 1. O total de publicações (com no mínimo um par de doutores ativos) foi de 602.170, e o número de colaborações corresponde a 2.590.497. Para estimar o modelo de dados de contagem, foram considerados 105 municípios porque concentraram mais de 80% das publicações e das colaborações científicas.

Foram detectadas 2.090.404 colaborações científicas (80,69%), considerando 969.512 colaborações intermunicipais e 1.120.892 intramunicipais. A partir desses dados, foram averiguados os principais fluxos de conhecimento científico, conforme Figura 1<sup>6</sup>:

---

<sup>5</sup> Doutores ativos referem-se aos pesquisadores que atualizaram o currículo na Plataforma Lattes nos últimos 2 anos, cujos dados foram obtidos por intermédio de *web scraping* desenvolvido em linguagem Python. Para obter os currículos em formato XML de forma automatizada, é necessário ter acesso ao IP institucional.

<sup>6</sup> Foram utilizados os softwares Gephi 0.9.2 e QGIS 3.18.1 para elaborar grafos e mapas.



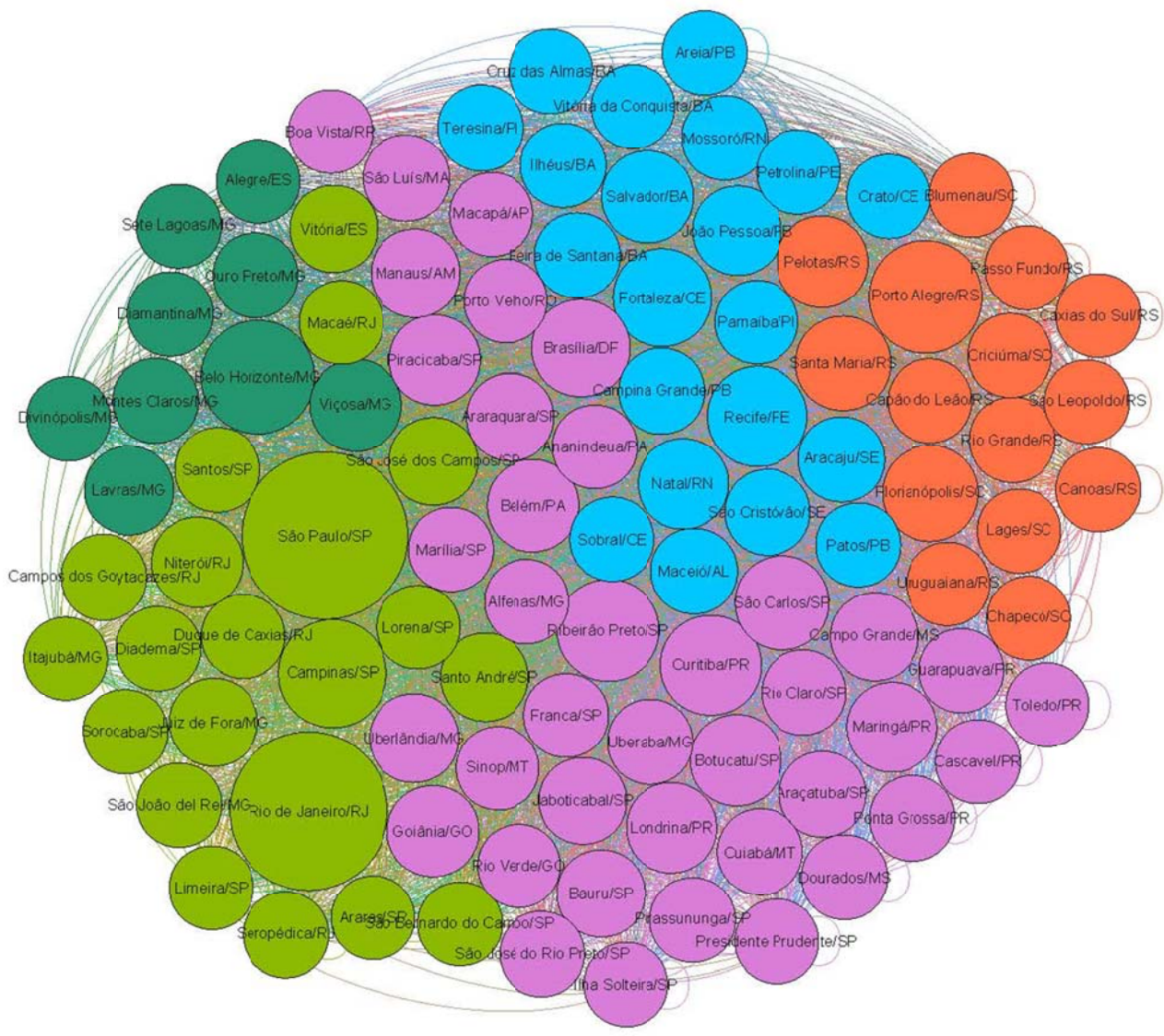
**Figura 1.** Grafo das principais colaborações científicas *intra* e *intermunicipais*.

Observa-se que o fluxo de conhecimento científico conecta diversas regiões do país. Contudo, é possível identificar colaborações científicas *intra* e *intermunicipais* concentradas em certas localidades, com destaque para Rio de Janeiro/RJ e São Paulo/SP. Tendo em vista que certos municípios não apareceram em todos os biênios analisados, foi necessário verificar a formação de *clusters* dos principais municípios em colaboração científica de acordo com a Figura 2<sup>7</sup>:

---

<sup>7</sup> O grafo não direcionado foi gerado com 105 nós e 5.209 arestas no Gephi. No apêndice desta pesquisa, há a relação dos 105 municípios principais em colaboração científica com os respectivos graus e modularidade de classe.





**Figura 2.** Grafo dos 105 municípios principais em rede de colaboração científica.

Foi identificada a seguinte quantidade de municípios por região na rede científica: Centro-Oeste (7), Norte (6), Nordeste (22), Sul (22) e Sudeste (48). Entre as localidades que apresentaram apenas 1 município como nó da rede, 4 são da região Norte, 2 são da região Nordeste e mais o Distrito Federal. Os estados que apresentaram maior quantidade de municípios na rede científica foram respectivamente: São Paulo (26), Minas Gerais (14), Rio Grande do Sul (10) e Paraná (7).

Observou-se também a ascensão do Nordeste, na medida em que a região apresentou a mesma quantidade de municípios do Sul (22), com destaque para a Bahia com 5 municípios na rede. Ressalta-se que alguns municípios não estiveram presentes em todo o período, mas apareceram em alguns biênios analisados, a exemplo de Marechal Cândido Rondon-PR, Governador Valadares-MG, Santarém-PA, Santo Antônio do Descoberto-GO e Palmas-TO.

Em média, houve colaboração científica em distâncias cada vez maiores entre 2010 e 2019, considerando a produção científica agregada. Como é esperado, porém, as

oito grandes áreas de conhecimento da Plataforma Lattes apresentaram distintos comportamentos nas colaborações científicas intermunicipais. As colaborações científicas das áreas de Ciências Humanas, Ciências Biológicas e Ciências da Saúde apresentaram aumento em redes mais distantes geograficamente, considerando que as faixas médias de distâncias variaram de 600 a 900 km.

**Tabela 1.** Distância média das colaborações científicas intermunicipais por área de conhecimento.

Área do conhecimento	Faixa de quilometragem em 2010	Faixa de quilometragem em 2019
Ciências Agrárias	800 km	800 km
Ciências Humanas	700 km	800 km
Ciências Biológicas	800 km	900 km
Ciências da Saúde	700 km	800 km
Engenharias	600 km	600 km
Ciências Sociais Aplicadas	700 km	700 km
Ciências Exatas e da Terra	700 km	600 km
Linguística, Letras e Artes	800 km	800 km

Fonte: extração de dados da Plataforma Lattes.

Quanto ao número de colaborações científicas, a área de Ciências Agrárias foi estável. Ciências Biológicas, Ciências da Saúde e Ciências Exatas e da Terra apresentaram o maior número de colaborações, o que pode explicar o aumento médio da distância na produção científica agregada. Linguística, Letras e Artes, Ciências Humanas e Ciências Sociais Aplicadas apresentaram o menor número de colaborações científicas, o que é esperado em virtude da característica inerente a essas áreas do conhecimento. Cumpre destacar que a área de Linguística, Letras e Artes apresentou o menor número de colaborações.

Os modelos de dados de contagem (Poisson e Binomial Negativa) indicam que o aumento da distância entre dois pesquisadores reduz a probabilidade de haver colaboração, conforme resultados da Tabela 2<sup>8</sup>:

<sup>8</sup> O software Stata 13.0 foi utilizado para os testes estatísticos.

**Tabela 2.** Estimativas dos modelos de dados de contagem para colaborações *intra* e *intermunicipais*.

		2010-2011	2012-2013	2014-2015	2016-2017	2018-2019
Poisson	Origem-destino	0,7677543***	0,9573585***	0,8628939***	0,9347968***	0,979404***
	( $\alpha_1=\alpha_2$ )	(0,0014993)	(0,0014305)	(0,0012673)	(0,0013362)	(0,0014129)
	Distância	-0,0020095***	-0,0019822***	-0,0019521***	-0,0022075***	-0,0019132***
	geográfica ( $\beta_1$ )	(5,14e-06)	(4,15e-06)	(3,90e-06)	(4,20e-06)	(3,58e-06)
	Distância	0,7520476***	0,7352618***	0,7263429***	0,7467253***	0,7989347***
	institucional ( $\beta_2$ )	(0,0045654)	(0,0039858)	(0,003703)	(0,0036757)	(0,0036608)
	Constante	-7,390428***	-7,264539***	-7,651489***	-7,116752***	-7,57211***
		(0,0199226)	(0,0165451)	(0,0159488)	(0,015768)	(0,0154468)
Binomial	Origem-destino	0,5917468***	0,6803572***	0,6461779***	0,6858878***	0,6617752***
	( $\alpha_1=\alpha_2$ )	(0,0000178)	(0,0214138)	(0,0202555)	(0,0239089)	(0,0211068)
	Distância	-0,0009116***	-0,0009504***	-0,000913***	-0,001018***	-0,0009325***
	geográfica ( $\beta_1$ )	(0,0000178)	(0,0000177)	(0,0000163)	(0,0000182)	(0,0000158)
	Distância	0,5091922***	0,5389609***	0,486361***	0,4820371***	0,5791102***
	institucional ( $\beta_2$ )	(0,0416911)	(0,0447524)	(0,0407158)	(0,0454206)	(0,0404023)
	Constante	-5,687193***	-6,070828***	-6,217018***	-5,933315***	-6,099909***
		(0,1863778)	(0,206314)	(0,1954895)	(0,2169725)	(0,2026092)
Negativa	Heterogeneidade	1,257136**	1,350501***	1,389171***	1,484337***	1,373415***
	( $\alpha$ )	(0,026071)	(0,0269265)	(0,026585)	(0,0290624)	(0,025973)
Número de observações		4.027	4.181	4.511	4.171	4.585

Nota: os parênteses indicam o erro-padrão. \*\*\*0.001 de significância estatística.

Os resultados dos biênios mantiveram similaridade e foram estatisticamente significativos. A distância geográfica (captada pelo parâmetro  $\beta_1$ ) foi negativa em Poisson e na Binomial Negativa, indicando que maior colaboração científica está associada à menor distância geográfica. A distância institucional (captada pelo parâmetro  $\beta_2$ ) foi positiva em Poisson e na Binomial Negativa, indicando que maior colaboração científica está associada à maior proximidade institucional. Nos biênios analisados, o parâmetro  $\beta_2$  oscilou entre 0,48 e 0,57 na Binomial Negativa, demonstrando a consistência do campus sede como critério de proximidade institucional.

O resultado dos termos de massa (origem e destino) apresentou melhor especificação em Poisson, aproximando-se mais da unidade. Contudo, a Binomial Negativa apresentou melhor especificação do modelo geral, pois a heterogeneidade foi captada pelo parâmetro  $\alpha$  em todos os biênios. Desse modo, evidenciou-se a superdispersão por meio do teste de razão de verossimilhança para  $\alpha = 0$ , com p-valor significativo (Prob>=chibar2 = 0.000). O parâmetro  $\beta_1 = -0,0009325$  da Binomial Negativa indica que, em média, há redução de 8,90% na colaboração científica a cada 100 km entre dois pesquisadores, mantendo-se as demais variáveis constantes.

## DISCUSSÃO

A produção científica mundial caracteriza-se por um processo de desconcentração geográfica (GROSSETTI et al., 2012). Sidone, Haddad e Mena-Chalco (2016) encontraram evidências desse processo na produção científica brasileira entre 1992 e 2009, ao observar a inserção de municípios com produção científica menos

expressiva. Os autores identificaram retraimento do processo de desconcentração a partir de 2007 e ressaltam que a proximidade geográfica constituiu fator significativo para as redes científicas brasileiras no período analisado, observando a concentração em torno de 105 municípios.

Ao mapear a rede de doutores ativos entre 2010 e 2019, os resultados desta pesquisa corroboram os achados de Sidone, Haddad e Mena-Chalco (2016), visto que a distância ainda influencia a rede de colaborações científicas no Brasil e a produção científica ainda se concentra em torno do mesmo número de municípios. Ainda que Sidone, Haddad e Mena-Chalco (2016) tenham recorrido à base de dados mais ampla por meio de distintas publicações bibliográficas de graduandos, mestres e doutores, o mapeamento desta pesquisa apresentou resultados análogos. Ademais, foi identificada a substituição gradual de certos municípios na seleta lista de 105 municípios em colaboração científica, ao analisar biênio por biênio.

Esses resultados talvez sejam explicados pelo fato de a rede científica ser conduzida pelos doutores ativos e também pelo fato de Sidone, Haddad e Mena-Chalco (2016) terem lidado com perda de informações, visto que o campo relativo à localização geográfica de alguns currículos não estava devidamente preenchido na Plataforma Lattes (43% de 1.131.192 currículos). Ressalta-se que — entre 2000 e 2009 — as universidades brasileiras da região Sul e Sudeste já concentravam cerca de 60 a 70% da produção científica indexada em periódicos internacionais. Em 2009, a Universidade de São Paulo (USP) foi responsável por aproximadamente 25% da produção brasileira (FAPESP, 2011; GLÄNZEL; LETA; THIJS, 2006; ROYAL SOCIETY, 2011; SIDONE, 2018).

Cabem oportunas discussões sobre os parâmetros do modelo empírico. A utilização do campus sede no parâmetro  $\beta_2$  (dimensão institucional) deve-se por dois motivos. Em primeiro lugar, o campus sede detém a maior parte da estrutura das universidades públicas e geralmente concentra a maior parte da produção científica, conquanto haja infraestrutura em diversos *campi* universitários espalhados pelo Brasil. Ao comparar universidades públicas e institutos federais, o problema torna-se mais complexo. Os *campi* dos institutos federais possuem estrutura similar se comparados às universidades. Entre as 661 unidades de institutos federais distribuídas entre as 27 unidades da Federação<sup>9</sup>, há mais de 1 milhão de alunos<sup>10</sup>. Entretanto, os *campi* do interior geralmente são os mais recentes e apresentam atividades acadêmicas voltadas à graduação e ao ensino técnico. Nesse sentido, o campus sede dos institutos federais reúne mais atributos, de forma análoga ao campus sede das universidades públicas. Não foi possível localizar arquivos com metadados com informações necessárias para fazer o georreferenciamento com vista a comparar o campus sede com os *campi* do interior. O aperfeiçoamento do censo da educação profissional no Brasil poderia suprir parte desses problemas informacionais.

Em segundo lugar, ainda que a maior parte das publicações científicas do Brasil provenha de universidades públicas, há publicações oriundas de universidades privadas e de outros centros de pesquisa. O caso mais emblemático analisado foi a Universidade de Sergipe, com campus sede em São Cristóvão. O fato de pesquisadores declararem Aracaju e São Cristóvão como endereço profissional no Lattes pode ser explicado pela presença de outros centros de pesquisa em Aracaju, uma vez que esses foram os dois

<sup>9</sup> <http://portal.mec.gov.br/rede-federal-inicial>.

<sup>10</sup> <http://plataformanilopecanha.mec.gov.br>.

únicos municípios de Sergipe que apareceram nos biênios analisados. Como consequência, a matriz binária institucional foi marcada com 0 nos pares de colaboração com Aracaju.

Conforme ressaltado na introdução desta pesquisa, a dimensão geográfica deve ser vista com precaução. A inclusão do parâmetro de dimensão institucional tem por finalidade ponderar o efeito geográfico no modelo empírico. Esse critério tem implicações sobre os termos de massa do modelo gravitacional, pois é esperado que as colaborações científicas entre duas localidades dependam do total de publicações de cada município.

## CONCLUSÃO

Nesta pesquisa, foram analisados os efeitos geográfico e institucional das redes científicas em um período de 10 anos. Métodos computacionais permitiram a extração de dados diretamente da Plataforma Lattes (CNPq), possibilitando o mapeamento de uma rede de colaborações com 136.939 pesquisadores distribuídos em 1.271 municípios brasileiros. Foram estabelecidos pares de municípios por intermédio de matrizes de adjacência ( $n \times n$ ) com vistas a analisar a intensidade das redes e a testar a probabilidade de colaborações em função da distância geográfica, de acordo com os modelos gravitacionais.

Os resultados revelam tendência de colaborações científicas intermunicipais em distâncias cada vez maiores em média, mas as áreas de conhecimento apresentam comportamentos distintos. Os resultados indicam que o aumento da distância média intermunicipal deve-se, em grande medida, às colaborações em Ciências Biológicas e Ciências da Saúde, áreas do conhecimento que apresentaram a maior quantidade de colaborações científicas. Ademais, foi possível identificar concentração de pesquisa científica em 105 municípios, cuja amostra permitiu a estimação por modelos de dados de contagem.

Os parâmetros do modelo estimado indicam que maior colaboração científica está associada à menor distância geográfica e à maior proximidade institucional. A Binomial Negativa revelou que, em média, há redução de 8,90% na colaboração nacional a cada 100 km de distância entre dois pesquisadores entre 2018 e 2019, mantendo-se as demais variáveis constantes (Tabela 2).

O Lattes ID e o endereço profissional declarado na Plataforma Lattes foram as *proxies* utilizadas para identificar os municípios como nós da rede de colaborações. Esse método é utilizado na cientometria espacial (FRENKEN; HARDEMAN; HOEKMAN, 2009; PAN; KASKI; FORTUNATO, 2012; SIDONE; HADDAD; MENA-CHALCO, 2016), cuja literatura mantém estreita relação com a geografia econômica. Nesse contexto, os dados viabilizaram a composição da rede nacional com base em artigos científicos publicados por doutores ativos entre 2010 a 2019, considerando oito grandes áreas da Plataforma Lattes (CNPq). Em contrapartida, foram enfatizadas três limitações ao extrair dados do Lattes: necessidade de ferramentas computacionais para obter e processar volumosa quantidade de dados, bem como a impossibilidade de quantificar colaboração internacional e de captar o efeito da migração de pesquisadores pelo território nacional.

Embora a geografia não seja o único fator de proximidade para a interação entre pesquisadores, a distância geográfica ainda exerce influência na composição das redes de colaboração científica no Brasil. Esta conclusão coaduna-se com os achados de Sidone, Haddad e Mena-Chalco (2016) para o contexto brasileiro e Hoekman, Frenken e Van Oort (2009) para o contexto europeu. Os resultados desta pesquisa cientométrica vão ao encontro de estudos sobre sistemas regionais de inovação, visto que é possível obter evidências acerca da concentração regional. Essa discussão fornece subsídios para formulação de políticas científicas, sobretudo por permitir a avaliação de eficiência das políticas regionais.

Sugere-se para futuros estudos: elaborar estudos bibliométricos que permitam convergência das áreas de conhecimento da Plataforma Lattes com as principais classificações internacionais, incluir outros parâmetros no modelo empírico e comparar os resultados da Plataforma Lattes com bases científicas indexadas em distintas formas de colaborações científicas, incluindo a colaboração internacional.

## Referências<sup>11</sup>

- ACOSTA, Manuel; CORONADO, Daniel; FERRÁNDIZ, Esther; LEÓN, M. Dolores. Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, [S. l.], v. 87, n. 1, p. 63–74, 2011. DOI: 10.1007/s11192-010-0305-6.
- BOSCHMA, Ron A. Proximity and Innovation: A Critical Assessment. *Regional Studies*, [S. l.], v. 39, n. 1, p. 61–74, 2005. DOI: 10.1080/0034340052000320887.
- BRÖCKER, J. How to eliminate certain defects of the potential formula. *Environment and Planning A*, [S. l.], v. 21, p. 817–830, 1988. DOI: 10.1068/a210817.
- CAMERON, A. Colin; TRIVEDI, Pravin K. *Regression Analysis of Count Data*. Second Edi ed. Cambrigde. DOI: <https://doi.org/10.1017/CBO9781139013567>.
- CARRINCAZEAUX, Christophe; LUNG, Yannick; VICENTE, Jerome. The scientific trajectory of the French School of Proximity: Interaction- and institution-based approaches to regional innovation systems. *European Planning Studies*, [S. l.], v. 16, n. 5, p. 617–628, 2008. DOI: 10.1080/09654310802049117.
- CHEN, Kaihua; ZHANG, Yi; FU, Xiaolan. International research collaboration: An emerging domain of innovation studies? *Research Policy*, [S. l.], v. 48, n. 1, p. 149–168, 2019. DOI: 10.1016/j.respol.2018.08.005. Disponível em: <https://doi.org/10.1016/j.respol.2018.08.005>.
- CHINCHILLA-RODRÍGUEZ, Zaida; VARGAS-QUESADA, Benjamín; HASSAN-MONTERO, Yusef; GONZÁLEZ-MOLINA, Antonio; MOYA-ANEGÓN, Félix. New approach to the visualization of international scientific collaboration. *Information Visualization*, [S. l.], v. 9, n. 4, p. 277–287, 2010. DOI: 10.1057/ivs.2009.31.
- DAVIDSON FRAME, J.; CARPENTER, Mark P. International Research Collaboration. *Social Studies of Science*, [S. l.], v. 9, n. 4, p. 481–497, 1979. DOI: 10.1177/030631277900900405.
- DE CASTRO ARAÚJO, Veneziano; GARCIA, Renato. Determinants and spatial dependence of innovation in Brazilian regions: evidence from a Spatial Tobit Model. *Nova Economia*, [S. l.], v. 29, n. 2, p. 375–400, 2019. DOI: 10.1590/0103-6351/4456.
- DIGIAMPIETRI, Luciano a; MENA-CHALCO, Jesús P.; PÉREZ-ALCÁZAR, José J.; TUESTA, Esteban F.; DELGADO, Karina V; MUGNAINI, Rogério; SILVA, Gabriela S. Minerando e Caracterizando Dados de Currículos Lattes. *I Brazilian Workshop on Social Network Analysis and Mining (BrasNAM)*, [S. l.], v. d, p. 12, 2012.
- ELSEVIER. *Research Metrics Guidebook*. [s.l: s.n.]. Disponível em: <https://www.elsevier.com/research-intelligence/resource-library/research-metrics-guidebook>.
- FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo. *Indicadores de ciência, tecnologia e inovação em São Paulo 2010*. [s.l: s.n.].
- FREEMAN, Christopher. *Economics of Research and Development*. London: Sage Publications, 1977.
- FRENKEN, Koen; HARDEMAN, Sjoerd; HOEKMAN, Jarno. Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, [S. l.], v. 3, n. 3, p. 222–232, 2009. DOI: 10.1016/j.joi.2009.03.005.
- GLÄNZEL, Wolfgang; LETA, Jacqueline; THIJS, Bart. Science in Brazil. Part 1: A macro-level comparative study. *Scientometrics*, [S. l.], v. 67, n. 1, p. 67–86, 2006. DOI: 10.1007/s11192-006-0055-7.

---

<sup>11</sup> Conforme ABNT (ECA-USP).

GLÄNZEL, Wolfgang; SCHUBERT, András. **Analysing Scientific Networks Through Co-Authorship**. In: Moed, H.F. et al. (Eds.). New York: Springer, 2005. DOI: 10.1007/1-4020-2755-9\_12.

GROSSETTI, Michel; ECKERT, Denis; GINGRAS, Yves; JÉGOU, Laurent; LARIVIÈRE, Vincent. The Geographical Deconcentration of Scientific Activities (1987-2007). **Proceedings of 17th International Conference on Science and Technology Indicators**, [S. l.], v. 1, n. January 2015, p. 348–356, 2012. Disponível em: [http://sticonference.org/Proceedings/vol1/Grossetti\\_Geographical\\_348.pdf](http://sticonference.org/Proceedings/vol1/Grossetti_Geographical_348.pdf).

HADDAD, Eduardo A.; MENA-CHALCO, Jesús P.; SIDONE, Otávio J. G. Scholarly Collaboration in Regional Science in Developing Countries: The Case of the Brazilian REAL Network. **International Regional Science Review**, [S. l.], v. 40, n. 5, p. 500–529, 2017. DOI: 10.1177/0160017615614898.

HILBE, J. M. **Negative binomial regression**. Cambridge, UK: Cambridge University Press, 2011. DOI: <https://doi.org/10.1017/CBO9780511973420>.

HOEKMAN, Jarno; FRENKEN, Koen; TIJSSSEN, Robert J. W. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. **Research Policy**, [S. l.], v. 39, n. 5, p. 662–673, 2010. DOI: 10.1016/j.respol.2010.01.012. Disponível em: <http://dx.doi.org/10.1016/j.respol.2010.01.012>.

JEONG, Seongkyoon; CHOI, Jae Young; KIM, Jang-Yun. On the drivers of international collaboration: The impact of informal communication, motivation, and research resources. **Science and Public Policy**, [S. l.], 2013. DOI: 10.1093/scipol/sct079.

KATZ, J. Sylvan. Geographical proximity and scientific collaboration. **Scientometrics**, [S. l.], v. 31, n. 1, p. 31–43, 1994. DOI: 10.1007/BF02018100.

KATZ, J. Sylvan; MARTIN, Ben R. What is research collaboration? **Research Policy**, [S. l.], v. 26, p. 1–18, 1997. DOI: 10.1016/S0048-7333(96)00917-1.

LANGRISH, John. The changing relationship between science and technology. **Nature**, [S. l.], v. 250, p. 614–616, 1974.

LETA, Jacqueline; CHAIMOVICH, Hernan. Recognition and international collaboration: The Brazilian case. **Scientometrics**, [S. l.], v. 53, n. 3, p. 325–335, 2002. DOI: 10.1023/A:1014868928349.

LEYDESDORFF, Loet; BORNEMANN, Lutz; WAGNER, Caroline S. The Relative Influences of Government Funding and International Collaboration on Citation Impact. **Journal of the Association for Information Science and Technology**, [S. l.], v. 70, n. 2, p. 198–201, 2019. DOI: 10.1002/asi.24109.

LONG, J. Scott; FREESE, Jeremy. **Regression models for categorical dependent variables using Stata**. [s.l.: s.n.].

MENA-CHALCO, Jesús Pascual; CESAR JUNIOR, Roberto Marcondes. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, [S. l.], v. 15, n. 4, p. 31–39, 2009. DOI: 10.1590/s0104-65002009000400004.

MENA-CHALCO, Jesús Pascual; DIGIAMPIETRI, Luciano Antonio; LOPES, Fabrício Martins; JUNIOR, Roberto Marcondes Cesar. Brazilian Bibliometric Coauthorship Networks. **Journal of the American Society for Information Science and Technology**, [S. l.], 2014. DOI: 10.1002/asi. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/asi.22883/abstract>.

NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, [S. l.], v. 98, n. 2, p. 404–409, 2001. DOI: 10.1073/pnas.98.2.404.



PAN, Raj Kumar; KASKI, Kimmo; FORTUNATO, Santo. World citation and collaboration networks: uncovering the role of geography in science. **Scientific Reports**, [S. l.], v. 2, p. 1–7, 2012. DOI: 10.1038/srep00902.

PONDS, Roderik; VAN OORT, Frank; FRENKEN, Koen. The geographical and institutional proximity of research collaboration. **Papers in Regional Science: The Journal of the Regional Science Association International**, [S. l.], v. 86, n. 3, p. 423–443, 2007. DOI: 10.1111/j.1435-5957.2007.00126.x.

ROCHA, Elisa Maria Pinto; FERREIRA, Marta Araújo Tavares. **Indicadores de ciência, tecnologia e inovação: mensuração dos sistemas de CTel nos estados brasileiros** *Ciência da Informação*, 2004. DOI: 10.1590/s0100-19652004000300008.

ROYAL SOCIETY. **Knowledge , networks and nations: Global scientific collaboration in the 21st century**. London. Disponível em: [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/Influencing\\_Policy/Reports/2011-03-28-Knowledge-networks-nations.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/Influencing_Policy/Reports/2011-03-28-Knowledge-networks-nations.pdf).

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**, [S. l.], n. 46, p. 247–266, 2011. DOI: 10.1007/s00168-009-0334-3.

SCHERNGELL, Thomas; HU, Yuanjia. Collaborative Knowledge Production in China: Regional Evidence from a Gravity Model Approach. **Regional Studies**, [S. l.], v. 45, n. 6, p. 755–772, 2011. DOI: 10.1080/00343401003713373.

SEN, Ashish; SMITH, Tony E. **Gravity models of spatial interaction behavior**. New York.

SIDONE, Otávio José Guerci. **Cientometria Espacial: a geografia do conhecimento no Brasil**. Rio de Janeiro: e-papers, 2018.

SIDONE, Otávio José Guerci; HADDAD, Eduardo Amaral; MENA-CHALCO, Jesús Pascual. Scholarly Publication and Collaboration in Brazil: The Role of Geography. **Journal of the American Society for Information Science and Technology**, [S. l.], 2016. DOI: 10.1002/asi. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/asi.22883/abstract>.

UBER. **H3—A Hexagonal Hierarchical Geospatial Indexing System**. 2020. Disponível em: <https://h3geo.org/>. Acesso em: 11 jun. 2021.

WAGNER, Caroline; JONKERS, Koen. Open countries have strong science. **Nature**, [S. l.], v. 550, n. 7674, p. 33, 2017. DOI: 10.1038/550032a.

WANG, Yan; WU, Yishan; PAN, Yuntao; MA, Zheng; ROUSSEAU, Ronald. Scientific collaboration in China as reflected in co-authorship. **Scientometrics**, [S. l.], v. 62, n. 2, p. 183–198, 2005. DOI: 10.1007/s11192-005-0013-9.

WINKELMANN, R. **Econometric Analysis of Count Data**. Berlin: Springer, 2008.

YE, Qi; WU, Bin; WANG, Bai. Visual Analysis of a Co-authorship Network and its Underlying Structure. **Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008**, [S. l.], v. 4, p. 689–693, 2008. DOI: 10.1109/FSKD.2008.436.

## APÊNDICE

Figura 2.

Top 105 municípios em colaboração	Grau	Modularidade de classe	Grau ponderado
São Paulo/SP	106	1	551423
Rio de Janeiro/RJ	106	1	494832
Belo Horizonte/MG	106	2	201863
Porto Alegre/RS	106	3	198113
Recife/PE	106	4	115046
Ribeirão Preto/SP	106	0	124454
Campinas/SP	106	1	165056
Curitiba/PR	106	0	110036
Fortaleza/CE	106	4	101831
Brasília/DF	106	0	93072
Florianópolis/SC	106	3	85730
Santa Maria/RS	105	3	74089
São Carlos/SP	106	0	77472
Belém/PA	106	0	62950
Botucatu/SP	106	0	58675
Viçosa/MG	106	2	60106
Salvador/BA	106	4	61061
Niterói/RJ	105	1	48547
Maringá/PR	105	0	49290
Pelotas/RS	106	3	55166
Goiânia/GO	106	0	52950
Londrina/PR	106	0	48909
Natal/RN	106	4	49833
Santo André/SP	99	1	35875
Lavras/MG	104	2	40646
Piracicaba/SP	105	0	46720
Uberlândia/MG	106	0	40588
João Pessoa/PB	106	4	41593
Araraquara/SP	101	0	35631
São José dos Campos/SP	106	1	33947
Manaus/AM	106	0	39337
Juiz de Fora/MG	106	1	35150
Bauru/SP	105	0	30131
Campo Grande/MS	106	0	33867
Vitória/ES	105	1	27947
São Luís/MA	105	0	26092
Araçatuba/SP	96	0	18798
Teresina/PI	104	4	24287
Campina Grande/PB	104	4	22069
Rio Grande/RS	100	3	21506

Jaboticabal/SP	105	0	24895
Campos dos Goytacazes/RJ	104	1	19116
Cuiabá/MT	106	0	22087
Ponta Grossa/PR	102	0	19263
Santos/SP	93	1	14238
Seropédica/RJ	105	1	21043
Capão do Leão/RS	91	3	9515
Maceió/AL	105	4	19848
Ouro Preto/MG	105	2	16904
São José do Rio Preto/SP	106	0	18050
Duque de Caxias/RJ	92	1	7958
Uberaba/MG	103	0	18061
São Cristóvão/SE	105	4	21536
Montes Claros/MG	94	2	14105
Aracaju/SE	105	4	16855
Dourados/MS	103	0	14707
Pirassununga/SP	96	0	13201
Presidente Prudente/SP	100	0	13965
Diadema/SP	103	1	9611
Mossoró/RN	103	4	13429
Sobral/CE	94	4	8452
Cascavel/PR	99	0	10490
São Bernardo do Campo/SP	98	1	6876
Alfenas/MG	103	0	13730
Passo Fundo/RS	96	3	10194
Sorocaba/SP	102	1	12356
Ilhéus/BA	106	4	11854
Diamantina/MG	100	2	11951
Canoas/RS	91	3	5907
Itajubá/MG	93	1	7106
Macaé/RJ	89	1	6365
Criciúma/SC	93	3	9070
Cruz das Almas/BA	95	4	9866
Chapecó/SC	89	3	8927
Caxias do Sul/RS	91	3	7502
Ananindeua/PA	66	0	6990
Crato/CE	87	4	7559
Feira de Santana/BA	102	4	13845
Rio Claro/SP	104	0	14040
Lages/SC	92	3	7402
Petrolina/PE	98	4	11647
Divinópolis/MG	90	2	6770
São Leopoldo/RS	91	3	7193
Ilha Solteira/SP	91	0	7480
Porto Velho/RO	95	0	8109

São João del Rei/MG	97	1	8225
Marília/SP	78	0	7311
Limeira/SP	85	1	6830
Franca/SP	82	0	6809
Alegre/ES	100	2	7910
Uruguaiana/RS	88	3	5673
Araras/SP	98	1	6663
Guarapuava/PR	94	0	7236
Macapá/AP	95	0	6661
Toledo/PR	87	0	5606
Patos/PB	81	4	5650
Rio Verde/GO	93	0	5561
Lorena/SP	83	1	5701
Areia/PB	89	4	5621
Sete Lagoas/MG	92	2	5579
Blumenau/SC	100	3	5517
Parnaíba/PI	95	4	5930
Boa Vista/RR	101	0	5486
Sinop/MT	101	0	6522
Vitória da Conquista/BA	99	4	5561