

# CONDICIONANTES DA COLABORAÇÃO ACADÊMICA NO SISTEMA CIENTÍFICO BRASILEIRO <sup>1</sup>

Daniel de Oliveira Costa <sup>2</sup>

Eduardo Gonçalves <sup>3</sup>

Renato de Castro Garcia <sup>4</sup>

**Resumo:** O objetivo deste trabalho é compreender a influência da proximidade social, institucional, organizacional, cognitiva e geográfica no contexto colaborativo entre pesquisadores brasileiros no período entre 2009 a 2017. Trabalhos empíricos discutem o papel de proximidades em países desenvolvidos, mas faltam trabalhos que abordem o tema em países em desenvolvimento. Além disso, embora os trabalhos sobre proximidades sejam numerosos, a maioria desses estudos não considera todas essas dimensões simultaneamente. Para a construção da rede de pesquisadores e a construção dos indicadores de proximidades, foram utilizados dados obtidos da plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Os resultados mostram que todas as dimensões de produtividade são relevantes para explicar as publicações de pares de pesquisadores no Brasil. Especificamente, o fato de estar no mesmo estado, pertencer à mesma instituição e à mesma organização possuem associação positiva com o número de publicações. Além disso, interações prévias e contínuas entre os pesquisadores também são condicionantes relevantes.

**Palavras-chave:** Conhecimento Científico; Análise de Redes Sociais; Proximidades; Plataforma Lattes

**Código JEL:** D85; D83; O30; O34

**Abstract:** This paper aims to understand the influence of social, institutional, organizational, cognitive and geographic proximities in the Brazilian collaborative academic production in the period between 2009 and 2017. Empirical works discuss these types of proximities in developed countries in which scientific and technological systems are mature. However, evidence for developing countries is sparse. Furthermore, although work on proximities is numerous, most of these studies do not consider all these dimensions simultaneously. For the construction of the network of researchers and the construction of proximity indicators, we use data obtained from the Lattes platform of the National Council for Scientific and Technological Development (CNPq). The results show that all productivity dimensions are relevant to explain the publications of peer researchers in Brazil. Specifically, the fact of being in the same state, belonging to the same institution and the same organization has a positive association with the number of publications. Besides, previous and continuous interactions between researchers are also relevant drivers.

**KEYWORDS:** Scientific knowledge; Social network analysis; Proximity; Lattes platform

**JEL CODES:** D85; D83; O30; O34

**Área 9 - Economia Industrial e da Tecnologia**

---

<sup>1</sup> Os autores agradecem o apoio da FAPEMIG, CNPq e CAPES.

<sup>2</sup> Mestre em Economia pelo PPGE/FE-UFJF. E-mail: daniellmg71@gmail.com

<sup>3</sup> Professor da Faculdade de Economia da Universidade Federal de Juiz de Fora (UFJF) e do PPGE, pesquisador do CNPq. E-mail: eduardo.goncalves@ufjf.edu.br

<sup>4</sup> Professor da Faculdade de Economia da Universidade de Campinas e do PPGE, pesquisador do CNPq. E-mail: rcgarcia@unicamp.br

## 1 Introdução

Diante de um Sistema Nacional de Inovação (SNI) imaturo as universidades assumem um papel de polos de interação especializados, que mantem a fronteira do conhecimento e o desenvolvimento tecnológico dos setores da indústria alinhados aos de países desenvolvidos e, desse modo, reduzem os riscos de maiores atrasos no SNI (Bernardes & Albuquerque, 2003; Rapini *et al*, 2006). Diante disso, a cooperação entre cientistas pesquisadores impulsiona a absorção de conhecimento e eleva o grau de importância das universidades como uma “antena”, com ligações internacionais capazes de absorver novos conhecimentos (Cohen & Levinthal, 1990).

Por sua vez, o mundo acadêmico tem apresentado um dinamismo cada vez mais acentuado, no qual os novos achados permeiam a comunidade científica por interação entre pesquisadores, divulgação em periódicos, congressos e coautorias de maneira a consolidar a difusão do conhecimento. A produção científica tem intensificado a realização de parcerias, o que tem levado ao aumento da produtividade (Sonnenwald, 2007). Há uma relação positiva entre artigos em coautorias e número de citações, sendo que este índice muitas vezes é associado ao impacto e qualidade do artigo (Ductor, 2015). No entanto, à medida que avança a fronteira do conhecimento, novos trabalhos tidos como inovadores e de qualidade acabam se tornando mais difíceis de serem atingidos. É exigido cada vez mais domínio de diferentes áreas do saber para estabelecer, analisar, aceitar ou rejeitar hipóteses, tornando a colaboração imprescindível ao avanço produtivo (Porac *et al*, 2004).

A literatura nacional traz extensa descrição do impacto da proximidade geográfica na ciência (Vanz & Stumpf, 2012; Sidone *et al*, 2016). Esses trabalhos incluem, muitas vezes, controles importantes a fim de identificar esse efeito espacial nas regiões brasileira como recursos e financiamento às pesquisas, parcerias internacionais, corpo acadêmico e proximidade tecnológica (McManus, 2020). Apesar de retratarem bem a realidade das colaborações brasileiras, a proximidade geográfica deve ser complementada com outras dimensões com o intuito de uma análise mais completa e não incorrer em possível endogeneidade, já que essas variáveis podem ser correlacionadas.

Logo, características individuais e institucionais podem ser uma fonte de identificação de cooperações (Boschma & Frenken, 2009). Boschma (2005) traduz essas dimensões em 5 proximidades, a saber, cognitiva, geográfica, organizacional, institucional e social, que serão descritas na revisão da literatura. A literatura ainda é escassa nos países emergentes com SNI imaturos, limitando-se à interação Universidade-Indústria e a estudos de Análise de Redes Sociais dessas organizações, não englobando todas as proximidades citadas acima.

Este trabalho busca interpretar o papel das proximidades social, cognitiva, geográfica, organizacional e institucional na decisão dos pesquisadores em colaborar, com base no modelo Binomial Negativo Inflado em Zeros (BNIZ). A variável dependente capta a contagem de artigos publicados em comum por pares de autores durante os triênios 2009 a 2011, 2012 a 2014 e 2015 a 2017. A priori, cabe ressaltar que a construção em nível nacional da proximidade social captou o trabalho em conjunto de pesquisadores no passado a fim de explicar as coautorias no período de análise corrente. As variáveis de proximidades aplicadas empiricamente à cooperação na área científica, as quais são recentes na literatura, sendo aplicadas somente a países com SNI desenvolvidos (Molina-Morales *et al*, 2015; Crescenzi *et al*, 2017; Crescenzi *et al*, 2016).

Este trabalho é composto por mais 6 seções, além desta introdução. A segunda seção revisa a literatura que abrange a aplicação das proximidades em rede em estudos de difusão de conhecimento. A terceira aborda os estudos empíricos acerca das proximidades citadas na seção anterior. Já a quarta seção aborda a estratégia empírica e apresenta estatísticas descritivas da base de dados. Os resultados são apresentados na quinta seção. Por fim, são apresentadas as conclusões e as limitações do trabalho.

## **2 Proximidade de rede no processo de colaboração científica**

Vários fatores influenciam o pesquisador na tomada de decisão com quem interagir a fim de maximizar a produção científica. O trabalho acadêmico em conjunto pode levar a estudos de maior qualidade, eficiência, compartilhamento da *expertise* e diminuição de custos. Mesmo os cientistas renomados estão à procura de coautoria que podem complementar seu domínio acadêmico a fim de fomentar novas ideias, aproveitar habilidades de ambos e compartilhar acesso a diferentes fontes de conhecimento (Smith & Sotala, 2011). Essa colaboração proporciona melhor entendimento de problemas complexos e que requerem união de várias áreas e pesquisadores.

As colaborações ligam diferentes áreas da academia incorporando nos estudos novas áreas férteis. O preenchimento das lacunas e avanço na fronteira do conhecimento trazido por essa colaboração resulta em artigos científicos inovadores. Trabalhos acadêmicos que possuem coautoria tendem a ser mais citados, especialmente se essas parcerias forem de diferentes países (Smith & Sotala, 2011). No entanto, mesmo com pares de pesquisadores que se complementam, as parcerias são restritas a diversos fatores sociais, físicos e institucionais. Estes são introduzidos em conjunto em cinco dimensões por Boschma (2005). A literatura teórica de proximidade em redes começa a ganhar força na década de 90 (Bellet *et al.*, 1993 *apud* Balland *et al.*, 2015), com uma variedade de proximidades, como a social, a cognitiva, a institucional, a organizacional e a geográfica. Esta última mais consolidada na literatura (Marshall, 1920; Audretsch & Feldman, 1996; Anselin *et al.*, 2000; Boschma, 2005; Belussi & Caldari 2009; Gonçalves *et al.*, 2019).

### **2.1 Proximidade Geográfica**

Devido a maior facilidade de transferência de conhecimento, principalmente tácito, pesquisadores geograficamente próximos têm maior propensão a colaborar. A tentativa de identificar o papel da proximidade geográfica vem evoluindo na literatura com diversas metodologias e abordagens. No entanto, deve-se atentar que esse efeito pode ser provocado pelo fato de que relações em rede tendem a ser mediadas geograficamente, o que causaria uma concentração de ligações em rede localizadas que explicaria esse fenômeno espacial. Uma parte do efeito locacional em citação de patentes se deve à autocitação (Criscuolo & Verspagen, 2008). Essa seria a indicação de difusão de conhecimento na mesma organização ou do mesmo indivíduo.

### **2.2 Proximidade Cognitiva**

A difusão do conhecimento muitas vezes está sujeita a conexões com instituições e agentes que têm domínio desse conhecimento. O bom entendimento entre as partes está sujeito à capacidade de absorção, a qual depende da área do conhecimento que esses indivíduos estão inseridos. Se ambos se situam em um mesmo campo tecnológico, esse processo é facilitado. Dado isso, a proximidade cognitiva mensura a distância que os agentes estão nas diversas áreas do conhecimento (Boschma, 2005).

### **2.3 Proximidade Social**

Experiências sociais informais nas quais o indivíduo tem ideias são consideradas interações sociais. Conferências, reuniões e encontros são meios de difusão de conhecimento social que, muitas vezes, são difíceis de serem captados em modelos empíricos. A proximidade social é moldada a partir de interações anteriores ou já existentes na rede. São baseadas em confiança e experiência. Logo, os processos tácitos de aprendizagem são facilitados. Relações de amizade, parentesco e interações passadas podem se basear em características em comum dos pesquisadores na rede, sejam elas já predefinidas ou que ocorreram após a interação (Boschma, 2005).

## 2.4 Proximidade Institucional

Instituições são normas, leis, hábitos em comum e incentivos que rege um grupo de indivíduos. Já a proximidade institucional é o inverso da distância que dois indivíduos têm em relação ao contexto institucional. É uma proximidade importante devido à diminuição da incerteza no contexto institucional com normas, leis e cultura similares a nível macro (Boschma, 2005). Apesar de fazer parte de uma rede social, não se deve confundi-la com a distância social, que ocorre ao nível micro (Edquist, 1997). Esta proximidade também leva em conta características locais e regionais. Indivíduos dentro de uma mesma região com características próprias são impactados por esta, além de haver outras estruturas institucionais que operam em conjunto (Balland *et al.*, 2014).

## 2.5 Proximidade Organizacional

A proximidade organizacional refere-se a agentes que se encontram em mesma situação organizacional ou que dividem um mesmo nível hierárquico, seja na mesma ou em diferentes organizações. Esta medida pode ser interpretada como uma variável contínua, onde seus extremos são indivíduos completamente autônomos e que possuem ligações fracas, proximidade baixa e indivíduos subordinados a um alto nível hierárquico com ligações fortes e proximidade alta (Boschma, 2005).

Agentes com proximidades organizacional alta exercem um papel relevante na difusão do conhecimento nas periferias da rede (Capaldo *et al.*, 2014), encurtando a distância de atores por terem menor custo de aprendizagem, tendo em vista que tanto a incerteza quanto a possibilidade de oportunismo são diluídos (Boschma, 2005). Esse mecanismo se torna importante na rede de colaboração, contribuindo na formação de *gatekeepers*<sup>5</sup> a nível regional e nacional (Legendijk & Lorentzen, 2006; Balland *et al.*, 2015).

## 3 Estudos empíricos sobre os tipos de proximidades

Trabalhos utilizam as interações prévias a fim de captar a proximidade social, mostrando maior probabilidade de colaborações futuras entre indivíduos que já interagiram (Molina-Morales *et al.*, 2015; Crescenzi *et al.*, 2016; Crescenzi *et al.*, 2017). O estudo de Hong & Su (2013) também corrobora este resultado ao utilizar as proximidades não-geográficas para inferir a probabilidade de interação universidade-indústria. Os autores concluem que outras proximidades tendem a diminuir o efeito da distância geográfica cujo efeito continua significativo mesmo após este controle.

As proximidades organizacional e institucional podem interferir na transferência de conhecimento e inovação. Nos Estados Unidos da América (EUA), há evidências de que parcerias formadas entre firmas na formação de estratégias de mercado facilitam a transferência de tecnologia (Mowery *et al.*, 1996). Kaygalak & Reid (2016) analisam a interação universidade-empresa dentro do contexto turco, fazendo-se a diferenciação de tipos organizacionais entre universidade técnica, universidade não-técnica e indústrias. Segundo o teste de independência entre os grupos em relação à medida de proximidade, existe diferença da média entre esses tipos organizacionais. Já na China, a estruturação das universidades que são vinculadas a ministérios específicos e províncias com tipos organizacionais diferentes apontam para uma menor importância da proximidade geográfica quando considerados dois agentes pertencentes ao mesmo ministério (Hong & Su, 2013). Embora estes trabalhos concordem com expressividade da dimensão organizacional, há indícios de uma menor contribuição desta dimensão, mas ainda significativo, e uma relação de substituição com a dimensão geográfica (Paci *et al.* 2014; Fernández *et al.*, 2016; Heringa *et al.*, 2016).

---

<sup>5</sup> *Gatekeepers* são pontes que ligam dois componentes importantes na rede (Borgatti *et al.*, 2013).

A colaboração entre pesquisadores muito próximos cognitivamente se torna prejudicial. É esperado que essa dimensão assuma um formato de U invertido na produtividade inventiva (Cohendet & Llerena, 1997). Devido a isso os acadêmicos enfrentam um *tradeoff* entre maior absorção de conhecimento e um grau maior de inovação no ato da colaboração, portanto, uma proximidade muito alta nem sempre reflete maior grau inventivo (Nooteboom, 2000; Nooteboom *et al.*, 2007; Broekel & Boschma, 2012).

#### 4 Estratégia Empírica

O objetivo desse trabalho é verificar o papel dos tipos de proximidades sobre as colaborações científicas. Como base nisso, será utilizado um modelo Binomial Negativo Inflado em Zeros a fim de inferir sobre o número de artigos publicados em colaboração por par de autor.

A equação 1 descreve o modelo e a estrutura a ser considerado nas regressões. A contagem do número de publicações em colaboração do par de pesquisadores é a variável dependente. As variáveis explicativas são as proximidades, a centralidade de entrelaçamento e o tempo de docência. Enquanto a contagem do número de publicações em colaboração do par de pesquisadores no triênio é a variável dependente. Os dois lados da equação ainda serão descritos na seção de metodologia. A contagem do número de publicações em colaboração do par de pesquisadores é a variável dependente. A descrição desta variável por triênio está na Tabela 1. Como é visto na tabela, ela possui uma dispersão maior que a média, o torna a distribuição de Poisson para dados de contagem não adequada, pois não suporta a heterogeneidade não observada dos valores preditos em relação à sobredispersão (Hilbe, 2011). Portanto, um modelo Binomial Negativo de contagem deve ser utilizado para explicar essas variáveis, já que ele acomoda a heterogeneidade da dispersão.

Há um número grande de pesquisadores na rede e sua maior parte não coopera entre si, caracterizando uma rede esparsa. A fim de inferir sobre a influência da proximidade geográfica, cognitiva, social, organizacional e institucional na rede, será preciso criar observações de ligações de pares fictícias para um contrafactual, as quais se enquadram no grupo de pares de pesquisadores que sempre assumem zero no valor observado. A definição do modelo contrafactual será através da seleção de uma rede aleatória com probabilidade de ligação igualmente distribuída, tal procedimento será descrito posteriormente.

$$\begin{aligned}
 & \text{Publicações do par}_{(ij)t} \\
 = & \beta_0 + \text{Proximidades}_{(ij)t}\beta_1 + \text{Formação}_{(ij)t}\beta_2 + \text{Tempo docência}_{it}\beta_3 \\
 & + \text{Tempo docência}_{jt}\beta_4 + \text{Betweenness}_{it}\beta_5 + \text{Betweenness}_{jt}\beta_6 + z_f + k_t + \varepsilon_{(ij)t}
 \end{aligned}$$

A variável dependente se trata de uma variável de contagem de ligações do par  $(i,j)$ , sendo, respectivamente, o autor e o coautor. As ligações foram agrupadas nos triênios de 2009 a 2011, 2012 a 2014 e 2015 a 2017. Cada colaboração do par  $(i,j)$  no triênio no mesmo artigo indica uma ligação. *Publicações do par* $_{(ij)t}$ , sendo  $i$  o autor,  $j$  o coautor,  $t$  é a referência do triênio e  $z_f$  é referente à grande área do conhecimento referente a ambos.

Tabela 1 – Estatísticas de dispersão em torno da média da variável dependente 2009-2017

Triênio	Média	Desvio-Padrão
2009-2011	1,07461	1,4392
2012-2014	1,27619	2,7954
2015-2017	1,29997	2,7241

Fonte: Elaboração própria a partir de dados da Plataforma Lattes.

#### 4.1 Descrição da base de dados

A base de dados utilizada é a da Plataforma Lattes, com o período de análise de 2000 a 2017. As estatísticas descritivas se estendem a todo o período, no entanto, por limitação computacional, as regressões incluem somente os 3 últimos triênios, de 2009 a 2017. Essa base se divide nas seguintes partes: publicações de artigos, depósito de patentes, endereço profissional e vínculos de atividades ou trabalho. O número de registros de artigos no período é de 5.822.614 atribuídos a 753.677 autores. Este primeiro valor pode ou não incluir mais de um registro de um mesmo artigo de indivíduos diferentes, ou seja, existem observações de autoria única, autoria solitária, sendo que neste caso só há um registro de um indivíduo, e também de coautoria com mais de um registro de um único título. O número de coautorias é determinado pela quantidade de registros de um mesmo artigo. A mesma lógica é atribuída às patentes. O número de depósito de patente repetido de diferentes indivíduos indica a cooperação no processo de depósito. O total de registro de depósitos é de 56.924 declarados por 24.114 indivíduos declarados na plataforma

Para construção do filtro que foi aplicado à base descrita foram considerados somente pesquisadores, professores universitários, revisores de periódicos e doutores que publicaram no período de análise. Com isso, a amostra retirada a partir da plataforma lattes conta com 445.390 pesquisadores. Metade destes tem até 5 anos de docência e a média é de 7 anos. A grande maioria, 86%, já teve experiência como docente e 45% tem doutorado. A região Sudeste é a que concentra maior número de pesquisadores, contando com 35% da amostra e a região Norte é a menor, com 4% da amostra. Esse cenário se repete em relação à centralidade *betweenness* que, em média, no Sudeste é maior que nas outras regiões, seguida por Sul, Nordeste e Centro-Oeste, respectivamente. A Tabela 2 descreve a representatividade de cada região na base, a média da centralidade *betweenness* e a centralidade de grau ponderada média dos pesquisadores dessa região.

Tabela 2 – Descrição das medidas de rede das regiões brasileiras

Região	Participação da base	Média centralidade ( <i>Betweenness</i> )	Grau médio ponderado
Norte	5,41%	14.412,79	13,04
Nordeste	20,04%	19.749,57	15,78
Sudeste	45,13%	28.317,34	24,05
Sul	21,32%	24.225,10	20,06
Centro-Oeste	8,11%	17.245,38	14,49

Nota: A participação da região na base indica a proporção de pesquisadores da região em relação ao total de pesquisadores, a soma das participações não equivale a totalidade já que há *missings*. As medidas de rede de centralidade *betweenness* e grau médio ponderado foram calculados a partir da média dos indivíduos de cada região brasileira que produziram artigos em colaboração.

Fonte: Elaboração própria a partir dos dados da Plataforma Lattes.

A Tabela 3 detalha as estatísticas de rede no período 2000 a 2017 das grandes áreas do conhecimento de pesquisas do indivíduo, sendo que o indivíduo pode ser multidisciplinar, ou seja, ter mais de uma grande área. As áreas são classificadas de acordo com a tabela do Conselho Nacional de Desenvolvimento Científico (CNPq). Somente as áreas da saúde, humanas e sociais aplicadas representam mais da metade da base. As áreas com maior colaboração são: ciências biológicas, ciências agrárias e ciências da saúde. A tendência de agrupamento medida pela média do coeficiente de *clustering* aponta para maior prevalência de grupos na biologia, seguido por ciências agrárias e engenharias.

O trabalho de Mena-Chalco *et al.* (2014) utiliza a plataforma Lattes para descrever as medidas de rede de cada área do conhecimento dos pesquisadores. Os autores constatam que as áreas com mais pesquisadores são Ciências da Saúde, Ciências humanas e Ciências Sociais, sendo que as áreas de Ciências Biológicas e Agrárias são as mais colaborativas.

Tabela 3 – Características das Grandes Áreas do Conhecimento na rede de colaboração

Grande área do conhecimento	% de indivíduos	Média do coeficiente de <i>cluster</i>	Média centralidade de grau
Ciências Exatas e da Terra	12,08%	0,35	11,50
Ciências Biológicas	12,66%	0,41	20,24
Engenharias	7,53%	0,35	8,06
Ciências da Saúde	21,41%	0,35	13,36
Ciências Agrárias	9,12%	0,38	16,19
Ciências Sociais Aplicadas	14,54%	0,14	2,18
Ciências Humanas	16,52%	0,14	2,15
Linguísticas, letras e artes	5,40%	0,70	1,01
Outros	0,74%	0,33	5,15

Nota: As medidas de rede de coeficiente *cluster* e centralidade de grau foram calculados a partir da média dos indivíduos de cada grande área do conhecimento brasileira que produziram artigos em colaboração.

Fonte: Elaboração própria a partir dos dados da Plataforma Lattes e do software *Gephi*.

A Tabela 4 descreve a evolução das métricas de rede no tempo. Há um aumento no coeficiente de *Cluster* que indica maior agrupamento. Isso pode ser consequência de uma rede mais densa e conectada e também da presença de ligações mais fortes indicado pela centralidade de grau.

Os dados mostram que a razão de autores solitários diminui em todos os anos, enquanto que a participação dos autores com parcerias aumenta no tempo. No ano de 2008, a participação dos autores únicos representava 24,32%. Este número, 10 anos depois, cai para 18,96%. No mesmo período, as participações em coautorias de mais de 3 membros aumentaram em mais de 10%.

Tabela 4 – Estatísticas descritivas da rede de colaboração dos triênios, Brasil, 2000-2017

Triênios	2000-2002	2003-2005	2006-2008	2009-2011	2012-2014	2015-2017
Número de autores	101.537	136.688	169.930	213.932	254.193	272.264
Grau médio <sup>1</sup>	3,12	3,86	4,77	6,08	5,18	7,77
Grau médio ponderado <sup>1</sup>	4,44	5,49	6,78	8,93	8,16	12,07
Número médio de artigos sem colaborações	3,03	2,86	2,36	2,07	1,77	1,65
% de nós isolados <sup>2</sup>	70,22	63,35	54,06	46,07	36,15	29,26
Coeficiente de cluster médio	0,60	0,62	0,65	0,66	0,63	0,66

Elaboração própria a partir dados da Plataforma Lattes. <sup>1</sup> Medidas de rede tomam como base somente os indivíduos que produziram artigos em cooperação. <sup>2</sup> Esse valor reflete os nós isolados identificados após a construção da rede, o que não retrata a verdadeira participação de pesquisadores solo.

Fonte: elaboração dos autores.

## 4.2 Variáveis explicativas

### *Proximidade social*

Esta variável será construída a partir de interações que o par de indivíduos teve no passado, como utilizado por Crescenzi *et al.* (2017) e Crescenzi *et al.* (2016). Este trabalho difere desta literatura, já que considera defasagens temporais de até terceira ordem. Tais interações podem ser em colaboração em artigos ou produção de patente. A construção levou em conta um período de defasagem de interações prévias de até 3 triênios anteriores.

O período de 2000 a 2008 foram 1.440.760 publicações registradas de 226.236 autores, sendo que 18% destes foram publicações solo. Após transformar essas observações em pares de autores e agregando todas as publicações é obtido 381.102 vértices de coautoria e 180.209 indivíduos que publicam sozinhos. Os Gráficos 1 a 6 mostram, em porcentagem, o número de cooperações por de 2000 a 2017. Há uma diferença absoluta entre os três períodos, sendo que o segundo e terceiro períodos tiveram um aumento na publicação de artigos em colaboração. O aumento de mais de 25%<sup>6</sup> dos trabalhos são de autoria única e a maior parte das interações ocorreu somente uma vez (67%). Interações repetidas representam 27% da amostra, estes são os potenciais laços mais fortes na rede. A Tabela 5 descreve a variável explicativa de proximidade social de coautorias segregada por triênios e por defasagem temporal. A construção desta foi por três variáveis binárias cujo valor assume 1 quando o par de autores já teve alguma interação prévia na defasagem de referência. Há um crescimento no tempo de interações prévias e isso condiz com a literatura, já que, com o tempo, a fronteira do conhecimento se torna mais complexa e demanda mais cooperações e interações repetidas são mais frequentes visto que há menos incerteza nessa parceria (Danell, 2014).

Tabela 5 – Descrição e distribuição da variável social de ligações prévias no tempo 2009-2017

Triênio	t-1	t-2	t-3
2015-2017	183.645	78.502	10.912
2012-2014	123.908	50.342	20.869
2009-2011	66.348	25.775	33.763
Média (Desvio-Padrão)	0,017 (0,130)	0,007 (0,084)	0,003 (0,055)

Fonte: Elaboração própria a partir de dados da Plataforma Lattes

É possível observar que as defasagens mais recentes são mais representativas. A frequência de presença de interação na primeira defasagem é mais que o dobro na segunda defasagem em todos os triênios. Essa tendência é lógica, visto que projetos desenvolvidos por pesquisadores podem produzir vários artigos os quais são aceitos e publicados subsequentemente. As interações prévias de primeira ordem estão correlacionadas com a de segunda ordem. Das 373.901 interações em *t-1*, 107.481 dessas se repetem em *t-2*, e 34.752 pares de indivíduos mantiveram relação nos 3 períodos de defasagem. Devido a isso, é preciso atentar para esta variável ao incluí-la no modelo, pois há risco de colinearidade. Esse raciocínio não é empregado no copatenteamento<sup>7</sup>, já que são poucas observações. Para esse caso, foi criado somente uma variável binária que acusa se o par de autores já patenteou em conjunto anteriormente.

<sup>6</sup> Este é o valor observado, criado a partir de das referências dos artigos publicados, ou seja, o autor acusou que este artigo é de autoria única. Este valor diverge do valor da rede criada, já que para montagem da rede foi necessário a correspondência das coautorias pelo título da publicação, o que nem sempre é exatamente igual em todos, mesmo depois de todas as correções manuais. A Tabela 4 mostra tal divergência, que nos triênios finais, aqueles que serão utilizados na regressão, convergem para o valor observado citado.

<sup>7</sup> Foram observadas 19.656 interações prévias em copatenteamento na rede com média de 0,001 e 0,030 de desvio-padrão.



Gráfico 1 – Distribuição de número de coautorias por artigo 2000-2002

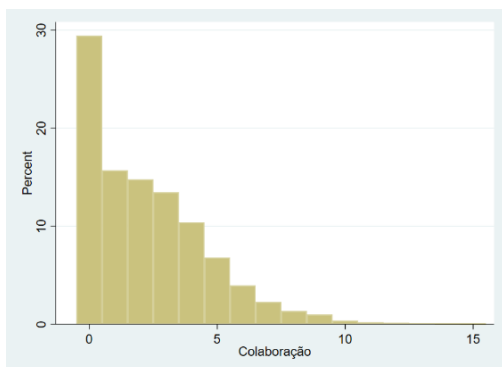


Gráfico 2 – Distribuição de número de coautorias por artigo 2003-2005

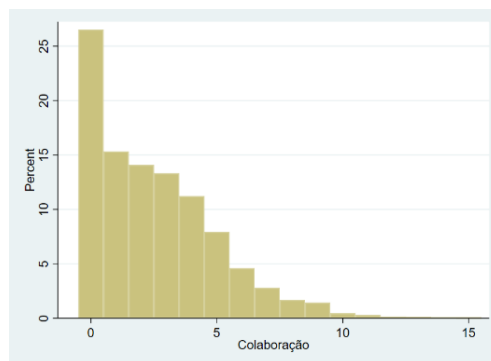


Gráfico 3 – Distribuição de número de coautorias por artigo 2006-2008

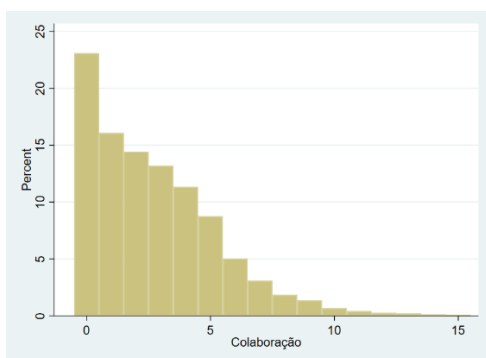


Gráfico 4 – Distribuição de número de coautorias por artigo 2009-2012

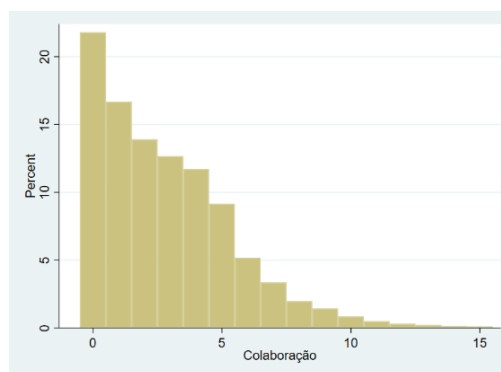


Gráfico 5 – Distribuição de número de coautorias por artigo 2012-2014

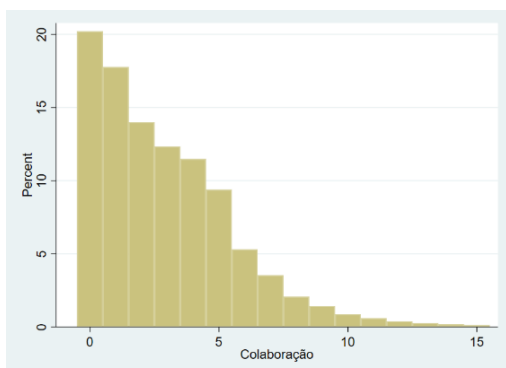
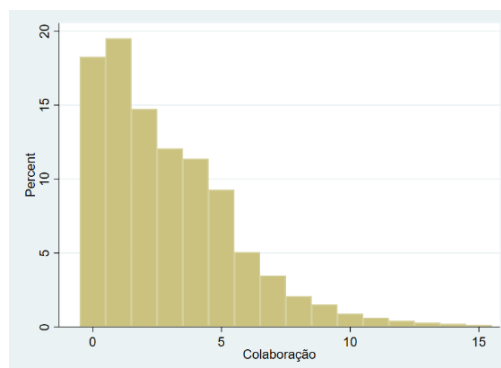


Gráfico 6 – Distribuição de número de coautorias por artigo 2015-2017



### *Proximidade institucional*

A variável de proximidade institucional será construída a partir de dados informados pelos pesquisadores do currículo Lattes. Estes informam à qual instituição estiveram vinculados durante sua vida acadêmica. Para isso, foram criadas *Dummies* que remetem às instituições mais frequentes na amostra, como institutos de pesquisa aplicados, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) e universidades públicas. Outra *Dummy* criada foi a de correspondência exata de instituição, que recebe valor igual a um quando o par já teve vínculo na mesma instituição.

### *Proximidade organizacional*

Indivíduos que possuem passagem por empresas ou universidade vivenciam tipos organizacionais diferentes, como normas, regras e incentivos que divergem. As variáveis criadas permitem entender se a relação intra e inter organizacionais são mais ou menos cooperativas. *Dummies* dos pares foram criadas a fim de discriminar se o par de indivíduos tem experiência universitária ou no setor produtivo, também se ambos tiveram carreira em universidade e em empresa. Essa análise de pertencimento ao setor produtivo ou à universidade e se estes grupos diferem no posicionamento da rede já fora feita por Kaygalak & Reid (2016).

### *Proximidade Cognitiva*

A variável cognitiva indica à qual área do conhecimento ambos os pesquisadores pertencem. Como visto na Tabela 3, há uma diferença grande entre os grupos que cooperam de acordo com sua grande área do conhecimento. A construção dessa variável leva a inclusão de várias *dummies* por grande área que assumem 1 quando ambos os indivíduos do par correspondem àquela grande área, caso contrário 0. No total são 9 grandes áreas do conhecimento definidas pela tabela do CNPq: Ciências exatas e da terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas; Linguísticas, letras e artes; e Outros.

### *Proximidade Geográfica*

Como a base de dados se limitava a um número de *missings* de cidades muito grande, optou-se por considerar a Unidade da Federação para construção da variável de proximidade geográfica. Esta é uma variável binária que, caso ambos os autores do par sejam do mesmo estado informado, assume o valor de 1, caso contrário 0.

### *Variáveis de controle*

De acordo com o vínculo institucional que os pesquisadores apresentaram na plataforma, foi possível identificar quais desses se enquadraram em atividades de docência. A partir desse indicador foram somados os meses que algum tipo de docência ocorreu. Essa variável determina o total de meses em logaritmo tanto do pesquisador *i* quanto do *j*. Seguindo o mesmo raciocínio, onde são consideradas duas variáveis para o par de pesquisadores, acrescenta-se as medidas de *betweenness* do par. A terceira variável de controle é binária, a qual indica se ambos do par são doutores.

Apresentadas as variáveis do modelo, a Tabela 8 apresenta as correlações entre estas. As variáveis que são mais correlacionadas com a dependente são as de interações prévias, a *dummy* do mesmo estado e o logaritmo dos *betweenness* de *i* e *j*. As variáveis independentes não apresentam correlação muito alta ao ponto de manifestar multicolinearidade na regressão.

Tabela 8 – Correlação entre as variáveis do modelo

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Contagem (dependente)	1,00																
(2) IntPrev1	0,38	1,00															
(3) IntPrev2	0,27	0,44	1,00														
(4) IntPrev3	0,19	0,29	0,43	1,00													
(5) MsmTipInst	0,03	0,03	0,03	0,02	1,00												
(6) MsmUni	0,02	0,02	0,02	0,01	0,01	1,00											
(7) MsmEmp	0,08	0,08	0,06	0,04	0,15	0,05	1,00										
(8) MsmUF	0,21	0,19	0,13	0,09	0,03	0,03	0,06	1,00									
(9) MsmUniPub	0,07	0,06	0,04	0,02	-0,02	0,22	0,03	0,03	1,00								
(10) MsmMest	-0,04	-0,04	-0,03	-0,02	-0,03	0,01	-0,05	-0,04	-0,04	1,00							
(11) AmbosDout	0,17	0,18	0,13	0,09	0,06	0,09	0,13	0,15	0,16	-0,20	1,00						
(12) LigaPat	0,08	0,07	0,06	0,05	0,01	0,00	0,02	0,04	0,01	-0,01	0,03	1,00					
(13) MsmInst	0,10	0,11	0,07	0,05	0,00	0,01	0,00	0,10	0,04	-0,01	0,03	0,02	1,00				
(14) LnBetween	0,14	0,14	0,11	0,08	0,04	0,02	0,08	0,10	0,05	-0,05	0,14	0,03	0,02	1,00			
(15) LnCobetwee	0,13	0,13	0,10	0,07	0,03	0,02	0,06	0,09	0,04	-0,04	0,12	0,02	0,01	0,07	1,00		
(16) LnTempdoce	0,04	0,05	0,05	0,05	0,02	0,06	0,04	0,05	0,01	-0,08	0,19	0,01	0,00	0,15	0,01	1,00	
(17) LnCotempdo	0,03	0,04	0,04	0,04	0,01	0,05	0,01	0,04	0,00	-0,08	0,19	0,01	0,00	0,01	0,10	0,01	1,00

Fonte: Elaboração própria a partir dos dados da plataforma Lattes

### 4.3 Construção do contrafactual

No período analisado de 2009 a 2017, foram mapeados 445.390 indivíduos que, na rede, são nós que têm potencial de ligação. Caso todos estes nós se ligassem, teríamos uma rede com 99.185.903.355 possíveis laços. Tal situação hipotética é irreal, visto que a coautoria de um único indivíduo com outros 445.389 autores é impossível. Dada tal situação, são criadas redes fictícias nas quais as ligações podem ser aleatórias ou definidas por características similares dos indivíduos (Chandrasekhar, 2015). Para comparação da base de dados onde há ligações entre os indivíduos, é preciso criar um contrafactual onde  $g_{ij} = 0$ . Em parte, tais modelos criam uma rede aleatória onde há uma probabilidade de ligação distribuída entre os indivíduos.

Neste trabalho será adotado o método para criação do contrafactual de Erdős-Renyi (ERM) (Erdős e Renyi, 1959), descrito por Chandrasekhar (2015). O modelo construído de forma aleatória conta com todos os pesquisadores filtrados, considerando os que publicam em solo e em coautoria. A probabilidade de formação de ligação dos nós na rede é de 0,05. A Tabela 6 mostra a divisão das observações entre o contrafactual, onde  $g_{ij} = 0$  e as ligações verdadeiras,  $g_{ij} \geq 1$ . A Tabela 7 traz valores das variáveis da base por completa. Sendo assim, as estatísticas levam em conta o contrafactual, que é a maior parte das observações, e os pares de pesquisadores observados.

Tabela 6 – Divisão do contrafactual nos triênios, 2009-2017

Y	2009-2011	2012-2014	2015-2017
Laço não formado ( $g_{ij} = 0$ )	94,49%	91,19%	89,44%
Laço formado ( $g_{ij} \geq 1$ )	5,51%	8,81%	10,56%
Total de laços	7.070.810	7.326.663	7.469.855

Fonte: Elaboração própria a partir de dados da Plataforma Lattes

Tabela 7 – Estatísticas descritivas das variáveis de interesse do total de observações, 2009-2017

Variável	Média	Desvio-Padrão	Mínimo	Máximo
ContArtigo	0,134	0,880	0	603
IntPrev1	0,017	0,130	0	1
IntPrev2	0,007	0,084	0	1
IntPrev3	0,003	0,055	0	1
IntPat	0,001	0,074	0	1
MesmaUF	0,134	0,341	0	1
MesmaTipoInst	0,306	0,461	0	1
MesmaInst	0,011	0,221	0	1
MesmaUni	0,946	0,226	0	1
MesmaEmp	0,314	0,226	0	1
MesmaUnivPubl	0,449	0,497	0	1
MsmGrandArea	0,052	0,222	0	1
LnBetweenness <sub>i</sub>	4,613	4,766	-6,557	17,303
LnBetweenness <sub>j</sub>	3,707	4,707	-6,497	17,303
LnTempDoceni	3,324	1,887	0	6,825
LnTempDocenj	3,672	1,746	0	6,825
AmbosDoutor	0,237	0,425	0	1

Fonte: Elaborado a partir de dados da plataforma lattes.

## 5 Resultados

Os resultados da regressão BNIZ em painel do período de 2009-2011 são descritos na Tabela 9. Os resultados foram divididos de acordo com o acréscimo de uma defasagem a cada modelo até que foram incluídas todas as variáveis de proximidade social. Devido a dificuldades computacionais, já que se tratam de mais de 20 milhões de observações e 27 parâmetros, foi feita uma amostra de 10% destes, totalizando 2.182.666 observações. Essa amostra deve ser representativa da base, pois já é um grande número de observações e, segundo a Lei dos Grandes Números, isso garante a consistência dos coeficientes da amostra.

Para melhor visualização as variáveis foram agrupadas em partes: proximidade social, proximidade geográfica, proximidade institucional, proximidade organizacional, e variáveis de controle. As variáveis da proximidade cognitiva foram omitidas para melhor visualização.

A primeira variável, referente à proximidade social, é acrescida a cada regressão com defasagem temporal até  $t-3$ . Em todos os modelos as variáveis são significativas. Assim como os outros parâmetros, os coeficientes de defasagem sofrem pouca alteração nos modelos. A regressão indica que as interações prévias até dois triênios anteriores influenciam positivamente a cooperação entre o mesmo par de indivíduos, embora o coeficiente diminua em  $t-2$ . A defasagem de 3 triênios prévios apresenta sinal negativo e significativo, mas um valor absoluto pequeno. Isto é, ligações prévias de triênios imediatos à observação são mais propícias à colaboração em contraste às ligações mais antigas. Vale ressaltar que essa variável é composta por maioria de ligações que se repetem em  $t-2$  e  $t-1$ . Esse resultado condiz com a literatura, que mostra uma relação benéfica entre a proximidade social e a cooperação (Heringa *et al*, 2014; Heringa *et al*, 2016; Molina-Morales, 2015; Crescenzi *et al*, 2016; Fernández *et al*, 2016; Crescenzi *et al*, 2017). Em acréscimo a essa análise, uma sétima coluna foi inserida a fim de distinguir somente interações prévias contínuas, aquelas que se mantiveram durante os 3 triênios. O sinal positivo e significativo sugere um efeito similar às anteriores, no entanto sua magnitude foi menor. Mesmo assim, o resultado atesta que pode haver confiança entre os pares que cooperam continuamente, com efeitos benéficos sobre a produção científica.

O coeficiente da variável de proximidade geográfica foi positivo e significativo. Isso mostra que indivíduos na mesma Unidade da Federação tendem a ser mais colaborativos. Tal resultado é condizente com a literatura (Heringa *et al*, 2014; Molina-Morales *et al*, 2015; Crescenzi *et al*, 2016). Já a terceira parte é referente à proximidade institucional, composta por duas variáveis binárias. A primeira indica quando o par de indivíduo está em instituições de pesquisa públicas e o segundo quando ambos estão na mesma instituição. Estas variáveis foram positivas e significativas. Vale ressaltar que o coeficiente da segunda variável possui uma magnitude maior. A literatura diverge a respeito desses resultados, sendo que estes resultados estão de acordo com os estudos de Crescenzi *et al*. (2017) e Broekel & Mueller (2018), enquanto Molina-Morales *et al* (2015) encontram um resultado negativo e pouco significativo após controlar por todas as proximidades e sugere que o papel da dimensão institucional reforça o da organizacional, podendo atuar como complementos.

As variáveis da quarta parte descrevem a proximidade organizacional. Essas variáveis binárias foram criadas a fim de captar a relação de pares que estão sob o mesmo tipo organizacional. A primeira *dummy*, que indica se o par já teve algum vínculo com universidade, mostra-se positiva e significava, assim como a *dummy* que indica se ambos declararam vínculo com empresas na plataforma. Estes resultados indicam que a proximidade organizacional ajuda na colaboração entre os agentes da rede, mesmo que em longas distâncias. O estudo de Hong & Su (2012) confirma esses resultados e fortalece o argumento de que a proximidade organizacional atua como um fator que “encurta” as distâncias geográficas entre pesquisadores.

A quinta parte mostra o coeficiente de variáveis binárias que indicam interação entre pesquisadores da mesma grande área do conhecimento. Como discutido na revisão de literatura, cada área do conhecimento tem sua particularidade e apresenta número de produções científicas em proporções diferentes. Por isso, é um fator importante de controle, pois mede a proximidade cognitiva entre os agentes. Os resultados mostram que todas as grandes áreas tiveram um resultado significativo e na direção de aumentar o número de ligações do par. As áreas que merecem destaque por terem coeficientes em valor absoluto maior são as Ciências Agrárias, as Ciências Exatas e da Terra, as Ciências Biológicas e Ciências da Saúde. Esses achados são condizentes com os números apresentados por Mena-Chalco *et al* (2014) que mostram maior produção e cooperação desses campos científicos.

As variáveis de controle do modelo são apresentadas na sexta parte. Optou-se por incluir as características dos dois pesquisadores para as variáveis de *betweenness* e tempo de docência. A centralidade de entrelaçamento apresentou um sinal positivo e significativo já esperado pela literatura, mas seu coeficiente foi pequeno em valor absoluto, pois os valores são contínuos e possuem uma média absoluta bem maior que as outras variáveis. Esse raciocínio também se emprega em relação ao tempo de docência, uma variável contínua. Entretanto, seu coeficiente é negativo e significativo. Isso é plausível na literatura, já que estudos mostram que a produtividade dos pesquisadores diminui com a idade (Bonaccorsi & Daraio, 2003).

Tabela 9 – Determinantes do número de publicações por pares de pesquisadores no Brasil (2009-2017).  
Estimação por Modelo Binomial Negativo Inflado. (continua)

Variáveis	1	2	3	4	5	6	7
<i>Proximidade social</i>							
Interação Prévia t-1	-	-	-	1,16*** (0,01)	1,11*** (0,01)	1,11*** (0,01)	-
Interação Prévia t-2	-	-	-	-	0,19*** (0,01)	0,20*** (0,01)	-
Interação Prévia t-3	-	-	-	-	-	-0,04** (0,01)	-
Interação Prévia Patente	-	-	0,30*** (0,01)	0,16*** (0,01)	0,16*** (0,01)	0,16*** (0,01)	-
Interação Contínua	-	-	-	-	-	-	0,86*** (0,02)
<i>Proximidade Geográfica</i>							
<i>Dummy</i> UF	1,05*** (0,01)	1,05*** (0,01)	1,00*** (0,01)	0,90*** (0,01)	0,90*** (0,01)	0,90*** (0,01)	0,96*** (0,01)
<i>Proximidade institucional</i>							
<i>Dummy</i> Tipo Institucional	-	-	0,02*** (0,01)	0,01** (0,01)	0,01 (0,01)	0,01 (0,01)	0,02*** (0,01)
<i>Dummy</i> Instituição	-	-	0,54*** (0,01)	0,41*** (0,01)	0,41*** (0,01)	0,41*** (0,01)	1,26*** (0,01)

Nota: Erro-padrão entre parênteses. Nível de significância: \* 5%, \*\* 1%, \*\*\*0,1%. Variável Dependente: Contagem do número de artigos publicados pelo par de pesquisadores. As áreas do conhecimento inclusas são Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas; Linguísticas, letras e artes; e Outros.

Tabela 9 – Determinantes do número de publicações por pares de pesquisadores no Brasil (2009-2017).  
Estimação por Modelo Binomial Negativo Inflado. (continuação)

Variáveis	1	2	3	4	5	6	7
<i>Proximidade organizacional</i>							
<i>Dummy</i> Universidade	-	0,26** (0,01)	0,05** (0,02)	0,07*** (0,02)	0,08*** (0,02)	0,08*** (0,02)	0,06** (0,02)
<i>Dummy</i> Empresa	-	0,09*** (0,01)	0,10*** (0,01)	0,08*** (0,01)	0,08*** (0,01)	0,08*** (0,01)	0,10*** (0,01)
<i>Dummy</i> Universidade Pública	-	0,26*** (0,01)	0,23*** (0,01)	0,23*** (0,01)	0,24*** (0,01)	0,24*** (0,01)	0,22*** (0,01)
<i>Controle</i>							
Ln( <i>Betweeness i</i> )	0,20*** (0,00)	0,20*** (0,00)	0,20*** (0,00)	0,19*** (0,00)	0,18*** (0,00)	0,18*** (0,00)	0,20*** (0,00)
Ln( <i>Betweeness j</i> )	0,23*** (0,00)	0,23*** (0,00)	0,23*** (0,00)	0,21*** (0,00)	0,21*** (0,00)	0,21*** (0,00)	0,23*** (0,00)
Ln(Tempo Docência <i>i</i> )	-0,04*** (0,00)	-0,05*** (0,00)	-0,05*** (0,00)	-0,05*** (0,00)	-0,06*** (0,00)	-0,06*** (0,00)	-0,05*** (0,00)
Ln(Tempo Docência <i>j</i> )	-0,09*** (0,00)	-0,09*** (0,00)	-0,09*** (0,00)	-0,10*** (0,00)	-0,10*** (0,00)	-0,10*** (0,00)	-0,09*** (0,00)
Ambos Doutores	0,19*** (0,01)	0,18*** (0,01)	0,18*** (0,01)	0,08*** (0,01)	0,08*** (0,01)	0,08*** (0,01)	0,17*** (0,01)
Constante	5,40*** (0,01)	-5,54*** (0,02)	-5,56*** (0,02)	-5,34*** (0,02)	-5,33*** (0,02)	-5,33*** (0,02)	-5,52*** (0,02)
<i>Dummies</i> de áreas do conhecimento <sup>1</sup>	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Nota: Erros-padrão entre parênteses. Nível de significância: \* 5%, \*\* 1%, \*\*\*0,1%. Variável Dependente: Contagem do número de artigos publicados pelo par de pesquisadores. <sup>1</sup>As áreas do conhecimento inclusas são Ciências Exatas e da Terra; Ciências Biológicas; Engenharias; Ciências da Saúde; Ciências Agrárias; Ciências Sociais Aplicadas; Ciências Humanas; Linguísticas, letras e artes; e Outros.



## 6 Conclusões

Este trabalho analisou, de forma pioneira, as dimensões de proximidade de rede que atuam de forma a facilitar a colaboração entre cientistas brasileiros, a saber, social, cognitiva, geográfica, organizacional e institucional. Em particular, o trabalho explorou essas dimensões ao nível de indivíduos, no contexto nacional onde há um claro aumento na produção científica. Essa abordagem já é explorada pela literatura internacional em países desenvolvidos, no entanto, em um país periférico, com um SNI em construção, como o Brasil, resultados como estes ainda são escassos. Cada vez mais é almejado entender o processo colaborativo entre os cientistas de diversas disciplinas, assim, pode-se entender a continuidade do avanço da fronteira do conhecimento. Neste contexto, a inserção das cinco proximidades de rede preenche uma lacuna, pois fornece evidências de um país em desenvolvimento. As proximidades organizacional e institucional também impactam positivamente o processo de colaboração, embora esta primeira não tenha um grande impacto em relação às outras dimensões. O maior coeficiente encontrado é o da variável *dummy* que mede o pertencimento à mesma instituição. Pertencer à mesma organização, como empresas, universidades e universidades públicas, mostra associação positiva com a produção científica dos pares. Estes achados mostram o papel importante da instituição a nível macro e micro, e podem indicar diluição da incerteza e oportunismo na formação de equipes de pesquisa.

As relações prévias mantidas pelos pesquisadores durante o triênio, que são representadas pela proximidade social, sugerem que essas interações proporcionam novas oportunidades de colaboração. O efeito é também positivo nos casos de pares de pesquisadores que já interagiram continuamente nos últimos dois triênios. Isso mostra que o processo de conhecimento mútuo entre os pares encoraja a produção científica exceto se os pares possuíram interações longínquas, ou seja, em mais de 6 anos no tempo (interações em  $t-3$ ). Por outro lado, interações contínuas dos mesmos pares de pesquisadores afetam positivamente a produção científica, o que também pode significar confiança mútua entre os pares de pesquisadores. A proximidade geográfica interfere positivamente na decisão de colaboração dos acadêmicos, ainda que seja medida neste estudo como pertencimento ao mesmo estado. Outros resultados como grau de centralidade dos pesquisadores na rede e tempo de docência possuem resultados que mostram associação positiva e negativa, respectivamente. Num caso, revela-se que a posição do pesquisador na rede afeta positivamente a colaboração dos pares. No outro, a produtividade decrescente com o tempo de docência revela a natureza do ciclo de vida do pesquisador, o qual com o do tempo se envolve com outras funções além da publicação de artigos científicos em si.

Este estudo tem algumas limitações que valem observar. Primeiro, leva em conta as colaborações que se formaram, ou seja, os insucessos nas produções em colaboração são desconsiderados. Segundo, somente interações acadêmicas que levam ao produto artigo são consideradas, sejam elas formais ou informais. Terceiro, as variáveis de proximidade são todas representadas por *dummies*; e a unidade de referência geográfica é a Unidade da Federação, o que deve implicar em menos precisão da dimensão geográfica, visto que os estados brasileiros têm grande dimensões.

## Referências

ALBUQUERQUE, E. M. Notas sobre os determinantes tecnológicos do processo de catching-up: uma introdução à discussão sobre o papel dos sistemas nacionais de inovação na periferia. **Estudos Econômicos**, São Paulo, v.27(2), 1997.

ANSELIN, L.; VARGA, A.; ACS, Z. Geographical Spillovers and University Research: A Spatial Econometric Perspective. **Growth and Change** v. 31, p. 501-515, 2000.

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

- BELUSSI, F., CALDARI, K. At the Origin of the Industrial District: Alfred Marshall and the Cambridge School.” **Cambridge Journal of Economics** v. 33(2), p. 335–355, 2009.
- BERNARDES, A.; ALBUQUERQUE, E. Cross-Over, thresholds and the interactions between Science and technology: lessons for less-developed countries. **Research Policy**, v. 32(5), p. 867-887, 2003.
- BONACCORSI, A.; DARAIIO, C. Age effects in scientific productivity The case of the Italian National Research Council (CNR) **Scientometrics**, v. 58(1), p. 49-90, 2003.
- BORGATTI, S. P.; EVERETT, M. G.; JOHNSON, J. C. **Analyzing Social Networks**, London: SAGE Publications Ltd, 2013.
- BOSCHMA, R. Proximity and Innovation: A Critical Assessment. **Regional Studies**, v. 39, 61-74, 2005.
- BOSCHMA, R. A.; FRENKEN, K. The special evolution of innovative networks: a proximity perspective. In: BOSCHMA, R. A.; MARTIN, R. L. (Ed). **The handbook of evolutionary economic geography**, Edward Elgar, Cheltenham, 20
- BROEKEL, T.; BOSCHMA, R. Knowledge networks in the Dutch aviation industry: the proximity paradox **Journal of Economic Geography**, v. 12(2), p. 409–433, 2012.
- BROEKE, T.; MUELLER, W. Critical links in knowledge networks – What about proximities and gatekeeper organizations? **Industry and Innovation**, v. 25(10), p. 919-939, 2018.
- CAPALDO, A.; PETRUZZELLI, M. Partner Geographic and Organizational Proximity and the Innovative Performance of Knowledge-Creating Alliances **European Management Review**, v. 11, p. 63–84, 2014.
- COHEN, W.M.; LEVINTHAL, D.A. Absorptive capacity: a new perspective of learning and innovation. **Administrative Science Quarterly**, v. 35, p. 128–152, 1990.
- COHENDET, P., LLERENA, P. (1997) Learning, technical change and public policy: how to create and exploit diversity. In C. Edquist (ed.) **Systems of Innovation**, p. 223–241, 1997.
- CRESCENZI, R.; NATHAN, M.; RODRIGUEZ-POSE, A. Do inventors talk to strangers? On proximity and collaborative knowledge creation **Research Policy** v. 45(1), p. 177-194, 2016.
- CRESCENZI, R.; FILIPPETTI, A.; IAMMARINO, S. Academic inventors: collaboration and proximity with industry **Journal of Technological Transfer**, 2017.
- DANELL, R. The contribution of authors: A study of the relationship between the size and composition of author teams **Ed Noyons**, Leiden, Netherlands: Leiden university, p. 123-13, 2014.
- DUCTOR, L. Does Co-authorship Lead to Higher Academic Productivity? **Oxford Bulletin of Economics and Statistics**, v. 77(3), p. 385-407, 2015.
- FERNÁNDEZ, A.; FERRÁNDIZ, E.; LEÓN, M. D. Proximity dimensions and scientific collaboration among academic institutions in Europe: The closer, the better? **Scientometrics**, v. 106(3), p. 1073-1092, 2016.

- HERINGA, P.; HESSELS, L. K.; VAN DER ZOUWEN The influence of proximity dimensions on international research collaboration: an analysis of European water projects **Industry and Innovation**, v. 23(8), p. 753-772, 2016.
- HERINGA, P.; HORLINGS, E.; VAN DER ZOUWEN, M.; VAN DEN BESSELAAR, P.; VAN VIERSSEN W. How dimensions of proximity relate to the outcome of collaboration? A survey of knowledge-intensive networks in the Dutch water sector **Economics of Innovation and New Technology**, v.23(7), p. 689-716, 2014.
- HILBE, Joseph M. Negative binomial regression. **Cambridge University Press**, 2011.
- HONG, W.; SU, Y. S. The effect of institutional proximity in non-local university–industry collaborations: An analysis based on Chinese patent data **Research Policy** v. 42, p. 454-464, 2013.
- KAYGALAK, I.; REID, N. Innovation and knowledge spillovers in Turkey: The role of geographic and organizational proximity: Innovation and knowledge spillovers in Turkey **Regional Science Policy & Practice** v. 8(1-2), p. 45-60, 2016.
- LAGENDIJK, A.; LORENTZEN, A. Proximity, Knowledge and Innovation in Peripheral Regions. On the Intersection between Geographical and Organizational Proximity **European Planning Studies**, v. 15(4), 2007.
- LATA, R.; PROFF, S.; BRENNER, T. The influence of distance types on co-patenting and co-publishing in the USA and Europe over time. **The Annals of Regional Science**. v. 61 (1), Heidelberg, 2018.
- MARSHAL, A. Principles of Economics: An Introductory Volume, Londres: Macmillan, 1920.
- MCMANUS, C.; BAETA NEVES, A.A.; MARANHÃO, A.Q.; SOUZA FILHO. A. G.; SANTANA, J. M. International collaboration in Brazilian science: financing and impact. **Scientometrics** v. 125, p. 2745–2772, 2020.
- MENA-CHALCO, J. P.; DIGIAMPIETRI, L. A.; LOPES, F. M.; CESAR JR R. M. Brazilian Bibliometric Coauthorship Networks. *Journal of the Association for Information Science and Technology*, v. 65(7), p. 1424–1445, 2014.
- NOOTEBOOM, B. Learning and Innovation in Organizations and Economies. **Oxford: Oxford University Press**, 2000.
- NOOTEBOOM, B., VAN HAVERBEKE, W., DUYSTERS, G., GILSING, V., VAN DEN OORD, A. Optimal cognitive distance and absorptive capacity. **Research Policy**, v. 36, p. 1016–1034, 2007
- PACI, R.; MARROCU, E.; USAI, S. The Complementary Effects of Proximity Dimensions on Knowledge Spillovers **Spatial Economic Analysis** v. 9(1), p. 9-30, 2014.
- PORAC, J. F.; WADE, J. B.; FISCHER, H. M.; BROWN, J.; KANFER, A.; BOWKER, G. Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: a comparative case study of two scientific teams **Research Policy**, v. 33(4), p. 661-678, 2004.

RAPINI, M. D.; RIGHI H. M. Interação Universidade-Empresa no Brasil em 2002 e 2004: Uma Aproximação a Partir dos Grupos de Pesquisa do CNPq. **Revista EconomiA**, v. 8(2), p. 248-268, 2006.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. **Transinformação**, v. 28(1), p. 15-32, 2016.

SORENSEN, O.; RIVKIN, J. W.; FLEMING, L. Complexity, networks and knowledge flow *Research Policy* v. 35 p. 994–1017, 2006.

SMITH, C; SOTALA, K. Knowledge, networks and nations Global scientific collaboration in the 21st century **The Royal Society** v. 03/11, 2011.

SONNENWALD, D. H. Scientific collaboration **Annual Review of Information Science and Technology**, v. 41, p. 643-681, 2007.