

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

DANIEL OLIVEIRA COSTA

JUIZ DE FORA

2021

DANIEL OLIVEIRA COSTA

Condicionantes da colaboração acadêmica no sistema científico brasileiro

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Economia Aplicada.

Área de concentração: Economia

Orientador: Prof. Dr. Eduardo Gonçalves

Coorientador: Prof. Dr. Renato de Castro Garcia

JUIZ DE FORA

2021

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Oliveira Costa, Daniel.

Condicionantes da colaboração acadêmica no sistema científico brasileiro / Daniel Oliveira Costa. -- 2021.

68 p.

Orientador: Eduardo Gonçalves

Coorientador: Renato de Castro Garcia

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2021.

1. Conhecimento científico. 2. C&T. 3. Análise de Redes Sociais. 4. Dimensões de proximidade. I. Gonçalves, Eduardo, orient. II. de Castro Garcia, Renato, coorient. III. Título.

Daniel Oliveira Costa

Condicionantes da colaboração acadêmica no sistema científico brasileiro

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Economia Aplicada.

Área de concentração: Economia

Aprovada em 11 de Agosto de 2021

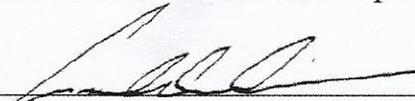
BANCA EXAMINADORA



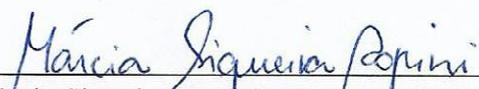
Dr. Eduardo Gonçalves - Orientador
Universidade Federal de Juiz de Fora



Dr. Renato de Castro Garcia - Co-orientador
Universidade Estadual de Campinas



Dra. Laura de Carvalho Schiavon – Membro titular interno
Universidade Federal de Juiz de Fora



Dra. Márcia Siqueira Rapini – Membro titular externo
Universidade Federal de Minas Gerais

AGRADECIMENTOS

Seria injusto resumir todos aqueles que me ajudaram nesta única página. Assim como me parece pequena essa sessão para demonstrar o quanto sou grato àqueles que incentivaram, apoiaram e fizeram que este trabalho tenha um significado. Tal feito não seria possível sem a contribuição de amigos, colegas e familiares que, me apoiaram incansavelmente, me mostraram um caminho que as vezes não me parecia claro e se prontificaram sem hesitar quando necessitava.

É difícil de mensurar a minha gratidão que tenho às duas mulheres mais importantes da minha vida: minha mãe Celeida e minha vó Aparecida. Duas mulheres que fizeram de tudo pelos seus filhos. Mulheres fortes e determinadas que me guiaram e lutaram por mim até a fase adulta. Obrigado por me instruírem da melhor forma possível e me ensinarem o caminho das pedras que me tornou o homem que sou hoje.

Agradeço ao meu pai por ter me apoiado nas melhores decisões e me inspirado com sua história de vida. Agradeço a meu irmão Miguel e ao meu primo Marcos, me ouviram desabafar muito nesses anos e foram muito importantes para que eu continuasse a me esforçar e dar o meu melhor. À minha família devo toda gratidão por ser meu alicerce e ser meu refúgio em momentos tensos.

A todos os meus amigos que encontrei no curso devo lembrá-los da importância de vocês na minha formação e por isso eu agradeço. Já em pouco tempo nos tornamos muito próximos e sabia que podia sempre contar, não importa o tipo de problema. Meu imenso obrigado: Beatriz, Heloísa, Lucas, Marina, Cláudio, Matheus e João Augusto. Me lembro da primeira semana onde tive o prazer de apresentar à Bia o famoso torresmo, esse foi o primeiro dia de muitos que demos muitas risadas e conseguimos nos descontrair. A Heloísa com seu estilo gótico suave sempre deixava nosso clima mais leve, mesmo nos momentos mais tensos da faculdade, exceto quando ela está com fome, melhor não interagir e esperar para quando tomarmos nosso cafezinho depois do RU. Lucas, obrigado por me ouvir e pelos seus memes sensacionais e pela fábrica figurinhas que nos representava em cada etapa do mestrado. Mauro, obrigado pelo seu ombro amigo que se dispôs a qualquer, seja para desabafar ou para dividir novas conquistas.

Meus sinceros agradecimentos aos meus mestres que tiveram muita paciência na minha orientação e que ensinou o quanto a ciência é importante. Obrigado, Eduardo e Renato. Agradeço à Verônica pela enorme ajuda no meu trabalho de dissertação, a FAPEMIG e a UFJF pelo apoio financeiro.

Meus sinceros agradecimentos.

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; C&T; Análise de Redes Sociais; Proximidade

ABSTRACT

The meaning of this work is to understand the influence of social, institutional, organizational, cognitive and geographic proximity in the collaborative context among Brazilian researchers in the period between 2009 and 2017. Empirical works discuss the role of proximity in developed countries, but there is a lack of works that address the theme in developing countries. Furthermore, although there is a plenty of papers about proximities, 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, data obtained from the Lattes Platform of the National Council for Scientific and Technological Development (CNPq) were used. The results show that all productivity dimensions are relevant to explain the publications of research pairs 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. Furthermore, previous and ongoing interactions between researchers are also relevant constraints.

Key words: Scientific knowledge; C&T; Social Network Analysis; Proximity

Lista de Gráficos

Figura 1 – Rede ilustrativa com nós e vértices fictícios.....	37
Figura 2 - Rede de colaborações dos pesquisadores 1% mais produtivos por diferentes áreas de conhecimento.....	44
Gráfico 1 – Distribuição de número de coautorias por artigo 2000-2002.....	31
Gráfico 2 – Distribuição de número de coautorias por artigo 2003-2005.....	31
Gráfico 3 – Distribuição de número de coautorias por artigo 2006-2008.....	31
Gráfico 4 – Distribuição de número de coautorias por artigo 2009-2012.....	31
Gráfico 5 – Distribuição de número de coautorias por artigo 2012-2014.....	31
Gráfico 6 – Distribuição de número de coautorias por artigo 2015-2017.....	31
Gráfico 7 – Média do número de autores por artigo ao longo do tempo 2000-2018.....	43

Lista de Tabelas e Quadros

Tabela 1 – Estatísticas de dispersão em torno da média da variável dependente 2009-2017...	29
Tabela 2 – Descrição e distribuição da variável social de ligações prévias no tempo 2009-2017.....	32
Tabela 3 - Valores das métricas de rede da Figura 1.....	39
Tabela 4 – Descrição das medidas de rede das regiões brasileiras.....	40
Tabela 5 – Características das Grandes Áreas do Conhecimento na rede de colaboração.....	41
Tabela 6 – Estatísticas descritivas e de rede em relação à titulação dos pesquisadores 2000-2017.....	42
Tabela 7 – Estatísticas descritivas da rede de colaboração dos triênios, Brasil, 2000-2017....	43
Tabela 8 – Correlação entre as variáveis do modelo.....	45
Tabela 9 – Divisão do contrafactual nos triênios.....	46
Tabela 10 – Estatísticas descritivas das variáveis de interesse do total de observações.....	47
Tabela 11 – Estatísticas descritivas das variáveis de interesse das observações em que houve ligação entre os pesquisadores.....	48
Tabela 12 – Condicionantes do número de publicações por pares de pesquisadores no Brasil (2009-2017). Estimação por Modelo Binomial Negativo Inflado.....	50
Tabela 13 – Condicionantes do número de publicações por pares de pesquisadores no Brasil em subamostras de acordo com Unidade da Federação e grande área do conhecimento (2015-2017). Estimação por Modelo Binomial Negativo Inflado.....	54
Quadro 1 – Literatura empírica de proximidades em rede.....	24
Quadro 2 – Descrição das variáveis independentes	34

LISTA DE SIGLAS

ARS	-	Aplicações de Análise de Rede Social
BNIZ	-	Binomial Negativo Inflado em Zeros (BNIZ)
CAPES	-	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	-	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CTI	-	Ciência, Tecnologia e Inovação
EMBRAPA	-	Empresa Brasileira de Pesquisa Agropecuária
ERM	-	Érdos-Renyi
SNI	-	Sistema Nacional de Inovação

Sumário

1. INTRODUÇÃO	10
2. REVISÃO DE LITERATURA	13
2.1 A colaboração na ciência e no contexto nacional.....	13
2.1 Aplicações de Análise de Rede Social (ARS) em processos de colaboração científica.....	14
2.2 Proximidade de rede na colaboração científica	17
2.2.1 Proximidade Geográfica	18
2.2.2 Proximidade Cognitiva	19
2.2.3 Proximidade Social	20
2.2.4 Proximidade Institucional	22
2.2.5 Proximidade Organizacional.....	22
2.3 Estudos empíricos sobre as proximidades na colaboração científica.....	23
2. ESTRATÉGIA EMPÍRICA	28
3.1 Modelo e método	28
3.2 Variáveis explicativas.....	30
3.3 Métricas de rede.....	36
3.4 Descrição da base de dados	39
3.5 Construção do contrafactual	46
3. RESULTADOS	49
3.1. Resultados da estimação de todas as proximidades	49
4. CONCLUSÃO	57
5. Referências	60

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 pode impulsionar 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 colaboração científica é definida por Sonnenwald (2007) como interações dentro de um contexto social entre cientistas que favorece o compartilhamento de ideias, significados e tarefas com um objetivo em comum. Além de a colaboração estar mais frequente, o número de pesquisadores em um artigo vem aumentando e as publicações em coautorias constantemente recebem mais citações, tornando-se métricas de qualidade (Persson *et al*, 2004).

A ciência já se mostrou como um dos pilares para o crescimento da economia em um país. Modelos clássicos de crescimento a longo prazo se baseiam na ideia de que a ruptura da estrutura tecnológica tem como insumo a produção científica nacional (Romer, 1994). As instituições acadêmicas contribuem significativamente para o desenvolvimento social, tecnológico e econômico por meio de externalidades que criam condições e oportunidades de crescimento. A maturação do meio científico impõe, de maneira crescente, maior complexidade às pesquisas. Devido a esse avanço na fronteira do conhecimento, a produção científica -se

voltado para a inclusão de diversos ramos do conhecimento acadêmico. Pesquisadores almejam colaborações interdisciplinares com objetivo de alcançar desenvolvimento na sua área de pesquisa (McCarty & Jawitz, 2013).

A produção científica brasileira vem crescendo e, em consonância outros países desenvolvidos, bem como a participação de trabalhos em coautoria também. Em uma visão global, as parcerias brasileiras também têm se mostrado crescentes e impactado positivamente a manutenção do estado-da-arte. Tal aumento ocorre em paralelo aos financiamentos de agências internacionais, e representa uma cooperação benéfica para ambas as partes. As colaborações incentivam a formação da Ciência, Tecnologia e Inovação (CTI) e de um SNI nacional (Leta & Chaimovich, 2002; Boekholt *et al*, 2009). No entanto, apesar do notório progresso em criar laços com outros países, o Brasil ainda apresenta níveis inferiores de produção e internacionalização científicos em comparação aos países iberos e enfrenta escassez em termos de recursos e incentivos nacionais (McManus, 2020).

A produção das universidades é movida por pesquisadores cientistas cujo interesse está focado em publicações em revistas de alto impacto, pautadas por um sistema de revisão por pares (Geisler & Rubenstein, 1989). Com intuito de aumentar a sua produtividade, o impacto do seu artigo e conquistar prestígio na sua área de pesquisa, muitos deles procuram outros pesquisadores para colaborar em seus trabalhos e complementar suas habilidades. Essas interações dependem das características individuais de cada pesquisador e estão sujeitas aos objetivos das instituições das quais fazem parte (Bodas Freitas et al, 2013).

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 nos triênios passados a fim de explicar as coautorias no período de análise corrente. Essa, em conjunto com as demais dimensões, fornecem 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 é apresentada a conclusão e as limitações do trabalho.

2. REVISÃO DE LITERATURA

2.1 A colaboração na ciência e no contexto nacional

A ciência acompanha o progresso e o desenvolvimento de uma sociedade. Através dela é possível influenciar o crescimento e a maturação da atividade econômica. À medida que se torna cada vez mais importante para a inovação e o bem estar da sociedade, cresce a necessidade de entender como se dá a evolução do descobrimento de novos conhecimentos no meio científico (Hey, Tansley, & Tolle, 2009). Essa análise pode ser aplicada por meio da identificação de padrões, de modelos econométricos e da visualização de relação entre os pesquisadores, os círculos sociais e grupos de pesquisa a partir de bases de dados de repositórios de produção científica (Martins, Gonçalves, Laender, & Ziviani, 2010; Persson & Beckmann, 1995; Uddin, Hossain, Abbasi, & Rasmussen, 2012).

O mundo acadêmico é guiado por etapas do método científico: observação, mensuração, formação de hipóteses, experimento, formulação, testes e conclusão. Esse sistema que rege a ciência desde seu início, passou a incluir a revisão por pares como critério técnico, apenas nas últimas décadas. Uma vez submetido a uma revista científica o artigo é analisado por especialistas da área e, mediante a qualidade e honestidade, poderá ser aprovado para publicação, rejeitado ou receber sugestão para alterações (Arraez, 2011).

A ciência tem se tornado nos últimos tempos cada vez mais colaborativa e, em consequência disso, novas organizações são formadas a fim de facilitar a colaboração entre os cientistas e aumentar a sua produtividade (Chompalov *et al*, 2001; Lee & Bozeman, 2005). Essa tendência tem modelado diferentes necessidades da comunidade acadêmica, como: os fundos de pesquisa, a profissionalização dos pesquisadores, a crescente necessidade de recursos à medida que o campo de pesquisa se torna mais complexo, a evolução de disciplinas com alta especialização, o uso de novas tecnologias de comunicação e o desejo de pesquisadores por visibilidade e produtividade profissional (Katz & Martin, 1997).

Em consonância a esse novo cenário, as colaborações trazem novo fôlego às pesquisas que necessitam de novas ideias, habilidades e coordenação de equipe. O grupo de pesquisa pode ser formado por indivíduos homogêneos, ou seja, que têm em comum interesse em resultados semelhantes e habilidades parecidas, o que leva a um melhor desempenho na comunicação e de uma equipe diversa (Bantel & Jackson, 1989). Esta última é apresentada pela literatura como indivíduos que têm habilidades, perspectivas e conhecimentos diferentes, adentraram na academia em tempos diferentes e, por consequência, conhecem pessoas de diferentes áreas. As

equipes diversas sofrem menos com resultados possivelmente redundantes (Ancona & Caldwell, 1992; Reagans & Zuckerman, 2001).

Existem diferenças significativas nos padrões de cada disciplina da ciência, a relação de colaboração entre as grandes áreas do conhecimento é uma delas. Ciências biológicas e física têm o trabalho em conjunto como algo corriqueiro devido aos seus métodos de pesquisas, já áreas como história e literatura são mais frequentes os trabalhos de autoria única (Farber, 2005). Colaborações em ciências sociais são bem menos frequentes se comparadas aos domínios científicos como biotecnologia, química e física (Yan, 2014). Este cenário da literatura é também parecido com o brasileiro (Mena-Chalco *et al*, 2014)

A ciência no contexto brasileiro também tem acompanhado o cenário internacional. Isto é, verifica aumento do número de publicações, os artigos possuem mais coautorias, há uma redução produções de autoria única e uma tendência para a estruturação de equipes multidisciplinares (Vanz & Stumpf, 2012; Mena-Chalco *et al*, 2014; Sidone *et al*, 2016). Há de se destacar uma grande concentração das publicações e das colaborações nos estados do Sul e Sudeste brasileiro, favorecidas pela grande quantidade de universidades tradicionais, pelo acesso aos recursos e disponibilidade de capital humano (Suzigan & Albuquerque, 2011). No entanto, essa centralização vem diminuindo e outras regiões vem apresentando um aumento na sua participação. As pesquisas por área de conhecimento são heterogêneas nas regiões brasileiras, disciplinas de Ciências da Saúde e Ciências Biológicas são as mais díspares, tendo uma concentração maior na região Sudeste, em especial no estado de São Paulo, enquanto Ciências Agrárias e Linguística, Letras e Artes têm um grau de concentração menor (Sidone *et al*, 2016).

2.1 Aplicações de Análise de Rede Social (ARS) em processos de colaboração científica

O alicerce de uma rede são os nós e representam os agentes da sociedade que interagem por meio de vértices ou ligações (Borgatti & Everett, 2013). No meio científico, essas ligações podem ser definidas pelas colaborações em coautoria, visto que o artigo pode ser feito por mais de um autor. Quando há colaboração na produção acadêmica, gera-se o chamado vértice. Isso se traduz em uma rede colaborativa acadêmica, facilitando o objetivo mútuo entre os agentes.

Dito como a rede é baseada, deve-se entender, dentro do contexto de coautoria, como as colaborações se tornam vértices. As colaborações em um texto publicado em revista recebem autorias de acordo com as participações nesse artigo. Sendo assim, são caracterizados como documentos científicos em coautoria os que têm dois ou mais envolvidos. Essas coautorias

provêm de diversas formas de colaboração, como por exemplo..., mas também podem resultar de outras causas, como a requisição dos nomes dos líderes de pesquisa sem que, de fato, tenham contribuído para o trabalho (Melin & Persson, 1996). Além do universo de autores com pouca ou nenhuma contribuição, existem colaborações que não resultam em coautoria, mas levam a outros produtos como patentes, formação de laços pessoais ou, até mesmo, nada. Os estudos que utilizam coautoria para construção de rede analisam os aspectos de interação entre os pesquisadores, seu grau de importância dentro da rede para transferência de conhecimento, levando em conta as métricas de rede e característica de seus componentes. Esse último tende a ter estruturas de rede divergentes quando analisados em separado de acordo com as disciplinas (Newman, 2001). Ainda assim, mesmo com essa diferenciação entre as grandes áreas do conhecimento, é de comum senso de que a colaboração vem se tornando cada vez mais frequente nos campos de pesquisa (Bahr & Zemon, 2000). Cientistas tendem a colaborar com o intuito de adquirir novos conhecimentos e aumentar a especialização dentro da sua área. Desse modo, é possível resolver problemas mais complexos com o novo conhecimento compartilhado (Sonnenwald, 2007; Katz & Martin 1997). De acordo com a cooperação formada entre os autores, a rede toma forma de um emaranhado, no qual existem atores principais – que possuem posições estratégicas na troca de conhecimento – e periféricos – que atuam na diversificação da base de conhecimento. Esses atores principais podem ser caracterizados por muitas ligações ou por ter ligações com atores em posições estratégicas na rede, o que lhes confere vantagens no acesso ao conhecimento. Contudo, a posição destas instituições na rede afeta a produtividade segundo a estrutura e composição de diferentes contextos interorganizacionais desses nós (Chen *et al*, 2017).

Ainda dentro do arcabouço metodológico de ARS, existem métricas que ajudam a entender o desempenho e as relações entre os cientistas. As medidas de centralidade de rede auxiliam no entendimento das posições dos pesquisadores na rede, assim como sua capacidade de formar novas coautorias. Autores com alta centralidade e que atuam como mediadores de fluxos de informação, conhecimento e comunicação são mais atrativos e até preferíveis para novas parcerias (Abassi *et al*, 2012). O número de coautorias de um par de pesquisadores determina a força desse laço na rede. Um forte grau de ligação entre dois cientistas em uma rede densa é necessário para uma troca de conhecimento mais complexa. As ligações mais fracas falham neste objetivo, embora, exerçam um papel muito importante na velocidade de transmissão do conhecimento, já que incorporam aprendizado de pesquisadores distantes (Hansen, 1999; Perkmann & Walsh, 2007; Bergenholtz & Bjerregaard, 2014).

As estruturas de ligação entre atores na rede influenciam no tipo de conhecimento produzido. Áreas metropolitanas com centros especializados em certo tipo de conhecimento tendem a se associar com redes sociais mais densas e robustas de coinventores, tanto para ligações internas quanto para externas, aproveitando os benefícios da urbanização e da aglomeração.

A área de biotecnologia é bastante referida pela literatura, por ter redes densas, ligações fortes com instituições do setor privado, as quais muitas se baseiam em relação de confiança, além de ligações fracas com intuito de procurar novos conhecimentos no exterior da rede (Kreiner & Schultz, 1993; Libesking *et al*, 1996; Bounty, 2000; Bjerregaard, 2014). Por outro lado, regiões metropolitanas com centros de conhecimentos diversificados têm relações sociais menos desenvolvidas que as cidades especializadas e não possuem grande impacto na taxa de invenção em relação às cidades com aglomeração (Van der Wounden & Rigby 2017).

As grandes áreas do conhecimento se mostram bem distintas em quantidade de colaboração. Mena-Chalco *et al* (2014) pontuam que áreas como Ciências Biológicas, Ciências Agrárias e Ciências da Saúde possuem um número muito maior de publicações e colaborações em relações a áreas como Ciências Sociais Aplicadas, Ciências Humanas e Linguísticas, Letras e Artes.

No Brasil, estudos de ARS utilizam patentes, grupos de pesquisas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e coautorias (Porto, 2011; Mena-Chalco *et al*, 2014; Reis *et al*, 2018; Araújo *et al*, 2018; Gonçalves *et al*, 2019). No entanto, o uso de coautoria na literatura brasileira é recente (Mena-Chalco *et al*, 2014). Este estudo descreve os artigos de grandes áreas do conhecimento e analisa as de maior impacto com medidas de qualidade estabelecidos pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

A literatura sugere que firmas que utilizam do conhecimento científico possuem maior qualidade em suas invenções (Meyer, 2000). Três fontes de interação entre o meio científico e o meio produtivo são essenciais para produção de novos inventos. Primeiramente, as externalidades geradas pelas universidades são aproveitadas de diversas formas pelas empresas, como estágios, formação de capital humano e externalidades. Uma segunda oportunidade econômica advém das conexões geradas por pesquisadores treinados por indústrias, o que facilita a comunicação entre essas instituições. A última fonte se alimenta do conhecimento

livre gerado pela universidade, a fim de aumentar o estoque de conhecimento da sociedade (David *et al*, 1992; David, 1993).

Empresas intensivas em Pesquisa e Desenvolvimento (P&D) tendem a ter mais ligações com universidades. Essa maior probabilidade de ligação com o setor acadêmico também é evidenciada em empresas que buscam outros tipos de conhecimento externo, como conferências, exposições, associações comerciais, instituições privadas e governamentais (Laursen & Salter, 2004). De fato, empresas investem em P&D com a intenção de assimilarem novos conhecimentos externos (Cohen & Levinthal, 1989; Cohen & Levinthal, 1990). O investimento em P&D e capital humano não necessariamente tem como foco novos inventos radicais, mas também visam desmitificar conhecimentos externos a partir do licenciamento de patentes ou por meio da engenharia reversa. A transferência de tecnologia da universidade à firma é um desafio em países com SNI imaturo, pois nem todas as firmas são intensas o suficiente em P&D para que essa interação seja bem sucedida, já que para a difusão do conhecimento necessita de capital humano capacitado. Entretanto, indústrias que não possuem esses atributos continuam contando com institutos de pesquisa públicos que atuam como substitutos do setor de P&D nas firmas (Albuquerque *et al*, 2015).

O sucesso no meio acadêmico no SNI é palco de debate na literatura. Há estudos que tentam relacionar a produção do conhecimento com elementos geográficos, sociais, culturais, organizacionais e institucionais (Hall *et al*, 2010; Gittelman, 2007; Branstetter & Ogura, 2005; Powell *et al*, 1996). Embora cada variável exerça uma influência já esperada, a intensidade e até a causalidade podem se alterar em regiões e setores distintos (Lata *et al*, 2018).

2.2 Proximidade de rede na colaboração científica

Vários fatores podem influenciar o pesquisador na tomada de decisão com quem interagir a fim de produzir novos artigos e maximizar o produto de inovação. O trabalho acadêmico em conjunto pode levar a estudos com maior qualidade, com mais eficiência, com o compartilhamento da *expertise* e a diminuição de custos. Mesmo os cientistas renomados estão à procura de coautorias que podem complementar seu domínio acadêmico no intuito de fomentar novas ideias, aproveitar habilidades de ambos e compartilhar acesso a diferentes fontes de conhecimento (Smith & Sotola, 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 o avanço na fronteira do conhecimento trazido por

essa colaboração resultam 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 (tecnológica), 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).

2.2.1 Proximidade Geográfica

A proximidade geográfica já é bastante discutida pela literatura. Devido a maior facilidade de transferência de conhecimento, principalmente tácito. Pesquisadores geograficamente próximos têm maior propensão a colaborar. A tentativa de identificá-la vem evoluindo na literatura com diversas metodologias e abordagens. Consistindo nos transbordamentos regionais, essa variável foi descrita por Marshall (1890) em seu trabalho seminal sobre difusão de conhecimento no desenvolvimento econômico. Após seu trabalho muitos economistas vêm tentando explicar esse fenômeno referido como “as ideias estão no ar”. Esta metáfora insinua que, em certas regiões, o espraiamento de conhecimento específico é dinâmico e por ser localizado, a absorção daquele saber é facilitado.

O desempenho inovativo de certas localidades sugere um papel transformador às universidades, mesmo que concentradas em setores específicos e atendendo à demanda industrial da região (Anselin *et al.*, 2002). A produção de externalidades positivas, formação de pessoal capacitado e geração de uma base de conhecimento robusta são fatores que explicam o desenvolvimento de setores intensivos em P&D (Karlsson & Gråsjö, 2012).

Embora os trabalhos apontem para uma concentração espacial inovativa, deve-se atentar a outros fatores que possam causar indiretamente a concentração geográfica desse polo tecnológico. O conhecimento utilizado na produção de novas invenções pode ser classificado em conhecimento tácito e codificado. Ao contrário do último, a difusão do conhecimento tácito requer contato direto com detentor deste saber (Nonaka & Krogh, 2009). Fazem-se necessárias a prática e a habilidade que são desenvolvidas presencialmente, cuja transferência envolve intuição e experiências físicas. Logo, invenções que dependem de conhecimento tácito e,

portanto, exigem muito contato com outros inventores podem tornar essa rede de colaboração concentrada espacialmente. Neste caso, a concentração espacial é causada pela cooperação em rede (Breschi & Lissoni, 2009).

A concentração geográfica de produção de conhecimento é amplamente amparada pela literatura, entretanto, 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. A mobilidade do trabalho é um fator que ajuda na difusão do conhecimento e está relacionada com esses achados, pois quando o indivíduo está em outra organização, o conhecimento pode se difundir (Kerr, 2008; Felsenstein, 2011), assim como as proximidades social, organizacional e institucional são variáveis que influenciam na localização. Devido a isso, para identificar o efeito locacional é preciso controlar estas variáveis (Boschma, 2005).

2.2.2 Proximidade Cognitiva

O conhecimento produzido pela academia nem sempre é acessível, o que o torna um bem não público. A difusão do conhecimento muitas vezes está sujeita a conexões com instituições e agentes que têm domínio desse saber. 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).

O setor produtivo toma a decisão de colaborar de acordo com as limitações de suas proximidades e sob incertezas. Este procura interagir com a academia para potencializar o produto do seu setor de P&D. Para isso, é considerado o conhecimento disponível em ambas as organizações, a capacidade de absorção e a aprendizagem do seu setor de P&D (Nooteboom, 2008). Contudo, a IUI enfrenta entraves por ser heterogênea em suas normas e objetivos. Foray & Lissoni (2010) descreve essas diferenças como:

“Transferências diretas de conhecimento entre as comunidades científicas acadêmicas e as organizações de P&D do setor privado são especialmente problemáticas de institucionalizar, uma vez que a coexistência de dois sistemas de recompensa típicos de cada sistema torna o comportamento dos participantes difícil de antecipar e tende a minar o

estabelecimento de normas culturais coerentes para a promoção da cooperação entre os membros da equipe (David et al., 1999). É evidente que as dificuldades de transferência de tecnologia não são causadas em primeira instância por estruturas institucionais, sistemas jurídicos ou normas culturais inadequados ou mal adaptados. Em vez disso, as dificuldades estão inerentemente associadas ao próprio processo, e todos os países enfrentam o mesmo problema, que consiste em administrar um *trade-off* entre duas coisas boas: obter mais conhecimento acadêmico utilizado pela economia versus manter as missões fundamentais (longo prazo pesquisa e educação) das universidades.” Foray & Lissoni (2010, p. 283, tradução livre)

A proximidade cognitiva se refere à base de conhecimento compartilhada pelos mesmos atores na rede. Indivíduos em diferentes organizações, mas no mesmo nicho tecnológico, tendem a absorver conhecimento um do outro com maior eficiência. Por estarem em uma área do conhecimento similar, a produtividade inventiva pode diminuir caso essa interação seja restrita aos mesmos agentes (Cohen & Levinthal, 1990).

Tal tipo de transferência de conhecimento pode se tornar prejudicial caso os atores estejam próximos demais. É esperado que a proximidade cognitiva assuma um formato de U invertido na produtividade inventiva (Cohendet & Llerena, 1997). Existe um *trade-off* entre maior absorção de conhecimento e um grau maior de inovação que é enfrentado no ato de colaboração, portanto, uma proximidade muito alta nem sempre reflete maior grau inventivo (Nooteboom, 2000; Nooteboom *et al.*, 2007; Broekel & Boschma, 2012).

2.2.3 Proximidade Social

A proximidade social é moldada a partir de interações anteriores, ou seja, já existentes na rede. As relações sociais 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 atores na rede, sejam elas já predefinidas ou que venham a ocorrer após a interação (Borgatti, 2013). Características similares entre os indivíduos, que os fazem interagir, são chamadas de homofílicas e exercem influência nas relações sociais (Alba & Kadushin, 1976).

Experiências sociais informais que levam o indivíduo a ter 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. Como citado na seção de proximidade geográfica, muitas interações sociais são geograficamente localizadas.

Estudos que desconsideram a relação social podem estar explicando parte deste efeito na proximidade geográfica.

A literatura seminal desta proximidade aponta para todas as relações econômicas como parte proveniente de contextos sociais. Tal pensamento leva comportamentos sociais a afetarem a economia (Polanyi, 1944). A maior socialização e interações em empresas resultam em melhor aprendizagem e, por consequência, a um desempenho inovativo. Entretanto, deve-se atentar, assim como as outras proximidades, ao alto nível desta proximidade, levando a redução do desempenho inovativo (Granovetter, 1985). Algumas situações de cooperação podem ser caracterizadas como oportunistas, visto que estão pouco pautadas pela busca de inovação, mas sim por casos de relações muito próximas que envolvem laços sentimentais, como as de amizade e parentesco. Assim como esta externalidade atribuída a alta proximidade social, o risco de *lock-in*¹ aumenta ao longo do tempo caso estes indivíduos não se atentem às novas abordagens.

Trabalhos anteriores utilizam as interações prévias para captar a proximidade social (Crescenzi *et al*, 2016; Crescenzi *et al*, 2017). Esses estudos indicam para a maior probabilidade de colaborações futuras entre indivíduos que já interagiram anteriormente. A pesquisa de Hong & Su (2013) também corrobora com esse resultado. As proximidades não-geográficas, utilizadas para inferir na probabilidade de interação Universidade-Empresa evidenciaram que as outras proximidades tendem a diminuir o efeito da distância geográfica cujo efeito continua significativo mesmo após este controle.

A proximidade social é descrita também por outras variáveis na literatura. A definição de pertencimento do grupo étnico dos pesquisadores é tema de debate dentro da proximidade social. Ligações cujos autores pertencem ao mesmo grupo étnico são mais propensas à transferência de conhecimento (Kerr, 2005; Agrawal *et al*, 2008). Grupos de etnia indiana nos EUA são estudados por deterem grande capital social o que facilita a troca de conhecimento em relações de confiança e com menores custos transacionais devido à proximidade social destes (Saxenian, 1999; Kalnins & Chung, 2006). Também há trabalhos que indicam a barreira linguística como um dificultador no processo de interação (Garfield, 1977; van Leeuwen *et al*, 2001; Liang *et al*, 2017).

¹ *Lock-in* é um processo de concentração de difusão de conhecimento localizado onde o nível de inovação se torna baixo por ser limitado a base de conhecimento local (Nooteboom, 2000).

2.2.4 Proximidade Institucional

Instituições denotam normas, leis, hábitos em comum e incentivos que regem um grupo de indivíduos. É uma proximidade importante devido a 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 misturar com a distância social, este último a 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).

Há uma linha tênue entre a proximidade institucional e a organizacional, tornando difícil sua identificação. Hong & Su (2013) separam ambas as proximidades com *proxy*'s no contexto chinês, a variável de proximidade institucional é construída a partir de dados da origem da administração local, a qual, a empresa ou universidade estão subordinadas, esta assumiu coeficiente positivo e significativo, indicando um papel importante da proximidade institucional na interação entre universidade e empresa.

2.2.5 Proximidade Organizacional

A proximidade organizacional refere-se a agentes que se encontram na 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, proximidade alta (Boschma, 2005).

Os mecanismos organizacionais gerenciam a coordenação de ações e transações entre entidades. Ademais são estruturas que integram seus agentes, possibilitando a transferência de conhecimento e de informação dentro de um ambiente repleto de incerteza (Cooke & Morgan, 1998). Considerada uma proximidade ampla pela literatura, muitos autores atribuem diferentes características a esta. Gilly & Torres (2000) atribuem significado de “mesmo espaço de relações”, não referente à espaço físico, podendo ser interpretado de várias formas, sempre incluindo similaridades as quais os atores são conectados pela mesma referência e pelo mesmo conhecimento. Já Kirat & Lung (1999, p. 30) denominam como um conjunto de interdependências cuja relação é financeira ou econômica. Tais relações não se limitam a nível intraorganizacional, mas também interorganizacional, caso essas dependências sejam similares entre os atores da rede.

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). Este mecanismo se torna importante na rede de colaboração, contribuindo na formação de *gatekeepers* em âmbito regional e nacional (Lagendijk & Lorentzen, 2006; Balland *et all*, 2015).

2.3 Estudos empíricos sobre as proximidades na colaboração científica

O Quadro 1 apresenta trabalhos empíricos relacionados às proximidades apresentadas acima. A maior parte dos estudos não abordam todas as proximidades conjuntamente como este trabalho faz. Há uma falta de consenso empírico na literatura mundial acerca dessas dimensões. Entretanto, seus resultados serão um guia para a estratégia empírica e obtenção dos resultados.

Quadro 1 – Literatura empírica de proximidades em rede

Autores	Região	Variável Dependente	Método estimação	Coefficientes / Resultados
Fernández <i>et al</i> , 2021	Espanha	Número de colaborações em artigos do par de pesquisadores	Modelo gravitacional	Cognitiva: Negativa significativa a 5%
				Geográfica: Positiva significativa
				Social: Positiva significativa
				Institucional: Positiva significativa
				Organizacional: Não significativo
Molina-Morales <i>et al</i> , 2015	Espanha	Matriz direcionada de ligações	Modelo exponencial aleatório gráfico (ERGM)	Cognitiva: Negativo significativo
				Geográfica: Negativo significativo
				Social: Positivo significativo a 10%
				Institucional: negativo significativo a 10%
				Organizacional: Positivo significativo
Crescenzi <i>et al</i> , 2017	Itália	Contagem de cooperações do par na criação de patentes.	Binomial Negativo Inflado em Zeros	Cognitiva: Incluída como controle
				Geográfica: Positivo significativo
				Social: Positivo significativo. Negativo quando considerado indivíduos dentro da mesma classe tecnológica
				Institucional: Positivo significativo. Negativo quando considerado indivíduos dentro da mesma classe tecnológica
				Organizacional: Positivo significativo
Crescenzi <i>et al</i> , 2017	Reino Unido	Contagem de cooperações do par na criação de patentes e <i>dummy</i> que indica colaboração	Modelo de eventos raros (Boschma <i>et al.</i> , 2013; King & Zeng, 1999) Binomial negativo	Cognitiva: Negativo significativo
				Geográfica: Positivo significativo
				Social: Positivo significativo
				Organizacional: Positivo significativo
Hong & Su, 2012	China Chinese Intellectual Property Press	Variável binária que acusa se houve colaboração entre universidade e empresa no desenvolvimento de patentes	Regressão logística baseado no modelo de eventos raros (King & Zeng, 1999)	As proximidades organizacional, institucional, social e geográfica se mostraram significativas. As proximidades institucional e organizacional atuam no sentido de encorajar novas laços entre os pesquisadores mesmo que longe, enquanto a proximidade social, construída a partir de interações prévias, está diretamente relacionada a maior probabilidade de futuras colaborações.

Fonte: Elaboração própria

Quadro 1 – Literatura empírica de proximidades em rede (Continuação)

Autores	Região	Variável Dependente	Método estimação	Coeficientes / Resultados
Nootboom <i>et al</i> , 2007	Dados de 116 companhias nos setores químicos, automotivos e farmacêuticos do Cooperative Agreements and Technology Indicators database, Maastricht Economic Research Institute on Innovation and Technology (MERIT-CAT).	Contagem do número de patentes concedidas na classe de patente que a firma foi ativa nos últimos 5 anos (Exploradora). Contagem do número de patentes concedidas na classe de patente que a firma não foi ativa nos últimos 5 anos (explorativo).	Modelo de Poisson e Binomial Negativo	Exploradora Cognitiva: Não significativo Explorativo Cognitiva: Positivo significativo (Cognitiva) ² : Negativo significativo ²
Kaygalak & Reid, 2016	Europa European Patent Office (EPO)	-	Teste de amostras independentes ANOVA	O estudo indica que existe diferença significativa na medida de rede de centralidade de proximidade entre os tipos organizacionais do setor privado e da universidade ³ .
Paci <i>et al</i> , 2014	Europa European Patent Office (EPO)	Patente per capita por região	Modelo Espacial Autorregressivo (SAR)	Os autores encontraram um efeito significativo nas proximidades: Social, Tecnológica (cognitiva), Geográfica e Organizacional. O objetivo central do artigo foi comparar o desempenho inovativo entre P&D e capital humano, os resultados considerando todas as proximidades indicam que capital humano agrega mais o desempenho inovativo da região.
Sorenson <i>et al</i> , 2006	Estados Unidos da América	<i>Dummy</i> de citação de patente do agente <i>i</i> para o <i>j</i>	Regressão logística baseado no modelo de eventos raros (King & Zeng, 1999)	O estudo corroborou com a literatura onde houve uma relação de U-invertido na variável de proximidade tecnológica (Cohen & Levinthal, 1990). Enquanto as variáveis de proximidade organizacional e geográfica foram positivamente significantes.

Fonte: Elaboração própria

² O resultado obtido nas duas variáveis de proximidade cognitiva e proximidade cognitiva ao quadrado refere-se a curva U-invertida. Isto é, proximidade cognitiva é benéfica, mas incorre o risco de *lock-in* caso esta os agentes estejam muito próximos.

³ O estudo infere sobre diferentes *clusters* de proximidade geográfica e organizacional. O primeiro foi criado a partir da divisão de regiões metropolitanas e periféricas. Já a proximidade organizacional refere-se ao tipo de organização que o indivíduo se insere, que são: universidades técnicas e não técnicas e setor privado

Quadro 1 – Literatura empírica de proximidades em rede (Continuação)

Autores	Região	Variável Dependente	Método estimação	Coefficientes / Resultados
Fernández <i>et al</i> , 2016	Europa	Número de coautorias entre o par de universidade <i>i e j</i> .	Binomial Negativo Inflado em Zeros	Todas as cinco proximidades foram significativas. Todavia, o coeficiente da proximidade organizacional influencia pouco na colaboração.
Heringa <i>et al</i> , 2016	Europa	<i>Dummy</i> indicando se houve colaboração entre o par de indivíduos <i>i e j</i> .	Regressão logística.	O estudo mostra que as proximidades geográfica, social e organizacional influenciam positivamente na probabilidade de colaboração, enquanto há um pequeno efeito de substituição entre a proximidade geográfica e organizacional.
Heringa <i>et al</i> , 2014	Holanda	6 regressões com diferentes variáveis dependentes: Inovação, publicações em conjunto, volume de negócios, elaboração de ideias, programas em colaboração e conhecimento compartilhado	Regressão logística e regressão logística multivariada.	As proximidades geográfica, social e cognitiva foram significativas e positivas nos ensaios apresentados, enquanto a proximidade organizacional afeta negativamente as variáveis dependentes.
Broekel & Mueller, 2018	Alemanha	Variável dicotômica que indica se a ligação entre as organizações <i>i e j</i> é crítica, ou seja, tem um papel de ponte entre as redes locais e globais.	Regressão logística baseado no modelo de eventos raros (King & Zeng, 1999)	As proximidades geográfica e social afetam positivamente as ligações críticas, enquanto a institucional tem efeito negativo. A proximidade organizacional não foi significativa. Ligações distantes institucionalmente são beneficiadas pelas proximidades geográfica e social.

Fonte: Elaboração própria

Os trabalhos utilizam as interações prévias a fim de captar a proximidade social, mostrando maior probabilidade de colaborações futuras entre indivíduos que já interagiram anteriormente (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 cuja resultante continua significativa mesmo após este controle.

Há divergências na influência e na magnitude das proximidades organizacional e institucional em relação à tomada de decisão do pesquisador em colaborar. Estas oscilam entre a influência negativa e a influência positiva a resultados não significativos. Um estudo recente que considera universidades mais novas da Espanha encontra que a proximidade organizacional não é significativa, no entanto uma possível explicação é de que esses centros procuram instituições mais tradicionais atrás de *expertise*, enquanto a proximidade institucional se mostrou positiva e significativa (Fernández *et al*, 2021).

Resultados divergentes a estes são encontrados na formação da rede de colaboração científica da China. As universidades que são vinculadas a ministérios específicos e províncias com tipos organizacionais diferentes apontam para significância das proximidades institucional e organizacional. Quando considerados dois pesquisadores do mesmo ministério, verificou-se uma menor importância em magnitude da distância geográfica (Hong & Su, 2013). Embora estes trabalhos concordem com a expressividade da dimensão organizacional, há indícios de uma menor contribuição desta dimensão, mas ainda significativa, e uma relação de substituição com a dimensão geográfica (Paci et al. 2014; Fernández et al, 2016; Heringa et al., 2016). Outro efeito analisado por Paci *et al* (2014), além da substitubilidade, é a complementariedade das proximidades social, organizacional, tecnológica e geográfica no transbordamento de conhecimento.

A colaboração entre pesquisadores muito próximos cognitivamente se torna prejudicial. É esperado que essa dimensão assumira um formato de U invertido na produtividade inventiva (Cohendet & Llerena, 1997). Devido a isso, os acadêmicos enfrentam um *trade-off* 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).

2. ESTRATÉGIA EMPÍRICA

O objetivo desse trabalho é esclarecer as relações individuais e interinstitucionais na atuação da cooperação no meio científico. Dito isto, as ligações dos pesquisadores, frequentemente, são resultadas de dinâmicas sociais, espaciais, culturais e institucionais. Para verificar tal hipótese o trabalho utiliza proximidades descritas na literatura de rede sociais descritas no Quadro 1 (Boschma, 2005; Crescenzi *et al*, 2016; Crescenzi *et al*, 2017; Molina-Morales *et al*, 2015).

3.1 Modelo e método

A equação 1 descreve o modelo e a estrutura a ser considerado nas regressões. 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 descrição desta variável a ser explicada está na Tabela 1. Outras contribuições também usam dados de contagem para cooperação de pares, as quais requerem modelos do tipo Binomial Negativo (CITAR). Este estudo adotará a mesma estratégia, já que se trata de uma estratégia bem aceita na literatura e 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. Tal fato é relevante, pois as ligações dos pares é a variável dependente e sua distribuição apresenta grande frequência de zeros, visto que a grande maioria da rede não coopera. Portanto, os dados indicam para um modelo Binomial Negativo Inflado em Zeros. Esse modelo predita um grupo na amostra que sempre assume zero. Em contraste as outras observações de pares de indivíduos que não são “sempre zero”, podem assumir zero no valor predito, contudo há uma probabilidade positiva destas observações assumirem valores não-zero. 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.

$$\begin{aligned} \text{Ligação do par}_{ijt} = & \beta_0 + \text{Prox.Social}_{ijt}\beta_1 + \text{Prox.Geográfica}_{ijt}\beta_2 + \\ & \text{Prox.Institucional}_{ijt}\beta_3 + \text{Prox.Organizacional}_{ijt}\beta_4 + \text{Prox.Cognitiva}_{ijt}\beta_5 + \end{aligned}$$

$$Formação_{ij}\beta_6 + Tempo\ docência_{it}\beta_7 + Tempo\ docência_{jt}\beta_8 + Betweenness_{it}\beta_9 + Betweenness_{jt}\beta_{10} + k_t + \varepsilon_{ijtf} \quad (1)$$

A variável dependente se trata de uma contagem de ligações dos pares i e 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 e j no triênio no mesmo artigo indica uma ligação. *Ligação do par* ij_t , sendo i o autor, j o coautor e t é a referência do triênio. Essa configuração da variável a ser explicada permite controlar características fixas ao longo dos triênios pela *dummy* k_t . O espaço de tempo de três anos como unidade de observação do painel junto com o par de pesquisadores contempla colaborações pontuais e recorrentes na rede. Esse formato da variável de resposta já é bastante utilizado na literatura de proximidades (Molina-Morales et al, 2015; Crescenzi et al, 2017; Crescenzi et al, 2016).

Tabela 1 – Estatísticas de dispersão em torno da média da variável dependente 2009-2017⁴

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

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Número de observações: 2.186.733.

Os períodos do painel foram divididos nos triênios: 2009 a 2011, 2012 a 2014 e 2015 a 2017. Houve um aumento de publicações registradas em todos os anos, assim como um crescimento da parcela de cooperação na produção de artigos.

A estimação BNIZ segue a formulação de Lawless (1987). Dado que este trabalho não tem intuito de agregar a tal metodologia, este não será tratado com muitos detalhes, reservando a literatura empírica citada. Seja Y a variável dependente de contagem de ligações, e x um vetor $p \times 1$ de variáveis explicativas. O modelo BNIZ é dado por:

$$P = (y = Y|x) = \frac{\Gamma(y + a^{-1})}{y! \Gamma(a^{-1})} \left(\frac{a\mu(x)}{1 + a\mu(x)} \right)^y \left(\frac{1}{1 + a\mu(x)} \right)^{a-1}$$

Sendo Γ a função distribuição binomial negativa, a e μ são parâmetros. Este modelo é estimado pela máxima verossimilhança.

⁴ Baseado na média e desvio-padrão de coautorias entre os dois pesquisadores no triênio.

3.2 Variáveis explicativas

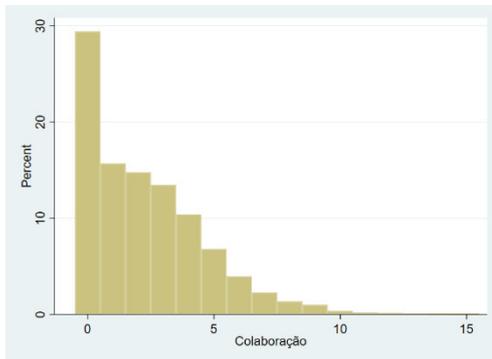
Proximidade social

Esta variável será construída a partir de interações prévias 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 somente em solo. Os Gráficos de 1 a 6 mostram, em porcentagem, o número de cooperações por artigo entre 2000 a 2005, 2006 a 2011 e 2012 a 2017, respectivamente. 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% dos trabalhos é de autoria única e a maior parte das interações ocorreu somente uma vez (67%). O aumento de mais de 25%⁵ 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.

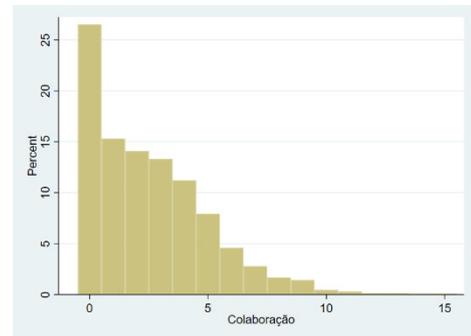
⁵ Este é o valor verdadeiro, criado a partir de das referências dos artigos publicados, ou seja, o autor acusou que este artigo é de autoria única. Este valor diverge com o valor da rede criada, já que para montagem da rede foi necessário a correspondência das coautorias pelo título da publicação, o que nem sempre é exatamente igual em todos, mesmo depois de todas as correções manuais. A Tabela 7 mostra tal divergência, que nos triênios finais, aqueles que serão utilizados na regressão, convergem para o valor verdadeiro citado.

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



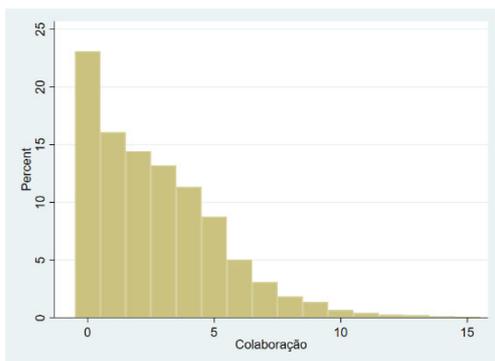
Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

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



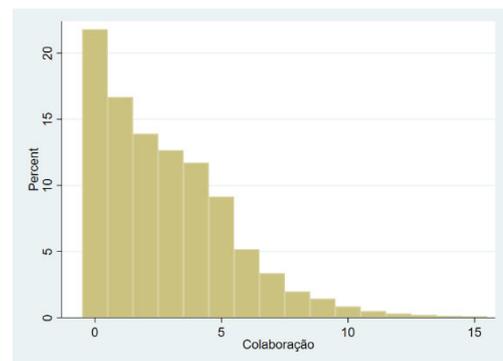
Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

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



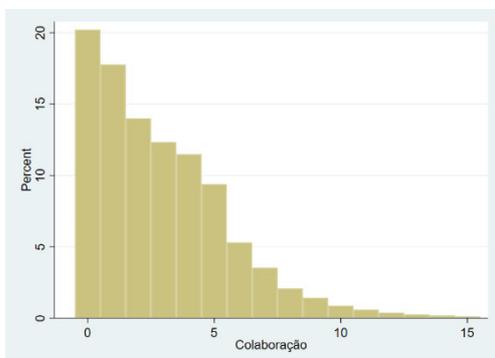
Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

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



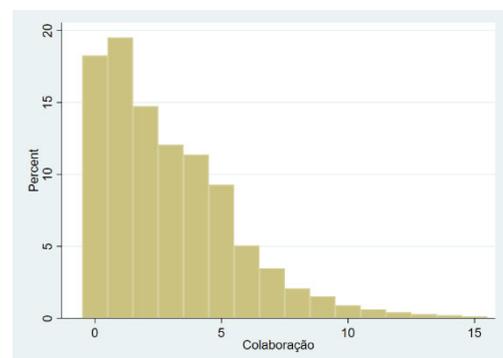
Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

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



Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

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



Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Nota: A primeira coluna referente ao 0, indica o número de artigo publicados em autoria única

A Tabela 2 descreve a variável explicativa de proximidade social de coautorias segregada por triênios e por defasagem temporal. A construção desta foi realizada por meio de 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, isso condiz com a literatura, já que, com o tempo, a fronteira do conhecimento se torna mais complexa e a demanda por mais cooperações e interações repetidas são mais frequentes visto que há menos incerteza nessa parceria (Danell, 2014).

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 a esta variável ao incluí-la no modelo, pois há risco de colinearidade dessas variáveis.

Esse raciocínio não é empregado no copatenteamento⁶, já que são poucas observações, para este foi criado somente uma variável binária que acusa se o par de autores já patentearam em conjunto anteriormente.

Tabela 2 – Descrição e distribuição da variável social de ligações prévias no tempo 2009-2017⁷

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: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Número de observações: 2.186.733.

Proximidade institucional

A variável de proximidade institucional será construída a partir de dados informados pelos pesquisadores do currículo Lattes. Dentre as informações relacionadas estão as

⁶ Foram observadas 19.656 interações prévias em copatenteamento na rede com média de 0,001 e 0,030 de desvio-padrão

⁷ A tabela mostra a distribuição da ocorrência de coautoria prévia até o terceiro triênio.

instituições as quais eles 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, esta é acionada quando o par já teve vínculo na mesma instituição.

Proximidade organizacional

Indivíduos que registram passagem por empresas ou universidade vivenciam tipos organizacionais diferentes, como normas, regras e incentivos divergentes, que podem levar a um conflito de interesses e objetivos. Embora este trabalho não tenha a intenção de medir o nível inovativo desta produção acadêmica, 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 universidade e se estes grupos se diferem no posicionamento da rede já fora feita por Kaygalak & Reid (2016).

Proximidade Cognitiva

A variável cognitiva indica a qual área do conhecimento ambos os pesquisadores pertencem. Como visto na Tabela 5, 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.

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 (UF) 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.

As variáveis de proximidades e de controle são descritas no Quadro 1 e no Quadro 2, exceto a proximidade social que já fora descrita nesta seção.

Quadro 2 – Descrição das variáveis independentes

Proximidade	Descrição da variável	Variáveis
Proximidade Institucional	As variáveis criadas a fim de indicar se há algum tipo de relação institucional entre os agentes. Para isso foram criadas variáveis binárias.	<i>Dummy</i> de mesma instituição
		<i>Dummy</i> institutos de pesquisa
Proximidade Organizacional	Indica se os atores são do mesmo tipo organizacional, se são de universidade pública e empresa. Duas <i>Dummies</i> foram criadas para indicar a proximidade organizacional, assume 1 se ambos do par de autores são do setor produtivo ou acadêmicos. Além disso foi criada outra <i>dummy</i> auxiliar a esta que aciona se ambos têm bolsa.	<i>Dummy</i> universidade
		<i>Dummy</i> empresa
		<i>Dummy</i> universidade pública
Proximidade Geográfica	<i>Dummies</i> de correspondência de mesma unidade da federação.	<i>Dummy</i> de mesma Unidade da Federação

Fonte: Elaboração própria

Quadro 2 – Descrição das variáveis independentes (Continuação)

Proximidade	Descrição da variável	Variáveis
Proximidade Cognitiva	<i>Dummies</i> de correspondência exata da área do conhecimento que o pesquisador pertence. Foram incluídas 9 <i>dummies</i> cada uma é acionada caso ambos os autores sejam da respectiva grande área do conhecimento.	Ciências exatas e da terra
		Ciências Biológicas
		Engenharias
		Ciências da Saúde
		Ciências Agrárias
		Ciências Sociais Aplicadas
		Ciências Humanas
		Linguísticas, letras e artes
		Outros
Variáveis controle		
Tempo de docência	Tempo de docência tanto do pesquisador i quanto do pesquisador j	Logaritmo do tempo de docência i
		Logaritmo do tempo de docência j
Nível de formação	<i>Dummies</i> indicando se ambos são ou doutores	<i>Dummy</i> indicando se ambos são doutores
<i>Betweenness</i>	<i>Betweenness</i> tanto do pesquisador i quanto do pesquisador j	Logaritmo do <i>Betweenness</i> de i
		Logaritmo do <i>Betweenness</i> de j

Fonte: Elaboração própria

3.3 Métricas de rede

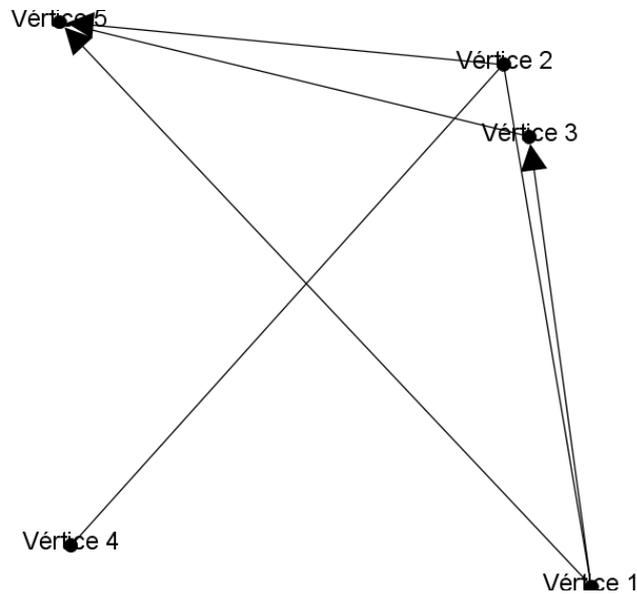
As métricas de rede são utilizadas para analisar a estrutura como um todo, somente o nó ou os componentes da rede de interesse. Estas medidas auxiliam a corroborar ou formar hipóteses, a apontar a importância de agentes decisivos e a comparar diferentes redes no tempo ou em regiões distintas. Nesta seção será explicado como são calculadas, interpretadas e posteriormente usadas em uma regressão.

Os nós em uma rede são agentes que podem ou não estar conectados a outros nós. Estes possuem características individuais que influenciam seu comportamento na rede. Nas relações de amizade, pessoas que são mais próximas a outras podem ser representadas pelo número de contatos pessoais os quais podem estar correlacionados com atributos, gostos e preferências. Uma representação dessa rede de amizade ilustra a correlação dessas proximidades de ligação com características individuais.

O grafo é a representação gráfica da rede e seus componentes. Um grafo $G(V,E)$ é composto por V vértices (nós ou pontos) e E laços (ligações ou conexões). Este último conecta um par de nós. Caso o nó i e o nó j estejam conectados por um vértice no grafo G , isso é expressado como $(i,j) \in E(G)$. Nós conectados por um laço são chamados de adjacentes. Os tipos de relação que definem os vértices formados estabelecem se a rede é direcionada ou não direcionada. Redes direcionadas assumem que o vértice do par de nó i e j pode estar ligado de 3 formas: $i > j$, onde somente o nó i se liga ao nó j ; $j > i$ somente o nó j está ligado ao nó i ; e por último $i > j > i$ onde ambos estão ligados (Borgatti *et al*, 2013).

A Figura 1 ilustra de forma didática como os nós são representados em grafos. É uma rede direcionada formada por 5 vértices, $V = (1, 2, 3, 4 \text{ e } 5)$ e 9 laços, $E = \{(1,2), (2,1), (1,3), (1,5), (2,4), (4,2), (3,5), (5,3) \text{ e } (2,5)\}$.

Figura 1 – Rede ilustrativa com nós e vértices fictícios



Fonte: Elaboração própria

De acordo com suas ligações e posicionamento de seus vizinhos, são criadas métricas de análise individual como: grau de ligação, medida de entrelaçamento, medida de proximidade e coeficiente de cluster (BORGATTI *et al.*, 2013).

2.1.1. Centralidade de entrelaçamento

A medida de centralidade de entrelaçamento (*betweenness centrality*) aponta, segundo o nó analisado, a contagem de caminhos geodésicos que este agente participa em toda a rede. Agentes bem posicionados possuem um grau de centralidade maior e, por isso, possuem maior fluxo de conhecimento.

$$C_B(v_k) = \sum_{1 \leq j \leq n \wedge i \neq k \neq j} \frac{g_{ij} v_k}{g_{ij}}$$

Na representação acima temos o grau de centralidade C_B do nó v_k que é determinado pelos caminhos geodésicos de todos os pares da rede que dependem deste nó. O termo g_{ij} mostra todos os caminhos geodésicos entre todos os nós, quando acrescentado o subscrito v_k ao termo g_{ij} estes caminhos só serão contabilizados caso passem pelo nó analisado v_k . Quando não há caminho entre um dos indivíduos do par o termo assume o valor de 0 no somatório.

2.1.2. Centralidade de grau

O grau de ligação é uma simples e poderosa medida de contagem de vértices que o nó tem. Seja k o nó analisado:

$$d_k = \sum_{j=1}^n a_{kj}$$

O somatório de todos os nós da rede que se ligam com k , a_{kj} , calcula o grau médio desse agente (d_k).

2.1.3. Centralidade de grau ponderada

Mede a soma de todas as ligações que o possui. Esta, diferentemente da centralidade de grau, considera os pesos que os vértices têm. Logo essa medida é sempre maior que a anterior.

$$wd_i = \sum_{j=1}^n w_{ij}$$

A Centralidade de grau ponderada, wd_i , é medida pelo somatório de todas as ligações do nó j : $\sum_{j=1}^n w_{ij}$.

2.1.4. Medida de proximidade

A medida de proximidade (*closeness*) é calculada pelo inverso da distância geodésica dos nós da rede. Sendo assim:

$$C_C(v_k) = \frac{1}{\sum_{j=1}^n dist(v_j, v_k)}$$

A centralidade de proximidade é representada por $C_C(v_k)$, que é o inverso da distância dos nós v_j e v_k .

2.1.5. Transitividade

A transitividade capta relações tríades na rede, como exemplo: “amigo do meu amigo também é meu amigo”. Se existem dois ligados a um indivíduo, vértices a_{ij} e a_{jk} , é esperado que haja uma maior probabilidade que a ligação a_{ik} exista, já que eles têm uma ligação de segundo grau. Esta medida pode assumir valores discretos de acordo com quantos possíveis tríades esses dois agentes estão inseridos.

2.1.6. Coeficiente médio de Cluster

Esta medida está relacionada também a ligações triádes. Ela se resume em quantos triádes fechados em relação a todos os possíveis triádes, tanto fechados quando abertos. Os triádes abertos são aqueles que não formam a terceira ligação em contraponto com o fechado que possui as 3 ligações.

$$C_{coef}C_{luster} = \frac{\text{Número de triádes fechados}}{\text{Número total de triádes possíveis (abertos e fechados)}}$$

A Tabela 3 descreve as medidas de rede no grafo da Figura 1.

Tabela 3 - Valores das métricas de rede da Figura 1

Tipo de medida	Valor
Centralidade de Grau	2,400
Centralidade de Grau Ponderada	2,400
Medida de Proximidade	0,600
Coeficiente Médio de Cluster	0,667

Fonte: Elaboração própria

3.4 Descrição da base de dados

A base de dados utilizada é a da Plataforma Lattes, com o período de análise de 2000 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, 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 na plataforma

O endereço informado pelo acadêmico da plataforma se refere ao endereço profissional atual. São 5.009.368 endereços de todos os indivíduos, porém cerca de 68% dos endereços estão incompletos. Esse número diminui quando é considerado somente os indivíduos que patenteiam e/ou publicam, caindo para 40% de *missings*. Para atenuar a presença destes, foram usados

descrições e nomes dos vínculos mais recentes de maior duração. Tal artifício subdimensionou os *missings* para 22%.

Os vínculos institucionais da Plataforma Lattes são informados pelos indivíduos que, em sua maioria, são titulações de mestrado e doutorado, estágios, iniciações científicas, docências, revisões de periódico, trabalhos, entre outros. Além da descrição do tipo de vínculo institucional, o indivíduo informa a data de início e término. Estes foram classificados como vínculos em instituições acadêmicas e do setor privado de acordo com o tipo de vínculo declarado e também pelos nomes das instituições. Entre os indivíduos cadastrados 86% tiveram vínculo com universidades e outros 43% tiveram algum vínculo com empresas⁸.

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 segundo. 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 a centralidade *betweenness* que, em média, no Sudeste é maior que nas outras regiões, seguida por Sul, Nordeste e Centro-Oeste respectivamente. A Tabela 4 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 4 – Descrição das medidas de rede das regiões brasileiras

Região	Participação da base	Média centralidade de <i>Betweenness</i>	Grau médio ponderado
Norte	4,2%	14.412,79	13,04
Nordeste	15,57%	19.749,57	15,78
Sudeste	35,06%	28.317,34	24,05
Sul	16,56%	24.225,10	20,06
Centro-Oeste	6,30%	17.245,38	14,49

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. 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

⁸ Caracteriza-se vínculo com empresa indivíduos que declararam no Plataforma Lattes que tiveram um vínculo com instituições que possuem nomes que denotam empresa.

médio, ponderado foram calculados a partir da média dos indivíduos de cada região brasileira que produziram artigos em cooperação.

A Tabela 5 detalha as estatísticas de rede no período entre 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. São 9 grandes áreas, a saber, ciências exatas e da terra, ciências biológicas, engenharias, ciências da saúde, ciências agrárias, ciências sociais aplicadas, ciências humanas, linguística, letras e artes e outros. Somente as áreas da saúde, humanas e sociais aplicadas representam mais da metade da base. Sendo as mais colaborativas respectivamente: 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.

Tabela 5 – 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%	0,7	1,01
Outros	0,74%	0,33	5,15

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. 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 cooperação.

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. Neste, assim como esse estudo, as áreas com mais pesquisadores são ciências da Saúde, ciências humanas e ciências sociais. As áreas de ciências biológicas e agrárias são as mais colaborativas, tal comportamento também confere a literatura citada.

A titulação de cada pesquisador é descrita na Tabela 6. Devido ao filtro aplicado, há uma proporção de mestres e graduados que produziram algum artigo em sua vida acadêmica e

que se encaixaram na descrição. Essas observações não foram excluídas a fim de deixar as observações mais heterogêneas e, também, há possibilidade de, em algumas observações, os pesquisadores não atualizaram seu currículo junto a plataforma. Tanto nas medidas de rede quanto na média de artigos destacam-se os doutores e professores, conforme o esperado.

Tabela 6 – Estatísticas descritivas e de rede em relação à titulação dos pesquisadores 2000-2017

Titulações	% na base	Média de Artigos	Média de artigos em colaboração	<i>Betweenness</i>	Grau Ponderado
Mestres	35,54%	5,68	3,07	3.006,16	6,19
Doutores	44,57%	18,40	10,66	44.419,76	34,17
Professor	86,31%	11,89	6,22	23.158,29	18,70
Graduação	7,56%	3,34	2,14	1.126,89	4,35
Mestres profissionais	1,73%	5,15	2,47	4.215,17	4,98
Residência Médica	1,09%	8,47	4,82	7.398,87	10,00

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq . Nota: Medidas de rede tomam como base somente os indivíduos que produziram artigos em cooperação

O Gráfico 7 mostra a média de autores em um artigo no tempo, o aumento do número de autores do mesmo artigo é nítido. A Tabela 7 descreve a evolução das métricas de rede no tempo. Há um aumento no coeficiente de *Cluster* que indica maior agrupamento e uma rede mais densa e também a presença de ligações mais fortes indicado pela centralidade de grau ponderada.

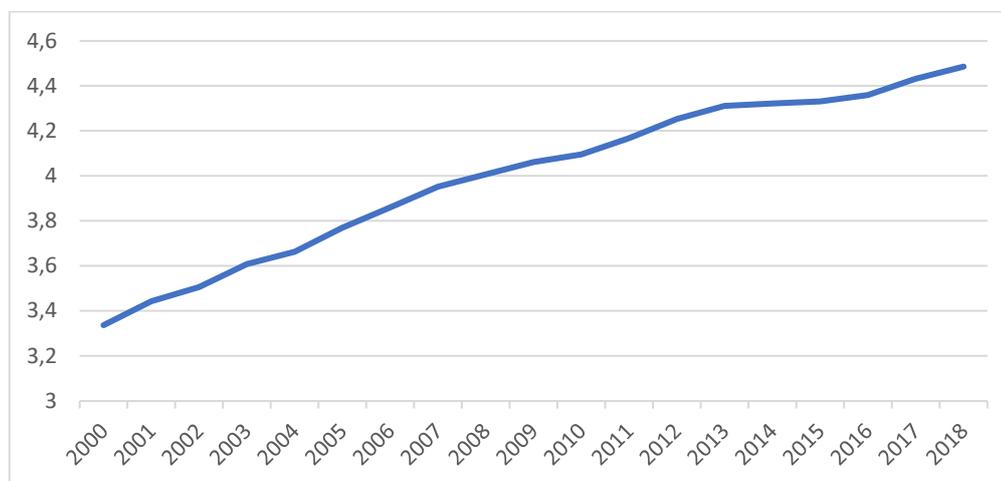
Tabela 7 – Estatísticas descritivas da rede de colaboração dos triênios, Brasil, 2000-2017

Triênio	2000- 2002	2003- 2005	2006- 2008	2009- 2011	2012- 2014	2015- 2017
Número de autores	101.537	136.688	169.93	213.932	254.193	272.264
Grau médio ¹	3,12	3,86	4,77	6,08	5,18	7,77
Grau médio ¹ ponderado	4,44	5,487	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 ²	70,22	63,35	54,06	46,07	36,15	29,26
Coefficiente de cluster médio	0,60	0,62	0,645	0,658	0,631	0,657

FFonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Notas: ¹ Medidas de rede tomam como base somente os indivíduos que produziram artigos em cooperação. ² Esse valor reflete os nós isolados identificados após a construção da rede, o que não retrata a verdadeira participação de pesquisadores solo.

Os dados mostram que a razão de autores solitários diminui em todos os anos enquanto 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, caiu para 18,96%. No mesmo período as participações em coautorias de mais de 3 membros aumentaram em mais de 10%.

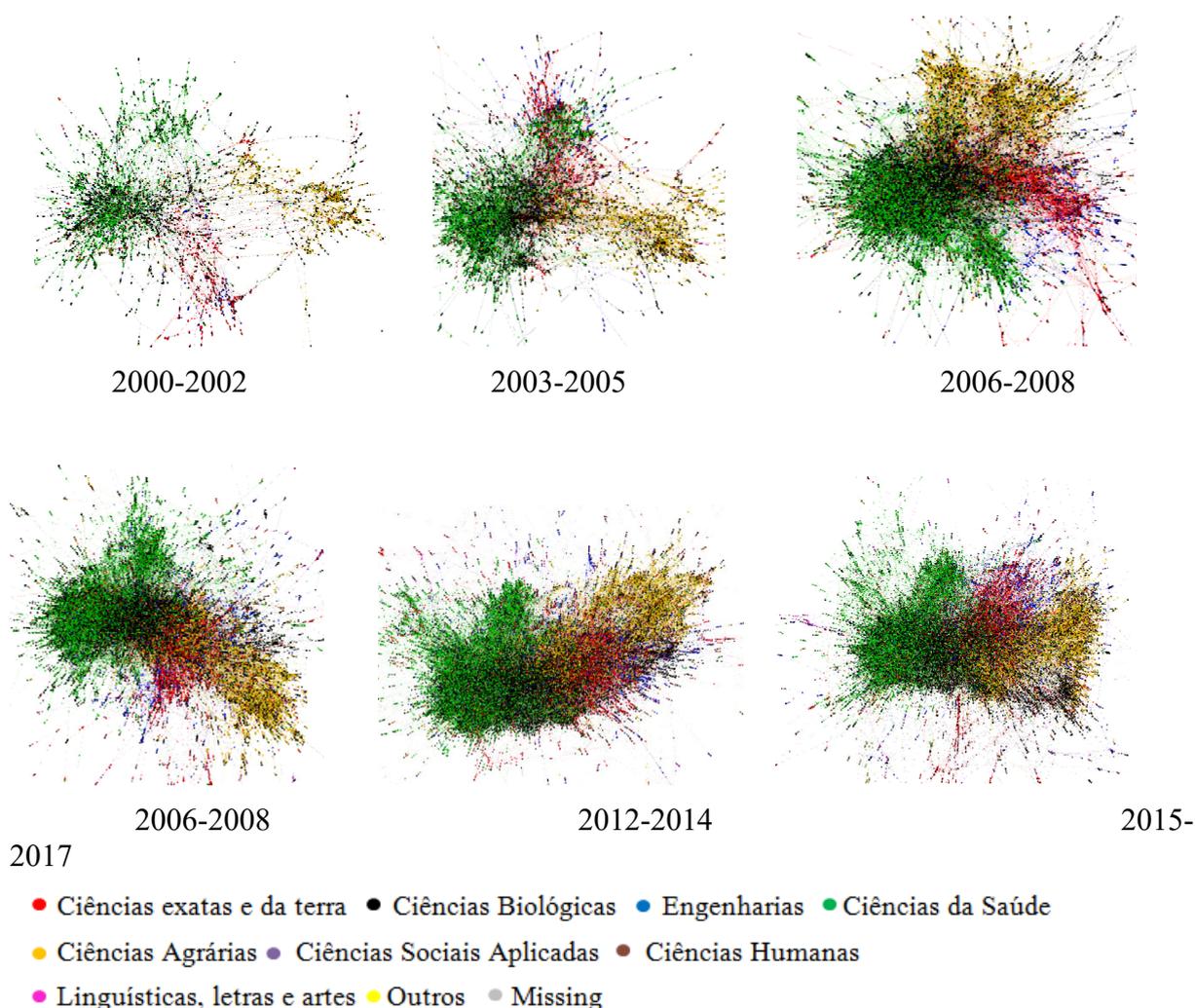
Gráfico 7 – Média do número de autores por artigo ao longo do tempo 2000-2018



Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

A Figura 2 apresenta as figuras de rede para os 1% pesquisadores mais produtivos e suas respectivas áreas do conhecimento. Estas figuras seguem a distribuição Atlas Force 2 que possui algoritmo baseado nas características da rede. Tal distribuição faz com que os nós repilam e os vértices se atraiam. Essas interações fazem com que *clusters* das áreas de conhecimento se formem indicando que as grandes áreas do conhecimento tendem a se agrupar na rede e que a sua periferia se cerque de outras áreas do conhecimento com as quais têm mais vínculos e afinidade (Jacomy *et al*, 2014). A rede cresce ao longo do tempo seguindo a tendência de aumento do número de pesquisadores e maior conexão entre eles. Dentre as áreas mais participadas estão as ciências exatas e da terra, ciências biológicas e ciências agrárias. Estes dois últimos apresentam um grande salto em suas proporções na rede assim como indica Mena-Chalco *et al* (2014).

Figura 2 - Rede de colaborações dos pesquisadores 1% mais produtivos por diferentes áreas de conhecimento.



Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. As distribuições seguem o padrão Force Atlas 2 (Jacomy *et al*, 2014)

Tabela 8 – Correlação entre as variáveis do modelo

	111	2	3	4	5	6	7	8	9	10	11	12						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(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) MsmCNPQ	0.10	0.11	0.10	0.09	0.07	0.02	0.07	0.06	0.04	1.00								
(11) MsmMest	-0.04	-0.04	-0.03	-0.02	-0.03	0.01	-0.05	-0.04	-0.04	-0.03	1.00							
(12) AmbosDout	0.17	0.18	0.13	0.09	0.06	0.09	0.13	0.15	0.16	0.11	-0.20	1.00						
(13) LigaPat	0.08	0.07	0.06	0.05	0.01	0.00	0.02	0.04	0.01	0.03	-0.01	0.03	1.00					
(14) MsmInst	0.10	0.11	0.07	0.05	0.00	0.01	0.00	0.10	0.04	0.01	-0.01	0.03	0.02	1.00				
(15) Between	0.14	0.14	0.11	0.08	0.04	0.02	0.08	0.10	0.05	0.12	-0.05	0.14	0.03	0.02	1.00			
(16) Cobetween	0.13	0.13	0.10	0.07	0.03	0.02	0.06	0.09	0.04	0.09	-0.04	0.12	0.02	0.01	0.07	1.00		
(17) Tempdocen	0.04	0.05	0.05	0.05	0.02	0.06	0.04	0.05	0.01	0.06	-0.08	0.19	0.01	0.00	0.15	0.01	1.00	
(18) Cotempdoc	0.03	0.04	0.04	0.04	0.01	0.05	0.01	0.04	0.00	0.05	-0.08	0.19	0.01	0.00	0.01	0.10	0.01	1.00

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Nota: Número de observações: 2.186.733.

3.5 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. O dispêndio de tempo para tais interações seria incompatível com a realidade do pesquisador

Dada tal situação, é abordado pela literatura a criação de redes fictícias, nas quais as ligações podem ser aleatórias ou definidas por características similares dos indivíduos (Chandrasekhar, 2015). Para comparação da base de dados, em que há ligações entre os indivíduos, é preciso criar um contrafactual onde $g_{ij} = 0$. Em parte, tais modelos criam uma rede aleatória em que há uma probabilidade de ligação distribuída entre os indivíduos.

Neste trabalho será adotado o método para criação do contrafactual de Ęrdos-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. Embora não seja o mais apropriado, visto que considera todas os autores com probabilidade igual de ligação, ele nos fornece um ganho de tempo no processamento, já que outros métodos que consideram características individuais demandam mais gasto de processamento. A Tabela 9 mostra a divisão da base entre o contrafactual, onde $g_{ij} = 0$ e as ligações verdadeiras, $g_{ij} \geq 1$.

Tabela 9 – Divisão do contrafactual nos triênios

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. Nota: Número de observações: 2.186.733.

A escolha do método descrito por Chandrasekhar (2015) nos leva às seguintes afirmações:

- (1) O total de ligações esperadas para um dado nó é de $(n - 1)p$;
- (2) A probabilidade que dois vizinhos de i, j e k sejam ligados é p ;

(3) A probabilidade de que i, j e k sejam mutuamente ligados é p^3 .

As observações cujos valores da dependente são nulos, ou seja, o contrafactual, têm estatísticas das variáveis explicativas diferentes do grupo não-nulo. Esse comportamento é esperado já que a dupla de pesquisadores que não produzem deve conter valores distintos de suas variáveis explicativas. As estatísticas dos dois grupos estão descritas nas Tabela 10 e Tabela 11. A primeira tabela traz valores da base por completa, isto é, quando $g_{ij} = 0$ e as ligações verdadeiras, $g_{ij} \geq 1$. Já a segunda tabela descreve as estatísticas dos pares de autores que realmente colaboraram $g_{ij} \geq 1$. A divergência desses valores é importante para identificação dos valores preditos para o contrafactual na primeira etapa do binomial negativo inflado em zeros cujo método é a regressão logística, identificando o grupo de pares de autores que sempre apresenta valor nulo na dependente.

Tabela 10 – Estatísticas descritivas das variáveis de interesse do total de observações

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
Betweenessi	4,613	4,766	-6,557	17,303
Betweenessj	3,707	4,707	-6,497	17,303
TempDoceni	3,324	1,887	0	6,825
TempDocenj	3,672	1,746	0	6,825
AmbosDoutor	0,237	0,425	0	1

Fonte: Elaborado a partir de dados da Plataforma Lattes. Nota: Número de observações: 2.186.733.

Tabela 11 – Estatísticas descritivas das variáveis de interesse das observações em que houve ligação entre os pesquisadores

Variável	Média	Desvio-Padrão	Mínimo	Máximo
ContArtigo	1,609	2,629	1	603
IntPrev1	0,205	0,404	0	1
IntPrev2	0,085	0,279	0	1
IntPrev3	0,036	0,186	0	1
IntPat	0,166	0,257	0	1
MesmaUF	0,476	0,499	0	1
MesmaTipoInst	0,129	0,649	0	1
MesmaInst	0,082	0,274	0	1
MesmaUni	0,980	0,138	0	1
MesmaEmp	0,527	0,499	0	1
MesmaUnivPubl	0,669	0,470	0	1
MsmGrandArea	0,391	0,488	0	1
Betweenness _i	9,788	3,491	-6,497	17,303
Betweenness _j	9,819	3,481	-6,497	17,303
TempDoceni	3,732	1,837	0	6,730
TempDocenj	3,779	1,820	0	6,730
MesmoDoutor	0,650	0,477	0	1

Fonte: Elaborado a partir de dados da Plataforma Lattes. Nota: Número de observações: 182.564.

3. RESULTADOS

3.1. Resultados da estimação de todas as proximidades

Os resultados da regressão BNIZ em painel do período de 2009-2011 são apresentados na Tabela 12. 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 aleatória 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 da proximidade cognitiva foram omitidas, as variáveis foram agrupadas em partes: proximidade social, proximidade geográfica, proximidade institucional, proximidade organizacional, e variáveis de controle.

Tabela 12 – Condicionantes do número de publicações por pares de pesquisadores no Brasil (2009-2017). Estimação por Modelo Binomial Negativo Inflado.

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>							
Dummy 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>							
Dummy Tipo Institucional	-	-	0,02*** (0,01)	0,01** (0,01)	0,01 (0,01)	0,01 (0,01)	0,02*** (0,01)
Dummy Instituição	-	-	0,54*** (0,01)	0,41*** (0,01)	0,41*** (0,01)	0,41*** (0,01)	1,26*** (0,01)

Nota: Erros-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 incluídas 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 12 – Condicionantes 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>Betweenness 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>Betweenness 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 ¹	Sim						
Número de observações	2.182.666	2.182.666	2.182.666	2.182.666	2.182.666	2.182.666	2.182.666

Nota: Erros-padrão entre parênteses. Nível de significância: * 5%, ** 1%, ***0,1%. Variável Dependente: Contagem do número de artigos publicados pelo par de pesquisadores. ¹As áreas do conhecimento incluídas 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.

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 colaboração (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 aos anteriores, no entanto a 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. Enquanto os estudos de Crescenzi *et al*. (2017) e Broekel & Mueller (2018), estão em consonância o trabalho de Molina-Morales *et al* (2015), após controlar todas as proximidades, aponta um resultado negativo e pouco significativo, sugerindo 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 Lattes. 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 os estudos mostram que a produtividade dos pesquisadores diminui com a idade (Bonaccorsi & Daraio, 2003).

A Tabela 13 apresenta os resultados de uma regressão que possui a mesma especificação da Tabela 12. No entanto, as colunas de resultados diferem por se basearem em subamostras. A coluna (1) e (2) são, respectivamente, pares de pesquisadores que são da mesma Unidade da Federação e pares de pesquisadores de Unidades da Federação diferentes. Já as duas últimas colunas de resultado são referentes à grande área do conhecimento, (3) par de pesquisadores na mesma área do conhecimento e (4) que divergem na sua formação. Esses resultados amparam a robustez dos coeficientes, além de trazerem variações destes que podem ser interpretadas como características daquela subamostra.

Tabela 13 – Condicionantes do número de publicações por pares de pesquisadores no Brasil em subamostras de acordo com Unidade da Federação e grande área do conhecimento (2015-2017). Estimação por Modelo Binomial Negativo Inflado

Variáveis	(1)	(2)	(3)	(4)
<i>Proximidade Social</i>				
Interação Prévia t-1	0,74*** (0,01)	1,43*** (0,16)	0,86*** (0,01)	1,45*** (0,02)
Interação Prévia t-2	0,22*** (0,02)	0,28*** (0,02)	0,20*** (0,02)	0,32*** (0,03)
Interação Prévia t-3	0,16*** (0,02)	0,06 (0,03)	0,11*** (0,02)	0,08* (0,04)
Interação Prévia Patente	0,38*** (0,04)	0,72*** (0,07)	0,28*** (0,04)	0,64*** (0,06)
<i>Proximidade Geográfica</i>				
Mesma Unidade Federativa	-	-	0,62*** (0,01)	1,15*** (0,01)
<i>Proximidade Institucional</i>				
Mesmo Tipo Institucional	-0,06*** (0,01)	0,04*** (0,01)	-0,05*** (0,01)	0,05*** (0,01)
Mesma Instituição	0,43*** (0,02)	1,65*** (0,03)	0,58*** (0,02)	1,64*** (0,03)
<i>Proximidade Organizacional</i>				
Mesma Universidade	0,24*** (0,04)	0,07* (0,11)	0,14*** (0,03)	0,10** (0,04)
Mesma Empresa	0,06*** (0,01)	0,11*** (0,01)	-0,01 (0,01)	0,15*** (0,01)
Mesma Universidade Pública	0,08*** (0,01)	0,23*** (0,01)	0,14*** (0,01)	0,29*** (0,01)

Nota: Erros-padrão entre parênteses. Nível de significância: * 5%, ** 1%, ***0,1%. Variável Dependente: Contagem do número de artigos publicados pelo par de pesquisadores. Coluna (1) se refere ao subgrupo de pares de pesquisadores que estão na mesma Unidade da Federação e a coluna (2) a pares com UF's diferentes. A coluna 3 e 4 são subgrupos em relação à área do conhecimento, sendo o par pertencente à mesma área e a diferentes áreas, respectivamente. ¹As 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 13 – Condicionantes do número de publicações por pares de pesquisadores no Brasil em subamostras de acordo com Unidade da Federação e grande área do conhecimento (2015-2017). Estimação por Modelo Binomial Negativo Inflado (Continuação)

Variáveis	(1)	(2)	(3)	(4)
<i>Controle</i>				
Betweeness <i>i</i>	0,11*** (0,00)	0,21*** (0,00)	0,14*** (0,00)	0,21*** (0,00)
Betweeness <i>j</i>	0,13*** (0,00)	0,24*** (0,00)	0,13*** (0,00)	0,25*** (0,00)
Tempo Docência <i>i</i>	-0,05*** (0,00)	-0,08*** (0,00)	-0,06*** (0,00)	-0,07*** (0,00)
Tempo Docência <i>j</i>	-0,07*** (0,00)	-0,12*** (0,00)	-0,05*** (0,00)	-0,14*** (0,00)
Ambos Doutores	-0,15*** (0,01)	-0,20*** (0,01)	-0,11*** (0,01)	-0,24*** (0,01)
Constante	-2,69*** (0,13)	-5,64*** (0,04)	-3,75*** (0,06)	-5,76*** (0,04)
<i>Dummies</i> de áreas do conhecimento ¹	Sim	Sim	-	Sim
Número de observações	293.086	1.893.647	342.684	1.844.049

Nota: Erros-padrão entre parênteses. Nível de significância: * 5%, ** 1%, ***0,1%. Variável Dependente: Contagem do número de artigos publicados pelo par de pesquisadores. Coluna (1) se refere ao subgrupo de pares de pesquisadores que estão na mesma Unidade da Federação e a coluna (2) a diferente a pares com UF's diferentes. A coluna 3 e 4 são subgrupos em relação a área do conhecimento, sendo o par pertencente a mesma área e a diferente área respectivamente. ¹As áreas do conhecimento incluídas 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.

Os resultados dessa nova regressão apontam para uma similaridade com o modelo anterior. No entanto, a variável binária do Mesmo Tipo Institucional apresenta um coeficiente negativo, embora pequeno e significativo. Os resultados dos coeficientes das variáveis de Proximidade Social nas colunas (2) e (4) sugerem um coeficiente de maior magnitude em relação às suas subamostras complementares cujas colunas são (1) e (3). Isso indica que pesquisadores de Unidades da Federação diferentes e que não são da mesma grande área do conhecimento estão mais propensos a colaborar caso já tenham produzido pelo menos um artigo em colaboração. Esses coeficientes também têm magnitudes ligeiramente maior em relação à regressão principal. O mesmo comportamento é identificado para a variável de *betweenness*, em que as colunas (2) e (4) possuem um coeficiente maior em relação aos seus complementos (1) e (3). Os pares da interação entre pesquisadores de outras UF's e diferentes áreas do conhecimento valorizam mais a centralidades de entrelaçamento.

4. CONCLUSÃO

Este trabalho analisou, de forma pioneira, nas 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, podemos 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 em termos de colaboração científica.

As métricas de rede foram fundamentais para analisar diferenças entre grupos de pesquisadores, como a área do conhecimento, titulação acadêmica e região. A medida de centralidade de entrelaçamento influencia positivamente a tendência de colaboração. Já o tempo de docência, que é a soma do tempo em que o indivíduo atuou lecionando em todos os seus vínculos, impacta negativamente a decisão de formar um laço com outro pesquisador. Tal resultado pode ser explicado pelo fato de que pesquisadores, em média, tendem a ser mais produtivos no início de sua carreira, envolvendo-se em encargos de promoção científica diferentes da publicação em si e mesmo em tarefas administrativas ao final da sua carreira.

As diferenças do número de colaborações entre grupos da mesma grande área do conhecimento e a proximidade cognitiva, são fundamentais para o entendimento de como as disciplinas se comportam, além de servir como controle para as outras proximidades. As diferentes disciplinas possuem peculiaridades na produtividade científica assim como na quantidade de coautores. Grupos de pesquisadores das áreas de Ciências Biológicas, Ciências da Saúde e Ciências Agrárias são as mais colaborativas, enquanto os grupos de Ciências Sociais Aplicadas, Ciências Humanas e Linguísticas, Letras e Artes são menos colaborativos na produção de artigos publicados em revistas. Esses resultados são semelhantes a trabalhos anteriores e confirmam esse padrão nacional. É possível concluir que a proximidade cognitiva tem influência na decisão de colaboração dos pesquisadores.

As proximidades organizacional e institucional também impactam positivamente o processo de colaboração, em especial quando os cientistas pertencem à mesma

instituição. A proximidade organizacional teve uma magnitude pequena em relação às outras dimensões e a significância da variável de vínculo universitário oscilou em alguns modelos para níveis baixos. Estes achados mostram o papel importante da instituição a nível macro e micro, ou seja, tanto as regras e normas quanto a organização integram o processo colaborativo. Isso indica que a incerteza e oportunismo são diluídos quando os pesquisadores estão na mesma instituição ou possuem normas, regras e hábitos em comum.

As relações prévias mantidas pelos pesquisadores durante o triênio, que são representadas pela proximidade social, sugerem que essas interações proporcionam novas oportunidades de colaboração. Embora menor, o efeito é também positivo nos casos de pares de pesquisadores que já interagiram continuamente nos últimos três triênios, mostrando que o processo de conhecimento mútuo entre os pares encoraja a produção científica. A proximidade geográfica interfere, como previsto, positivamente na decisão de colaboração dos acadêmicos. Por fim, pesquisadores que estão na mesma Unidade da Federação tendem a colaborar mais.

A partir das conclusões obtidas deste trabalho é possível ter um panorama de como as colaborações no meio acadêmico são formadas, seus principais catalizadores e entraves. Os fatores apresentados nesta pesquisa focam no atual cenário brasileiro, podendo vir a servir de base no desenvolvimento de metodologias ou análises das instituições cujos objetivos abarcam o aperfeiçoamento da ciência brasileira e desenvolvimento científico, como a CAPES e o CNPq. Em vista disso, o desenvolvimento do SNI perpassa pela intensificação de políticas facilitadoras à colaboração entre pesquisadores, com a promoção de intercâmbios universitários e o financiamento de cientistas para a participação em congressos, criando um ambiente fértil para uma produção científica colaborativa.

Este estudo tem algumas limitações que valem observar. Primeiro, levam-se em conta as colaborações em artigos que se formaram, ou seja, os insucessos nas produções em colaboração são desconsiderados, assim como outros resultados de colaboração como publicação em conferências, capítulos livros e *working papers*. Segundo, não é considerado a qualidade dos artigos, seus fatores de impacto e número de citações. Por último, as variáveis de proximidade são todas representadas por *dummies*; e a unidade de referência geográfica é a Unidade da Federação, o que implica em menor precisão da proximidade geográfica, visto que os estados brasileiros têm grande dimensões. Todavia,

mesmo considerando esse nível de agregação é possível utilizar esses resultados para avaliar o efeito total das dimensões de proximidade.

5. Referências

- ABASSI, A.; HOSSAIN, L.; LEYDESDORFF, L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks **Journal of Infometrics**, v. 6(3), p. 403-412, 2012.
- ALBUQUERQUE, E. M.; SILVA, L. A.; RAPINI, M. S.; SOUZA, S. G. A. Interactions Between Firms and Universities In An Immature System Of Innovation: a survey of industrial R&D-performers firms in Minas Gerais, Brazil Belo Horizonte: **UFMG/Cedeplar**: Texto para discussão, 2005.
- ALBUQUERQUE, E. M.; SILVA, L. A.; PÓVOA, L. Diferenciação Intersetorial na interação entre empresas e universidades no Brasil. **São Paulo em Perspectiva**, v. 19(1), p. 95-104, 2005.
- ALBUQUERQUE, E.; SUZIGAN, M.; KRUSS, G.; LEE, K. **Developing National Systems of Innovation**, Edward Elgar, 2015.
- AGRAWAL, D.; KAPUR, D.; MCHALE, J. How do spatial and social proximity influence knowledge flows? Evidence from patent data **Journal of Urban Economics**, v. 64(2), p 258-269, 2008.
- ALBA, R.; KADUSHIN, C. The Intersection of Social Circles: A New Measure of Social Proximity in Networks **Sociological Methods & Research** v. 5(1), 1976.
- ANCONA, D. G.; CALDWELL, D. F. Demography and design: Predictors of new product team productivity. **Organization Science** v. 3, p. 321-341, 1992.
- ANSELIN, L.; VARGA, A.; ACS, Z. Geographic and sectoral characteristics of academic knowledge externalities. **Papers in Regional Science** v. 79, p. 435–445, 2000.
- ANSELIN, L.; VARGA, A.; ACS, Z. Geographical Spillovers and University Research: A Spatial Econometric Perspective. **Growth and Change** v. 31(4), p. 501-515, 2000.
- ARRAEZ, M. A. Background and Perspectives of the Peer Review System **World neurosurgery**, v. 76(1), p. 33-34, 2011.
- AUDRETSCH, D. B., FELDMAN, M. P. R&D spillovers and the geography of innovation and production. **The American Economic Review**, v. 86(3), p. 630–640, 1996.
- BALLAND, P. A.; BOSCHMA, R.; FRENEN, K. Proximity and Innovation: From Statics to Dynamics, **Regional Studies**, v. 49(6), 2015.

- BAHR, A.; ZEMON, M. Collaborative Authorship in the Journal Literature: Perspective for Academic Librarians Who Wish to Publish **College & Research Libraries** v. 61(5), p. 410-419, 2000.
- BANTEL, K. A.; JACKSON, S. E. Top management and innovations in banking: Does the composition of the top team make a difference? **Strategic Management** v. J. 10, p. 107-124, 1989.
- 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.
- BERGENHOLTZ, C.; BJERREGAARD, T. How institutional conditions impact university–industry search strategies and networks, **Technology Analysis & Strategic Management**, v. 26(3), p. 253-266, 2014.
- 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.
- 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.
- BOEKHOLT, P.; EDLER, J.; CUNNINGHAM, P.C.; FLANAGAN, K. Drivers of international collaboration in research. Study for the European commission, **DG research**, Amsterdam, Brussels, Manchester: European commission, 2009.
- 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.
- BRANSTETTER, L.; OGURA, Y. Is academic science driving a surge in industrial Innovation? Evidence from patent citations. **National Bureau of Economic Research Working Paper Series** n. 11561, 2005.
- BRESCHI, S.; LISSONI, F. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows **Journal of Economic Geography**, v. 9(4), p. 439-468, 2009.

- 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.
- BROEKE, T.; MUELLER, W. Critical links in knowledge networks – What about proximities and gatekeeper organisations? **Industry and Innovation**, v. 25(10), p. 919-939, 2018.
- BODAS FREITAS, I. M.; GEUNA, A.; ROSSI, F. Finding the right partners: Institutional and personal modes of governance of university–industry interactions, **Research Policy**, 42(1), p. 50-62, 2013.
- BOSCHMA, R. Proximity and Innovation: A Critical Assessment. **Regional Studies**, v. 39, 61-74, 2005.
- 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.
- CHEN, C. J. The effects of knowledge attribute, alliance characteristics, and absorptive capacity on knowledge transfer performance **R&D Management**, v. 34 (3), p. 311-321, 2004)
- CHEN, K.; ZHANG, Y.; ZHU, G.; MU, R. Do research institutes benefit from their network position in research collaboration network with industries or/and universities? **Technovation**, p. 94-95, 2020.
- CHOMPALOV, I.; GENUTH, J.; SCHRUM, W. The organization of scientific collaborations. **Research Policy**, v. 31, p. 749–767, 2001.
- 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** v. 42, p. 730-762, 2017.
- COHEN, W.M.; LEVINTHAL, D. A. Innovation and learning: the two faces of R&D. **The Economic Journal** v. 99, p. 569–596, 1989.
- 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.
- 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. 123-131, 2014.
- COOKE, P.; MORGAN, K. 1998 **The Associational Economy. Firms, Regions, and Innovation**. Oxford University Press, Oxford, 1998.
- DAVID, P. A. Knowledge property and the system dynamics of technological change. **Proceedings of the World Bank Annual Conference on Development Economics**, The World Bank, Washington, 1993.
- DAVID, P. A.; MOWERY, D. C.; STEINMUELLER, W. E. Analyzing the economic payoffs from basic research **Economics of Innovation and New Technology**, v. 2, p. 73-90, 1992.
- DUCTOR, L. Does Co-authorship Lead to Higher Academic Productivity? **Oxford Bulletin of Economics and Statistics**, v. 77(3), p. 385-407, 2015.
- EDQUIST, C.; JOHNSON, B. Institutions and organisations in systems of innovation, **Systems of Innovation: Overview and Basic Concepts - Capítulo 2**, London: Pinter Publishers, pp. 41-63, 1997.
- FARBER, M. Single-authored publications in the sciences at Israeli universities. **Journal of Information Science**, v. 31(1), p. 62–66, 2005.
- FELSENSTEIN, D. Handbook of Regional Innovation and Growth: **Human Capital and labour mobility determinants of regional innovation**, Edward Elgar Publishing, 2011.
- FERNÁNDEZ, A.; FERRÁNDIZ, E.; LEÓN, M. D. Are organizational and economic proximity driving factors of scientific collaboration? Evidence from Spanish universities, 2001–2010 **Scientometrics**, v. 126, p. 579-602, 2021.
- GARFIELD, E. English An International language for science. **The Information Scientist**, v. 1, p. 19-20, 1977.
- GEISLER, E.; RUBENSTEIN, A. H. University—Industry Relations: A Review of Major Issues **Cooperative Research and Development: The Industry—University—Government Relationship** p. 43-62, 1989.

- GILLY, J. P.; TORRE, A. Proximity Relations: Elements for an Analytical Framework. **Industrial Networks and Proximity**. Aldershot: Ashgate Publishing, p. 1-17, 2000.
- GITTELMAN, M. “Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology”. **Organization Science** v. 18, p. 724–741, 2007.
- GONÇALVES, E.; OLIVEIRA, P. M.; ALMEIDA, E. Spatial determinants of inventive capacity in Brazil: the role of inventor networks **Spatial Economic Analysis** v. 15(2), p 186-207, 2019.
- GRANOVETTER, M. Economic action and social structure. The problem of embeddedness, **American Journal of Sociology**, v. 91, p. 481-510, 1985.
- HALL, B. H.; MAIRESSE, J.; MOHNEN, P. Measuring the return to R&D. **Handbooks in Economics**, v. 2(24), 2010.
- HALL, B. H.; ROSENBERG, N. (Eds.), Handbook of the Economics of Innovation. **Amsterdam: Elsevier**, v.2, p. 275-314, 2010.
- HANSEN, M. T. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. **Administrative Science Quarterly**, v. 44 (1), p. 82-111, 1999.
- 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 VIERSEN 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.
- HEY, T.; TANSLEY, S.; TOLLE, K. **The fourth paradigm: Data intensive scientific discovery**. Redmond, WA, 2009.
- 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.

- JACOMY, M.; TOMMASO, V.; HEYMANN, S; MATHIEU, S ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software **PLoS ONE** v. 9(6), e98679, 2014.
- KALNINS, A.; CHUNG, W. Social Capital, Geography, and Survival: Gujarati Immigrant Entrepreneurs in the U.S. Lodging Industry. **Management Science**. v. 52, p. 233-247, 2006.
- KARLSSON, C.; GRÅSJÖ, U. Knowledge Flows, Knowledge Externalities, and Regional Economic Development. **Handbook of Regional Science**, p. 413-437, 2012.
- KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy** v. 26, p. 1-18, 1997.
- 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.
- KERR, W., 2005. Ethnic scientific communities and international technology diffusion. **Working paper, Harvard University**, 2005.
- KIRAT, T, LUNG, Y. Innovation and proximity. Territories as loci of collective learning processes, **European Urban and Regional Studies** v. 6, p. 27–38, 1999.
- KREINER, K.; SCHULTZ, M. Informal Collaboration in R & D. The formation of Networks Across Organizations **Organizations Studies** v. 14(2), p. 189-209, 1993.
- 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), p. 457-466, 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.
- LAURSEN, K.; SALTER, A. Searching high and low: what types of firms use universities as a source of innovation? **Research Policy** v. 33, p. 1201-1215, 2004.
- LAWLESS, Jerald F. Negative binomial and mixed Poisson regression. **Canadian Journal of Statistics**, v. 15(3), p. 209-225, 1987.

- LEE, S.; BOZEMAN, B. The impact of research collaboration on scientific productivity. **Social Studies of Science**, v. 35, p. 673–702, 2005.
- LETA, J.; CHAIMOVICH, W. Recognition and international collaboration: The Brazilian case. **Scientometrics**, v. 53(3), p. 325–335, 2002.
- LIANG, W.; DONG-MEI, Z.; GUO-XIN, L.; DI, H.; LI-MEI Z. Measuring Effects of Language Barriers on Knowledge Spillover from Scientific Journals, **2017 International Conference on Management Science and Engineering (ICMSE)**, p. 287-294, 2017.
- Workshop on University Linkages in Europe and North America**, Downing College, Cambridge, 2005.
- MCCARTY, C., & JAWITZ, J. W. Attitudes about publishing and normal science advancement. **Journal of Informetrics**, v. 7(4), p. 850–858, 2013.
- MARSHAL, A. **Principles of Economics: An Introductory Volume**, Londres: Macmillan, 1920.
- MARTINS, W.S.; GONÇALVES, M.A.; LAENDER, A.H.F.; ZIVIANI, N. Assessing the quality of scientific conferences based on bibliographic citations. **Scientometrics**, v. 83, p. 133–155, 2010.
- 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.
- MELIN, G.; PERSSON, O. Studying Research Collaboration using co-authorships **Scientometrics**, v. 36(3), p. 363-377, 1996.
- 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.
- MEYER, M. What is Special about patent citation? Difference between scientific and patent citation. **Scientometrics**. v 49(1), p. 93-123, 2000.
- NEWMAN, M. E. J. Scientific Collaboration network. Network construction and fundamental results. **Physical Review E**, v. 64, 2001.

NONAKA, I., VON KROGH, G. Perspective—Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. **Organization Science**, v. 20(3), p. 635-652, 2009.

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

NOOTEBOOM, B. Cognitive Distance in and Between Communities of Practice and Firms: Where Do Exploitation and Exploration Take Place, and How Are They Connected? **Community, Economic Creativity, and Organization** Oxford Scholarship Online, 2008.

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.

PERKMANN, M.; WALSH, K. University-industry relationships and open innovation: Towards a research agenda. **International Journal of Management Reviews** v. 9(4), p 259–80, 2007.

PERSSON, O.; BECKMANN, M. Locating the network of interacting authors in scientific specialties. **Scientometrics**, v. 33(3), p. 351–366, 1995.

PERSSON, O.; GLANZEL, W.; DANELL, R. Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. **Scientometrics** v. 60, p. 421-432, 2004.

POLANYI, **The Great Transformation**. Beacon, Boston, 1944.

PORTO, G. S.; KANNEBLEY, S.; SELAN, B.; BARONI, J. P. M. T. Rede de interações universidade-empresa no Brasil: uma análise de redes sociais. **Revista de Economia**. Editora UFPR, v.37(especial), p. 51-84, 2011.

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.

- POWELL, W. W.; KOPUT, K.W.; SMITH-DOERR, L. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. **Administrative Science Quarterly** v. 41, p. 116–145, 1996.
- 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, 2007
- REAGANS, R.; ZUCKERMAN, E.W. Networks, diversity, and productivity: the social capital of corporate R&D teams. **Organization Science**, v. 12, p. 502–517, 2001.
- ROMER, P. M. The Origins of Endogenous Growth **Journal of Economic Perspectives**, v. 8(1), p. 3-22, 1994.
- SAXENIAN, A. Silicon Valley's New Immigrant Entrepreneurs, **Public Policy Institute of California**, 1999.
- 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.
- SONNENWALD, D. H. Scientific collaboration **Annual Review of Information Science and Technology**, v. 41, p. 643-681, 2007.
- SORENSEN, O.; RIVKIN, J. W.; FLEMING, L. Complexity, networks and knowledge flow **Research Policy** v. 35 p. 994–1017, 2006.
- SUZIGAN, W.; ALBUQUERQUE, E. The underestimated role of universities for the Brazilian system of innovation. **Brazilian Journal of Political Economy**, v.31 (1), p. 3-30, 2011.
- TEIXEIRA, A. L. S. A percepção dos benefícios e dificuldades na interação Universidade-Empresa: O caso dos grupos de pesquisa mineiros **Gestão e Sociedade**, v. 10(26) p. 1360-1385, 2016.
- UDDIN, S.; HOSSAIN, L.; ABBASI, A.; RASMUSSEN, K. Trend and efficiency analysis of co-authorship network. **Scientometrics**, v. 90, p. 687–699, 2012.
- VAN DER WOUDE, F.; RIGBY, D. L. Co-inventor network and knowledge production in specialized and diversified cities. **Papers in Regional Science**, v. 98(4) p.1833- 1853, 2019.

VANZ, S. A. S.; STUMPF, I. R. C. Scientific Output Indicators and Scientific Collaboration Network Mapping in Brazil, **Collnet Journal of Scientometrics and Information Management**, v. 6(2), p. 315-334, 2012.

YAN, E. Finding knowledge paths among scientific disciplines **Journal of the Association for Information Science and Technology**, v. 65(11), p. 2331–2347, 2014.