



ENEI

Encontro Nacional de Economia Industrial e Inovação

FACE-UFMG

Inovação, Sustentabilidade e Pandemia

10 a 14 de maio de 2021

Colaboração científica e produtividade: uma análise de redes sociais no período 2000-2017

Verônica de Castro Lameira (UFJF/UFSJ);

Eduardo Gonçalves (UFJF);

Daniel Oliveira Costa (UFJF);

Carolina Teixeira Saldanha (UFJF)

resumo:

A comunidade científica brasileira apresentou crescimento de sua produção nos últimos anos. Esse aumento foi acompanhado pela ampliação das colaborações de seus pesquisadores. Este estudo investiga o papel das redes de coautorias de artigos científicos na alta produtividade científica no Brasil entre 2000 e 2017. A fonte de dados é a plataforma Lattes/Cnpq, utilizada para avaliar a produção científica e para construção das redes de colaborações entre pesquisadores. Modelos de contagem são aplicados para explorar como o desempenho científico está relacionado a diferentes medidas de rede de coautoria. Os resultados sugerem papel preponderante das redes de colaboração científica na produção científica. Pesquisadores com maior centralidade na rede de colaborações científicas são mais produtivos, enquanto o tempo de atividade docente está associada à menor produtividade. A supervisão de teses e o vínculo com universidades públicas podem incrementar a publicação de pesquisadores.

palavras-chave:

Produtividade de pesquisa, colaboração científica, redes sociais, Plataforma Lattes.

Código JEL: D85, J24, O34.

Área Temática: redes e sistemas urbanos, regionais e nacionais

1. Introdução

A produção científica brasileira vem crescendo nos últimos anos, acompanhando a tendência mundial. Grande parte desse crescimento é sustentado pela colaboração científica (Royal Society, 2011; Sidone, Haddad e Mena-Chalco, 2016) estimulada por políticas destinadas a fomentar a parceria e o trabalho em rede entre os vários componentes do sistema de pesquisa (Abramo, D'Angelo e Costa, 2009).

As redes de colaboração científica desempenham papel fundamental na produção de novos conhecimentos em virtude de sua influência positiva sobre a pesquisa e sua difusão, e tem se tornado cada vez mais proeminente dentro e entre diferentes áreas do conhecimento nas últimas décadas, ampliando seu caráter coletivo (He, Ding e NI, 2011). Diante disso, são crescentes os esforços de pesquisa direcionados à compreensão dos determinantes da produção científica. Alguns desses estudos se dedicam à explicação da quantidade de publicações pela comunidade acadêmica brasileira (Glänzel, Leta, e Thijs, 2006; Leta e Chaimovich, 2002).

Outro grupo de autores possui como foco a qualidade das publicações, bem como o impacto das mesmas através de citações (Leta e Chaimovich, 2002; Durieux e Gevenois, 2010; Leta e Ribeiro, 2012; Leta *et al.*, 2013; Sooryamoorthy, 2009; Mingers e Leydesdorff, 2015; Ioannidis, Boyack e Baas, 2020), a proximidade geográfica (Boschma, 2005; Ponds, Oort, e Frenken, 2009) e sua distribuição desigual no espaço geográfico (Royal Society, 2011), ou ainda, sua relação com a endogenia acadêmica (Pelegri e França, 2020).

De maneira geral, é patente a importância assumida pelas redes de colaboração científica na produtividade acadêmica (Glänzel, 2000; Glänzel e Schubert, 2001; Egghe, Goovaerts e Kretschmer, 2008) uma vez que tais redes de pesquisa contribuem para determinar a estrutura cognitiva e social dos campos científicos, difusão do conhecimento e consequente ampliação na produção científica (Horta, Veloso e Grediaga, 2010; Bordons *et al.* 2015).

À vista disso, diversos estudos têm se dedicado à compreensão dos relacionamentos entre os pesquisadores nas redes de colaboração científica. Estudos prévios de redes de coautorias analisam diferentes aspectos, como, por exemplo, a evolução da rede de coautorias (Barbarasi *et al.*, 2002; Moody, 2004; Newman, 2001; Leydesdorff e Wagner, 2008), a interdisciplinaridade das colaborações (Abramo *et al.*, 2018; Iglič *et al.*, 2017; Júnior, 2019; Lopes *et al.*, 2012; Mena-Chalco *et al.*, 2014; Shi *et al.*, 2018) e os padrões geográficos das redes de colaboração científica (Sidone, Haddad e Mena-Chalco, 2016).

Não obstante, embora extensa literatura tenha sido realizada para analisar diferentes aspectos da produção científica, a maioria dos estudos anteriores não dedicou atenção especial ao nicho de pesquisadores altamente produtivos com foco na composição das suas redes sociais e explorando diferentes informações da plataforma Lattes como supervisões, publicação de artigos científicos e idade docente. Diante disso, este artigo analisa a estrutura das redes de coautoria de artigos científicos no período de 2000 a 2017 e explora a relação entre a produção de pesquisa de cientistas, medida pelo número de artigos publicados, e as características de suas redes de coautorias.

A compreensão do papel das redes de colaboração para a alta produção científica é importante, pois pode subsidiar decisões de políticas públicas relacionadas ao estímulo e fortalecimento de laços estratégicos ao desenvolvimento da pesquisa brasileira. Além disso, pode auxiliar aos formuladores de políticas no fomento de projetos colaborativos que insiram atores altamente produtivos e centrais ao desenvolvimento da produção acadêmica.

As informações profissionais dos pesquisadores são coletadas do currículo vitae (CV) na Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). As evidências obtidas revelam o papel preponderante que as redes de colaboração científica possuem na produção científica (Glänzel, 2001; Lee e Bozeman, 2005; Abramo, D'Angelo, Di Costa, 2009; Bordons, Aparicio e Costas, 2013; Haddad, Mena-Chalco e Sidone, 2017).

As colaborações são intensificadas ao longo dos triênios, com ampliação da rede em tamanho. Pesquisadores de ciências da saúde, ciências biológicas e ciências agrárias aparecem em maior expressividade na rede pois essa área figura com maior número de artigos produzidos em colaboração.

De outro modo, pesquisadores associados à linguística, letras e artes apresentam poucos coautores ao publicar seus artigos. Adicionalmente, as estimativas revelaram que mestres e doutores com maior centralidade na rede de colaborações científicas são mais produtivos, enquanto o tempo de docência diminui a produtividade. Por sua vez, a supervisão de teses e o vínculo com universidades públicas estão associados à maior produção de artigos científicos.

Propondo-se a contribuir para os já referenciados estudos brasileiros voltados à colaboração científica e à produtividade acadêmica, este estudo é dividido da seguinte forma, juntamente com esta introdução. A segunda seção recupera a literatura sobre produtividade, redes de colaborações e o papel que estas representam no desempenho dos pesquisadores altamente produtivos. Na terceira seção são apresentadas a metodologia e as métricas de redes sociais adotadas, a base de dados e amostra. A quarta seção apresenta os resultados e seus desdobramentos. Na última seção, são apresentados os comentários finais, contribuições desse estudo e sugestões para trabalhos futuros.

2. Produtividade científica e redes sociais

A produtividade científica vem sendo objeto de estudo desde a década de 1920. Pioneiramente, Lotka (1926) se debruçou sobre a quantidade de publicações de químicos e físicos entre 1907 e 1916, analisando as distribuições de frequência de produtividade desses grupos de pesquisadores. O autor identificou já naquela época o que ainda observamos nos dias de hoje: grande parte da produção científica é resultado do trabalho de um seleto grupo de pesquisadores, enquanto um grupo maior contribui pouco para o desenvolvimento da ciência. A descoberta foi denominada Lei de Lotka.

Desde a publicação de Lotka em 1926 até os dias de hoje, a literatura científica sobre os determinantes da produtividade de um pesquisador avançou e tem mostrado uma variedade de atributos associados à produtividade científica: relacionados às características do pesquisador, como sexo, idade, titulação, associados às instituições, como o tamanho do corpo docente e infraestrutura disponível, ou ainda, aos recursos públicos e privados existentes e à quantidade de alunos para auxiliar a pesquisa (Fox, 1992; Alvarez, Júnior e Carolo, 2013; Bordons *et al.*, 2015; Abramo, D'Angelo e Murgia, 2017).

Fox (1992) analisou uma amostra de professores de economia, ciência política, psicologia e sociologia para investigar a relação entre a produtividade acadêmica e a sobreposição das atividades de pesquisa e docência. Os resultados sugerem que atividades de ensino estão associadas à menor produtividade docente, enquanto a pesquisa e todas as atividades não associadas com o ensino aumentam a produtividade do pesquisador.

Mais recentemente, Nunes-Silva *et al.* (2019) estudaram a produtividade com o objetivo de identificar o perfil e o grau de produtividade dos pesquisadores de Gestão Universitária entre 2012 e 2017 com informações coletadas na Plataforma Lattes (CNPq). Os principais resultados indicaram que existe uma forte colaboração entre os autores no desenvolvimento da pesquisa, com indícios da existência de uma elite de pesquisadores em Gestão Universitária, em sua maioria, professores mestres e doutores, graduados ou pós-graduados em administração.

A colaboração facilita a geração e seleção de ideias originais devido às sinergias que podem ser obtidas entre cientistas com formações complementares, ou mesmo de diferentes disciplinas (Rigby e Edler, 2005; Katz e Martin, 1997). Além disso, o contato face a face na colaboração permite que os pesquisadores compartilhem conhecimentos, experiências e técnicas, transferindo conhecimentos tácitos, acelerando o processo de pesquisa e ampliando sua visibilidade (Katz e Martin, 1997; Sonnenwald, 2007; Bordons *et al.*, 2015).

Em um horizonte de tempo mais longo, a colaboração permite superar as lacunas do indivíduo em competências, por meio da ativação de processos de aprendizagem, incluindo a aprendizagem de conhecimento tácito (He, Geng e Campbell-Hunt, 2009; Beaver, 2001). Além disso, pode permitir o uso mais eficiente do tempo, devido à divisão do trabalho entre os diferentes membros da equipe (Barnett, Ault e Kaserman, 1988).

Outro aspecto positivo da colaboração é o fortalecimento das motivações de um cientista para realizar pesquisas e publicações, pois oferece uma maneira de superar o isolamento intelectual e de

compartilhar seu entusiasmo (Beaver e Rosen, 1978). No entanto, às vezes um cientista desenvolve colaborações apenas alavancar seu próprio capital social, e não para aumentar sua produtividade (Pezzoni, Sterzi e Lissoni, 2012; Ynalvez e Shrum, 2011).

A despeito de todos os benefícios da colaboração científica supracitados, um cientista pode publicar sozinho, indicando que o pesquisador provavelmente possui todos os ingredientes necessários para a realização de publicações sobre determinados temas, podendo o mesmo selecionar suas redes colaborativas da maneira mais vantajosa possível (Medoff, 2003; Hollis, 2001). Por esse motivo, um cientista com menor propensão a colaborar pode geralmente ser caracterizado por uma maior produtividade de pesquisa. Por outro lado, é possível ainda que o tempo e o custo necessários para a coordenação de uma rede de colaboração sejam tais que reduzam a produtividade de pesquisa de um cientista (Landry, Traore e Godin, 1996; Ynalvez e Shrum, 2011).

Paralelamente ao crescimento da comunidade científica e do número de publicações houve incremento no número de colaborações científicas, fator preponderante ao incremento da produtividade dos pesquisadores. Desse modo, diversos estudos têm intensificado a investigação da relação entre colaborações científicas e produtividade (Landry, Traore e Godin, 1996; Abramo, D'Angelo e Costa, 2009; Ynalvez e Shrum, 2011; Bordons et al., 2015; Souza, Amorim e Rêgo, 2016; Abramo, D'Angelo e Murgia, 2017).

Landry, Traore e Godin (1996) avaliam o impacto de diferentes relações de colaboração (entre pesquisadores universitários, entre pesquisadores universitários e a indústria, e entre pesquisadores universitários e outras instituições) na produtividade acadêmica de pesquisadores universitários. Os resultados mostram que a colaboração, seja ela empreendida com universidades, indústrias ou instituições, pode de fato aumentar a produtividade dos pesquisadores. O efeito da colaboração na produtividade varia de acordo com a proximidade geográfica dos cientistas com seus parceiros e com seu campo de pesquisa. Verificou-se, ainda, que a colaboração entre pesquisadores e a indústria teve um impacto significativamente maior na produtividade do que as colaborações entre pesquisadores e seus pares ou pesquisadores e outras instituições.

Com o intuito de medir os efeitos da colaboração no desempenho de pesquisadores das áreas tecnológicas de 78 universidades italianas entre 2001 e 2003, Abramo, D'Angelo e Costa (2009) empregam a publicação científica em periódico internacional como *proxy* para a produção da pesquisa acadêmica. A coautoria na extensão universitária foi utilizada como *proxy* da colaboração científica entre as organizações de pesquisa. Os resultados mostram que a maior qualidade da publicação está associada a colaborações e o grau de correlação entre produtividade e intensidade de colaboração varia substancialmente entre as diferentes áreas.

Em um contexto de instituições de pesquisa com recursos limitados e empregando uma amostra de cientistas filipinos, Ynalvez e Shrum (2011) examinam fatores associados ao tamanho e localização das redes de colaboração e investigam suas relações com a publicação em periódicos estrangeiros e nacionais. Possuir título de doutor está associado a aumento na produção em periódicos locais, enquanto a proporção de contatos nos países desenvolvidos está negativamente relacionada à publicação local, sugerindo que redes maiores e altamente domésticas são o que importa para a produtividade local. Ser doutor e ter maior rede de contatos nos países desenvolvidos são importantes fatores determinantes das publicações em periódicos estrangeiros. Além disso, os autores verificam que a colaboração científica não leva à produtividade da publicação local e estrangeira.

Bordons *et al.* (2015) analisam a estrutura das redes de coautoria em três campos diferentes (Nanociência, Farmacologia e Estatística) na Espanha durante um período de três anos (2006-2008) e exploram a relação entre o desempenho de pesquisa de cientistas e sua posição em redes de coautoria. Uma rede de coautoria mais densa é encontrada nos dois campos experimentais em relação à Estatística, onde a rede é de natureza menos conectada e mais fragmentada. Usando o índice *g* como um *proxy* para o desempenho de pesquisa individual, um modelo de regressão de Poisson é usado para explorar como o desempenho está relacionado a diferentes medidas de rede de coautoria e para revelar diferenças entre campos. O número de coautores (centralidade de grau) e a força dos links mostram uma relação positiva com o índice *g* nos três campos.

Os efeitos da interação universidade-empresa na produtividade de pesquisadores docentes é investigado por Alvarez, Júnior e Carolo (2013). Os resultados apontam que pesquisadores colaboradores publicaram em média um artigo a mais, por período, do que aqueles que não interagem. O tempo de pesquisa é importante para explicar a produtividade docente em termos do número de artigos publicados, confirmando a hipótese de um pico de produtividade docente em pesquisa em torno de 29 anos posteriormente à obtenção do Doutorado. Os achados indicam ainda uma relação positiva entre o número de orientações de Doutorado e a produtividade docente, em que cada orientação de tese de Doutorado concluída no período corrente proporciona um aumento médio do número de artigos publicados no período posterior entre 7,8% e 25,2%.

Em uma análise aprofundada da relação entre os diferentes tipos de colaboração e produtividade em pesquisa, Abramo, D'Angelo e Murgia (2017) mostram que apenas a colaboração em nível interno e doméstico tem um efeito positivo na produtividade da pesquisa. Os autores empregam dados de 30.866 professores de universidades italianas no período de 2001-2012, aplicando diferentes modelos de painel cross-lag. Levando em consideração a quantidade e qualidade das publicações, mostram que todas as formas de colaboração são afetadas positivamente pela produtividade da pesquisa.

Em estudo análogo, porém para diferentes áreas, Souza, Amorim e Rêgo (2017) analisam a rede de coautoria entre todos os bolsistas de produtividade do CNPq no Brasil na área de probabilidade e estatística no período 2009 a 2013 e descrevem como medidas da rede influenciam a produtividade de pesquisadores a partir da Plataforma Lattes do CNPq. Os resultados obtidos sugerem que as posições de centralidade de um autor são importantes para explicar sua produtividade. A centralidade de proximidade tem um efeito negativo na produtividade, enquanto a centralidade de grau tem efeito positivo.

A coautoria acadêmica dos docentes permanentes vinculados aos PPGS da área de Economia no Brasil é objeto de estudo de Haddad, Mena-Chalco e Sidone (2017). A partir de métricas de redes sociais, os autores caracterizam a colaboração entre instituições da área, analisando as coautorias de artigos em periódicos, livros, e capítulos de livros com a construção de um índice multidimensional de desempenho científico para o período 2004-2015. A pesquisa revela que economistas com maior impacto na profissão adotam uma estratégia caracterizada por um equilíbrio entre qualidade, quantidade e colaboração científica.

Ao que parece, a relação entre a produção científica e a estrutura das redes de colaboração científica ainda não foi exaustivamente investigada. Há predominância de pesquisas exploratórias e em áreas individuais ou envolvendo departamentos, faculdades, instituições, e amostra de indivíduos em contextos particulares.

Esse trabalho contribui no campo da literatura sobre produtividade científica com foco na colaboração em publicações, além de explorar a contribuição das supervisões acadêmicas na produtividade. Investigação adicional é proposta aos pesquisadores doutores altamente produtivos, controlando pelas áreas do conhecimento e demais características ocupacionais. Nota-se papel central nas colaborações por parte desses grupos de pesquisadores. Esse fato pode ser explicado pela maior habilidade que estes possuem em gerir efetivamente as colaborações (Abramo, D'Angelo e Murgia, 2017).

Conforme aludido, a colaboração relaciona-se à produtividade e à qualidade da pesquisa, portanto, a seção a seguir designa os procedimentos metodológicos utilizados para determinar se ocorre colaboração e se esse fenômeno afeta a produtividade de pesquisadores.

3. Estratégia metodológica

3.1. Medidas de colaboração e hipóteses

Este estudo avalia a produtividade da pesquisa com ênfase na estrutura das redes de colaboração dos pesquisadores. As hipóteses do estudo são formuladas a partir de contribuições da literatura quanto à correlação entre características das redes sociais para difusão do conhecimento (Coleman, 1988; Bettencourt *et al.*, 2007; Maggioni e Uberti, 2007; Lobo e Strumsky, 2008; Miguelléz e Moreno, 2011).

A colaboração é analisada por meio da coautoria¹ em publicações científicas. E, embora este seja um indicador limitado da dinâmica obscura das colaborações científicas, uma vez que nem todas as coautorias são baseadas em contribuições colaborativas e nem todos os autores que colaboram tornam-se coautores (Kartz e Martin, 1997; Laudel, 2002; Sidone, Haddad e Mena-Chalco, 2016), evidências na literatura mostram uma correlação positiva entre colaboração e coautorias, sendo este indicador útil ao estudo de diferentes aspectos da colaboração científica (Glanzel e Schubert, 2004; Bordons et al., 2015).²

Medidas de intermediação ou entrelaçamento dos nós - ou importância dele nas pontes de ligação (*betweenness*), proximidade (*closeness*) possibilitam uma análise a respeito da posição de cada ator na rede social. Essas medidas são fundamentais para mensurar a relação das colaborações científicas na produtividade dos pesquisadores.

A centralidade de intermediação (*betweenness*) calcula o número de vezes que o nó está no caminho geodésico (menor caminho entre dois nós) de todos os nós na rede. Calculada por:

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

Assumindo-se que $C_B(v_k)$ representa a centralidade de intermediação; n o número de nós da rede; e v_k o nó examinado. Ressalta-se que não há caminho entre os nós i e j , $\frac{g_{ij} v_k}{g_{ij}}$ assume o valor de 0. g_{ij} é a distância entre os nós i e j ponderados pelo nó central k .

A proximidade leva em consideração a distância dos nós, mas para que essa medida não seja simplesmente a distância dos nós e, por isso, desproporcional à proximidade, a distância é elevada a -1, para que quanto maior esta, menor o coeficiente de proximidade, ou seja:

$$C_C(v_k) = \frac{1}{\sum_{j=1}^n \text{dist}(v_j, v_k)} \quad (2)$$

A centralidade de proximidade é representada por $C_C(v_k)$, que é o inverso da distância dos nós v_j e v_k . Nós com maior centralidade compartilham conhecimento com muitos outros na rede (Burt, 1992). Em outras palavras, essas organizações podem servir como organizações de ponte que coordenam recursos e informações entre atores desconectados ou organizações que operam em diferentes locais e clusters (Freeman, 1979; Abbasi *et al.*, 2012; Saban, 2015; Jung, 2017).

- **Hipótese 1 - *Betweenness***: pesquisadores com alto grau de entrelaçamento tendem a se ligar com outros que também possuem elevado *betweenness*, corroborando para o ampliar sua produtividade.
- **Hipótese 2 - *Closeness***: proximidade dos agentes na rede fortalece as colaborações e, por conseguinte, sua produtividade.

3.2. Estratégia empírica

A variável dependente é uma variável de contagem inteira e não negativa. Essa particularidade faz com sua distribuição não se aproxime da distribuição normal e requer a utilização de modelos de regressão especiais, uma vez que os tradicionais modelos lineares não acomodam esses dados.

¹ Diferentes indicadores foram propostos para quantificar a colaboração em pesquisas (para detalhamento consultar, por exemplo, Egghe, 1991; De Haan, 1997; Glänzel & Schubert, 2004; Vinkler, 2010).

² A coautoria retrata apenas a parte mais formal das múltiplas atividades envolvidas na colaboração. De Haan (1997) sugere seis indicadores de colaboração científica: coautoria; editoria compartilhada de publicações; supervisão compartilhada de projetos de doutorado; projetos de pesquisa em conjunto; participação em programas formais de pesquisa; e organização compartilhada de conferências científicas. Roa e Cellis (2002) sugerem como fonte alternativa de registro da cooperação os agradecimentos presentes nas publicações. Laudel (2002) relata que cerca de metade das colaborações científicas são formalmente invisíveis.

Diferentes alternativas surgem para estimação desse tipo de dado, sendo o modelo de Poisson a principal referência para acomodar eventos aleatórios associados a observações de contagem.

No entanto, a escolha entre as opções será dada pela característica dos dados, sendo imprescindível a análise descritiva da variável dependente. Nesse estudo, há sobredispersão dos dados de publicações, uma vez que as publicações estão concentradas em uma parcela pequena de pesquisadores, fazendo com que a média seja inferior à variância, descartando-se então o uso do modelo de Poisson, o qual presuppõe equidispersão.

Em virtude da variável dependente utilizada, o número de publicações, a metodologia considerada mais adequada pela literatura é o modelo binomial negativo, uma vez que permite acomodar as características de contagem e sobredispersão, permitindo melhor modelagem da variância (Cameron e Trivedi, 2008). Essa abordagem trata a sobredispersão como resultado de heterogeneidade não observada, incorporando ao modelo de Poisson um componente aleatório.

O modelo de Poisson possui especificação expressa na equação (3):

$$f_y(y, \mu) = e^{-\mu} * \frac{\mu^y}{y!}, y = 0, 1, 2, \dots (3)$$

Em que neste estudo y representa o número de artigos publicados e μ o número esperado de artigos publicados, denominado também de parâmetro intensidade ou taxa. Como há apenas um parâmetro na distribuição de Poisson, a igualdade entre média e variância é assumida (equidispersão de dados). Todavia, como supracitado, a média não é idêntica à variância em nosso estudo, sendo na verdade sobredispersos. Desse modo, as características do modelo binomial negativo podem ser expressas como segue, relaxando a hipótese de equidispersão do modelo de Poisson (Greene, 2008):

$$Pr(Y = y|\mu, \alpha) = \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}} \frac{\mu}{\mu+\alpha^{-1}} \left(\frac{\mu}{\mu+\alpha^{-1}}\right)^y, y = 0, 1, 2, \dots; \alpha > 0, \mu > 0 (4)$$

$$E[y_i|X_i, \varepsilon_i] = h_i \mu_i; \mu_i = E[y_i|X_i] (5)$$

Em que o Γ corresponde à função Gama padrão, e α ao grau de dispersão dos dados. Quanto maior α , maior a dispersão dos dados. Quando $\alpha = 0$, o modelo binomial negativo retoma ao modelo de regressão de Poisson. Assume-se que $h_i = \exp \varepsilon_i$ segue uma distribuição gama com um valor médio de 1 e uma variância gama. Nesse caso, a média e a variância da distribuição binomial negativa marginal podem permitir a sobredispersão como segue.

$$E[y_i|X_i = \mu_i] (6)$$

$$Var[y_i|X_i] = \mu_i[1 + \alpha\mu_i] (7)$$

Diante do objetivo desse estudo, qual seja, analisar a relação entre publicações científicas e a estrutura de redes de colaboração dos pesquisadores, uma regressão binomial negativa é estimada pelo método da Máxima Verossimilhança e possui a seguinte forma:

$$Y_i = \alpha + x_i' \beta + \varepsilon_i (8)$$

Em que Y_i é o número de artigos produzidos pelo pesquisador i ao longo do período 2000-2017. As variáveis explicativas x_i' corresponde ao vetor de variáveis explicativas descritas na tabela 1, enquanto ε_i corresponde ao termo de erro aleatório.

Em algumas situações, o resultado “zero” do processo que gera os dados é qualitativamente diferente dos resultados positivos; isto constitui uma limitação dos modelos de Poisson ou do binomial negativo. Diante disso, o modelo hurdle é uma alternativa, uma vez que possui duas partes: a primeira de resposta binária, geralmente logit, e a segunda um modelo de contagem truncado (Hilbe, 2011). Tal partição permite a interpretação de que observações positivas são derivadas do cruzamento do ponto de

corte em 0; a primeira parte modela a probabilidade de que o ponto de corte seja cruzado³. A especificação hurdle lida com modelos mistos da seguinte forma:

$$f_1(\cdot), \text{ de modo que } Pr(y = 0) = f_1(0) \text{ e } Pr(y > 0) = 1 - f_1(0) \text{ (9)}$$

A contagem positiva é proveniente da densidade truncada $f_2(y|y > 0) = f_2(y)/\{1 - f_2(0)\}$, que é multiplicado por $Pr(y > 0)$ a fim de garantir a soma das probabilidades idêntica a um (Cameron e Trivedi, 2008). No caso desse estudo, uma probabilidade binomial determina a resposta binária se o pesquisador tem publicação zero ou valor positivo; se a publicação é positiva, o ponto de corte (hurdle) é cruzado e a distribuição condicional dos positivos determinada por um modelo de contagem truncado em zero.

Os modelos de Poisson, Binomial Negativo e Binomial Negativo Logit Hurdle são calculados para examinar a relação entre colaborações científicas acadêmicas e produtividade em publicação, em que este último permite a partição da variável dependente em dois grupos independentes, com a probabilidade de apresentar um artigo publicado ou maior número de artigos publicados. A comparação de qualidade de ajuste dos modelos é feita pelos critérios AIC, em que um menor AIC indica melhor qualidade de ajuste do modelo aos dados.

Existem várias pesquisas sobre a relação entre colaborações inventivas e produtividade acadêmica. O modelo de Poisson tem sido comumente utilizado, mas suas limitações foram recorrentemente apontadas na literatura (Hill *et al.*, 2008). Nesse sentido, nos últimos anos, muitos estudos adotaram modelos binomiais negativos. Este artigo propôs um modelo binomial negativo logit hurdle. Como resultado da estimativa, foi identificada a presença de sobre-dispersão. Além disso, o teste de Wald confirmou que a configuração de barreira era válida e havia uma diferença entre os resultados dos modelos de barreira e os de configurações binomiais negativas gerais.

A estratégia empírica tem limitações. Primeiro, são analisados dados autodeclarados sobre publicações na plataforma Lattes. Apesar de sua relativa confiabilidade (Dias, Moita e Dias, 2016), não há como avaliar a qualidade dos artigos, uma vez que esses dados são faltantes. Isso não significa necessariamente que a ausência de diferenças significativas no número de artigos entre colaboradores e não colaboradores signifique que não haja diferença em sua qualidade. Pesquisadores colaboradores e não colaboradores podem ter um número igual de artigos de qualidades diferentes. Com nossos dados não é possível testar essa hipótese. Estudos futuros comparando colaboradores e não colaboradores precisam mensurar não apenas o número de publicações, mas também a qualidade das publicações.

Em segundo lugar, analisamos apenas um tipo de publicação - artigos. Este tipo de publicação abrange apenas uma parte da produção e não é igualmente importante em diferentes campos de pesquisa (Larivière *et al.*, 2006). Nossas regressões incluem *dummies* para grandes áreas do conhecimento para levar o campo de pesquisa em consideração.

3.3. Dados e criação da rede

Os dados utilizados neste estudo são provenientes da base de dados da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e correspondem às informações de artigos publicados, supervisões acadêmicas, vínculos profissionais (e suas respectivas durações) e endereço profissional de pesquisadores, professores universitários, revisores de periódicos e doutores, conformando um total de 1.826.002 indivíduos, 5.822.614 artigos publicados e 753.677 autores. Destes, 97% são brasileiros, e este mesmo percentual apresentou vínculo empregatício com universidade, sendo 1.465.062 docentes, 431.794 possuem mestrado e 260.918 possuem título de doutor.

No que se refere à titulação dos autores, 44,5% são doutores, 35% são mestres e 7,5% são graduados. Ressalta-se o elevado percentual de professores entre autores, 86%, e destes, 67% possui vínculo com uma universidade pública, corroborando estudos anteriores de que universidades são

³ Modelos de contagem hurdle e inflado de zero são os dois os principais métodos usados para lidar com dados de contagem com excesso de zeros. Mas, embora os modelos inflacionados de zero exijam contagens excessivas de zeros nos dados, modelos de barreira podem ser ajustados para modelos com poucas contagens zero (Hilbe, 2011).

responsáveis pela maioria da atividade científica, padrão típico de países em desenvolvimento (Sidone, Haddad e Mena-Chalco, 2016). Dos vínculos de pesquisadores com universidades, 86% ocorreu com finalidade docente, e 54% já trabalhou em alguma empresa.

Os artigos são escritos solitariamente ou em parceria, a qual pode conter diversos autores, sendo esta identificação feita a partir do registro do nome do artigo, em que o número de coautorias é dado pela quantidade em que um mesmo artigo aparece registrado na plataforma Lattes/CNPq.⁴

3.3.1 Variável Dependente

A variável dependente corresponde ao número de publicações do pesquisador no período de 2000 a 2017. Além disso, efeitos do tempo de carreira docente são controlados (Horta e Veloso 2010) junto aos efeitos de sobreposição de atividade docente e de pesquisa através do número de orientações de teses e dissertações (Pelegri e França, 2020).

3.3.1 Variáveis de controle

A análise inclui variáveis de controle que estão potencialmente relacionadas com o número de publicações dos pesquisadores mestres e doutores. O primeiro grupo de variáveis de controle inclui características profissionais do pesquisador associadas à atividade docente: se é docente, o tempo de docência, o número de teses orientadas. O segundo grupo de variáveis de controle corresponde à características de vínculo empregatício do pesquisador: universidade pública ou não. A grande área do conhecimento é controlada usando 9 *dummies* para cada área, usando outros como categoria de referência. O terceiro grupo de variáveis inclui medidas de redes sociais. As estatísticas descritivas das variáveis de controle são apresentadas na Tabela 1.

Tabela 1 – Estatísticas descritivas

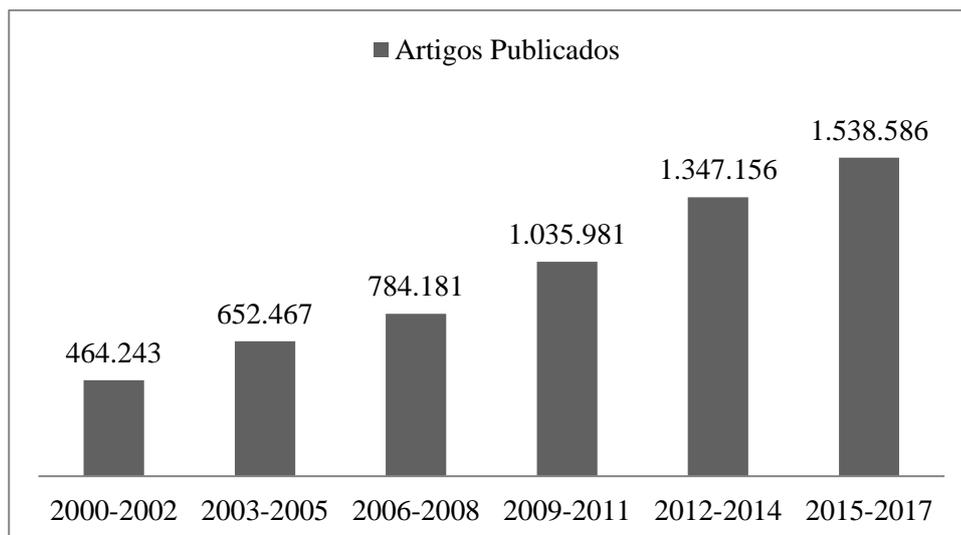
Variável	Observações	Média	Desvio-padrão
<i>Dependente</i>			
Número de Publicações (npublic)	691.803	8,43	24,46
<i>Controles</i>			
Professor (prof)	691.803	0,82	0,39
Tempo docente (tempd)	691.803	79,55	110,08
Número de orientações de teses (ntese)	691.803	0,44	2,36
Universidade Pública	691.803	0,57	0,49
<i>ln Betweennes</i>	691.803	2,56	4,10
<i>Closeness</i>	691.803	0,00	0,02
Área do conhecimento	691.803	-	-

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

As ligações dos pares autor (i) e coautor (j) correspondem a informações trienais. Cada colaboração de autoria para o mesmo artigo no triênio indica uma ligação entre um par de autores. A partir dos dados das figuras 1 e 2 é possível notar um aumento do número de publicações ao longo do período sob análise, bem como crescimento da cooperação científica.

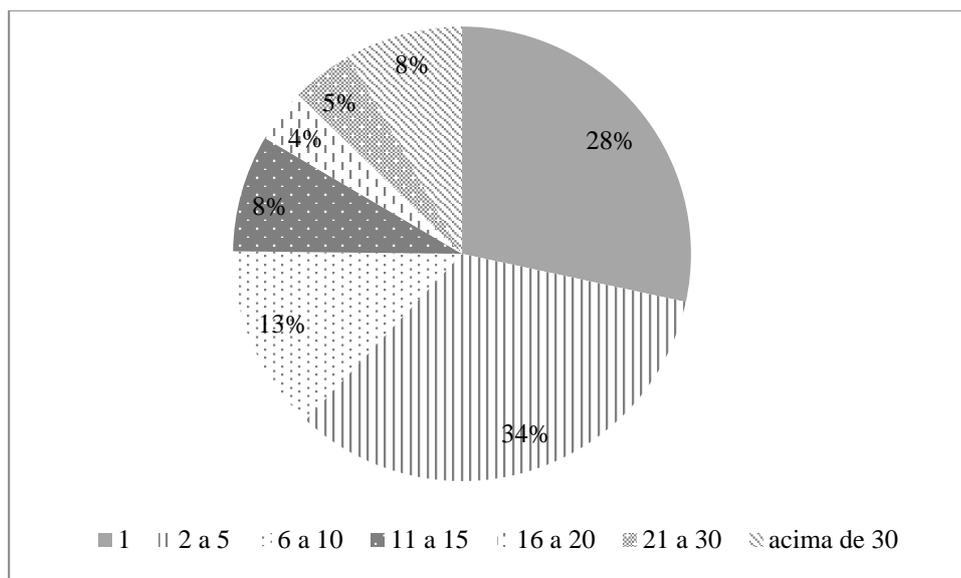
⁴ Vale lembrar que por ser uma fonte de dados em que as informações são autodeclaradas há a possibilidade de erros de digitação do título do artigo, inviabilizando sua identificação como fruto de uma parceria científica e consequentemente subestimando tais colaborações.

Figura 1- Número de artigos científicos produzidos no Brasil por triênio (2000-2017)



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

Figura 2 – Quantidade de colaborações dos pesquisadores no período 2000 a 2017



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

Tabela 2 - Distribuição dos pesquisadores, por situação de publicação 2000-2017

Não publicam		Publicam		Total Absoluto
Nº absoluto	%	Nº absoluto	%	
1.380.612	75,6	445.390	24,4	1.826.002

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

Durante o período analisado é possível observar que apenas 24% dos pesquisadores apresentaram ao menos uma publicação (tabela 2), denotando que a publicação de artigos científicos no Brasil é altamente seleta. Além disso, entre os pesquisadores que fazem parte desse grupo 71%

produzem junto a outros pesquisadores (tabela 3), sugerindo grande importância dos laços entre pesquisadores.

Tabela 3 - Distribuição dos pesquisadores, por situação de colaboração 2000-2017

Não cooperam		Cooperam		Total
Nº absoluto	%	Nº absoluto	%	Absoluto
128.683	28,9	316.656	71.1	445.390

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

As áreas do conhecimento são agrupadas em nove grandes áreas, seguindo a classificação da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (Capes), quais sejam: Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências da Saúde, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas, Linguística, Letras e Artes, e outros.⁵ Os dados dos pesquisadores, publicações e colaborações se mostram bem distintas em quantidade de produção e colaboração, corroborando estudos, como o de Mena-Chalco *et al.* (2014), o qual identifica as áreas de Ciências da Saúde, Ciências Biológicas e Ciências Agrárias como as mais produtivas e coletivas, como apresentado na tabela 4.

Tabela 4 – Número de artigos publicados por área do conhecimento (2000-2017)

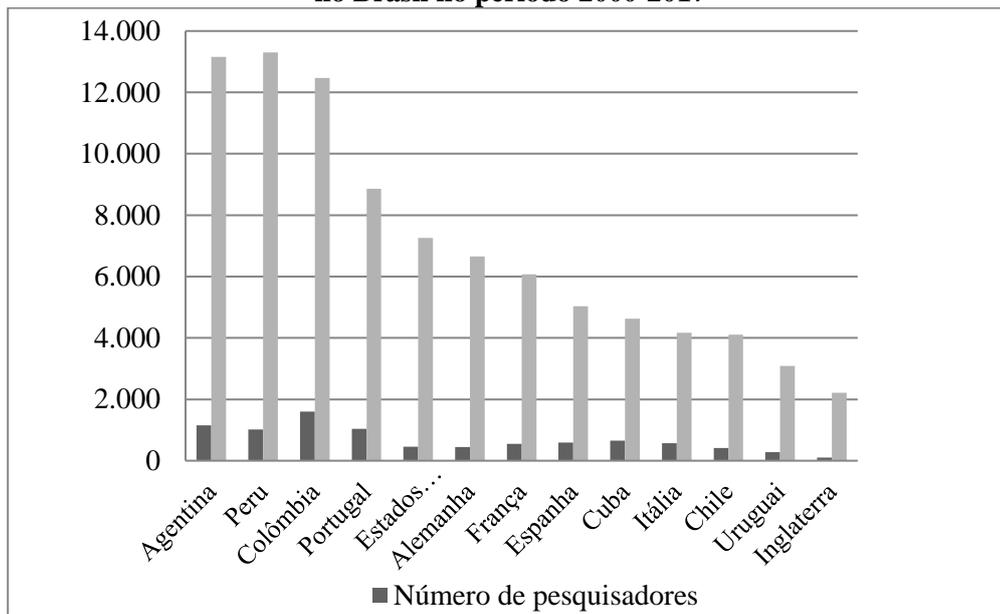
Área do conhecimento	Número de pesquisadores	Número de Artigos publicados por área	Número de artigos publicados com cooperação
Ciências Exatas e da Terra (EXA)	187.095	822.628	49.126
Ciências Biológicas (BIO)	116.950	947.069	51.496
Engenharias (ENG)	120.654	417.245	30.629
Ciências da Saúde (SAUDE)	293.401	1.582.453	87.086
Ciências Agrárias (AGR)	90.014	751.487	37.083
Ciências Sociais Aplicadas (CSA)	257.240	555.465	59.152
Ciências Humanas (CH)	286.648	618.395	67.176
Linguísticas, letras e artes (LLA)	110.051	178.497	21.961
Outros	17.510	22.151	3.013
Total	1.826.002 ¹	5.895.390	406.722

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. Nota 1: 346.439 pesquisadores não possuem informação de área do conhecimento. Além desse total, 232.736 artigos foram publicados sem que fosse registrada a área dos autores.

No âmbito da participação internacional nas colaborações (figura 3), observam-se 123.425 publicações com participação estrangeira, com maior participação do Peru, com 13.303 colaborações, seguido por Argentina e Colômbia com 13.157 e 12.473, respectivamente. Tais colaborações, são importantes à visibilidade e ao valor agregado da produção científica brasileira (Leta, Thijs e Glänzel, 2013; Sooryamoorthy, 2009), porém, como salienta Royal Society (2011), o Brasil, assim como assim como os países cientificamente emergentes, apresenta predominância de colaboração doméstica frente às colaborações internacionais.

⁵ A área “outros” é composta dos cursos multidisciplinares Biotecnologia, Ciências Ambientais, Ensino, Interdisciplinar e Materiais.

Figura 3- Número de artigos científicos produzidos em colaboração estrangeira no Brasil no período 2000-2017



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

As colaborações internas sobre o total de colaborações científicas foram de aproximadamente 94%, ou seja, quase a totalidade de artigos publicados com esforços colaborativos são de pesquisadores internos. O baixo percentual de internacionalização da colaboração científica é usual e pode ser reflexo de fatores culturais, como barreiras linguísticas, e geográficos para esse tipo de colaboração (Sidone, Haddad e Mena-Chalco, 2016). Outro aspecto relevante a ser destacado na colaboração internacional são as parcerias com países maduros cientificamente, como Estados Unidos e países europeus, uma vez que estas parcerias tendem a receber maior visibilidade mundial (Barjak e Robinson, 2008; Sud e Thelwall, 2016).

O número de supervisões concluídas pelos docentes é apresentado na Tabela 5. A análise evidencia características importantes do sistema científico brasileiro. Primeiramente, nota-se ampliação do número de supervisões em todos os níveis de titulação ao longo de todos os triênios, reflexo das políticas públicas da década de 2000 voltadas ao ensino superior, com maior quantidade de bolsas de auxílio, expansão de vagas na graduação e pós-graduação⁶, bem como o aumento da oferta de cursos na pós-graduação, em que nos últimos anos, cada vez mais alunos têm feito a transição direta graduação/pós-graduação (Pelegriani e França, 2020). Observa-se ainda a maior expansão das supervisões de pós-doutorado, cujo crescimento no período foi da ordem de 1.540%. Em geral, o aumento das supervisões pode ainda estar associado à sobreposição da atividade docente junto à pesquisa e, por conseguinte, à produtividade do pesquisador.

⁶ No período de abrangência desse estudo, o número de programas de pós-graduação no Brasil apresentou crescimento de 148%, e o volume de bolsas de estudo da pós-graduação cresceu 394% (Geocapes, 2020).

Tabela 5 - Supervisões concluídas pelos docentes por triênio no período 2000-2017

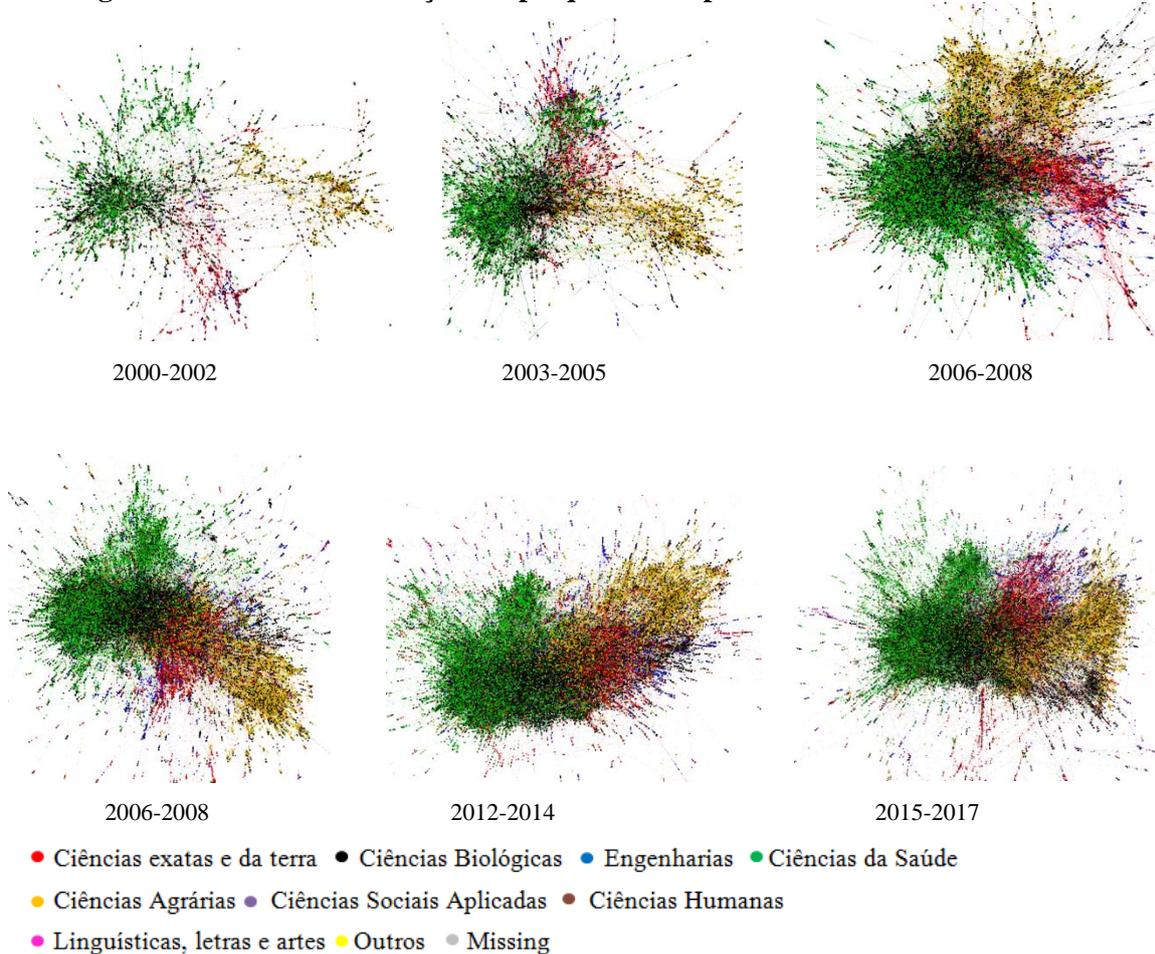
Orientações	Triênio					
	2000-2002	2003-2005	2006-2008	2009-2011	2012-2014	2015-2017
Monografia	203.288	511.747	786.438	910.827	946.064	1.009.310
Dissertação	80.966	111.809	136.123	167.151	206.032	230.323
Tese	24.106	35.587	44.097	53.665	71.326	75.406
Pós-doutorado	1.056	2.369	5.890	9.852	14.758	17.319

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

4. Análise de redes

A figura 4 apresenta as redes de colaborações entre pesquisadores a partir das publicações ocorridas entre 2000 e 2017 pertencentes à amostra de 1% dos pesquisadores das diferentes áreas do conhecimento. Os pesquisadores são representados por nós, coloridos segundo a área do conhecimento; as coautorias são representadas por arestas. As redes são criadas considerando publicações produzidas nos triênios entre 2000 e 2017. Pesquisadores sem colaborações não são mostrados. Como pode ser observado, as colaborações são intensificadas ao longo dos triênios, com ampliação da rede em tamanho. Pesquisadores de ciências da saúde aparecem em maior expressividade na rede pois essa área figura com maior número de artigos produzidos em colaboração.

Figura 4 - Rede de colaborações de pesquisadores por diferentes áreas de conhecimento.



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

As interações entre as principais áreas do conhecimento também são mostradas na figura 4. Um ponto de destaque são as colaborações de pesquisadores das ciências biológicas e ciências da saúde, com integração de ambas em todos os triênios. Há ainda acentuada colaboração entre as ciências exatas e da terra e ciências biológicas. É provável que essas conexões façam parte grupos de pesquisa multidisciplinares. Nota-se ainda, que a área de ciências biológicas posiciona-se no meio da rede, assim como as ciências agrárias, apresentando alto grau de colaboração quando comparados com as demais grandes áreas do conhecimento, corroborando os achados de Mena-Chalco *et al.* (2014).

A Tabela 6 apresenta as métricas da rede de colaborações científicas. As áreas do conhecimento com maiores colaborações são as áreas de ciências biológicas, ciências agrárias e ciências da saúde. A tendência de agrupamento é medida pelo coeficiente de *clustering*. Pesquisadores da biologia, seguido por ciências agrárias e engenharias formam agrupamentos com maior frequência. Cada rede de coautoria reflete as características inerentes a cada grande área do conhecimento brasileira (Mena-Chalco *et al.*, 2014). Pesquisadores associados à linguística, letras e artes apresentam poucos coautores ao publicar seus artigos. Por outro lado, pesquisadores associados às ciências biológicas e às agrárias geralmente tendem a ter vários coautores em suas publicações.

Tabela 6 – métricas da rede de colaborações

Métrica de rede	Área do conhecimento								
	EXA	BIO	ENG	SAUDE	AGR	CSA	CH	LLA	Outros
Coeficiente Clustering	0,35	0,41	0,35	0,35	0,38	0,14	0,14	0,7	0,33
Média centralidade de grau	11,50	20,24	8,06	13,36	16,19	2,18	2,15	1,01	5,15

Fonte: Elaboração própria a partir dos dados do Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq.

A Tabela 7 mostra as métricas das redes colaborativas por triênios. Observa-se elevação na média de autores em um mesmo artigo ao longo do tempo. Nota-se evolução das métricas de rede no tempo, com aumento do coeficiente de *cluster*, sugerindo maior agrupamento de pesquisadores e além de uma rede mais densa e também a presença de ligações mais fortes indicado pela centralidade de grau ponderada.

Tabela 7 – Métricas da rede de colaborações por triênios (2000-2017)

Informações da rede	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	3,12	3,86	4,77	6,08	5,18	7,77
Grau médio 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	70,22	63,35	54,06	46,07	36,15	29,26
Coeficiente de cluster médio	0,60	0,62	0,645	0,658	0,631	0,657

Fonte: Elaboração própria a partir dos dados do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

5. Discussão dos resultados econométricos

Essa seção apresenta a análise dos resultados dos modelos poisson, binomial negativo e binomial negativo logit hurdle sobre a produtividade dos pesquisadores com título de mestrado e doutorado. Tendo em vista que o modelo binomial negativo logit hurdle apresentou menor valor para o critério de informação de Akaike, a interpretação dos resultados se concentra no mesmo. De modo geral, espera-se que indivíduos com maior centralidade na rede de colaborações científicas sejam mais

produtivos em quantidade de artigos científicos publicados (Bordons *et al.*, 2015; Souza, Amorim e Rêgo, 2017). Os resultados são apresentados na tabela 8.

Como pode ser notado, a atividade docente possui relação negativa com a produtividade, sugerindo que a sobreposição das atividades de pesquisa e docência podem comprometer o desempenho de professores em termos de publicação, *ceteris paribus*. A parte de contagem não zero do modelo hurdle binomial positiva e estatisticamente significativa demonstra que os professores possuem menor probabilidade de publicarem artigos em relação aos pesquisadores que não são professores. Segundo Fox (1992), atividades de ensino estão associadas à menor produtividade em pesquisa, e se o objetivo for aumentar a produtividade do pesquisador, deve ocorrer incentivo às atividades não relacionadas ao ensino.

Tabela 8 – Modelos de contagem para produtividade científica

Variável dependente: <i>N_{public}</i>	Poisson	Binomial Negativo	Binomial Negativo Hurdle	
Variáveis explicativas	Estimativa	Estimativa	Contagem	Zero
Professor	-0,159*** (0,002)	-0,168*** (0,006)	-0,199*** (0,007)	1,247 (1,054)
Tempo de docencia	0,004*** (0,000)	0,003*** (0,000)	0,004*** (0,000)	0,006 (0,004)
Tempo de docencia ²	-0,000*** (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000 (0,000)
Tese	0,032*** (0,000)	0,086*** (0,000)	0,088*** (0,000)	-0,202 (0,165)
<i>Lnbetweennesscentrality</i>	0,178*** (0,000)	0,153*** (0,001)	0,168*** (0,000)	-0,085 (0,040)
<i>Closeness</i>	0,379*** (0,014)	0,624*** (0,063)	0,656*** (0,074)	-103,312 (333232,52)
Unipub	-0,065*** (0,001)	0,014*** (0,003)	0,022*** (0,004)	0,962** (0,451)
Area do conhecimento				
Exatas	Omitida	Omitida	Omitida	Omitida
Biológicas	0,012*** (0,001)	-0,035*** (0,006)	-0,036*** (0,007)	-0,205 (0,027)
Engenharias	-0,180*** (0,002)	-0,175*** (0,007)	-0,197*** (0,008)	-0,134*** (0,025)
Saúde	0,169*** (0,001)	0,089*** (0,005)	0,091*** (0,006)	0,012*** (0,023)
Agrárias	0,135*** (0,001)	0,101*** (0,006)	0,112*** (0,007)	-0,050 (0,031)
Ciências Sociais	0,013*** (0,002)	0,139*** (0,005)	0,169*** (0,007)	-0,743*** (0,023)
Ciências Humanas	-0,034*** (0,002)	0,107*** (0,006)	0,136*** (0,006)	-0,887 (0,021)
Linguística e Letras	0,138*** (0,003)	0,217*** (0,008)	0,264*** (0,010)	-0,802 (0,027)
Outros	-0,213*** (0,007)	-0,162*** (0,019)	-0,193*** (0,022)	-0,414 (0,075)
Constante	1,374*** (0,002)	1,412*** (0,006)	1,208*** (0,007)	-10,436*** (0,024)
Observações	332.960	332.960	332.960	332.960
AIC	4905840	2230289	2177926	2177926
Ln Alpha (a)	-	-0,539***	-0,187***	-

Fonte: Elaboração própria a partir dos dados do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

Quanto maior o tempo transcorrido de docência maior é a produtividade do pesquisador, no entanto essa produtividade alcança um pico em determinado ponto da carreira docente e passa a se tornar decrescente, mostrando menor quantidade de publicações (contagem) para o termo quadrático do tempo de docência, o que pode indicar menos incentivos à produção científica associados à progressão na carreira docente.

Em relação ao número de teses orientadas, variável que combina elementos de ensino e pesquisa, os resultados demonstram resultados estatisticamente significativos em todos os modelos, controlando os demais preditores, sugerindo que supervisionar alunos na pós-graduação pode ser favorável à produção científica. Conforme Horta, Veloso e Grediaga (2010) salientam, o vínculo de pesquisa formado entre os pesquisadores e seus orientandos durante o doutorado é muito forte e tende a permanecer durante a carreira acadêmica, sustentando maior produtividade de ambas as partes.

Vale lembrar, no entanto, que o conceito de produtividade empregado é definido como o número de artigos publicados e nestes há, por trás, a quantidade de orientandos que o professor possui. Assim, pesquisadores com mais orientandos geralmente possuirão mais artigos com o nome do pesquisador aparecendo como coautor.

As medidas de rede de colaborações, entrelaçamento (*Betweenness*) e proximidade dos agentes na rede (*closeness*) apresentam relação positiva com a produtividade científica, em linha com a literatura (Abramo, D'Angelo e Murgia, 2017; Souza, Amorim e Rêgo, 2017). Esses resultados podem ser explicados levando-se em consideração a transmissão mais fácil do conhecimento tácito e a criação mais simples de um clima de confiança através das conexões entre pesquisadores, e sugere ainda a capacidade geral de pesquisadores para aproveitar os ganhos potenciais da colaboração, como melhor visibilidade e acesso a recursos e experiências complementares

Cumprir lembrar que tais colaborações científicas não respondem apenas à necessidade dos pesquisadores de aumentar sua produtividade, mas também ao prestígio decorrente da interação com seus pares, facilitando acesso a participações em bancas e projetos de pesquisa, os quais também contribuem para ascensão profissional e produtividade.

Há uma diferença no nível de publicação de acordo com o status de vínculo com universidades públicas, em que a parte de contagem (não zero) demonstra que os pesquisadores de universidades públicas são mais propensos a produzir mais de um artigo. Esse resultado é coerente com a hipótese de que as universidades públicas configuram o cerne da produção científica brasileira. De acordo com trabalhos recentes (Sidone *et al.*, 2016; Cross, Thomson e Sinclair, 2019), os pesquisadores de universidades públicas têm uma maior propensão à colaboração em relação aos pares das demais universidades e tendem a publicar mais artigos em parceria, o que aumenta o número de publicações.

Além disso, esse resultado pode estar associado ao fato de que muitos periódicos nacionais não possuem destaque nas grandes bases de dados indexadas internacionais e apresentam um fator de impacto muito baixo, sendo necessária a inclusão do fator de impacto das revistas para verificar a relevância dessas diferenças. De maneira geral, instituições privadas estão voltadas para uma formação profissional direcionada ao mercado e possuem docentes cujas atividades se concentram no ensino, enquanto a qualidade de seus cursos é comparativamente menor do que em universidades públicas (Pelegrini e França, 2020).

6. Conclusões

Para atingir excelência no desempenho científico, um dos caminhos é a colaboração acadêmica, uma vez que estimula a produtividade e a qualidade das publicações. O presente estudo contribui para esclarecer ainda mais esse tópico por meio de uma análise de uma amostra de mestres e doutores brasileiros com currículos cadastrados na plataforma Cnpq/Lattes. A partir de um modelo binomial negativo, analisamos a relação entre colaboração científica e produtividade desses pesquisadores entre 2000 e 2017, empregando medidas de redes sociais. Embora a literatura tenha investigado profundamente essa relação ao longo dos anos, apenas alguns artigos brasileiros empregam medidas de rede para doutores brasileiros de todas as áreas. A análise dessa relação para grupos de pesquisadores

mais produtivos, como os doutores e mestres, permite uma melhor compreensão dos processos de transferência de capital humano entre cientistas.

Em particular, nosso trabalho contribui para o preenchimento de uma lacuna na literatura sobre determinantes da produtividade, mostrando que a alta produtividade, de doutores, recebe influência direta das características das redes de colaborações científicas. Este fato pode ser explicado pela habilidade que os cientistas com maior titulação possuem para gerenciar suas redes, extraindo delas maior efetividade em termos de publicações.

Em conformidade com a literatura, mostramos que as centralidades de rede têm impacto positivo na produtividade, assim como o número de orientações e o vínculo com universidades públicas. Nesse sentido, será útil estender a análise a outros grupos de pesquisadores com diferentes perfis de vínculos institucionais a fim de compreender se esse resultado pode ser generalizado ou se está relacionado especificamente a mestres e doutores e universidades públicas, por exemplo.

Outro resultado que deve ser confirmado por análises semelhantes em outros recortes amostrais contemplando diferentes áreas do conhecimento é o que diz respeito à relação entre tempo de atividade docente e produtividade acadêmica. Embora seu impacto na colaboração esteja alinhado com a literatura, o fraco impacto poderia ser explicado como o resultado de um mecanismo de contratação e promoção não adequadamente vinculado à produtividade ou de uma perda de motivação após a nomeação para uma progressão acadêmica associada ao tempo de docência e não à produtividade.

Além de um retorno à literatura sobre relação entre colaboração científica e produtividade, esse estudo oferece alguns *insights* úteis para o desenvolvimento de políticas relacionadas a essas questões. Em primeiro lugar, esses resultados confirmam que a definição de uma política destinada a maximizar a produtividade da pesquisa deve levar em conta a questão dos diferentes tipos de colaboração, avaliando seus benefícios. Em segundo lugar, os resultados sugerem a necessidade de políticas que reforcem a produtividade de maneira efetiva, sobretudo por parte de acadêmicos com maior tempo de carreira.

Um dos principais limites da literatura sobre colaborações científicas é dado pela impossibilidade de identificar todos os tipos de colaboração que geraram uma publicação. Nesse sentido, em trabalhos futuros, poderíamos ampliar nosso conjunto de dados incluindo informações de outras formas de colaborações científicas, como publicação de livros, artigos apresentados em congressos e a qualidade das publicações. Além disso, poderíamos tentar ampliar nosso conjunto de dados com informações das universidades às quais estão vinculados os pesquisadores, por exemplo, infraestrutura, volume de bolsas de pesquisa, número de alunos, o que poderia influenciar tanto a produtividade quanto a colaboração dos indivíduos. Pesquisas futuras podem ainda concentrar esforços no impacto da colaboração entre acadêmicos de diferentes nacionalidades, as quais vem crescendo no Brasil ao longo dos últimos anos.

Scientific collaboration and productivity: an analysis of social networks in the period 2000-2017

Abstract: The Brazilian scientific community has grown in production in recent years. This increase was accompanied by an increase in the collaborations of its researchers. This study investigates the role of co-authoring networks of scientific articles in high scientific productivity in Brazil between 2000 and 2017. The data source is the Lattes / Cnpq platform, used to evaluate scientific production and to build networks of collaborations between researchers. Counting models are applied to explore how scientific performance is related to different co-authoring network measures. The results suggest a preponderant role of networks of scientific collaboration in scientific production. Researchers with greater centrality in the network of scientific collaborations are more productive, while teaching time is associated with lower productivity. The supervision of theses and the link with public universities can increase the publication of researchers.

Keywords: Research productivity, scientific collaboration, social network, Lattes Platform.

Referências

- ABRAMO, G.; D'ANGELO, C. A.; DI COSTA, F. Research collaboration and productivity: is there correlation?. **Higher education**, v. 57, n. 2, p. 155-171, 2009.
- ABRAMO, G.; D'ANGELO, A. C.; MURGIA, G. The relationship among research productivity, research collaboration, and their determinants. **Journal of Informetrics**, v. 11, n. 4, p. 1016-1030, 2017.
- ALIPOVA, O.; LOVAKOV, A. Academic inbreeding and publication activities of Russian faculty. **Tertiary Education and Management**, v. 24, n. 1, p. 66-82, 2018.
- ALVAREZ, R. B. P.; JÚNIOR, S. K.; CAROLO, M. D. O impacto da interação universidade-empresa na produtividade dos pesquisadores: uma análise para as ciências exatas e da terra nas universidades estaduais paulistas. **Revista Brasileira de Inovação**, v. 12, n. 1, p. 171-206, 2013.
- BARJAK, F.; ROBINSON, S. International collaboration, mobility and team diversity in the life sciences: impact on research performance. **Social geography**, v. 3, n. 1, p. 23-36, 2008.
- BEAVER, D. D. Reflections on scientific collaboration (and its study): past, present, and future', **Scientometrics**, n.52, v.3, p. 365-377, 2001.
- BORDONS, M.; APARÍCIO, J.; GONZÁLEZ-ALBO, B.; DÍAZ-FAES, A. A. The relationship between the research performance of scientists and their position in co-authorship networks in three fields. **Journal of Informetrics**, v.9, n. 1, p. 135-144, 2015.
- BOSCHMA, R. Proximity and innovation: a critical assessment. **Regional studies**, v. 39, n.1, p. 61-74, 2005.
- CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics using stata**. College Station, TX: Stata press, 2010.
- CROSS, D.; THOMSON, S.; SINCLAIR, A. **Research in Brazil**. A report for CAPES by Clarivate Analytics.
- DIAS, T. M. R.; MOITA, G. F.; DIAS, P. M. Adoção da plataforma lattes como fonte de dados para caracterização de redes científicas. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 21, n. 47, p. 16-26, 2016.
- DURIEUX, V.; GEVENOIS, P. A. Bibliometric Indicators: Quality Measurements Of Scientific Publication. **Radiology**, v. 255, n.2, p. 342-351, 2010.
- EGGHE, L.; GOOVAERTS, M.; KRETSCHMER, H. Collaboration and Productivity: an investigation into 'Scientometrics' journal and 'UHasselt'repository. **COLLNET Journal of Scientometric and Information Management**, v. 2, n. 1, p. 83-89, 2008.
- FOX, M. F. Research, teaching, and publication productivity: mutuality versus competition in Academia. **Sociology of Education**, v. 65, n. 4, p. 293-305, Out. 1992.
- GLÄNZEL, W.; LETA, J.; THIJS, B. Science in Brazil. Part 1: A macro-level comparative study. **Scientometrics**, v. 67, n. 1, p. 67-86, 2006.
- GLÄNZEL, W.; SCHUBERT, A.: Double Effort = Double Impact? A Critical View at International Co-authorship in Chemistry, **Scientometrics**, 50, 2, 199-214, 2001.
- HADDAD, E. A.; MENA-CHALCO, J. P.; SIDONE, O. Produção científica e redes de colaboração dos docentes vinculados aos programas de pós-graduação em Economia no Brasil. **Estudos Econômicos**, v. 47, n. 4, p. 617-679, 2017.
- HE, B.; DING, Y.; NI, C. Mining enriched contextual information of scientific collaboration: A meso perspective. **Journal of the American Society for Information Science and Technology**, v. 62, n. 5, p. 831-845, 2011.
- HILBE, J. M. **Negative binomial regression**. Cambridge University Press, 2011.
- HILL, R. C.; GRIFFITHS, W. E.; LIM, G. C.; ADKINS, L. C. **Principles of Econometrics**, NJ: Wiley Hoboken, 2008.
- HORTA, H.; VELOSO, F.; GREDEIAGA, R. Navel gazing: Academic inbreeding and scientific productivity. **Management Science**, v. 56, n. 3: p. 414-429, 2010.
- IGLIĆ, H et al. With whom do researchers collaborate and why?. **Scientometrics**, v. 112, n. 1, p. 153-174, 2017.
- IOANNIDIS, J. P.A.; BOYACK, K. W.; BAAS, J. Updated science-wide author databases of standardized citation indicators. **PLoS Biology**, v. 18, n. 10, p. e3000918, 2020.
- KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research policy**, v. 26, n. 1, p. 1-18, 1997.
- LANDRY, R.; TRAORE, N.; GODIN, B. An econometric analysis of the effect of collaboration on academic research productivity. **Higher Education**, v. 32, n. 3, p. 283-301, 1996.

- LARIVIÈRE, V.; ARCHAMBAULT, É.; GINGRAS, Y.; VIGNOLA-GAGNÉ, É. The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. **Journal of the American Society for Information Science and Technology**, v. 57, n. 8, p. 997-1004, 2006.
- LAUDEL, G. What do we measure by co-authorships?. **Research evaluation**, v. 11, n. 1, p. 3-15, 2002.
- LETA, J.; CHAIMOVICH, H. Recognition and international collaboration: The Brazilian case. **Scientometrics**, v.53, n.3, p.325-335, 2002.
- LETA, J.; THIJS, B.; GLÄNZEL, W. A macro level study of science in Brazil: seven years later. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 18, n. 36, p. 51-66, 2013.
- LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington Academy of Sciences**, v.16, n.12, p.317-323, 1926.
- MALI, F.; KRONEGGER, L.; DOREIAN, P.; FERLIGOJ, A. Dynamic scientific co-authorship networks. In **Models of science dynamics**, p. 195-232, Springer, Berlin, Heidelberg, 2012.
- 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, n.7, p. 1424–1445, 2014.
- MINGERS, J.; LEYDESDORFF, L. A review of theory and practice in scientometrics. **European journal of operational research**, v. 246, nº.1, p. 1-19, 2015.
- MOODY, J. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. **American sociological review**, v. 69, n. 2, p. 213-238, 2004.
- PONDS, R.; OORT, F.V.; FRENKEN, K. Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach". **Journal of Economic Geography**, v. 10, no.2: 231-255, 2009.
- RIGBY, J.; EDLER, J. Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. **Research policy**, v. 34, n. 6, p. 784-794, 2005.
- ROYAL SOCIETY. **Knowledge, networks and nations: Global scientific collaboration in the 21st century**. London: The Royal Society, 2011.
- 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**, Campinas, v. 28, n. 1, p. 15-32, Apr. 2016.
- SOORYAMOORTHY, R. Collaboration and publication: How collaborative are scientists in South Africa?. **Scientometrics**, v. 80, n. 2, p. 419-439, 2009.
- SOUZA, F. C. D; AMORIM, R. M.; RÉGO, L. C. A Co-authorship network analysis of CNPq's productivity research fellows in the probability and statistic area. **Perspectivas em Ciência da Informação**, v. 21, n. 4, p. 29-47, 2016.
- SUD, P.; THELWALL, M. Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. **Journal of the Association for Information Science and Technology**, v. 67, n. 8, 1849–1857, 2016.
- YNALVEZ, M. A.; SHRUM, W. M. Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. **Research Policy**, v. 40, n. 2, p. 204-216, 2011.