

**UM ESTUDO SOBRE AS COLABORAÇÕES
INTERDISCIPLINARES NA COMUNIDADE
CIENTÍFICA BRASILEIRA**

GERALDO JOSÉ PESSOA JUNIOR

**UM ESTUDO SOBRE AS COLABORAÇÕES
INTERDISCIPLINARES NA COMUNIDADE
CIENTÍFICA BRASILEIRA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ALBERTO HENRIQUE FRADE LAENDER

Belo Horizonte

Julho de 2019

© 2019, Geraldo José Pessoa Júnior.
Todos os direitos reservados

Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG

Pessoa Júnior, Geraldo José.

P475e Um estudo sobre as colaborações interdisciplinares
na comunidade científica brasileira / Geraldo José
Pessoa Júnior — Belo Horizonte, 2019.
xxii, 51 p. il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas
Gerais – Departamento de Ciência da Computação.

Orientador: Alberto Henrique Frade Laender

1. Computação – Teses. 2. Abordagem interdisciplinar do
conhecimento. 3. Bibliometria. 4. Coautoria. 5. Publicações
científicas. I. Orientador. II. Título.

CDU 519.6*(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

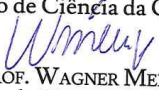
FOLHA DE APROVAÇÃO

UM ESTUDO SOBRE AS COLABORAÇÕES INTERDISCIPLINARES
NA COMUNIDADE CIENTÍFICA BRASILEIRA

GERALDO JOSÉ PESSOA JUNIOR

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:


PROF. ALBERTO HENRIQUE FRAIDE LAENDER - Orientador
Departamento de Ciência da Computação - UFMG


PROF. WAGNER MEIRA JÚNIOR
Departamento de Ciência da Computação - UFMG


PROFA. MIRELLA MOURA MORO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 31 de Julho de 2019.

Aos meus pais, Geraldo e Almira, que já se foram, mas continuam sendo meu exemplo e fonte de força para lutar pelos meus sonhos de vida.

Agradecimentos

Em primeiro lugar, agradeço a Deus que me presenteou com o dom da vida e me iluminou durante toda esta caminhada. Igualmente agradeço aos meus pais (in memoriam), Geraldo e Almira, que sempre me ensinaram a lutar pelos meus sonhos e, mesmo longe, sei que de alguma forma sempre estarão comigo.

Agradeço também ao meu orientador, Professor Alberto Henrique Frade Laender, por toda a paciência e empenho com que me orientou durante o desenvolvimento deste trabalho. Estendo os agradecimentos a todos os professores que fizeram parte de minha vida acadêmica.

À Universidade Federal de Viçosa pela licença capacitação, me possibilitando aperfeiçoar meus conhecimentos técnicos em prol de um melhor desenvolvimento das minhas atividades na instituição.

Por último, agradeço a todos os meus familiares e amigos (que são a família que escolhi), que sempre estiveram do meu lado, me incentivando e apoiando a iniciar a pós-graduação, e a me manter firme nos momentos em que ousei duvidar da minha capacidade, bem como aos novos amigos reconhecidos durante os dois anos que morei em Belo Horizonte, a principal razão de me sentir em casa nesse novo ambiente. E à Priscilla, que nessa reta final foi minha companheira, dividindo os meus anseios e multiplicando as minhas alegrias.

*“Com grandes poderes, vêm
grandes responsabilidades”*
(Stan Lee)

Resumo

Colaborações interdisciplinares têm recentemente chamado a atenção de estudiosos, uma vez que tais relações acadêmicas contribuem para fortalecer a comunidade científica. No entanto, estudos anteriores se concentraram em caracterizar grupos específicos, em vez de estudar uma comunidade científica completa e robusta. Nesta dissertação, em vez de analisar cenários particulares, caracterizamos essas colaborações com relação às oito comunidades científicas brasileiras definidas de acordo com o nível superior do esquema de classificação de áreas do conhecimento do CNPq. Os resultados mostram que a rede brasileira de colaboração cresceu e se tornou especialmente interdisciplinar. Em particular, de acordo com nossa análise, 35,2% de todas as colaborações nessa rede são interdisciplinares e 57,6% dos pesquisadores têm pelo menos uma interação científica deste tipo. Além disso, nosso estudo também explora aspectos geográficos e temporais, identificando padrões distintos de colaboração envolvendo as comunidades científicas mencionadas.

Palavras-chave: Colaborações Interdisciplinares, Redes de Colaboração, Comunidades Científicas, Plataforma Lattes.

Abstract

Interdisciplinary collaborations have recently drawn the attention of scholars, since such relationships contribute to make scientific communities stronger. However, previous studies have focused on characterizing specific groups rather than on studying a complete and robust scientific community. In this MSc dissertation, instead of analyzing particular scenarios, we characterize these collaborations with respect to the eight Brazilian scientific communities defined according to the upper level of the CNPq knowledge area classification scheme. The results show that the Brazilian collaboration network grew and became especially interdisciplinary. In particular, according to our analysis 35.2% of all collaborations in this network are interdisciplinary and 57.6% of the researchers have at least one such scientific interaction. In addition, our study also explores geographical and temporal aspects, thus identifying distinct collaboration patterns involving the aforementioned scientific communities.

Keywords: Interdisciplinary Collaborations, Collaboration Networks, Scientific Communities, Lattes Platform.

Lista de Figuras

2.1	Trecho de um Currículo Lattes	10
2.2	Trecho de um Currículo Lattes no Formato XML	11
2.3	Trecho de um arquivo CSV derivado de um Currículo Lattes	12
2.4	Exemplo de uma Rede Complexa	15
2.5	Grau de um vértice	16
2.6	Exemplo de um Componente Gigante	16
2.7	Componentes Isolados	17
2.8	Caminho Mínimo e Diâmetro	18
3.1	Redes Global e Interdisciplinar de Colaboração Científica no Brasil	20
3.2	Percentual de Colaborações Interdisciplinares por Grande Área	23
3.3	Taxa de Colaboração Interdisciplinar por Grande Área	25
3.4	Regiões Geográficas Brasileiras.	28
3.5	Distribuição das Colaborações Interdisciplinares por Região	29
3.6	Padrões de Colaboração entre Regiões Brasileiras	30
3.7	Padrões de Colaboração entre as Grandes Áreas por Região	32
4.1	Distribuição dos Pesquisadores por Tipo de Colaboração e Idade Acadêmica	36
4.2	Evolução Temporal das Grandes Áreas com Taxas Baixas de Colaboração Interdisciplinar	38
4.3	Evolução Temporal das Grandes Áreas com Taxas Médias de Colaboração Interdisciplinar	40
4.4	Evolução Temporal das Grandes Áreas com Taxas Altas de Colaboração Interdisciplinar	41

Lista de Tabelas

2.1	Exemplos de Chaves Hash para Identificação de Colaborações	13
3.1	Distribuição dos Pesquisadores por Grande Área	19
3.2	Comparação entre as Redes Gobal e Interdisciplinar	21
3.3	Publicações por Grande Área	24
3.4	Dez Principais Padrões de Colaboração Interdisciplinar entre Regiões Bra- sileiras	31
4.1	Distribuição dos Pesquisadores por Idade Acadêmica	35

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Objetivos	2
1.2 Motivação e Questões de Pesquisa	3
1.3 Trabalhos Relacionados	4
1.4 Contribuições	7
1.5 Organização da Dissertação	8
2 Metodologia	9
2.1 Definição do Esquema de Classificação	9
2.2 Seleção dos Pesquisadores	9
2.3 Coleta e Preparação dos Dados	10
2.4 Identificação das Colaborações	12
2.4.1 Redes Complexas: Conceitos Básicos	13
3 Análise das Redes de Colaboração	19
3.1 Colaborações Interdisciplinares por Grande Área	22
3.2 Colaborações Interdisciplinares por Região Geográfica	27
4 Análises Adicionais	35
4.1 Influência da Idade Acadêmica nas Colaborações Interdisciplinares	35

4.2	Análise Temporal das Colaborações Interdisciplinares	37
4.2.1	Grandes Áreas com Taxas Baixas de Colaboração Interdisciplinar	37
4.2.2	Grandes Áreas com Taxas Médias de Colaboração Interdisciplinar	39
4.2.3	Grandes Áreas com Taxas Altas de Colaboração Interdisciplinar	40
4.3	Considerações Finais	42
5	Conclusões e Trabalhos Futuros	45
5.1	Conclusões	45
5.2	Trabalhos Futuros	46
	Referências Bibliográficas	47

Capítulo 1

Introdução

Estudos recentes têm abordado o papel importante das colaborações interdisciplinares nas redes de coautoria científica [Abramo et al., 2018; Iglič et al., 2017; Lopes et al., 2012; Mena-Chalco et al., 2014; Shi et al., 2018], uma vez que tais relações contribuem para fortalecer a compreensão da evolução da ciência. De fato, isso chamou a atenção dos estudiosos e os esforços anteriores concentraram-se em colaborações externas (por exemplo, cooperação entre grupos de pesquisa [Freire & Figueiredo, 2011; Silva et al., 2014], migração de pesquisadores [Kato & Ando, 2013] e influência de áreas de pesquisa distintas [Lima et al., 2013]) como fatores importantes na evolução das comunidades científicas. No entanto, tais estudos têm focado principalmente na caracterização de áreas ou grupos de pesquisa específicos [Kato & Ando, 2013; Silva et al., 2014].

Assim, ao invés de analisarmos cenários específicos (por exemplo, um departamento universitário ou acadêmico, uma área de pesquisa ou um grupo de pesquisa), avançamos um passo ao construir as redes de colaboração científica global e interdisciplinar de todo o Brasil, utilizando para isso uma fonte de dados específica, a Plataforma Lattes¹. Mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), essa plataforma é uma iniciativa de renome internacional [Lane, 2010] que fornece um repositório de currículos de pesquisadores e grupos de pesquisa integrados em um único sistema. Como todos os pesquisadores no Brasil (juniores e sêniores) são obrigados a manter seus currículos atualizados nessa plataforma, ela fornece uma grande quantidade de informação sobre suas atividades de pesquisa e produção científica.

Nesta dissertação, analisamos as colaborações interdisciplinares da comunidade científica brasileira com base em um esquema específico de classificação de áreas de pesquisa. Particularmente, utilizamos o nível superior do esquema de classificação de

¹Plataforma Lattes: <http://lattes.cnpq.br>

áreas do conhecimento do CNPq, que está organizado em quatro níveis²: *grande área* (por exemplo, Ciências Exatas e da Terra), *área* (por exemplo, Ciências da Computação), *subárea* (por exemplo, Teoria da Computação) e *Especialidade* (por exemplo, Linguagens Formal e Autômatos) [de Siqueira et al., 2017]. Entretanto, nossa análise considera somente o primeiro nível desse esquema de classificação, que divide a comunidade acadêmica brasileira nas seguintes oito grandes áreas de pesquisa: *Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Ciências Humanas, Ciências Sociais e Aplicadas, Engenharias, e Linguística, Letras e Artes*.

Para a nossa análise, construímos as redes de colaboração científica brasileira (global e interdisciplinar) a partir dos currículos de todos os pesquisadores com doutorado cadastrados na Plataforma Lattes (um total de 263.264 currículos coletados em abril de 2018). Como veremos, 35,2% de todas as colaborações dessa rede são interdisciplinares, ou seja, envolvem pesquisadores de distintas grandes áreas, e 57,6% dos pesquisadores têm pelo menos uma interação científica desse tipo, o que enfatiza a importância da interdisciplinaridade nas colaborações científicas.

1.1 Objetivos

Conforme mencionado anteriormente, vários estudos têm focado apenas no aspecto global da produção científica de um país ou grupo de pesquisa [Araújo et al., 2014; Boaventura et al., 2014; de Farias et al., 2012; de Souza Vanz & Stumpf, 2012; Mena-Chalco & Cesar Junior, 2009; Mena-Chalco et al., 2014; Dias, 2016], dando a mesma importância para as arestas que conectam partes específicas de uma rede de colaboração. Em contraste, este trabalho enfatiza a importância de minerar mais profundamente padrões interdisciplinares para caracterizar a força dessas ligações internas, isto é, possibilitar uma melhor compreensão da evolução das colaborações científicas [Freire & Figueiredo, 2011; Kato & Ando, 2013; Mooney et al., 2013; Silva et al., 2014]. Neste contexto, o objetivo desta dissertação é analisar a interdisciplinaridade nas redes de colaboração científica e como isso influencia a produção científica de modo geral, considerando como escopo a comunidade científica brasileira.

Para a construção e análise da rede de colaboração científica brasileira, o trabalho apresentado nesta dissertação foi dividido nas seguintes etapas:

- Seleção do conjunto de pesquisadores objeto deste estudo;
- Coleta na Plataforma Lattes dos currículos dos pesquisadores selecionados;

²Para mais detalhes: <http://lattes.cnpq.br/web/dgp/arvore-do-conhecimento>

- Mineração e tratamento dos dados dos currículo coletados;
- Identificação das colaborações entre os pesquisadores por meio da análise dos dados tratados;
- Construção das redes de colaboração científica, global e interdisciplinar, com base nas coautorias entre os pesquisadores;
- Análise das redes de colaboração científica construídas;
- Geração de visualizações gráficas e tabulares dos principais resultados encontrados durante a análise das redes de colaboração científica;
- Avaliação do impacto das colaborações interdisciplinares na produção científica brasileira.

1.2 Motivação e Questões de Pesquisa

Atualmente, a Internet tem contribuído de forma significativa para facilitar o acesso de seus usuários à produção científica, por meio das bibliotecas digitais, redes de relacionamentos e repositórios bibliográficos, contribuindo assim para a expansão do conhecimento [Dias, 2016]. Além da procura por produções científicas específicas, um pesquisador pode utilizar a Internet para encontrar outros pesquisadores que possam colaborar em suas pesquisas [Digiampietri et al., 2012; Ströele et al., 2012].

Colaborações científicas podem ser descritas como interações entre cientistas com o intuito de compartilhar atividades para alcançar objetivos comuns [Lane, 2010; Sonnenwald, 2007]. Nesse contexto, análises baseadas na teoria de redes complexas têm sido realizadas em redes de colaboração científica para entender o significado de tais interações nos mais diferentes cenários [de Souza Vanz & Stumpf, 2012; Freire & Figueiredo, 2011; Hua & Haughton, 2012; Mena-Chalco et al., 2014; Mooney et al., 2013; Silva et al., 2014].

A principal fonte de dados utilizada para análises das colaborações científicas no contexto brasileiro é a Plataforma Lattes. Uma vez que todos os pesquisadores que atuam no Brasil são obrigados a manter seus currículos atualizados nessa plataforma, ela fornece um grande volume de dados sobre as atividades de pesquisa e a produção científica desses indivíduos. Desse modo, minerar os padrões de colaboração científica existentes permite explorar o ambiente envolvido, possibilitando, por exemplo, a extração de relacionamentos complexos e a obtenção de informações sobre a produção científica desses pesquisadores [Mena-Chalco et al., 2014].

Neste contexto, esta dissertação teve como motivação construir as redes de colaboração científica brasileira, global e interdisciplinar, visando analisar a importância da interdisciplinaridade na produção científica brasileira e o padrão de colaboração estabelecido por tais relações nas diferentes comunidades de pesquisa do país, conforme proposto inicialmente [Pessoa Jr. et al., 2018]. Assim, as principais questões de pesquisa a serem respondidas nesta dissertação são:

- Quão relevantes são as colaborações científicas interdisciplinares na comunidade científica brasileira?
- Qual é o padrão de colaboração interdisciplinar entre as oito grandes áreas de pesquisa propostas pelo CNPq?
- Como as características de uma região geográfica brasileira afetam os padrões de colaboração interdisciplinar?
- Como a idade acadêmica influencia as colaborações interdisciplinares de um pesquisador?
- Como a interdisciplinaridade das colaborações variou ao longo do tempo?

1.3 Trabalhos Relacionados

O significativo volume de publicações geradas pelas pesquisas científicas deu origem a inúmeras iniciativas, como a Plataforma Lattes e a DBLP³ (Digital Bibliography & Library Project), que armazenam esses dados de forma organizada, tornando-os amplamente disponíveis a toda comunidade científica. A DBLP é um repositório bibliográfico que tem sido utilizado como fonte de dados para diversos estudos em nível internacional, desde a classificação de especialistas acadêmicos [Moreira et al., 2015] até tópicos mais específicos como a investigação de um fenômeno semelhante ao *paradoxo da amizade* nas redes de colaboração científica [Benevenuto et al., 2016], a desambiguação de nomes de autores [Franzoni et al., 2019; Kim, 2018; Santana et al., 2017] e mesmo o impacto da mobilidade acadêmica na qualidade dos programas de pós-graduação [Silva et al., 2017].

Neste contexto, inúmeros trabalhos também têm sido desenvolvidos utilizando dados sobre as colaborações entre pesquisadores (ou grupos de pesquisa) contidos nesses repositórios. Por exemplo, Freire & Figueiredo [2011] caracterizam e comparam

³dblp: <https://dblp.uni-trier.de/>

as redes de colaboração científica global e brasileira na área de Ciência da Computação, focando na intensidade das relações entre os pesquisadores para identificar e classificar os indivíduos (e grupos) mais influentes. Já Silva et al. [2014] exploram os relacionamentos entre grupos como uma ponte para trazer novas ideias para diferentes contextos, enquanto Kato & Ando [2013] reforçam o benefício da mobilidade internacional dos pesquisadores em seu desempenho. Em um outro trabalho, Boaventura et al. [2014] analisam as redes de colaboração científica de algumas universidades brasileiras (UFAM, UFMG, UFPE, UFRGS, UFRJ, UNB, UNICAMP e USP), de modo a determinar o impacto de suas respectivas produções científicas e o surgimento de novas áreas de pesquisa.

Hua & Haughton [2012] analisam uma rede de interações entre médicos (postagens e comentários) com base em suas especialidades, a fim de investigar como eles se comportam em uma rede social *on-line*, enquanto Mooney et al. [2013] avaliam a evolução das interações entre pesquisadores de ciências naturais e sociais no contexto de programas de pesquisa de mudanças globais. Benevenuto et al. [2016] investigam o *paradoxo da amizade* nas colaborações científicas em diferentes comunidades de ciência da computação e apresentam evidências empíricas da existência de um fenômeno semelhante em tais comunidades, discutindo as suas potenciais consequências.

Assim, a mineração de padrões de colaboração nos permite explorar toda uma comunidade científica para, por exemplo, extrair relacionamentos complexos e informações específicas sobre os seus membros [Hua & Haughton, 2012; Mena-Chalco et al., 2014]. Nesse contexto, Mena-Chalco & Cesar Junior [2009] descrevem o ScriptLattes, um software para extração de dados de currículos Lattes previamente selecionados e que permite a visualização das informações relacionadas na forma de relatórios e gráficos. Mais especificamente, Mena-Chalco et al. [2014] apresentam um estudo sobre as grandes áreas de pesquisa da comunidade científica brasileira, no qual caracterizam as oito grandes áreas aqui abordadas através de métricas topológicas.

Na literatura é possível encontrar estudos sobre a interdisciplinaridade na pesquisa científica. Por exemplo, os resultados obtidos por uma equipe responsável por avaliar um programa de apoio à pesquisa interdisciplinar nos Estados Unidos, o NAKFI (National Academies Keck Futures Initiative)⁴, é apresentado por Porter et al. [2008]. Essa equipe desenvolveu e testou duas métricas quantitativas, a *integração* (dispersão entre as áreas dos periódicos citados por uma pesquisa) e a *especialização* (disseminação de áreas em que a pesquisa é publicada), para avaliar os resultados das pesquisas do programa. Por meio dessas métricas os avaliadores observaram um alto grau de

⁴<https://www.keckfutures.org/>

interdisciplinaridade nas pesquisas desse programa.

Huang & Chang [2011] apresentam um estudo em que investigam as mudanças interdisciplinares na área de ciência da informação entre 1978 e 2007, utilizando a citação direta e as coautorias. Os autores utilizaram o índice de Brillouin [Peet, 1975] para mensurar o grau de interdisciplinaridade nessa área em cada ano do período considerado. Os resultados mostram que o grau de interdisciplinaridade nas citações diretas e nas coautorias apresentou um crescimento ao longo dos anos, sendo a taxa de crescimento das coautorias mais evidente do que a das citações diretas.

Em um estudo recente, Abramo et al. [2018] apresentaram recentemente um estudo sobre como a produção científica pode ser afetada pela multidisciplinaridade das áreas dos integrantes de um grupo de pesquisa. O trabalho analisa as publicações de todos os professores das universidades italianas entre os anos de 2004 e 2008. Assim como no Brasil, os pesquisadores italianos são divididos em oito grandes áreas do conhecimento⁵. Os resultados alcançados mostram que produções científicas específicas são desenvolvidas por pesquisadores de uma mesma área e as diversificadas por grupos multidisciplinares e, além disso, que quanto maior a quantidade de áreas envolvidas maior a diversidade da produção científica.

Um outro estudo avalia se o grau de reputação de uma instituição científica influencia a atuação interdisciplinar no fluxo de conhecimento Shi et al. [2018]. Para mensurar o grau de prestígio de uma instituição, os autores analisaram as produções científicas ali desenvolvidas. Uma métrica que combina a abordagem IDR (Irreproducible Discovery Rate) [Li et al., 2011] com o algoritmo *P-Rank* foi então usada para analisar o grau de importância das produções científicas de uma instituição. Os resultados mostram que instituições mais conceituadas fomentam uma maior interdisciplinaridade no fluxo de conhecimento.

Esses estudos corroboram a ideia da força das colaborações interdisciplinares como pontes importantes que conectam toda uma rede. No entanto, vários estudos tendem a concentrar sua análise apenas na visão global da rede ou mesmo negligenciam partes importantes dela. Da mesma forma que os estudos mencionados acima, esta dissertação considera as relações interdisciplinares como um meio de alavancar o significado social dos relacionamentos (por exemplo, laços fracos [Granovetter, 1973]).

De fato, os relacionamentos externos de um determinado grupo (definido por área de pesquisa, comunidade científica ou qualquer outro critério) podem ser interpretados como sendo uma troca de ideias ou conhecimento entre seus indivíduos. Por exemplo, sociólogos podem discutir conceitos sociais com pesquisadores da Ciência da

⁵Biológia, Ciências da Terra e Espaço, Engenharia, Física, Matemática, Medicina Clínica, Pesquisa Biomédica e Química

Computação para entender porque e como a estrutura de uma rede social é formada. Dessa forma, a colaboração entre eles pode ser extremamente benéfica para a ciência, uma vez que proporciona mais coesão e permite acesso a diferentes informações dentro de uma perspectiva social [Granovetter, 1973]. Alguns estudos também consideram os potenciais benefícios da mobilidade acadêmica para promover a transferência de conhecimento [Kato & Ando, 2013; Silva et al., 2016].

Nesse contexto, é positiva a realização de estudos para conhecer melhor como as redes de colaboração entre pesquisadores são formadas e como elas influenciam o desenvolvimento científico e tecnológico, bem como para entender como a ciência tem se comportado com o passar dos anos e com a mudança das formas de colaboração e dos meios para troca de informação. Além disso, torna-se também importante avaliar o impacto da colaboração entre pesquisadores de áreas distintas na pesquisa realizada.

Finalmente, vale ressaltar que algumas ferramentas também foram desenvolvidas para automatizar o acesso e a coleta de dados da Plataforma Lattes. Por exemplo, Fernandes et al. [2011] desenvolveram uma ferramenta que importa no formato XML currículos cadastrados na Plataforma Lattes, permitindo assim maior flexibilidade na manipulação de seus dados. Tendo como base estratégias utilizadas nessa ferramenta, Dias [2016] desenvolveu um arcabouço para coleta, tratamento e análise de dados que permite a criação de grandes repositórios específicos a partir de dados da Plataforma Lattes e que foi utilizado para a coleta dos dados usados nesta dissertação.

1.4 Contribuições

Em resumo, as principais contribuições desta dissertação são:

- Um estudo de caracterização das oito grandes comunidades acadêmicas brasileiras por meio da construção das redes global e interdisciplinar de colaboração científica. Esta dissertação explora não apenas as propriedades de cada rede, mas também fornece uma análise visual delas.
- Uma investigação da intensidade das colaborações interdisciplinares entre comunidades distintas, enfatizando a importância de tais colaborações para o fortalecimento de toda a rede. Especificamente, identificamos padrões de colaboração distintos que nos permitem desvendar aspectos importantes subjacentes a tais interações.
- Uma análise das colaborações interdisciplinares tendo como base a localização geográfica dos pesquisadores. Essa análise fornece visualizações gráficas e tabu-

lares referentes ao percentual de colaborações interdisciplinares entre as regiões geográficas brasileiras, ao padrão das colaborações interdisciplinares entre as regiões e grandes áreas, e ao padrão de colaboração por grande área em cada região brasileira.

- Uma análise das colaborações interdisciplinares focada na idade acadêmica dos pesquisadores. Essa análise fornece visualizações gráficas e tabulares sobre a taxa de interdisciplinaridade por idade acadêmica e a taxa de interdisciplinaridade por faixa de idade acadêmica e grande área.
- Uma análise da quantidade anual de publicações resultantes das colaborações interdisciplinares desde de 1990. Essa análise provê visualizações gráficas e tabulares sobre a variação das taxas de colaboração interdisciplinar ao longo dos anos.

1.5 Organização da Dissertação

O restante desta dissertação está organizado da seguinte forma. O Capítulo 2 descreve os aspectos metodológicos que sustentam os objetivos específicos da dissertação. O Capítulo 3 apresenta resultados gerais da análise realizada sobre as redes de colaboração científica (global e interdisciplinar) brasileiras, levando em conta a região geográfica e a idade acadêmica dos pesquisadores considerados. O Capítulo 4 apresenta análises adicionais que consideram evolução temporal das colaborações interdisciplinares. Finalmente, o Capítulo 5 apresenta as nossas conclusões e discorre sobre trabalhos futuros.

Capítulo 2

Metodologia

Este capítulo descreve a metodologia utilizada para caracterização das comunidades científicas brasileiras, seleção e coleta dos dados dos currículos dos pesquisadores, e identificação das coautorias necessárias para a construção das redes de colaboração científica. Para um melhor entendimento das análises que são feitas nos Capítulos 3 e 4, são apresentados também conceitos básicos sobre redes complexas.

2.1 Definição do Esquema de Classificação

Como já mencionado, este trabalho adotou o nível superior do esquema de classificação de áreas do conhecimento do CNPq que considera oito grandes áreas: *Ciências Agrárias*, *Ciências Biológicas*, *Ciências da Saúde*, *Ciências Exatas e da Terra*, *Ciências Humanas*, *Ciências Sociais e Aplicadas*, *Engenharias*, e *Linguística, Letras e Artes*. De fato, este estudo abrange um amplo espectro de campos do conhecimento, fornecendo, assim, um arcabouço robusto para analisar como a interdisciplinaridade ocorre em grandes comunidades científicas.

2.2 Seleção dos Pesquisadores

Como a Plataforma Lattes armazena dados tanto de jovens pesquisadores (por exemplo, estudantes de graduação) quanto de pesquisadores mais experientes, neste trabalho consideramos apenas os pesquisadores que possuem doutorado. Assim, nosso objetivo foi enfatizar apenas colaborações interdisciplinares maduras, ou seja, colaborações estabelecidas por pesquisadores mais experientes. Note-se que o nosso esforço envolve

toda a rede de colaboração acadêmica brasileira, que compreende os pesquisadores com doutorado que têm alguma associação com uma instituição de pesquisa no país.

2.3 Coleta e Preparação dos Dados

Inicialmente, foram coletadas da Plataforma Lattes as versões XML dos currículos de todos os pesquisadores com doutorado (um total de 280.465 em abril de 2018), desconsiderando todas as entradas correspondentes a pesquisadores que não indicaram a sua grande área de pesquisa, conforme o esquema de classificação mencionado na Seção 2.1, reduzindo assim para 263.264 o total de currículos usados para geração das redes de colaboração científica brasileiras. Como ilustração, a Figura 2.1 apresenta um trecho da seção de identificação do currículo de um pesquisador conforme parcialmente visualizado na Plataforma Lattes, enquanto a Figura 2.2 apresenta a versão XML do mesmo trecho desse currículo.

Identificação	
Nome	Alberto Henrique Frade Laender
Nome em citações bibliográficas	LAENDER, A. H. F.; Laender, Alberto H F; Laender, Alberto H. F.; Laender, Alberto H.F.
Endereço	
Endereço Profissional	Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação. Av. Antônio Carlos, 6627 Pampulha 31270-010 - Belo Horizonte, MG - Brasil Telefone: (31) 34095860 Fax: (31) 34095858
Formação acadêmica/titulação	
1980 - 1984	Doutorado em Ciência da Computação. University of East Anglia, UEA, Inglaterra. Título: A Framework for the Definition and Manipulation of Database Views by End Users, Ano de obtenção: 1984. Orientador: Peter M. Stocker. Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil. Palavras-chave: Bancos de Dados; Banco de Dados Relacional; Visões; Interfaces para Usuários Finais. Grande área: Ciências Exatas e da Terra Grande Área: Ciências Exatas e da Terra / Área: Ciência da Computação / Subárea: Metodologia e Técnicas da Computação / Especialidade: Banco de Dados. Grande Área: Ciências Exatas e da Terra / Área: Ciência da Computação / Subárea: Metodologia e Técnicas da Computação / Especialidade: Sistemas de Informação.

Figura 2.1. Trecho de um Currículo Lattes

```

<CURRICULO-VITAE SISTEMA-ORIGEM-XML="LATTES, OFFLINE" NUMERO-IDENTIFICADOR="9089204821424223" DATA-ATUALIZACAO="24012019" HORA-ATUALIZACAO="222015">
+ <DADOS-GERAIS NOME-COMPLETO="Alberto Henrique Frade Laender" NOME-EM-CITACOES-BIBLIOGRAFICAS="LAENDER, A. H. F.;Laender, Alberto H. F.;Laender, Alberto H.F."
NACIONALIDADE="B" PAIS-DE-NASCIMENTO="Brasil" UF-NASCIMENTO="MG" CIDADE-NASCIMENTO="Belo Horizonte" PERMISSAO-DE-DIVULGACAO="NAO" DATA-FALECIMENTO="" SIGLA-PAIS-
<DADOS-GERAIS>
- <PRODUCAO-BIBLIOGRAFICA>
+ <TRABALHOS-EM-EVENTOS></TRABALHOS-EM-EVENTOS>
- <ARTIGOS-PUBLICADOS>
- <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="1" ORDEM-IMPORTANCIA="">
<DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="MER+: Uma Extensão do Modelo de Entidades e Relacionamentos Para Projeto Conceitual de Bancos de Dados." ANO-DO-
ARTIGO="1989" PAIS-DE-PUBLICACAO="Brasil" IDIOMA="Português" MEIO-DE-DIVULGACAO="IMPRESSO" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="NAO" DOI="" TITULO-DO-ARTIGO-
INGLES="" FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
<DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="RBC. Revista Brasileira de Computação" ISSN="01010883" VOLUME="5" FASCICULO="1" SERIE="" PAGINA-INICIAL="5"
PAGINA-FINAL="14" LOCAL-DE-PUBLICACAO="RIO DE JANEIRO"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="A.C. FERNANDES" NOME-PARA-CITACAO="FERNANDES, A. C." ORDEM-DE-AUTORIA="1" NRO-ID-CNPQ=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Alberto Henrique Frade Laender" NOME-PARA-CITACAO="LAENDER, A. H. F.;Laender, Alberto H. F.;Laender, Alberto H. F.;Laender, Alberto H.F." ORDEM-DE-
AUTORIA="2" NRO-ID-CNPQ="9089204821424223"/>
<PALAVRAS-CHAVE PALAVRA-CHAVE-1="Bancos de Dados" PALAVRA-CHAVE-2="Modelo E-R" PALAVRA-CHAVE-3="Projeto Conceitual" PALAVRA-CHAVE-4="" PALAVRA-CHAVE-5="" PALAVRA-CHAVE-
6=""/>
- <AREAS-DO-CONHECIMENTO>
<AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS EXATAS E DA TERRA" NOME-DA-AREA-DO-CONHECIMENTO="" NOME-DA-SUB-AREA-DO-
CONHECIMENTO="Metodologia e Técnicas da Computação" NOME-DA-ESPECIALIDADE=""/>
</AREAS-DO-CONHECIMENTO>
</ARTIGO-PUBLICADO>
- <ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="2" ORDEM-IMPORTANCIA="">
<DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="Mapping from a Conceptual Schema to a Target Internal Schema" ANO-DO-ARTIGO="1985" PAIS-DE-PUBLICACAO=""
IDIOMA="Ingles" MEIO-DE-DIVULGACAO="IMPRESSO" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="NAO" DOI="" TITULO-DO-ARTIGO-INGLES="" FLAG-DIVULGACAO-
CIENTIFICA="NAO"/>
<DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="Computer Journal" ISSN="00104620" VOLUME="28" FASCICULO="5" SERIE="" PAGINA-INICIAL="508" PAGINA-FINAL="517"
LOCAL-DE-PUBLICACAO=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Alberto Henrique Frade Laender" NOME-PARA-CITACAO="LAENDER, A. H. F.;Laender, Alberto H. F.;Laender, Alberto H. F.;Laender, Alberto H.F." ORDEM-DE-
AUTORIA="1" NRO-ID-CNPQ="9089204821424223"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="D.J. FLYNN" NOME-PARA-CITACAO="FLYNN, D. J." ORDEM-DE-AUTORIA="2" NRO-ID-CNPQ=""/>
<PALAVRAS-CHAVE PALAVRA-CHAVE-1="Conceptual Schema" PALAVRA-CHAVE-2="Database" PALAVRA-CHAVE-3="Internal Schema" PALAVRA-CHAVE-4="" PALAVRA-CHAVE-5="" PALAVRA-CHAVE-
6=""/>
- <AREAS-DO-CONHECIMENTO>
<AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS EXATAS E DA TERRA" NOME-DA-AREA-DO-CONHECIMENTO="" NOME-DA-SUB-AREA-DO-
CONHECIMENTO="Metodologia e Técnicas da Computação" NOME-DA-ESPECIALIDADE=""/>
</AREAS-DO-CONHECIMENTO>
</ARTIGO-PUBLICADO>

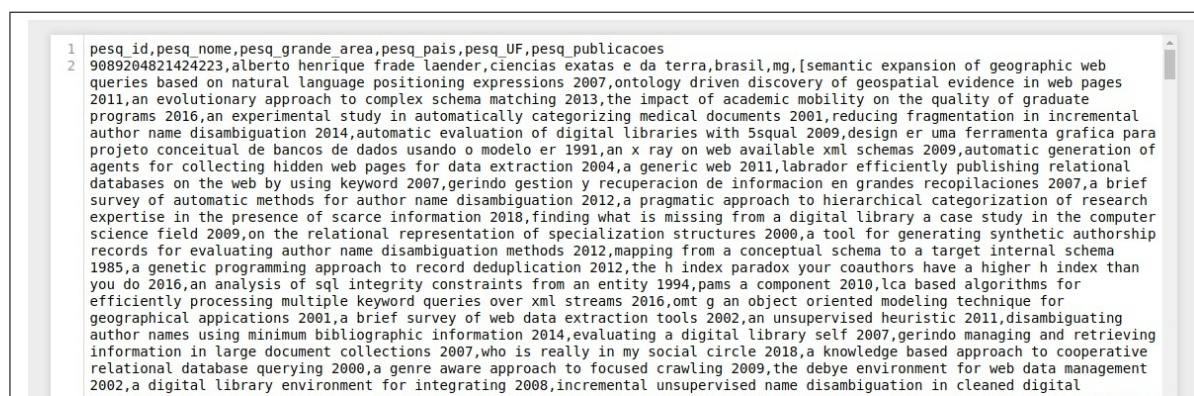
```

Figura 2.2. Trecho de um Currículo Lattes no Formato XML

Em seguida, os respectivos documentos XML foram transformados em arquivos CSV contendo apenas os campos necessários para identificar as colaborações científicas dos pesquisadores, reduzindo, assim, a quantidade de dados necessária para uma manipulação posterior. Os campos necessários para identificar as colaborações são descritos a seguir:

- `pesq_id` - identificação do pesquisador durante a definição das colaborações.
- `pesq_nome` - identificação do pesquisador durante a definição das colaborações.
- `pesq_grande_area` - identificação da comunidade científica do pesquisador durante grande parte das análises;
- `pesq_pais` - identificação dos pesquisadores residentes no Brasil durante as análises por região geográfica;
- `pesq_UF` - identificação da região geográfica dos pesquisadores residentes no Brasil durante as análises por região geográfica;
- `pesq_publicacoes` - identificação das publicações dos pesquisadores, primordial para a definição das colaborações. Necessária também para a análise da idade acadêmica dos pesquisadores e para a análise temporal das publicações interdisciplinares.

A Figura 2.3 apresenta parte de um currículo Lattes no formato CSV após a mineração dos dados necessários para a identificação das colaborações.



1	pesq_id,pesq_nome,pesq_grande_area,pesq_pais,pesq_UF,pesq_publicacoes
2	9089204821424223,alberto henrique frade laender,ciencias exatas e da terra,brasil,mg,[semantic expansion of geographic web queries based on natural language positioning expressions 2007,ontology driven discovery of geospatial evidence in web pages 2011,an evolutionary approach to complex schema matching 2013,the impact of academic mobility on the quality of graduate programs 2016,an experimental study in automatically categorizing medical documents 2001,reducing fragmentation in incremental author name disambiguation 2014,automatic evaluation of digital libraries with 5squal 2009,design er uma ferramenta grafica para projeto conceitual de bancos de dados usando o modelo er 1991,an x ray on web available xml schemas 2009,automatic generation of agents for collecting hidden web pages for data extraction 2004,a generic web 2011,labrador efficiently publishing relational databases on the web by using keyword 2007,gerindo gestion y recuperacion de informacion en grandes recopilaciones 2007,a brief survey of automatic methods for author name disambiguation 2012,a pragmatic approach to hierarchical categorization of research expertise in the presence of scarce information 2018,finding what is missing from a digital library a case study in the computer science field 2009,on the relational representation of specialization structures 2000,a tool for generating synthetic authorship records for evaluating author name disambiguation methods 2012,mapping from a conceptual schema to a target internal schema 1985,a genetic programming approach to record deduplication 2012,the h index paradox your coauthors have a higher h index than you do 2016,an analysis of sql integrity constraints from an entity 1994,pams a component 2010,lca based algorithms for efficiently processing multiple keyword queries over xml streams 2016,omt g an object oriented modeling technique for geographical appications 2001,a brief survey of web data extraction tools 2002,an unsupervised heuristic 2011,disambiguating author names using minimum bibliographic information 2014,evaluating a digital library self 2007,gerindo managing and retrieving information in large document collections 2007,who is really in my social circle 2018,a knowledge based approach to cooperative relational database querying 2000,a genre aware approach to focused crawling 2009,the debye environment for web data management 2002,a digital library environment for integrating 2008,incremental unsupervised name disambiguation in cleaned digital

Figura 2.3. Trecho de um arquivo CSV derivado de um Currículo Lattes

2.4 Identificação das Colaborações

Como a inserção de dados na Plataforma Lattes é feita manualmente pelos próprios pesquisadores, não há garantia de que seus coautores estejam devidamente identificados em cada publicação registrada. Em razão disso, foi necessário processar todos os currículos para identificar as colaborações, comparando os títulos das publicações e os nomes dos seus autores. Mais especificamente, para evitar o processamento de todas as publicações para todos os autores, este trabalho adotou a estratégia proposta por Dias & Moita [2015] que gera um dicionário com um tipo de identificação de coautoria. Essa estratégia primeiro remove todos os caracteres especiais e *stop words* dos títulos das publicações. Em seguida, ela reduz cada palavra do título ao seu radical. Por fim, ela concatena todos os *tokens* gerados e adiciona o ano de publicação à cadeia de caracteres final, criando assim uma chave *hash* para cada publicação. No final deste processo, as chaves *hash* são usadas como um identificador para as colaborações dos autores, ou seja, há uma colaboração envolvendo dois ou mais pesquisadores se eles compartilham a mesma chave *hash*. A Tabela 2.1 apresenta exemplos de chave *hash* geradas a partir das publicações do currículo de um pesquisador.

Por exemplo, a publicação “*A digital library environment for integrating, disseminating and exploring ecological data*” de 2008, cuja chave *hash* é “*digitlibraryenvironintegratedisseminateexploreecologydata2008*”, está associada aos seguintes autores: Evandrino Gomes Barros (5601476436148018), Alberto Henrique Frade Laender (9089204821424223), Marcos André Gonçalves (3457219624656691) e Francisco Antônio Rodrigues Barbosa (0932673194762377). Como as grandes áreas do conhecimento associadas a esses pesquisadores são, respectivamente, *Ciências Exatas e da Terra* (no caso dos três primeiros) e *Ciências Biológicas*, temos um total de seis colaborações, sendo três delas interdisciplinares. Ou seja, as colaborações do pesquisador da grande

Tabela 2.1. Exemplos de Chaves Hash para Identificação de Colaborações

Chave <i>hash</i>	Autores
briefsurveywebdataextracttools2002	3405503472010994 9089204821424223 5461069167314414
digitallibraryenvironintegratedisseminateexploreecologydata2008	5601476436148018 9089204821424223 3457219624656691 0932673194762377
...	
...	
timeawareranksportsocialnetworks2012	9089204821424223 3457219624656691 3621433615334969
...	

área de *Ciências Biológicas* com os demais da grande área de *Ciências Exatas e da Terra* são interdisciplinares. Em resumo, quatro pesquisadores de duas grandes áreas do conhecimento geraram um total de seis colaborações, sendo três interdisciplinares, e uma produção científica interdisciplinar.

Para fins das análises realizadas nesta dissertação, foram consideradas apenas as colaborações que correspondem a coautorias identificadas a partir de artigos de periódicos, uma vez que esse tipo de publicação é o mais comum nas principais áreas de pesquisa.

2.4.1 Redes Complexas: Conceitos Básicos

Redes complexas têm sido utilizadas em vários contextos para representar diversos aspectos do mundo real [Kleinberg & Easley, 2010] com o intuito de estudar o comportamento de uma comunidade e as relações entre os seus indivíduos [Granovetter, 1973]. Em nosso contexto, depois que as colaborações foram identificadas corretamente, duas redes complexas foram construídas para representar essas relações científicas entre pesquisadores brasileiros. A primeira delas, denominada *Global*, inclui um vértice para cada um dos 263.264 pesquisadores em nosso conjunto de dados, sendo criada uma aresta entre dois vértices quando os respectivos pesquisadores têm pelo menos uma colaboração científica. A segunda rede, denominada *Interdisciplinar*, inclui a mesma quantidade de vértices da rede anterior, sendo criada uma aresta apenas quando os pesquisadores que representam os vértices são de grandes áreas distintas.

Uma rede complexa é definida como um grafo $\mathbf{G}(\mathbf{V}, \mathbf{A})$, onde $\mathbf{V}$ é um conjunto não vazio de objetos denominados vértices (nós) e $\mathbf{A}$ é um conjunto de pares não ordenados dos vértices contidos em $\mathbf{V}$, denominadas arestas [Bondy & Murty, 1976; Granovetter, 1973; Scott, 2017]. Em resumo, uma rede complexa pode ser definida como um conjunto de entidades (vértices) e as conexões entre elas (arestas). As redes complexas podem ser consideradas um campo de estudo interdisciplinar, já que são usadas em estudos nas mais distintas áreas da ciência, como Biológica, Ciência da Computação e Física, entre outras [Strogatz, 2001].

A Figura 2.4 apresenta um exemplo de rede complexa que representa parte da rede de colaboração científica Global construída nesta dissertação, onde $\mathbf{V}$ é o conjunto de pesquisadores e $\mathbf{A}$ é o conjunto de coautorias par-a-par. Na rede ilustrada na Figura 2.4 é possível observar a formação de três comunidades por meio das colaborações dos pesquisadores. Pela posição dos vértices vemos que as duas primeiras comunidades são mais fortemente conectadas do que a terceira, o que indica que são mais colaborativas. Pelas cores nota-se que pesquisadores de diferentes grandes áreas do conhecimento (segundo a classificação do CNPq) se relacionam para formar novas comunidades. Como podemos ver nesse exemplo, mesmo quando uma comunidade formada a partir dessas colaborações é majoritariamente de pesquisadores de uma mesma grande área, ela é interdisciplinar.

Na literatura são encontrados vários tipos de métrica que são usados em diferentes análises para a extração de informações topológicas sobre as redes, tanto em relação aos vértices quanto às arestas [Bollobás, 1998; Coleman & Moré, 1983; Granovetter, 1973; Hatala, 2006; Havel, 1955]. A seguir são definidas algumas das métricas que serão consideradas durante a análise das redes de colaboração científica abordadas nesta dissertação.

2.4.1.1 Grau Médio

O grau de um vértice é definido pela quantidade de arestas incidentes sobre ele, ou seja, é a quantidade de relações que um vértice tem na rede [Granovetter, 1973]. Então quanto maior o grau de um vértice, maior é o número de suas relações e maior é a sua importância na rede [Havel, 1955]. A Figura 2.5 apresenta um exemplo com cinco vértices e quatro arestas. O vértice ao centro (azul) tem quatro relações e os demais têm uma relação cada. Então o grau do vértice em azul é igual a quatro e os demais igual a um.

O grau médio é definido pelo somatório do grau de todos os vértices de uma rede sob a quantidade total de vértices [Granovetter, 1973]. No exemplo da Figura 2.5, o

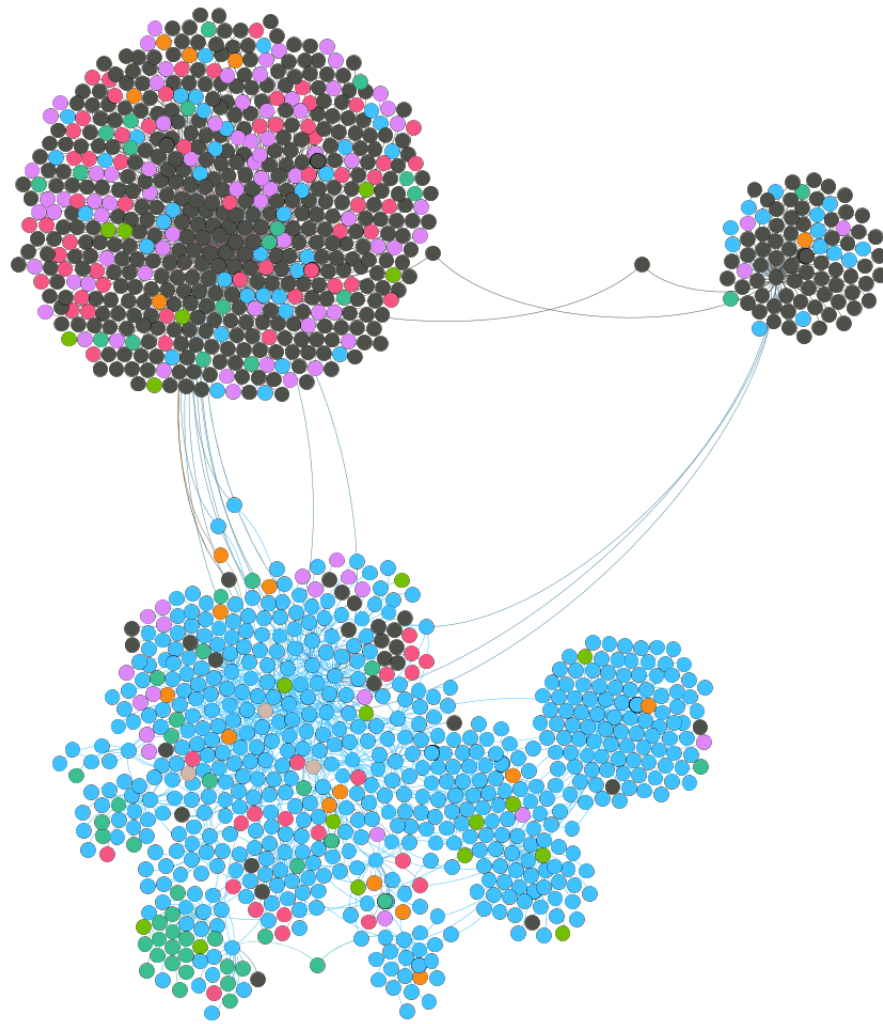


Figura 2.4. Exemplo de uma Rede Complexa

grau médio é 1,6, já que a soma dos graus dos cinco vértices da rede é igual a oito. No caso desta dissertação, o grau médio de uma rede representa a quantidade média de colaborações entre os pesquisadores.

2.4.1.2 Componente Gigante

Um componente gigante corresponde ao subgrafo fechado de uma rede com maior número de vértices e no qual existe um caminho entre quaisquer dois vértices [Bollobás, 1998; Gabardo, 2015]. No componente gigante é possível instanciar algumas métricas como, por exemplo, o seu tamanho (número de vértices), o percentual de seus vértices em relação à rede completa e o número total de arestas. A Figura 2.6 ilustra o componente gigante (vértices em azul) de uma rede complexa.

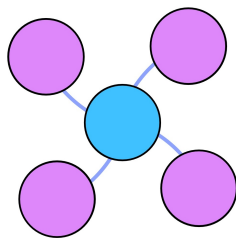


Figura 2.5. Grau de um vértice

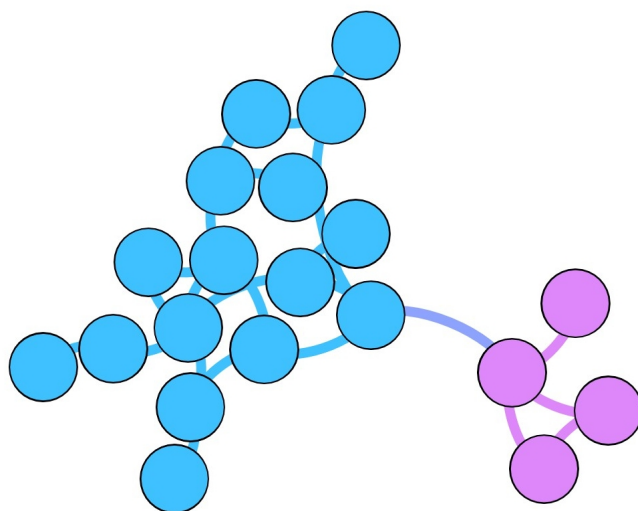


Figura 2.6. Exemplo de um Componente Gigante

O tamanho do componente gigante é dado pela quantidade de vértices que fazem parte desse subgrafo formado na rede. No caso desta dissertação, o tamanho do componente gigante é dado pela quantidade de pesquisadores presentes no maior grupo de colaboração científica formado dentro da rede. O percentual de vértices do componente gigante é a razão da quantidade de vértices dessa sub-rede em relação ao total de vértices da rede. Neste estudo, essa métrica mensura o percentual de pesquisadores presentes no maior grupo de colaboração. O número de arestas do componente gigante é a quantidade de arestas presentes nesse subgrafo. Em uma rede de colaboração científica essa métrica quantifica as interações no componente gigante.

2.4.1.3 Componente Isolado

Um componente isolado é um subgrafo, composto de um ou mais vértices, que não tem arestas com os demais componentes da rede [Granovetter, 1973]. A Figura 2.7 apresenta um exemplo de uma rede com componentes isolados, os três vértices (rosa) sem arestas.

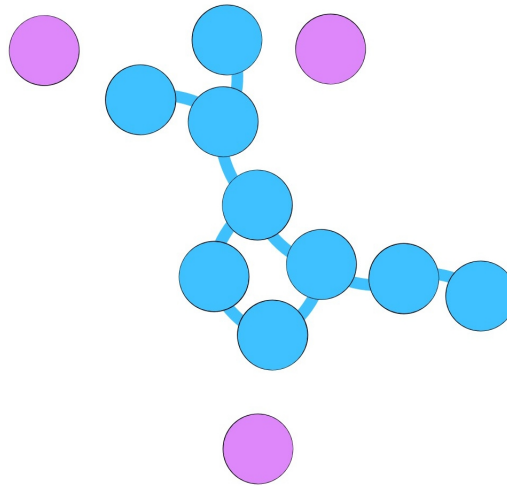


Figura 2.7. Componentes Isolados

O total de componentes isolados corresponde à quantidade de vértices (ou conjuntos de vértices) que estão isolados em uma rede. No exemplo ilustrado na Figura 2.7, há três componentes isolados (vértices em rosa). Na rede de colaboração científica construída nesta dissertação, o total de componentes isolados corresponde à quantidade de pesquisadores que não colaboram com os demais.

2.4.1.4 Caminho Mínimo e Diâmetro

Em uma rede complexa, um caminho é definido por uma sequência (finita ou não) de arestas que ligam diferentes vértices. O tamanho do caminho entre dois vértices pode ser mensurado pelo número total de arestas percorridas (ou pelo número de vértices visitados menos um). É possível que dois vértices sejam ligados por mais de um caminho com diferentes tamanhos, sendo o menor deles denominado *caminho mínimo*. O diâmetro que determina o comprimento de uma rede é definido pelo maior caminho entre todos os caminhos mínimos possíveis entre os vértices [Bollobás, 1998; Gabardo, 2015; Maia et al., 2013].

A Figura 2.8 apresenta uma rede com oito vértices destacados em azul apresentando um caminho ao longo de uma sequência numérica de 1 a 6 que, por ser o maior caminho mínimo possível, representa o diâmetro dessa rede. O tamanho do caminho médio de uma rede é a média do tamanho de todos os caminhos mínimos possíveis entre os vértices [Maia et al., 2013]. Essa métrica e o diâmetro são usadas para mensurar a força das conexões de uma rede, ou seja, quanto menor seus valores mais próximas são as relações entre os componentes da rede. No caso desta dissertação, o tamanho do

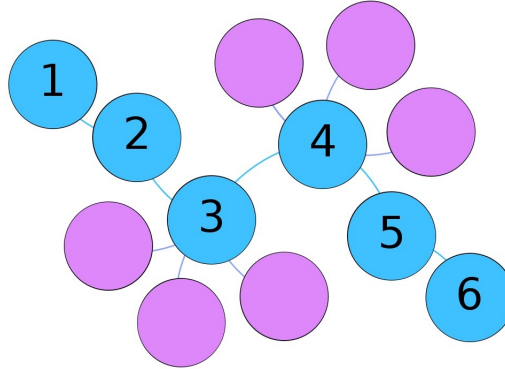


Figura 2.8. Caminho Mínimo e Diâmetro

caminho médio e o diâmetro de uma rede determinam quão fortes são as relações entre os pesquisadores que formam uma rede de colaboração científica. Em outras palavras, uma rede com um número maior de colaborações apresenta valores menores para o seu diâmetro e para o tamanho do seu caminho médio do que uma rede com um número menor de interações, já que na primeira as relações entre os pesquisadores são mais coesas e na segunda mais esparsas.

2.4.1.5 Densidade de uma Rede

A densidade de uma rede indica o grau de conectividade dentro dela e é calculada por meio da razão entre o o total de relacionamentos existentes e o total de relacionamentos que podem ocorrer na rede [Bollobás, 1998; Coleman & Moré, 1983; Granovetter, 1973]. Essa métrica pode variar entre 0, rede sem arestas, e 1, rede completa [Coleman & Moré, 1983; Hatala, 2006]. Por exemplo, uma rede com densidade igual a 0,3 mostra que o número total de relacionamentos é 30% da quantidade potencial de arestas entre todos os vértices. Por meio do cálculo dessa métrica é possível analisar quão densa é a rede, ou seja, quanto maior a densidade, maior a coesão entre o grupo de indivíduos representados por ela [Mena-Chalco et al., 2012; Hatala, 2006]. A densidade das redes de colaboração científica fornece uma visão de como são encorpados os relacionamentos entre os pesquisadores. A Equação 2.1 define como é calculada a densidade de uma rede complexa, onde $|A|$ é o número de arestas e $|V|$ é o número de vértices do grafo $G(V,A)$.

$$d(G(V,A)) = \frac{2 \cdot |A|}{|V| \cdot (|V| - 1)} \quad (2.1)$$

Capítulo 3

Análise das Redes de Colaboração

Neste capítulo é realizada uma análise geral das redes brasileiras de colaboração científica Global e Interdisciplinar. Essas redes fornecem uma visão das colaborações interdisciplinares envolvendo as oito grandes áreas que caracterizam a comunidade científica brasileira, abrangendo os últimos 66 anos. A Tabela 3.1 mostra a distribuição final do número de pesquisadores em cada uma das grandes áreas em nosso conjunto de dados. Como podemos ver, o número de pesquisadores em cada grande área é bastante desigual e basicamente reflete a popularidade de cada uma delas.

Tabela 3.1. Distribuição dos Pesquisadores por Grande Área

Grande Área	Pesquisadores
Ciências Agrárias	26.953 (10,2%)
Ciências Biológicas	36.356 (13,8%)
Ciências da Saúde	45.990 (17,5%)
Ciências Exatas e da Terra	39.231 (14,9%)
Ciências Humanas	44.743 (17,0%)
Ciências Sociais e Aplicadas	29.146 (11,1%)
Engenharias	24.746 (9,4%)
Linguística, Letras e Artes	16.099 (6,1%)
Total	263.264 (100.0%)

A seguir, são apresentados os resultados obtidos na análise das propriedades das redes brasileiras de colaboração científica global e interdisciplinar construídas a partir dos dados da Plataforma Lattes. Além disso, é descrito como essas redes estão interligadas, mostrando a proporção de colaborações em cada grande área. Por rede de colaboração global entende-se aquela cujos vértices correspondem aos 263.264 pesquisadores considerados e as arestas indicam as suas coautorias. Já a rede interdisciplinar possui o mesmo conjunto de vértices, sendo descartadas as arestas que correspondem

a coautorias dentro da própria grande área dos pesquisadores. Assim, esta dissertação enfatiza a importância de considerar colaborações interdisciplinares ao analisar as redes sociais acadêmicas.

Como ilustração dessa afirmação, a Figura 3.1 revela o quão volumosas e complexas são essas colaborações em cada comunidade científica. O gráfico à esquerda mostra a rede de colaboração global enfatizando as colaborações dentro de cada grande área, enquanto o gráfico à direita mostra apenas as colaborações interdisciplinares. É nítido que a rede Global é mais densa, com conexões fortes no componente gigante, enquanto a rede Interdisciplinar possui um conjunto mais disperso de nós conectados em suas extremidades. Embora as grandes áreas sejam bem definidas quando olhamos de perto para a rede Global, notamos claramente um componente conectado de forma bastante diferente na rede Interdisciplinar. Particularmente, existe uma forte sobreposição das grandes áreas (isto é, uma maior mistura de cores na vizinhança das mesmas), correspondendo cada cor a uma grande área específica.

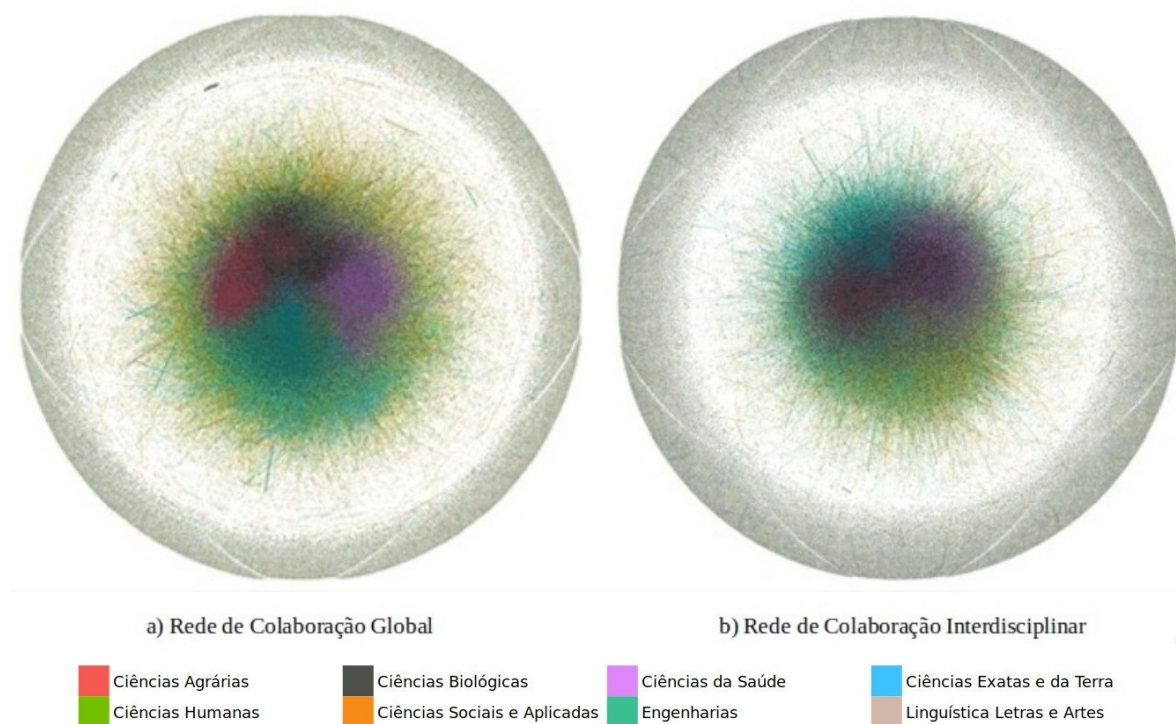


Figura 3.1. Redes Global e Interdisciplinar de Colaboração Científica no Brasil

Note que este esforço envolve a rede completa de colaboração acadêmica brasileira, compreendendo todos os indivíduos com doutorado (um total de 263.264 em abril de 2018) ora atuando em atividades de pesquisa no país. A Tabela 3.2 mostra os valores de algumas métricas comuns de redes complexas considerando as redes Global

Tabela 3.2. Comparação entre as Redes Global e Interdisciplinar

Métrica	Global	Interdisciplinar
Número Total de Vértices	263.264	263.264
Número de Vértices com Relações	211.594	151.664
Número de Vértices Isolados (Componentes Isolados)	51.670	111.600
Número de Arestas	2.563.017	900.992
Grau Médio	11,9	6,8
Tamanho do Componente Gigante	207.583	167.324
Percentual de Vértices no Componente Gigante	78,8%	63,6%
Arestas no Componente Gigante	2.447.681	791.161
Comprimento do Caminho Médio	5,2	7,1
Diâmetro da Rede	15	18
Densidade da Rede	3.98E-5	1.30E-5

e Interdisciplinar. Primeiramente, cabe destacar que 35,2% (900.992 de 2.563.017) das colaborações correspondem a arestas interdisciplinares, ou seja, envolvem pesquisadores de diferentes grandes áreas, o que enfatiza a importância da interdisciplinaridade no cenário atual de ciência e tecnologia. Este número é muito expressivo, já que se refere à rede de toda uma comunidade científica. Além disso, o fato de que um grande número de arestas no componente gigante (sub-gráfico máximo que inclui um caminho conectando cada par de nós) da rede Global é mantido na rede Interdisciplinar é muito relevante, já que 95,5% das arestas pertencem ao componente gigante. Além disso, isso reforça o quão fechadas e importantes essas colaborações são em um contexto acadêmico.

Como esperado, o número de arestas e o grau médio dos vértices diminuíram na rede Interdisciplinar, enquanto que o número total de componentes isolados aumentou. Apesar disso, o diâmetro e o comprimento do caminho médio em ambas as redes estão próximos. Nota-se que tais valores são ligeiramente maiores na rede Interdisciplinar que possui menos arestas, tornando-a mais esparsa. Em relação à densidade das redes (razão entre o número de arestas existentes e o número de arestas possíveis no gráfico completo), a rede Interdisciplinar tende a ser naturalmente menos densa, uma vez que somente as colaborações interdisciplinares são consideradas.

Como mencionado no Capítulo 2, foram considerados como vértices apenas os pesquisadores com doutorado, uma vez que nosso objetivo é avaliar colaborações mais consistentes. Por outro lado, em seu estudo pioneiro, Mena-Chalco et al. [2014] analisaram uma rede que inclui outros tipos de vértice (ou seja, pesquisadores de todos os níveis, variando de estudantes de graduação àqueles mais seniores). Assim, para efeitos de comparação com os nossos resultados na Tabela 3.2, são apresentados aqui alguns

números da rede por eles gerada: número de vértices igual a 1,131,912, grau médio igual a 4,4, densidade da rede igual $3,92\text{E-}6$, comprimento do caminho médio igual a 5,8 e porcentagem de vértices no componente gigante igual a 35,4%. Embora eles não tenham relatado nós isolados em sua caracterização, acredita-se que as diferenças observadas com relação a esses números se devem à presença de alunos e colaboradores não acadêmicos na respectiva rede. Ressalta-se, ainda, que o conjunto de dados por eles utilizado foi coletado em maio de 2011, enquanto que o nosso é bem mais recente. Assim, tais diferenças também podem ser explicadas pela evolução natural da comunidade científica brasileira. De fato, ao comparar o valor da densidade das redes dos dois estudos, eles diferem em uma ordem de grandeza (de $3.9\text{E-}6$ para $5.8\text{E-}5$), pois a rede do primeiro estudo inclui pesquisadores com distintos níveis de formação, enquanto que a nossa inclui apenas aqueles que possuem doutorado.

Vale ressaltar que Mena-Chalco et al. [2014] também realizaram uma análise topológica de cada grande área na rede. Mais especificamente, eles observaram padrões diferentes em todas elas, sustentando a hipótese de que toda a rede de colaboração científica brasileira apresenta uma interdisciplinaridade natural. Entretanto, o estudo deles apenas sugere a ideia que a rede de colaboração científica brasileira esteja se tornando mais interdisciplinar, em contraste com esta dissertação que tem o objetivo de estudar essa mesma rede e analisar a sua importância na produção científica nacional. Corroborando com os objetivos desta dissertação, de Souza Vanz & Stumpf [2012] também mostraram que as grandes áreas de pesquisa no Brasil apresentam diferentes padrões de colaboração. Assim, tais evidências mostram a necessidade de análises adicionais dessas colaborações interdisciplinares, particularmente para entender a motivação dos pesquisadores envolvidos [Iglič et al., 2017] e a dinâmica de sua produção científica [Silva et al., 2015].

3.1 Colaborações Interdisciplinares por Grande Área

Esta seção investiga as características das relações interdisciplinares encontradas na rede brasileira de colaboração científica, considerando as oito grandes áreas definidas de acordo com o esquema de classificação de áreas de conhecimento do CNPq. A Figura 3.2 mostra os percentuais de tais colaborações para cada grande área. Por exemplo, *Ciências Agrárias* (primeira barra) apresenta uma distribuição de colaboração em relação às outras grandes áreas de 2,4% (*Ciências Humanas*), 13,0% (*Ciências da Saúde*), 0,2% (*Linguística, Letras e Artes*), 53,1% (*Ciências Biológicas*), 9,7% (*En-*

genharias), 2,4% (*Ciências Sociais Aplicadas*) e 19,2% (*Ciências Exatas e da Terra*).

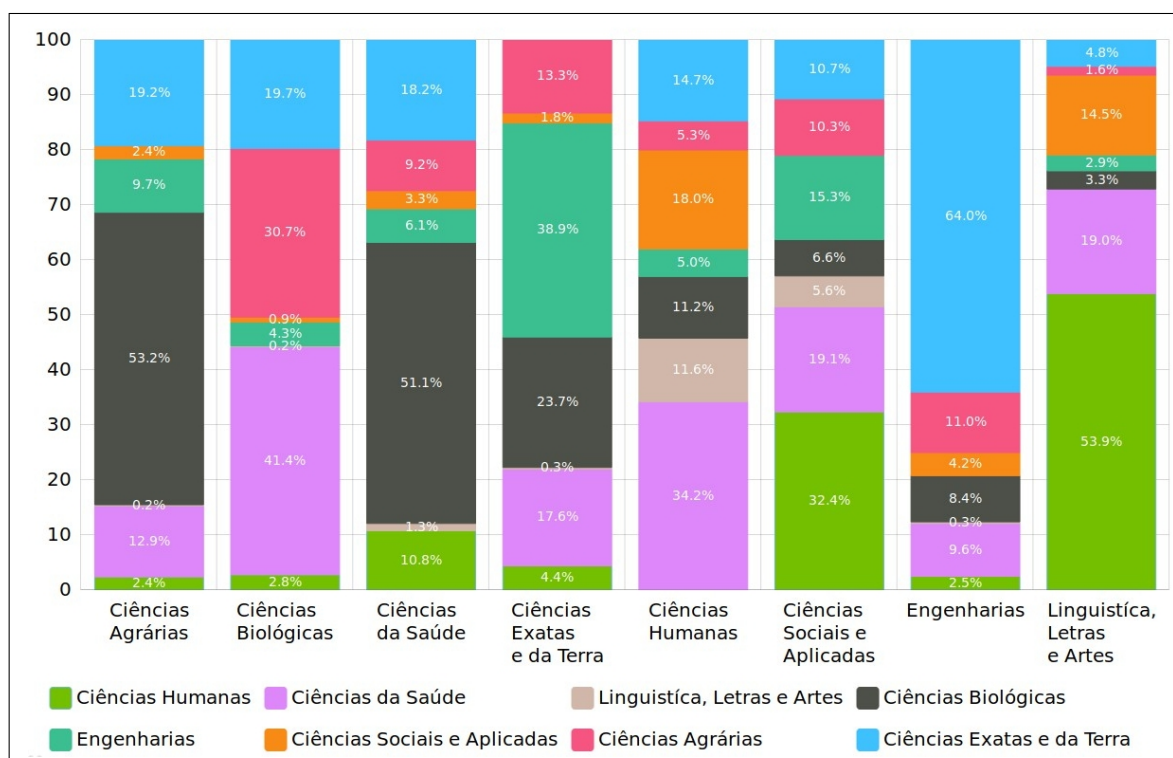


Figura 3.2. Percentual de Colaborações Interdisciplinares por Grande Área

Observa-se que em cada grande área, as colaborações interdisciplinares mais expressivas envolvem principalmente as grandes áreas de *Ciências Biológicas* (barras pretas), *Ciências da Saúde* (barras cor-de-rosa), *Ciências Exatas e da Terra* (barras azuis) e *Ciências Agrárias* (barras vermelhas), enquanto que as grandes áreas de *Linguística, Letras e Artes* (barras cinzas) e *Ciências Sociais e Aplicadas* (barras laranjas) apresentam percentuais menos expressivos. Mais especificamente, o maior percentual de colaborações interdisciplinares ocorre entre pesquisadores das grandes áreas de *Engenharias* e *Ciências Exatas e da Terra* (64,0%), seguido pelas colaborações entre os pesquisadores das grandes áreas de *Linguística, Letras e Artes* e *Ciências Humanas* (53,9%), *Ciências Agrárias* e *Ciências Biológicas* (53,1%), *Ciências da Saúde* e *Ciências Biológicas* (51,2%), *Ciências Biológicas* e *Ciências da Saúde* (41,4%), *Ciências Exatas e da Terra* e *Engenharias* (39,0%) e, finalmente, *Ciências Humanas* e *Ciências da Saúde* (34,2%). Nota-se que tais valores são bastante consistentes com a realidade dos pesquisadores envolvidos nessas colaborações, uma vez que essas grandes áreas possuem uma base teórica bastante próxima (por exemplo, há uma forte interação entre as grandes áreas de *Engenharias* e *Ciências Exatas e da Terra*). Vale ainda ressaltar que os pesquisadores da grande área de *Ciências Biológicas* são os que mais se envolvem

em colaborações interdisciplinares. De fato, a grande área de *Ciências Biológicas* não só provê uma forte base teórica para a grande área de *Ciências da Saúde*, mas também serve como fundamentação para novas disciplinas como a Bioinformática.

Em relação à diversidade das colaborações, destacam-se as grandes áreas de *Ciências Sociais e Aplicadas* e *Ciências Humanas* como as mais democráticas dentre todas, uma vez que seus pesquisadores tendem a colaborar de maneira mais equilibrada com colegas de outras áreas. Em contraste, pesquisadores das grandes áreas de *Engenharias*, *Ciências Agrárias*, *Ciências da Saúde* e *Linguística, Letras e Artes* tendem a concentrar suas colaborações com aqueles de uma grande área específica (*Ciências Exatas e da Terra*, *Ciências Biológicas*, *Ciências Biológicas* e *Ciências Humanas*, respectivamente).

A Tabela 3.3 mostra o número de colaborações para cada grande área nas redes Global e Interdisciplinar. Por exemplo, *Ciências Agrárias* apresenta 220.443 colaborações distintas (ou seja, publicações) que envolvem pelo menos um coautor daquela área específica, tendo 94.495 (42,9%) dessas relações, pelo menos, um coautor de outra grande área. Nota-se que o número de colaborações globais não reflete necessariamente o tamanho das comunidades em questão. Por exemplo, embora as *Ciências Exatas e da Terra* e as *Ciências Humanas* estejam entre as três grandes áreas mais populosas (ver Tabela 3.1), elas mostram uma grande divergência em termos do número de colaborações globais por pesquisador (respectivamente 5,3 e 2,0), o que reforça o fato de que existem diferentes padrões de colaboração nas grandes áreas.

Tabela 3.3. Publicações por Grande Área

Grande Área	Total [Média por Pesquisador]	
	Global	Interdisciplinar
Ciências Agrárias	220.443 [8,2]	94.495 [3,5] (42,9%)
Ciências Biológicas	255.572 [7,0]	164.084 [4,5] (64,2%)
Ciências da Saúde	340.933 [7,4]	131.303 [2,9] (38,5%)
Ciências Exatas e da Terra	206.724 [5,3]	106.485 [2,7] (51,5%)
Ciências Humanas	90.440 [2,0]	52.118 [1,2] (57,6%)
Ciências Sociais e Aplicadas	68.409 [2,3]	33.125 [1,1] (48,4%)
Engenharias	113.630 [4,6]	64.111 [2,6] (56,4%)
Linguística, Letras e Artes	15.374 [1,0]	9.060 [0,6] (58,9%)
Total	1.311.525 [5,0]	654.781 [2,5] (49,9%)

Em relação às colaborações interdisciplinares, a grande área de *Ciências da Saúde*, a mais populosa, apresenta o menor percentual (38,5%). Como discutido anteriormente, essa grande área está entre aquelas com maior proporção de colaborações com as demais. Por outro lado, as *Ciências Sociais e Aplicadas* mostram uma maior taxa de colaboração interdisciplinar (48,4%), apesar de apresentar um pequeno número de

pesquisadores e colaborações, o que endossa seu padrão diversificado.

A Figura 3.3 apresenta a variação temporal da taxa de colaboração interdisciplinar de cada grande área entre os anos de 1952 e 2017. Observa-se que a grande área de *Ciências da Saúde*, que tem o menor percentual de publicações com colaborações interdisciplinares (38,5%), apresenta em todos os anos a partir de 1980 a menor taxa anual de colaboração interdisciplinar entre todas as grandes áreas. Em contraste, a grande área de *Ciências Biológicas* que possui o maior percentual de interdisciplinaridade (64,2%), passa a ter a maior taxa de colaboração interdisciplinares a partir do início dos anos 2000.

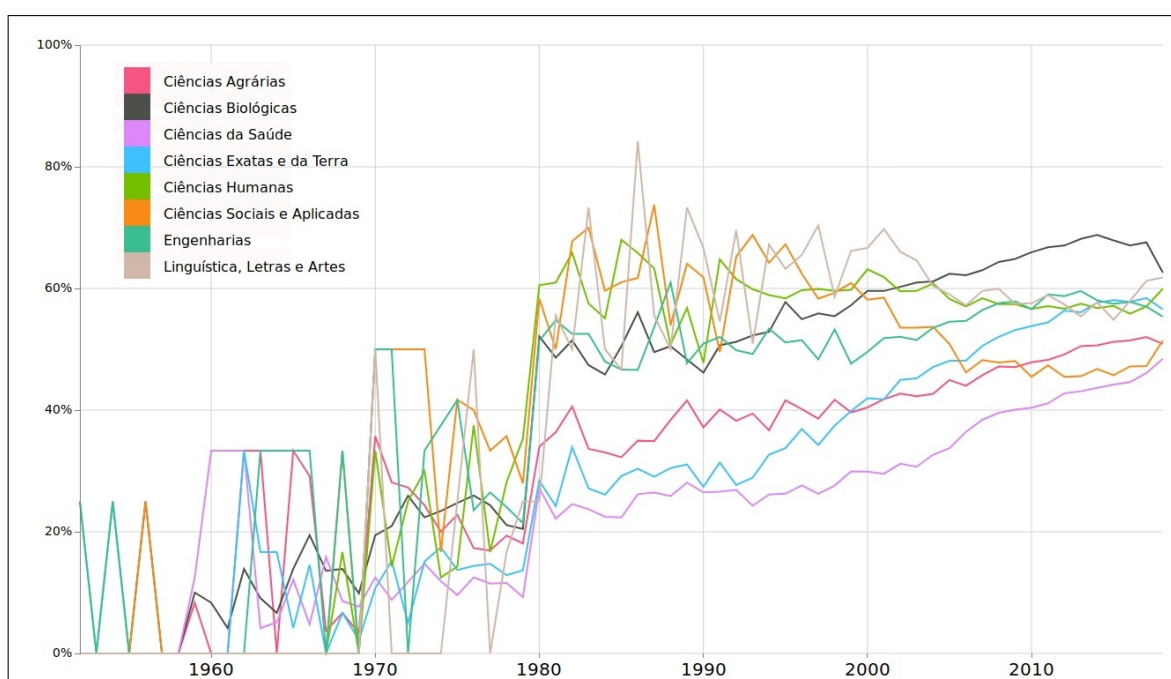


Figura 3.3. Taxa de Colaboração Interdisciplinar por Grande Área

Pode-se notar uma oscilação brusca nas taxas de colaboração interdisciplinar até 1980 nas oito grande áreas que, de modo geral foram baixas, sendo que somente as grandes áreas de *Ciências Sociais e Aplicadas*, *Engenharias* e *Linguística, Letras e Artes* ultrapassam a marca de 40% nesse período. Nos anos seguintes, a oscilação manteve uma certa constância até 2000, quando a taxa de colaboração interdisciplinar presenciou um crescimento na maioria das grandes áreas.

Analisando a Figura 3.3, observa-se que em nenhum ano no período entre 1980 e 2017 a taxa de colaboração interdisciplinar esteve abaixo de 20% em todas as oito grandes áreas, chegando essa taxa a ultrapassar os 60% nas grandes áreas de *Ciências Biológicas*, *Ciências Humanas*, *Ciências Sociais e Aplicadas* e *Engenharias*, e 80%

na grande área de *Linguística, Letras e Artes*. De modo geral, observa-se que as colaborações se tornaram mais interdisciplinares nos últimos anos.

Considerando as taxas de colaboração interdisciplinar a partir de 1980, a grande área de *Ciências Exatas e da Terra* é a que apresenta o maior crescimento. Em contraste, a grande área de *Ciências Sociais e Aplicadas* é a que apresenta o maior decréscimo. A partir do ano de 2000, as grandes áreas com padrões de colaboração interdisciplinar mais expressivos (ver Figura 3.2), *Ciências Agrárias*, *Ciências Biológicas*, *Ciências da Saúde*, *Ciências Exatas e da Terra* e *Engenharias*, apresentaram um aumento na taxa de colaboração interdisciplinar ao longo dos anos. Em contraste, as grandes áreas de *Ciências Humanas*, *Ciências Sociais e Aplicadas*, e *Linguística, Letras e Artes*, as menos expressivas em termos do total de publicações (ver Figura 3.2), tiveram uma queda na taxa de colaboração interdisciplinar.

Nos últimos 10 anos, as grandes áreas de *Ciências Agrárias* e *Ciências da Saúde*, aquelas com as maiores médias de publicação por pesquisador (8,2 e 7,4) e com os menores percentuais de colaboração interdisciplinar (42,9% e 38,5%) de acordo com a Tabela 3.3, são as que apresentam menores taxas anuais de colaboração interdisciplinar (ver Figura 3.3). Já a grande área de *Ciências Biológicas*, com a terceira maior média de publicação por pesquisador (0,7) e o maior percentual de colaboração interdisciplinar (64,2%), é aquela com a maior taxa anual de colaboração interdisciplinar.

Por outro lado, as duas grandes áreas com as menores médias de publicação por pesquisador, *Ciências Humanas* (2,0) e *Linguística, Letras e Artes* (1,0), e que possuem, respectivamente, a terceira (57,6%) e a segunda (58,9%) maiores taxas de publicação interdisciplinar (ver Tabela 3.3), estão entre as maiores taxas anuais de colaboração interdisciplinar (ver Figura 3.3). Em contrapartida, a grande área de *Ciências Sociais e Aplicadas*, a terceira grande área menos expressiva em termos de média de publicações por pesquisador (2,3), e que possui uma das menores taxas de colaboração interdisciplinar, está entre as três com menores taxas anuais de colaboração interdisciplinar.

De modo geral, as colaborações interdisciplinares não são necessariamente reflexo do tamanho das respectivas comunidades ou dos padrões de colaboração entre os seus pesquisadores, mas também de características específicas subjacentes a cada uma delas, exigindo, portanto, uma discussão mais detalhada. Assim, em seguida, analisamos duas dessas características: como as colaborações interdisciplinares ocorrem nas regiões geográficas brasileiras e como elas evoluíram ao longo do tempo.

3.2 Colaborações Interdisciplinares por Região Geográfica

Dado que há um número expressivo de colaborações interdisciplinares envolvendo pesquisadores de todas as grandes áreas, o nosso próximo passo é investigar como tais interações ocorrem nas cinco regiões geográficas brasileiras¹. Como o Brasil é um país continental composto por 26 estados, caracterizar tais colaborações científicas em termos de suas regiões é muito importante para entender melhor como o conhecimento científico se dissemina pelo país.

Para melhor compreensão da importância de cada uma dessas regiões, a Figura 3.4 apresenta alguns dados demográficos (disponibilizado pelo IBGE, Instituto Brasileiro de Geografia e Estatística²), a quantidade de programas de pós-graduação classificados com nível 7 pela CAPES³ e a quantidade de pesquisadores em cada uma dessas regiões. Vale ressaltar que, embora a rede interdisciplinar envolva 151.664 pesquisadores (ver Tabela 3.1), para esta análise específica foram considerados apenas os pesquisadores com endereço profissional no Brasil, contabilizando um total de 136.711 currículos.

A Figura 3.5 apresenta o percentual de colaborações interdisciplinares envolvendo os pesquisadores de cada região. Como se pode observar, a região Sudeste (em roxo) é a que possui o maior percentual de colaborações interdisciplinares, próximo a 50% em quase todos os cenários. Uma primeira explicação é que essa é a região mais rica e populosa do país (ver Figura 3.4) e, portanto, a que possui maior acesso a recursos, tanto governamentais como privados, para financiamento de pesquisa. De fato, existe uma relação intrínseca entre a riqueza e o tamanho da população dessas regiões quando se considera a distribuição de colaborações interdisciplinares. Outra questão relacionada refere-se ao fato de que um número substancial de universidades brasileiras de alto nível, considerando-se ensino e pesquisa, está localizado nessa região. Para ilustrar essa afirmação, de acordo com dados apresentados na Figura 3.4, 75,3% dos programas de pós-graduação com o mais alto grau de excelência (nível 7 de acordo com a classificação da CAPES) estão localizados na região Sudeste. Portanto, é possível afirmar que há uma forte relação entre a quantidade de programas de pós-graduação de qualidade em determinadas regiões e a distribuição das colaborações interdisciplinares.

Além disso, nota-se que a proximidade geográfica é um fator importante, já que o percentual das colaborações interdisciplinares dentro da própria região é significativo

¹Veja: https://en.wikipedia.org/wiki/Regions_of_Brazil

²IBGE: <https://www.ibge.gov.br/>. Números da Área a partir de 2018, População a partir de 2018 (estimativa) e PIB a partir de 2016.

³Plataforma Sucupira: <https://sucupira.capes.gov.br>

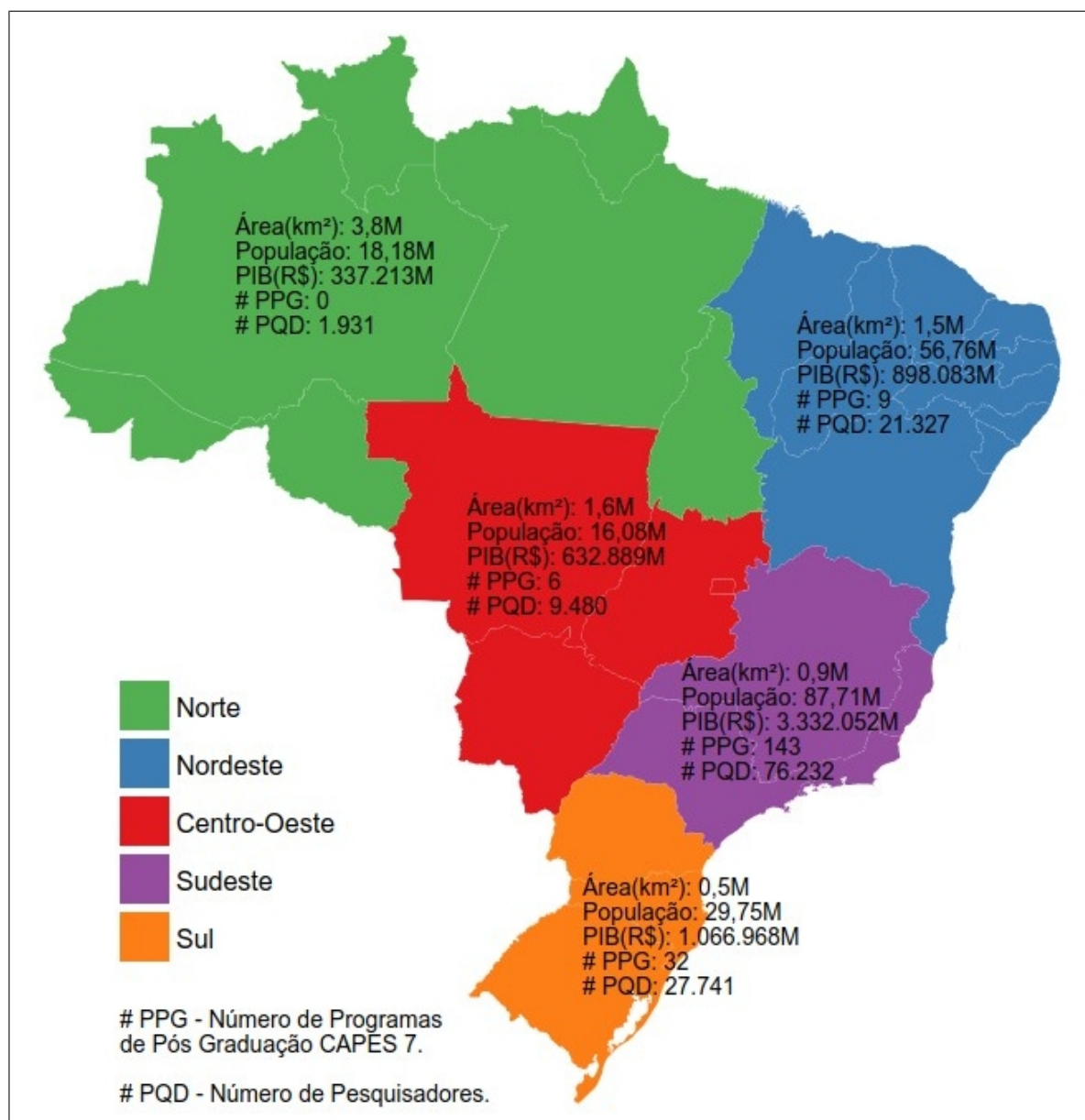


Figura 3.4. Regiões Geográficas Brasileiras.

em todos os cenários. Uma exceção aos padrões de colaboração discutidos acima é o caso da Região Centro-Oeste. Novamente, há um aspecto geográfico que sustenta esse fato, já que essa região é próxima às regiões Sudeste e Sul (as duas regiões mais ricas e com as melhores instituições acadêmicas), o que facilita o intercâmbio acadêmico com instituições dessas duas regiões. Em contraste, embora os pesquisadores da região Norte também tendam a manter uma colaboração expressiva com colegas das regiões Sudeste e Sul, isso ocorre em menor grau do que na região Centro-Oeste.

A fim de apresentar uma visão mais detalhada de tais padrões de colaboração

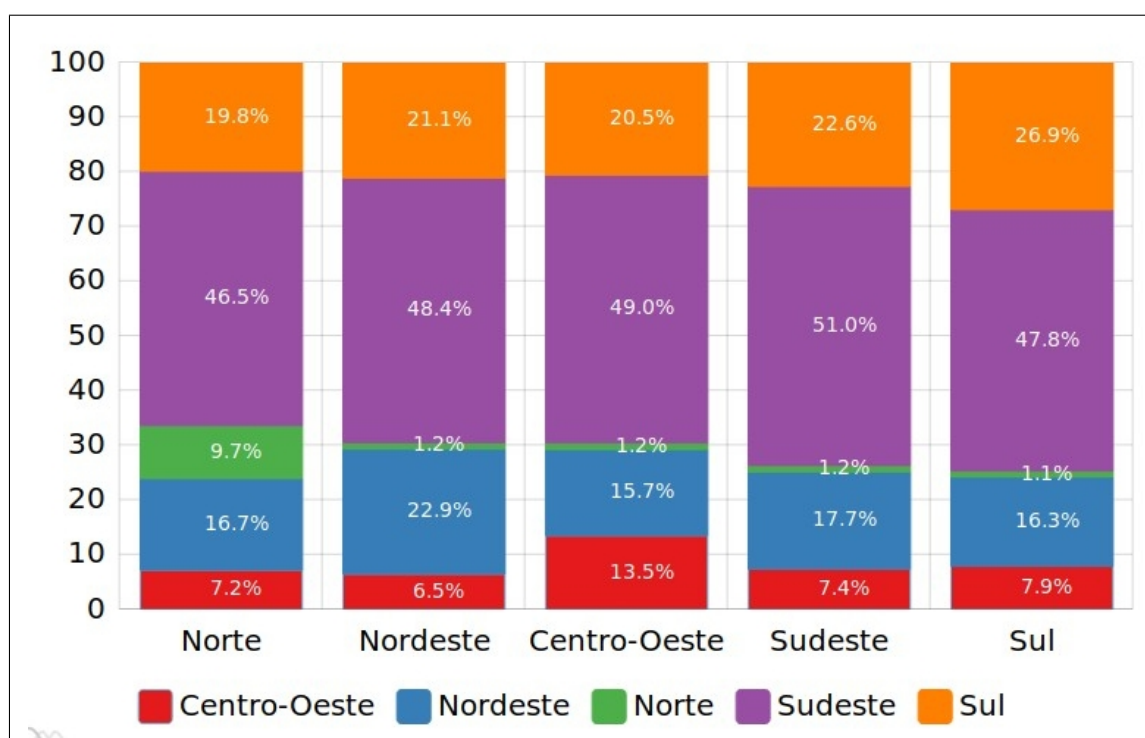


Figura 3.5. Distribuição das Colaborações Interdisciplinares por Região

interdisciplinar, a Figura 3.6 fornece uma visualização de como eles ocorrem entre as regiões do país. Para isso, cada vértice representa um grupo de pesquisadores de uma região associados a uma grande área, que são contextualizados pelo rótulo e cor do vértice, respectivamente. O tamanho do vértice e a espessura da aresta correspondem, respectivamente, à média do número de colaborações interdisciplinares normalizada pelo número de pesquisadores na respectiva região e ao número de colaborações entre as duas regiões representadas pelos dois vértices envolvidos. Por exemplo, a aresta entre o vértice rosa que representa a região Sudeste e o vértice cinza escuro que representa a região Sul indica que existe uma forte colaboração interdisciplinar entre pesquisadores das grandes áreas de *Ciências da Saúde* e *Ciências Biológicas* associados a instituições dessas duas regiões.

A Figura 3.6 descreve também a intensidade dos padrões de colaboração em relação ao total de pesquisadores de cada região. Por exemplo, a grande área de *Linguística, Letras e Artes* (bege), a menor comunidade em termos de pesquisadores, parece mais isolada, envolvendo colaborações menos intensas. Já a Tabela 3.4 resume os dez padrões de colaboração mais relevantes em termos do número de colaborações envolvidas. Por exemplo, a primeira linha mostra uma forte colaboração entre pesquisadores de *Ciências Biológicas* da região Sudeste e pesquisadores de *Ciências da Saúde* da região Sul. Além disso, a segunda linha confirma que a recíproca também é verdadeira,

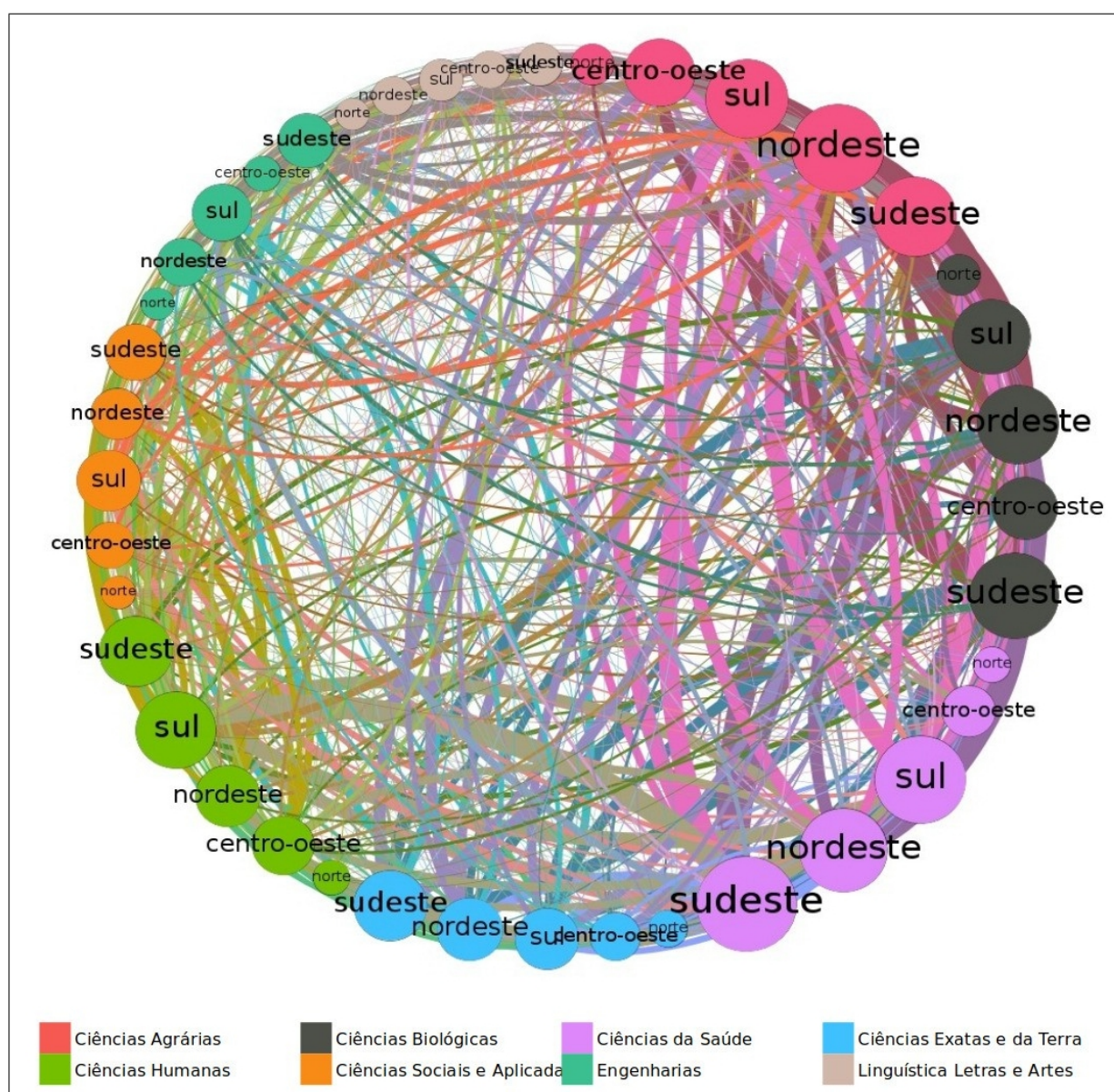


Figura 3.6. Padrões de Colaboração entre Regiões Brasileiras

ou seja, mesmo tendo acesso a outros pesquisadores devido à proximidade geográfica (que, como já foi mostrado, facilita as colaborações), há uma busca intensa pelo conhecimento de outros domínios. Da mesma forma, as grandes áreas de *Ciências Biológicas* (preta) e *Ciências Agrárias* (vermelho) têm considerável reciprocidade envolvendo as regiões Sudeste e Nordeste. Essas evidências reforçam que os pesquisadores brasileiros também buscam estabelecer novas parcerias científicas, tornando a rede brasileira de colaboração mais coesa e o acesso à informação mais amplamente disponível.

Uma análise mais profunda de cada grupo de vértices que representam os pesquisadores de uma grande área na Figura 3.6 mostra que eles possuem um tamanho

Tabela 3.4. Dez Principais Padrões de Colaboração Interdisciplinar entre Regiões Brasileiras

Contexto 1	Contexto 2	Colaborações
Ciências Biológicas/Sudeste	Ciências da Saúde/Sul	0,998
Ciências Biológicas/Sul	Ciências da Saúde/Sudeste	0,909
Ciências Humanas/Sul	Ciências da Saúde/Sudeste	0,908
Ciências Biológicas/Sudeste	Ciências da Saúde/Nordeste	0,906
Ciências Biológicas/Sudeste	Ciências Agrárias/Sul	0,877
Ciências Biológicas/Sudeste	Ciências Agrárias/Nordeste	0,867
Ciências Biológicas/Nordeste	Ciências da Saúde/Sudeste	0,839
Ciências Biológicas/Nordeste	Ciências Agrárias/Sudeste	0,812
Ciências Biológicas/Sul	Ciências Agrárias/Sudeste	0,758
Ciências Agrárias/Nordeste	Ciências da Saúde/Sudeste	0,752

bastante desigual. Embora isso reflita a popularidade de cada grande área, o gráfico nos permite entender a intensidade entre as colaborações. Por exemplo, considerando as grandes áreas de *Engenharias* (verde escuro) e *Ciências Sociais e Aplicadas* (laranja), que são duas comunidades comparáveis em número de pesquisadores (24.746 e 29.146, respectivamente), elas têm padrões de colaboração bastante diferentes. Mais especificamente, os pesquisadores das *Engenharias* (verde escuro) apresentam fortes laços com os das *Ciências Exatas e da Terra* (azul), enquanto aqueles das *Ciências Aplicadas e Sociais* (laranja) e das *Ciências Humanas* (verde claro) possuem vínculos menos intensos.

Vale ressaltar que as *Ciências Biológicas* (preta) é a grande área mais envolvida em colaborações interdisciplinares. Embora seja apenas a terceira maior grande área em termos de número de pesquisadores (ver Tabela 3.1), ela tem uma presença marcante em todos os cenários. Além disso, também podemos ver que as *Ciências da Saúde* (rosa) e as *Ciências Agrárias* (vermelho) estão entre as grandes áreas mais presentes nas principais colaborações interdisciplinares entre as regiões.

A Figura 3.7 apresenta os padrões de colaboração interdisciplinar envolvendo as grandes áreas em cada uma das cinco regiões geográficas brasileiras. O tamanho dos vértices é função do percentual de colaborações interdisciplinares entre os pesquisadores da região e a espessura das arestas expressa a quantidade de tais colaborações entre os pesquisadores das duas grandes áreas ligadas por elas. Os resultados mostram que a grande área com maior número de pesquisadores envolvidos em colaborações interdisciplinares em todas as regiões é a de *Ciências Biológicas*, corroborando a afirmação que essa é a grande área mais envolvida nesse tipo de colaboração. Em contraste, a grande área com a menor quantidade de colaborações em todos os cenários é a de *Linguística*,

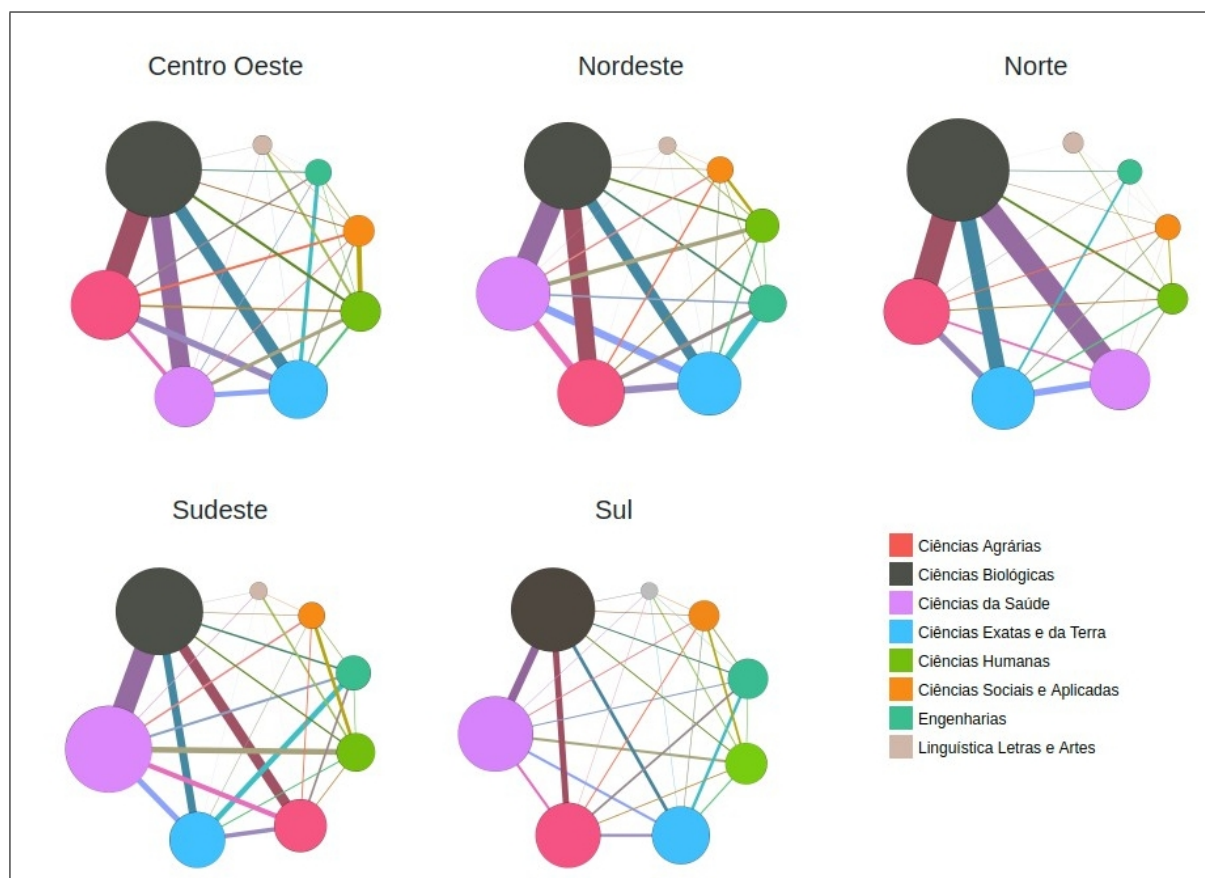
Letras e Artes.

Figura 3.7. Padrões de Colaboração entre as Grandes Áreas por Região

Considerando somente as seis grandes áreas que apresentam diferentes padrões de colaboração interdisciplinar entre as regiões brasileiras (*Ciências Agrárias*, *Ciências da Saúde*, *Ciências Exatas e da Terra*, *Ciências Humanas*, *Ciências Sociais e Aplicadas* e *Engenharias*), observa-se que as grandes áreas de *Ciências Agrárias*, *Ciências da Saúde* e *Ciências Exatas e da Terra* apresentam em todas as regiões entre os três percentuais mais relevantes. Por meio da análise da Figura 3.7, pode-se notar que a grande área de *Ciências da Saúde* tem um maior apelo nas regiões mais ricas e populosas do país (Sudeste, Sul e Nordeste - Figura 3.4), enquanto grande área de *Ciências Agrárias* destaca-se nas regiões com as maiores extensões territoriais (Norte e Centro Oeste - Figura 3.4). Isso reforça que as peculiaridades de cada região geográfica influenciam nos padrões de colaboração interdisciplinar.

Analisando a Figura 3.7, observa-se em termos gerais que apenas na região Norte o percentual de colaborações interdisciplinares envolvendo pesquisadores da grande área de *Ciências Exatas e da Terra* é maior do que aquele envolvendo pesquisadores da grande área de *Ciências da Saúde*. Já a região Nordeste é a única onde a grande

área das *Engenharias* está a frente da *Ciências Humanas* no percentual de colaborações interdisciplinares. Além disso, em todas as regiões, as *Ciências Humanas* tem um maior percentual de colaborações interdisciplinares do que as *Ciências Sociais e Aplicadas*. Pelo tamanho do vértice (verde escuro) observa-se que o maior percentual de colaborações interdisciplinares da grande área de *Engenharias* ocorre na região Nordeste e o menor na região Centro Oeste. Em contrapartida, a região Nordeste é a única onde a grande área de *Ciências Humanas* tem um percentual menor de colaborações interdisciplinares do que a grande área das *Engenharias*.

Em relação à diversidade das colaborações interdisciplinares, nota-se que a região Sul é a mais homogênea em relação ao número de pesquisadores e de colaborações entre todas as regiões, ou seja, ela envolve um maior número de grandes áreas nas principais colaborações interdisciplinares. Em contraste, a região Sudeste concentra a maioria de suas colaborações interdisciplinares entre duas grandes áreas, *Ciências Biológicas* e *Ciências da Saúde*.

Por fim, a espessura das arestas mostra que as colaborações entre pesquisadores das grandes áreas de *Ciências Biológicas* e *Ciências Agrárias*, e aquelas entre pesquisadores das grandes áreas de *Ciências Biológicas* e *Ciências da Saúde* estão entre as mais relevantes em todos os cenários (ver Figura 3.7). Já as colaborações entre pesquisadores das grandes áreas de *Ciências Exatas e da Terra* e das *Engenharias* é mais intensa nas regiões Nordeste e Sudeste. Na região Sul, citada como a mais democrática, nota-se que a espessura das arestas é mais similar em todos os cenários.

Capítulo 4

Análises Adicionais

Neste capítulo são apresentadas análises adicionais que procuram elucidar como a idade acadêmica dos pesquisadores influencia as suas colaborações interdisciplinares e como as colaborações interdisciplinares evoluem ao longo dos anos.

4.1 Influência da Idade Acadêmica nas Colaborações Interdisciplinares

Esta seção investiga como a idade acadêmica de um pesquisador influencia as suas colaborações interdisciplinares. Para isso, a idade acadêmica é definida a partir de sua primeira publicação, por exemplo, um pesquisador que teve sua primeira publicação em 1990 teria hoje (2019) a idade acadêmica de 29 anos. Em seguida, os pesquisadores foram divididos em três grupos de acordo com suas idades acadêmicas: até 10 anos, mais de 10 e até 20 anos, e mais de 20 anos. A Tabela 4.1 apresenta a distribuição dos pesquisadores brasileiros de acordo com essa estratificação.

Tabela 4.1. Distribuição dos Pesquisadores por Idade Acadêmica

Idade Acadêmica	Não Interdisciplinar	Interdisciplinar
[0,10]	55.1%	44.9%
(10,20]	42.0%	58.0%
> 20	41.5%	58.5%
Total	42.4%	57.6%

Como pode-se ver na Tabela 4.1, a interdisciplinaridade cresce com a idade acadêmica dos pesquisadores, ou seja, pesquisadores mais experientes tendem a se envolver mais em colaborações interdisciplinares. Outro ponto relevante a ser analisado é como a idade acadêmica dos pesquisadores impacta a interdisciplinaridade em cada grande

área. A Figura 4.1 mostra a distribuição dos pesquisadores em relação às colaborações não-interdisciplinares e interdisciplinares, considerando tanto a idade acadêmica quanto a grande área de atuação.

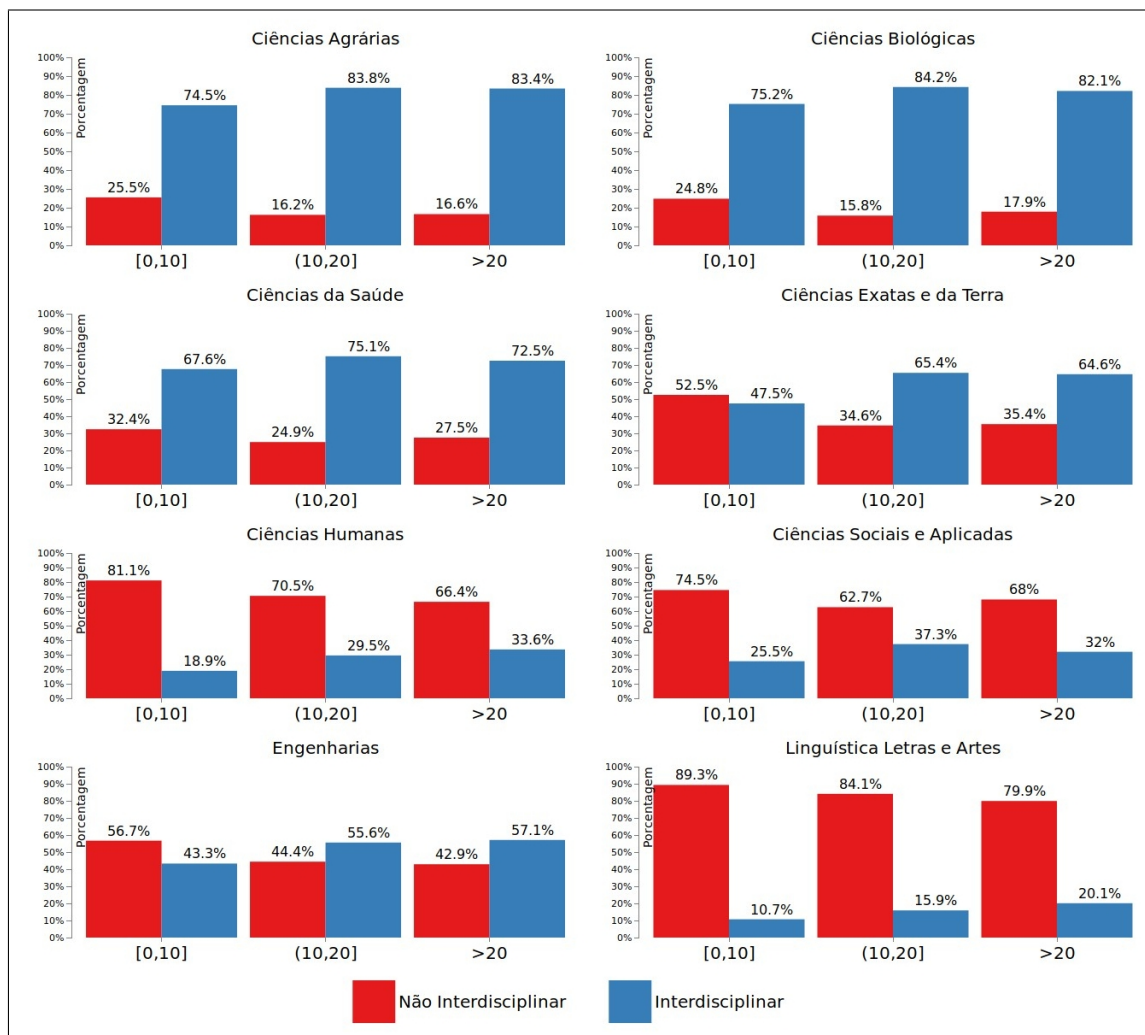


Figura 4.1. Distribuição dos Pesquisadores por Tipo de Colaboração e Idade Acadêmica

As grandes áreas de *Ciências Agrárias*, *Ciências Biológicas* e *Ciências da Saúde* apresentam o maior número de pesquisadores envolvidos em colaborações interdisciplinares em todas as faixas etárias acadêmicas (ou seja, barras azuis mais altas do que as vermelhas). Isso já era esperado, uma vez que essas grandes áreas mantêm fortes colaborações entre si (veja a Figura 3.2). Em contraste, as grandes áreas de *Ciências Humanas*, *Linguística*, *Letras e Artes*, e *Ciências Sociais e Aplicadas* são aquelas que, para todas as faixas etárias acadêmicas, têm menos pesquisadores envolvidos em colaborações interdisciplinares (ou seja, barras vermelhas mais altas do que as azuis).

Como discutido anteriormente (ver novamente a Figura 3.2), essas grandes áreas são aquelas que apresentam o menor nível de interdisciplinaridade em relação às demais.

Em geral, todas as grandes áreas mostram um crescimento na interdisciplinaridade nas duas primeiras faixas etárias acadêmicas ($[0,10]$ e $(10,20]$), que variam entre 5,2% (*Linguística, Letras e Artes*) e 17,9% (*Ciências Exatas e da Terra*). No entanto, no último intervalo (>20) as colaborações interdisciplinares permanecem praticamente estáveis. Além disso, apesar dos distintos padrões de interdisciplinaridade em algumas áreas importantes (por exemplo, *Ciências Biológicas* e *Linguística, Letras e Artes* com porcentagens extremamente divergentes), os resultados mostram que as interações de um pesquisador tendem a se tornar mais interdisciplinares ao longo do tempo, estabilizando-se no final de sua carreira.

4.2 Análise Temporal das Colaborações Interdisciplinares

Nesta seção é realizada uma análise da evolução temporal da taxa de colaboração interdisciplinar em cada grande área. Como essas taxas oscilaram bruscamente até 1980 (ver Figura 3.3), e considerando que a Internet, que hoje tem um papel fundamental na interação entre os grupos de pesquisa, passou a ser utilizada pelas universidades brasileiras em 1988¹, a análise realizada nesta seção tem como base as colaborações ocorridas a partir de 1990. Para uma melhor comparação entre as grandes áreas, esta seção está dividida em três subseções, definidas de acordo com os padrões de colaboração interdisciplinar observados no Capítulo 3.

4.2.1 Grandes Áreas com Taxas Baixas de Colaboração Interdisciplinar

Como visto anteriormente, as grandes áreas de *Ciências Humanas, Ciências Sociais e Aplicadas*, e *Linguística, Letras e Artes* são aquelas com os mais baixos percentuais de colaboração interdisciplinar (Figura 3.2), com as menores médias de publicação por pesquisador (Tabela 3.3), com menos pesquisadores em todas as regiões (Figuras 3.6 e 3.7) e com o menor número de pesquisadores envolvidos em colaborações interdisciplinares em todas as faixas etárias acadêmicas (Figura 4.1). A Figura 4.2 apresenta a evolução temporal das colaborações interdisciplinares dessas três grandes áreas.

¹https://pt.wikipedia.org/wiki/Internet_no_Brasil

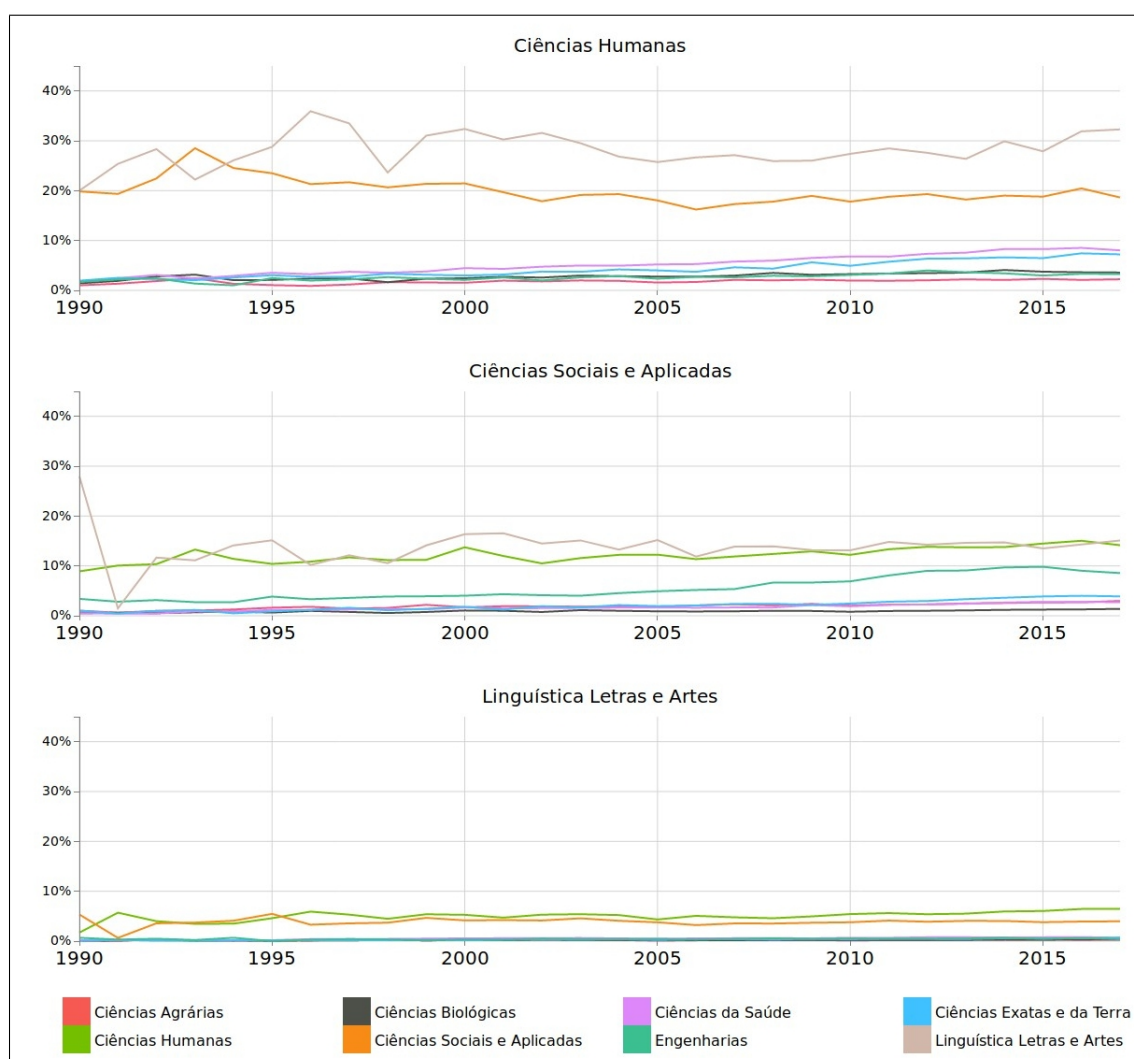


Figura 4.2. Evolução Temporal das Grandes Áreas com Taxas Baixas de Colaboração Interdisciplinar

Como pode ser visto na Figura 4.2, as grandes áreas de *Ciências Humanas*, *Ciências Sociais e Aplicadas* e *Linguística, Letras e Artes* possuem uma relação bastante constante desde os anos 90. Além disso, como pode ser observado, as taxas de colaboração interdisciplinar envolvendo essas grandes áreas não tiveram, de modo geral, oscilações muito bruscas, exceto no caso específico das colaborações entre as grandes áreas de *Ciências Sociais Aplicadas* e *Linguística, Letras e Artes* nos primeiros anos da década de 1990.

No caso das *Ciências Humanas* (primeiro gráfico da Figura 4.2), as taxas de colaboração interdisciplinar com as grandes áreas de *Ciências Sociais e Aplicadas* e *Linguística, Letras e Artes* têm se mantido em patamares próximos a 20% e 30%, respectivamente, em particular nos últimos anos. Entretanto, nota-se que as *Ciências*

Humanas possui taxas de colaboração interdisciplinar com as demais grandes áreas bem abaixo dos 10% em quase todo o período considerado, com uma pequena exceção nos casos das colaborações com as *Ciências da Saúde* e as *Ciências Exatas e da Terra* que a partir do final dos anos 2000 alcançaram patamares acima de 5%.

Já a grande área de *Ciências Sociais e Aplicadas* (segundo gráfico da Figura 4.2) possui taxas de colaboração interdisciplinar com a *Linguística, Letras e Artes* e as *Ciências Humanas* similares, e praticamente estáveis, entre 10% e 20% desde os anos de 1990. Em relação às demais grandes áreas, as taxas de colaboração interdisciplinar permanecem estáveis em patamares bem baixos, exceto no caso das colaborações com as *Engenharias* que a partir de 2010 atingiram patamares próximos a 10%.

Finalmente, o terceiro gráfico da Figura 4.2 mostra que as colaborações interdisciplinares da grande área de *Linguística, Letras e Artes* estão praticamente limitadas às Ciências Humanas e às Ciências Sociais e Aplicadas, com taxas abaixo dos 10%. As taxas de colaboração com outras grandes áreas, como a de Ciências Exatas e da Terra, são praticamente irrelevantes.

4.2.2 Grandes Áreas com Taxas Médias de Colaboração Interdisciplinar

De modo geral, as grandes áreas de *Ciências Exatas e da Terra* e de *Engenharias* possuem um percentual médio de colaborações interdisciplinares, exceto quando se considera a relação direta entre elas (Figura 3.2). Além disso, elas também possuem um volume médio de publicações (Tabela 3.3), vértices com tamanhos heterogêneos (Figuras 3.6 e 3.7) e um percentual de pesquisadores com e sem colaborações interdisciplinares sem bruscas diferenças em todas as faixas etárias acadêmicas. (Figura 4.1). A Figura 4.3 mostra a evolução da taxa de colaboração interdisciplinar dessas duas grandes áreas nos últimos 28 anos.

Conforme pode-se observar nos gráficos da Figura 4.3, as grandes áreas de *Ciências Exatas e da Terra* e de *Engenharias* possuem uma forte relação entre si, apresentando taxas de colaboração interdisciplinar bem superiores àquelas com as demais grandes áreas, com valores crescentes acima dos 20% no caso das Ciências Exatas e da Terra e acima dos 10% no caso das Engenharias. Além disso, pode-se notar que, as taxas de colaboração interdisciplinar envolvendo essas duas grandes áreas e as demais estão, no geral, em crescimento constante.

No caso específico da grande área de *Ciências Exatas e da Terra*, nota-se que a sua taxa de colaboração interdisciplinar com as *Engenharias* apresenta um crescimento vertiginoso, chegando à marca de 40% nos últimos anos, conforme mostra o gráfico

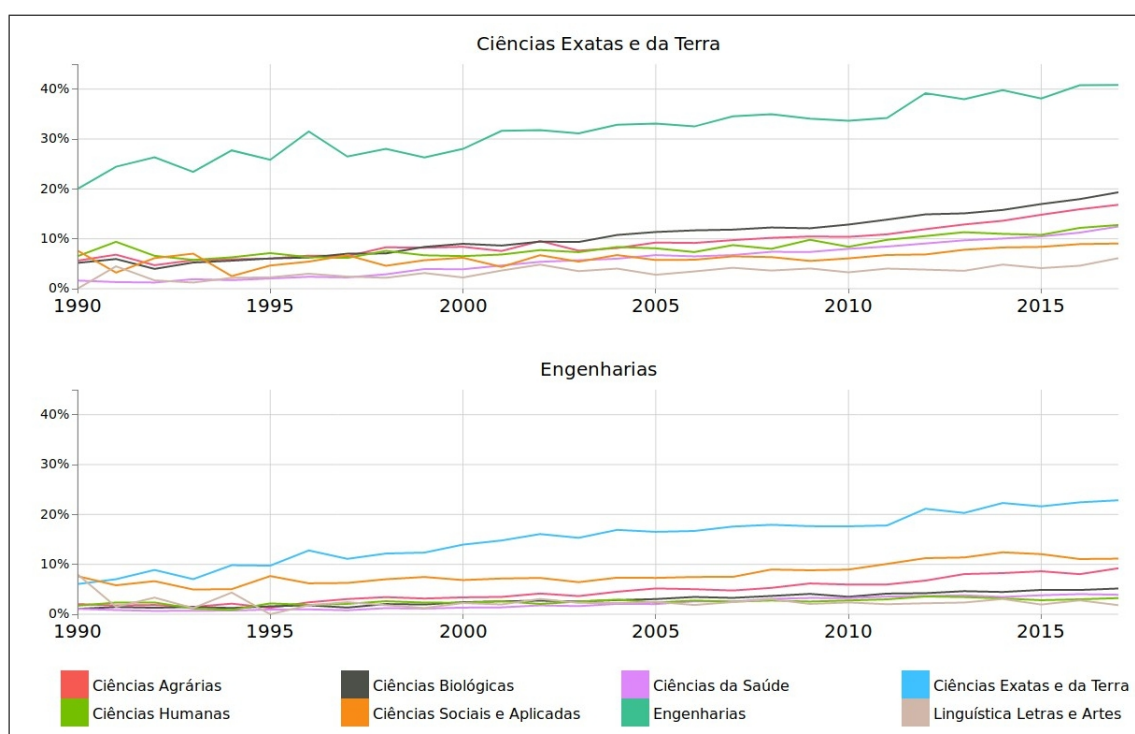


Figura 4.3. Evolução Temporal das Grandes Áreas com Taxas Médias de Colaboração Interdisciplinar

superior da Figura 4.3. Já em relação às suas colaborações com as grandes áreas de *Ciências Biológicas*, *Ciências Agrárias*, *Ciências da Saúde* e *Ciências Humanas*, essa taxa teve um crescimento mais discreto e recente, chegando a superar os 10% a partir de 2010.

Por outro lado, como pode ser visto no gráfico inferior da Figura 4.3, a grande área das *Engenharias*, apesar de forma menos intensa do que ocorre no cenário anterior, também apresenta uma taxa de colaboração interdisciplinar com as *Ciências Exatas e da Terra* superior àquelas praticadas com as demais grandes áreas, chegando a superar a marca de 20% pouco depois de 2010. Considerando as demais grandes áreas, as colaborações interdisciplinares das *Engenharias* com as *Ciências Sociais e Aplicadas* e com as *Ciências da Saúde* também apresentam taxas relevantes no mesmo período e que chegam a ultrapassar os 10% no caso da grande área de Ciências Sociais e Aplicadas.

4.2.3 Grandes Áreas com Taxas Altas de Colaboração Interdisciplinar

Finalmente, como pode ser visto na Figura 3.2, as grandes áreas de *Ciências Agrárias*, *Ciências Biológicas* e *Ciências da Saúde* são aquelas que apresentam, de modo geral,

as maiores taxas de colaboração com as demais. Além disso, elas também estão entre aquelas com maior volume de publicações (Tabela 3.3) e com o maior número de pesquisadores em quase todas as regiões (Figura 3.6 e Figura 3.7), além de contarem com os maiores percentuais de pesquisadores envolvidos em colaborações interdisciplinares em todas as faixas etárias (Figura 4.1). A Figura 4.4 analisa a evolução da taxa de colaboração interdisciplinar dessas grandes áreas ao longo dos anos.

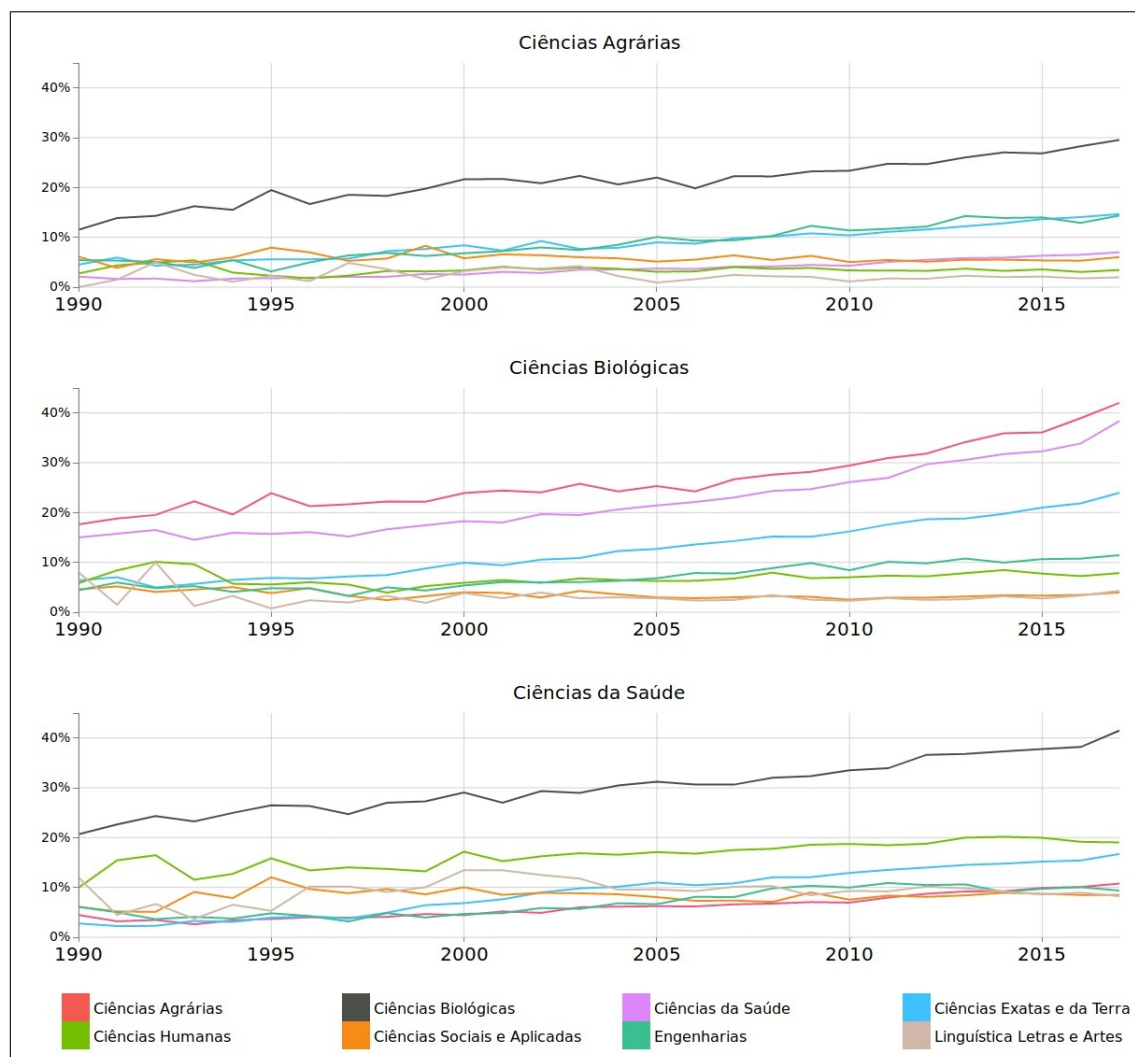


Figura 4.4. Evolução Temporal das Grandes Áreas com Taxas Altas de Colaboração Interdisciplinar

Conforme mostra o primeiro gráfico da Figura 4.4, as *Ciências Agrárias* apresenta uma taxa de colaboração interdisciplinar com as *Ciências Biológicas* com crescimento relevante e valores bem superiores aos das demais grandes áreas, aproximando-se dos 30% nos últimos anos. Em relação às *Ciências Exatas e da Terra* e às *Engenharias*, as taxas de colaboração interdisciplinar apresentam um crescimento contínuo e que

ultrapassa os 10% nos últimos 10 anos. Já no caso das demais grandes áreas, a taxa de colaboração interdisciplinar apresenta valores constantes e similares, que se firmaram próximos a 5% nos últimos 15 anos.

Já em relação às *Ciências Biológicas* (segundo gráfico da Figura 4.4), nota-se que as taxas de colaboração interdisciplinar com as *Ciências Agrárias* e com as *Ciências da Saúde* tiveram um crescimento vertiginoso, saltando de cerca de 20% para próximo de 40% nos últimos anos. Com um crescimento menos significativo, as taxas de colaboração interdisciplinar com as *Ciências Exatas e da Terra* se mantêm desde do final dos anos 1990 na terceira posição com valores acima dos 20% a partir do final da década de 2010. Em relação às demais grandes áreas, *Engenharias*, *Ciências Humanas* e *Ciências Sociais e Aplicadas*, as taxas de colaboração interdisciplinar se mantiveram praticamente constantes em patamares abaixo dos 10%, exceto no caso das Engenharias em que essa taxa ultrapassou ligeiramente esse valor a partir de 2010.

Por fim, o terceiro gráfico da Figura 4.4 mostra a grande área de *Ciências da Saúde* com taxas de colaboração interdisciplinar crescentes com todas as demais grandes áreas, com destaque para as colaborações com as Ciências Biológicas (variação de 20 para 40%) e com as *Ciências Humanas* (variação de 10 para 20%). Vale ainda ressaltar o crescimento da taxa de colaboração interdisciplinar com as Ciências Exatas e da Terra que desde 2000 se mantém crescente, estando atualmente próxima à taxa das Ciências Humanas, pouco abaixo dos 20%.

4.3 Considerações Finais

Este capítulo mostrou que a idade acadêmica dos pesquisadores influencia em suas colaborações interdisciplinares. De acordo com os dados considerados, nos primeiros 10 anos de carreira aproximadamente 45% dos pesquisadores se envolveram em colaborações interdisciplinares, sendo que nos 10 anos seguintes esse percentual subiu para 58.0%. Em relação às grandes áreas, observou-se que as *Ciências Agrárias*, *Ciências Biológicas* e *Ciências da Saúde*, que são as mais colaborativas (ver Capítulo 3), também são aquelas com os maiores percentuais de pesquisadores envolvidos em colaborações interdisciplinares em todas as faixas etárias. Já as *Ciências Exatas e da Terra* e as *Engenharias* são as grandes áreas com maior equilíbrio entre pesquisadores envolvidos em colaborações interdisciplinares e não interdisciplinares. Por fim, as grandes áreas de *Ciências Humanas*, *Ciências Sociais e Aplicadas* e *Linguística, Letras e Artes*, aquelas que menos colaboram (ver Capítulo 3), também são aquelas que possuem menos pesquisadores com colaborações interdisciplinares em todas as faixas etárias.

Vale ressaltar também que em todas as grandes áreas, inclusive aquelas com as menores taxas de colaboração interdisciplinar, há pelo menos um caso em que essa taxa se mostra crescente nos últimos anos. Nota-se ainda que as *Ciências Biológicas*, a grande área mais colaborativa (ver Capítulo 3), apresenta também os maiores valores para a taxa de colaboração interdisciplinar com as *Ciências Agrárias* (10% a 30%) e as *Ciências da Saúde* (20% a 40%). Nas *Ciências Exatas e da Terra* é possível ainda observar a maior diferença entre a maior taxa de colaboração interdisciplinar (com as *Engenharias*) em relação às colaborações com as demais grandes áreas. Finalmente, a grande área de *Linguística, Letras e Artes* é a única que apresenta taxas de colaboração interdisciplinar com as demais áreas com valores inferiores a 10% em todo o período considerado.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

Esta dissertação procurou caracterizar a importância das colaborações interdisciplinares na comunidade científica brasileira. Para isso, foi analisada a rede brasileira de colaborações científicas de acordo com as oito grandes áreas de pesquisa definidas pelo CNPq. Em geral, os resultados mostram que essa rede cresceu de forma consistente, tornando-se mais interdisciplinar. Particularmente, foi mostrado que 35,2% de todas as colaborações são interdisciplinares e 57,6% dos pesquisadores têm pelo menos uma interação científica desse tipo. Além disso, os resultados também mostram como essas áreas são integradas, enfatizando a força das colaborações envolvendo seus respectivos pesquisadores.

Ao explorar a rede interdisciplinar considerando as grandes áreas de pesquisa, observamos perfis distintos que demonstram a dinâmica e a peculiaridade de cada uma delas. Em geral, nosso estudo reforça a ideia de que existe um esforço claro para estabelecer novas relações, tornando a rede brasileira de colaboração científica mais interconectada e robusta. Considerando o potencial interdisciplinar das grandes áreas, nota-se que todas elas relacionam-se entre si. Vale ressaltar também que a grande área de *Ciências Biológicas*, a terceira mais populosa, destaca-se como a mais democrática em termos de diversidade interdisciplinar e que a grande área de *Linguística, Letras e Artes*, a menos populosa, possui os menores percentuais de colaboração interdisciplinar. No entanto, é importante notar que o tamanho das grandes áreas não é um fator chave na promoção da interdisciplinaridade, já que, por exemplo, as *Ciências Humanas* têm um percentual de colaboração interdisciplinar menor do que as *Ciências Agrárias*, mesmo sendo mais populosa.

Conforme mostrado, cada região do país possui fortes laços com suas próprias

raízes, reforçando, assim, que a localização geográfica é um fator determinante para a interdisciplinaridade. Além disso, as regiões mais ricas tendem a ser mais atraentes e, portanto, têm uma proporção muito expressiva de colaborações interdisciplinares em todos os cenários estudados, o que corrobora alguns trabalhos anteriores [Furtado et al., 2015; Silva et al., 2016]. As análises também mostram que as grandes áreas de *Ciências Agrárias*, *Ciências Biológicas* e *Ciências da Saúde* são as mais envolvidas em colaborações interdisciplinares nas diferentes regiões. Além disso, há uma tendência dos pesquisadores em colaborar mais de forma interdisciplinar com o passar dos anos, mesmo aqueles de grandes áreas menos colaborativas. Finalmente, vale ressaltar que, de modo geral, as taxas de colaboração interdisciplinar estão em crescimento desde o final dos anos de 1990, chegando, nos melhores casos, a patamares na faixa de 40%.

5.2 Trabalhos Futuros

Como trabalho futuro, pretendemos analisar a formação de pesquisadores que tendem a migrar para adquirir mais conhecimento para disseminar em suas próprias regiões ou transmitir para outras áreas por meio de colaborações de pesquisa de interesse mútuo [Furtado et al., 2015; Leydesdorff & Persson, 2010; Mooney et al., 2013; Silva et al., 2016]. Outra perspectiva é explicitar a força das interações, por exemplo, destacando os principais atores (pesquisadores ou grupos de pesquisadores) que tornam as redes de colaboração mais integradas ou mesmo identificando as colaborações que dão origem a diversas comunidades acadêmicas devido à sua força [Leão et al., 2017]. As evidências existentes nos levam a especular que há um compartilhamento implícito de conhecimento interdisciplinar entre as regiões, de modo que, para explicar melhor a força dessas colaborações, seria importante identificar aqueles pesquisadores que são realmente relevantes para a difusão do conhecimento e, portanto, para promover a evolução de uma comunidade acadêmica [Furtado et al., 2015; Kato & Ando, 2013].

Referências Bibliográficas

- Abramo, G.; D'Angelo, C. A. & Di Costa, F. (2018). The effect of multidisciplinary collaborations on research diversification. *Scientometrics*, 116(1):423–433.
- Araújo, E. B.; Moreira, A. A.; Furtado, V.; Pequeno, T. H. & Andrade Jr, J. S. (2014). Collaboration Networks from a Large CV Database: Dynamics, Topology and Bonus Impact. *PLOS One*, 9(3):e90537.
- Benevenuto, F.; Laender, A. H. F. & Alves, B. L. (2016). The H-index paradox: your coauthors have a higher H-index than you do. *Scientometrics*, 106(1):469–474.
- Boaventura, M.; Bonson, K.; Silva, A.; Veloso, A. & Meira Jr, W. (2014). Caracterização Temporal das Redes de Colaboração Científica nas Universidades Brasileiras: Anos 2000-2013. In *Proceedings of the III Brazilian Workshop on Social Network Analysis and Mining*, Brasília, DF, Brasil.
- Bollobás, B. (1998). *Modern graph theory (Graduate texts in mathematics)*. Springer, New York, USA.
- Bondy, J. A. & Murty, U. S. R. (1976). *Graph theory with applications*, volume 290. Macmillan, London, UK.
- Coleman, T. F. & Moré, J. J. (1983). Estimation of sparse jacobian matrices and graph coloring blems. *SIAM journal on Numerical Analysis*, 20(1):187–209.
- de Farias, L. R.; Vargas, A. P. & Borges, E. N. (2012). Um Sistema para Análise de Redes de Pesquisa Baseado na Plataforma Lattes. *Anais da VIII Escola Regional de Banco de Dados*. Curitiba, PR, Brasil.
- de Siqueira, G. O.; Canuto, S. D.; Gonçalves, M. A. & Laender, A. H. F. (2017). Automatic Hierarchical Categorization of Research Expertise Using Minimum Information. In *Research and Advanced Technology for Digital Libraries - 21st International*

- Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, September 18-21, 2017, Proceedings*, pp. 103–115.
- de Souza Vanz, S. A. & Stumpf, I. R. C. (2012). Scientific Output Indicators and Scientific Collaboration Network Mapping in Brazil. *Collnet Journal of Scientometrics and Information Management*, 6(2):315–334.
- Dias, T. M. R. (2016). *Um Estudo sobre a Produção Científica Brasileira a partir de Dados da Plataforma Lattes*. Tese de doutorado, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Brasil.
- Dias, T. M. R. & Moita, G. F. (2015). A method for the identification of collaboration in large scientific databases. *Em Questão*, 21(2).
- Digiampietri, L.; Mena-Chalco, J.; de Jesús Pérez-Alcázar, J.; Tuesta, E. F.; Delgado, K. & Mugnaini, R. (2012). Minerando e Caracterizando Dados de Currículos Lattes. In *Proceedings of the I Brazilian Workshop on Social Network Analysis and Mining*, Curitiba, PR, Brasil.
- Fernandes, G.; Oliveira, J. & Souza, J. (2011). XMLattes - A Tool for Importing and Exporting Curricula Data. In *Proceedings of the International Conference on Information and Knowledge Engineering*, San Diego, California, USA.
- Franzoni, V.; Lepri, M. & Milani, A. (2019). Topological and Semantic Graph-based Author Disambiguation on DBLP Data in Neo4j. *arXiv preprint arXiv:1901.08977*.
- Freire, V. P. & Figueiredo, D. R. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, 17(4):255–266.
- Furtado, C. A.; Davis Jr, C. A.; Gonçalves, M. A. & de Almeida, J. M. (2015). A Spatiotemporal Analysis of Brazilian Science from the Perspective of Researchers' Career Trajectories. *PLOS One*, 10(10):e0141528.
- Gabardo, A. C. (2015). *Análise de redes sociais: uma visão computacional*. Novatec Editora, São Paulo, Brasil.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.
- Hatala, J.-P. (2006). Social network analysis in human resource development: A new methodology. *Human Resource Development Review*, 5(1):45–71.

- Havel, V. (1955). A remark on the existence of finite graphs. *Casopis Pest. Mat.*, 80:477–480.
- Hua, G. & Haughton, D. (2012). A network analysis of an online expertise sharing community. *Social Network Analysis and Mining*, 2(4):291–303.
- Huang, M.-H. & Chang, Y.-W. (2011). A study of interdisciplinarity in information science: Using direct citation and co-authorship analysis. *Journal of Information Science*, 37(4):369–378.
- Iglič, H.; Doreian, P.; Kronegger, L. & Ferligoj, A. (2017). With whom do researchers collaborate and why? *Scientometrics*, 112(1):153–174.
- Kato, M. & Ando, A. (2013). The relationship between research performance and international collaboration in chemistry. *Scientometrics*, 97(3):535–553.
- Kim, J. (2018). Evaluating author name disambiguation for digital libraries: a case of DBLP. *Scientometrics*, 116(3):1867–1886.
- Kleinberg, J. & Easley, D. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, Cambridge, UK.
- Lane, J. (2010). Let’s make science metrics more scientific. *Nature*, 464(7288):488–489.
- Leão, J. C.; Brandão, M. A.; de Melo, P. O. V. & Laender, A. H. F. (2017). Mineração de Perfis Sociais em Redes Temporais. In *Anais do XXXII Simpósio Brasileiro de Banco de Dados - Short Papers*, pp. 264–269, Uberlândia, MG, Brasil.
- Leydesdorff, L. & Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for information Science and Technology*, 61(8):1622–1634.
- Li, Q.; Brown, J. B.; Huang, H.; Bickel, P. J. et al. (2011). Measuring reproducibility of high-throughput experiments. *The annals of applied statistics*, 5(3):1752–1779.
- Lima, H.; Silva, T. H. P.; Moro, M. M.; Santos, R. L. T.; Meira Jr, W. & Laender, A. H. F. (2013). Aggregating Productivity Indices for Ranking Researchers Across Multiple Areas. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 97–106, Indianapolis, IN.
- Lopes, G. R.; da Silva, R.; Moro, M. M. & de Oliveira, J. P. M. (2012). Scientific Collaboration in Research Networks: A Quantification Method by Using Gini Coefficient. *International Journal on Computational Science & Applications*, 9(2):15–31.

- Maia, G.; de Melo, P. O. V.; Guidoni, D. L.; Souza, F. S.; Silva, T. H.; Almeida, J. M. & Loureiro, A. A. (2013). On the analysis of the collaboration network of the brazilian symposium on computer networks and distributed systems. *Journal of the Brazilian Computer Society*, 19(3):361.
- Mena-Chalco, J. P. & Cesar Junior, R. M. (2009). ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, 15(4):31–39.
- Mena-Chalco, J. P.; Digiampietri, L. A. & Cesar-Jr, R. M. (2012). Caracterizando as redes de coautoria de currículos Lattes. In *Proceedings of the I Brazilian Workshop on Social Network Analysis and Mining*, Curitiba, PR, Brasil.
- Mena-Chalco, J. P.; Digiampietri, L. A.; Lopes, F. M. & Cesar, R. M. (2014). Brazilian Bibliometric Coauthorship Networks. *Journal of the Association for Information Science and Technology*, 65(7):1424–1445.
- Mooney, H. A.; Duraiappah, A. & Larigauderie, A. (2013). Evolution of natural and social science interactions in global change research programs. *Proceedings of the National Academy of Sciences*, 110(Supplement 1):3665–3672.
- Moreira, C.; Calado, P. & Martins, B. (2015). Learning to rank academic experts in the DBLP dataset. *Expert Systems*, 32(4):477–493.
- Peet, R. K. (1975). Relative diversity indices. *Ecology*, 56(2):496–498.
- Pessoa Jr., G. J.; Dias, T. M. R.; Silva, T. H. R. & Laender, A. H. F. (2018). Importância das Colaborações Interdisciplinares nas Redes de Coautoria Científica. In *Proceedings of the VII Brazilian Workshop on Social Network Analysis and Mining*, Natal, RN, Brasil.
- Porter, A. L.; Roessner, D. J. & Heberger, A. E. (2008). How interdisciplinary is a given body of research? *Research evaluation*, 17(4):273–282.
- Santana, A. F.; Gonçalves, M. A.; Laender, A. H. F. & Ferreira, A. A. (2017). Incremental author name disambiguation by exploiting domain-specific heuristics. *Journal of the Association for Information Science and Technology*, 68(4):931–945.
- Scott, J. (2017). *Social network analysis*. Sage Publications, London.
- Shi, S.; Zhang, W.; Zhang, S. & Chen, J. (2018). Does prestige dimension influence the interdisciplinary performance of scientific entities in knowledge flow? Evidence from the e-government field. *Scientometrics*, 17(2):1237–1264.

- Silva, T. H. P.; Laender, A. H. F.; Davis, C. A.; da Silva, A. P. C. & Moro, M. M. (2017). A profile analysis of the top Brazilian Computer Science graduate programs. *Scientometrics*, 113(1):237–255.
- Silva, T. H. P.; Laender, A. H. F.; Davis Jr, C. A.; da Silva, A. P. C. & Moro, M. M. (2016). The Impact of Academic Mobility on the Quality of Graduate Programs. *D-Lib Magazine*, 22(9/10).
- Silva, T. H. P.; Moro, M. M. & Silva, A. P. C. (2015). Authorship Contribution Dynamics on Publication Venues in Computer Science: an Aggregated Quality Analysis. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pp. 1142–1147, Salamanca, Spain.
- Silva, T. H. P.; Moro, M. M.; Silva, A. P. C.; Meira, W. & Laender, A. H. F. (2014). Community-based Endogamy as an Influence Indicator. In *Proceedings of the IEEE/ACM Joint Conference on Digital Libraries*, pp. 67–76, London, UK.
- Sonnenwald, D. H. (2007). Scientific collaboration. *Annual review of information science and technology*, 41(1):643–681.
- Ströele, V.; Zimbrão, G. & Souza, J. M. (2012). Análise de Redes Sociais Científicas: Modelagem Multi-relacional. In *Proceedings of the I Brazilian Workshop on Social Network Analysis and Mining*, Curitiba, PR, Brasil.
- Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825):268.